

REQUEST FOR
EXPRESSION OF INTEREST (EOI)
FOR
“NEXT-GEN DATA WAREHOUSE
SOLUTION”

STATE BANK OF INDIA

Request for EOI No.:
SBI/GITC/Data Warehouse/2018/2019/34
Dated: 23.03.2019

Data Warehouse Department,
GLOBAL IT CENTRE,
SECTOR-11, CBD BELAPUR,
NAVI MUMBAI-400614
(MAHARASHTRA)

Contents

Introduction.....	3
Schedule of Events.....	3
Background.....	4
Applicant’s Eligibility Criteria	5
Technical Parameters of the solution/Scope of Work.....	5
Process before submission of EOIs.....	5
Format and Signing of EOI.....	7
Process after submission of EOIs	7
Terms & Conditions.....	8
Disclaimer	9
Annexure A - Eligibility Criteria	11
Annexure B - Technical Criteria/Scope of Work	12
Annexure C - Monthly Data processed in DWH Warehouse	38
Annexure D - Existing Data Warehouse Architecture.....	39
Annexure E - Functional Use Cases	40
Annexure F - Summary of SBI Internal IS Policies	44
Annexure G - Next Gen Data Warehouse Sizing	46
Annexure H - Acronyms	47
Annexure I - Definitions	50

Introduction

State Bank of India (SBI) Group is the biggest financial services conglomerate in India. Headquartered in Mumbai, SBI provides a wide range of products and services to individuals, commercial enterprises, large corporates, public bodies and institutional customers through its various branches and outlets, joint ventures, subsidiaries and associate companies. The Group comprises of State Bank of India (SBI), its various non-banking subsidiaries/ joint ventures, and foreign banking subsidiaries/ joint ventures.

A Fortune 500 company, SBI has entered into the league of top 50 global banks with a balance sheet size of over Rs 30 lakh crore, over 24,000 branches and 59,000+ ATMs serving over 42 crore customers after the merger of its five Associate Banks and Bharatiya Mahila Bank on 1st April 2017. SBI has an overseas presence through 195 foreign offices spread across 36 Countries.

SBI invites Expression of Interest (**EOI**) from companies/firms (**Applicants**) wishing to be considered for providing solution for Next-Gen Data Warehouse Solution, hereinafter referred to as the **solution** as detailed in this EOI document.

Please note, the objective of this Request for EOI is to identify all possible solution (s) for the scope of work defined in this document.

Schedule of Events

1	Date of commencement of EOI Process	23 rd Mar 2019
2	Last date and time for receipt of queries (through emails only) for clarification from applicants	1 st Apr 2019 5.00 PM IST
3	Raise a gate-pass request to join briefing session	3 rd Apr 2019, 5.00PM IST
4	Briefing Session	5 th Apr 2019 3.00PM – 5.00PM IST
5	The queries received (without revealing source of query) and response of the Bank thereof will be posted on the Bank's website by	10 th Apr 2019
6	Last Date and Time for EOI Submission along with all supporting documents.	22 nd Apr 2019 up to 04.00PM IST In case the designated day happens to be a holiday; the next working day will be deemed as the last date for submission of EOI

Background

State Bank of India (SBI) is keen on embarking on a Data Analytics journey and gain an advantage over its competitors in terms of innovation, time to market, user delight, cost of service, despite its behemoth size and scale.

Current model of data processing focuses on reporting needs and captures data needed for end of day reports / credits - debits / assets - liabilities etc. There is a potential of losing out on important insights / patterns in data from this approach. It also means that analysts will not be able to use an exploratory / experimental workbench for data analysis.

There is an increasing trend globally in large Banks to use customer data in near real time from their mobile click-stream / touch-stream data, purchase patterns to make customized offers / improve service experience.

With data at SBI growing at an astronomical rate, a robust, horizontally scalable and cost-effective infrastructure for data storage and processing is the need of the hour. The ability to handle complex data from semi-structured and unstructured data sources, in addition to the conventional structural data processing, maximizes value potential and quality of services in the financial sector. Operational efficiency in the digital age is one of the factors that determine a Bank's growth. By managing petabytes of data efficiently, SBI has the potential to be a trend setter for smart data management globally. Empowering representatives of Bank at every level with analytics, thus standardizing and speeding up the decision-making process, will be a natural outcome of smart data management.

Expression of Interest (EOI) are invited in sealed envelope superscripted as **<Expression of Interest – Next-Gen Data Warehouse Solution>**

- a. From the applicants who meet the eligibility criteria as set out in **Annexure-‘A’**,
- b. Who have solution strictly in line with the technical parameters as set out in **Annexure-‘B’** and
- c. Agree to abide by the terms and conditions contained in this Request for EOI document.

Sealed envelope containing complete set of hard copy of EOI and a soft copy thereof (in a CD/DVD) should be submitted by Post to or delivered in person at the below mentioned office: -

The Dy. General Manager (Data Warehouse),
State Bank Global IT Centre, Data Warehouse Department,
Fourth Floor, Tower 7, Belapur Railway Station Building,
Sector-11, CBD Belapur, Navi Mumbai – 400 614.
Phone: Ms. Y Sobha (022 2752 4102)

Applicant's Eligibility Criteria

This process is open to all applicants who fulfill the eligibility criteria as set out in **Annexure 'A'** of this document. The applicants should furnish information on the lines of Annexure-A in their EOI proposal.

Technical Parameters of the solution/Scope of Work

The applicant should describe how their solution will fulfill the requirements as desired in **Annexure 'B'** of this document. The applicants should furnish information on the lines of Annexure-B in their EOI proposal.

Process before submission of EOIs

- i. **Raising of queries/clarifications on Request for EOI document:** The applicants requiring any clarification on this document should submit their written queries through email only to:

Main SPOC:

Mr. Pushkar Kulkarni (Project Manager)

022 2752 3772

pushkar.kulkarni@sbi.co.in

Keeping in CC:

Ms. Y Sobha (Deputy General Manager)

022 2752 4102

dgmit.dwh@sbi.co.in

Mr. B Kirouchenamourty (Assistant General Manager)

022 2752 3730

agm.dwh@sbi.co.in

Mr. Rohit Kasle (Assistant Manager)

022 2752 3912

rohit.kasle@sbi.co.in

- ii. Vendor requiring any clarification of the bidding Document may notify the Bank in writing **strictly as per the format** given in below the address/by e-mail given in this document within the date/time mentioned in the schedule of events.

Vendor Name	Sl. No	EOI Page No	EOI Clause No.	Existing Clause	Query/ Suggestions

- iii. **Modification in Request for EOI document:** At any time prior to the deadline for submission of EOIs, SBI may modify any part of this document. Such change(s) if any may be in the form of an addendum/corrigendum and will be uploaded in Bank's website -<https://bank.sbi>. All such change(s) will automatically become part of this Request for EOI and binding on all applicants. Interested applicants are advised to regularly refer the Bank's URLs referred above.
- iv. **Extension of date of submission of EOIs:** Request for extension of date for submission of EOIs will not be entertained. However, the Bank at its discretion may extend the deadline to allow prospective applicants a reasonable time to take the amendment/changes, if any into account.
- v. **Briefing session:** SBI will arrange for a briefing session to answer queries at below venue:

SBI Red Conference Meeting Room,
 Tower 7, Fourth Floor,
 Belapur, Railway Station Building,
 CBD Belapur, Navi Mumbai - 400614

Two representatives from every participating Vendor will be allowed to join this briefing session. Please raise a gate-pass request on Email-ids given in Point-i of the same section to join this meeting. Format for raising gate-pass request as below:

Name
E-mail ID
Mobile Number
Organization
Laptop Serial Number (if required)

Format and Signing of EOI

- i. The applicant should prepare EOI strictly as desired in this Request for EOI document.
 - a. EOI should be typed and submitted on A4 size paper, spirally and securely bound and with all pages therein in serial order. EOI needs to be submitted in softcopy format also. EOI submitted in both hard copy & soft copy formats will be considered by the Bank.
 - b. All pages of the EOI should be signed by only the authorized person(s) of the company/firm. All interlineations erase, or overwriting shall be valid only if the person(s) signing the EOI authenticates them. The EOI should bear the rubber stamp of the applicant on each page except for the un-amendable printed literature.
 - c. Contact detail of the authorized signatory and an authorized contact person on behalf of the applicant is to be provided as under:

Particulars	Authorized signatory for signing the EOI	Authorized contact person.
Name		
Designation		
Email id		
Landline		
Mobile No.		
Fax No.		
Address		

- ii. The applicants should demonstrate in EOIs that they meet all parameters given in **Annexure 'A'** and **Annexure 'B'** of Request for EOI.
- iii. In case any discrepancy is observed between hard and soft copy, the hard copy will be considered as the base document.

Process after submission of EOIs

- i. All EOIs received by the designated date and time will be examined by the Bank to determine if they meet criteria/terms and conditions mentioned in this document including its subsequent amendment(s), if any and whether EOIs are complete in all respects.
- ii. On scrutiny, the EOIs found NOT in desired format/illegible/incomplete/not containing clear information, in view of SBI, to permit thorough analysis or failing to fulfill the relevant requirement will be rejected for further evaluation process.
- iii. SBI reserves the right, at any time, to waive any of the requirements of this Request for EOI document if it is deemed in the interest of SBI.

- iv. If deemed necessary, the Bank may seek clarifications on any aspect of EOI from the applicant. If a written response is requested, it must be provided within 3 days beyond the response received, if any will not be considered. However, that would not entitle the applicant to change or cause any change in the substances of their EOI document already submitted. Bank may also make enquiries to establish the past performance of the applicants in respect of similar work. All information submitted in the application or obtained subsequently will be treated as confidential.
- v. After examining the EOI, some or all the applicants may be asked to make presentation of the solution.
- vi. Nothing contained in this EOI shall impair the Bank's Right to issue 'Open Tender' on the proposed solution.

Terms & Conditions

- i. Lodgement of an EOI is evidence of an applicant's consent to comply with the terms and condition of Request for EOI process and subsequent bidding process. If an applicant fails to comply with any of the terms, its EOI may be summarily rejected.
- ii. Willful misrepresentation of any fact in the EOI will lead to the disqualification of the applicant without prejudice to other actions that the Bank may take. The EOI and the accompanying documents will become property of SBI. The applicants shall be deemed to license, and grant all rights to SBI, to reproduce the whole or any portion of their product/solution for evaluation, to disclose the contents of submission to other applicants and to disclose and/ or use the contents of submission as the basis for EOI process.
- iii. SBI reserves the right to accept or reject any or all EOIs received without assigning any reason therefore whatsoever and the Bank's decision in this regard will be final. No contractual obligation whatsoever shall arise from the EOI process.
- iv. Any effort on the part of applicant to influence evaluation process may result in rejection of the EOI.
- v. SBI is not responsible for non-receipt of EOIs within the specified date and time due to any reason including postal delays or holidays in between.
- vi. SBI reserves the right to verify the validity of information provided in the EOIs and to reject any bid where the contents appear to be incorrect, inaccurate or inappropriate at any time during the process of EOI or even after award of contract.
- vii. Applicants shall be deemed to have:
 - a. examined the Request for EOI document and its subsequent changes, if any for the purpose of responding to it.

- b. examined all circumstances and contingencies, having an effect on their EOI application and which is obtainable by the making of reasonable enquiries.
 - c. satisfied themselves as to the correctness and sufficiency of their EOI applications and if any discrepancy, error or omission is noticed in the EOI, the applicant shall notify the Bank in writing on or before the end date/time.
- viii. The vendor shall bear all costs associated with submission of EOI, presentation/POC desired by the Bank. Bank will not be responsible or liable for any cost thereof, regardless of the conduct or outcome of the process.
- ix. Applicants must advise the Bank immediately in writing of any material change to the information contained in the EOI application, including any substantial change in their ownership or their financial or technical capacity. Copies of relevant documents must be submitted with their advices. For successful applicants, this requirement applies until a contract is awarded because of subsequent bidding process.
- x. Participating applicants must not advertise or publish the same in any form without the prior written consent of SBI.
- xi. Brief overview of the proposed procurement/scope of work given in this document may be further elaborated, viz., more details may be included in the Request for Proposal (RFP) document to be issued because of evaluation process of EOIs.
- xii. SBI shall have the right to cancel the EOI process itself at any time, without thereby incurring any liabilities to the affected Applicants. Reasons for cancellation, as determined by SBI in its sole discretion include but are not limited to, the following:
 - a. Services contemplated are no longer required.
 - b. Scope of work not adequately or clearly defined due to unforeseen circumstance and/or factors and/or new developments.
 - c. The project is not in the best interest of SBI.
 - d. Any other reason.

Disclaimer

SBI is not committed either contractually or in any other way to the applicants whose applications are accepted. The issue of this Request for EOI does not commit or otherwise oblige the Bank to proceed with any part or steps of the process.

Subject to any law to the contrary, and to the maximum extent permitted by law, SBI and its directors/officers/employees/contractors/agents and advisors

disclaim all liabilities (including liability by reason of negligence) from any loss or damage, cost or expense incurred or arising by reasons of any person using the information and whether caused by reasons of any error, omission or misrepresentation in the information contained in this document or suffered by any person acting or refraining from acting because of any information contained in this Request for EOI document or conduct ancillary to it whether or not the loss or damage arises in connection with any omission, default, lack of care or misrepresentation on the part of SBI or any of its officers, employees, contractors, agents or advisors.

Please Note: Since this is not a Request for Proposal (RFP), commercials are not required to be submitted at this stage.

Annexure A - Eligibility Criteria

Eligibility Criteria

S. No.	Eligibility Criteria	Compliance (Yes/No)	Documents to be submitted
1	The vendor must be a firm/ company / organization registered under Companies Act/Partnership Act/LLP Act etc.		Copy of the Certificate of Incorporation issued by Registrar of Companies and full address of the registered office
2	Vendor should have existing Next-Gen Data Warehouse solution as mentioned in the EOI		The details to be provided as per Annexure- 'B'.
3	The solution should have been implemented in at least 2 large scale organizations.		Two reference with following details for each reference to be provided: <ol style="list-style-type: none"> 1. Name of the Organization 2. Name of the Official 3. Contact number of Official 4. E-mail Id of Official
4	The company/firm should be profit making organization for last 3 years.		Audited Balance Sheet for the last 3 years should be enclosed.
5	The company/firm should not have been declared ineligible for corrupt and fraudulent practices by the Govt. of India / State Governments / Regulatory Agencies/ PSU/ Private Company.		Declaration to be provided

Annexure B - Technical Criteria/Scope of Work

<Next-Gen Data Warehouse Solution> Technical Criteria/Scope of Work
--

Below provided are key requirements depicting different capabilities expected from new set-up. Vendors are free to propose their architecture to meet Bank's requirements in this EOI.

- **End state objectives** from the architecture;
 - Data Warehouse, Data Marts and Data Stores
 - Use case driven and scalable Data Lake
 - Data Virtualization/Federation
 - Data Archival and backup solution
 - Sandboxes for research and development purpose
 - Migration from existing setup to proposed solution
 - Ensuring Data Quality & Data Reconciliation
 - Data Governance including Data Dictionary and Data Lineage
 - Data Science Platform/Tool with AI/ML capabilities and real time analytics
 - Business Intelligence Tools and real time reporting
 - Compliance to Global standards and SBI Internal policies
 - Framework for regulatory reporting
 - Data protection, Data security, Data privacy
 - Real time Data ingestion with spontaneous reconciliation
 - Log Storage/Archive Solution

Vendor may propose one or multiple solutions to meet the scope of work of this EOI. Vendors must give comparison between multiple options highlighting pros and cons (In case they propose more than one option).

EOI Proposal should include following items for each proposed option;

- The Following Specifications for each of PROD, DEV, UAT and DR environments;

- Architecture Diagram
 - Network Diagram
 - Process/Data Flow Diagram
 - Proposed software specification along with OEM and features of each software product
 - Proposed hardware specification
 - Proposed internal and external network specification
 - Rackspace, Power, Cooling, Network connectivity & Bandwidths
 - Any other relevant artefacts
- Detailed Migration Plan including timelines from existing to new setup
 - Cost model (How the licensing will be done)- Actual commercials are not required at this stage
 - Team structure (without actual profiles)
 - Performance benchmark for Next Gen DWH
 - Tentative Project Timeline

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 tables as Compliant (Yes/No).

Note – We have given indicative numbers of Jobs, Tables, Storage, Users etc in this EOI. However new system proposed by vendors should be flexible enough to support additional Jobs, Tables, Storage, Users etc as required by the Bank.

• **Critical Functional Requirements:**

Sr. No.	Parameters	Compliant (Yes/No)
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.	
2	Data may be structured, semi-structured, and unstructured. It may come from internal or external sources. It may come in batches, incremental additions or real-time feeds. There should be no limitation on the type, format and size of data ingested. Data may include log, feeds, audio, video, image, NOSQL, RDBMS, unstructured text, through ERP systems, etc	

3	GUI based framework to configure sources to NEXT-GEN DW	
4	Ingestion subsystem should allow to configure ingestion processes from single / multiple source system, single / multiple files, single / multiple operational input files	
5	Tools used for Data Ingestion should be platform and database independent and should be compatible to ingest and replicate data on parallel processing	
6	End objective for the data ingestion is to publish the dashboards for end users or any job related to reporting and analytics max by 8.00am on next business day 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	
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.	
8	Existing ETL Jobs to be Fine Tuned. Re-runability checkpoints should be present in ETL jobs. New ETL jobs should be able to parallel read and write data.	
9	Auto Resume – Aborted ETL jobs must resume from check point without any manual clean up in associated table or file being written	
10	Ingestion subsystem should generate audit and diagnostic logs	
11	One of the most important feature is the richness of the transformations to do day-to-day tasks, such as; Data conversion, lookup, expression, joining records, splitting data, filtering, ranking, sorting, grouping, looping, and combining data, pivot/unpivot, converting dates, setting variables based on parameter files, merging rows, finding the latest file, and splitting data based on certain conditions, running web methods, transforming XML documents, rebuilding indexes, sending emails, profiling data, handling arrays and records, processing unstructured data, masking, monitoring the inbound data flow for completeness, consistency and accuracy, wizards to assist creating complex packages, like loading fact tables, or type two slowly changing dimensions (SCD – T2)	
12	An alerting report and monitoring utility about the ingest pipelines should be available as part of the solution.	
13	Trigger mechanisms in identifying any structural changes at source	
14	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.	
15	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.	

16	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 ~ 2 TB each scalable up to 10000 logs. Proposed solution should be capable of scrapping logs generated by any type of Database. E.g. Oracle Database, IBM DB2 Database etc.	
17	Solution should be able to handle DDL change without manual reorg/runstat. It should handle network fluctuations and hindrances.	
18	A job scheduler, along with process management controls that provide things like runtime monitoring and error alerting, handling, and logging.	
19	When an ETL package runs, it is very important to be able to log how it executed i.e. we need to know how long it takes, what time it started/ended, who triggered it, was it successful or failed, what was the error message, etc. Apart from logging, when an ETL tool runs, it also needs to be able to run at scheduled times, re-run when failed, and limit the execution duration. ETL meta data repository should be query-able by designated user to extract this information.	
20	Daily and monthly data volumes for ingestion into the current DWH given in (Annexure - C), to provide an idea of the requirement for your calculations. These are only indicative numbers of the usage; NEXT-GEN DW ingestion requirements will be not only higher in volume but also have far more variety compared to these numbers.	
21	Vendor should propose which technology is suitable for each kind of upstream data ingestion like Data Warehouse, Data Marts, Data Lake, Use Data Virtualization/Federation layer, etc.	
22	Tool should have in-built capability to optimize/prioritize the ingestion pipeline system as a whole. Foe example if multiple pipelines are batched in the system tool should determine the cost of the pipeline and optimize & prioritize accordingly.	
23	Tool should have the capability to notify any transmission loss through connectivity, network, system, hardware or any other failures. It should automatically resume from the checkpoints with minimal or no manual intervention.	
Data Storage		
1	Vendor should propose effective number of data storage layers in NEXT-GEN DW between data ingestion and data consumption.	
2	Vendor should propose the type of storage to opt for in NEXT-GEN DW (SQL, NO-SQL, etc) and provide details of the hardware requirements, supported open source / proprietary components, skill requirements for storage management and any other relevant artifacts for each storage subsystem	
3	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.	

4	Storage replication (e.g. RAID) should be automatically managed by the platform.	
5	Tool should have capability to store/swap data in memory, disk and distributed storage areas depending on the age of the data determined by its usages through user queries.	
6	User should be able to work on DB even while backup is in progress. They should be able to run statistics and reorganize their tables. Any background process including backup must not hamper performance of user queries.	
7	Processing engines should have capabilities to process all kinds of Data Storage formats supported by the platform.	
8	Storage should support data compression. It should be possible to perform both fast compression and efficient compression based on data processing needs.	
9	The storage should be horizontally and vertically scalable. Redistribution of data across the NEXT-GEN DW should be possible automatically and seamlessly.	
10	Ensuring real time health checks, monitoring and alerting about data storage / utilization of storage / failure handling of storage components. Actionable dashboard must be available to designated users to monitor health checks and tool should automatically issues alerts to users.	
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).	
12	Downstream departments (data consumer) to be given separate processing power, storage to undertake their requirements with separate DB snapshot, Audit trails should be available for any user accessing the Databases. Construction of this separate Database snapshot and enabling this audit trails must not cause any major systemic issues/challenges in smooth functioning of primary DB.	
13	It should be possible to project and view data through multiple modes using the Storage on NEXT-GEN DW. Varieties of GUIs should be available to project or view the output generated through analytic processes. For instance: The Bank may decide to implement use-cases that project transactions data as a graph data structure. The Storage solution on NEXT-GEN DW should allow for such projections.	
14	The NEXT-GEN DW will have storage intensive, compute intensive and balanced workload scenarios. The storage mechanism should serve as a single entity that can be utilized by multiple processing engines or compute nodes designated for a specific purpose in the NEXT-GEN DW.	
15	The storage layer should be able to accommodate data used for specific processing needs. For example: Storage is expected to allow data reside in multiple layers. The NEXT-GEN DW should be flexible to handle data needs for Staging, direct data ingestion / landing, projections of data, organizing data and aligning it with specific schema type(s) etc.	

16	Vendor should list out comprehensively the types of risks they expect from the storage subsystem (e.g., loss of data blocks due to insufficient replication, data corruption, problems of insufficient storage during operations, inefficient retrieval of blocks or files, etc). For each such risk, they should describe the processes and mechanisms that would be put in place to avoid and/or mitigate the risk.	
Data Processing Framework		
1	For the data to be accessible and consumable by businesses / downstream applications, the NEXT-GEN DW should have robust, highly efficient and parallel execution of data transformation jobs.	
2	The NEXT-GEN DW ecosystem should have state of the art data processing engines that can perform in-memory processing to reduce the time for data transformations and query in case of real time requirements.	
3	Framework should allow joining multiple sources/tables/inputs etc.	
4	Framework should be able to put the processed data in a predefined area separate from the raw data, while minimizing data replication.	
5	Framework should be capable of performing validation checks pre-and post-processing.	
6	Have a workflow management and scheduling solution to schedule data transformation, data acquisition or data delivery jobs. Allocations of separate workload channel to designated queries	
7	Should have failure, retry, alert and escalation logic defined also successful file transfer should be automatically acknowledged	
8	Should have audit and error logs for auditing and troubleshooting	
9	Automatic recovery of data after failure/rejection of record needs to happen without any manual intervention	
10	Framework should deter improper user practices and facilitate to implement best data governance structure without impacting performance of the system.	
11	Framework should have mechanism to protect data at rest and at motion from unauthorized user access and amendments.	
12	ETL/ELT tool for data extraction should be AI/ML features for suggesting / improving Query / ETL / ELT Stages	
13	Existing reports and extracts generation jobs on DWH should be analyzed and transformed to the NEXT-GEN DW. The vendor should use preferably off-the-shelf tools and not resort to building from scratch.	
14	Data transformations should be triggered in parallel. The NEXT-GEN DW should be capable to run multiple transformation jobs in parallel. The NEXT-GEN DW should be able to run at-least 1500 jobs in parallel, scalable up to 5000 in next 5 years, of varying complexity - simple, medium, complex, in batch or near real time mode every day.	
15	There should be provision to monitor, log and assess the progress of each transformation job. Job design (s) should be re-usable. There should be a provision to organize individual transformations as user defined operators. Provision should be there in framework to tie these operators to data structures and organize the entire process as a single job.	

16	Organizing jobs, scheduling jobs and triggering jobs should be possible from an easy to use interface. Reports on job status and success / failure / retrigger should be sent to concerned stakeholders on a continuous basis. The tool(s) used for data transformation should be aligned to data sources and types / formats of data residing on NEXT-GEN DW.	
17	The processing pipelines for ETL/ELT jobs also include real time, daily, weekly, monthly, quarterly and annual reports, feeding data structures for downstream consumption. These activities are in-scope for this engagement.	
18	<p>Transformations for this activity can be categorized into the following types:</p> <ul style="list-style-type: none"> Existing transformations in DWH that needs to be migrated to NEXT-GEN DW New transformations for data sources that are not sourced by DWH Transformations and data processing pipelines for real time data capture 	
19	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, 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.	
20	Tool should be flexible enough to implement the data privacy and security policies of the Bank while transforming and storing the data.	
Data Federation/Virtualization		
1	Vendor to propose a 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	
2	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.	
3	Semantic integration of structured & unstructured Data.	
4	Data virtualization should support the use of APIs.	
5	All data should be made discoverable and integrable easily through a single virtual layer which will expose redundancy and quality issues faster.	
Migration from Existing Setup to Proposed Solution		
1	Vendor should propose a detailed seamless automated migration plan from existing setup to proposed solution. Plan should focus on less manual intervention, data reconciliation between the systems and minimum parallel run of existing and proposed solution.	
2	Data migration from Staging and Data Marts, user tables and any other schemas identified by Bank.	
3	Data migration from existing archival solution to new one.	
4	Migration of existing data sourcing ETL jobs.	

5	Migration of existing data extraction and reporting jobs.	
6	Migration of monitoring dashboard data points.	
7	Migration of user details.	
8	Migration of Data Governance, Data Lineage and Data Quality rules and policies	
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.	
10	Vendor should list out all types of risks they expect during the migration. Vendor should provide justification if any downtime is required on existing or proposed system during migration. Vendor should provide all the pre-requisites for the migration in the proposal.	
11	Vendor to review the existing architecture during migration and remove duplication of data and recommend improvements in overall setup if any	
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 Next-Gen DW which will save cost to the Bank. (Annexure D gives the technology architecture of the current setup)	
Disaster Recovery		
1	Bank proposes to setup only functional DR to start with. At later stage Bank may take decision to setup full scale 100% DR.	
2	Functional 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 (Statistics provided in Next Gen DW Sizing section).	
3	The NEXT-GEN DW – DR solution needs to be set up at a remote location at Hyderabad.	
4	The DR solution should be synced with production NEXT-GEN DW. The SLA for RTO should be maximum 2Hrs as per Bank’s defined policy.	
5	Fully functional NEXT-GEN DW – DR can either run the jobs and have checkpoints for data synchronization or have a solution by replicating data from NEXT-GEN DW to NEXT-GEN DW – DR	
6	The proposed solution is expected to have a monitoring engine that can determine the health of production NEXT-GEN DW and raise alerts / trigger remedial actions to bring NEXT-GEN DW – DR as the default NEXT-GEN DW	
7	Switching over to NEXT-GEN DW – DR must have limited manual intervention. The handover should be soft and must not incur any data loss / data corruption. Jobs in progress must be restarted with minimal impact (including data loss) with migration & automation tool.	
8	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.	
9	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 NEXT-GEN DW	

10	The Bank reserves the right to enforce mock up scenarios for failure to ensure that the NEXT-GEN DW – DR solution is working as expected. The vendor should support the Bank to demonstrate such scenarios	
11	Vendor to propose frequency of updating functional DR.	
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	
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	
3	All the applications connected to the non-archived data should be available with archived as well	
4	Automated process for archiving data as per Banks policy	
5	Store backup of entire ecosystem on suitable cost-effective, fast recovery infrastructure (Currently tape backup is taken)	
6	Mixing and Merging data from Current/Live to and from Archival must not result in any significant loss of performance and response time	
7	Archival and Backup setup must support automated Data Reconciliation whenever movement from Current/Live happens	
Cloud Integration and Migration		
1	NEXT-GEN DW should be able to consume data from external cloud-based infrastructures.	
2	Cloud integration / data transfer to and from public/private/hybrid cloud should be available using all standard protocols. (Web requests / secure transfer channels etc)	
3	The architecture of NEXT-GEN DW should enable interaction between public cloud and designated edge servers alone. NEXT-GEN DW non-edge server cannot be directly exposed to the public cloud.	
4	Transfer out of NEXT-GEN DW to public cloud should not be possible by all roles in NEXT-GEN DW. All activities with data transfer from Public cloud should be logged for audit and monitoring.	
5	In view of the intent to reduce the hardware footprint (in future), the technical architecture of NEXT-GEN DW solution should be flexible to accommodate provisioning of NEXT-GEN DW on cloud. The Bank understands that there can be differences in services offered by cloud service providers. The NEXT-GEN DW solution architecture should be designed considering as-is infrastructure availability in cloud.	
6	Adherence to global standards related to cloud	
Monitoring Dashboard		
1	Real time data flow in dashboard	
2	Health monitoring of NEXT-GEN DW ecosystem	
3	Drag and Drop feature to customize data	

4	Monitoring of all data sourcing tables/data marts on daily basis. Data Ingestion status pass / fail / in-progress for all source systems	
5	Operational logs analysis highlights	
6	Scalable distributed monitoring system for high-performance computing systems such as clusters and grids	
7	Industry best tools for performance (slowness) monitoring	
8	Controlling threshold breaches in pre-defined standard ways	
9	End to end performance monitoring with automatic alerts	
10	Threshold control to kill the high resource consumption query	
11	Monitoring of allocations of separate workload channel to designated queries	
12	Status of all applications / interfaces running on NEXT-GEN DW along with memory, disk, CPU usage etc	
13	List of daily missing files from source systems	
14	Email alerts from portal to downstream applications if data not loaded	
15	Status of NEXT-GEN DW clusters, hardware status/software status etc. with need-based access	
16	Back Dated Data changes needs to be updated on portal	
17	Source System should be having access to portal for seeing the processing (rejection) status of the data provided by them	
18	Monitor privileged users to monitor DBA activities & share history	
19	Automated alerts for jobs failure, jobs successful should be available	
20	List of Job created not compiled, executed – should be available	
21	Mechanism to ensure that file transfer failures are automatically detected	
22	DB health monitoring tool as well as monitor of N/W health	
23	Usage statistics of tables, jobs, reports, dashboards etc.	
24	Data Reconciliation status for every data movement on real time basis	
25	All above parameters to be displayed in console with real time updates without any manual intervention	
Data Governance		
1	Provide traceability – it should be possible to track and visualize any data transformation or any rule applied to data in the source system -> Next Gen DWH -> Downstream systems	
2	Provide trust – The system should be able to ensure the users that they are accessing data from the right source of information.	
3	Provide auditability – the solution should record any access to the data to satisfy compliance audits. For example, it should be able to check on who touched the data, when did they touch it, is there a chain of custody issue, is there transparency in terms of data privacy and protection etc.	
4	Enforce security and privacy – Data inside the NEXT-GEN DW will be accessed by only authorized users. Data at rest/in-motion should be encrypted	
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 NEXT-GEN DW when required by only authorized ID's.	

6	Capability to define clear roles and access management rules to user ID's	
7	Automated propagation of changes to NEXT-GEN DW Data Dictionary and business glossary by multiple sources as and when changes occur in source.	
8	Capability for house-keeping focusing on purging of tables/jobs/reports which are not in use.	
9	Capability to review the conformance to governance policies and processes periodically (monthly/quarterly) and make necessary changes.	
10	Provide detailed information about interfacing (upstream and downstream) applications.	
11	Data modeling capabilities to be provided by the tool	
12	Metadata Management Capability: Tool should cater to three broad categories of metadata; Business metadata, Technical metadata and Operational metadata	
13	Masterdata Management Capability: Master Data Management tool (s) should deliver consolidated, complete and accurate view of business-critical master information to all the operational and analytical systems across the Bank.	
14	Parallel processing: Data governance tool should be able to handle 500 concurrent users, scalable up to 1000 users in next 5 years, running any kind of job (eg: Data Lineage on simple/medium/complex jobs running on multiple tables)	
15	Architecture Review: The architecture including all the applications and policies of the Next-Gen DW ecosystem to be reviewed by Vendor and make necessary changes to get the maximum benefits from the setup. Vendor should provide statistics which Bank may demand in connection with this like user details who is using the services, hit rates, application runtime and downtime, hardware memory and storage usages, etc	
16	Vendor should provide a robust change management framework to track versioning, change history with date and user information, track migration of code from development to production, security. Source code, IPR or customized IPR of Next Gen DWH will be owned by the Bank.	
17	Performance benchmark of all components of Next-Gen DW to be given by participating Vendors	
18	Data lineage management - Framework should enable the Bank to map the phases of the information flow and document the transformation applied to data instances along the flow.	
Web-based Data Dictionary		
1	Designated SPOC at source system and DWH should be able to update Data Dictionary in Real-time.	
2	Search by keyword	
3	Management of referential integrity	
4	Tying various Data Dictionary items to broader business concepts which can then be used for profiling and analysis of business attributes.	
5	Standard services for accessing the data dictionary.	
6	Audit trail/log	
7	Help facility – this helps to instruct users on how to use the data dictionary	

8	Security features: - to help in protecting the information contained in the data dictionary.	
9	Available to end users to put their remarks and data dictionaries of user created tables.	
10	Auto generated reminders to SPOC's for periodic reviews of Data Dictionary.	
11	Data Dictionary must have the provision of marking golden source of data.	
Data Quality		
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	
2	Parsing and standardization — Decomposition of text fields into component parts and formatting of values into consistent layouts based on industry standards, local standards	
3	Generalized “cleansing” — Highlight data values which are failing to meet domain restrictions, integrity constraints or other business rules for user review	
4	Recommend Matching —Identification, linking or merging related entries within or across sets of data for user review	
5	Profiling — Analysis of data to capture statistics (metadata) that provide insight into the quality of the data	
6	Monitoring — Deployment of controls to ensure ongoing conformance of data to business rules that define data quality for the organization	
7	Recommend Enrichment — Enhancing the value of internally held data by appending related attributes from external sources	
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.	
9	Connectivity to multiple data sources - Tool (s) should be able to connect to a wide selection of data source types, both from a data management system -- e.g., RDBMS vs. NoSQL database -- and from a platform -- e.g., on premises vs. cloud -- basis.	
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.	
11	Continuous ETL/ELT and Analytical data quality checks	
12	Mechanism to capture feedback from end users to report Data Quality issues	
13	Dedicated Data Quality team who will thoroughly verify the Data Quality across setup	
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)	

15	<p>Each month 5 data quality use cases will be developed and implemented on Next Gen DW. Examples of use cases are as given below;</p> <ul style="list-style-type: none"> • Profiling of Customer Master table for verifying PAN, Mobile Number, Date of Birth, Address, Pin Code, etc • Profiling of Branch Master for verifying branch address, contact information, branch manager/staff information, etc 	
16	System detectable Data Quality rules using business concepts defined at Data governance framework	
Data Reconciliation		
1	Developing automated reconciliation methodology suitable for each source system depending on the nature of data. 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. Inherent database capabilities of ingestion pipeline content management DB must be leveraged for this purpose	
2	In the case of failure in reconciliation, tool should be able to drill down to the stage to identify the error and should be able to rectify the errors on near real time basis. AI/ML capabilities can be leveraged for reject processing wherever applicable.	
3	<p>Typical errors we face for record rejection during sourcing;</p> <ul style="list-style-type: none"> • Null values assigned to non-null column • Special characters • Data range out of bound • Connection with source system lost, etc 	
4	Near real time dashboard to be published displaying records in source system and records received in DWH and action take for any discrepancies	
5	Data should be reconciled in every movement across table in any/all layers including data marts in automated way at the time of loading itself.	
Security and 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.	
2	Authorization and access control - Role-based access controls should be provided for authorization of account-related and data-related activities.	
3	Data protection - It should be possible to protect the data in the NEXT-GEN DW throughout its lifecycle including data at rest and data in motion.	

4	Auditing - Audit or diagnostic logs should be used to log management-related activities or data-related activities. Log management and auditing of all critical activities on NEXT-GEN DW is a critical requirement. The Bank reserves right to ask the vendor to produce / analyze logs for reporting purposes.	
5	Data Democratization - Secure access of PROD Database to LHOs/GOCs	
6	Data Leakage - Security CIA parameters should be achieved, and tools should be able to find and alert on Data leakage	
7	Compliance with SBI IS Policies – some of the key areas are as under (Summary of Key policies is given in Annexure F):	
8	Compliance to Global Standards <ul style="list-style-type: none"> • 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. 	
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.	
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.	
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	
4	Dashboards – Web browser dashboards can be used to present information and manage processes, such as data being used, data change history, job versions, current status, etc	
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.	
6	Drill Down and 100% Auditability – Every step in the data collection, enrichment and consolidation should be tracked and recorded, including all comments.	
7	Electronic Submission – Should support for all regulators globally in all required formats, including XBRL, XML or other file-based electronic submission.	

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.	
9	High Volume, High Performance and Reliability – Scalable and resilient architecture which will handle all volume and performance demands.	
10	Leveraging Data and Processes – All data, calculations and other processes defined for one purpose are stored and should be reused wherever required to address new regulations and internal management 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.	
12	Tool must support generation of reports in XBRL format	
13	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.	
Audit and Log Management		
1	Logging of operational activities: Support the logging of all user activities without slowing down the performance.	
2	Should have metadata enabled reporting mechanism on run time log.	
3	Should have audit and error logs for auditing and troubleshooting	
4	Extracting useful information from system logs to understand the efficiency of system and any fraud	
5	User should be able to control the events/items being logged. For e. g user can opt for enabling logging of failures only.	
6	Vendor to propose solution with cheap storage options for log storage of all the Bank’s applications and mechanisms to extract requested information from the logs as and when required	
Data Encryption		
1	The sensitive and personal identification data on NEXT-GEN DW should be encrypted	
2	Data encryption applies to the following data on NEXT-GEN DW: <ul style="list-style-type: none"> • Data at rest • Data in motion • Data available over an API to external sources / applications • Data at end users • Data from source system during ingestion 	
3	Standard encryption/decryption techniques / policies should be implemented for this activity. Decryption should be allowed only to selective users/roles by complying the global standards like DES, SKC, PKC, 3DES, MD5 cryptographic hash, etc.	

4	The vendor team will determine tables / columns/ Cell-level encryption/Table space-level encryption that needs encryption by discussing with data owners before implementing the solution.	
5	The processing time for data should include time needed for encrypting and decrypting information needed for downstream / report generation. The overall SLA for data processing should be adhered to, keeping data encryption as an important activity.	
6	Proposed Solution should be capable on ingesting encrypted data from source system. It should support the encryption/decryption mechanism implemented at source system.	
Data Masking		
1	Capability to disable unmasking of masked data according to users' requirements	
2	Tool should enable selective masking of columns in user specified tables	
3	The output must be repeatable - The same source data, masked repeatedly by the same masking methodology, must yield the same output	
4	Maintain referential integrity - Make sure the masked data is usable	
5	Data Masking solution should support various methodologies like but not limited to below examples; <ul style="list-style-type: none"> • Substitution • Nullifying and spacing • Number and date variance • Format-preserving encryption 	
User Management		
1	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.	
2	User Access Management should allow IT administrators to securely manage access to services/information/ resources for all the users.	
3	Access rights to a user should be allocated and modified on the principle of least privilege and "Need to know" or "Need to do/have" basis.	
4	The access privileges associated with each system product, e.g. operating system, network, database, application and system utilities, and the users to which these privileges need to be allocated should be clearly identified and documented.	
5	Privileges granted to the user should not conflict with Access Management - Segregation of Duties policy of the Bank.	
Business Continuity Plan		
1	Plans, measures and arrangements to ensure the continuous delivery of critical services and products, which permits the organization to recover its facility, data and assets.	

2	Identification of necessary resources to support business continuity, including personnel, information, equipment, financial allocations, legal counsel, infrastructure protection and accommodations.	
3	Points to be taken care during preparation of BCP plan: <ul style="list-style-type: none"> • Plans must be updated and tested frequently • All types of threats must be considered • Dependencies and interdependencies should be carefully analyzed • Availability of key personnel • Network and telecommunications • Alternate sites • Employee support 	
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.	
2	Should support bulk data and targeted data extracts via statistical tools and APIs	
3	NEXT-GEN DW should support open source library / proprietary / supported open source statistical and machine learning tools configured for analysis and reporting	
4	Facility to generate and distribute canned / automatic bursted reports from NEXT-GEN DW to downstream end users like BID, Analytics, CRM, YONO, OFSAA, etc	
5	Self-service portal to extract the data on their own (Should support Data Democratization)	
6	Dedicated high-performance department wise sandboxes allocated to end users for R&D	
7	Dedicated Data marts to be created as per BFSI industry Standard models and previous experience. All the reporting and analytics request to be sufficed with the same. Requirements for Data Marts to be reviewed periodically.	
8	Real time reporting can be done through Staging area of Data Warehouse	
9	The vendor as mandated by the Bank will streamline ad-hoc reports / visualizations using the data discovery capability or assist business teams build pipelines for analysis using data discovery capability.	
Data Science Platform with AI/ML Capabilities		
1	Implementing end to end analytics use-cases as mandated by the Bank	
2	Power data / objects to existing analytics models built on proprietary tools (IBM SPSS). Migration of such models to new solution	
3	Availability Pre-build models which can be directly used with Bank's data to get insights	
4	Model building capabilities on agile mode	
5	AI/ML framework that run on scalable architectures	
6	Templates for common data processing / end to end applications, tag data sets and re-use operators across machine learning pipelines	

7	In-memory computing & integration with Spark, Redis, etc	
8	Analytics on real-time data in real-time/near real-time	
9	GPUs to be incorporated in solution if possible using HDFS Hadoop like environment for better analytical results	
10	Analytics notebooks which should be shared, and experiments be replicable across users / teams performing similar analysis	
11	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	
12	Platforms facilitate the automation of tasks such as feature engineering and hyper parameter tuning	
13	Access management at the data, workflow and models	
14	Automate workflow execution and scaling of the servers	
15	Build and Publish detailed reports, insights on web portal	
16	Deploy models/workflows as industry standard web service	
17	Centralized monitoring and management	
18	All machine-learning platforms either support multiple models out of the box or provide an option to custom-code the same	
19	Integration with R, Python, Keras, Tensorflow, Theano, scikit-learn etc and other frameworks / languages	
20	Stable Product	
21	Support on demand basis.	
22	Performance and scalability	
23	Enables ease and speed with which the user can move models from a developed environment to a production environment, or embed them in a business process	
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.	
25	Vendor to provide solution / tool (s) for below scope of activities on SBI data sets; <ul style="list-style-type: none"> • Benchmarking • Predictive & Prescriptive Analytics • Social Media Analytics • Web Analytics • Geolocation Analysis • Ad-Hoc Analysis • Trend Indicators • Profit Analysis • In-Memory Analysis • Statistic Analytics • Data Mining 	

	<ul style="list-style-type: none"> Machine Learning 	
26	Pipelines to orchestrate the entire machine learning process from data preparation to deployment	
27	The ability to launch and redirect training to CPU and GPU-enabled resources.	
28	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.	
Business Intelligence Tools		
1	Capability to connect to various data sources: Access to various databases and file types, such as comma-separated values files, text, Excel and XML, SQL/NoSQL databases. Bank may raise specific requirements as and when required.	
2	Data filters and drilldown: The product should enable user to filter the contents in a tabular report or visualization by data values. The product should also enable the user to drill down from summarized to more detailed data and then drill up	
3	Web-based client user interface: The product's client user interface for the BI consumer role should be web-based.	
4	Independent and interconnected mashups: When the business intelligence style enables a single-screen to display of multiple visualizations, including tabular reports, the software should allow these visualizations to be either independent of each other or interconnected. If they're interconnected, data filters and selections will affect all the visualizations; for example, if any attribute is selected, all the visualizations will share that attribute.	
5	Visualizations: BI tools must provide below different types of visualizations; <ul style="list-style-type: none"> Animations, Barcodes Bar, line, pie, area and radar chart types Tables, Graphs, Infographics, Filters Widgets Drag and Drop Creation, Customization Templates Freehand SQL Command Geospatial Integration Layouts Themes Ability to mix and match various combinations 	
6	Reporting on all types of available of Data Formats; <ul style="list-style-type: none"> Structured, semi-structured, unstructured Click stream data Audit Logs Documents Multimedia data (Images/Videos/Audios) XBRL format IRIS iFILE framework 	
7	Security: Both user and user role-based security, designating who can create, modify, publish, use and administer the BI applications.	

8	Print and export: The product must enable you to export print visualizations and tabular reports to PDF or other graphics. Tabular reports need to be exportable to text files at a minimum and, preferably, to spreadsheets.	
9	Select data for analysis. BI tools must enable the user to select the data used in decision-making analysis and present it as a pivot table-style interface where dimension attributes are placed in rows and columns, measures are selected, and filters are applied.	
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.	
11	Reporting framework / model – Tool must support “easy to create and maintain” reporting framework / model containing all business concepts / measures used in reporting.	
12	Create measures: The product must enable the user to create and save measures or calculations for use in analysis. These are also referred to as performance measures or key performance indicators.	
13	Create hierarchies: The product must enable the user to create dimensional hierarchies, such as by geography or product, to group and summarize data. This establishes the drill-down paths.	
14	Save queries and analysis: The product should enable the BI user to save the data filters, selections and drill-down paths used in decision-making process so that they can be reused.	
15	Create and publish by business users: The product must enable the user to save and share his or her analysis with other BI consumers.	
16	Context-based filters: Filters will list only the choices that have values that fit the current selection of facts and dimensions.	
17	Context-based visualizations: Only visualizations or chart types that are relevant to the data selected will be listed as options.	
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.	
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.	
20	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.	
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.	

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.	
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.	
24	Performance monitoring: BI applications that monitor report and data usage enable a BI group to improve analytical performance for the business, eliminating bottlenecks and enabling users to assess infrastructure needs.	
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.	
26	Ease of analytical use: There should be different criteria defined for each type of user, such as information consumer, business analyst and IT.	
27	Ease of creating BI applications: There should be different criteria for each type of analytics creator, such as business analysts and IT.	
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.	
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.	
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.	
31	There should be separate criteria for BI user online help versus technical documentation.	
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.	
Hardware Specifications		
1	Vendor to supply, install, test, commission, manage and maintain the required IT Hardware for Next Gen DW.	
2	Software and Solution proposed by vendors should be compatible with all types of Hardware.	
3	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.	

4	Vendor proposed hardware is expected to be enterprise class, best of breed, tested and stable release.	
5	The proposed architecture considers vertical and horizontal scalability as one of the most important design principles.	
6	Vendor to propose hardware specifications for each component of Next-Gen DW ecosystem like Data Warehouse, Data Marts, Data Lake, Data Archival, Data Federation/Virtualization, Data Science Platform, Backup, Sandboxes, Functional DR, etc. for PROD, DEV and UAT environment as applicable	
7	Vendor must ensure that the proposed servers are fault-resilient with the most comprehensive features and functionalities to ensure maximum system uptime.	
8	The proposed hardware must not fall into 'End of Support' for at least 7 years from the date of delivery to the Bank.	
9	The hardware will be delivered in a staggered manner and Vendor to provide a plan for the same	
10	Vendor must provide detailed configuration of the proposed Hardware, including Hosting Space Requirements, Racks, Power, Cooling and any other requirement for the fulfillment of the Vendor's obligation in this EOI.	
11	Vendor will supply hardware resources and related services at the desired locations (Production and DR)	
12	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	
13	Vendor to provide all necessary hardware and software required to make the solution work	
14	The Vendor shall be responsible for the following duly verified by the Bank's team <ul style="list-style-type: none"> • Rack Layout along with cable layout • IP Schema details and VLAN layout • Ensuring proper connectivity and access to Next-Gen DW setup • Network and Security Hardening • Server Setup and configuration • Storage and backup configuration • Hardware Failover, Functional DR architecture and IT continuity approach (Only related to hardware) • Performance and benchmarking approach and plan 	
15	Architecture diagram Deployment plan - Vendor to submit architecture diagram of entire setup with network and security equipment required. Bank may change it after vetting by Information Security Dept and / or Enterprise architecture Dept. It will be binding on vendor.	

16	<p>The Vendor shall ensure all Installations & Implementation to be done by OEM badged resources only which will include;</p> <ul style="list-style-type: none"> • Preparation of racks for hosting including all required cabling & all other activities required for installation of hardware • Installation and implementation of all the servers (Production, Development, UAT, DR) including network cabling. • BIOS-configuration configure management port and assigning IP's for managing systems remotely, RAID Configuration. • OS (Linux) Installation, OS Configuration, OS Management Capabilities, Formatting the disks, OS and Network configuration • OS packages Repository Configuration. • Installation and Configuration of Network equipment • Installation and Configuration of Storage and Backup equipment with Hot, warm and Cold data segregation • Installation and Configuration of Security equipment • Setting up the testing of the platform and UAT. • All work related to cabling to utilize network infrastructure will be done by Vendor. • All work related to patch panels will be done by Vendor. • 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 	
17	<p>One Time Review Services for Production, DR, UAT and Dev environment;</p> <ul style="list-style-type: none"> • Review the HLD & LLD of the hardware setup • Reviewing the Platform security requirements & security architecture • Reviewing the implementation & administration approach • Reviewing Platform administration and management requirements • Post Hardware Setup Review & sign off based on the design document provided by the Bank • Ensure the High Availability & uptime requirements in line with the SLA 	
18	<p>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.</p>	
19	<p>Vendor will create/update process and process manuals for the management of complete environment. This will also include Installation, Housekeeping, Recovery and other operations manuals and share with the Bank</p>	
20	<p>Bank will perform a third-party validation of Hardware at its own cost to ensure all framework specifications, deployments are as per guidelines and other parameters as decided by the Bank to ensure proper functioning after commissioning of all equipments. Vendor will have to close all observations identified in third party validation. Approval from third party vendor of the Bank will be treated as completion of installation successfully.</p>	
21	<p>Vendor's team should monitor all supplied components and environment, pro actively report and resolve issues as a part of warranty and AMC obligations.</p>	

22	The Hardware solution must be compatible to integrate with various systems in the Bank including but not limited to SOC, PIMS, NOC, Command Centre, ITAM, Service Desk, ADS, and SSO etc. at no extra cost. Vendor will have to give appropriate support to the Bank during integration with various components of IT environment.	
23	System Administration Support- Service Provider must provide 24X7 supports for the Administration, Maintenance, Up-gradation (related to hardware) and other related activity to keep system running so that high availability can be assured.	
24	Vendor needs to provide Helpdesk support 24X7 to the Bank for end to end support for hardware maintenance.	
25	Vendor will have to provide support for Inventory, Asset register, License for Hardware and bundled software.	
26	Vendor will have to promptly follow Bank's processes/policies for movement of HW, Manpower Resources in and out of the Data centre.	
27	Vendor needs to give full support to the Bank related to any security upgrades, implementations, VAPT closure, SCD, support to audits, patch management/TL upgrade of HW & bundled software on time to time basis as deemed by the Bank and close all the gaps identified in such activities within given timelines by the Bank as a part of warranty and AMC obligations.	
28	Vendor will be required to comply with cyber security guidelines of the Bank which shall be revised from time to time as a part of warranty and AMC obligations.	
29	New patches are to be proactively identified, tested and applied by the Vendor during the period of contract. A proper record of testing, installation will be required to be maintained by the Vendor.	
30	Vendor is required to provide the minimum resources to monitor & manage the infrastructure, however it is the Vendor's responsibility to right size the resources to meet the SLA	
31	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.	
32	Development should be made in DEV only and testing in UAT. Framework to be implemented by Vendor to avoid improper user practices.	
33	Next-Gen DW 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)	
34	Next Gen DW should support at-least 200 concurrent users, scalable up to 500 users in next 5 years, on Data Science platform performing advanced analytics and running models. GPUs can be proposed for running Machine Learning/Neural Networks models.	

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	
36	NEXT-GEN DW is expected to have more users and the solution should not be bound by any license model for number of users	
37	Ad-hoc jobs of any complexity should not hamper the scheduled jobs performance.	
38	The NEXT-GEN DW solution at SBI should be enabled for exact point-in-time recovery of files or directories in the cluster from accidental deletions or corruption due to user or application error.	
39	While building NEXT-GEN DW, failure handling should be a design characteristic. There should be tool(s) that can notify, alert and if possible, predict node, disk or application failure.	
40	Necessary corrective action should be in place for each failure. Failure of node, disk or any network component should not affect the functionality of NEXT-GEN DW.	
41	Failure tolerance is not restricted only to server and application failure. Failures of network, hardware, software or nodes are considered failure scenarios in scope.	
42	Standard practices of configuring NEXT-GEN DW for High Availability and storage namespace Federation/Virtualization (at least 99.99% - calculated monthly basis) should be implemented. NEXT-GEN DW solution should serve users even when confronted with node failures or network partitions.	
43	New system should be available 24x7 without any failure.	
44	Vendor need to propose a solution for data migration / transfer between Existing DWH (Navi Mumbai Location 1) and NEXT-GEN DW-PR (Navi Mumbai Location 2) and also between NEXT-GEN DW-PR (Navi Mumbai Location 2) and Hyderabad (DR) or any other places for PR and DR decided by the Bank.	
45	Database should be linearly scalable which can expand the database capacity by just adding more nodes to the existing database. If the data volume grows more hardware can be added and expand the database capacity	
46	DEV and UAT environments are to be set up at the PR site	
47	The Vendor also needs to provide the configuration for setting up of functional DR along with DEV and UAT for each and every component / application of Next Gen DW ecosystem.	
48	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.	
49	It will be appreciated if the vendor comes with Green computing technology across infrastructure setup (which minimizes the electricity consumption of high computing computers).	

50	Annexure G gives the NEXT-GEN DW 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.	
51	Vendor to submit all back-to-back agreement copies between Vendor and SI / OEM / Parent company etc if any and tenure of the back-to-back agreement should be same as selected Vendor's agreement with the Bank	

Note: The applicants may please note that the above parameters may be further elaborated, viz., more details may be included in the Request for Proposal (RFP) document to be issued because of evaluation process of EOIs.

Annexure C - Monthly Data processed in DWH Warehouse

Sr. No.	Source system	Monthly Data Flow Volume Dec-18 (GB)
1	Archived log extract CBS (SBI) + TF (SBI)	SBI-CBS => 3500 (Avg Daily Logs) * 1.8 GB(AvgLogSize) * 31 (days) = 195300 (GB monthly) = (Approx) 190 TB (Monthly)
2	Flat file extract for SBI INVM	Daily File =>151 (Total Files) * 0.9 (GB Avg File Size) * 31 (days) = 4212.9 (GB) Monthly File => 2TB(Approx)
3	Flat file extract from ATM, INB, PSG, and SBI Card	1.PSG Weekly=>102 MB per week*4=408 MB (Avg File Size) 2.SBI Card Weekly=>1 GB per week*4=4 GB (Avg File Size) 3. [~22 GB per day*31=~682 GB]load for file ATM,INB base]=1147
4	Contribution from other source systems like DMAT, CMP, SBI Life, LOS, etc.	Monthly 200 GB(Approx)

Above figures are approximate sizes of data being processed by ETL / other tools at DWH and may not translate to actual incremental monthly loads into DWH.

Annexure D - Existing Data Warehouse Architecture

Sr. No.	Purpose	Statistics
1	Database appliance	500+ TB of data with compression index of 2.5 Comprises of 4+ data marts and 20000+ tables
2	Data Sourcing and Extraction Jobs	17000 Data sourcing and data extraction jobs
3	Scrapping of Jobs from Source	Scrapping of daily 5TB + logs
4	Reporting	300+ Interactive reports 100+ Busted reports
5	Software Version Control	
6	Hardware Monitoring	Server memory and compute monitoring of total 80+ production servers
7	User Management	User database of 30000+ officials
8	Job Scheduling	Running 2000+ scheduled jobs daily and additional 1000+ for weekly/bi-weekly/monthly/quarterly
9	Backup	
10	Web GUI	
11	File Sharing	
12	Archival	270+ TB Data with compression index of 2.5
13	Data Governance and Metadata	
14	Data Quality	
16	Analytics	
17	Budgeting	
18	Bank Master	10TB Oracle database

Annexure E - Functional Use Cases

Sr. No.	Group	Use Case	Problem Statement	Objective
1	Sales & Service	Behavioral Analytics: Customer Segmentation & Profiling	Acquiring and retaining 'profitable' customers is challenging in the competitive financial services market and in view of the rising customer acquisition costs, it is valuable to retain loyal customers	SME & Retail Customer segmentation based on Transaction behavior
2	Sales & Service	Customer Analytics- Up-sell	Customers may have the same or a similar product but might be very different in profitability and marketing efforts can leverage this information to sell a higher margin or higher value product to a profitable customer. An upsell model evaluates this insight at a customer level	Identify customers for specific product offer
3	Sales & Service	Expanding Book Size Through Cross Sell Programs	Various campaigns are run to target existing customers for cross-selling other products. Budgets allocated for running such campaigns are limited. Thus, models to improve response rate to these cross-sell campaigns are significant. Higher product penetration per customer also discourages customer attrition, improves customer loyalty	To predict the likelihood of customer's response to a specific product campaign/offer
4	Sales & Service	Down-Sell Analytics	Although Banks often look for ways to sell pricier products, offering the customer a lower-end product or service can be more profitable through volumes and the best way to build loyalty for some customers	Identify the subset of customers who are likely to be more active/profitable with a lower value product example lower ticket size loan, lower interest rate
5	Sales & Service	Customer Propensity Modelling	Identify New Business opportunity.	Identify effects of underlying factors on customer's affinity towards product, service and/or channel.

6	Sales & Service	Customer Churn Prediction for Effective Retention Programs	The switching of customers to other financial service providers, results in loss of profits (possibly for the entire customer lifetime), and loss from other value contributions from the customer, for example, potential recommendations	Reliable, early, and point-in-time detection of customer's attrition
7	Sales & Service	Win Back Propensity Modeling	For any Bank, a key variable is the rate of churn and vast sums are spent on marketing to try to replace all those defectors. New research shows that companies should focus more on winning back lapsed customers by smart strategies	Analyze dormant or lost customers and Identify target customers for win back campaigns
8	Sales & Service	Customer Acquisition – Market Segmentation & Application Scoring	Develop real-time insights into individual prospects to more effectively align the organization's interactions with them as they make their way through the sales process	Creating prospect segments based on their propensity to respond for customer and determine a cut-off score for high risk prospects.
9	Sales & Service	Lead Allocation/ Routing Analytics	Ensure better allocation of resources	Allocate right human resource to convert the lead based on Historical trend, ability of human resource
10	Sales & Service	Lead conversion Forecasting & Scoring	Lead conversion rate is not as expected	Identify internal & external factors affecting lead conversion and forecast the factors for upcoming quarters
11	Sales & Service	Sales Target / Budget Setting	Sales targets / Budgets should be based on customer insights to increase sales revenue and profitability and operational efficiency and to manage risk more effectively.	Setting up sales targets based on customer preference & behavior
12	Sales & Service	Customer Profitability	Increasing wallet share by optimizing costs	Create customer centric strategies and plans that maximize the Bank's profitability
13	Sales & Service	Performance & Profitability – Branch	Branch operations - towards business improvement	Analysis of Branch performance towards the increase of ROI

14	Sales & Service	Performance & Profitability – Channels	Channel optimization towards business improvement	Optimizing channel (i.e. YONO/ INB / MB / Branch / SMS / Email, etc) strategy to maximize profits by product; location etc.
15	Sales & Service	Budgeting models		Budget would-be set-in conjunction with the analysis of targets for the marketing and sales functions.
16	Risk	Credit Scoring		Credit decisioning process while sanctioning a loan
17	Risk	Pre-Delinquency and Loss forecasting		Proactively resolving potential delinquencies before they materialize.
18	Risk	Delinquency Management – Default/NPA analysis		Prediction of Default/NPA customers
19	Risk	Collection Optimization- Process and Cost optimization		Identify slow paying & non-paying customers and Improve cash collection strategy
20	Risk	Organized Fraud/ Collusion Detection	Fraud/ AML practice	Analysis and identification of factors causing Fraud/AML practice
21	Risk	Opportunistic Fraud Detection	Identification of Fraud and its prevention	Identify and predict potential fraud across the Bank (I e. Cheque fraud, Remittance fraud, Card fraud, Online fraud, etc)
22	Risk	AML detection and alert management	Identification of Money Laundering activities and its prevention	Identification of money laundering activities based on account transaction and behavior
23	Risk	Asset Quality Optimization	Optimization of Asset allocation based on valuation and risk	Optimization of Asset allocation based on valuation & risk. Improve the risk strategy of the organization
24	Risk	Concentration & Exposure Management		Identification of the diversified needs across portfolios and respective allocation anomalies

25	Wealth	Portfolio Recommendation: Creation of investment profiles for individuals based on their social media footprint and other digital touch points of the Enterprise		Investor's segmentation and risk profiling to create an individual specific investment profile matching risk –return expectation and economic behavior.
26	Wealth	Portfolio Monitoring & Servicing: Recommending / Selling the Right Product or Service to a Customer based on their Life Stage		Analysis of Customer Age, Employment Status, their Transaction behavior and push a recommendation to the Customer on their preferred channel of a Product or Service
27	ATM	Approach for ATM location optimization		Create a plan for ATM location by understanding demand in each sub-location for all customers for expected profitability.
28	ATM	ATM Network Usage		Analysis of ATM usages, cash levels, types of transactions, usage by other Bank customers, usage by violators, etc
29	Marketing	Campaign Performance Analytics – Market Mix Modelling	The optimal allocation of marketing funds has become an increasingly complex problem in view of the multi channel options for promotions.	To quantify the impact of marketing decisions of the past and predict future sales
30	Marketing	Marketing Mix Optimization		identification of appropriate mix of marketing channels for optimal utilization of available resources
31	Product	Product Penetration: Market Basket Analysis		Identify product penetration across dimensions such as; geography, product groups, customer base, channels
32	Risk	Transaction Fraud Prediction	Identification of Suspicious transactions on Real Time Basis for all digital channels	Identification of Suspicious transactions on Real Time Basis for all digital channels

Annexure F - Summary of SBI Internal IS Policies

Sl. No	Policy Statement
1	Post Implementation evaluation of the benefits of procurement should be done for all purchases The evaluation should compare the extent of benefits achieved against the initial requirements and also rate the performance of vendor/ solution for future reference.
2	Service Level Agreements (SLA) together with other supporting service agreements, supplier contract and corresponding procedures should be defined with suppliers providing services related to management of State Bank's information assets.
3	Change control committee (CCC) should be established for all application groups to take decisions on changes to be implemented. For the changes which spans across departments, the Change Control Co-ordination Committee should be established with representatives from the respective areas. Modifications to existing applications shall follow Bank's change control procedures
4	Users should not install any software, freeware, shareware or application on their desktop/laptop that is not authorized for State Bank's business.
5	Whenever connecting to the LAN, users should ensure that anti-virus agent is updated with latest signatures on the portable device.
6	There should not be generic User ID or more than one user ID for a user on a single system with same/ different privilege(s).
7	All systems should be updated with the latest signatures
8	Applications shall be subjected to periodic security reviews as per criticality criteria and as and when the application is modified
9	Applications shall be subjected to periodic security reviews as per criticality criteria and as and when the application is modified
10	OS should be configured as per secure configuration document.
11	Content and information published on internet and intranet should be accurate, consistent, and current.
12	Bank's IT infrastructure should be updated with the latest patches including security patches and up-gradation patches through a centralized patch management system.
13	Business Continuity (BCP) plan shall be defined and implemented for all the applications and processes as per the latest version of Bank's "BUSINESS CONTINUITY PLANNING POLICY".
	BCP (Business continuity Planning) committee shall be formed as the decision-making body for the business continuity planning as per the latest version of Bank's "BUSINESS CONTINUITY PLANNING POLICY".
14	NTFS should be used on all partitions
15	Responsibilities for data and application privacy and confidentiality
16	Responsibilities on system and software access control and administration
17	Custodial responsibilities for data, software and other assets of the Bank being managed by or assigned to the Vendor

18	Physical Security of the facilities
19	Physical and logical separation from other customers of the Vendor
20	Incident response and reporting procedures
21	Password Policy of the Bank
22	Data classification and Data Encryption/Protection requirements of the Bank.
23	In general, confidentiality, integrity and availability must be ensured.
24	Both table and row/column level encryption solution
25	Masking and Unmasking feature for privilege data for Compliance purpose
26	Data Access Policy
27	Proactive Patch management
28	Cyber Insurance
29	Data Leakage Prevention policy

Annexure G - Next Gen Data Warehouse Sizing

PROD in TB			
	DWH	Archival	Data Lake
Dec-19	592.16	346.00	730.10
Dec-20	717.66	557.00	730.58
Dec-21	1075.72	612.00	731.25
Dec-22	1541.63	734.66	732.19
Dec-23	2251.96	838.71	733.51

UAT (25% of Sourced Data) in TB			
	DWH	Archival	Data Lake
Dec-19	94.41	NA	NA
Dec-20	120.43	NA	NA
Dec-21	204.04	NA	NA
Dec-22	314.03	NA	NA
Dec-23	484.48	NA	NA

DEV (10% of Sourced Data) in TB			
	DWH	Archival	Data Lake
Dec-19	25.50	NA	NA
Dec-20	37.77	NA	NA
Dec-21	48.17	NA	NA
Dec-22	81.62	NA	NA
Dec-23	125.61	NA	NA

DR (Critical Data identified by Bank) in TB			
	DWH	Archival	Data Lake
Dec-19	19.59	NA	NA
Dec-20	27.43	NA	NA
Dec-21	38.40	NA	NA
Dec-22	53.75	NA	NA
Dec-23	75.26	NA	NA

Annexure H - Acronyms

As used in the Document	Description
ADS	Active Directory System
AML	Anti-Money Laundering
API	Application programming interface
AI/ML	Artificial Intelligence/ Machine learning
ADF	Automated data flow
ATM	Automated Teller Machine
BFSI	Banking, financial services and insurance
BCBS	Basel Committee on Banking Supervision's standard
BIOS	basic input/output system
BCP	Business continuity planning
BID	Business Intelligence Department
CPU	Central Processing Unit
CDC	Change Data Capture
CIA	Confidentiality Integrity and Availability
CBS	Core Banking Solution
CRM	Customer Relationship Management
DAS	Data Archival Solution
DC	Data Centre
DCT	Data Collection Template
DDL	Data Definition language
DES	Data Encryption Standard
DL	Data Lake
DLP	Data Loss Prevention
DQ	Data Quality
DWH	Data Warehouse
DBA	Database Administrator
DEV	Development
DFRA	Digital forensic and readiness assessment
DR/PR	Disaster recovery /Production
DR	Disaster Recovery Centre
EOD	End of Day
EOI	Expression of Interest
XBRL	extensible Business Reporting Language
XML	Extensible Markup Language
ETL	Extract Transform Load
ELT	Extract Load Transform

FinCEN	Financial Crimes Enforcement Network
GDPR	General Data Protection Regulation
GUI	Graphical user interface
GPU	Graphics Processing Unit
HDFS	Hadoop Distributed File System
HLD	High Level Design
SCAPM	IBM Smart Cloud
IA	Information Analyzer
IGC	Information Governance Catalog
IS	Information security
ITAM	Information Technology Asset Management
IPR	Intellectual property Rights
ISO	International Organization for Standardization
INB	Internet Banking
IDS	Intrusion Detection System
IPS	Intrusion Prevention System
KRI	key risk indicator
LDAP	Lightweight Directory Access Protocol
LLD	Low Level Design
MDM	Master data management
MD5	Message-digest 5
NOC	Network Operations Centre
NTFS	New Technology File System
NEXT-GEN DW	Next Generation Data warehouse
NPA	Non-performing asset
OFAC	Office of Foreign Assets Control
OCC	Office of the Comptroller of the Currency
OLAP	Online analytical processing
OFSA	Oracle Financial Services Analytical Application
OEM	Original equipment manufacturer
PCIDSS	Payment Card Industry Data Security Standard
PSG	Payment System Gateway
PIMS	Privileged Identity Management System
POC	Proof of Concept
PKC	Public-key cryptography
RTC	Rational Team Concert
RPO	Recovery Point Objective
RTO	Recovery Time Objective
RAID	Redundant Array of Independent Disks

RDBMS	Relational database management system
RFP	Request for Proposal
R&D	Research and development
ROI	Return on investment
SCD	Secure Configuration Document
SEC	Securities and Exchange Commission
SOC	Security operations Centre
SLA	Service Level Agreement
SPOC	Single Point of Contact
SSO	Single Sign on
SME	Small and Medium Enterprise
SBI	State Bank of India
SPSS	Statistical Package for the Social Science
SQL	Structured Query Language
TL	Technology Level
TWS	Tivoli Workload Scheduler
UAT	Unit Testing
UAM	User Access Management
VLAN	Virtual LAN
VAPT	Vulnerability Assessment and Penetration Testing
YONO	You only need one

Annexure I - Definitions

As used in the Document	Description
ADS	Microsoft Active Directory services of the Bank for controlling user access
Audit Logs	Security related chronological record of events in various systems.
Automation	Process or procedure performed with minimum or no human assistance
Bank	State Bank of India (Domestic, Foreign branches/subsidiaries) combined
Bank IS Policy	Latest IS policy of the Bank
Business Glossary	Glossary of Business concepts and terms defined and used by Bank time to time
Business Intelligence	Creation of architecture to create and deliver business values from the raw data of the Bank
Business Metadata	Business attributes to help, promote the context meaning and search ability. It gives business users a way to search and understand data attributes based on business concepts
Business Units	Branches, Extension Counters, Service Units, Corporate Offices, Head Offices, Zonal Offices, Regional Offices, Other Back Offices/ Offices and any new branches/offices
CDC	Change Data Capture or any process of scrapping prevailing in any of the Bank's system either existing or proposed
Checkpoints	Snapshot at a specific time from which activity can be resumed without any loss of transformation/modification of data
Cloud	Data centres available to many users over internet or intranet
Columnar	Column based storage of data either in in-memory or in secondary storage
Data Stores	Repository of heterogeneous data
Data Warehouse	Existing Data Warehouse system of the Bank
Diagnostic Logs	Detailed report of activities performed / being performed by the System which will be used to diagnose any problem faced
DLP	Data Loss Prevention is a strategy for making sure that users do not send sensitive or critical information outside the corporate network
Edge Servers	A device providing entry point into the network
Error Logs	Detailed report of errors in the System which will be used to resolve same

Existing Setup	Existing Data Warehouse ecosystem of the Bank along with upstream and downstream links/interfaces/systems
Functional DR	DR with all the schemas/applications of existing setup with limited data as identified by the Bank
LHOs/GOCs	Local Head Offices/GITC Offsite Centre of the Bank
Near Real Time	Near real time refers to processing data as and when available with an acceptable processing time of a few minutes.
Operational Metadata	Lineage of data along with audit trails of success and failure of ingestion pipelines duly updated in a pre-determined frequency
Processing Pipelines	ETL, ELT, any data processing / transformation / sourcing/ extraction jobs
Project	The Vendors' delivery of the services, deliverables, supply, installation, testing and commissioning, integration of computer hardware / software and services with support under Warranty and annual maintenance contract, if required for the contract period.
Proposed Solution	Solution to be proposed by Vendors vide this EOI
Record Rejection	Rejection of full or part of data due to non-conformance to various rules set in system
Regulatory Reporting	Mandatory reporting to various regulators by the Bank
Reject Processing	Successful loading of reject records in reasonable time ensuring no impact on downstream processing/reporting
Semantic Integration	Process of integrating heterogeneous sources of data and interrelating them by leveraging the semantic information embedded within these sources
Services	Services ancillary to the supply of the equipment/product, such as transportation, transit insurance, installation, commissioning, customization, integration, provision of technical assistance, training, maintenance and other such obligations of the Vendor covered under the Contract.
Solutions/ Services/ Work/ System	“Solution” or “Services” or “Work” or “System” or “IT System” means all services, scope of work and deliverable to be provided by a Vendor as described in the EOI and include services ancillary to the development of the solution, such as installation, commissioning, integration with existing systems, provision of technical assistance, training, certifications, auditing, maintenance and other obligation of the Supplier covered under the EOI.
Storage Subsystem	Includes all the components of proposed data storage like Data Warehouse, Data Marts, Data Lake, Data

	Stores, and any other storage proposed by Vendor
Supported Open Source	Open source software supported by Vendor or his partners/joint ventures/parent company
System	Any existing legacy system in the Bank and proposed NEXT-GEN DW by Vendor
Technical Metadata	Technical attributes including data type, length, precision, format, nullability, etc
Tranche 1	Mandatory reporting to Reserve Bank of India (RBI) for Risk Based Supervision (RBS)
Validation	The process of checking or establishing the validity or accuracy of data/hardware/software
