

**Request for Expression of Interest (EOI)**  
**for**  
**Procurement of Integrated Solution for Data Lake and Data Warehouse**  
**Ref: SBI/GITC/ Data Warehouse/2023/2024/48 dated 08.12.2023**

**Bank's responses to pre-bid queries (received upto 05.01.2024)**

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
1	42 of 106	Compliance Matrix	End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics maximum within 3 hours from day end for batch data loads on all days including month-end / quarter-end / year-end. Vendor may propose suitable solution to achieve the objective. Refer Annexure – C for existing data flow in DWH	We request you to share the Annexure – C, D, E, F, G mentioned in the EOI document.	Architecture diagram has been presented in pre-bid meeting
2	52 of 106	Compliance Matrix - Disaster Recovery	Vendor should provide a feasible plan for best use of existing infrastructure which is procured during last 10 years in staggered manner during the implementation of ISDLDW which will save cost to the Bank.	Please could you share the details of all the infrastructure that has been procured by the Bank in the last 10 years that is expected to be leverage for the new solution.	Existing infrastructure details has been presented in pre-bid meeting.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
3	67 of 106	Compliance Matrix - Data Encryption	Data available over an API to external sources / applications	Does the Bank already have an API infrastructure or the vendor is expected to propose the API infrastructure as part of the solution ?	Vendor is expected to propose the API infrastructure as part of the solution.
4	35 of 106	Volumetric of Existing Data Warehouse:	Production DB Size	The EOI mentions Production DB data volume mentioned as ~1 Peta Byte. Is this data compressed or not?	Data is compressed (compression factor is approx 4)
5	63 of 106	Appendix-C: Technical Criteria/Scope of Work C4. Functional Requirements Security and Compliance - Point No. 6	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	To ensure Availability under CIA triad, it is essential to deploy a dedicated On-Premise DDoS Protection. On-Premise DDoS Protection solution detects, alerts and protects against Volumetric, State Exhaustion, Application Layer and Outbound DDoS Attacks and thus ensures Availability of Critical Systems, Services and Applications in Data Center.  So, could you please confirm if you need On-Premise Anti-DDoS Protection Systems ?	DDoS protection is not required for this solution, but bidder need to integrate with proposed solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
6	63 of 106	Appendix-C: Technical Criteria/Scope of Work C4. Functional Requirements Security and Compliance - Point No. 6	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	As a part of Solutions to identify and alert on Data leakage, it is essential to deploy a packet based Network Detection and Response (NDR) Solution. A packet based NDR enables Enterprises to achieve complete security visibility into their network and identify vulnerabilities and threats with high accuracy.  So, could you please confirm if you need a Network Detection and Response Solution ?	Proposed solution should be integrated with Bank's centralized security Solutions. No additional solution is required for this purpose.
7	56 of 106	Monitoring Dashboard Point No. 7	Industry best tools for performance (slowness) monitoring	As Industry best practices the tool or solution should be agent less as it will be non-intrusive and dose not impact on performance on the Servers or the network.	Suggestion from bidder
8	56 of 106	Monitoring Dashboard Point No. 9	End to end performance monitoring with automatic alerts	End to end visibility from client to Application and the backend database visibility is possible via traffic monitoring solution with automatic alerting.	Suggestion from bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
9	57 of 106	Monitoring Dashboard Point No. 21	DB health monitoring tool as well as monitor of N/W health	Monitoring of DB application health and network health will be possible only via Application aware Network Performace monitoring (Traffic based Agent less solution) for this requirement please provide the capacity/ bandwidth for network traffic monitoirng.	Infrastructure and Health and performarncce monitoring is required. Network Monitoring is not required.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
10	16 & 24 & 28 and 29	12 & 19 & Appendix A	<p>ii. Prior to the detailed evaluation, the Bank will determine the responsiveness of each Bid to the EOI. For purposes of these Clauses, a responsive Bid is one, which conforms to all the terms and conditions of the EOI in toto, without any deviation.</p> <p>v. SBI may shortlist the Bidders who fulfill the eligibility and technical criteria specified in this EOI, have solution/services as per the requirement of the Bank and are agreeing to abide by the terms and conditions of the Bank. Bank's judgment in this regard will be final.</p> <p>&amp;</p> <p>i. Lodgement of an EOI is evidence of a Bidder's consent to comply with the terms and condition of Request for EOI process and subsequent bidding process. If a Bidder fails to comply with any of the terms, its EOI may be summarily rejected.</p>	<p>While xxxxx has endeavored to accept all terms and conditions without any deviation, we request the opportunity to put forward certain deviations that we expect to discuss with Bank. As such, we submit our response with the below caveat – “Subject to deviations submitted or to be submitted as part of the subsequent RFP/bidding process, we agree with EOI and subsequent terms and conditions”.</p> <p>Request your clarification and confirmation that Bidder will have a chance to propose deviations at the RFP stage that we expect will be mutually discussed and agreed at such RFP stage.</p>	<p>Deviations are not required at this EOI Stage. Discussion will be done at the RFP Stage.</p>

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
11	63 & 64	Security and Compliance & Regulatory Reporting	<p>8. Compliance to Global Standards • GDPR, BCBS239, PCIDSS, DFRA and similar relevant standards • Office of Foreign Assets Control (OFAC) • Financial Crimes Enforcement Network (FinCEN) • Securities and Exchange Commission (SEC) • Office of the Comptroller of the Currency (OCC) etc.</p> <p>&amp;</p> <p>Vendor should follow the RBI guideline in developing the solution with which it will be easier for the Bank to migrate to the element-based data reporting envisaged by the RBI.</p>	<p>In case the Bank requires Bidder or the Services to comply with any other laws, rules, regulations, guidelines, or regulatory requirements applicable to the Bank, it shall be the Bank's responsibility to clarify and provide to Bidder those other laws, rules, regulations, guidelines, and regulatory requirements in writing to Bidder before commencement of services or under RFP as part of the scope, as Bidder is not in the business of providing legal advisory services to identify and ensure compliance or advise the Bank in this regard.</p> <p>Request Bank to confirm the above understanding.</p> <p>Further, request the Bank to clarify whether Bidder will have access to personal data of EU citizens' personal data as solution is expected to comply with GDPR.</p>	<p>Bidder need to comply with all the guidelines/laws/instructions issued by Bank/Regulators/Govt. from time to time.</p>
12	64	Security and Compliance	10. Data privacy compliance	Request Bank to elaborate on this requirement.	Data privacy compliance will be provided in the RFP stage.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
13	34 of 106	C2.1 Goals and Objectives	d)Optimization of solution / tool (s) for Data Federation/Virtualization to ensure seamless integration of data in real time when stored in multiple sources without physical movement of data sets for the purpose of reporting / analytics. Data federation should be enhanced with more intelligent real- time query optimization, caching, in-memory and hybrid strategies that are automatically (or manually) chosen based on source constraints, application need and network awareness	Data Federation: Do we still need Data Virtualization/Federation if all the analytics needs can be served with a single platform covering Data Lake, Data warehouse, Operation DB, and Object storage?	Yes it will be required.
14	35 of 106	Volumetric of Existing Data Warehouse:	Production DB Size	Can we get the architecture and HW configuration for the current setup, no of servers and core count, memory count, etc?	Architecture diagram has been presented in pre-bid meeting
15	42 of 106	5	Solution proposed for Data Ingestion should be compatible to ingest and replicate data simultaneously on all available regions.	Can you please list the total number of regions in production and DR site?	One region each for Dev, UAT, PR and DR sites. Proposed Sites are: PR: Navi Mumbai DR: Hyderabad

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
16	34	Appendix - C - C2.1.f	Implementable solution for availability of DR	Understanding : DC in Navi Mumbai (Location 2) & DR is in Hyderabad with RTO as 2 Hrs. What is banks RPO requirement. Are we looking for high availability for every component in the architecture including ingestion, data storage / DBMS & Analytics etc. Is the HA configuration required at both sites DC and DR?	RPO should be 0. Bidder should ensure high availability (HA) both sites.
17	38	Appendix C C4, Functional Requirements - pt 23	Governance: The ability to forecast, budget usage, monitor and control costs by throttling, workload or user prioritization or other means.	What is the expectation from the solution for this points? “Governance: The ability to forecast, budget usage, monitor and control costs by throttling, workload or user prioritization or other means.”	From the governance point of view, the tool should have the capability to forecast and suggest the workload for the model , monitor the usage and enable threshold level prioritization.
18	38	Appendix C C4, Functional Requirements - pt 31	Platform should support running models on both CPU and GPU. GPU to be provisioned accordingly.	Are we talking about GPU instance for on-premise solution?	Yes
19	52	Disaster Recovery - pt 1	Bank proposes to setup replication based DR capable of replication of select tables, schema and jobs.	Are we talking about complete site replication (App + database) or object replication?	Bank is expecting complete site replication for DR.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
20	35	C3. Current Components - C3.2, pt 2	Archival DB Size	1. Where archival data is available like Online/Offline/Tape/NFS 2. Is archival database compressed?	Online Yes. Compressed
21	35	C2. Project Mandate - C2.1, K	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	What is the ratio of dev/uat/DR/Sandbox with respect to Production	Bidder to propose ratio for dev/uat/sandbox as per their solution. DR should be exact replica of PR
22	84	Hardware Specifications - Pt 18	The proposed hardware is mission critical for the proposed project and support of 24 X 7 with an uptime of 99.99 % to be ensured by providing support at PR, and DR site for a period of 5 years.	What would be the DB growth for Prod/Archive/Data Lake for Next 5 year	It would be refined in RFP phase but expecting volume of 10 PB.
23	43	Data Ingestion - Pt 13	A source system table with ~ 300 columns in an Oracle database generates data of ~ 800 GB on a daily basis	As per Table C3.2 monthly data ingestion ~20 TB, but with given clause 800 GB daily become ~24 TB Per Month, which one is correct	These are approximate values (20-24 TB). Please refer to the following clause: A source system table with ~ 300 columns in an Oracle database generates data of ~ 800 GB on a daily basis with ~ 400 columns in DWH. 100 columns in DWH are <b>compressed</b> in a single column at source in Oracle.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
24	53	Data Archival and Backup - Pt 3	All the applications connected to the non-archived data should be available with archived as well	Understanding is non-archived prod data should be online. Pls clarify if data which are archived also needs to be available instantly	Archived data should be available instantly for reporting / querying historic data without moving to main storage.
25	37	16	Analytics on Streaming Data: Analytical capabilities specifically tied to stream data, including windowing analytics, analytics on in-flight data, joins to static data and support for streaming standards, among others.	Please provide the list of Streaming Data sources and the approximate frequency at which events are generated from each source.	At present streaming date is not being sourced at DWH level. Bank is expecting this capability from each digital channels.
26	37	19	Operational Intelligence: The ability to handle large numbers of concurrent users running fixed short analytical queries while still providing appropriate performance SLAs for operational workloads.	Please clarify what is the maximum number of concurrent user queries/interactions expected so that the solution can be sized to meet desired SLAs.	Please refer to sr no 11 of page no 46 and 13 of page no 48.
27	40	6	Note: All components must be based on open standard/open source software, customized for enterprise use (Open Core)	Please clarify, does it mean that only open source platform can be proposed.	Bank is open for OSS components subject to bidder will provide complete support (compliance with Bank's security requirements)

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
28	33	C1	With the implementation of project, DWH has been shifted from PDOA (Pure data for Operation Analytics) technology to IIAS (IBM Integrated Analytical System) technology, along with a small-scale Data Lake and Data Virtualization capabilities. Since our present Data Warehouse system is on-prem, it has limited scope of scalability and it requires complex forecasting to determine how to scale the data warehouse for future needs	What are your top challenges for this initiative , in terms of performance, flexibility etc?	existing project is approaching end of life.
29	33	C1	With the implementation of project, DWH has been shifted from PDOA (Pure data for Operation Analytics) technology to IIAS (IBM Integrated Analytical System) technology, along with a small-scale Data Lake and Data Virtualization capabilities. Since our present Data Warehouse system is on-prem, it has limited scope of scalability and it requires complex forecasting to determine how to scale the data warehouse for future needs	Are there any known dependencies or gaps that we should be aware of ?	existing project is approaching end of life.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
30	33	C2	With the implementation of project, DWH has been shifted from PDOA (Pure data for Operation Analytics) technology to IIAS (IBM Integrated Analytical System) technology, along with a small-scale Data Lake and Data Virtualization capabilities. Since our present Data Warehouse system is on-prem, it has limited scope of scalability and it requires complex forecasting to determine how to scale the data warehouse for future needs	What level of documentation current exists today for system study / knowledge acquisition phase ? (Architecture related, system related, current Hardware and software specs, Integration details, API documentation, etc. )	All required documents are present.
31	34	C1	Capabilities of data ingestion from all source system available with the Bank	Can you provide a list of source systems and their respective technologies and how these currently get integrated with data warehouse/data lake	Architecture diagram & Data ingestion has been presented in pre-bid meeting
32	41	C4	Data may be structured, semi-structured, and unstructured.	Can you please provide the % of structured vs unstructured data in terms of storage	current structure is ~1/4 th of total storage is object storage
33	35	C3.2	Data lake - ~730 TB	Is the data lake currently used as a storage of raw data before it is loaded into the datawarehouse. Does the 730 TB of data is also available in the datawarehouse in structure form	In Current architecture, most of the data present is data lake is sourced from data warehouse.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
34	36	C3.2	Total ETL Jobs ~27000, Scheduled daily jobs ~5000	Can you provide the approximate of number of data ingestion jobs, streaming jobs in the daily jobs. Considering the daily jobs is ~5000 can you provide a high level break-up of the remaining ~22000 ETL Jobs.	Approximate no of jobs are: data ingestion jobs: ~4000 Streaming jobs: NIL Other jobs :
35	36	C4	C4, #2 - Enables different teams to use unified platform across all the enterprise data for a range of products including data science, machine learning and business intelligence.	Can you provide the various user personas who will be using the analytics from this data warehouse and how will they be delivered with the analytics (direct through BI or through embedded analytics)	Users from IT and business are expected to access to the proposed solution as per their requirements using all possible ways.
36	36	C4	C4. Functional Requirements #8 - Integrated Business Intelligence solution	Can you please explain more about the functionality of -"Integrated Business Intelligence solution."	Solution should serve requirements of BI.
37	46	C4	C4 #7 Processing engines should have capabilities to process all kinds of Data Storage formats used by the Bank.	Can you provide the data storage formats used by the bank currently.	All prevailing formats available in the market are currently being used.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
38	62	C4	Security & Compliance #1 Authentication and Identity Management - A comprehensive identity and access management system should be available for centralized management of users and groups. It should be possible to quickly create and revoke the identity of a user or a service by simply deleting or disabling the account in the directory. Multi-factor authentication is desired as an additional layer of security for user sign-in and transactions.	Can you provide the current Centralized Identify & access management system used across the organization. Is the requirement to integrate with the Centralized IAM system	Bank is currently using Microsoft ADS. Yes.
39	71	C4	Implementing end to end analytics use-cases as mandated by the Bank	Which use cases are the most critical and need immediate implementation? This will help prioritize	Tool should have the capability to implement end to end analytics use cases.
40	71	C4	Power data / objects to existing analytics models built on proprietary tools. Migration of such models to new solution	Could you share the specific technologies and frameworks used for existing models (excluding IBM SPSS Statistics) to understand the migration feasibility	IBM SPSS Modeler, Open source -Python, SPARK and R are used at present. For data pipelines, ETL tool - IBM datastage is used.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
41	71	C4	Availability of Pre-built models which can be directly used with Bank's data to get insights	Could you clarify the desired functionalities and performance criteria for pre-built models. Are these models required for structured data only or unstructured and semi-structured data as well	Pre-built models are required for all types of data including GenAI foundation models.
42	72	C4	Templates for common data processing / end to end applications, tag data sets and re-use operators across machine learning pipelines	Could you describe the specific types of data processing and application templates needed, and how they should facilitate reusability across pipelines.	Tool should have the capability to process all types of data and workflow template for MLOps which can be reused across different projects.
43	72	C4	Analytics on real-time data in real-time	Could you provide more clarity on the latency requirements and data sources for real-time analytics use cases.	Details will be provided in RFP phase
44	72	C4	The ability to automate the process of iteratively searching for the best model from a set of candidates. This feature has also been called "model factory." Platform provide this functionality	Would it be possible to add more details around the specific functionalities that are expected from the "model factory", for example features like automatic hyperparameter tuning and model selection criteria?	Details will be provided in RFP phase

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
45	74	C4	Pipelines to orchestrate the entire machine learning process from data preparation to deployment	Can it be assumed that the preferred deployment option would be cloud and not on-premise. What are the security requirements expected for models and workflows?	Bank is looking for scalable solution that will be on-prem or Bank's own Meghdoot cloud. However, solution to be scalable to public cloud if require in future (minimize effort and impact to move to cloud in future). Suitable architecture to be proposed by bidder
46	74	C4	DevOps - ability to automate build and deploy on all elements of the data science project, from the pipelines to building the models to model deployment.	Could you specify the level of detail and granularity required for monitoring the model performance.	Bidder to suggest possible options of this requirement. Details will be provided in the RFP
47	58	C4	Data Governance - #10 Provide detailed information about interfacing (upstream and downstream) applications.	Can you provide the list of upstream and downstream applications.	will be shared with finally selected bidder



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
48	42	C4	End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics maximum within 3 hours from day end for batch data loads on all days including month-end / quarter-end / year-end. Vendor may propose suitable solution to achieve the objective. Refer Annexure – C for existing data flow in DWH	Annexure C- existing data flow in DWH - Can you provide the annexure listed	architectue has been presented during pre-bid meeting.
49	51	C4	Migration of All the remaining components of existing ecosystem (Mentioned in Annexure - D) as and when identified by Bank like job scheduler, reports, history of version control, existing tape backup, etc.	Can you please provide the details of the existing data eco-system components including technology components, job scheduler	architectue has been presented during pre-bid meeting.
50	63	C4	Compliance with SBI IS Policies – some of the key areas are as under (Summary of Key policies is given in Annexure F):	Can you please provide the key IS policies of SBI	Details will be presented to selected bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
51	73	C4	Annexure E gives sample use cases which are to be implemented on Next Gen Data Warehouse using structured and/or unstructured and/or semi-structured and/or any other kind of data gathered from either Data Warehouse or Data Lake or Data Virtualization or all together or any other source.	Can you provide details of the use cases that will be need to be implemented in the Next Gen Data Warehouse	Existing models to be migrated and new models will be developed by the Bank In-house team
52	52	C4	Vendor should provide a feasible plan for best use of existing infrastructure which is procured during last 10 years in staggered manner during the implementation of ISDLDW which will save cost to the Bank. (Annexure D gives the technology architecture of the current setup)	Can you provide the details on the existing technology architecture	Details will be presented to selected bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
53	89	C4	<p>Annexure G gives the ISDLDW sizing forecasted for next five years. Sizing given in the Annexure G is in compressed form with compression factor of 2.5. Vendors are expected to propose storage forecast over and above given sizing in the solution to ensure fast performance of system. Vendors shall propose performance benchmark for this system as part of the solution. This is indicated projected growth, but the solution should be scalable enough to store any volume of data</p>	<p>Can you provide the ISDLDW sizing forecasted over the next five years</p>	<p>Sizing detail will be provided in RFP</p>

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
54	42	C4	End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics maximum within 3 hours from day end for batch data loads on all days including month-end / quarter-end / year-end. Vendor may propose suitable solution to achieve the objective. Refer Annexure – C for existing data flow in DWH	Can you provide details of the existing data flow in the DWH	Architectue has been presented during pre-bid meeting.
55	41	C4	Present Architecture of existing Data Warehouse and technology stack will be shared with the intended bidders who participate in this EOI proposal on execution of Non-Disclosure Agreement (NDA). Copy of NDA will be shared in due course.	Can you provide the current data warehouse architecture and technology stack	Architectue has been presented during pre-bid meeting.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
56	50	C4	<p>The workflows should work with standard schedulers. Monitoring and management of workflows should be possible from an easy to use interface. Workflow management tool(s) should have connectors / pluggable interfaces to already existing / in-use proprietary software available with the Bank. These could be (and not restricted to) data repositories,</p> <p>Expression of Interest for procurement of integrated solution for Data Lake and Data Warehouse</p> <p>Page 50 of 106</p> <p>reporting tools, data analysis tools and generic interfaces for data transfer. Scheduled jobs status should be made available to the Bank in Monitoring dashboard on real time basis</p>	Can you please provide the existing/in-use proprietary workflow monitoring softwares available with the bank	Architectue has been presented during pre-bid meeting.
57	51	C4	Data migration from existing archival solution to new one.	Can you provide details on the existing archival solution	Architectue has been presented during pre-bid meeting.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
58	53	C4	User jobs will be categorized into Critical/Normal priorities as decided by the Bank so that jobs can be moved into DR based on assigned priorities.	Can you provide the % of jobs that can be categorized as Critical vs Normal	will be shared with finally selected bidder
59	58	C4	Automated propagation of changes to ISDLDW ecosystem Data Dictionary and business glossary by multiple sources as and when changes occur in source.	Can you provide the tool, architecture of the current data dictionary, business glossary. Can it be re-used in the new ISDLDW setup.	Bank is proposing DR to be exact replica of PR
60	61	C4	Dedicated Data Quality team who will thoroughly verify the Data Quality across setup	Is the expectation here that the vendor will provide a data quality team to monitor the data quality metrics and take the necessary action like following up with source systems for correction	Bidder should propose DQ tool and team to work in tandem with Bank DQ team.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
61	62	C4	Developing automated reconciliation methodology suitable for each source system depending on recon implementation supported by source systems. Reconciliation should be automated such that, while loading data from source data should be reconciled. Reconciliation reports should be displayed real time for monitoring. Any discrepancies in reconciliation should be notified through alerts.	Can you provide the details on the recon implementation supported by the source systems.	Recon details are being shared by source systems.
62	63	C4	Data Democratization - Secure access of PROD Database to LHOs/GOCs	Is the requirement here to give direct DB access to the databases for SQL quering	Yes
63	64	c4	Automation –Tool should automate analytics and reporting workflow end-to-end, including all data collection, enrichment, and management, as well as all calculations, processes to final report submission. Currently 500+ jobs are being used for Tranche 1 DCT generation along with 500 more for other regulatory reports/returns.	Can you provide more information about the Tranche 1 DCT Generation	details will be shared with finally sected bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
64	64	C4	Change Management –Tool for handling ongoing change in regulation or business requirements without the need for programming expertise.	Does the bank use any existing Change Management tool which can be reused for managing ISDLDW change requests.	ITSM tool is used for change management. RTC tool is being used for version control. Bidder can propose any suitable better solution / tool in their proposal.
65	65	c4	Electronic Submission – Should support for all regulators globally in all required formats, including XBRL, XML or other file-based electronic submission.	Please help us understand what are the other global regulators being referred here apart from RBI / IBA ? Our understanding is the format of the reports will be provided by the Bank. Kindly confirm	Yes. Data format will be provided by the Bank to selected bidder.
66	66	C4	Pre-submission Review – Multiple report writers should allow users to review reports in various formats before submission, with the ability to drill down and make manual adjustments where necessary.	Does this manual adjustment required in the data warehouse or will this be done in the source systems only.	Details are decided case to case basis.
67	66	c4	Security and Control – Security and Compliance framework should provide a completely controlled and secure environment by providing appropriate permission to view granular data to designated users.	Is this requirement to provide controls/framework for direct SQL access to the data to designated users	Yes.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
68	67	c4	Proposed Solution should be capable on ingesting encrypted data from source system. It should support the encryption/decryption mechanism implemented at source system.	Can you provide the various encryption/decryption mechanism available at source systems	Bidder need to propose best suitable industry practise encryption techniques.
69	69	c4	Privileges granted to the user should not conflict with Access Management - Segregation of Duties policy of the Bank.	Can you provide the Segregation of Duties policy of the bank	will be shared with finally selected bidder
70	70	C4	Capability to easily connect with any downstream applications currently used at the Bank and be flexible to support all end users in the Bank based on future needs.	Can you provide integrations used by downstream systems	Architectue has been presented during pre-bid meeting.
71	80	C4	There should be separate criteria for BI user versus administration training. Training may include in-person classes, online classes -- live or prerecorded -- or web recordings for specific features or processes.	Is the expectation that the vendor has to provide trainings to the users and administrators.	bidder is expected to provide KT for use fo ETL and BI tools proposed in the solution.
72	104	C4	Appendix F - Value of Work Order (In Lakh) (only single work order)	We typically do not provide value of work order as it is confidential information with our clients. Can you please let us know this value can be omitted	This clause is applicable to all the bidders interested in bidding

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
73	40	C4	All components must be based on open standard/open source software, customized for enterprise use (Open Core)	Can you confirm if you are looking at enterprise paid versions of open source products like Confluent for Kafka, Cloudera for HDFS	Bank is open for OSS components subject to bidder will provide complete support (compliance with Bank's security requirements)
74	39	C4	Architectural Requirements	Is SBI open to host the proposed platforms in some other cloud apart from SBI's own cloud? Does SBI has any preferred cloud hosting center?	Please refer to sr no 'e' of page no 34 of the EOI document
75	39	C4	Architectural Requirements	Is SBI open for any integration between SBI owned cloud and any other cloud platform?	Please refer to sr no 'e' of page no 34 of the EOI document
76	57	C4	Metadata Management Capability: Tool should cater to three broad categories of metadata; Business metadata, Technical metadata and Operational metadata	Does your current landscape has MDM solutions ? What is the tech stack ? Does the solution requires MDM solution ?	No.
77	71	C4	Implementing end to end analytics use-cases as mandated by the Bank	Does your current landscape has Gen AI solutions ? What is the tech stack ? Does the solution requires GenAI solution ?	Bidder need to propose the best suitable solution for SBI. Since EOI is an ideation process, RFP will be released with final requirement.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
78	80	C4	Software and Solution proposed by vendors should be compatible with all types of Hardware. The proposed solution envisages use of commodity hardware, and if any proprietary components are used, should be listed in the response with details and justification.	Is SBI open to having software defined hardware (appliance driven software) like Dell Powerscale/ObjectScale storage	Bidder can propose the best suit of software and hardwares suitable for SBI environment.
79	40	C4	All components must be based on open standard/open source software, customized for enterprise use (Open Core)	Is SBI open to open source tools where open core is not available?	Bank is open for OSS components subject to bidder will provide complete support (compliance with Bank's security requirements)
80	34	C2.1	Data Federation	Do we still need Data Virtualization/Federation if all the analytics needs can be served with a single platform covering Data Lake, Data warehouse, Operation DB, and Object storage ?	Yes it will be required.
81	35	C3.2	Volumetric of Existing Data Warehouse	The data size given for Production DB is compressed or raw. Can we get the raw data size? Similarly, we would need the same for Archival and Data Lake	Compressed.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
82	35	C3.2	Volumetric of Existing Data Warehouse	Can we get the HW configuration for the current setup, no of servers and core count, memory count, etc?	Details have been presented in pre-bid meeting.
83	36	C4	Bank is planning to upgrade existing Data Warehouse and Data Lake solution to a modern and state of art integrated solution	Functional Requirement: It's mentioned Bank is planning to upgrade the existing Data Warehouse and Data Lake solution to a modern and state-of-the-art integrated solution. In addition to project mandates mentioned in section C3, the solution should provide the following additional capabilities: Instead of upgrading, Is the bank open to replacing the existing setup completely with modern capabilities built into it?	Yes. Bank is open for replacement of the existing setup. Bidder can propose suitable solution in their EOI proposal.
84	43	C4	The proposed solution should be able to scrap encrypted log	What does this mean, please elaborate.	Change Data Capture (CDC) should be able to scrap encrypted logs as well from TDE enabled database.
85	46	C4	Processing engines should have capabilities to process all kinds of Data Storage formats used by the Bank	. Can you please elaborate on the Data formats used.	All prevailing formats available in the market should be supported.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
86	53	C4	Data Archival solution should not be visible to end user, but Archived data should be available for all end users. For end user it should be a single view with Data Federation/Virtualization Layer	It is asked that Archived data should be visible to Virtualized layer. What will be format of archived data. considering ieverything is open sourc can it be archived in parquet format?	Yes it will be required.
87	50	C4	Semantic integration of structured & unstructured Data.	It is being asked that " Semantic integration of structured & unstructured Data.", please suggest some unstructured data type here. Can we assume unstructured data means csv/json/excel files.	Unstructure data may captured through streaming or clickstream sources.
88	50	C4	All data should within eco-system and across other system whenever possible be made discoverable and integrable easily through a single virtual layer which will expose redundancy and quality issues faster.	It says "All data should within eco-system and across other system whenever possible be made discoverable", However it should be that all connected database sources and datalake sources in virtulization tool should be made discoverable/ querable.	all connected database sources and datalake sources in virtulization tool should be made discoverable/ querable
89	34	C2	the proposed solution should be implementable in Physical hardware / Bank's own cloud (Meghdoot)	We understand the entire IT infrastructure for the project scope would be implemented in SBI's datacenter / cloud (Meghdoot). Please confirm	Please refer to sr no 'e' of page no 34 of the EOI document
90	34	C2	the proposed solution should be implementable in Physical hardware / Bank's own cloud (Meghdoot)	Please confirm where the SBI datacenter / cloud is located	PR: Navi Mumbai DR: Hyderabad

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
91	34	C2	the proposed solution should be implementable in Physical hardware / Bank's own cloud (Meghdoot)	Please confirm if SBI would provide the required rack space for hosting the IT infra at their datacenter	If deployment is on Banks' own private cloud (ie. Meghdoot), Bank will provide the required infrastructure.
92	34	C2	the proposed solution should be implementable in Physical hardware / Bank's own cloud (Meghdoot)	Please confirm if SBI has the required network (LAN and WAN) infrastructure with adequate capacity already in place or that needs to be deployed.	Bank is having enterprise wide network infrastructure in place. However, Bidder need to mention the infrastructure capacities required as per scope of the requirement.
93	34	C2	the proposed solution should be implementable in Physical hardware / Bank's own cloud (Meghdoot)	Please confirm if the IT infra for this scope would be part of SBI ADFS / domain	Yes.
94	34	C2	the proposed solution should be implementable in Physical hardware / Bank's own cloud (Meghdoot)	Please confirm if SBI has sufficient internet BW at their datacenter for accessing the applications or internet needs to be provisioned.	Bidder can propose their bandwidth requirements in the proposal
95	37	C4	Operational Intelligence: The ability to handle large numbers of concurrent users running fixed short analytical queries while still providing appropriate performance SLAs for operational workloads.	Do you have any current SLAs in place ? Please share details	Yes. SLA is in place which is ending on Feb 2025
96	85	C4	Vendor needs to provide Helpdesk support 24X7 to the Bank for end to end support for hardware maintenance	Do you share the current incident management flow diagram with the current Helpdesk team and the respective resolver groups for reference?	details will be shared with finally selected bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
97	63	C4	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	What are the external 3rd party systems outside of SBI environment accessing the DWH ? Please share what kind of data is accessed (Is there any SPI/ PII/ etc. )	at present, existing DWH is not accessible to outside entities.
98	64	C4	Compliance of Bank's Information security policy, Data governance policy.	Can you provide the Bank's IS and Data Governance policy	will be presented to finally selected bidder
99	64	C4	Data privacy compliance	Can you provide the Bank's privacy policy	will be presented to finally selected bidder
100	30	Point No.3	The Bidder must have an average turnover of minimum Rs. 1275 crore during last 03 (three) financial year(s) (2022-23, 2021-22, 2020-21)	The turnover required for the project is extremely high which shall restrict a fair competition in the bidding process, request the Authority to consider:  The Bidder must have an average turnover of minimum Rs. 25 crore during last 03 (three) financial year(s) (2022-23, 2021-22, 2020-21)	Criteria mentioned in the EOI are final

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
101	31	Point No.5	Bidder should have experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake preferably in BFSI sector.	We request the Authority to confirm that any bidder who do not possess the specific experience in BFSI sector but have experience of carrying out projects in large enterprises handling large volumes of data shall be considered during evaluation.	Preference will be given to bidders having experience in the BFSI sector.
102	30	Point 3 Appendix-B: Bidder's Eligibility Criteria	The Bidder must have an average turnover of minimum Rs. 1275 crore during last 03 (three) financial year(s) (2022-23, 2021-22, 2020-21)	We humbly requesting bank to relax this criteria considering GOI guideline for MSEs. <b>{Ministry of Micro, Small &amp; Medium Enterprises (MSMEs) vide Policy Circular No. 1(2)(1)/2016-MA dated 10" March, 2016 has clarified that all Central Ministries / Departments / Central Public Sector Undertakings (CPSUs) may relax condition of prior turnover and prior experience with respect of Micra &amp; Small Enterprises (MSEs) in all public procurements subject to meeting of quality and technical specifications.}</b>	Criteria mentioned in the EOI are final



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
103	31	Point 5 Appendix-B: Bidder's Eligibility Criteria	Bidder should have experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake preferably in BFSI sector.	We humbly requesting bank to relax this criteria for MSEs towards data-scale.	Criteria mentioned in the EOI are final
104	31	Point 5 Appendix-B: Bidder's Eligibility Criteria	Client references and contact details (email/ landline/ mobile) of customers for whom the Bidder has executed similar projects in India. (Start and End Date of the Project to be mentioned) in the past <b>(At least two client references are required)</b>	We humbly requesting bank to kindly lower this criteria to One client references for MSEs.	Criteria mentioned in the EOI are final
105	-	General	Last Date & Time for Bid submission: 10.01.2024 up to 5 PM	We humbly request bank to extend the bid submission for atleast 10-15 days after the pre-bid clarification is shared.	Bid submission date has been extended till 29.01.2024

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
106	8	8	Data Warehouse has migrated on new infrastructure on 10th May 2023 as per fiveyear Agreement between Bank and M/s IBM India Private Limited in February 2020 for implementation and maintenance of the Data Warehouse. With the implementation of project, DWH has been shifted from PDOA (Pure data for Operation Analytics) technology to IIAS (IBM Integrated Analytical System) technology, along with a small-scale Data Lake and Data Virtualization capabilities. Data Lake is implemented with a view to enrich the capabilities of sourcing unstructured and semi-structured data for advance analytics with AI / ML tools.	We would like to seek clarity whether the proposed solution will replace the existing infra or bank is going to use it in addition to it. Also will there be any integration with existing infra?	Proposed solution will replace existing infra. However, existing and new systems will work parallely initially for few months.
107	10	10	If any Bidder submits Bid on behalf of Principal/OEM, the same Bidder shall not submit a Bid on behalf of another Principal/OEM under the EOI. Bid submitted with options of multiple OEMs shall also be considered Bid submitted on behalf of multiple OEMs.	Please clarify if bank is assuming Principal / OEM as in Datawarehouse software OEM and not the infrastructure OEM.	Bank is open for all possibilities (infra and software OEMs could be same or different).

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
108	11	11	Either the Bidder on behalf of Principal/OEM or Principal/OEM itself is allowed to Bid, however both cannot Bid simultaneously.	Can a hardware OEM bid with multiple System Integrators?	Yes. Hardware OEM can participate with multiple SI. In this case SI will be considered as a principle bidder
109	31	31	Client references and contact details (email/ landline/ mobile) of customers for whom the Bidder has executed similar projects in India.	Request bank to consider global references as well.	Please refer to Corrigendum-2.
110	62	62	Authentication and Identity Management - A comprehensive identity and access management system should be available for centralized management of users and groups. It should be possible to quickly create and revoke the identity of a user or a service by simply deleting or disabling the account in the directory. Multi-factor authentication is desired as an additional layer of security for user sign-in and transactions.	Does vendor has to provide IAM solution or needs to integrate with banks existing IAM tool.	Needs to integrate with Bank's existing tool (MS-AD).
111	70	70	Capability to easily connect with any downstream applications currently used at the Bank and be flexible to support all end users in the Bank based on future needs.	Request bank to provide details of such applications	Details can be shared with selected bidder post RFP.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
112	81	81	The Vendor is required to supply, install, test, commission, monitor, manage and maintain the IT System along with operating system and other peripherals with one-year warranty and AMC for 4 years from the date of delivery at data centers advised by the Bank`	We suggest banks to take minimum 36 months warranty upfront as today most of the leading hardware OEM's supply IT hardware with 36 months warranty.	Suggestion can be included in EOI proposal by Bidder
113	50	Data Federation/Virtualization	Data virtualization should support the use of APIs.	What could be the possible sources to integrate apart from APIs?	Other sources would be other storages in the whole ecosystem (archival system etc.) ; and if required in future should be able to connect to various source systems (upstreams).
114	51	Migration from Existing Setup to Proposed Solution	Migration of All the remaining components of existing ecosystem (Mentioned in Annexure - D) as and when identified by Bank like job scheduler, reports, history of version control, existing tape backup, etc.	No Annexure D provided	Architecture has been presented in pre-bid meeting
115	52	Migration from Existing Setup to Proposed Solution	Vendor should provide a feasible plan for best use of existing infrastructure which is procured during last 10 years in staggered manner during the implementation of ISDLDW which will save cost to the Bank. (Annexure D gives the technology architecture of the current setup)	No Annexure D provided	Architecture has been presented in pre-bid meeting

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
116	52	Disaster Recovery	The ISDLDW – DR solution needs to be set up at a remote location at Hyderabad.	Will the bank provide any necessary Hardware / OS and NW connectivity to deploy the disaster recovery solution prescribed by bidder ?	DR solution should be included in the proposal. However, network connectivity will be provided by the Bank.
117	52	Disaster Recovery	The DR solution should be synced with production ISDLDW. The SLA for RTO should be maximum 2Hrs as per Bank's defined policy.	Bidder would like to define the RPO and RTO in alignment with the business refreshes as per criticality, keeping Bank's best interests in mind	Suggestion can be included in EOI proposal by Bidder
118	52	Disaster Recovery	Bank proposes to setup replication based DR capable of replication of select tables, schema and jobs.	Are we talking about complete site replication (App + database) or object replication?	Complete Site Replication; any suggestions can be included in the EOI proposal by bidder.
119	53	Disaster Recovery	User jobs will be categorized into Critical/Normal priorities as decided by the Bank so that jobs can be moved into DR based on assigned priorities	Can bidder suggest the prioritisation of the activities as per the industry standards?	Suggestion can be included in EOI proposal by Bidder
120	53	Data Archival and Backup	Data older than specific duration as identified by Bank to be archived in low cost cold storage. Changing data archival rules should be easily configurable. Vendor to propose solution for the same with cheap and flexible storage and processing	Storage system procurement will be done by Vendor or bank to provide?	Depends on the solution proposed (if solution in on top of Meghdoot, Bank will provide; otherwise it should be part of proposal).

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
121	53	Data Archival and Backup	Data older than specific duration as identified by Bank to be archived in low cost cold storage. Changing data archival rules should be easily configurable. Vendor to propose solution for the same with cheap and flexible storage and processing	Are there any performance and scalability concerns that we need to keep in mind before suggesting the same> For example: what is the volume of records we are looking at when the archival happens	No considerable impact on operations while archival is done at present. This is adhoc process presently, which needs to be streamlined and automated.
122	53	Data Archival and Backup	Automated process for archiving data as per Banks policy	What is the current data archival policy for the data? What is the period of retention of the archived data?	Details can be shared with selected bidder post RFP.
123	53	Data Archival and Backup	All the applications connected to the non-archived data should be available with archived as well	Understanding is non-archived prod data should be online. Pls clarify if data which are archived also needs to be available instantly	Archived data should be available instantly for reporting / querying historic data without moving to main storage.
124	54	Cloud Integration and Migration	In view of the intent to reduce the hardware footprint (in future), the technical architecture of ISDLWD solution should be flexible to accommodate moving of ISDLWD on cloud and also should support cloud bursting for heavy workloads. The Bank understands that there can be differences in services offered by cloud service providers. The ISDLWD solution architecture should be designed considering as-is infrastructure availability in cloud.	Does this mean that the public cloud e.g. AWS, Azure, Snowflake, etc. will be used only for Infrastructure/platform scaling only and the data architecture will be reused?	Bank is looking for scalable solution that will be on-prem or Bank's own Meghdoot cloud. However, solution to be scalable to public cloud if require in future (minimize effort and impact to move to cloud in future). Suitable Data architecture to be proposed by bidder and solution should include effort for changing architecture.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
125	54	Cloud Integration and Migration	General query	What could be the possible data volume to be integrated from public cloud?	At present, no integration with public cloud. Futuristic data volume cannot be ascertained at this stage.
126	54	Data Archival and Backup	Store backup of entire ecosystem on suitable cost-effective, fast recovery infrastructure (Currently tape backup is taken)	Is there any expectation of moving the tape back up to the suggested solution? What is the period for which tape back up has been taken till date?	Yes. Now full backup is taken.
127	54	Data Archival and Backup	Archival and Backup setup must support automated Data Reconciliation whenever movement from Current/Live happens	As automated Data Recon is requested - is there a monitoring dashboard required here that publishes the status of archival at regular fixed intervals?	Yes
128	54	Cloud Integration and Migration	General query	Cloud integration required data to be fetched from cloud data lake storage or can it be cloud RDMS and Datawarehouse also?	Bank is looking for scalable solution that will be on-prem or Bank's own Meghdoot cloud. However, solution to be scalable to public cloud if require in future (minimize effort and impact to move to cloud in future). At present, no integration with public cloud.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
129	54	Cloud Integration and Migration	General query	Is there any performance requirements for cloud data integration?	Bank is looking for scalable solution that will be on-prem or Bank's own Meghdoot cloud. However, solution to be scalable to public cloud if require in future (minimize effort and impact to move to cloud in future). At present, no integration with public cloud.
130	54	Cloud Integration and Migration	General query	Data Migration from cloud would require any transformations or ETL?	Bank is looking for scalable solution that will be on-prem or Bank's own Meghdoot cloud. However, solution to be scalable to public cloud if require in future (minimize effort and impact to move to cloud in future). At present, no integration with public cloud.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
131	54	Data Archival and Backup	Store backup of entire ecosystem on suitable cost-effective, fast recovery infrastructure (Currently tape backup is taken)	Pls specify backup and recovery window, so that all bidders provide solution accordingly For such large workload (PB in size) database aware online, incremental forever and highly compressed backup are recommended. Please clarify if bidders can propose such solution to meet bank's RTO requirement.	Suggestion can be included in EOI proposal by Bidder
132	54	Data Archival and Backup	Archival and Backup setup must support automated Data Reconciliation whenever movement from Current/Live happens	Please provide more details around this point	Once data is moved to archival system, it is reconciled for data quality (e.g. Data completeness - all data is archived and not missed).
133	55	Monitoring Dashboard		Will there be a consortium of SI/Vendor partners created for procurement	Bidder should propose appropriate solution for the same
134	55	Monitoring Dashboard		There is a need to refresh data in the reporting layer w/real time data. Would this be near to real time or real time? If either, what are the data sources that would be real time and would be at a fixed intervals?	Present reporting layer(in DWH ecosystem) is logical layer (non persistent) so data availability is instantaneous. Regarding ingestion from upstream, there are all freq.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
135	64	Regulatory Reporting - 2	Automation –Tool should automate analytics and reporting workflow end-to-end, including all data collection, enrichment, and management, as well as all calculations, processes to final report submission. Currently 500+ jobs are being used for Tranche 1 DCT generation along with 500 more for other regulatory reports/returns.	We believe that the Bank has third part solutions for risk regulatory returns such as OFSAA for RCAIII, ALM Returns, etc. Does Bank wants to migrate from OFSAA to new solution and idea is to develop the data model at each line item level in each regulatory return and host the same in the new solution?	Present scope is with DWH eco-system.
136	64	Regulatory Reporting - 2	Automation –Tool should automate analytics and reporting workflow end-to-end, including all data collection, enrichment, and management, as well as all calculations, processes to final report submission. Currently 500+ jobs are being used for Tranche 1 DCT generation along with 500 more for other regulatory reports/returns.	Request confirmation on the list of regulatory returns which Bank wants to automate in the new solution	Details can be shared with selected bidder post RFP.
137	65	Regulatory Reporting - 7	Electronic Submission – Should support for all regulators globally in all required formats, including XBRL, XML or other file-based electronic submission.	Is the Bank looking for hosting all the regulatory returns which are submitted by the Bank entities in other jurisdictions in the new solution?	Present scope is with DWH eco-system.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
138	65	Regulatory Reporting - 8	Flexibility to deal with any level of Complexity – Tool should handle the flow of information particular to any client’s organization, such as specific Key Risk Indicators (KRIs), syntaxes, data taxonomies and process operating models. Further, as demands like BCBS 239 require near-constant fine-tuning to address data governance, Business Rules should enable adjustments to regulatory and management changes with its out-of-the-box configurability and slice-and dice of data.	Does the vendor is required to develop the data taxonomy and/or regulatory interpretations for each line item level in regulatory returns or same will be provided by the Bank to leverage and create data model?	Bank will provide data dictionary.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
139	66	Regulatory Reporting - 11	Pre-submission Review – Multiple report writers should allow users to review reports in various formats before submission, with the ability to drill down and make manual adjustments where necessary.	Making the manual adjustments as the pre-submission review generally leads to inconsistencies in the reporting considering users making changes in specific regulatory returns which might result into validation failures while cross validation with other returns. Considering the same, discussions will be held involving all stakeholders responsible for reg returns submissions to identify the best way to make any manual interventions in the tool. Request confirmation if Bank is ok with this.	Bidder to propose suitable approach
140	66	Audit and Log Management	General query	What is the scope of logging? Is it to be implemented for all the users, applications, jobs and activities associated to complete data landscape?	Audit logs should be maintained for all the users, applications, jobs and activities and same to be retained / available for minimum 6 months.
141	66	Audit and Log Management	User should be able to control the events/items being logged. For e. g user can opt for enabling logging of failures only	Admin Portal for log required where admin can manage what log to generate?	Yes

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
142	66	Audit and Log Management	Should have metadata enabled reporting mechanism on run time log.	Fraud analytics and reporting required on the top of logs? What would be the scope of reporting and analytics?	Audit logs should be maintained for all the users, applications, jobs and activities and same to be retained / available for minimum 6 months.
143	66	Audit and Log Management	Logging of operational activities: Support the logging of all user activities without slowing down the performance.	Is there any performance benchmark?	No
144	67	Data Encryption	The sensitive and personal identification data on ISDLWDW should be encrypted	What type of personal data is available? Encryption would be required on which layer (Data Lake or Warehouse)?	All type of personal data (PII and SPI) is available. Encryption is required in all the layers (rest, motion, use)
145	67	Data Encryption	General query	Is there any encryption technique followed within organization or vendor can select any technique which will be a best fit?	Bidder need to provide industry standard best suitable encryption technique in their solution.
146	67	Data Encryption	General query	Are there industry-specific regulations or compliance standards that mandate data encryption for your organization?	Bidder need to comply with all the guidelines/laws/instructions issued by Bank/Regulators/Govt. from time to time.
147	67	Data Encryption	General query	How to handle data encryption during backup and archival processes?	Bidder need to provide industry standard best suitable encryption technique in their solution.
148	67	Data Encryption	General query	Do you use a dedicated key management solution, and what is your key rotation policy?	SKLM is being use.policy will be advise to selected bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
149	67	Data Encryption	General query	How do you handle encryption when data is shared with third-party services or partners?	Bidder need to provide industry standard best suitable encryption technique in their solution.
150	67	Data Encryption	General query	Do you implement multi-factor authentication or other security measures for user access?	No.
151	68	Data Masking	General query	Is there any General Data Protection Regulation (GDPR) policy?	Yes
152	68	Data Masking	General query	What types of data do you need to mask? (e.g., structured databases, unstructured files, cloud storage)	As per prevailing regulations, all sort of PII, SPDI, etc stored in structure , unstructured formats needs to be encrypted / masked.
153	68	Data Masking	General query	How do you handle data masking for analytics and reporting purposes?	Data masking is done at DB2 (DB2 native masking) level.
154	68	Data Masking	General query	Are there requirements for dynamic data masking, where sensitive data is dynamically masked based on user roles or permissions?	Yes. Dynamic data masking will be required based on the user roles or permissions
155	69	User Management	User Access Management should allow IT administrators to securely manage access to services/information/ resources for all the users.	Would different UAM privileges required for different area of data architecture?	Yes. Depending on the applicable roles, relevant previledges only to be granted.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
156	69	User Management	Privileges granted to the user should not conflict with Access Management - Segregation of Duties policy of the Bank.	What are the Duties policy of the Bank related to access management?	Conservative approach followed by providing only need to know basis access as per defined roles of the user.
157	69	User Management	Vendor should propose automated solution / tool (s) of User Access Management (UAM) for administration of giving access to individual users within a system access to the tools they need at the right time.	How many users are we talking about (approx.)?	Pleaser refer Sr. no 33 of page no 86
158	70	Downstream Data Consumption 1	Capability to easily connect with any downstream applications currently used at the Bank and be flexible to support all end users in the Bank based on future needs.	What is the current downstream application being used ?	Business Intelligence department, CRM and Analytics are primary downstreams users of DWH. Application being used by these departments will be mentioned in the RFP.
159	70	Downstream Data Consumption 2	Should support bulk data and targeted data extracts via statistical tools and APIs	What are the current statistical tools and API being used ?	IBM SPSS tool. For data extraction, API is being used.
160	71	Downstream Data Consumption 6	Dedicated high-performance department wise sandboxes allocated to end users for R&D	What is the consumption pattern and frequency of the end users here?	queries / reports can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experiece and may vary depending on future load /requirement

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
161	84	Hardware Specifications	The proposed hardware is mission critical for the proposed project and support of 24 x 7 with an uptime of 99.99 % to be ensured by providing support at PR, and DR site for a period of 5 years.	What would be the DB growth for Prod/Archive/Data Lake for Next 5 year	Details will be mentioned in RFP
162	104	Appendix G-	Certificate of Local Content	Local content is usually provided as RFP responses which requires auditor approval on estimates. Estimates are not ready as part of EOI response, hence please consider to not have as part of EOI submission	Compliance of local content is based on total cost of project
163	59, 60			Only defined quality rules are included in scope or this list can expand?	Yes.
164	59, 60			Would there be any business/functional data quality check also?	Yes.
165	74-80	Business Intelligence Tools		The owners of the source system should have visibility of the data processing status. For this would the data visibility be restricted at source system level.	Data loading is published in the dashboard for monitoring purpose
166	74-80	Business Intelligence Tools		As per the EOI we understand that the ISDLDW will be on prem but the BI dashboard/report has to web-based with user level restrictions. Is this correct?	Bidder can propose best suitable solution.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
167	74-80	Business Intelligence Tools		What are the different parameters on which the performance of the product will be gauged? Is it scalability, user-accessibility or pure-play im-memory consumption and performance?	Bidder need ot propose the best suitable solution.
168	74-80	Business Intelligence Tools		Are all reporting needs to be addressed through the web-based interface or there could be a need for desktop user access as well?	Web based tools will be recommended.
169	74-80	Business Intelligence Tools		Does the reporting product need to be centralized across the organization or may vary based upon the nature of data and work environment?	Reporting product need to be centralised for all the businees units.
170	74-80	Business Intelligence Tools		How much of plug and play is required in the reporting layer from the business user? Are the reports to be generated from scratch?	Bidder need ot propose the best suitable solution.
171	74-80	Business Intelligence Tools		Do the business users who will be the consumers of the plug and play feature of the BI tool have prior knowledge on any of the dashboarding or story-boarding technologies ?	Bidder need ot propose the best suitable solution.
172				How many years of data is required in ISDLDW	Details will be provided in the RFP.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
173				Considering the magnitude of the asks in the EOI, the bidder team requests for an extension on submission of the EOI Response from 10th January 2024 to 31st January 2024	Bid submission date has been extended till 29.01.2024
174			General Query	Bidders are currently expected to propose complete on-prem solution which can be scalable to public/hybrid cloud in future, kindly confirm.	Please refer to sr no 'e' of page no 34 of the EOI document
175			General Query	What is the reason for looking at new solution? Is it the same as mentioned in Pt. e on Pg. 34.	Existing contract is coming to the end of life
176			General Query	Please share details of all data sources (~100 systems) for creating a DWH, including database software, operating systems, versions, OEM.	Architecture diagram has been presented in pre-bid meeting
177			General Query	Request you to please share details on files types, sources and volume of data ingested daily as well as volume to be migrated to new platform.	Architecture diagram has been presented in pre-bid meeting
178			General Query	Please share the existing DWH architecture diagram and architecture component details, including software, versions, OS etc.	Architecture diagram has been presented in pre-bid meeting

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
179			General Query	What is the use volumetrics - total users, concurrent users - who will be accessing the DWH system?	Please refer to sr no 11 of page no 46 and 13 of page no 48.
180			General Query	How much is the percentage of data in Structured, Semi structured and Unstructured form.	Mostly data is structured.Data used in Datalake is unstrcutred at source cbut can be converted to structured data.
181			General Query	What is the expected user mix - reporting users, analytical users, data mining users?	Users are bifurctaed based on applications which they are using.
182			General Query	Can we add Global credentials with Self declaration as supportings?	Please refer to Corrigendum-2.
183	7	1 (i)	This Request for EOI is issued by the Bank on behalf of SBG for procurement	Please clarify if scope is to implement data lake for SBG (incl. SBI, its subsidiaries, JVs, foreign offices or only for SBI India banking	Its required for SBG.
184	8	2	Data Warehouse has migrated on new infrastructure on 10th May 2023 as per five- year Agreement between Bank and M/s IBM India Private Limited in February 2020	Will data warehouse be dismantled after expiry of 5 year agreement or will it function along side new data lake. Is this 5 year agreement expiring in Feb 2025 or May 2028	Present agreemnt with the existing vendor/OEM is expiring on Feb 2025

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
185	31	Appendix B	<p>Bidder should have experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake preferably in BFSI sector.</p> <p>Copy of the order and / or Certificate of completion of the work. The Bidder should also furnish user acceptance report from client(s)</p>	<p>We request to please accept at this stage bidder self declaration as proof of implementing large scale EDW / data lake. Clients, especially foreign ones, donot typically allow showing documents pertaining to their projects. We may at best manage client publication on best avaiability basis. Also, peta byte implementations in BFSI have been limited, we request to please delete this preference clause and accept ref. of industry other than BFSI</p>	<p>Please refer to Corrigendum-2.</p>
186	31	Appendix B	<p>Client references and contact details (email/ landline/ mobile) of customers for whom the Bidder has executed similar projects in India. (Start and End Date of the Project to be mentioned) in the past (At least two client references are required)</p>	<p>With limited implementation in India of data lakes / EDW of peta byte type scale (there have been very few in BFSI), we request to accept bidder self declaration on EDW implementation experience of data size in TB. Please modify suitably. Contact details of client reference persons cannot be shared due to NDA signed with clients</p>	<p>Please refer to Corrigendum-2.</p>

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
187	24 28	19 Appendix A	i.Lodgement of an EOI is evidence of a Bidder's consent to comply with the terms and condition of Request for EOI process and subsequent bidding process  v.It is further certified that the contents of our Bid are factually correct. We have not sought any deviation to the terms and conditions of this EOI	As we come close to submission, it is possible some assumption, other clarification is to be included in submission. Please allow consideration of those for mutual agreement	Bidder may submit their proposal with assumptions at this stage and further clarification may be sought upto presenttion stage.
188	28		general	Can any existing component be used. Please give detail of such components - license available quantity, type, version, if SBI will manage with OEM on its own or through new vendor etc.	Details have been presented in pre-bid meeting.
189	36	C 3.2	Number of Cognos Reports	Is no. of total reports out of EDW is about 100. Total reports requirement in a bank is much more. Do other reports come out of any other system	These are the reportes generated through cognos only. Other reports are not covered under this head.
190	82	C4	Vendor shall ensure all Installations & Implementation to be done by OEM badged resources only	Please clarify if this clause is only for hardware installation /implementation or entire solution	terms will be applicable for entire solution proposed by the bidder.
191	89	C4	Sizing given in the Annexure G is in compressed form with compression factor of 2.5	Please specify if bank has any guidance on minimum data compression factor to be used for proposed solution	Compression ratio should be around 4.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
192	42	5	Solution proposed for Data Ingestion should be compatible to ingest and replicate data simultaneously on all available regions.	<p>Although SBI has office and branches across all regions of the world, but assumption is that this ISDLDW setup would primarily be done in India as centralized Data Warehouse repository and would be getting data from source systems located across the world (either on cloud or on-prem). Please validate that bank will provide data of all regions / geographies to the proposed centralised data lake.</p> <p>Will there be any mechanism to improve the network traffic from USA/Europe to centralized India ISDLDW environment.</p>	<p>Bank will provide data of all the regions / geographies to the proposed centralised data lake.</p> <p>mechanism to improve the network traffic from USA/Europe to centralized India ISDLDW environment - not applicable</p>

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
193	42	6	End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics maximum within 3 hours from day end for batch data loads on all days including month-end / quarter-end / year-end. Vendor may propose suitable solution to achieve the objective	<p>Three hour threshold for data processing End of Day for each region /respective geography.</p> <p>This would mean that DW environment is loading data/ETL processing during India business hours when Users are running their queries.</p> <p>Please let us know the volume of data coming from different geographies?</p> <p>Is there any need to have a separate cluster for ETL and compute for reporting purpose?</p>	Architecture along with data flow has been presented in pre-bid meeting
194	48	12	Existing reports and extracts generation jobs on DWH should be analyzed and transformed to the ISDLDW. The vendor should use preferably off-the-shelf tools and not resort to building from scratch.	<p>Does this means that bank wants to have a new data model for the ISDLDW which would effectively mean creation of new reports and ETL mappings as per new data model?</p> <p>Basically, need to understand how much existing ETL, DM and reporting components bank wants to reuse and also current issues faced by bank for the same.</p>	It is expected that all the existing reports and jobs to be migrated to the new system.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
195	53	2	Data Archival solution should not be visible to end user, but Archived data should be available for all end users. For end user it should be a single view with Data Federation/Virtualization Layer	Does it means that reporting tool will also connect to the underlying storage through Virtualization/Federation tool to make use of both current and archive data?  Is the same true for all other applications consuming archive and current data?	Yes it will be required.
196			In view of the intent to reduce the hardware footprint (in future), the technical architecture of ISDLDW solution should be flexible to accommodate moving of ISDLDW on cloud and also should support cloud bursting for heavy workloads. The Bank understands that there can be differences in services offered by cloud service providers. The ISDLDW solution architecture should be designed considering as-is infrastructure availability in cloud.	For cloud bursting, do we need to provide the hardware sizing for the cloud right now as that would be purely on the need basis at that point in time. Same goes for the cost of having the public cloud capabilities basis the requirements at that time.	System should be capable to migrate on cloud. So, hardware sizing estimation is not required at this stage.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
197			Data Governance	<p>The data governance section is only dedicated towards technical implementation of tool to get the Data Definition, business rules, integration of technical/business metadata etc into the tool.</p> <p>This is not concerned with setting up a DG office and defining its members and policies. Please clarify that defining DG policy is not in scope?</p>	Bidder need to comply with the data governance policy of the Bank.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
198	52	11	Vendor to review the existing architecture during migration and remove duplication of data and recommend improvements in overall setup if any	<p>The existing setup has more than 1 PB each in current and Archive DB and few other strages. Is the expectation to review the entire Data Model, reporting and ETL setup. Same goes for hardware setup in terms of creating environment based on Object Storage and Kubernetes cluster?</p> <p>Is the bank looking for re development of entire architecture or reuse existing components. If yes, which components can be re-used else it would be a time consuming task to do the review especially for DM, ETL, Reporting and Analytical models</p>	<p>Bidder needs to review existing data model and propose better option, if any.</p> <p>Infrastructure changes will be required based on solution provided by the bidder.</p>
199	57	1	Provide traceability – it should be possible to track and visualize any data transformation or any rule applied to data in the ISDLDW eco-system	A single tool may not be able to provide traceability over entire pipeline consisting of CDC (+ file movement), ETL, Reporting and ML Engine. Is the bank fine with the staggered lineage provided by individual tools for these components.	Bidder can propose the same in the EOI proposal

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
200	58	5	Capability to classify and store (personal identifiable information) sensitive data in encrypted /masked form and should have capability to decrypt/unmask such information in ISDLWD when required by only authorized ID's.	The tools will classify the information as PII basis the rules defined, but the final approval on classified needs to be done by SBI functional team as to whether a provided elements is PII or not.	Yes
201	68	6	Proposed Solution should be capable on ingesting encrypted data from source system. It should support the encryption/decryption mechanism implemented at source system.	Is it correct to say that data in movement from source to ISDLWD or other application should be encrypted. Which tool does the encryption may not be a concern i.e. encryption can be done by source system, ETL Tool, DB of ISDLWD or reporting tool, but the data should stay encrypted during the flow from source to ISDLWD?	Yes . (minimum TLS 1.2 and AES256 to be used at present).
202	68		User Management	Which tool is currently being used for the User Access management?	Bank is using ITSM at present. Bidder can also propose other tools in their solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
203	71		Data Science Platform with AI/ML Capabilities	<p>How many AI/ML use cases are currently in production that needs to be migrated to new platform?</p> <p>How many new AI/ML usecases to be created as part of the project?</p> <p>How much time it takes currently to process the AI/ML models?</p> <p>Size of the current data science setup in terms of number of nodes, RAM on each node, storage, networking capability?</p> <p>What is the number of real-time models amongst these and platform sizing for these?</p>	Approx. 100 models are presently running. Details will be provided in the RFP.
204	80	30	There should be separate criteria for BI user versus administration training. Training may include in-person classes, online classes -- live or prerecorded -- or web recordings for specific features or processes.	<p>Number of days for which training needs t be provided?</p> <p>Can the training be provided virtually or in-person?</p> <p>In case of in-person training, location where training needs to be provided?</p>	details will be shared in the RFP.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
205	63	Appendix-C: Technical Criteria/Scope of Work C4. Functional Requirements Security and Compliance - Point No. 6	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	<p>To ensure Availability under CIA triad, it is essential to deploy a dedicated On-Premise DDoS Protection. On-Premise DDoS Protection solution detects, alerts and protects against Volumetric, State Exhaustion, Application Layer and Outbound DDoS Attacks and thus ensures Availability of Critical Systems, Services and Applications in Data Center.</p> <p>So, could you please confirm if you need On-Premise Anti-DDoS Protection Systems ?</p>	DDoS protection is not required for this solution, but bidder need to integrate with proposed solution.
206	63	Appendix-C: Technical Criteria/Scope of Work C4. Functional Requirements Security and Compliance - Point No. 6	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	<p>As a part of Solutions to identify and alert on Data leakage, it is essential to deploy a packet based Network Detection and Response (NDR) Solution. A packet based NDR enables Enterprises to achieve complete security visibility into their network and identify vulnerabilities and threats with high accuracy.</p> <p>So, could you please confirm if you need a Network Detection and Response Solution ?</p>	Proposed solution should be integrated with Bank's centralized security Solutions. No additional solution is required for this purpose.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
207	56	Monitoring Dashboard Point No. 7	Industry best tools for performance (slowness) monitoring	As Industry best practices the tool or solution should be agent less as it will be non-intrusive and dose not impact on performance on the Servers or the network.	Suggestion by bidder
208		Monitoring Dashboard Point No. 9	End to end performance monitoring with automatic alerts	End to end visibility from client to Application and the backend database visibility is possible via traffic monitoring solution with automatic alerting.	Suggestion by bidder
209	57	Monitoring Dashboard Point No. 21	DB health monitoring tool as well as monitor of N/W health	Monitoring of DB application health and network health will be possible only via Application aware Network Performace monitoring (Traffic based Agent less solution) for this requirement please provide the capacity/ bandwidth for network traffic monitoirng.	Suggestion by bidder
210	21-22	Clause 14(v)(b): Debarment from participation including removal from empanelled list	Other than in situations of force majeure, technically qualified Bidder withdraws from the procurement process or after being declared as successful bidder: (i) withdraws from the process; (ii) fails to enter into a Contract; or (iii) fails to provide performance guarantee or any other document or security required in terms of the RFP documents;	The said reason for removal from list of empanelled vendors to be removed since the next process of RFP along with its relevant terms and conditions is not issued to the Bidders. Therefore this reason has to be deleted from this EOI document	EOI terms are final

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
211	24-25	S.No - ii of Clause 19: Other Terms and Conditions	ii. Willful misrepresentation of any fact within the Bid will lead to the disqualification of the Bidder without prejudice to other actions that the Bank may take. All the submissions, including any accompanying documents, will become property of the Bank. <b>The Bidders shall be deemed to license, and grant all rights to the Bank, to reproduce the whole or any portion of their Bid document for the purpose of evaluation and to disclose the contents of submission for regulatory and legal requirements.</b>	Request Bank to clarify the bold portion of the clause	EOI terms are final
212	104-106	Appendix - G	RFP Number reference mentioned in the said Appendix	Request Bank to either remove the said Appendix - G from EOI or provide EOI reference number in the said Appendix as RFP is still not issued to the Bidders.	bidder can consider RFP no as EOI no.
213	104	Appendix F	Value of Work Order Name of the contact person and designation Phone number of the contact person E-mail address of the contact person Start - end date of project Present status	Due to NDA with clients, these details will not be possible to be provided. We request to please delete these points / accept submission without these details. These may be looked into at a later stage as feasible	Client references are required.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
214	104	Appendix G	Certificate of Local Content	This certificate may be taken during RFP submission. We request you to kindly delete this for EOI.	Compliance of local content is based on total cost of project
215	10	5. ELIGIBILITY AND TECHNICAL CRITERIA/ SCOPE OF WORK: a	If any Bidder submits Bid on behalf of Principal/OEM, the same Bidder shall not submit a Bid on behalf of another Principal/OEM under the EOI. Bid submitted with options of multiple OEMs shall also be considered Bid submitted on behalf of multiple OEMs.	Does this clause imply that the bidder with single or multiple Principal / OEMs are bound within a single bid and none of the parties can be part of any other bid with another bidder.	If any Bidder submits Bid on behalf of Principal/OEM, the same Bidder shall not submit a Bid on behalf of another Principal/OEM under the EOI. Bid submitted with options of multiple OEMs shall also be considered Bid submitted on behalf of multiple OEMs.
216	34	C21e	The proposed solution should be implementable in Physical hardware / Bank's own cloud (Meghdoot) and scalable to public / hybrid cloud in future.	Is the bank flexible to choose a hybrid solution comprising of one or more OEMs with on-premise / Meghdoot deployment as well as with Cloud service providers such as Azure, Google, AWS.	At present bank is not looking for public / hybrid cloud solution. However, proposed solution should be scalable to public / hybrid cloud in future, if decided to migrate by the Bank.
217	37	C414	Capability of In-memory and real-time analytics.	Please provide examples of the usecases requiring this capability.	Bidder may suggest suitable architecture and financial sector use-cases for this requirement. Details will be provided in the RFP
218	37	C418	Compatible with Oracle CDC (Goldengate) at near real-time	Is Oracle Goldengate the Chosen CDC tool or bidder can propose alternatives for the same	Bidder can provide alternate better solution.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
219	31	Appendix B 5	Bidder should have experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake preferably in BFSI sector.	We presume that our global / international experience for implemtations of large-scale data warehouses and lakes is also acceptable to suit this criteria	Yes. EOI terms are final.
220	89	C 50	Annexure G gives the ISDLDW sizing forecasted for next five years. Sizing given in the Annexure G is in compressed form with compression factor of 2.5.	Can you please provide us with Annexure G as mentioned in this clause	Details have been presented in pre-bid meeting.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
221	35	C 3.2	Volmetric of existing data warehouse	<p>Can you please provide distribution of the various workloads that are currently in place for the said volumetrics such as below. Nature of workloads and compute / storage needed for the same would help understand the volumetrics better</p> <ul style="list-style-type: none"> <li>- Warehousing jobs,</li> <li>- Number of marts and types of marts</li> <li>- Analytical jobs and storage</li> <li>- Feature model data sets</li> <li>- Ingestion jobs</li> <li>- Unstructured data ingestion</li> <li>- Third party interfaces or Integration</li> <li>- BI - management reporting, Finance reporting</li> <li>- Regulatory reporting</li> <li>- Collections</li> <li>- offer marts etc</li> </ul>	Please refer to section C3.2 at page no. 35 and 36.
222	53	9	The Bank reserves the right to enforce mock up scenarios for failure to ensure that the ISDLDW – DR solution is working as expected. The vendor should support the Bank to demonstrate such scenarios	Please provide a sample Mockup Scenario	Some mockup scenarios will be part of RFP document as illustrations. DR should be based on realtime replication.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
223	40	C4 - Architecture Requirements	Cost model (e.g. how licensing will be done)	Our understanding of costing model would mean basis of providing commercials for the solution such as compute power, storage, number of jobs, number of users etc. Note that the basis would not provide the final costing as the details needed for the same are not available yet	Suggestion from bidder
224	48	Appendix-C: Technical Criteria/Scope of Work	Automatic recovery of data after failure/rejection of record needs to happen without any manual intervention	We need more clarification on which type of recovery of data, i.e from which type of failure or rejection	Expectation is recovery from all type of failures.
225	48	Appendix-C: Data Processing framework	Should have audit and error logs for auditing and troubleshooting	Timelines to hold audit logs need to be defined	Audit logs should be kept for minimum 6 months.
226	51	Appendix-C: Migration from Existing Setup to Proposed Solution	Migration of user details.	What kind of user details need to be migrated ? If ETL tool is replaced user need to be available to rewrite in different language	ETL pipelines migrations are also in the scope, if technology is changed.
227	54	Appendix C: Data Archival and Backup	Store backup of entire ecosystem on suitable cost-effective, fast recovery infrastructure (Currently tape backup is taken)	Backup Recovery via tape is cost effective but time consuming . Faster mechanisms would have higher cost, Safe to assume higher costs are acceptable	Bidder should provide faster mechanism for data backup & recovery and commercial should be inline with the proposed solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
228	59	Appendix C: Data Quality	Vendor should propose end-to-end solution for Data Quality Management starting from data origin till the data consumption. These tool (s) to be used for addressing various aspects of the data quality problem mentioned below on SBI data set during data ingestion, data processing or data consumption as advised by Bank on case by case basis	Please share examples of data quality challenges that are encountered. We are assuming that data quality issues origination at source will not be fixed in the DWH or data lake. Please confirm this assumption	Anything that need to be fixed at data warehouse level has to be fixed by the bidder.
229	64	Appendix C: Security and Compliance	Compliance of Bank's Information security policy, Data governance policy.	Please provide details of Bank;s information security policy (apart from what is followed at present), data governance policy (not much defined as per our knowledge)	details will be shared with finally sected bidder
230	83	Appendix C: Hardware Specification	The Vendor shall also carry out OS Hardening, Anti-Virus installation, Create Super user for the Production, DR and UAT/Dev environment according to Bank's policy and secured configuration document	Please provide overview of Banks Policy and Secured config document in addition to what is followed at present	SCD documents will be changed from time to time and same will be shared with finally sected bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
231	86	Appendix C:Hardware Specification	UAT setup should be exact replica of production (all the business schemas and software / applications / services installed) and to be refreshed every month. Data to be put up in UAT will be finalized by Bank. Dummy dataset to be inserted in UAT to maximize the benefits. Either Data Anonymization on PROD data or any other suitable approach to be proposed by Vendor.	Please share the details of data privacy policies that need to be complied with for dataset that needs to be pushed to UAT. In addition, please share the mechanism as per bank's policy for making the dataset available in the UAT environment	details will be shared with finally sected bidder
232	87	Appendix C:Hardware Specification	Ad-hoc jobs of any complexity should not hamper the scheduled jobs performance.	It is assumed bank understands such features come in with high cost and there is always a limit to complexity it can not be 'any complexity'	Sugession by bidder
233	88	Appendix C:Hardware Specification	DEV and UAT environments are to be set up at the DR site	Right now Dev & UAT set up at Rabale and DR at Hyderabad. Are we continue with this arrangement?	DR , Dev and UAT also to be setup at hyderabad.
234	42	Data Ingestion	Data ingestion system should be capable to configure data ingestion from various sources sharing all type of data through various methods e.g streaming, batch, files, archive logs, etc.	What could be the expected maximum size and retention period for a record in stream. And also what is the maximum number of consumers can be expected for the stream.	Details will be provided in RFP phase. Bidder to suggest scalable architecture

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
235	45	Data Storage	A multi-temperature data management solution to be proposed by vendor where data that is frequently accessed on fast storage—hot data—compared to less frequently accessed data stored on slightly slower storage—warm data—and rarely accessed data stored on the slowest storage —cold data. System should also be capable automated storage tiering and seamless data transfer between hot, warm and cold storage. Data residing in any of these storage areas must be seamlessly mixed / merged according to requirements without impacting performance.	<p>a) What is the minimum storage duration between hot, warm and cold strategy?</p> <p>b) Hot, Warm, Cold nature of storage has a cost benefit and impact on data retrieval based on consumption pattern. Please clarify and elaborate the requirement - "Data residing in any of these storage areas must be seamlessly mixed / merged according to requirements without impacting performance."</p>	<p>a) will be decided by the Bank as per policy.</p> <p>B) Suggestion by bidder</p>
236	61	C4. Data Quality #16	System detectable Data Quality rules using business concepts defined at Data governance framework	a) Business concept is used to derive the business process or object. Is reference is about business process or business object? Please elaborate this with an example	Both (e.g. mobile number, capturing of form16, etc ).

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
237	35	C3.2	Prod DB Size, Archival DB Size, Data Ingestion Size etc	Some volumes are listed in compressed form and some in uncompressed form - it should call out uncompressed size (or include compression ratio other wise)	Compression ratio should be around 4.
238	42	C4 - 5	Solution proposed for Data Ingestion should be compatible to ingest and replicate data simultaneously on all available regions.	What regions are referred here - OnPrem or Public Cloud Regions, Does it refer to Hyderabad, Pune and Mumbai as regions ?	Regions(on-prem/private cloud) to be DR, PR for replication
239	42	C4. Data Ingestion #6	End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics maximum within 3 hours from day end for batch data loads on all days including month-end / quarter-end / year-end.	<p>a) Workload, complexity, number of jobs in daily, week-end, month, quarter, year-end are different. Is 3 hour SLA applicable for all frequencies</p> <p>b) what is the rationale for 3 hours SLA for month end , Quarter End Scenario - Does linear dependencies because of Banks Functional requirement even make it feasible - Please include functional requirement that mandate sequential processing of data right from landing to end user report creation</p> <p>c) Clarify the projected daily data ingestion volumes</p>	<p>a) yes (This is Bank's expectation)</p> <p>b) Should not be more than 3 hours from data availability from source side.</p> <p>C)Incremental data ingestion ~ 2 TB</p>

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
240	43	C4. Data Ingestion #12	Trigger mechanisms in identifying any structural changes at source	a) What is the ask here as Source has to share metadata to detect every structural changes without which no product can meet this requirement b) Are the source systems amenable to necessary changes for implementing this?	Existing CDC tool detects the changes and alerts; Bidder can provide alternate improved solution.
241	44	C4 - 19	Tool should have in-built capability to optimize/prioritize the ingestion pipeline system as a whole.	Not clear as what kind of optimization and prioritization is expected - please provide examples	Example is already mentioned in the same point (C4.19)
242	46	C4 Data Storage 7	Processing engines should have capabilities to process all kinds of Data Storage formats used by the Bank	What is meant by all kinds - please be specific ?	All prevailing formats available in the market should be supported.
243	46	C4 Data Storage 11	scalable upto 6,00,000 concurrent queries	Extremely generic requirement that can mean different capacity based on different interpretations . what is complexity definitions and complexity levels of these queries ? Can this be standardized on query cost parameters ?	Bank's expectation are final; Queries can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
244	48	C4 Data Processing 13	Data transformations should be triggered in parallel. The ISDLDW should be capable to run multiple transformation jobs in parallel. The ISDLDW should be able to run at-least 500 jobs in parallel (concurrently), scalable up to 1500 in next 5 years, of varying complexity - simple, medium, complex, in batch or near real time mode every day.	a) Number of jobs is not a standard measure unless quantified with complexity levels and complexity definitions. E.g. 1 jobs can consume resources equivalent to 100s of jobs b) What is the YoY concurrency %age increase expected? c) Please clarify this requirement and its relation to 3 hours of SLO mentioned earlier d) Please specify the complexity breakup (in percentages)	Details will be mentioned in RFP
245	55	C4 Cloud Integration 6	Adherence to global and local standards related to cloud as well as compliance to applicable regulatory requirements.	which local and global standards - there are many standards out there	Bidder should comply with all applicable standards
246	84	C4 Hardware 18	24 X 7 with an uptime of 99.99 %	Is there a provision for planned downtime for specific types of releases and upgrades	yes
247	87	C4 Hardware 35	scalable up to 500 users in next 5 years	What kind of activity per user with complexity levels and complexity definitions	SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement
248	88	C4 Hardware 46	DEV and UAT environments are to be set up at the DR site	does it mean DC is not supposed to have any Dev Environment or Dev and UAT is required on both side ?	Duplicate query by same Bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
249	89	C4 Hardware 47	DR, UAT and DEV should be exact replica of production.	Does that mean capacity and configuration also must match - how can it match exactly if config is different	Bidder to propose ratio for dev/uat as per their solution. DR should be exact replica of PR
250	53	Disaster Recovery 8	Jobs developed to be deployed in production, should be deployed both in production and DR. The same applies to any admin activity, security policy and other activities on ISDLDW	Does this it will be an Active-Active Setup ? For example If an artifact like a machine learning model created by Notebook is saved and deployed simultaneously deployed in DR ?	DR should be always sync with PR and users should be able to query data if required.
251	34	C2.1	a. "Integrated solution with real time / near real time sourcing of structured and unstructured data"	a) Please clarify if you mean all sources and targets to be integrated real time/near real time? b) What percent/volume of data is currently required by business teams real-time / near real time? C) What unstructured data do you have today? What are the types, volume and number of systems? What percent / volume of the unstructured data do you foresee to be integrated in real time with the new data platform solution in near term?	a) Real time / near real time is one of the capabilities expected by the Bank. B) percentage of volume required under real-time/near real-time may vary time to time c) unstructured data sourcing is expected in near future.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
252	39	C4. Functional Requirements #42	42) Support ML, Deep Learning, AI, Generative AI Models	<p>a) Can you please elaborate at a high level the Gen AI capabilities / Use Cases you are looking for in near / mid term?</p> <p>b) Pls elaborate on the type and no of models of Generative AI and deep learning models.</p>	Bidder to suggest all types of GenAI capabilities which are scalable. Details will be provided in the RFP
253	41	C4. Data Ingestion #1	Capable of ingesting data from any source system in automated manner currently implemented in the Bank, or any future standard source systems that the Bank will decide to use with high throughput and low latency. Vendor to propose performance benchmarking for the same.	<p>a) Please clarify what have you meant by "Vendor to propose performance benchmarking for the same."? Will the vendor benchmark and define "high throughput" and "Latency" for different systems by tech stack?</p> <p>b) Please clarify the term "future standard source systems". Does this mean any future system of SBI based on technologies available with Indian banks as of today? Future systems may be on tech stack not available today.</p> <p>c) What is the estimated growth in data volumes over next 5 years?</p>	<p>a) Bidder to mention performance benchmarking in the EOI. Based on this final parameters will be provided in the RFP.</p> <p>B) both may be considered.</p> <p>C) details will be mentioned in the RFP.</p>
254	42		3. Graphical User Interface (GUI) based framework to configure sources to ISDLDW	Please clarify what is meant by "configure sources" through GUI? Is it setting up connections and/or throughput or other parameters relevant for data pipelines?	Bidders need to provide suitable solution to fulfill the requirement.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
255	NA	NA	NA	Many of the questions in compliance sheet are relevant for older technologies such as DWs OR ETL tools deployed on traditional MPP / RDBMS OR on-prem setup. Some of these may not be relevant on the cloud native architecture of the new platform chosen. Can we answer "NA" with an explanation instead of "Yes" or "No"?	Yes
256	NA	NA	NA	The EOI speaks about a lot of Capabilities? What is your preferred approach? "Big Bang" or starting with some foundational capabilities and then expanding and addition of additional capabilities in phases? For Open Source tooling based approach, a lot of POCs are usually required based on sample use cases, migration or co-existence of old & new platform, integration with legacy systems and so on. In such a case "Big Bang" may present its own risks.	Bidder to propose suitable approach for the tech stack.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
257	37	C4	Analytical capabilities specifically tied to stream data, including windowing analytics, analytics on in-flight data	Clients ask for this capability. But how much of Analytics on flight use cases do you have today or foresee in near term? Can you provide a high level view of your near term vs long term capability needs for us to work on a cost optimized platform architecture, incrementally expandable in terms of capabilities?	Bidder may suggest suitable architecture and financial sector use-cases for this requirement. Details will be provided in the RFP
258	44	Appendix C-7	Data sanity checks, automated reject processing, validations and reconciliation of data should be available as part of data ingestion solution to ensure the integrity of data.	<p>a) Does "automated reject processing," refer to automated rejection?</p> <p>b) How is this being done in the current solution? Are there rejection, reprocessing, recon of data happening today?</p> <p>c) Sanity checks, reject processing, validations and reconciliation are usually implemented programmatically. Can we assume that after migration/refactoring of existing applications to the new platform these capabilities should function in similar fashion as it is doing today? Are you currently experiencing any problem in these areas today? Please clarify.</p>	Solution required for some suitable workflow/tool/solution to monitor and intimate stakeholder for any suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
259	54	Appendix C - Data Federation/Virtualization-> 4	Data virtualization should support the use of APIs.	Is the support expected to get data from APIs?	API based sourcing capability is also expected in the solution
260	54	Appendix C - Data Federation/Virtualization-> 4	All data should within eco-system and across other system whenever possible be made discoverable and integrable easily through a single virtual layer which will expose redundancy and quality issues faster.	Is data cataloguing feature expected?	Yes ( Expecting data federation and data virtualization along with data discovery across systems).
261	55	Appendix C- Migration from existing Setup to Proposed Solution	Migration of All the remaining components of existing ecosystem (Mentioned in Annexure - D) as and when identified by Bank like job scheduler, reports, history of version control, existing tape backup, etc	Why is tape backup to be migrated ? The existing lake/dwh will have more latest data.  Which is the tool for version control? Is the history exportable or available in standard (csv/ json/ xml format )?	Expectation is restoration should be possible from tape backup to new solution (in case such scenario arise). Bidder can propose any suitable version control solution
262	55	Appendix C- Migration from existing Setup to Proposed Solution	Data migration from existing archival solution to new one.	When migrated current archived data will move to low cost cold storage is the understading. Please clarify.	Yes (expectation is cold storage data should also be accesible for user)
263	56	Appendix C - Replication	Bank proposes to setup replication based DR capable of of select tables, schema and jobs.	What is the relevance of selective replication for DR? Is there data that does not need to be replicated and available in DR?	DR is exact replica of PR (1:1); Bidder to propose suitable solution

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
264	58	Appendix C - Data Archival and Backup	All the applications connected to the non-archived data should be available with archived as well	Does this mean the data storage format of Active DWH and archived should be same?  Does all application require Archived data?	Not necessary; Bidder can propose suitable solution
265	59	Appendix C - Data Archival and Backup	Mixing and Merging data from Current/Live to and from Archival must not result in any significant loss of performance and response time	What is the performance loss accepted for archival data query? Please quantify. Is it 25% or 50% less than Active DWH?	Bidder can propose optimal solution with least performance impact.
266	55	Appendix C- Monitoring Dashboard	Operational logs analysis highlights	Please elaborate the requirement. Is this for required for DWH audit logs ? Is this required only for 1 day or 1 week? What is the volume of operation logs?	As per information security guidelines of Bank.
267	64	Appendix C - Data Governance	Automated propagation of changes to ISDLDW ecosystem Data Dictionary and business glossary by multiple sources as and when changes occur in source.	Is Schema evolution feature expected here? Will sources make available the changes?	Bidder can propose suitable approach/customized solution to address challenges due to schema/object changes in source.
268	67	Appendix C- Data Quality	Recommend Enrichment — Enhancing the value of internally held data by appending related attributes from external sources	What are the external sources to be considered for Enrichment? Please share list of external sources bank currently uses.	Bank is looking for capability for data enrichment. Bidder to provide suitable solution
269	68	Appendix C- Data Quality	Dedicated Data Quality team who will thoroughly verify the Data Quality across setup	How many people are required for the DQ team? Any specific experience level expected? Is this scope for tis team to define 5 New DQ Rules?	Bidder should propose DQ tool and team to work in tandem with Bank DQ team.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
270	66	Appendix C-Data Quality	Parsing and standardization — Decomposition of text fields into component parts and formatting of values into consistent layouts based on industry standards, local standards	Please elaborate on the global, local standards called out here. Is this requirement for semi/unstructured text data?	Yes (for semi/unstructured data). Applicable industry standards as per Bank's requirement
271	72	Appendix C-Regulatory Reporting		what is the current regulatory reporting tool in use at bankd?	BI tool (cognos) is used for reporting. Bidder can propose any suitable solution to specifically address the regulatory reporting
272	73	Appendix C-Regulatory Reporting	Pre-submission Review – Multiple report writers should allow users to review reports in various formats before submission, with the ability to drill down and make manual adjustments where necessary.	Is this functionality existing in current tool?	Yes (customized solution). Bidder can propose any suitable solution
273	77	Appendix C - Data Encryption	Proposed Solution should be capable on ingesting encrypted data from source system. It should support the encryption/decryption mechanism implemented at source system	What are the encryption mechanisms used in source? Will the source provide the decryption keys or expose decryption apis?	Details can be shared with selected bidder post RFP.
274	82	Appendix C-Downstream data consumption	Real time reporting can be done through Staging area of Data Warehouse	How is this different from operational reporting from reporting database of CBS?	TBD



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
275	83	Appendix C- Data Science Platform with AI/ML Capabilities	Implementing end to end analytics use-cases as mandated by the Bank	Please elaborate on meaning of end to end analytics? Assume integrations for analytics consumptions will be done by the downstream application.	Automated end to end
276	82	Appendix C- Downstream data consumption	Dedicated high-performance department wise sandboxes allocated to end users for R&D	How many departments are envisaged to have sandboxes? What is the indicative size of sandbox for department.	Details can be shared with selected bidder post RFP.
277	104	Appendix C- Hardware Specifications	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	Does this access also consider the published reports? What is the total no of reporting users?	This is Bank's requirement; Bidder can propose suitable solution
278	56	Appendix C - Replication	Annexure D	There are reference to Annexures in document. When will they be shared?	Architecture details has been presented in the pre-Bid demo
279	37	NA	Compatible with Oracle CDC (Goldengate) at near real-time	Does bank have Golden Gate EULA ? Is Golden Gate for bigdata license existing with bank?	Details can be shared with selected bidder post RFP, however Bidder can propose solution with any tech stack
280	34	C2.1	Real time integration with upstream and downstream users with latest technologies like streaming and APIs for data movement / utilization.	What is the extent of real-tme integration expected between upstream, downstream systems. Is there a count of upstream, downstream systems that can be shared for which real-time integration will be required	Details can be shared with selected bidder post RFP.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
281	34	C2.1	Low cost integrated archival solution with support for easy movement of data between active and archive data sets.	Are there any archival solution in place today? Active and Archive, are they defined by time periods or data size, please clarify	Archival solution is available. Archival process is manual based on defined practice.
282	35	C2.1	Support for effective data governance for data discovery, data modelling, data quality, data catalogue, data profiling, data reconciliation and data lineage.	Are there any existing Data Governance solution, Data catalogue in place? Is there an Enterprise Data Dictionary available? How is data profiling, data recon, data lineage being done in the current DWH project?	IGC, QS toos are available in eco-system. Bank has defined data dictionary. Data profiling tool - IA Data recon tool - no tool exists Data lineage tool - IGC
283	37	C4. Functional Requirements #18	Compatible with Oracle CDC (Goldengate) at near real-time	a) What is the current CDC solution available and the need for compatibility with Oracle Golden Gate? b) Please explain what is meant by Compatible?	Proposed solution / tool should have the capability to source data from Oracle logs in near real-time basis.
284	37	C4	Operational Intelligence: The ability to handle large numbers of concurrent users running fixed short analytical queries while still providing appropriate performance SLAs for operational workloads.	Please clarify on the no of concurrent users, no of analytial queries and the expected performance SLAs	Details are provided in the EOI document at Page 46, 61, 86. Further details can be provided in RFP.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
285	37	C4. Functional Requirements #20	Data Science: The ability to build, train and deploy models in the DBMS, the inclusion of a data science workbench, the ability to implement a smooth flow of model development and deployment, support for feature stores and other capabilities relevant to the exercise of data science workloads.	<p>a) Build, Train and deploy models in DBMS? Please clarify.</p> <p>b) Meaning of "In the DBMS"? Does it mean Data Scientists have to be given access to Database</p> <p>c) Is there a MLOps, Feature store currently available as part of the existing solution?</p>	Bidder need to propose the best suitable solution for SBI. Since EOI is an ideation process, RFP will be released with final requirement.
286	37	C4. Functional Requirements #21	Intercloud: The ability to deploy and operate analytical activities across nodes on multiple private/public cloud environments and on-premises. Inter-cloud means the ability to use data and execute operations across multiple clouds as a single logical entity	<p>a) The requirement of inter cloud is not clear, Data sources across different clouds can be federated or virtualized as a single logical entity, please clarify the requirement or the need for a intercloud</p> <p>b) What is meaning of single logical entity. Please elaborate with a business case</p>	<p>Bank is looking for capability for seamless integration with multiple cloud environments in future, If required.</p> <p>Bidders need to provide suitable solution to fulfill the requirement.</p>

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
287	43	C4. Data Ingestion #13	The vendor should be able to design solutions to handle data volumes and complexity in source data with decompression logic wherever required. An example of such workload is provided below (Workloads of this type are in scope of this engagement); A source system table with ~ 300 columns in an Oracle database generates data of ~ 800 GB on a daily basis with ~ 400 columns in DWH. 100 columns in DWH are compressed in a single column at source in Oracle. The ingestion tools should be able to perform a change data capture on source systems of this nature with run time decompression functionality.	a) Please clarify this requirement, how many source systems are of this nature that will need support for decompression, compression? b) Please clarify. The example indicates data that is concatenated + compressed. Is CDC required to decompress plus deconcatenate?	(a)presently there are 2 such source systems, in future there could be more. (b)Yes; final data should be readable and available for reporting
288	43	C4	Vendor should list out all types of risks they expect from the ingestion subsystem (e.g., dropping of data packets during ingestion, security loopholes, unprotected personally identifiable information, etc.) along with mechanisms and processes they would implement for mitigating such risks.	Please share current Network diagram of all subsystems, encryption mechanisms and the confidentiality, sensitivity nature of data transmitted	Details have been explained in Pre-bid meeting.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
289	44	C4. Data Ingestion #16	Solution should be able to handle DDL change without manual reorg/runstat.	a) Please clarify if the current solution is schema on read or schema on write, is this requirement applicable for DWH or DL? b) Will supplemental logging be enabled for all the tables to be replicated?	(a)present solution in schema on read; Bidder may propose suitable solution (b) for all tables - yes on primary key columns only
290	46	C4. Data Storage - point 11	The storage system should be robust to handle at least 1,50,000 concurrent queries (Select/DML) by processing engines / ETL jobs / end users scalable up to 6,00,000 concurrent queries in next 5 years (assuming parallelism of 100 degree).	Please clarify if the no of concurrent queries requested is across all different data layers or specific to DWH, DL?	Bank is looking for integrated solution (so combined)
291	48	C4. Data Processing Framework - point 10	Framework should have mechanism to encrypt data at rest, data at motion and data in use.	Please clarify data in use encryption requirement	Bidder should bring Industry best practice Encryption techniques.
292	39	"Architectural Requirements" on page 39 of the EOI, point no. 2	Commodity H/w with ability to integrate Private Cloud (VMWare Tanzu /ESXI) nodes into cluster for compute or storage.	We feel, in the context with a project of such a scale and complexity, Redhat openshift platform plus is more suitable. We assume bank will be open to this recommendation.	Bidder may propose suitable solution

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
293	37	C4. Functional Requirements #10	Supporting multi-programming languages for data lake and data warehouse functioning.	Please Elaborate	Along with SQL, it should support R, python, pyspark, c, c++, java based interfaces as well as programs etc.
294	37	C4. Functional Requirements #13	Ability to relate data across diverse type of data environment.	Please Elaborate	data virtualization and federation capability
295	38	C4. Functional Requirements #23	Governance: The ability to forecast, budget usage, monitor and control costs by throttling, workload or user prioritization or other means.	Is this referring to Finops (Or cost governance)	From compute and resources perspective
296	39	C4. Architectural Requirements #1	Cloud Native (Container Clusters, with on-demand instances across onpremises/private cloud/public cloud)	Which components of DWH do you want to be cloud native?	bidder need to propose suitable solution
297	42	C4. Data Ingestion #5	Solution proposed for Data Ingestion should be compatible to ingest and replicate data simultaneously on all available region	Elaborate whether you are having distributed and decentralized database	Ability to replicate data ingestion in PR and DR regions simultaneously.
298	43	C4. Data Ingestion #11	An alerting report and monitoring utility about the ingest pipelines should be available as part of the solution.	What are the monitoring and alerting scenarios expected?	bidder need to propose suitable solution

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
299	43	C4. Data Ingestion #15	Proposed solution should be able to scrap encrypted log, capture Metadata changes at source level completely, scrapping 4000-5000 logs daily having log size of ~ 32 GB each scalable up to 20000 logs. Proposed solution should be capable of scrapping logs generated by any type of Database. E.g. Oracle Database, IBM DB2 Database etc.	Please share the projected PEAK log throughputs for each source database.	Change Data Capture (CDC) should be able to scrap encrypted logs as well from TDE enabled database.
300	45	C4. Data Processing Framework #4	Framework should be able to put the processed data in a predefined area separate from the raw data, while minimizing data replication.	Please elaborate "while minimizing replication"?	Minimize copies of data across eco-system (looking for integrated solution). Presently multiple copies of data is spread across DWH, DL, BID reporting databases. Proposed solution to eliminate/minimize such data movements. Bidder can propose solution on enhancing present data model as well.
301	46	C4. Data Processing Framework #7	Should have failure, retry, alert and escalation logic defined. Successful file transfer should also be automatically acknowledged	Acknowledgement to be logged or emailed?	feasibility of both to be proposed.
302	46	C4. Data Processing Framework #9	Automatic recovery of data after failure/rejection of record needs to happen without any manual intervention	Is this to do with insertion of data to DWH or some file transfer mechanism	Both

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
303	46	C4. Data Processing Framework #11	ETL/ELT tool for data extraction should be AI/ML features for suggesting / improving Query / ETL / ELT Stages	Are the suggestions expected real-time or offline?	both
304	53	C4. Data Archival and Backup #1	Data older than specific duration as identified by Bank to be archived in low cost cold storage. Changing data archival rules should be easily configurable. Vendor to propose solution for the same with cheap and flexible storage and processing	Please specify the retention periods for archival	details will be shared with finally sected bidder
305	55	C4. Monitoring Dashboard #3	Drag and Drop feature to customize dashboard	Please elaborate on what features. What is the purpose of this dashboard?	Integrated BI tool is in the scope of this EOI. Bidder should propose BI tool for monitoring purpose with drag and drop features.
306	56	C4. Monitoring Dashboard #8	Controlling threshold breaches in pre-defined standard ways	Please elaborate of what type of threshold breaches	Dashboard tools proposes should have such capabilities.
307	57	C4. Data Governance #1	Provide traceability – it should be possible to track and visualize any data transformation or any rule applied to data in the ISDLDW eco-system	Apart from transformation and source to target mapping, is any other traceability needed?	Bidder may propose suitable solution / ability.
308	58	C4. Data Governance #10	Provide detailed information about interfacing (upstream and downstream) applications.	Please specify what information is needed	Bidder may propose suitable solution based on their implementation experience



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
309	61	C4. Data Quality #14	Parallel processing: Data quality tool should be able to handle 500 concurrent users running any kind of job (e.g. Data Profiling on small/medium/large tables as advised by Bank in the tool)	Please specify the job complexity breakup and definition	Details will be mentioned in RFP
310	61	C4. Data Quality #15	Each month 5 data quality use cases will be developed and implemented on Next Gen DW. Examples of use cases are as given below: <input type="checkbox"/> Profiling of Customer Master table for verifying PAN, Mobile Number, Date of Birth, Address, Pin Code, etc <input type="checkbox"/> Profiling of Branch Master for verifying branch address, contact information, branch manager/staff information, etc	Considering this is PII, to be discussed with Bank at the time of design	Suggestion from bidder
311	64	C4. Regulatory Reporting #1	Vendor should follow the RBI guideline in developing the solution with which it will be easier for the Bank to migrate to the element-based data reporting envisaged by the RBI.	Our assumption is RBI guidenlines can be interpreted in terms of data availability and reporting requirements based on customer data to be used for report	Bidder need to comply with all the guidelines/laws/instrcutions issued by Bank/Regulators/Govt. from time to time.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
312	64	C4. Regulatory Reporting #2	Automation –Tool should automate analytics and reporting workflow end-to-end, including all data collection, enrichment, and management, as well as all calculations, processes to final report submission. Currently 500+ jobs are being used for Tranche 1 DCT generation along with 500 more for other regulatory reports/returns.	Please clarify if the need is for ETL workflow or business users driven maker checker workflow	Bidder can propose suitable solution (tool/technology/custom solution) to address this requirement
313	64	C4. Regulatory Reporting #3	Change Management –Tool for handling ongoing change in regulation or business requirements without the need for programming expertise. On and average logic for 5% of jobs being changed monthly. Data used for regulatory reporting changes on any frequency like daily / weekly / bi-weekly / monthly, etc	Please provide example of expected change	details will be shared with finally sected bidder
314	65	C4. Regulatory Reporting #5	Data Lineage and Transparency – Tool should retrace the journey of the source data through every single workflow processes or calculations across siloes systems all the way to disclosures.	Please provide more details on lineage requirements	Data lineage should capture all data flows inside DWH eco-system (across various layers)
315	65	C4. Regulatory Reporting #6	Drill Down and 100% Auditability – Every step in the data collection, enrichment and consolidation should be tracked and recorded, including all comments.	Please provide details on steps in data collection	Data Ingestion, reject monitoring, reject processing, reconciliation, mart population, etc.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
316	65	C4. Regulatory Reporting #8	Flexibility to deal with any level of Complexity – Tool should handle the flow of information particular to any client’s organization, such as specific Key Risk Indicators (KRIs), syntaxes, data taxonomies and process operating models. Further, as demands like BCBS 239 require near-constant fine-tuning to address data governance, Business Rules should enable adjustments to regulatory and management changes with its out-of-the-box configurability and slice-and dice of data.	Share more details on taxonomy and BCBS requirements	Solution should be flexible to cater futureistic requirements from regulators like BCBS.
317	65	C4. Regulatory Reporting #9	High Volume, High Performance and Reliability – Scalable and resilient architecture which will handle all volume and performance demands.	Please provide volumetric assumptions on Users and Data volumes	bidder to propose suitable scalable architecute to handle all volume and performance demands
318	68	C4. Data Masking #5	Data Masking solution should support various methodologies like but not limited to below examples; <input type="checkbox"/> Substitution <input type="checkbox"/> Nullifying and spacing <input type="checkbox"/> Number and date variance <input type="checkbox"/> Format-preserving encryption	Kindly clarify 'Nullifying and spacing', 'Number and date variance' and 'Format-preserving encryption'	Bidder to propose suitable tool/tech for data masking compatible with the proposed solution

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
319	70	C4. Downstream Data Consumption #1	Capability to easily connect with any downstream applications currently used at the Bank and be flexible to support all end users in the Bank based on future needs.	Please clarify 'future needs'.	Bidder need to propose suitable solution to have a capability to easily connect with any downstream applications currently used at the Bank and be flexible to support all end users in the Bank based on future needs
320	70	C4. Downstream Data Consumption #5	Self-service portal to extract the data on their own (Should support Data Democratization)	Can you please elaborate in context with an use case or provide some examples?	Solution should have a capability of self service BI so that downstream users / branch users can able to extract data by their own.
321	71	C4. Data Science Platform with AI/ML Capabilities #2	Power data / objects to existing analytics models built on proprietary tools (IBM SPSS). Migration of such models to new solution	Migration to what solution? SPSS models can run on Cloud Pak for Data	Bank has no preference for any technology; Any new tech/tool proposed by Bidder should consider migration effort for existing models.Bank is open for OSS components subject to bidder will provide complete support (compliance with Bank's security requirements)
322	72	C4. Data Science Platform with AI/ML Capabilities #9	GPUs to be incorporated in solution if possible using HDFS Hadoop like environment for better analytical results	Understand this is an architectural consideration	Bidder need to propose suitable architecture / solution

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
323	72	C4. Data Science Platform with AI/ML Capabilities #13	Access management at the data, workflow and models	Please elaborate on the workflows	Details will be provided in the RFP.
324	72	C4. Data Science Platform with AI/ML Capabilities #17	Centralized monitoring and management	Please elaborate on the kind of monitoring being referred to? Is it model Monitoring for Model Risk Management?	All kinds of monitoring
325	73	C4. Data Science Platform with AI/ML Capabilities #24	Annexure E gives sample use cases which are to be implemented on Next Gen Data Warehouse using structured and/or unstructured and/or semi-structured and/or any other kind of data gathered from either Data Warehouse or Data Lake or Data Virtualization or all together or any other source.	Can you please elaborate in context with an use case or provide some examples?	there are no use cases defined in this EOI. Please refer to Appendix-C (Scope of Work).
326	74	C4. Data Science Platform with AI/ML Capabilities #27	The ability to launch and redirect training to CPU and GPU-enabled resources.	Can you please elaborate in context with an use case or provide some examples?	Details will be provided in the RFP.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
327	74	C4. Business Intelligence Tools #1	Capability to connect to various data sources: Access to various databases and file types, such as commaseparated values files, text, Excel and XML, SQL/NoSQL databases. Bank may raise specific requirements as and when required.	Can you please provide data source details?	Capability to connect to various data sources: Access to various databases and file types, such as commaseparated values files, text, Excel and XML, SQL/NoSQL databases. Bank may raise specific requirements as and when required.
328	76	C4. Business Intelligence Tools #10	Data blending: The product must permit the user to blend data from various data sources. This includes accessing the data and mapping or creating relationships with data from multiple sources.	Can you please provide more details or examples of source data that requires blending? Can you please give an exaple with a use case?	Bank is looking for capability; Bidder can propose suitable use case along with solution for data blending
329	77	C4. Business Intelligence Tools #14	Save queries and analysis: The product should enable the BI user to save the data filters, selections and drilldown paths used in decision-making process so that they can be reused.	How many users will need this functionality?	Solution proposed should have these capabilities. Bidder need to propose suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
330	77	C4. Business Intelligence Tools #18	Collaboration and social interaction: BI tools enable the creation of a business community that can share and discuss their decision-making analysis. This would include annotating analysis to share observations and social media, which can enable discussion threads or chats.	Can you please elaborate in context with an use case or provide some examples?	Bank is looking for capability; Bidder can propose suitable use case along with solution for data blending
331	78	C4. Business Intelligence Tools #19	Storyboarding. Business analysis often involves a process or workflow to analyze different data from different perspectives. Storyboarding enables a series of reports or visualizations to be tied together in a workflow that can be shared.	Can you please elaborate in context with an use case or provide some examples?	Bank is looking for capability; Bidder can propose suitable use case along with solution for data blending
332	78	C4. Business Intelligence Tools #22	In-memory analytics: The product should pull data into an in-memory or locally cached data store preferably columnar is an increasingly popular feature that enables very fast analytics once the data is loaded.	Can you please elaborate "in memory needs" in context with an use case or provide some examples?	The proposed solution should have capability to pull data into an in-memory or locally cached data store preferably columnar is an increasingly popular feature that enables very fast analytics once the data is loaded.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
333	78	C4. Business Intelligence Tools #23	Offline updates: BI tools, when storing copies of the source data in an online analytical processing (OLAP) cube or in-memory columnar data store, should enable business users to schedule automatic data updates.	Can you please elaborate "offline update" in context with an use case or provide some examples?	Data refresh should happen in automated fashion. Bidder can propose suitable solution
334	79	C4. Business Intelligence Tools #25	Business intelligence platform administration: Although all BI software should provide code and version management, there are many application development features, such as team development and user administration, that are useful for larger BI deployments.	Can you please elaborate in context with an use case or provide some examples?	Bank is looking for this capability; Bidder can suitable solution
335	79	C4. Business Intelligence Tools #26	Ease of analytical use: There should be different criteria defined for each type of user, such as information consumer, business analyst and IT.	How many users will need self service?	Details will be provided in the RFP.
336	79	C4. Business Intelligence Tools #27	Ease of creating BI applications: There should be different criteria for each type of analytics creator, such as business analysts and IT.	How many users will need self service?	Details will be provided in the RFP.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
337	79	C4. Business Intelligence Tools #28	Speed of access: Query performance will vary based on the complexity of the queries and the amount of data involved. Dashboards with multiple visualizations will need to get query results from many queries. The best practice is to create several prebuilt query scenarios and compare how each product performs based on these specific examples. The worse practice is to just arbitrarily rate the speed.	Can you please elaborate more on what is the ask here?	more details will be provided in the RFP
338	79	C4. Business Intelligence Tools #29	The best practice is to establish a testing environment to determine scalability in terms of both the number of concurrent users and data metrics, such as volumes, variety and veracity.	Please elaborate on requirements for testing	more details will be provided in the RFP
339	80	C4. Business Intelligence Tools #30	There should be separate criteria for BI user versus administration training. Training may include in-person classes, online classes -- live or prerecorded -- or web recordings for specific features or processes.	Please provide more details on training requirements	more details will be provided in the RFP
340	80	C4. Business Intelligence Tools #31	There should be separate criteria for BI user online help versus technical documentation.	Please elaborate on "separate criteria" ask.	Online help for business users should be in functional/business terms so that user can go ahead their self service requirements

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
341	80	C4. Business Intelligence Tools #32	Ability to handle and summarize huge volumes of data. E.g. 30-40 million rows accessed on index and summarized over 5 to 8 metrics.	Please provide more details on source data structure and framework. Can you please elaborate in context with an use case or provide some examples?	more details will be provided in the RFP
342	General	General	General	<p><b>We request SBI to please include clause for credentials/ past work experience.</b></p> <p>Experience of the Parent/Holding Company/its subsidiary or subsidiary companies/its branch offices/its associates/member of same network or member of same global firm shall also be acceptable provided a certificate is given by the bidding entity mentioning that they would be able to draw international resources/expertise from them as and when required to fulfil commitments of this tender.</p>	EOI terms are final

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
343	31	Appendix B (Point 5)	Copy of the order and / or Certificate of completion of the work. The Bidder should also furnish user acceptance report from client(s).	The requirement of certificates from all banks/ customers will be difficult to execute as all consultants, including us will be bound by confidentiality requirements and will be restricted from disclosing information about our past work. Would request you to accept "self-certification supported by CA certificate" instead of "work orders" as we work with NDA with our clients.	Please refer to Corrigendum-2.
344	34	C2.1 Goals and Objectives	Capabilities of data ingestion from all source system available with the Bank or any future standard source systems that the Bank will decide to use with high throughput and low latency of the Bank in automated manne	How many source systems are supposed to be integrated with the new target solution?	Please refer to section C3.2 at page no 35
345	35	C2.1 Goals and Objectives	Support for effective data governance for data discovery, data modelling, data quality, data catalogue, data profiling, data reconciliation and data lineage	Is there a data dictionary and critical data elements available with bank?	Yes. Data dictionay is available.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
346	35	C2.1 Goals and Objectives	Self-service BI reporting by end users	Is there a count on the number of reports required by end users?	Bank expectation is capability of self-service BI reporting.
347	36	C4. Functional Requirements	Support of Artificial Intelligence, Machine Learning, Business Intelligence, Data Engineering, Data Management features together on a single platform	Is there a need to use Large Language Models / Natural Language Processing in the target solution?	Yes. Bidder to suggest possible best approach
348	55	Monitoring Dashboard	Drag and Drop feature to customize dashboard	Is bank open to explore Low Code No Code platforms also in the target solution?	Yes.
349	40	C4	Proposal Requirements	Thoughts on implementing a multi-cloud solution. Are you open to considering multiple options for a cloud solution?	No. Bank is expecting solution on-prem / internal cloud.
350	40	C4	Proposed hardware specification	whether the requirement mentioned in the outline will still apply if we go with the proposed cloud solution. Is there any additional information or clarification that can be provided on this matter?	No. Bank is expecting solution on-prem / internal cloud.
351	40	C4. ii	Proposed internal and external network specification.	whether the requirement mentioned in the outline will still apply if we go with the proposed cloud solution. Is there any additional information or clarification that can be provided on this matter?	No. Bank is expecting solution on-prem / internal cloud.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
352	40	C4 vii	Rackspace, Power, Cooling, Network connectivity & Bandwidths	whether the requirement mentioned in the outline will still apply if we go with the proposed cloud solution. Is there any additional information or clarification that can be provided on this matter?	No. Bank is expecting solution on-prem / internal cloud.
353	51		Migration on ETL pipeline	What technologies are used currently for ETL pipelines What is meant by Analytical models here - MDDDB like Cubes or Machine Learning models or both? What are the technologies involved in these areas already used by the Bank?	Details has been presented in pre-bid meeting
354	40		Architectural Requirements	Are we looking for primarily OSS components only for all components?	Bank is open for OSS components subject to bidder will provide complete support (compliance with Bank's security requirements)
355	51		Data migration from existing archival solution to new one	Can you provide additional details?	Data migration from present archival system to new archival system as proposed.
356	35		Self-service BI reporting by end users	Is there one or more BI reporting tools currently being used by the bank users?	Yes.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
357	14	14(v)(b)	Bidders are obliged under code of integrity to Suo-moto proactively declare any conflicts of interest (pre-existing or as and as soon as these arise at any stage) in EOI process or execution of contract. Failure to do so would amount to violation of this code of integrity.	Bidder requests for the Bank to provide it a copy of the code of integrity for the Bidder's review and confirmation.	EOI terms are final.
358	21	14(v)(b)	<p>The Bidder fails to abide by the terms and conditions or to maintain the required technical/operational staff/equipment or there is change in its production/service line affecting its performance adversely, or fails to cooperate or qualify in the review for empanelment;</p> <p>Employs a Government servant or the Bank's Officer within two years of his retirement, who has had business dealings with him in an official capacity before retirement; or</p>	Bidder requests for deletion of this two sub-clauses. Any failure to abide by the terms and condition under this EOI should not lead to debarment of the Bidder (including their related entities) for a period (one to two years) from the Bank's procurements including removal from empanelment.	EOI terms are standard and can not be changed

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
359	28	Appendix-A: BID FORM (TECHNICAL BID)	It is further certified that the contents of our Bid are factually correct. We have not sought any deviation to the terms and conditions of this EOI. We also accept that in the event of any information / data / particulars proving to be incorrect, the Bank will have right to disqualify us from the EOI or any subsequent bidding process without prejudice to any other rights available to the Bank.	Bidder requests for modification of this clause. The bidder should be allowed to give deviations to the terms and conditions of this EOI.	EOI terms are standard and can not be changed
360	23	TENDER FEE:	Non-refundable Tender Fee should be directly credited to the designated account as mentioned in Schedule of Events. Proof of remittance of Tender Fee in the designated account should be enclosed with the technical bid. The Bids without tender fee will not be considered valid.	Tender fee to be paid through demand draft	for this EOI, Bank is not charging any tender fee.
361	34	C 2.1	Optimization of solution / tool (s) for Data Federation/Virtualization to ensure seamless integration of data in real time when stored in multiple sources without physical movement of data sets for the purpose of reporting / analytics	d) Data Federation: Do we still need Data Virtualization/Federation if all the analytics needs can be served with a single platform covering Data Lake, Data warehouse, Operation DB, and Object storage ?	In future data may be required to be federated across multiple source systems.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
362	35	C 3.2	Volumetric of Existing Data Warehouse	The data size given for Production DB is compressed or raw. Can we get the raw data size? Similarly, we would need the same for Archival and Data Lake	Compressed with factor of ~4
363	35	C 3.2	Volumetric of Existing Data Warehouse	Can we get the HW configuration for the current setup, no of servers and core count, memory count, etc?	Details has been presented in pre-bid meeting
364	36	C 4		Functional Requirement: It's mentioned Bank is planning to upgrade the existing Data Warehouse and Data Lake solution to a modern and state-of-the-art integrated solution. In addition to project mandates mentioned in section C3, the solution should provide the following additional capabilities: Instead of upgrading, Is the bank open to replacing the existing setup completely with modern capabilities built into it?	Bank is open for replacement of existing setup completely.
365	43	15		The proposed solution should be able to scrap encrypted log, What does this mean, please elaborate	Change Data Capture (CDC) should be able to scrap encrypted logs as well from TDE enabled database.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
366	46	7		Processing engines should have capabilities to process all kinds of Data Storage formats used by the Bank. Can you please elaborate on the Data formats used.	All prevailing formats available in the market should be supported.
367	31	7	Bidder should have experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake preferably in BFSI sector.	Request rephrase to Bidder should have experience in implementation or managing large-scale (~petabytes) enterprise data warehouse / data lake in BFSI sector.	EOI terms are final
368	50	Migration from Existing Setup to Proposed Solution	"Vendor should provide a feasible plan for best use of existing infrastructure which is procured during last 10 years in staggered manner during the implementation of ISDLDW which will save cost to the Bank. (Annexure D gives the technology architecture of the current setup)"	Based on the point, please confirm if SI can re use the the existing ETL, DQ, Data Governace, BI tools in the suggested architecture and only replace the DWH/DL ( IBM IAS) and SPSS (AI/ML Model build and Operationalisation)?	Bidder need to propose best suitable solution in the EOI. Existing infrastructure/components can be used only after sunset of existing setup subject to not reachig end of life.
369	36-39	C4		Does requirement include development for onboarding new sources (semi structured, unstructured, real time, streaming etc that is not part of existing warehouse)??	Yes
370	35-36	C3.2		What kind of data model is there in existing data platform?	Mostly ditional and customized.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
371	35-36	C3.2		What is data ingestion/ ETL tool in current data platform? - to find accelerator for re-engineering	Details have been presented in pre-bid meeting
372	36-39	C4		Are there any AI/ ML use cases to be implemented in new platform? or just ability to develop on AI/ ML use cases?	Yes. Other than migrating existing models, platform should provide the capability to develop, implement and monitor new use cases
373	36-39	C4		Is data profiling required? Or DQ issues to be handled are known and prioritized?	Yes.
374	36-39	C4		What is tentative timeline for Platform set up and migration?	bidder to propose considering migration efforts along with platform implementation.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
375	34-35	C2.1	"Migration of existing ETL pipelines / jobs and analytical models"	<p>Is it the expectation to migrate all of the existing ETL jobs into a newer suggested platform.</p> <p>If so provide the following details</p> <ol style="list-style-type: none"> <li>1. What is the ETL tool currently used</li> <li>2. There is a mention of 27000 ETL jobs in the volumetrics. Is the expectation to migrate all of these jobs into the newer platform</li> <li>3. Does these ETL jobs carry out only Ingestion or does it perform data quality, data processing tasks as well</li> <li>4. Does these ETL jobs cover ingestion of unstructured data as well as streaming data as well</li> <li>5. Is the expectation to convert these 27000 ETL jobs one to one into a newer platform ( Lift and Shift) or re engineering and optimization to be considered</li> </ol>	<ol style="list-style-type: none"> <li>1) Details have been presented in pre-bid meeting</li> <li>2) YES (bidder can propose approach)</li> <li>3) Both (ingestion as well as DQ etc.)</li> <li>4) No (at present no unstructured data source)</li> <li>5) YES (Optimization and re-engineering to be proposed)</li> </ol>
376	35-36	C3.2	Number of Cognos Reports - 100	<p>Does the scope include to migrate these 100 Cognos reports into a newer BI platform. Please confirm. Also please confirm if there are new reports to be built as well apart from migration of these reports</p>	<p>Migration - Yes</p> <p>New Reports developments - Not as part project implementation but as part of maintenance phase</p>

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
377	35-36	C3.2	Total Tables ~50000	Does the scope include lift and shift of these tables into a newer data platform or does it involve rationalization and re engineering of the data model itself. Please explain	YES (Optimization and re-engineering to be proposed)
378	34-35	c2.1	Optimization of solution / tool (s) for Data Federation/Virtualization to ensure seamless integration of data in real time when stored in multiple sources without physical movement of data sets	Is the expectation to Data virtualization between Structured and Unstructured data. What kind of use cases to be implemented.	Bidder to suggest Data virtualization possibilities between all types of data. Details will be provided in the RFP
379	31	Appendix B	Bidder's Eligibility Criteria	Request to accommodate below changes in clause:  Bidder/OEM should have experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake preferably in BFSI sector.	EOI terms are final

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
380	31	Appendix B	Bidder's Eligibility Criteria	Request to accommodate below changes in clause:  Client references and contact details (email/ landline/ mobile) of customers for whom the Bidder/OEM has executed similar projects in India/Global. (Start and End Date of the Project to be mentioned) in the past (At least two client references are required)	EOI terms are final
381	71		"Power data / objects to existing analytics models built on proprietary tools (IBM SPSS). Migration of such models to new solution "	How many models exist today. Ideally these models have to be re built in the newer platform for migrating these from the existing ones. Please confirm the understanding	Approx. 100 models are presently running. Details will be provided in the RFP.
382	37	C4 Pt.18	"Compatible with Oracle CDC (Goldengate) at near real-time"	In the existing current architecture Is Oracle CDC is being used or IBM CDC . Please clarify	IBM CDC

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
383		Generic		<p>Please clarify if the scope is only to replatform only the existing DWH platform from IBM to newer solution and retain all the existing ETL, DQ , Data governance, BI, AI/ML tools of IBM by repointing these tools into the newer DWH/DL solution</p> <p>or</p> <p>Is the expectation to replatform and migrate all of ETL, DQ, Data Governance, BI , AI/ML tools into a newer tool set solution along with DWH/DI</p>	Expectation is, bidder to propose any suitable platform with migration efforts
384		Generic		Can we leverage the existing Tivoli backup and recovery solution in the new architecture as well.	Bidder can propose the same in the EOI proposal
385		Generic		BCP solution , should we align with the existing SBI solution or is it the expectation to provide a newer BCP solution. Please Confirm	Align to Bank's BCP requirements

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
386	31	6	Client references and contact details (email/ landline/ mobile) of customers for whom the Bidder has executed similar projects in India. (Start and End Date of the Project to be mentioned) in the past (At least two client references are required)	Bidder values confidentiality of information of its clients. At this point we will not be able to furnish the details of the individuals who will be entitled for reference checks. All reference checks are done through Peer Advantage. Should SBI asks for a reference check, LTIM through Peer Advantage will be happy to assist.	EOI terms are final
387	31	6	Client references and contact details (email/ landline/ mobile) of customers for whom the Bidder has executed similar projects in India. (Start and End Date of the Project to be mentioned) in the past (At least two client references are required)	Please specify if global references of similar nature are acceptable?	Please refer to Corrigendum-2.
388	31	6	Client references and contact details (email/ landline/ mobile) of customers for whom the Bidder has executed similar projects in India. (Start and End Date of the Project to be mentioned) in the past (At least two client references are required)	Please let us know if the references can be a mix of the software OEM and the SI partner involved in the bid?	Yes (for Bidder has executed similar projects)
389	35	C3.2	Volumetric Details	Segregation of hot/warm/cold data	details will be provided in RFP phase
390	35	C3.2	Volumetric Details	Can SBI throw light on the incremental growth for the next 3-5years?	It would be refined in RFP phase but expecting volume of 10 PB.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
391	35	C3.1	Present Architecture of Data warehouse	Requesting SBI to give us an understanding of the current architecture and lay of land	Details have been explained in Pre-bid Meeting.
392	39	C4, 36	Catalogue of pre-packaged ML use cases	Need more clarification of Pt. 36 pg-39: "Catalog of pre-packaged ML use cases, models & algorithms... which can be directly used with Bank's data.."	Platform may have the capability of the pre-built models specific to financial sector which can be used by the Bank on the Bank's data
393	40	C4, 6		Need more clarification of Pt. 6 pg-40: "Note: All components must be based on open standard/open source software, customized for enterprise use (Open Core)"	Bank is open for OSS components subject to bidder will provide complete support (compliance with Bank's security requirements)
394	40		Cost Model	Would SBI looking for a term(subscription based alias yearly) or perpetual license?	Bidder can propose suitable approach mechanism based on their solution



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
395	41	C4.e	Performance Benchmark: Performance benchmark for proposed solution with comparable solution available in the market	We strongly believes that benchmark should be done on customer data. This would ascertain the capability of the platform in handling scale/complexity. We are happy to do a POC on SBI data set and prove our point. We typically does performance benchmark on customer data based on request. Hence we do not have any standardized performance benchmark as per industry standards. Is it ok if we can submit performance benchmark after white labelling the account(since we are bound by NDA with clients)?	TBD Post EOI phase
396	42		Point 6: Annexure C	Please share the details around flow of data in the present DWH	Architecture and high level data flows will be presented in the demo
397	51		Point 9: Migration	Please share details of the existing setup(architecture diagram) of the current setup	Details have been explained in Pre-bid Meeting.
398	64	1	Regulatory Reporting	Regulatory reporting Point 1-clarification of RBI guideline of element based data reporting. Please direct us to the specific clause on RBI guidelines.	details will be shared with sected bidders

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
399	73	24	Annexure E	Need Annexure E for the sample analytical use cases	Use cases has not been defined for this EOI
400	73		Point 24: Annexure E	Please share details on the sample use cases.	Existing models to be migrated and new models will be developed by the Bank In-house team
401	104	Appendix-F	Reference: Value of Work Order	Our contracts with customers are confidential. We will not be able to furnish the value of work order with SBI. However during reference checks, SBI is free to ask this question to the client. Hope this is ok?	Bidder can skip that field (if covered under any NDA with their customer)
402	41-89	C4	Functional Requirements	Is SBI expecting explanations to each of the line items or just "Yes/No"?	Yes. Bidder can also provide explanation of each line items.
403		Generic	NDA	Please share the template	Architecture diagram along with data flow has been presented in pre-bid meeting.
404		Generic	Bid submission request	EOI is quite exhaustive. Request you to grant an extension of atleast two weeks to comply with the ask of this EOI	Bid submission date has been extended till 29.01.2024
405		Generic	Evaluation Criteria	Could SBI give us insights to the evaluation criteria/scoring for this EOI?	Please refer to sr no 'v' at page no 16
406		Generic	Existing Platform	Pls let us know the end of license of the current data platform which is in use at SBI	Please refer to para C1 at page no 33

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
407		Generic	Open Source	Is the bank looking for an open source ecosystem?	Bank is open for OSS components subject to bidder will provide complete support (compliance with Bank's security requirements)
408		Generic	Other components	Can the OEM use opensource components to cater to the bank requirements or the solutions have to be proprietary	Bank is open for OSS components subject to bidder will provide complete support (compliance with Bank's security requirements)
409		Generic	Existing Infrastructure	Is the bank expecting the selected vendor to make best use of existing infrastructure?	Yes, (provided sufficient support is available by OEM on the product and it is well suited to proposed solution)
410		Generic	Existing applications	Are the vendors expected to use existing applications or integrate with existing applications	Existing application can be used provided sufficient support is available by OEM on the product and it is well suited to proposed solution
411		Generic	Architecture	There are many components in EDW Architecture , If SBI team can highlight 1) which all component this ( EOI/RFP in future ) should replace and 2) which all component the given solution should integrate.	Bidder will suggest best approach as per their solution. List of major components used by the Bank will be presented in pre-BID meeting

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
412	63	Appendix-C: Technical Criteria/Scope of Work C4. Functional Requirements Security and Compliance - Point No. 6	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	<p>To ensure Availability under CIA triad, it is essential to deploy a dedicated On-Premise DDoS Protection. On-Premise DDoS Protection solution detects, alerts and protects against Volumetric, State Exhaustion, Application Layer and Outbound DDoS Attacks and thus ensures Availability of Critical Systems, Services and Applications in Data Center.</p> <p>So, could you please confirm if you need On-Premise Anti-DDoS Protection Systems ?</p>	DDoS protection is not required for this solution, but bidder need to integrate with proposed solution.
413	63	Appendix-C: Technical Criteria/Scope of Work C4. Functional Requirements Security and Compliance - Point No. 6	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	<p>As a part of Solutions to identify and alert on Data leakage, it is essential to deploy a packet based Network Detection and Response (NDR) Solution. A packet based NDR enables Enterprises to achieve complete security visibility into their network and identify vulnerabilities and threats with high accuracy.</p> <p>So, could you please confirm if you need a Network Detection and Response Solution ?</p>	Proposed solution should be integrated with Bank's centralized security Solutions. No additional solution is required for this purpose.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
414	56	Monitoring Dashboard Point No. 7	Industry best tools for performance (slowness) monitoring	As Industry best practices the tool or solution should be agent less as it will be non-intrusive and dose not impact on performance on the Servers or the network.	Bidder to propose suitable approach for the tech stack.
415		Monitoring Dashboard Point No. 9	End to end performance monitoring with automatic alerts	End to end visibility from client to Application and the backend database visibility is possible via traffic monitoring solution with automatic alerting.	Bidder to propose suitable approach for the tech stack.
416	57	Monitoring Dashboard Point No. 21	DB health monitoring tool as well as monitor of N/W health	Monitoring of DB application health and network health will be possible only via Application aware Network Performace monitoring (Traffic based Agent less solution) for this requirement please provide the capacity/ bandwidth for network traffic monitoirng.	Infrastructure and Health and performarncce monitoring is required. Network Monitoring is not required.
417	Page 37	Functional Req, Point 13	Ability to relate data across diverse type of data environment.	Please elaborate more	Data Federation/virtualization capability across hetrogenous data systems and data types
418	Page 37	Functional Req, Point 15	Analytics: The product's ability to perform advanced analytic operations within the dbPaaS	Please elaborate more	Bidder to propose suitable solution based on on-prem.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
419	Page 35	C3.2 , first 1 PB ,second 1 PB and third points 730 TB in the table	Volumetrics of existing Datawarehouse	The data size given for Production DB is compressed or raw. Can we get the raw data size? Similarly, we would need the same for Archival and Data Lake	Compressed with factor of ~4
420	Page 71	Data Science Platform with AI/ML capabilities	Which all use cases to be developed	Please share the Annexures	Existing models to be migrated and new models will be developed by the Bank In-house team
421	Page 35	C3.1	Present Architecture of DWH	Please share the existing Architecture and Tech specifications , Can we get the HW configuration for the current setup, no of servers and core count, memory count, etc? (Preferable IBM Appliance Details including Datastage)	Details have been explained in Pre-bid Meeting.
422	Page 36	9,10	Scheduled Workload (jobs per day), Ad-hoc Workload (jobs per day)	Please share the breakup of Jobs (% wise) on following bais: Complex, Medium and Light	SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experiece and may vary depending on future load /requirement

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
423	Page 36	11	Number of Cognos Reports	Please share the breakup of reports (% wise) on following bais: Big, Medium and small	SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experiece and may vary depending on future load /requirement
424	Page 59	1,5		How much Data will be processed in total for Data Quality per day? How much Data will you be Profiling per day?	At present - Profiling is done on need basis. Data Quality rules are executed on daily basis.
425	Page 34	C2.1 D	Goals and Objectives	Data Federation: Do we still need Data Virtualization/Federation if all the analytics needs can be served with a single platform covering Data Lake, Data warehouse, Operation DB, and Object storage ?	In furure data may be required to be federated across multiple source systems.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
426	Page 36/37	C4 / C3	Functional Requirements	Functional Requirement: It's mentioned Bank is planning to upgrade the existing Data Warehouse and Data Lake solution to a modern and state-of-the-art integrated solution. In addition to project mandates mentioned in section C3, the solution should provide the following additional capabilities: Instead of upgrading, Is the bank open to replacing the existing setup completely with modern capabilities built into it?	Yes. Bank is open for replacement of the existing setup. Bidder can propose suitable solution in their EOI proposal.
427	Page No 43	15 Data Ingestion	Data Ingestion	15) Page no., 43 . The proposed solution should be able to scrap encrypted log, What does this mean, please elaborate.	Change Data Capture (CDC) should be able to scrap encrypted logs as well from TDE enabled database.
428	Page No 46	7	Data Ingestion	Processing engines should have capabilities to process all kinds of Data Storage formats used by the Bank. Can you please elaborate on the Data formats used.	All prevailing formats available in the market should be supported.
429	Section C1	33		Is it reasonable to understand that IBM IIAS contact would expire by Feb2025	Yes.
430	Section C2	35		IS the new DataLake and DWH expected to retain the same data structures as they are now	Bidder is expected to propose suitable solution in their EOI proposal



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
431	Section C2	35		Migration of existing ETL pipelines / jobs and analytical models - What ETL tools are used, and what are the analytical models deployed - Kindly share details	Details have been presented in pre-bid meeting
432	Section C4	37		Data Cleansing - Given that the data is currently in DWH, what sort of data cleansing is expected	Bank expectation is to minimise data redundancy and improve quality.
433	-	73	point 24	There is reference to sample use cases in Annexure E yet Annexure E In Page 102 is about pre bid queries	Use cases has not been defined for this EOI
434	C2. Project Mandate	34	C2. Project Mandate - point a	Please elaborate on the requirement for NFS storage in the solution.	NFS storage is required by Bank to store data/files. Bidder can propose any better solution with their EOI submission.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
435	C2. Project Mandate	34	C2. Project Mandate - point e	Request you to please help relax the condition of on-prem/banks own cloud to make it open for public cloud/hyperscaler based DW solution as it will be able to scale to the bank's requirements and also be secure than having on-prem solution. The cloud based platform will be easy to manage as well as give bank lots of capabilities which may be difficult to deploy and scale in on-prem solution.	EOI terms are final
436	C3.2 Volumetric of Existing Data Warehouse:	35	-	Please confirm if number of ETL jobs are 27000? Are these monthly/yearly run numbers or unique ETL jobs? Also confirm on file types to be supported in Data lake of 730 TB.	27000 number of scheduled and ad-hoc. Data lake presently supports all types of preveilig file types. Further details will be provided in RFP document.
437	Architectural Requirements	39	-	Request you to please help relax this requirement - Commodity H/w with ability to integrate Private Cloud (VMWare Tanzu /ESXI) nodes into cluster for compute or storage as this restricts the platform choice which can benefit SBI.	Bidder to suggest possible options of this requirement. Details will be provided in the RFP

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
438	Architectural Requirements	39	-	Request you to please help confirm if SI can provide hybrid cloud/Cloud at customer offerings which will ensure that SBI gets benefit of extreme performance, scalability and ease of management without any security concerns. It will be OPEX model of cloud consumption.	EOI terms are final
439	EOI evaluation process	16	12. BIDDING PROCESS/OPENING OF TECHNICAL BIDS:	Request you to please help share any scoring methodology for EOI responses to shortlist SIs for RFP? This will ensure that every SI gets best fitment opportunity to present solution basis the methodology.	Please refer to sr no 'v' at page no 16
440	Annexures	-		Request you to please help share Annexures A-F which are not available online on SBI website.	Please refer to Appendix-A to Appendix-G. Annexures are not present in this document.
441	References	31	pt 7 - Customer references	Request you to please relax the requirement to have petabyte scale DW requirements in India as it may restrict competition. This should be relaxed to have similar scale customer refereneces worldwide - so global deployments should also be ok for response to EOI.	Please refer to Corrigendum-2.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
442	Self-Certification of Local Content	105	-	This self certification may not be applicable to cloud based platforms as its available as a service. So, this certification should not be applicable to the software and hybrid cloud/cloud at customer platforms.	Compliance of local content is based on total cost of project
443	5. ELIGIBILITY AND TECHNICAL CRITERIA/ SCOPE OF WORK:	11	(b) Either the Bidder on behalf of Principal/OEM or Principal/OEM itself is allowed to Bid, however both cannot Bid simultaneously.	As this is an EOI, request for the product/platform principal/OEM to be also allowed to independently submit EOI response, while Service Providers/System Integrators submit their respective EOI responses with same principal/OEM as part of their solution and services scope.	EOI terms are final
444	3	6	Last date and time for Bid submission 10.01.2024 (5:00 PM)	Please let us know the possibilities of extension for Last Date of Bid Submission. We would request an extension of at least 2 weeks owing to worldwide annual break for the firm. It will be difficult to get the required approvals within time.	Bid submission date has been extended till 29.01.2024

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
445	31	5	Bidder should have experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake preferably in BFSI sector.	Will experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake of Non BFSI & Network client club will be considered?	EOI terms are final
446	34	C2.1 - h	Support for effective data governance for data discovery, data modelling, data quality, data catalogue, data profiling, data reconciliation and data lineage.	Are we using data governance tools as of now? If yes, what tools are we using? Do we have a DG Council in Place with identified roles and responsibilities? Do we need to Set up One? What are the domains in scope?	yes. Presently IBM IGC tool is being used. Yes. Bank is having DG council.
447	34	C2.1.d	Optimization of solution / tool (s) for Data Federation/Virtualization	Do we still need Data Virtualization/Federation if all the analytics needs can be served with a single platform covering Data Lake, Data warehouse, Operation DB and Object storage ?	In future data may be required to be federated across multiple source systems.
448	35	C2.1 Goals & Objectives	Migration of existing ETL pipelines / jobs and analytical models	What will be the number of analytical models in-scope for migration?	Approximate 100 models are to be migrated. Details will be provided in the RFP
449	35	C2.1 Goals & Objectives	Migration of existing ETL pipelines / jobs and analytical models	What are the names of the existing models (persistence model, prediction model and so on) , brief description and products in-scope for the models	Appropriate details will be provided in the RFP

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
450	35	C2.1 Goals & Objectives	Migration of existing ETL pipelines / jobs and analytical models	What are the methodology/algorithms (regression, neural networks, decision tree, naive bayes and so on) used in the existing models?	Appropriate details will be provided in the RFP
451	35	C2.1 Goals & Objectives	Migration of existing ETL pipelines / jobs and analytical models	What are the programming language (R, python, SAS and so on) used for training & deploying the models?	R, Python are used in addition to IBM SPSS
452	35	C2.1 Goals & Objectives	Migration of existing ETL pipelines / jobs and analytical models	What is the volume of the training data and number of databases/tables used in the existing models?	Appropriate details will be provided in the RFP
453	35	C2.1 Goals & Objectives	Migration of existing ETL pipelines / jobs and analytical models	Do we need to retrain or retune the models during migration process?	Appropriate details will be provided in the RFP
454	35	C2.1 Goals & Objectives	Migration of existing ETL pipelines / jobs and analytical models	Where are the existing models hosted? Other than input data, are there any impacts to the models during migration to next gen DwH?	Existing models are hosted in on-prem Data lake / Private cloud servers.
455	35	C2.1 Goals & Objectives	Migration of existing ETL pipelines / jobs and analytical models	What are the parameters/criteria wrt analytical models for evaluating successful migration?	Appropriate details will be provided in the RFP

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
456	35	C3.2	Volumetric of Existing Data Warehouse	The data size given for Production DB is compressed or raw. Can we get the raw data size? Similarly, we would need the same for Archival and Data Lake also. Can we also get the HW configuration for the current setup, no of servers and core count, memory count, etc?	Datalake - no compression.
457	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	Is your current deployment of Data Lake inline with PROD , UAT, DEV and DR having shared Storage and DB Servers or separate Storage & DB servers across all environments?	Present data lake setup has only PR and small size UAT.Appropriate details will be provided in the RFP
458	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	Which are the current DB servers OEMs (My SQL, Oracle, MS SQL, Postgre SQL, etc?)	IBM DB2
459	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	Is the current set-up Active-Active or Active-Passive for all environments?	At present, DR is not fully operational
460	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	What is the current Network Architecture for the Data Lake deployment traffic flow is from Perimeter Layer, MZ or DMZ Zones?	Present data lake has been implemented within Bank's private network.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
461	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	Is there any Loadbalancer at Newtork layer for traffic balancing or it has been set-up from proxy server for existing Data Lake set-up?	Load balancer is available in data lake
462	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	Is the MPLS and ILL having Dual connectivity inline with existing Data Lake set-up ?	not applicable
463	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	Is the MPLS connectivity is Hud & Spoke and Full Mesh?	not applicable
464	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	Is the current Data Lake set-up is being public facing or traffic is routed from Public to private network?	Present data lake has been implemented within Bank's private network.
465	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	For all new PROD, DEV, UAT and DR set-up are you looking for HA pair components that will be working on Active-Active or Active-Passive mode?	Yes. HA in active-active mode.
466	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	Will the connection between DC and DR for the new Data Lake set-up will be on Active-Active or Active-Passive Mode?	Yes. in active-active mode. Also, bidder can propose better solution/approach.
467	35	C2.1 Goals & Objectives	Dev, UAT, Production and DR setup to be done along with sandbox for analytics.	What is the bandwidth you are looking for P2P connectivity between DC and DR for data Lake or what is the existing P2P bandwidth between DC and DR for data lake?	At present, DR is not fully operational



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
468	36		Business Intelligence Tools	What is the expected no of reports and dashboards (including complexities) to be built apart from the migration of 100 reports build in Cognos	Migration of existing reports - Yes New Reports developments - Not as part project implementation but as part of maintenance phase
469	36	C4	Functional Requirements	It's mentioned Bank is planning to upgrade the existing Data Warehouse and Data Lake solution to a modern and state-of-the-art integrated solution. In addition to project mandates mentioned in section C3, the solution should provide the following additional capabilities: Instead of upgrading, Is the bank open to replacing the existing setup completely with modern capabilities built into it?	Yes. Bank is open for replacement of the existing setup. Bidder can propose suitable solution in their EOI proposal.
470	43	15	Proposed solution should be able to scrap encrypted log	What does this mean, please elaborate.	Change Data Capture (CDC) should be able to scrap encrypted logs as well from TDE enabled database.
471	46	7	Processing engines should have capabilities to process all kinds of Data Storage formats used by the Bank.	Can you please elaborate on the Data formats used	All prevailing formats available in the market should be supported.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
472	51	8	Migration of Data Governance, Data Lineage and Data Quality rules and policies	What is the scope of migration( quantum of data )? For Data Governance, Lineage and Data quality	Bank is expecting for migration of data quality rules and data lineages from existing tools (technology) to the solution proposed by bidder ensuring continuity in the metadata availability.
473	57	1	Provide traceability – it should be possible to track and visualize any data transformation or any rule applied to data in the ISDLDW eco-system	Please provide inputs on any traceability feature preference - Lineage Visualization (end-to-end lineage visualization from source to target capturing all logic), Automated Scanning Capability, Setp-wise Debugging (Capability to debug various points and hops across the data lineage diagram), Ability to scan both structured and unstructured datasets, Automated lineage flow update.	No preferences. Bidder to propose appopriate solution.
474	57	2	Provide trust – The system should be able to ensure the users that they are accessing data from the right source of information.	Is there an exitsing single source of truth or primary trusted system/solution already mplemented or used? If yes, please provide details about the data domains in the system. If not, is there an existing MDM solution already implemented ?	SSOT has been defined for critical data points.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
475	58	9	Capability to review the conformance to governance policies and processes periodically (monthly/quarterly) and make necessary changes.	Do we have existing DG policies? Do we need to define new policies?	Yes. Yes, as per requirement.
476	58	10	Provide detailed information about interfacing (upstream and downstream) applications.	Please provide list of source systems and consuming system/applications for assessing the connector/scanning capability.	Details have been presented in pre-bid meeting
477	58	7	Automated propagation of changes to ISDLDW ecosystem Data Dictionary and business glossary by multiple sources as and when changes occur in source.	Are there any existing Business Glossaries defined for key reports, data domains, mater data? If yes, how are they currently mainatined and updated? Is there an SOP and/or policy defined for creation and maintainence of business glossary and data dictionary?	Yes.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
478	58	8	Capability for purging/achieving of tables/jobs/reports which are not in use.	Please specify how are 'the tables/jobs/reports not in use' currently handled in the existing systems/applications/DBs. Are the dormant tables/jobs/reports purged or archived on periodic basis? Is the impact assessed on the other associated systems/applications/dbs performed before purging of dormant/not in use tables? Are there any existing SOPs and/or policies available for archival and purging of tables/jobs/reports?	Process has been defined and being followed.
479	58	12	Metadata Management Capability: Tool should cater to three broad categories of metadata; Business metadata, Technical metadata and Operational metadata	Are Critical Data Elements across key data domains and/or reports already identified? Are technical and operational metadata across all applications, systems or outsourced/third party systems metadata have restricted access to SBI team?	Yes. Solution should be capable to capture such information.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
480	59	1	Vendor should propose end-to-end solution for Data Quality Management starting from data origin till the data consumption. These tool (s) to be used for addressing various aspects of the data quality problem mentioned below on SBI data set during data ingestion, data processing or data consumption as advised by Bank on case by case basis	We assume we are asking for a new DQ tool implementation? Are we using any DQ tool currently? What is the number of CDEs in scope?	Bank is presently using Informaton Analyzer (IA) tool. Rules will be mentioned in RFP.
481	59	2	Parsing and standardization — Decomposition of text fields into component parts and formatting of values into consistent layouts based on industry standards, local standards	How many DQ rules are currently in place which needs to be migrated?	Appx 100
482	60	7	Recommend Enrichment — Enhancing the value of internally held data by appending related attributes from external sources	How many external sources need to be integrated?	Details have been presented in pre-bid meeting
483	60	8	Data quality dashboarding - A data quality dashboard can aggregate the status of continuously monitored data quality rules, as well as generate alerts to notify data stewards when they need to address an issue.	Dashboards to be developed in the DQ tool itself or in any other platform like Power BI/Tableau?	Bidder should propose suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
484	60	10	Identity resolution - Identity resolution is the process of linking various records and is the main engine for record de-duplication, which can enable some aspects of data cleansing.	How many new identity resolutions are part of the scope?	Number can not be ascertained at this stage.
485	61	13	Dedicated Data Quality team who will thoroughly verify the Data Quality across setup	Do we need a focused DQ support team to support post implementation? If Yes, what would be the support period? Or do we need a DQ team trained on your end?	Bidder should propose DQ tool and team to work in tandem with Bank DQ team.
486	62	NA	Authentication and Identity Management - - A comprehensive identity and access management system should be available for centralized management of users and groups.	Is the expectation here to bring in a IDAM solution separately for DWH or we can leverage the existing IAM solutions that the Bank has?	No. Bank is currently using Microsoft ADS. Bidder need to integrate their solution with ADS
487	62	NA	Authentication and Identity Management - - A comprehensive identity and access management system should be available for centralized management of users and groups.	For managing the access controls, can we use Bank's existing ITSM solutions for workflow management?	Yes

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
488	64		Automation –Tool should automate analytics and reporting workflow end-to-end, including all data collection, enrichment, and management, as well as all calculations, processes to final report submission. Currently 500+ jobs are being used for Tranche 1 DCT generation along with 500 more for other regulatory reports/returns.	How will be expected no of regulatory reporting by end of 5 years	Number can not be ascertained at this stage.
489	73	Data Science Platform with AI/ML Capabilities	Annexure E gives sample use cases which are to be implemented on Next Gen Data Warehouse	No annexure on the sample use cases mentioned in the document	Existing models to be migrated and new models will be developed by the Bank In-house team
490	76	NA	Reporting on all types of available of Data Formats; Structured, semi-structured, unstructured Click stream data Audit Logs Documents Multimedia data Images/Videos/Audios) XBRL format IRIS iFILE framework	Requesting the bank to qualify the types of reports to be developed for Multimedia data and share the data format for IRIS iFile	Bidder need ot propose the best suitable solution.
491	76	NA	Data blending: The product must permit the user to blend data from various data sources. This includes accessing the data and mapping or creating relationships with data from multiple sources.	Please elaborate, Need more clarity	Bidder need ot propose the best suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
492	77		Context-based visualizations: Only visualizations or chart types that are relevant to the data selected will be listed as options.	Please elaborate on this requirement	Bidder need ot propose the best suitable solution.
493	77	NA	Collaboration and social interaction: BI tools enable the creation of a business community that can share and discuss their decision-making analysis. This would include annotating analysis to share observations and social media, which can enable discussion threads or chats.	Requesting bank to share details on the actual use cases	Bidder need ot propose the best suitable solution.
494	78		Microsoft Office real-time data integration. Beyond simple import and export, the product should provide real-time data integration with Microsoft Office products, which enables business people to embed analytics from the business intelligence tool into a PowerPoint or Excel presentation, for example, and refresh it automatically as the data is updated.	What will be the scope (# of reports) for the real-time reporting	Bidder need ot propose the best suitable solution.



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
495	78	NA	Offline updates: BI tools, when storing copies of the source data in an online analytical processing (OLAP) cube or in-memory columnar data store, should enable business users to schedule automatic data updates.	Requesting bank to share details on the actual use cases	Bidder need ot propose the best suitable solution.
496	79		Ease of analytical use: There should be different criteria defined for each type of user, such as information consumer, business analyst and IT.	Please elaborate on this requirement	Bidder need ot propose the best suitable solution.
497	79		Ease of creating BI applications: There should be different criteria for each type of analytics creator, such as business analysts and IT.	Please elaborate on this requirement	Bidder need ot propose the best suitable solution.
498	79	NA	Ease of creating BI applications: There should be different criteria for each type of analytics creator, such as business analysts and IT	Requesting bank to share details on the actual use cases	Bidder need ot propose the best suitable solution.
499	80	NA	Ability to handle and summarize huge volumes of data. E.g. 30-40 million rows accessed on index and summarized over 5 to 8 metrics.	Modern BI platforms utilize the scale and performance of backend Data Warehouse which is best suited to deliver the required performance to handle huge volumes of data. This architecture not only helps in improving performance but also ensure a stronger security posture and reduces data duplicacies	Bidder can provide suggestions

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
500	NA	EOI Binding Type	NA	Whether this a non-binding EOI, through which the client seeks to shortlist potential bidders? Can we decide at a later stage not to bid without any consequences, e.g. Financial penalty/ Being black-listed/ Forfeiture of EMD, performance security?	This EOI is an ideation stage. RFP released after EOI may be open or closed. Please refer clause 11. MODIFICATION AND WITHDRAWAL OF BIDS, Page 15 of EOI.
501	NA	EOI Declaration	NA	Do we need to submit any declarations or undertakings related to debarment or pending litigation at the EOI response stage?	Yes Bidder need to provide documents as per the eligibility criteria
502	NA		Business Intelligence Tools	Apart from Regulatory reporting, is there any requirement of pixel reporting ?	Bidder need ot propose the best suitable solution.
503	33	C 2	DWH has been shifted from PDOA (Pure data for Operation Analytics) technology to IIAS (IBM Integrated Analytical System) technology, along with a small-scale Data Lake and Data Virtualization capabilities. Data Lake is implemented with a view to enrich the capabilities of sourcing unstructured and semi-structured data for advance analytics with AI / ML tools	what is the current Datalake being used ? Does it have Spark framework on top of this ?	IBM CP4D, CP4I : Spark framework is there.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
504	34	C 2.1.b	Real time integration with upstream and downstream users with latest technologies like streaming and APIs for data movement / utilization.	<p>Can you please list the number of</p> <p>API Jobs ~</p> <p>RealTime Jobs ~</p> <p>BigData Jobs ~</p> <p>Data Integration Jobs ~</p> <ul style="list-style-type: none"> <li>•No of APIs to be hosted as a part of the project.</li> <li>•Mediation complexity, simple medium or complex</li> <li>•Average message request size</li> <li>•Messages or requests/second to be processed.</li> <li>•Messages/Data size processed per day.</li> </ul>	will be shared with finally selected bidder
505	35	C 2.1.j	Migration of existing ETL pipelines / jobs and analytical models.	what is the current pain points on ETL processing and Data governance & existing DW ? please share the top challenges	Objective is meeting 3 hours timeline for data availability
506	35	C3.2	Volumetric of Existing Data Warehouse:	can you please share the % split of Realtime jobs, API, Jobs, BigData jobs and normal DI Jobs. What is the time window for each processing per day	Please refer section C3.2 Processing approx 5 hours per day

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
507	35	C3.1	Present Architecture of existing Data Warehouse and technology stack will be shared with the intended bidders who participate in this EOI proposal on execution of Non-Disclosure Agreement (NDA). Copy of NDA will be shared in due course.	Can you please share the current system architecture	Details have been explained in Pre-bid Meeting.
508	36	C3.2. 11	Number of Cognos Reports ~100	Number of simple/medium/complex reports?	Reports can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement
509	36	C3.2. 11	Number of Cognos Reports ~100	Number of simple/medium/complex dashboards?	Dashboards can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement
510	36	C3.2. 11	Number of Cognos Reports ~100	How many scheduled reports are required?	Approx 100+ Cognos Reports
511	36	C3.2. 11	Number of Cognos Reports ~100	How many Reports on the email on daily/hourly/weekly and monthly ?	Approx 20 Reports: 10 Daily, 9 Weekly, 2 Monthly
512	36	C3.2. 11	Number of Cognos Reports ~100	What is the frequency of these dashboards and reports?	Frequency of these Reports is Daily

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
513	42	5	Solution proposed for Data Ingestion should be compatible to ingest and replicate data simultaneously on all available regions.	can you please list the total number of regions in production	Only one Region In production at present. Solution can propose multiple regions.
514	42	6	End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics maximum within 3 hours from day end for batch data loads on all days including month-end / quarter-end / year-end. Vendor may propose suitable solution to achieve the objective. Refer Annexure – C for existing data flow in DWH	can we get Annexure C	Architecture along with data flow has been presented in pre-bid meeting
515	43	15	Proposed solution should be able to scrap encrypted log, capture Metadata changes at source level completely, scrapping 4000-5000 logs daily having log size of ~ 32 GB each scalable up to 20000 logs	can provide more details on what are the log types are required to process ?	Here log implies CDC scrapping logs (Archive logs generated by Source)

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
516	51	4	Migration from Existing Setup to Proposed Solution Migration of existing data sourcing ETL jobs.	can you please list the details on ETL workload ? 1) Type of ETL Jobs 2) Type of ETL tools 3) Total no of Production Jobs 4) Type of ETL jobs eg Basic, BigData etc 5) Split % Simple, Medium and Complex Jobs need to be migrated	1,2) Data Stage Jobs of various frequencies. 3) Total Number of production Jobs approx 27000. 4) Approx 30 Big Data Jobs (Huge Data Size) 5) SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experiece and may vary depending on future load /requirement
517	51	9	Migration of All the remaining components of existing ecosystem (Mentioned in Annexure - D) as and when identified by Bank like job scheduler, reports, history of version control, existing tape backup, etc.	can we get Annexure D	Components has been presented in Pre-Bid Meeting.
518	51	8	Migration of Data Governance, Data Lineage and Data Quality rules and policies	what are the TOP challenges for current Data governance process and tool ?	Generating Data Lineages which takes time, Regular Data profiling.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
519	52	2	DR will contain jobs, scripts, models, snapshots, metadata, all software requires for managing and maintaining PR cluster along with critical data defined by the Bank	Is there existing DB sync happens between Prod and DR can we leverage or customer DB for metadata storage ? Is there NFS or Shared storage can leverage between Prod and DR ? existing Code repository & Artificat respostory can be reused ?	At present, DR is not fully operational
520	59	1	Vendor should propose end-to-end solution for Data Quality Management starting from data origin till the data consumption. These tool (s) to be used for addressing various aspects of the data quality problem mentioned below on SBI data set during data ingestion, data processing or data consumption as advised by Bank on case by case basis	What are the types of sources need to profiled ? eg of few list of systems	CBS , LOS, LLMS (these are sample source systems). Also, on need basis, data from other sources may also be required to be profiled for ensuring Data quality
521	61	14	Parallel processing: Data quality tool should be able to handle 500 concurrent users running any kind of job (e.g. Data Profiling on small/medium/large tables as advised by Bank in the tool)	what is the total number of DQ jobs runs as a part of normal day to day operations ? How much data need to be profiled per day est ?	~50 Data Stage jobs ( rules ~ 100)

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
522	61	14	Parallel processing: Data quality tool should be able to handle 500 concurrent users running any kind of job (e.g. Data Profiling on small/medium/large tables as advised by Bank in the tool)	Where the current no of adhoc users currently in production performing this profiling ? Just to confirm these are 500 DQ profiling jobs or 500 users standalone submitted the jobs on day to day basis ?	At present data profiling is done on adhoc basis
523	73	24	Annexure E gives sample use cases which are to be implemented on Next Gen Data Warehouse using structured and/or unstructured and/or semi-structured and/or any other kind of data gathered from either Data Warehouse or Data Lake or Data Virtualization or all together or any other source.	can we get Annexure E	Components will be presented in Pre-Bid Meeting.
524	80	32	Ability to handle and summarize huge volumes of data. E.g. 30-40 million rows accessed on index and summarized over 5 to 8 metrics.	Relevant analytics ready data volume (Size in TB or PB , no of tables)for Dashboarding and reporting layer?	Reporting layer : ~ 3000 tables
525	86	33	ISDLDW should support at-least 500 concurrent users, scalable up to 1000 users in next 5 years, running ETL/ELT jobs or doing ad-hoc data extraction requests on database (Not including API based access or scheduled job connections to database)	Can you please let us know the current adhoc users for the ETL jobs ? API , BigData and Streaming jobs users ?	Adhoc Users (directly accessing DB) => ~300 users API => 2 departments Streaming jobs => use cases under implementation



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
526	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	No of Users / No of developers/ No of self-service users/ No of consumers only?	Bidder need ot propose the best suitable solution.
527	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	Can you share the role/persona wise peak user concurrency? For example Sales manager/RMs/HO/Zos etc.	Bidder need ot propose the best suitable solution.
528	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	What is the concurrency on dashboards as of today?	Bidder need ot propose the best suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
529	1. INVITATION TO BID:	1. INVITATION TO BID:	iii. Bidder shall mean any entity (i.e. juristic person) who meets the eligibility criteria given in Appendix-B of this EOI and willing to provide the product and services required by the Bank through RFP process in pursuant to this EOI. The interested Bidders who agree to all the terms and conditions contained in this EOI may submit their Bids with the information desired in this EOI. Consortium bidding is not permitted under this EOI.	<p><b>Change Request</b> We request bank to allow consortium with 2 partners bidding jointly</p> <p><b>Background</b> As we understand, this project requires deliverables related to infrastructure (Hardware, Virtualisation) management, administration, coordination, FMS activities.</p> <p>These will be apart from the Database and Application level activities.</p> <p>An SI having experience around infrastructure, various tools, public cloud migration and management can add value</p>	As per Point No: 3, Page No: 7 of EOI Consortium bidding is not permitted.
530	8	2	Background	What are the Key Challenges with the Current DWH/DL and What are the Key Drivers for a New/Improved DWH/DL?	Present agreement with the existing vendor/OEM is expiring on Feb 2025: DR unavailability, Scalability for DataLake.
531	8	2	Background	Can you provide a brief business functionality coverage of the Existing DW/DL	MIS, operational reports, Regulatory Reporting, Analytics etc.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
532	8	2	Background	Are there any specific use cases to be implemented from this new platform which were not being catered by earlier platform?	1) DR 2) Scalability
533	34	C2.1	Goal and Objectives	Does SBI currently have a full fledged DR or only a Functional DR? Can you provide some views on the same, as to is it 10% of Functionality Covered in FDR etc.	Bank is looking for full fledged DR solution.
534	34	C2.1	Goal and Objectives	Can the bank give some view on various types of sources, their key functionalities , that will source data to the DWH/DL	will be shared with finally selected bidder
535	34	C2.1	Goal and Objectives	Is there any preference on any SaaS or COTS platform? Or the SI is free to choose one that best fits the solution ecosystem.	Please refer to Section C2.1 (e ) in this regard. Bidder should propose best suitable solution.
536	34	C2.1	Goal and Objectives	Please let us know if there is any preference of cloud provider?	Bank is looking for scalable solution that will be on-prem or Bank's own Meghdoot cloud. However, solution to be scalable to public cloud if require in future (minimize effort and impact to move to cloud in future). At present, no integration with public cloud.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
537	35	C3.1	Present Architecture of Data Warehouse	Can the bank provide a technology architecture of various components within the DL/DWH?	Details have been explained in Pre-bid Meeting.
538	35	C3.2	Volumetric of Existing Data Warehouse	With such a huge count of ETL jobs(27000), is bank looking for migrating to same technology stack to save fresh build efforts or looking for new tech stack altogether?	Bank is open for any new tech stack; Bidder needs to consider migration effort as well in the proposal (including rationalization)
539	35	C3.1	Present Architecture of Data Warehouse	Will SBI's existing DWH and new DWH platform run parallelly or will there be sunset for existing DWH platform?	Once new platform is fully functional and reconciliation is done, existing will be sunset.
540	34	C2.1	d) Data Federation	Do we still need Data Virtualization/Federation if all the analytics needs can be served with a single platform covering Data Lake, Data warehouse, Operation DB, and Object storage ?	Yes it will be required.
541	35	C3.2	Volumetric of Existing Data Warehouse:	The data size given for Production DB is compressed or raw. Can we get the raw data size? Similarly, we would need the same for Archival and Data Lake	Compressed with factor of ~4. Further details will be provided to final bidders.
542	35	C3.2	Volumetric of Existing Data Warehouse:	Can we get the HW configuration for the current setup, no of servers and core count, memory count, etc?	Details have been explained in Pre-bid Meeting.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
543	36	C4	Functional Requirements	It's mentioned Bank is planning to upgrade the existing Data Warehouse and Data Lake solution to a modern and state-of-the-art integrated solution. In addition to project mandates mentioned in section C3, the solution should provide the following additional capabilities: Instead of upgrading, Is the bank open to replacing the existing setup completely with modern capabilities built into it?	Yes
544	43	C4	15. The proposed solution should be able to scrap encrypted log	What does this mean, please elaborate.	Change Data Capture (CDC) should be able to scrap encrypted logs as well from TDE enabled database.
545	46	C4	7. Processing engines should have capabilities to process all kinds of Data Storage formats used by the Bank.	Can you please elaborate on the Data formats used.	All prevailing formats available in the market should be supported.
546	89	C4	The vendor should provide EXACT size needed for production in the 1st year and estimated sizes for consecutive years keeping in view the growth rate predicted by Bank in this section and provide empirical evidence for the calculation of growth rate.	Is Bank expecting to provide Hardware only for 1st Year ? OR Bank expecting to provide Hardware with five year sizing. Please confirm	Bidder to propose suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
547	16	12. (ii)	After examining the EOI, some or all of the Bidders may be asked to make presentation of the solution and demonstrate proof of concept (POC).	What will be the scope of POC?	Details will be provided with final bidders.
548	34	C2.1 (e)	With a view to overcome these limitations, the proposed solution should be implementable in Physical hardware / Bank's own cloud (Meghdoot) and scalable to public / hybrid cloud in future. This solution should also support cloud bursting for heavy workload on demand.	1) Bank has mentioned "Physical hardware / Bank's own cloud (Meghdoot) and scalable to public / hybrid cloud in future". Does this mean solution should be implementable on all these options? Any preferences of the bank? 2) Cloud bursting would imply "application runs in a private cloud or data center and bursts into a public cloud when the demand for computing capacity spikes". Our understanding from this is that the bank is looking for a stack similar to cloud and connected to cloud with on demand extension. Please clarify.	Bidder need to propose the best suitable solution for SBI. Since EOI is an ideation process, RFP will be released with final requirement.
549	33	C1	Data Warehouse has migrated on new infrastructure on 10th May 2023 as per five-year Agreement between Bank and M/s IBM India Private Limited in February 2020 for implementation and maintenance of the Data Warehouse.	By when does the bank require the new setup to be ready and migration to be done?	Bidder need to propose the Tentative project implementation timelines in their EOI proposal.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
550	34	Appendix C - C2 : b	Real time integration with upstream and downstream users with latest technologies like streaming and APIs for data movement / utilization.	What is the data mentioned here, where is it being moved to?	Real time integration with upstream for data sourcing and with downstream for data extraction
551	35	Appendix C - C2 : k	Dev, UAT, Production and DR setup to be done along with sandbox for analytics	Who will provide these environments ?	Bidder has to implement these environment
552	41	Appendix C - C4	The applicant should describe how their solution will meet the required parameters and provide details thereof in their EOI proposal on the following lines by answering all the points given in below table as capability (Yes/No)	Below points to be updated by delivery team.	these points need to be provided by bidder under EOI response.
553	63	Appendix C - C4 : Security and Compliance - 6	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	What is the tool mentioned here to alert on data leakage ?	Details will be provided in RFP
554	63	Appendix C - C4 : Security and Compliance - 8	Compliance to Global Standards	Is it required for the project associates to be compliant to the following standards, if so then Bidder HR to be contacted.	Query is not for bank
555	67	Appendix C - C4 : Data Encryption	Data Encryption	Where will the customer data reside, in bidder or Customer environment ?	Data will be reside at Bank's environment.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
556	33	C 2	DWH has been shifted from PDOA (Pure data for Operation Analytics) technology to IIAS (IBM Integrated Analytical System) technology, along with a small-scale Data Lake and Data Virtualization capabilities. Data Lake is implemented with a view to enrich the capabilities of sourcing unstructured and semi-structured data for advance analytics with AI / ML tools	what is the current Datalake being used ? Does it have Spark framework on top of this ?	IBM CP4D, CP4I : Spark framework is there.
557	34	C 2.1.b	Real time integration with upstream and downstream users with latest technologies like streaming and APIs for data movement / utilization.	Can you please list the number of API Jobs ~ RealTime Jobs ~ BigData Jobs ~ Data Integration Jobs ~ •No of APIs to be hosted as a part of the project. •Mediation complexity, simple medium or complex •Average message request size •Messages or requests/second to be processed. •Messages/Data size processed per day.	will be shared with finally selected bidder



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
558	35	C 2.1.j	Migration of existing ETL pipelines / jobs and analytical models.	what is the current pain points on ETL processing and Data governance & existing DW ? please share the top challenges	Objective is meeting 3 hours timeline for data availability
559	35	C3.2	Volumetric of Existing Data Warehouse:	can you please share the % split of Realtime jobs, API, Jobs, BigData jobs and normal DI Jobs. What is the time window for each processing per day	Please refer section C3.2 Processing approx 5 hours per day
560	35	C3.1	Present Architecture of existing Data Warehouse and technology stack will be shared with the intended bidders who participate in this EOI proposal on execution of Non-Disclosure Agreement (NDA). Copy of NDA will be shared in due course.	Can you please share the current system architecture	Details have been explained in Pre-bid Meeting.
561	36	C3.2. 11	Number of Cognos Reports ~100	Number of simple/medium/complex reports?	Reports can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
562	36	C3.2. 11	Number of Cognos Reports ~100	Number of simple/medium/complex dashboards?	Dashboards can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement
563	36	C3.2. 11	Number of Cognos Reports ~100	How many scheduled reports are required?	Approx 100+ Cognos Reports
564	36	C3.2. 11	Number of Cognos Reports ~100	How many Reports on the email on daily/hourly/weekly and monthly ?	Approx 20 Reports: 10 Daily, 9 Weekly, 2 Monthly
565	36	C3.2. 11	Number of Cognos Reports ~100	What is the frequency of these dashboards and reports?	Frequency of these Reports is Daily
566	42	5	Solution proposed for Data Ingestion should be compatible to ingest and replicate data simultaneously on all available regions.	can you please list the total number of regions in production	Only one Region In production at present. Solution can propose multiple regions.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
567	42	6	End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics maximum within 3 hours from day end for batch data loads on all days including month-end / quarter-end / year-end. Vendor may propose suitable solution to achieve the objective. Refer Annexure – C for existing data flow in DWH	can we get Annexure C	Architecute along with data flow has been presented in pre-bid meeting
568	43	15	Proposed solution should be able to scrap encrypted log, capture Metadata changes at source level completely, scrapping 4000-5000 logs daily having log size of ~ 32 GB each scalable up to 20000 logs	can provide more details on what are the log types are required to process ?	Here log implies CDC scrapping logs (Archive logs generated by Source)

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
569	51	4	Migration from Existing Setup to Proposed Solution Migration of existing data sourcing ETL jobs.	can you please list the details on ETL workload ? 1) Type of ETL Jobs 2) Type of ETL tools 3) Total no of Production Jobs 4) Type of ETL jobs eg Basic, BigData etc 5) Split % Simple, Medium and Complex Jobs need to be migrated	1,2) Data Stage Jobs of various frequencies. 3) Total Number of production Jobs approx 27000. 4) Approx 30 Big Data Jobs (Huge Data Size) 5) SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experiece and may vary depending on future load /requirement
570	51	9	Migration of All the remaining components of existing ecosystem (Mentioned in Annexure - D) as and when identified by Bank like job scheduler, reports, history of version control, existing tape backup, etc.	can we get Annexure D	Components has been presented in Pre-Bid Meeting.
571	51	8	Migration of Data Governance, Data Lineage and Data Quality rules and policies	what are the TOP challenges for current Data governance process and tool ?	Generating Data Lineages which takes time, Regular Data profiling.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
572	52	2	DR will contain jobs, scripts, models, snapshots, metadata, all software requires for managing and maintaining PR cluster along with critical data defined by the Bank	Is there existing DB sync happens between Prod and DR can we leverage or customer DB for metadata storage ? Is there NFS or Shared storage can leverage between Prod and DR ? existing Code repository & Artificat repository can be reused ?	At present, DR is not fully operational
573	59	1	Vendor should propose end-to-end solution for Data Quality Management starting from data origin till the data consumption. These tool (s) to be used for addressing various aspects of the data quality problem mentioned below on SBI data set during data ingestion, data processing or data consumption as advised by Bank on case by case basis	What are the types of sources need to profiled ? eg of few list of systems	CBS , LOS, LLMS (these are sample source systems). Also, on need basis, data from other sources may also be required to be profiled for ensuring Data quality
574	61	14	Parallel processing: Data quality tool should be able to handle 500 concurrent users running any kind of job (e.g. Data Profiling on small/medium/large tables as advised by Bank in the tool)	what is the total number of DQ jobs runs as a part to normal day to day operations ? How much data need to be profiled per day est ?	~50 Data Stage jobs ( rules ~ 100)

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
575	61	14	Parallel processing: Data quality tool should be able to handle 500 concurrent users running any kind of job (e.g. Data Profiling on small/medium/large tables as advised by Bank in the tool)	Where the current no of adhoc users currently in production performing this profiling ? Just to confirm these are 500 DQ profiling jobs or 500 users standalone submitted the jobs on day to day basis ?	At present data profiling is done on adhoc basis
576	73	24	Annexure E gives sample use cases which are to be implemented on Next Gen Data Warehouse using structured and/or unstructured and/or semi-structured and/or any other kind of data gathered from either Data Warehouse or Data Lake or Data Virtualization or all together or any other source.	can we get Annexure E	Components will be presented in Pre-Bid Meeting.
577	80	32	Ability to handle and summarize huge volumes of data. E.g. 30-40 million rows accessed on index and summarized over 5 to 8 metrics.	Relevant analytics ready data volume (Size in TB or PB , no of tables)for Dashboarding and reporting layer?	Reporting layer : ~ 3000 tables
578	86	33	ISDLDW should support at-least 500 concurrent users, scalable up to 1000 users in next 5 years, running ETL/ELT jobs or doing ad-hoc data extraction requests on database (Not including API based access or scheduled job connections to database)	Can you please let us know the current adhoc users for the ETL jobs ? API , BigData and Streaming jobs users ?	Adhoc Users (directly accessing DB) => ~300 users API => 2 departments Streaming jobs => use cases under implementation

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
579	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	No of Users / No of developers/ No of self-service users/ No of consumers only?	Bidder need ot propose the best suitable solution.
580	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	Can you share the role/persona wise peak user concurrency? For example Sales manager/RMs/HO/Zos etc.	Bidder need ot propose the best suitable solution.
581	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	What is the concurrency on dashboards as of today?	Bidder need ot propose the best suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
582	20	V(a)	<p>(a) Holiday Listing (Temporary Debarment - suspension): Whenever a Bidder is found lacking in performance, in case of less frequent and less serious misdemeanors, the Bidders may be put on a holiday listing (temporary debarment) for a period upto 12 (twelve) months. When a Bidder is on the holiday listing, he is neither invited to bid nor are his bids considered for evaluation during the period of the holiday. The Bidder is, however, not removed from the list of empaneled vendors, if any.</p> <p>Performance issues which may justify holiday listing of the Vendor are:</p> <ul style="list-style-type: none"> <li>• Vendors who have not responded to requests for quotation/tenders consecutively three times without furnishing valid reasons, if mandated in the empanelment contract (if applicable);</li> <li>• Repeated non-performance or performance below specified standards (including after sales services and maintenance services etc.);</li> <li>• Bidders undergoing process for</li> </ul>	(	EOI terms are final



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
583	21	V(b)	<p>(b) Debarment from participation including removal from empaneled list  Debarment of a delinquent Bidder (including their related entities) for a period (one to two years) from the Bank's procurements including removal from empanelment, wherever such Bidder is empaneled, due to severe deficiencies in performance or other serious transgressions. Reasons which may justify debarment and/or removal of the Bidder from the list of empaneled vendors are:</p> <ul style="list-style-type: none"> <li>• Without prejudice to the rights of the Bank under Clause 14(i) hereinabove, if a Bidder is found by the Bank to have directly or indirectly or through an agent, engaged or indulged in any corrupt/fraudulent/coercive/undesirable or restrictive practices during the bidding Process, such Bidder shall not be eligible to participate in any EOI/RFP issued by the Bank during a period of 2 (two) years from the date of debarment.</li> <li>• The Bidder fails to abide by the terms and conditions or to maintain the required technical/operational</li> </ul>		EOI terms are final

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
584	22	V (c )	(c) Banning from Ministry/Country-wide procurements For serious transgression of code of integrity, a delinquent Bidder (including their related entities) may be banned/debarred from participation in a procurement process of the Bank including procurement process of any procuring entity of Government of India for a period not exceeding three years commencing from the date of debarment.		EOI terms are final
585	24	19 (ii)	19. OTHER TERMS & CONDITIONS ii. Willful misrepresentation of any fact within the Bid will lead to the disqualification of the Bidder without prejudice to other actions that the Bank may take. All the submissions, including any accompanying documents, will become property of the Bank. The Bidders shall be deemed to license, and grant all rights to the Bank, to reproduce the whole or any portion of their Bid document for the purpose of evaluation and to disclose the contents of submission for regulatory and legal requirements.	ii. Willful misrepresentation of any fact within the Bid will lead to the disqualification of the Bidder from the current bidding process without prejudice to other actions that the Bank may take. All the submissions, including any accompanying documents, will become property of the Bank. The Bidders shall be deemed to license, and grant all rights to the Bank, to reproduce the whole or any portion of their Bid document for the purpose of evaluation and to disclose the contents of submission for regulatory and legal requirements.	EOI terms are final

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
586	27-29	Appendix A	<p>Appendix-A: BID FORM (TECHNICAL BID)</p> <p>v. It is further certified that the contents of our Bid are factually correct. We have not sought any deviation to the terms and conditions of this EOI. We also accept that in the event of any information / data / particulars proving to be incorrect, the Bank will have right to disqualify us from the EOI or any subsequent bidding process without prejudice to any other rights available to the Bank.</p> <p>vi. We certify that while submitting our Bid document, we have not made any changes in the contents of the EOI, read with its amendments/clarifications provided by the Bank.</p> <p>x. We hereby certify that on the date of submission of Bid, we do not have any Service Level Agreement pending to be signed with the Bank for more than 6 months from the date of issue of purchase order.</p>	<p>v. It is further certified that the contents of our Bid are factually correct. We also accept that in the event of any information / data / particulars proving to be incorrect, the Bank will have right to disqualify us from the EOI without prejudice to any other rights available to the Bank.</p> <p>vi..</p> <p>x. We hereby certify that on the date of submission of Bid, we do not have any Service Level Agreement pending to be signed with the Bank for more than 6 months from the date of issue of purchase order.T</p>	EOI terms are final
587	33	C1		Which is the technology on which the Data Lake has been implemented?Is it part of the existing IBM stack of solutions	Details have been explained in Pre-bid Meeting.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
588	33	C1		Is SBI looking for a Data Lakehouse kind of solution which can be compatible with its existing architecture?	Bidder should propose best suitable solution in the EOI.
589	33	C1		Has SBI identified any suitable vendors for their Data Lake in the new architecture?	Bidder should propose best suitable solution in the EOI.
590	35	C3.2 Volumetric of Existing Data Warehouse:		Which is the tool which is used for ETL jobs? Do you have any preferred ETL tool?	Details have been explained in Pre-bid Meeting. Bidder should propose best suitable solution.
591	36	C3.2 Volumetric of Existing Data Warehouse:		In which database are the tables currently? Are they on any RDBMS and if yes which one?	IBM DB2
592	37	C3.2 Volumetric of Existing Data Warehouse:		What are the different Data source systems ?	Details will be provided in the RFP.
593	38	C3.2 Volumetric of Existing Data Warehouse:		The scheduled workload and ad-hoc workloads are on which platform currently?	IBM data Stage

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
594	39	C3.2 Volumetric of Existing Data Warehouse:		Have you identified any reporting tool for the new architecture on which the Cognos reports need to be migrated?	Bidder should propose Best suitable solution.
595	35	C2. Project Mandate	C2.1 Goals and Objectives	Have you identified any tools for Data Governance?Are you using any existing tools for Data Goveranance and Data Quality?	Bank has no preference. Bidder to propose suitable solution / tool
596	34	C2. Project Mandate	C2.1 Goals and Objectives	Have you identified any tools for Data Virtualization?Which is the existing tool which you are using for the same?	Bank has no preference. Present tool CP4D virtualization.
597	36	C4. Functional Requirements	C4. Functional Requirements	Have you identified any tools for Data Science and ML related work?	Bank has no preference.
598	34	C2. Project Mandate	C2.1 Goals and Objectives	Are you having any preferences of Cloud vendor like AWS,Google or Azure or Oracle cloud for scaling in the future?	Bidder need to propose best suitable solution.
599	71	C4. Functional Requirements	IBM SPSS	It seems that SBI is currently using IBM SPSS as a ML platform. Are you looking to change the platform ?	Bidder should propose suitable solution. In case of new technology, migration is involved in the scope
600	71	C4. Functional Requirements	IBM SPSS	Has SBI decided on whether to continue with IBM SPSS or have you decided on an alternate platform for AI/ Machine Learning	Bidder need to propose best suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
601	39	C4. Functional Requirements	Data governance	Do you have any existing data governance framework in place and what is the current maturity level?	Various tools are being used for data lineage, data quality.
602	51	C4. Functional Requirements	Architectural Requirements - Migration from Existing Setup to Proposed Solution	Do you have any existing Data Catalog/Governance/Data Quality tool implemented into your ecosystem? If yes, please provide details.	Details have been presented in pre-bid meeting.
603	51	C4. Functional Requirements	Architectural Requirements - Migration from Existing Setup to Proposed Solution	Total No. of Business glossary, policies,rules,sourcesystems details to be migrated ?	Details to be provided in the RFP.
604	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Current data architecture,data flow and their technology stack details?	Details have been explained in Pre-bid Meeting.
605	36	C3.2 Volumetric of Existing Data Warehouse		The Total no. of source systems is specified as ~100, Does SBI wants all of them Cataloged or only a portion of them?	Yes. All systems need to be considered for new solution.
606	36	C3.2 Volumetric of Existing Data Warehouse		Approximate Splitup of master,reference,transactional,dimensional & fact tables out of the total 50000 tables?	Details will be shared in RFP phase

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
607	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Please specify the technology stack of your source systems?	Details have been explained in Pre-bid Meeting.
608	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Do you have any existing Data Catalog/Governance/Data Quality tool implemented into your ecosystem? If yes, please provide details.	Details have been presented in pre-bid meeting.
609	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Please confirm if a Data Governance team/Council is in place, and if yes what are the different roles that exist in the team.	Yes
610	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Have you identified the DG Roles and Responsibilities ? Eg. Data Stewards, Data Custodian,Data Owner,etc.,	Yes, DG Roles and responsibilities has been defined.
611	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Are there any existing Policies, standards,Processes, or Guidelines defined with relation to the DG framework?	Yes, these have been defined.
612	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Is the Business Glossary (unified Enterprise level) or yet to be curated by SBI ?	Bank has defined data dictionary.
613	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Do you have centralized metadata repository or platform for knowledge sharing across organization?	Yes

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
614	57	C4. Functional Requirements	Architectural Requirements - Data Governance	How is data governed currently?	As per Data Governance policy of the Bank.
615	57	C4. Functional Requirements	Architectural Requirements - Data Governance	How is personal data currently captured & maintained?	All data including personal data are sourced from upstream systems
616	57	C4. Functional Requirements	Architectural Requirements - Data Governance	Do you have any built-in AI/ML/LLM models that needs to be integrated to the data ecosystem?	Yes. Model migraton is also in the scope
617	59	C4. Functional Requirements	Architectural Requirements - Data Quality	What is your current maturity status in terms of data Quality?	Defined.
618	59	C4. Functional Requirements	Architectural Requirements - Data Quality	How is data quality & its remediation handelled currently?	IGC, IA, QS toos are available in eco-system.
619	59	C4. Functional Requirements	Architectural Requirements - Data Quality	Any existing DQ issues which SBI wants us to address	Present DQ framework need to be migrated and also capability to develop new DQ rules.
620	59	C4. Functional Requirements	Architectural Requirements - Data Quality	What are type of Data Quality issues seen in the current Eco system ?	Present DQ process handles few functional scenarios. Details will be provided in RFP



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
621	59	C4. Functional Requirements	Architectural Requirements - Data Quality	Are there any business and data quality rules defined currently?	Present DQ process handles few functional scenarios. Details will be provided in RFP
622	59	C4. Functional Requirements	Architectural Requirements - Data Quality	Would the Data quality implementation include remediation of bad quality data back to source, or is purely reporting on the bad data in scope currently?	Reporting of bad data is in the scope
623	59	C4. Functional Requirements	Architectural Requirements - Data Quality	Total no. of KPI's and Critical Data Elements(CDE)?	Details can be provided in RFP
624	59	C4. Functional Requirements	Architectural Requirements - Data Quality	How are Reference Data & MDM managed currently?	Presently there is no MDM solution exists. It is not in the scope of this EOI
625	33	C1. Background		Please provide more clarity on the need for ISDLDW when you have already migrated the DW to IBM IAS and implemented a Data Lake. Can you provide the details of uses cases implemented and elaborate on the additional business benefits envisaged? Also, can you provide the gaps and challenges that exists and you want to address with ISDLDW.	Scope of work is defined in EOI.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
626	34	C2. Project Mandate		Could you please elaborate the scope of ISDLDW? Which line of businesses and/or functions of the bank, does ISDLDW is aimed at supporting? What is the current process of extracting analytics from the DW and Data Lake?	Scope fo work is defined in EOI.
627	40	C4. Functional Requirements	EOI Proposal should include following items for each proposed solution: -> a) to e)	In the absence of available data about the systems, is the expectation to share the reference architecture/ solution / best practices from the industry perspective?	Bidder need to propose best suitable solution in the EOI.
628	31	Appendix-B: Bidder's Eligibility Criteria	5. Bidder should have experience of minimum 3 years in implementation and managing large-scale (~petabytes) enterprise data warehouse / data lake preferably in BFSI sector.  Copy of the order and / or Certificate of completion of the work. The Bidder should also furnish user acceptance report from client(s).	We request bank to allow letter from client stating details of the project/ publicly available case studies as part of reference criteria.	Please refer to Corrigendum-2.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
629	31	Appendix-B: Bidder's Eligibility Criteria	6. Client references and contact details (email/ landline/ mobile) of customers for whom the Bidder has executed similar projects in India. (Start and End Date of the Project to be mentioned) in the past (At least two client references are required)	We understand that the Indian client reference is not specific to any data size. Please confirm.	Yes. As per EOI terms
630	42	C4. Functional Requirements	6. End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics maximum within 3 hours from day end for batch data loads on all days including month-end / quarter-end / year-end. Vendor may propose suitable solution to achieve the objective. Refer Annexure – C for existing data flow in DWH	We could not find Annexure C in the EOI. Please check and confirm.	Details have been presented in pre-bid meeting.
631	46	C4. Functional Requirements	11. The storage system should be robust to handle at least 1,50,000 concurrent queries (Select/DML) by processing engines / ETL jobs / end users scalable up to 6,00,000 concurrent queries in next 5 years (assuming parallelism of 100 degree).	Please specify the concurrency of type of query categories (Shallow, Simple, Medium, Complex)	queries can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
632	48	C4. Functional Requirements	12. Existing reports and extracts generation jobs on DWH should be analyzed and transformed to the ISDLDW. The vendor should use preferably off-the-shelf tools and not resort to building from scratch	Some of the reports and extracts cannot be converted. This can be done using tool+ human efforts. Request to acknowledge.	Bidder should propose best suitable solution
633	51	C4. Functional Requirements	9. Migration of All the remaining components of existing ecosystem (Mentioned in Annexure - D) as and when identified by Bank like job scheduler, reports, history of version control, existing tape backup, etc.	We could not find Annexure D in the EOI. Please check and confirm.	Components of existing eco systems have been presented in pre-bid meeting.
634	51	C4. Functional Requirements	9. Migration of All the remaining components of existing ecosystem (Mentioned in Annexure - D) as and when identified by Bank like job scheduler, reports, history of version control, existing tape backup, etc.	Please confirm whether banks existing vendor can extract and restore on disk ?	Bidder should propose best suitable solution
635	52	C4. Functional Requirements	12 Vendor should provide a feasible plan for best use of existing infrastructure which is procured during last 10 years in staggered manner during the implementation of ISDLDW which will save cost to the Bank. (Annexure D gives the technology architecture of the current setup)	We could not find Annexure D in the EOI. Please share details of existing infrastructure.	Technology architecture has been presented in pre-bid meeting

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
636	53	C4. Functional Requirements	1. Data older than specific duration as identified by Bank to be archived in low cost cold storage.	Please confirm whether bank is expecting old tape data to be restored?	Tape backup need to be restored, if required.
637	53	C4. Functional Requirements	3 All the applications connected to the non-archived data should be available with archived as well	Request to clarify.	Archived data should be available instantly for reporting / querying historic data without moving to main storage.
638	54	C4. Functional Requirements	6 Mixing and Merging data from Current/Live to and from Archival must not result in any significant loss of performance and response time	Request to define SLA and remove significant , it will be ambiguous and lead to sizing based on discretion	EOI terms are final
639	61	C4. Functional Requirements	15. Each month 5 data quality use cases will be developed and implemented on Next Gen DW.	We understand that bank will develop and implement the use case basis the platform provided by the bidder. Please confirm.	Bank has not defined any use case for this EOI.
640	63	C4. Functional Requirements	7 Compliance with SBI IS Policies – some of the key areas are as under (Summary of Key policies is given in Annexure F):	We could not find Annexure F in the EOI. Please check and confirm.	Brief if Bank's security guidelines will be highlighted in RFP
641	73	C4. Functional Requirements	24 Annexure E gives sample use cases which are to be implemented on Next Gen Data Warehouse using structured and/or unstructured and/or semi-structured and/or any other kind of data gathered from either Data Warehouse or Data Lake or Data Virtualization or all together or any other source.	We could not find Annexure E in the EOI. Please check and confirm.	Bank has not defined any use case for this EOI.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
642	73	C4. Functional Requirements	Vendor to provide solution / tool (s) for below scope of activities on SBI data sets; <ul style="list-style-type: none"> <li>• Benchmarking</li> <li>• Predictive &amp; Prescriptive Analytics</li> <li>• Social Media Analytics</li> <li>• Web Analytics</li> </ul>	Can web analytics be referred as web log analytics?	It refers to user behaviour analytics on web sites
643	78	C4. Functional Requirements	21 Mobile version: BI tools should be able to differentiate between viewing BI applications on a web browser on a mobile device versus a mobile BI application.	Please confirm the number of users accessing mobile BI application.	This should be capability of proposed BI tool
644	89	C4. Functional Requirements	50 Annexure G gives the ISDLDW sizing forecasted for next five years. Sizing given in the Annexure G is in compressed form with compression factor of 2.5. Vendors are expected to propose storage forecast over and above given sizing in the solution to ensure fast performance of system. Vendors shall propose performance benchmark for this system as part of the solution. This is indicated projected growth, but the solution should be scalable enough to store any volume of data.	We could not find Annexure G in the EOI. Please check and confirm.	Projected volumetric will be mentioned in the RFP.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
645	33	C 2	DWH has been shifted from PDOA (Pure data for Operation Analytics) technology to IIAS (IBM Integrated Analytical System) technology, along with a small-scale Data Lake and Data Virtualization capabilities. Data Lake is implemented with a view to enrich the capabilities of sourcing unstructured and semi-structured data for advance analytics with AI / ML tools	Bidder requests bank to confirm the details of the current Datalake being used ? Does it have Spark framework on the top ?	IBM CP4D, CP4I : Spark framework is there.
646	34	C 2.1.b	Real time integration with upstream and downstream users with latest technologies like streaming and APIs for data movement / utilization.	Bidder requests bank to provide following details- a) Number of API Jobs ~ b) Number of Real time Jobs ~ c) Number of Big Data Jobs ~ d) Number of Data Integration Jobs ~ e) No of APIs to be hosted as a part of the project. f) Mediation complexity, simple medium or complex g) Average message request size h) Messages or requests/second to be processed. i) Messages/Data size processed per day.	will be shared with finally selected bidder

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
647	35	C 2.1.j	Migration of existing ETL pipelines / jobs and analytical models.	Bidder requests bank to provide the current pain points on ETL processing and Data governance & existing DW ? Please share the top challenges	Objective is meeting 3 hours timeline for data availability
648	35	C3.2	Volumetric of Existing Data Warehouse:	Bidder requests bank to provide the % split of Realtime jobs, API, Jobs, BigData jobs and normal DI Jobs. What is the time window for each processing per day?	Please refer section C3.2 Processing approx 5 hours per day
649	35	C3.1	Present Architecture of existing Data Warehouse and technology stack will be shared with the intended bidders who participate in this EOI proposal on execution of Non-Disclosure Agreement (NDA). Copy of NDA will be shared in due course.	Bidder requests bank to share the current system architecture	Architecture has been presented in pre-bid meeting
650	36	C3.2. 11	Number of Cognos Reports ~100	Bidder requests bank to share the number of simple/medium/complex reports?	Reports can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement



Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
651	36	C3.2. 11	Number of Cognos Reports ~100	Bidder requests bank to share the number of simple/medium/complex dashboards?	Dashboards can be grouped in SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experience and may vary depending on future load /requirement
652	36	C3.2. 11	Number of Cognos Reports ~100	Bidder requests bank to confirm the number of scheduled reports required?	Approx 100+ Cognos Reports
653	36	C3.2. 11	Number of Cognos Reports ~100	Bidder requests bank to confirm the number of reports required on the email on daily/hourly/weekly and monthly ?	Approx 20 Reports: 10 Daily, 9 Weekly, 2 Monthly
654	36	C3.2. 11	Number of Cognos Reports ~100	Bidder requests bank to confirm on the frequency of dashboards and reports?	Frequency of these Reports is Daily
655	42	5	Solution proposed for Data Ingestion should be compatible to ingest and replicate data simultaneously on all available regions.	Bidder requests bank to confirm on the total number of regions in production.	Only one Region In production at present. Solution can propose multiple regions.
656	43	15	Proposed solution should be able to scrap encrypted log, capture Metadata changes at source level completely, scrapping 4000-5000 logs daily having log size of ~ 32 GB each scalable up to 20000 logs	Bidder requests bank to provide more details on what are the log types that are required to process ?	Here log implies CDC scrapping logs (Archive logs generated by Source)

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
657	51	4	Migration from Existing Setup to Proposed Solution Migration of existing data sourcing ETL jobs.	Bidder requests bank to provide following inputs on ETL workload- 1) Type of ETL Jobs 2) Type of ETL tools 3) Total no of Production Jobs 4) Type of ETL jobs eg Basic, BigData etc 5) Split % Simple, Medium and Complex Jobs need to be migrated	1,2) Data Stage Jobs of various frequencies. 3) Total Number of production Jobs approx 27000. 4) Approx 30 Big Data Jobs (Huge Data Size) 5) SIMPLE(50%), MEDIUM(30%) and COMPLEX (20%). These are approximate numbers based on experiece and may vary depending on future load /requirement
658	51	8	Migration of Data Governance, Data Lineage and Data Quality rules and policies	Bidder requests bank to provide the TOP challenges for current Data governance process and tool	Components has been presented in Pre-Bid Meeting.
659	52	2	DR will contain jobs, scripts, models, snapshots, metadata, all software requires for managing and maintaining PR cluster along with critical data defined by the Bank	Bidder requests bank to confirm whether there is existing DB sync happening between Prod and DR which we can leverage for customer DB for metadata storage ? Is there NFS or Shared storage which bidder can leverage between Prod and DR ? Is bank having any existing Code repository & Artificat repository which can be reused ?	At present, DR is not fully operational

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
660	59	1	Vendor should propose end-to-end solution for Data Quality Management starting from data origin till the data consumption. These tool (s) to be used for addressing various aspects of the data quality problem mentioned below on SBI data set during data ingestion, data processing or data consumption as advised by Bank on case by case basis	Bidder requests bank to confirm on the types of sources that need to profiled ? Please provide list of systems.	CBS , LOS, LLMS (these are sample source systems). Also, on need basis, data from other sources may also be required to be profiled for ensuring Data quality
661	61	14	Parallel processing: Data quality tool should be able to handle 500 concurrent users running any kind of job (e.g. Data Profiling on small/medium/large tables as advised by Bank in the tool)	Bidder requests bank to confirm on the total number of DQ jobs that runs as a part of normal day to day operations ? Also confirm estimated volume of data that needs to be profiled ?	~50 Data Stage jobs ( rules ~ 100)
662	61	14	Parallel processing: Data quality tool should be able to handle 500 concurrent users running any kind of job (e.g. Data Profiling on small/medium/large tables as advised by Bank in the tool)	Bidder requests bank to confirm the no. of adhoc users currently in production performing this profiling ? Please confirm whether these are 500 DQ profiling jobs or 500 users standalone submitted the jobs on day to day basis ?	At present data profiling is done on adhoc basis
663	80	32	Ability to handle and summarize huge volumes of data. E.g. 30-40 million rows accessed on index and summarized over 5 to 8 metrics.	Bidder requests bank to confirm on the relevant analytics ready data volume (Size in TB or PB , no of tables)for Dashboarding and reporting layer?	Components has been presented in Pre-Bid Meeting.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
664	86	33	ISDLDW should support at-least 500 concurrent users, scalable up to 1000 users in next 5 years, running ETL/ELT jobs or doing ad-hoc data extraction requests on database (Not including API based access or scheduled job connections to database)	Bidder requests bank to confirm on the current number of adhoc users for the ETL jobs ? API , BigData and Streaming jobs users ?	Reporting layer : ~ 3000 tables
665	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	Bidder requests bank to provide following inputs- a) No of Users b) No of developers c) No of self-service users d) No of consumers only	Adhoc Users (directly accessing DB) => ~300 users API => 2 departments Streaming jobs => use cases under implementation
666	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	Bidder requests bank to confirm on the role/persona wise peak user concurrency? For example Sales manager/RMs/HO/Zos etc.	Bidder need ot propose the best suitable solution.
667	87	35	The web portal of Business Intelligence tool should support at-least 25000 concurrent users, scalable up to 75000 in next 5 years, accessing various reports generated	Bidder requests bank to provide the concurrency on dashboards as of today?	Bidder need ot propose the best suitable solution.

Sr No	EOI Page No.	EOI Clause No.	Existing Clause	Query/Suggestions	Bank Response
668	74	C4. Functional Requirements	24 Annexure E gives sample use cases which are to be implemented on Next Gen Data Warehouse using structured and/or unstructured and/or semi-structured and/or any other kind of data gathered from either Data Warehouse or Data Lake or Data Virtualization or all together or any other source.	Data virtualisation does not support unstructured data , request to remove.	Bidder need ot propose the best suitable solution.