Pakistan Revenue Automation (PVT.) Limited – PRAL

Proposal Central Data Hub (CDH)

Hardware, Software Supply & Infrastructure Implementation

1. Purpose of RFP

This RFP solicits bids for the supply, licensing, and infrastructure implementation of a production-ready Modern Data Platform, deployed strictly on on-premises infrastructure. The Platform shall serve as the Central Data Hub, the permanent intelligence backbone for data-driven, intelligence-led tax administration at national scale. This RFP covers hardware supply, software licensing, and infrastructure commissioning; application-level configuration, data pipeline build, and platform go-live are covered under a separately contracted Data Platform & Application Implementation Services engagement, coordinated per Section 19 (Vendor Coordination & Interface Management).

Current systems are optimized for transaction recording and filing processing. They do not support economic behavior analysis, compliance risk prediction, or enforcement optimization. The Central Data Hub shall close this gap by establishing an integrated, governed, and auditable data foundation across all tax domains.

The following components shall be supplied and licensed in their entirety, with the underlying infrastructure installed and commissioned by the Vendor:

- On-premises compute, storage, and networking hardware.
- Enterprise ELT pipeline tool license, with schema enforcement and data quality control capabilities.
- Distributed S3-compatible object storage software license, constituting the Lakehouse foundation.
- A distributed compute cluster (hardware and software license) for AI/ML workloads and micro-batch streaming; the Vendor shall propose the technology stack for real-time data processing.
- Massively Parallel Processing (MPP) Enterprise Data Warehouse license, with native object storage connectivity.
- High-availability in-memory caching cluster (hardware and software license) for low-latency operational queries.
- All associated commercial and open-source software licenses required to operate the complete Platform.
- A Data Governance Tool license.

The Platform shall be architected to AI-ready standards, including a Feature Store and semantic layer, to support integration of AI components under future procurement instruments.

1.1 Compliance and Vendor Accountability

- The selected Vendor shall assume single-point accountability for hardware provisioning, software licensing and installation, infrastructure commissioning, infrastructure acceptance testing, and structured knowledge transfer of the infrastructure environment to PRAL's Data and Operations team.
- Infrastructure Acceptance, as defined in Section 7 (Infrastructure Implementation & Commissioning), shall be achieved upon project completion, at which point the environment shall be ready for the separately contracted Implementation Vendor to commence data platform and application implementation.

2. Project Scoping & General Obligations

2.1 In Scope

- Supply and provisioning of all on-premises hardware including compute, storage, and networking infrastructure for entire solution.
- Supply of all commercial and open-source software licenses required to operate the complete platform for the full contract duration; no component shall be left unlicensed or subject to separate procurement.
- Installation, configuration, and commissioning of all supplied hardware and infrastructure components (compute, storage, and networking) up to the point of Infrastructure Acceptance, in accordance with Section 7 (Infrastructure Implementation & Commissioning).

2.2 Out of Scope

- Modifications, upgrades, or patches to any FBR/PRAL source system, including IRIS, eFBR, RegSys, TAMS, and VAT; source systems shall be consumed as-is.
- Development or modification of any FBR transactional or operational application.
- Procurement or setup of end-user devices, workstations, or operator terminals.
- Any data processing, storage, or transmission outside on-premises network perimeter.
- Data platform and application-layer implementation, including ELT/ETL pipeline development, data governance configuration, BI/AI-ML implementation, and any activity beyond Infrastructure Acceptance; these are covered under the separately contracted Data Platform & Application Implementation Services engagement

2.3 General Obligations

- The awarded Vendor shall assume full responsibility for the completeness, quality, and functional readiness of all hardware, software licensing, and infrastructure commissioning deliverables under this contract. Partial or component-only delivery shall not be accepted, and infrastructure performance shortfalls shall not be attributed to factors outside the Vendor's supplied scope.
- All hardware and infrastructure delivered under this contract shall operate exclusively within the designated on-premises network perimeter, and no data shall be transmitted, stored, or processed outside that boundary at any point.
- The Vendor shall coordinate infrastructure delivery and commissioning timelines with the separately contracted Data Platform & Application Implementation Services Vendor, in accordance with Section 19 (Vendor Coordination & Interface Management), to ensure Infrastructure Acceptance is achieved without delay to downstream implementation activities.

3.Business Goals

The objective of this procurement is to establish a robust, secure, and scalable on-premises hardware, software licensing, and infrastructure foundation for the Central Data Hub (CDH), enabling FBR and PRAL's subsequent transition to a data-driven, intelligence-led, and AI-enabled tax administration ecosystem under a separately contracted Data Platform & Application Implementation Services engagement.

This procurement shall deliver the following foundational capabilities:

- **Infrastructure Capacity & Scalability:** compute, storage, and networking infrastructure sized and licensed to support national-scale data volumes and concurrent workloads, with capacity for onboarding future sources and domains through configuration alone, without structural hardware changes.
- **Software Licensing Completeness:** all commercial and open-source software licenses required to operate the complete Platform for the full contract duration, with no component left unlicensed or subject to separate future procurement.
- **Infrastructure Security & Compliance:** hardware- and OS-level security controls (encryption at rest, network segmentation, access control) implemented in compliance with applicable regulatory and legal requirements, forming the secure foundation on which platform-level governance controls (RBAC, PII masking, audit logging) will subsequently operate.
- **High Availability & Resilience:** infrastructure architected for high availability, redundancy, and disaster recovery consistent with the reference architecture (Section 5.2) and sizing requirements (Section 6).

- **Infrastructure Readiness for Platform:** infrastructure installed, configured, and commissioned to a state of Infrastructure Acceptance (Section 7), enabling the separately contracted Implementation Vendor to commence data platform and application build activities without infrastructure-related delay.
- **Technology Refresh & Long-Term Viability:** a Vendor-delivered technology refresh roadmap covering the full warranty period, ensuring the infrastructure and licensed software remain supportable and extensible as CDH usage scales.

4.Success Criteria

The success of this procurement shall be measured against the following outcomes:

Infrastructure Delivery & Commissioning

- All hardware and software components delivered, installed, and commissioned per the approved Bill of Materials and the sizing specifications defined in Section 6.
- Infrastructure Acceptance achieved within the committed project timeline, with no critical or major defects outstanding at sign-off.

Licensing Completeness

- 100% of required commercial and open-source software licensed for the full contract duration; zero components left unlicensed or requiring separate future procurement.

Security & Compliance

- Hardware- and OS-level security baseline (encryption at rest, network segmentation, access control, directory service integration) implemented and validated.
- Zero tolerance for infrastructure-level data leakage; all infrastructure operating exclusively within the designated on-premises perimeter.

Availability & Performance

- Infrastructure availability and performance (network throughput, latency, HA failover) validated against the targets defined in Section 8 (Performance & Scalability) and Section 10 (HA, Backup & Disaster Recovery).
- Infrastructure sized and validated to sustain the workloads and concurrency levels defined in Section 6, without requiring re-architecture for at least the five-year sizing horizon.

Scalability & Future Readiness

- Modular infrastructure capable of onboarding additional compute and storage capacity through configuration alone, without structural changes or service interruption.
- A Vendor-delivered technology refresh roadmap in place, covering the full warranty period, identifying planned upgrades, end-of-life component replacements, and recommended evolution paths for all major infrastructure components.

Handover Readiness

- Structured knowledge transfer of the infrastructure environment completed to PRAL's Data and Operations team.
- Infrastructure documentation (as-built diagrams, runbooks, configuration records) delivered and accepted.

5.Functional and Technical Requirements

This section contains the Functional and Technical requirements that would enable the vendors to design the proposed solution.

5.1 As-Is Architecture

The current data ecosystem at FBR primarily comprises of two database systems. These are:

- Microsoft SQL Server
- Oracle Exadata

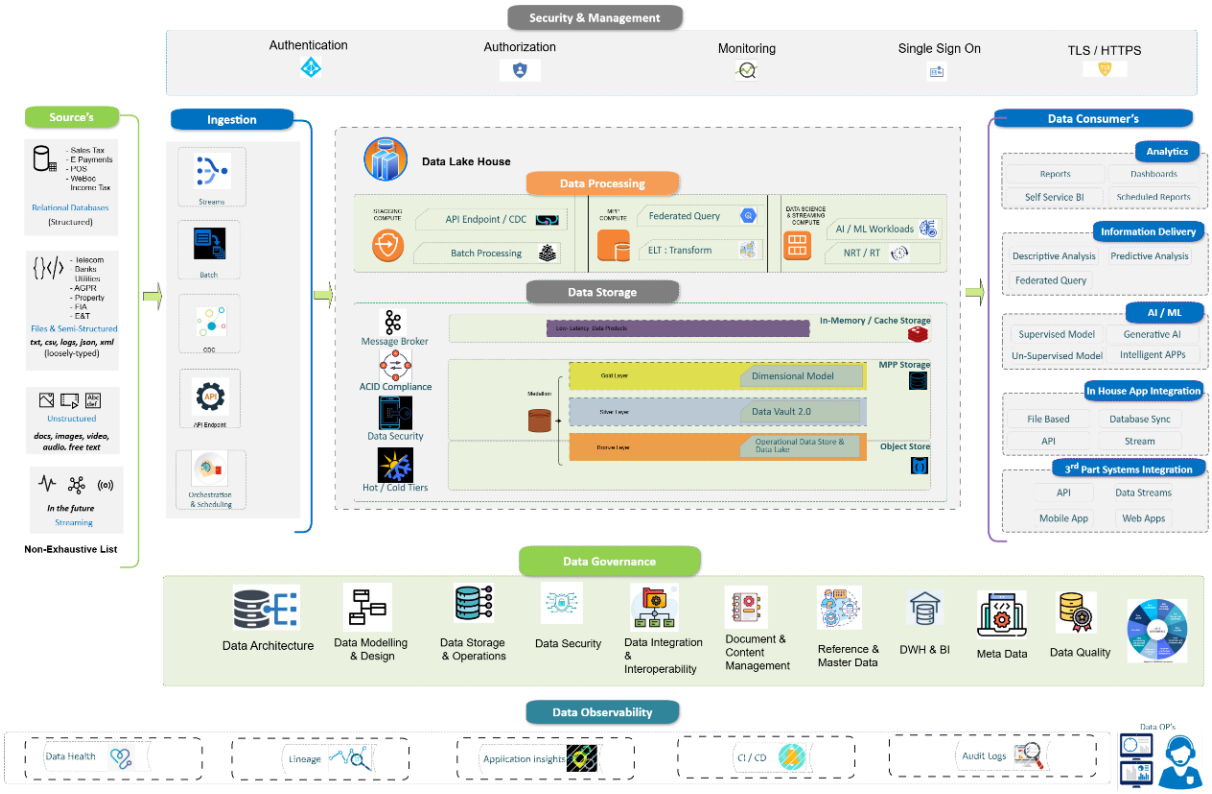
The SQL Server is a legacy system which was developed to serve the initial set of FBR applications. The web applications transactional data was stored in this instance. Apart from storing and serving the applications, the SQL Server also serves as an ODS (Operational Data Store) for downstream BI & Reporting. There is no proper Enterprise Datawarehouse construct for analytical reporting. Once all the data is collected in the ODS, it serves downstream dashboards, reports and replicated to Oracle Exadata ODS.

The Oracle Exadata serves the latest set of FBR applications like IRIS is also a transactional system. It is one appliance which is partitioned into three DB instances (Transactional System for IRIS, ODS & FBRDWH). Just like the SQL Server, the ODS in Oracle Exadata collects the replicated data from the Oracle Transactional tables and further serves it downstream for BI & reporting. The third DB instance FBRDWH is not a proper Datawarehouse but only serves Data Marts. In summary FBR today does not have a proper Enterprise Datawarehouse that can be used for Analytical reporting.

In this RFP, one of the primary requirements is to set up a comprehensive Modern Data Platform that ingests data from the SQL Server, Oracle transactional systems, Third Party, Commercial Projects and a non-exhaustive list of data sources. We will not use the existing ODS instances both SQL Server & Oracle as data sources.

Please see Annexture A listing data sources that will need to be ingested.

5.2 Proposed Reference Architecture



Data Sources

The architecture supports three categories of source data. Structured data comes from relational databases connected to operational systems like sales tax, e-payments, POS, and income tax. Semi-structured and unstructured data includes files, logs, JSON, XML, documents, images, audio, and free text. There's also a forward-looking "In the future" streaming category, covering IoT and real-time event sources.

Data Ingestion

Four ingestion patterns are supported. Streaming handles real-time data flows. Batch covers bulk periodic loads. CDC (Change Data Capture) tracks database changes incrementally. Consumption from API endpoints. Scheduling and Orchestration coordinate and automate the pipelines.

Data Processing

The Data Lakehouse houses three processing layers. The Staging Compute layer handles API-based data processing, CDC transformations, and batch processing. The MPP (Massively Parallel Processing) Compute layer supports Federated Queries and ELT transformations. Lastly compute clusters for AI/ML workloads, near real-time and real-time data processing.

Data Storage

"Storage is organized into three tiers. The Object Store (Bronze Layer) holds raw operational data and Data Lake files in open file and open table formats. The Silver Layer stores curated data using a structured, auditable modeling approach. The Gold Layer contains the Dimensional Model for analytics-ready consumption. Supporting this are an MPP Storage engine, In-Memory/Cache Storage for low-latency data products, and a Message Broker for event streaming.

Data Governance

This layer spans the full data lifecycle across ten domains: Data Architecture, Data Modelling & Design, Data Storage & Operations, Data Security, Data Integration & Interoperability, Document & Content Management, Reference & Master Data, DWH & BI, Metadata management, and Data Quality. This ensures consistency, compliance, and trustworthiness across all data assets.

Data Consumers

The right-hand side of the architecture represents all downstream users and systems. The Analytics layer serves reports, dashboards, self-service BI, and scheduled reports. Information Delivery supports descriptive and predictive analysis as well as federated queries. The AI/ML layer enables supervised and unsupervised models, generative AI, and intelligent applications. In-House App Integration covers file-based, database sync, API, and stream-based connectivity. Third-Party Systems Integration extends reach through APIs, data streams, mobile apps, and web apps.

Security & Management

A cross-cutting layer that spans the entire architecture. Authentication integration with on-premises Active Directory and LDAP. Vendor should propose components for Authorization and Data Access Policy Enforcement. Monitoring to be powered by SIEM tools. Single Sign-On is a requirement managed through an Identity Provider. All data in transit is protected via TLS (v1.3/HTTPS).

Data Observability

Provides end-to-end visibility across the platform. Data Health monitoring tracks pipeline and dataset quality. Lineage maps data origin and transformations. Application Insights monitors runtime behavior. CI/CD pipelines automate deployment and testing. Audit Logs capture access and change history for compliance.

Data Ops

Represented at the bottom-right of the observability layer, DataOps bridges the technical and operational aspects of the platform. It involves dedicated operations personnel who monitor pipeline health, manage incidents, coordinate deployments via CI/CD, and ensure SLA adherence across ingestion, processing, and storage layers. They work in close coordination with the observability tooling to maintain reliability and data trust across the entire ecosystem.

API Framework

The CDH shall incorporate a bidirectional, centrally managed API Framework that governs both inbound data ingestion via APIs and outbound data serving to internal and external consumers. This framework is a first-class platform component and must be designed, implemented, and governed as part of the core architecture.

Inbound API Integration (Data Ingestion)

- Support for REST, SOAP, GraphQL, and webhook-based API connections for ingesting data from source systems and external data providers.
- Configurable API connectors with scheduled, event-triggered, and real-time polling modes - manageable without bespoke development per source.
- Secure credential management, token rotation, and OAuth 2.0 / API key authentication for all inbound connections.
- Automatic schema detection, payload validation, and error handling with dead-letter queue management for failed API calls.
- Rate limiting, throttling controls, and retry logic to protect source systems from overload during ingestion.
- Full lineage tracking for all API-ingested data - origin endpoint, ingestion timestamp, payload version, and transformation history maintained end-to-end.

Outbound API Services (Data Serving)

- A managed API gateway layer enabling the CDH to expose governed data products, aggregates, and analytical outputs as secure, versioned REST APIs - consumable by

internal FBR applications, operational systems, and authorized downstream platforms.

- API catalogue maintained within the data governance layer - documenting all published endpoints, data products served, consumer access policies, SLOs, and versioning history.
- Business-requirement-driven API provisioning: new outbound APIs definable and publishable through a governed request and approval workflow without requiring core platform changes.
- Field-level access control on all outbound API responses - consumers receive only the data fields their role and clearance permit, enforced at the API gateway layer.
- API versioning and deprecation management - backward compatibility guaranteed for active API consumers with a minimum notice period before breaking changes or deprecations.
- Rate limiting, usage monitoring, and consumer-level throttling enforced across all outbound APIs - with usage dashboards available to the CDO Office.
- Support for both synchronous (request-response) and asynchronous (event-driven / webhook push) outbound API patterns to accommodate varied consumer system architectures.

API Governance & Operations

- All APIs (inbound and outbound) registered in the central API catalogue with ownership, SLA, data classification, and consumer mapping documented.
- API health monitoring integrated into the Data Observability layer - latency, error rates, availability, and throughput tracked per endpoint with alerting on SLA breach.
- API security scanning and vulnerability assessment included in the CI/CD pipeline for all new and updated API endpoints.
- Audit logs maintained for all inbound API calls and outbound API responses - capturing consumer identity, timestamp, endpoint, and payload size for compliance and forensic purposes.

5.3 Data Integration and Ingestion Platform

We require an Enterprise-grade Data Integration and Ingestion Platform comprising an Extract, Load & Transform (ELT) pipeline engine and, a distributed real-time message broker. The Integration Platform shall serve as the primary data movement and orchestration layer for the Central Data Hub, replacing an existing proprietary ELT tool

currently executing 3000 + active jobs. The platform shall be deployed strictly on-premises and shall satisfy all requirements specified in this section in their entirety.

Current-State Baseline

The specifications below are informed by an operational analysis of the incumbent ELT environment.

Metric	***Current State (Existing Tool)***
Total ETL Jobs	3,000 + active jobs across ODS and BI layers
Source Systems	IRIS, eFBR, RegSys, TAMS, VAT, and further domains (100+)
Scheduling Patterns	Interval (1-min to 10-hr), daily time-triggered (overnight batch), weekly triggers, Majority Jobs are Interval based.
CDC Dependency	Dominant ingestion pattern is CDC with File Loading also available.
Agent Distribution	4 agents

Universal Connectivity and Native Adapters

Generic JDBC/ODBC bridges shall not satisfy connectivity requirements where native connectors are commercially available. The following native connector capabilities are mandatory:

Connector Category	**Mandatory Requirements**
Relational Databases	Native adapters for Oracle (including RAC), Microsoft SQL Server, PostgreSQL, Teradata, and Netezza/IBM Pure Data. Each adapter shall support bulk extraction, parallel partition reads, and pushdown query optimization.
Object Storage and Lakehouse	Native read/write to S3-compatible on-premises object stores. Shall support Parquet, ORC, Avro, JSON, Delta, and Iceberg formats without third-party intermediaries.
REST and Web Service APIs	Ingestion from RESTful and SOAP APIs with native support for OAuth 2.0, API key, and mutual-TLS authentication. Pagination handling, rate limiting, and schema inference from JSON/XML shall be supported natively.
Flat File and Unstructured Sources	Ingestion of CSV, TSV, fixed-width, JSON, XML, and binary formats delivered via SFTP, shared network paths, or object storage. Schema-on-read shall be supported for semi-structured and unstructured sources.

Workflow Orchestrator Integration	Native integration with Apache Airflow and equivalent DAG-based orchestrators via published REST APIs or a certified operator/plugin library.

Change Data Capture (CDC)

CDC shall be treated as a first-class, architecturally central capability of the platform, not an optional module. Log-based CDC is the mandatory extraction pattern.

Requirement	Specification
Log-Based CDC	Redo log mining for Oracle; transaction log parsing for SQL Server. Zero impact on source system performance. Sub-second capture latency for INSERT, UPDATE, and DELETE operations.
CDC to Object Store	Delivery of CDC event streams to S3-compatible on-premises object storage. Shall support both append-only event log and merge/upsert patterns for Lakehouse current-state tables.
Schema Evolution Handling	Automatic detection and handling of column additions, type changes, and table rename. Configurable alerting with graceful pipeline continuation or controlled suspension based on change severity.
Message Broker Integration	The platform shall include or integrate a distributed, fault-tolerant message broker as a first-party suite component supporting at-least-once or exactly-once delivery semantics, configurable retention, consumer group management, and air-gapped on-premises deployment.

Job Orchestration, Scheduling and Dependency Management

The platform shall support all existing scheduling patterns (interval-based, daily time-triggered, on-startup, and weekly) and shall add event-driven orchestration capabilities.

Requirement	Specification
Hybrid Scheduling Engine	Interval-based scheduling from seconds to hours, event-driven triggers (file arrival, API webhook, message broker event, row-count threshold), and on-demand manual execution within a single unified interface.
DAG-Based Dependency Orchestration	All pipelines shall be expressible as Directed Acyclic Graphs (DAGs) with explicit dependency declarations. Downstream jobs shall not execute until all upstream dependencies complete successfully.

Parallel and Multi-Threaded Execution	Configurable parallelism at pipeline, step, and partition level. Automatic workload partitioning across available compute nodes.
Checkpoint and Resumption	Failed jobs shall resume from the last successfully committed micro-batch or partition. Full source reprocessing shall not be required upon failure.

High Availability and Fault Tolerance

Requirement	Specification
Clustered Worker Architecture	Horizontally scalable worker/agent cluster with no single point of failure. Workload automatically redistributed upon agent failure. In-flight jobs re-queued without data loss within RTO not exceeding 60 seconds.
Hardware Redundancy	Underlying hardware shall be deployed in a clustered, redundant configuration with no single point of failure at the server, storage, or network layer.
Automatic Load Balancing	Load-aware scheduler distributing jobs by domain affinity, resource availability, and priority class.
Failure Detection and Alerting	Detection of agent failures, job anomalies, SLA breaches, and source connectivity loss within sub-minute intervals. Automated alerts dispatched via email, SMTP gateway, and webhook. Escalation policies configurable by severity tier, domain, and business hours.

Data Lineage, Observability and Quality Profiling

Requirement	Specification
End-to-End Data Lineage	Automatic field-level and table-level lineage capture for every pipeline from source through all transformation steps to target. Lineage shall be query able, visualizable as a directed graph, and exportable for audit purposes. Impact analysis shall identify all downstream datasets affected by a source change or failure in real time.
Job Monitoring Dashboard	Live dashboard displaying job status, execution duration trends, row-count throughput per pipeline, and SLA compliance metrics. Refresh interval shall not exceed 30 seconds. Accessible via standard web browser with no client-side software installation.

Data Quality Profiling	Automated profiling aligned to all six DAMA-DMBOK quality dimensions: Accuracy, Completeness, Consistency, Integrity, Timeliness, and Uniqueness. Profiling shall execute at ingestion time and publish quality scorecards per pipeline and source system. Records failing quality thresholds shall be quarantined and routed for remediation without halting the primary pipeline.
Audit Logs	Immutable audit logs capturing all job executions, configuration changes, user access, data exports, and alert dispositions. Application performance metrics (CPU, memory, queue depth, connector latency) shall be exposed via a published metrics API compatible with standard enterprise monitoring systems.

Access Control, Developer Experience and AI-Assisted Intelligence

Requirement	**Specification**
Role-Based Access Control (RBAC) and Attributes-Based Access Control (ABAC).	Domain-scoped both RBAC & ABAC allowing each business domain team to view and manage only their assigned pipelines. Platform administrators shall retain cross-domain visibility. Integration with FBR/PRAL enterprise directory via LDAP or SAML 2.0 is mandatory
Low-Code and Code-First Interface	Visual drag-and-drop pipeline design canvas for standard source-to-target mappings. Full code-first interface supporting SQL, Python, or platform-native scripting for complex transformations. Both modes shall produce identical runtime artefacts with no functional disparity.
AI-Assisted Mapping and Anomaly Detection	Intelligent column mapping recommendations based on schema similarity, naming conventions, and data type compatibility. Statistical anomaly detection monitoring row-count volumes, execution durations, and null-rate trends per pipeline, with alerts triggered upon material deviation from established baselines.

5.4 Data Modeling and Platform Design

The Platform shall implement a structured, on-premises Lakehouse architecture organized as an Extended Medallion Data Model. This model shall comprise four named layers Bronze, Silver, Gold, and Cache. The architecture shall unify the scalability and schema flexibility of object storage with the query performance and ACID compliance of a Massively Parallel Processing (MPP) Enterprise Data Warehouse, connected through a federated query engine accessible across all layers without data movement. The model is designed to serve

national-scale tax data workloads spanning ingestion from 30+ heterogeneous source systems through to low-latency analytical serving.

The Reporting Workload is 80% Transactional reports that require very quick response time and 20% Analytical Reports this would be a key metric in solution finalization.

The Platform shall provide native multi-modelled processing capability, handling both Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) data models within a single unified engine, coupled with Hybrid Transactional and Analytical Processing (HTAP), enabling transactional and analytical workloads to be served concurrently from the same platform without ETL-driven duplication between systems.

Layer	Storage Engine	Primary Purpose	Storage Tier
Cache Layer	In-Memory Distributed Cluster	Low-latency data objects, API serving, feature serving	Enterprise In-Memory Store
Gold Layer	MPP Enterprise Data Warehouse	Dimensional Model Star/Snowflake schemas for BI & reporting	MPP Storage
Silver Layer	MPP Enterprise Data Warehouse	Structured, auditable, historized, multi-source truth	MPP Storage
Bronze Layer	On-Premises S3-Compatible Object Store	Raw source replicas schema-on-read, full fidelity landing	Object Store / MPP Storage

Federated Query and Open Data Format

Requirement	Specification
Federated Query Engine	MPP Data Warehouse shall execute SQL queries spanning both the object store (Bronze) and MPP storage (Silver and Gold) within a single query plan, with no prior data movement. Query pushdown optimization shall minimize object store scans.
Open Table Format	Native read/write support for Apache Iceberg as the primary open table standard. Parquet as the default columnar format. ORC and Avro supported for ingestion compatibility. Full time-travel, snapshot management, and schema evolution shall be supported without table rewrites.
Open Metadata and Interoperability	All table definitions, schema registries, and catalogue entries shall be stored in open, non-proprietary formats. Metadata shall be exposed via open catalogue API ensuring interoperability with a compute cluster and future AI/ML tooling.

ACID Transactions	Full ACID transaction semantics guaranteed across all layers including concurrent write isolation, read consistency, and point-in-time recovery at both the object store and MPP warehouse layers.

Bronze Layer: Raw Landing and Operational Data Store

The Bronze Layer is the primary landing zone for all inbound source data, persisted in full fidelity on on-premises S3-compatible object storage. No business transformation shall be applied at this layer.

Requirement	Specification
Source-Faithful Raw Storage	All data from relational databases, APIs, flat files, and CDC streams shall be persisted in source schema. Ingestion timestamp, source system identifier, and batch or event sequence metadata shall be appended as system columns.
Business Abstraction Sub-Layer	An optional sub-layer shall provide human-readable aliases, domain-aligned naming conventions, and logical groupings over technically complex source schemas without altering underlying raw data.
Schema-on-Read and Partition Management	Schema-on-read semantics shall be supported for raw data querying without prior schema registration. Automatic partitioning by ingestion date, source system, and domain shall be enforced.
Retention and Immutability	Bronze Layer data shall be retained for a configurable period aligned with statutory and regulatory requirements.

Silver Layer: Conformed and Historized Data (MPP)

The Silver Layer is the enterprise-wide source of truth hosted on the MPP Data Warehouse. A structured, auditable modelling standard shall be applied at this layer, to be finalized during the design phase.

Requirement	Specification
Silver Layer Modelling	Mandatory implementation of a structured, auditable modelling approach comprising: core business entities (e.g., NTN, CNIC, STRN, invoice identifiers), cross-domain entity relationships, and descriptive attributes with historized change records. The modelling approach shall absorb source schema changes without structural model rewrites, with the specific technique to be finalized during the design phase.

Multi-Source Identity Resolution	The Silver Layer modelling approach shall support identity-resolution linking across duplicate or related records, point-in-time (PIT) structures for temporal joins, and bridge structures for cross-domain relationship traversal, enabling a 360-degree Taxpayer View across IRIS, eFBR, RegSys, and TAMS
Alternative Modelling Support	An alternative normalized modelling approach shall be supported for reference data, lookup tables, and regulatory code sets, where appropriate. These structures shall coexist with the primary Silver Layer modelling structures within the same MPP layer without performance degradation
ACID Compliance and Time-Travel	Full ACID compliance across all Silver Layer tables. Point-in-time query capability shall enable reconstruction of any entity's exact state at any historical timestamp, as required for legal defensibility of tax assessments.

Gold Layer: Dimensional Model and Analytical Serving (MPP)

The Gold Layer provides the consumption-ready analytical model hosted on the MPP Data Warehouse. All data shall be derived exclusively from the Silver Layer.

Requirement	Specification
Dimensional Modelling	Star Schema and Snowflake Schema models aligned to tax domain data products: Taxpayer 360 Profile, Compliance Behavior, Revenue Performance, Audit and Enforcement, and Refund Management. Fact tables shall carry pre-aggregated measures to sustain Seconds query response times.
MPP Query Performance	Interactive query performance at full production data volumes. Workload management shall prioritize executive and officer-facing dashboards over bulk extract workloads. Result caching and partition pruning shall be employed to meet performance SLAs without manual query tuning.

Cache Layer: In-Memory Distributed Cluster

The Cache Layer is a first-class architectural component serving as the low-latency tier for data objects requiring sub-second access patterns unsuitable for MPP or object storage retrieval.

Requirement	Specification
High-Availability Cluster	Minimum three nodes in a primary-replica configuration with automatic failover. No single node shall constitute a single point of failure. Data persistence shall be enabled to support cache recovery without full reload from the MPP warehouse upon node restart.

Low-Latency Data Product Serving	Pre-computed data objects including Taxpayer 360 profiles, risk scores, lookup tables, and reference datasets shall be served at sub-millisecond read latency under concurrent API load. Shall function as the primary serving store for the platform's REST API tier.
Feature Store Interface	Key-value and hash map interface for serving pre-computed ML feature vectors to AI/ML model inference workloads. Feature TTL policies shall be configurable per feature group. Cache invalidation shall be triggered programmatically upon Silver or Gold Layer refresh events.
Access Control and Namespace Isolation	Namespace-level isolation between domain data objects. RBAC enforced at namespace and key-pattern level, aligned with Silver and Gold Layer access policies. All cache access events shall be logged to the platform's central audit log.

5.5 Data Enrichment

To transform, enhance, and augment data in CDH to bring in higher-value information assets that directly support FBR's analytical, operational, and decision-making needs. The proposed platform shall provide the capabilities, frameworks, and processes required to systematically enrich data across all business domains.

Enrichment Capabilities

The platform shall support the following categories of enrichment:

- **Derived Attributes & Computed Fields**: Generation of calculated fields, ratios, trends, and aggregations not present in source systems (e.g., year-on-year filing growth, payment-to-declaration ratios).
- **Classification & Segmentation**: Automated tagging and categorization of taxpayers, transactions, and entities based on defined business rules and ML-driven models (e.g., risk tiers, taxpayer segments, sector classification).
- **Scoring & Indexing**: Generation of composite scores such as risk scores, compliance scores, and tax scores, combining multiple underlying indicators into business-consumable metrics.
- **Reference Data Matching & Standardization**: Matching and standardizing entity records (taxpayers, businesses, addresses) against reference and master data to resolve duplicates, correct inconsistencies, and establish golden records.
- **Geospatial Enrichment**: Appending geographic and jurisdictional context (tax office, region, geocoded address) to support geospatial analytics.

- **Lifecycle Enrichment**: Appending time-based context such as fiscal periods, filing cycles, and historical comparison windows to support trend and lifecycle analysis.
- **Cross-Domain Linkage**: Establishing relationships and linkages between datasets across business domains (e.g., linking sales tax filings to income tax declarations for the same taxpayer) to enable cross-subject area analysis and Taxpayer 360 profiles.

5.6 Object Storage On-Premise

Deployment Architecture and Infrastructure Compatibility

Requirement	Specification
Deployment Model	Fully on-premises deployment using NVMe SSD as the primary storage media to support I/O-intensive, high-throughput Big Data workloads per manufacturer-published performance specifications.
Infrastructure Integration	Solution shall integrate natively with the existing platform infrastructure without requiring proprietary interconnects or vendor-specific hardware dependencies.
Platform Compatibility	Solution shall be vendor-certified for full integration with compute engines providing a unified storage layer across the platform.
Hardware Failure Tolerance	The storage layer shall implement erasure coding as the default data protection scheme, with a minimum EC 8+3 configuration or equivalent, ensuring continued read/write operations and zero data loss upon simultaneous failure of up to 33% of independent storage nodes or drives. The solution shall additionally support configurable replication factors for latency-sensitive datasets where erasure coding overhead is operationally unacceptable. Rebuild operations following drive or node failure shall proceed autonomously without manual operator intervention and shall not degrade foreground query performance below 80% of baseline throughput.
Integration with AI/ML Compute Cluster	The object storage needs to connect with the core AI/ML Compute cluster for data processing and all types of AI/ML workloads.

Native S3 Compatibility and Open Format Support

Requirement	Specification
Native S3 Protocol	Built-in S3-protocol compatibility with no reliance on third-party translation layers, middleware, or gateway software. Shall support object PUT, GET, DELETE, multipart upload, bucket lifecycle policies, and pre-signed URL generation.
Open Table Format Support	Native support for open table format like Delta Lake, Apache Iceberg, and Apache Hudi and others.

Open Columnar File Formats	Native support for open file formats like Apache Parquet and Apache ORC and others enabling direct query access from compute and warehousing layers without data movement or format conversion overhead.

Scalability and Vendor-Neutral Extensibility

Requirement	Specification
Horizontal Scalability	Seamless horizontal scaling of raw usable storage capacity to a minimum of two Petabytes. Storage node expansion shall be performed online with minimal operational disruption and zero unplanned downtime.

Monitoring, Observability and Operational Management

Requirement	Specification
Built-In Observability	Native monitoring suite providing visibility into storage utilization, I/O throughput, latency metrics, node health, and fault events without dependency on external monitoring tools.
Alerting and Integration	Configurable alerting thresholds with support for enterprise-grade SNMP, Syslog, or REST-based notification system integration.

5.7 MPP Enterprise Data Warehouse

The MPP Enterprise Data Warehouse shall serve as the central analytical engine of the Central Data Hub, deployed entirely on-premises with no dependency on cloud-hosted services or external network connectivity. All capabilities shall conform to latest DAMA-DMBOK data management principles.

Parallelism and Query Optimization

Requirement	Specification
Parallel Architecture	Processing shall be distributed across all nodes in the cluster, whether via a shared-nothing, shared-storage, or shared-memory architecture, such that all data operations including filtering, aggregation, joining, and sorting scale horizontally across nodes with no single processing bottleneck. Where the proposed engine natively supports Multi-Model / HTAP processing (see Multi-Model & HTAP Processing below), the underlying storage and memory architecture may be shared or unified across nodes as required to support real-time transactional-analytical consistency, provided horizontal scalability and fault tolerance are preserved.
Multi-Level Parallelism	Shall support inter-query parallelism (simultaneous independent queries), intra-query parallelism (parallel execution of individual query steps), and intra-operator parallelism (parallel processing within a single operator).

Cost-Based Query Optimizer	Execution plans shall be evaluated using real-time and historical statistics on data distribution, cardinality, column selectivity, and node resource availability. Partition-wise joins and partition pruning shall be supported.
Adaptive Query Re-Optimization	Execution engine shall revise query plans mid-flight when actual runtime statistics deviate materially from optimizer estimates, eliminating performance degradation caused by data skew or parameter-sensitive queries.
Hybrid Execution	Native support for combined row-oriented and column-oriented access patterns within a single query, adapting dynamically to workload characteristics.

Multi-Model & HTAP Processing

Multi-Model Engine	The MPP Enterprise Data Warehouse shall natively support both Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) data models within a single unified engine, without requiring separate specialized databases for each processing model.
Hybrid Transactional and Analytical Processing (HTAP)	Transactional and analytical workloads shall be served concurrently against the same underlying data with no batch-driven replication delay between an operational copy and an analytical copy, consistent with the mixed 80% transactional / 20% analytical reporting profile defined in Section 5.4.
No Federation Dependency	Multi-Model and HTAP capability shall be delivered natively by the platform engine itself, rather than achieved by federating or stitching together separate specialized systems (e.g., a distinct OLTP database plus a distinct OLAP warehouse) through external integration or ETL layers.

Advanced Workload Management

Requirement	Specification
Workload Classification	Classification shall operate on user identity, group membership, role, database object, query complexity, estimated resource consumption, submission time, and application source identifier. Rules shall be configurable without system restart.
Resource Pools and Prioritization	Configurable resource pools with limits on memory, CPU, and concurrency. High-priority tax assessment and audit workloads shall be guaranteed resource availability under peak load.
Automatic Query Demotion	Long-running or resource-intensive queries shall be dynamically repositioned to lower-priority queues upon exceeding pre-defined resource thresholds.
Operational Visibility	Real-time visibility into queue depth, query wait time, resource pool saturation, and active session counts via a dedicated administrative interface and programmatic monitoring APIs.

Federated Query Capability

Requirement	Specification
Native Federated Execution	Single SQL statements shall transparently retrieve and join data from heterogeneous external sources alongside locally stored data. Supported sources shall include object storage (Iceberg, Parquet), relational databases (JDBC/ODBC), columnar file stores, and structured flat-file repositories.
Predicate Pushdown	Predicate pushdown to external sources shall be performed wherever technically supported, minimizing data movement and ensuring only required result sets are transported across the network.
External Metadata Catalogue	External data sources shall be registered with schema definitions and presented as logical schema objects indistinguishable from native internal tables. Federated access shall be subject to identical RBAC, audit logging, and session controls as internally stored data.

Native Object Storage Connectivity

Requirement	Specification
Direct Object Storage Integration	Native high-throughput read/write connectivity to the on-premises S3-compatible object storage tier with no intermediary file staging, manual export/import cycles, or third-party data movement utilities.
Support for Open table	Full support for open table format as the primary open table format including ACID transactions, time-travel queries, schema evolution, and partition evolution without data reorganization.
DML on Object Storage	Shall support CREATE TABLE, INSERT, UPDATE, DELETE, and MERGE operations directly targeting object storage-backed Iceberg tables. Concurrent reads and writes shall be managed through optimistic concurrency control with transparent conflict resolution.

Data Encryption and Anonymization

Requirement	Specification
Encryption at Rest and in Transit	Industry-standard encryption applied uniformly to all data files, temporary storage, transaction logs, audit logs, and backup artefacts. All data in transit between cluster nodes, object storage, and client applications shall be encrypted.
Dynamic Data Masking	Column-level dynamic masking enforced at query runtime without modifying stored data. Masking policies shall apply selectively based on the querying user's role, enabling privileged users to access unmasked values while non-privileged users receive masked outputs from the same table.
Static Masking and Tokenization	Built-in support for static data masking and format-preserving tokenization for PII protection aligned with DAMA-DMBOK Data Security principles.
Data Classification Tagging	Column-level classification tagging to facilitate automated PII discovery, masking policy assignment, and downstream lineage tracking.

Role-Based Access Control

Requirement	Specification
RBAC Model	Comprehensive RBAC governing all interactions with databases, schemas, tables, columns, views, stored procedures, and administrative operations. Access rights shall be granted exclusively through role assignments, not direct user-to-object grants.
Role Hierarchies and Least Privilege	Role inheritance from parent roles shall be supported. Least privilege shall be enforced such that no user or process possesses access rights exceeding those required for their assigned function.
Column and Row-Level Security	Column-level security enforced natively by the query engine. Row-level security enforced through dynamic session context policies enabling multi-tenancy on shared physical tables.
Enterprise Directory Integration	Integration with enterprise directory services via LDAP and SAML 2.0 for centralized identity management and single sign-on.

Audit Logging

Requirement	Specification
Mandatory Audit Events	Logs shall capture all user authentication, privilege escalation, DDL and DML operations on sensitive objects, administrative configuration changes, bulk export operations, federated query executions, and access to encrypted or masked columns.
Audit Record Content	Each record shall include event timestamp, authenticated user identity, source IP address, client application identifier, full SQL statement text, objects accessed, row and column scope, and outcome status.
SIEM Integration	Streaming audit log forwarding to external SIEM systems via industry-standard protocols for real-time threat detection. Audit logging shall operate with negligible performance impact on production query workloads.

Query Performance, Storage and SQL Compliance

5.10	Data Governance
Query Performance	Consistently high performance across ad-hoc exploratory queries, complex multi-join analytical queries, scheduled batch reports, and real-time dashboard queries requiring sub-second response latency. Query profiling and explain plan tooling shall be available for routine tuning without vendor assistance.
Data Compression	Adaptive, lossless compression applied automatically at the storage layer per column based on data type, cardinality, and value distribution. Compression shall be transparent to all SQL operations with no manual decompression steps required.
SQL Standard Compliance	Full conformance to ANSI SQL:2016 including all standard DML and DDL constructs, set operators, comprehensive window functions (RANK, DENSE_RANK, ROW_NUMBER, LAG, LEAD, FIRST_VALUE, LAST_VALUE, NTILE), and recursive CTEs for hierarchical data traversal.

Storage Organization and Distribution	Support for both row-oriented and columnar storage formats. Distribution strategies shall include hash distribution, range distribution, and replicated distribution for lookup tables. Table partitioning on date, numeric range, or list values with automatic partition pruning enforced by the query optimizer.

5.8 Data Science & AI/ML

The Central Data Hub shall provide a fully enterprise AI platform supporting the complete machine learning lifecycle including data preparation, feature engineering, model training, deployment, monitoring, explainability, governance, auditability, and retirement.

All platform components shall operate exclusively within the on-prem network perimeter, support air-gapped deployment, and have no dependency on external networks or services for any aspect of their operation. The platform shall not require any GPU hardware for its operation or for the development of AI/ML models. The platform will provide AI/ML workload compute cluster using the underlying object storage, MPP storage and other data storage layers when applicable.

The following functional requirements shall apply to the development of AI/ML models.

Section	**Category**	**Specification**
Core Platform Requirements	Foundation	The platform shall provide a reliable, scalable foundation capable of handling AI/ML workloads, with automatic recovery from failures and consistent performance under high load. All software shall be fully licensed for on-premises, air-gapped operation with no dependency on external networks.
Feature Store and Data Quality	Data Pipeline	The platform shall provide a centralized Feature Store that calculates, stores, and serves model-ready data features to all AI models with consistent values and sub-millisecond retrieval for real-time scoring. The platform shall enforce automated data quality checks on all data feeding AI models, covering completeness, validity, consistency, and freshness. Records failing defined thresholds shall be quarantined and an alert raised before reaching any model.
MLOps and Model Governance	AI Development	The platform should provide an MLOps framework covering experiment tracking, model registry, version control, and automated CI/CD pipelines for training, validation, and deployment of AI models, with full lineage and audit trail maintained for every model version.

DataOps and Pipeline Governance	Data Pipeline	The platform should maintain version control for all datasets used in AI model training, with each model traceable to the exact dataset version used. Pipelines feeding AI models shall be governed by defined data contracts specifying schema, refresh frequency, and quality thresholds, with automated quality checks quarantining records that fail these thresholds
AI and Advanced Analytics Platform	AI Development	The AI/ML Development Platform shall provide a scalable environment covering the complete model development lifecycle including data preparation and feature engineering through training, validation, deployment, and monitoring that supports both code-based and low-code/no-code development with AutoML capabilities.
Data Sharing Hub	Consumer Layer	The platform should provide a Data Sharing Hub enabling FBR/PRAL departments and authorized external agencies to access governed data products and AI model outputs through a single secure interface, including API gateways. All access shall be logged, role-based, and auditable.
Predictive Analytics and AI-Driven Insights	AI Intelligence	The platform shall provide predictive analytics capabilities enabling FBR/PRAL to anticipate taxpayer behavior, identify risk, and prioritize compliance activities based on historical data patterns. This includes risk scoring, anomaly detection, forecasting, segmentation, and case prioritization models built and managed through the AI/ML Development Platform.
Security and Access Control	Security	Attribute-based access control (ABAC) enforced at the object store policy layer. No path to training data shall bypass the access control boundary. Identity federation via SSO/IAM, network isolation, and secrets management shall be mandatory. Data at rest should be encrypted using platform-managed or customer-managed keys. All inter-node traffic between executors and the object store shall be encrypted in transit via SSL/TLS. RBAC, audit logging, lineage tracking, data catalogue integration, and policy enforcement for all sensitive data assets across the AI/ML layer.
Observability and Monitoring	Operational monitoring	The platform should provide cross-cutting operational monitoring across all functional components, covering operational monitoring, ML model monitoring, data health, data lineage, and application insights.
Platform Scale-Up	Foundation	The platform should be designed to support future infrastructure growth, covering capacity planning, compute

		expansion, storage growth, dynamic scaling, and workload separation, based on the functional scope defined above.
AI Platform Subscription and Support	Delivery & Support	The vendor should provide ongoing support and subscription for the AI platform after go-live, covering the subscription term, offline updates, annual reviews, support for new AI use cases, and ongoing support for deployed models.

5.9 Stream Data Processing

The platform shall address both Near Real-Time (NRT) and Real-Time (RT) stream processing requirements. NRT processing shall utilize a micro-batching approach to balance data freshness with system throughput, delivering results within seconds to minutes. RT processing shall handle record-by-record event processing with sub-second latency, imposing distinct requirements on event sourcing, state management, and fault tolerance at every architectural layer.

Core Processing Requirements

Requirement	NRT Specification	RT Specification
Latency	Data processed within seconds to minutes using configurable micro-batch windows.	Data processed within milliseconds to seconds using continuous, record-by-record streaming.
Processing Engine	Micro-batch processing engine with configurable time and size-based windowing.	Event-by-event stream processing architecture with stateful and stateless transformation support.
Data Ingestion	Scheduled streaming or buffered event ingestion from applications, databases, APIs, and logs.	High-throughput continuous event streaming via low-latency message brokers with no buffering delay.
Scalability	Elastic horizontal scaling for streaming and analytics workloads handling variable data volumes and peak loads.	Auto-scaling distributed architecture optimized for ultra-low latency under concurrent high-volume event loads.
Fault Tolerance	Checkpointing, automated retries, and replay mechanisms ensuring pipeline continuity on failure.	High-availability architecture with state recovery and exactly-once processing guarantees.
AI/ML Integration	Continuous feature refresh, model monitoring, and automated retraining triggers based on drift signals.	Real-time feature serving, online inference APIs, and instant prediction responses at sub-second latency.

Data Governance, Security and Observability

Requirement	Specification
Access Control and Security	RBAC and real-time policy enforcement across all stream processing components. Secure streaming with encrypted data in transit. Low-latency access controls enforced without introducing processing overhead.
Audit Logging and Lineage	Audit logging and lineage tracking maintained across all NRT and RT pipeline stages. All processed events shall be traceable from source ingestion through transformation to downstream consumption.
Data Quality and Validation	Quality and validation checks enforced prior to processing to detect missing, duplicate, delayed, or malformed data. Records failing quality thresholds shall be quarantined and flagged without halting the pipeline.
Observability and Alerting	Real-time telemetry covering throughput, latency, and processing lag. SLA monitoring with automated alerting when data freshness, accuracy, or system thresholds are breached.

Downstream Integration

Requirement	Specification
Consumption Targets	Processed data shall be delivered to dashboards, operational applications, ML models, REST APIs, decision engines, and data warehouses. Both NRT and RT outputs shall be available as consumable data products through standardized interfaces.
Data Persistence	All processed outputs shall be persisted to the designated storage layer in open formats with ACID compliance, supporting downstream replay, reprocessing, and historical analysis.

Implementation Scope

The vendor shall be responsible for provisioning the infrastructure required to support NRT and RT stream processing, covering the following activities:

- Provisioning and deployment of stream and micro-batch processing clusters including installation, configuration, and network setup
- Sizing and capacity planning of compute infrastructure to handle variable data volumes, peak loads, and concurrent workloads, per Section 6

- Installation and base configuration of infrastructure-level fault tolerance mechanisms (clustering, redundancy, storage replication) up to the point of readiness for the Data Platform Implementation Vendor to configure pipeline-level checkpointing, retry logic, and replay
- Network and connectivity readiness for downstream integration points (dashboards, ML models, APIs, MPP Data Warehouse), excluding actual pipeline/application-level integration
- Deployment and base configuration of observability infrastructure (agents, log shippers, monitoring endpoints), excluding application-level SLA/latency instrumentation, which is the Data Platform Implementation Vendor's responsibility

5.10 Data Governance

Data Governance, as defined by the latest Data Management Association's *DMBOK* framework, is the exercise of authority, control, and shared decision-making over the management of data assets encompassing policies, processes, standards, roles, and metrics that ensure data is managed as a strategic organizational resource across its full lifecycle. Grounded in DMBOK's knowledge areas from Data Quality, Metadata Management, Data Security, Privacy and Master Data Management. Governance establishes the stewardship structures, accountability frameworks, and operational controls that enable an organization to trust its data, demonstrate regulatory compliance, and derive consistent, measurable value from its information assets.

The Platform shall implement a comprehensive, enterprise-grade Data Governance capability that establishes and enforces policies, roles, responsibilities, processes, and controls across the entire data lifecycle. All requirements in this section are mandatory unless stated otherwise.

Governance Framework & Policy Enforcement

Req. ID	Functional Requirement	Priority
DG-001	The Platform shall provide a centralized data governance module capable of defining, publishing, and enforcing data policies across all integrated data domains without requiring manual intervention per dataset.	Mandatory
DG-002	The Platform shall support a configurable data governance operating model encompassing a minimum of three (3) tiers: Executive Steward, Domain Data Owner, and Data Custodian with role-specific dashboards and task queues.	Mandatory

DG-003	The Platform shall enforce data governance policies at ingestion, transformation, and query layers. Policy violations shall be intercepted, logged, and escalated without corrupting the operational data pipeline.	Mandatory
DG-004	The Platform shall provide a policy versioning engine that maintains a full audit trail of policy creation, modification, approval, and retirement events, with a minimum retention of five (5) years.	Mandatory
DG-005	The Platform shall support integration with FB R's organizational directory to assign governance roles based on organizational unit and position classification. (AD/LDAP info will be shared with vendor).	Mandatory

Data Stewardship & Accountability

Req. ID	Functional Requirement	Priority
DG-006	The Platform shall provide a stewardship workflow engine enabling data stewards to raise, route, approve, and resolve data issues, with configurable SLA timers and escalation paths per issue severity class.	Mandatory
DG-007	The Platform shall maintain a stewardship accountability register, recording all stewardship assignments, actions taken, resolution timestamps, and residual risk signoffs.	Mandatory
DG-008	The Platform shall generate stewardship performance metrics on a configurable schedule (daily, weekly, monthly), exportable in structured format (CSV, JSON, PDF) for executive reporting.	Mandatory

Access Governance & Role-Based Access Control (RBAC)

Req. ID	Functional Requirement	Priority
DG-009	The Platform shall enforce granular Role-Based Access Control (RBAC) at a minimum of five (5) hierarchical levels: Platform, Domain, Dataset, Table/Object, and Column/Field.	Mandatory
DG-010	The Platform shall enforce data classification labels (e.g., Public, Internal, Confidential, Restricted, Top Secret) that propagate automatically from metadata through all downstream data products and query surfaces.	Mandatory

DG-011	The Platform shall support attribute-based access control (ABAC) extensions, permitting policy conditions based on user attributes (clearance level, organizational unit, geographic region) in addition to role assignments.	Mandatory
DG-012	The Platform shall generate a complete, tamper-evident access governance audit log capturing: user identity, role at time of access, data object accessed, operation performed, timestamp (UTC), and outcome. Logs shall be immutable and retained for a minimum of five (5) years.	Mandatory
DG-013	The Platform shall enforce PII access restrictions by default. Access to PII fields shall require explicit data owner approval, with approvals recorded in the stewardship accountability register (see DG-007).	Mandatory

Regulatory & Audit Compliance

Req. ID	Functional Requirement	Priority
DG-014	The Platform shall provide a dedicated regulatory compliance dashboard mapping all active data governance controls to the applicable FBR obligations. (Compliance Rules will be provided to vendor).	Mandatory
DG-015	The Platform shall produce on-demand regulatory evidence packages (audit packs) in a structured, machine-readable format (PDF/JSON) summarizing: data lineage, policy enforcement records, access logs, and open issue logs for any defined date range.	Mandatory
DG-016	The Platform shall support air-gapped deployment mode, in which all governance operations, policy enforcement, and audit logging function without any dependency on external network connectivity or cloud-hosted services.	Mandatory

Governance Metrics & Reporting

Req. ID	Functional Requirement	Priority
DG-017	The Platform shall expose a configurable data governance scorecard tracking, at minimum: policy compliance rate (%), open stewardship issues (count/age), RBAC exceptions (count), PII access events (count), and audit pack generation frequency.	Mandatory

DG-018	The Platform shall support integration with external BI reporting tools via a documented REST API or JDBC/ODBC endpoint for governance metric consumption by the national fiscal authority's enterprise reporting layer.	High

Metadata Repository & Active Catalogue

Req. ID	Functional Requirement	Priority
MM-001	The Platform shall provide a centralized, actively maintained Metadata Repository (Business Glossary + Technical Catalogue) capable of storing, versioning, and serving meta-data for all data assets ingested into or produced by the Platform.	Mandatory
MM-002	The Platform shall auto-discover and register metadata for new datasets upon ingestion, capturing source system identifier, ingestion timestamp, schema (column names, data types, nullability), record volume, and partition strategy.	Mandatory
MM-003	The Platform shall maintain three (3) distinct metadata layers: (i) Business Metadata business definitions, data owner, classification; (ii) Technical Metadata schemas, storage paths, file formats, statistics; (iii) Operational Metadata job runs, row counts, SLA status, quality scores.	Mandatory
MM-004	The Platform shall version all metadata changes, maintaining a complete history of schema evolution events with actor identity, change timestamp, and change description.	Mandatory
MM-005	The Platform shall expose the Metadata Repository via a documented, versioned REST API supporting CRUD operations, enabling programmatic integration with external governance tools, reporting systems, and auditors.	Mandatory

Business Glossary & Term Management

Req. ID	Functional Requirement	Priority
MM-006	The Platform shall include a Business Glossary module enabling data stewards to create, review, approve, publish, and retire business terms. Each term record shall include term name, definition, synonyms, related terms, owning domain, status, and approval history.	Mandatory
MM-007	The Platform shall link Business Glossary terms to corresponding technical metadata objects (tables, columns,	Mandatory

	reports, pipelines) via a term-to-asset association model, providing a bi-directional navigation path.	
MM-008	The Platform shall support FBR fiscal-domain taxonomies applicable to tax administration, including classification hierarchies for: taxpayer entity types, tax product codes, filing period definitions, and transaction categorization schemas. (Applicable FBR taxonomy will be shared with vendor).	Mandatory

End-to-End Data Lineage

Req. ID	Functional Requirement	Priority
MM-009	The Platform shall automatically capture and maintain end-to-end data lineage at both column-level and dataset-level granularity across the complete data pipeline, from source ingestion through ELT transformation, object storage, compute clusters, MPP warehouse, and reporting layers.	Mandatory
MM-010	The Platform shall render data lineage as an interactive directed acyclic graph (DAG) accessible via the governance interface, with the ability to trace any field in any downstream report back to its originating source system record.	Mandatory
MM-011	The Platform shall support impact analysis queries given any upstream change (schema modification, source system migration, pipeline restructure); the system shall immediately identify and surface all downstream datasets, reports, and AI model feature sets affected.	Mandatory
MM-012	Data lineage records shall be immutable once committed. Corrections shall be implemented via superseding lineage records, with full retention of the superseded record and actor attribution.	Mandatory

Metadata Quality & Completeness

Req. ID	Functional Requirement	Priority
MM-013	The Platform shall continuously monitor and score Metadata Completeness across all registered data assets. Assets with Metadata Completeness below a configurable threshold (default: 80%) shall trigger a stewardship task in the governance workflow engine (see DG-006).	Mandatory

MM-014	The Platform shall validate metadata entries against configurable rules (e.g., mandatory field population, valid value lists, term linkage requirements) at point of registration and on scheduled re-validation cycles.	Mandatory
MM-015	The Platform shall produce a Metadata Health Report on a configurable schedule, quantifying: total registered assets, completeness score distribution, orphaned assets (no owner), stale metadata (not updated within configurable threshold), and unlinked technical objects.	High

Search, Discovery & Integration

Req. ID	Functional Requirement	Priority
MM-016	The Platform shall provide a full-text search interface across the Metadata Repository, supporting search by business term, technical object name, data owner, classification label, data domain, and free-text keyword with ranked relevance results.	Mandatory
MM-017	The Platform shall support metadata federation, enabling the Metadata Repository to harvest and synchronize metadata from registered external source systems on a configurable schedule without requiring manual import.	Mandatory
MM-018	The Platform shall maintain a Sensitive Data Register within the Metadata Repository, auto-tagging fields classified as PII, Tax-Sensitive, or Legally Restricted upon ingestion, using pattern-based classification rules. Human steward review shall be required before classification is finalized.	Mandatory

Data Quality Rules Engine

Req. ID	Functional Requirement	Priority
DQ-001	The Platform shall provide a centralized Data Quality Rules Engine enabling authorized data stewards to define, version, deploy, and retire quality rules without requiring software development or platform redeployment.	Mandatory
DQ-002	The Platform shall support a minimum of six (6) rule types: (i) Completeness, null/blank detection; (ii) Accuracy value conformance to authoritative reference; (iii) Consistency cross-field/cross-dataset agreement; (iv) Uniqueness duplicate detection; (v) Timeliness freshness relative to defined SLA; (vi) Reasonability statistical range/outlier validation.	Mandatory

DQ-003	Quality rules shall be applicable at three (3) enforcement points: (a) at ingestion (pre-load validation); (b) at transformation (mid-pipeline validation); (c) at publication (pre-consumption certification).	Mandatory
DQ-004	The Platform shall support quarantine mode: records failing critical quality rules shall be automatically routed to an isolated quarantine zone with full metadata retention, pending stewardship review and remediation, without halting the primary pipeline.	Mandatory

DAMA Quality Dimensions Detailed Requirements

Dimension	Req. ID	Platform Requirement	Measurement KPI
Accuracy	DQ-005	The Platform shall validate field values against authoritative reference datasets (e.g., national taxpayer register, financial institution codes, tax product master). Discrepancies shall be flagged, scored, and routed to the stewardship workflow.	*Accuracy Rate ≥ 99.5% for Mandatory fiscal fields*
Completeness	DQ-006	The Platform shall compute column-level and dataset-level completeness scores on every pipeline execution. Datasets falling below a configurable completeness threshold (default: 98%) shall be blocked from the publication layer.	*Completeness Score ≥ 98% (configurable)*
Consistency	DQ-007	The Platform shall enforce cross-domain consistency rules ensuring that equivalent data attributes (e.g., taxpayer ID, entity name, registered address) are reconciled to a single canonical value across all source systems ingested.	*Cross-domain conflict rate < 0.1%*
Integrity	DQ-008	The Platform shall enforce referential integrity constraints across all relational and dimensional models within the MPP warehouse, rejecting or quarantining records that violate defined foreign key relationships.	*Integrity violation rate = 0% for certified datasets*
Reasonability	DQ-009	The Platform shall apply configurable statistical profiling (min, max, mean, standard deviation, inter-quartile range) to numeric and date fields, flagging records whose values deviate beyond a configurable z-score threshold (default: ±3σ) as potential anomalies for stewardship review.	*Anomaly detection sensitivity: configurable; default ±3σ*

Timeliness	DQ-010	The Platform shall track data freshness for every registered dataset against a defined Maximum Tolerable Data Age (MTDA) SLA. Datasets exceeding their MTDA shall trigger an automated stewardship alert within five (5) minutes of breach detection.	*MTDA breach alert latency ≤ 5 minutes*
Uniqueness	DQ-011	The Platform shall execute duplicate detection algorithms (exact match and configurable fuzzy match) on defined entity keys (taxpayer identifiers, transaction references, filing records) across the full dataset scope. Detected duplicates shall be flagged, not silently deleted.	*Duplicate rate ≤ 0.01% on key entity identifiers*

Data Quality Scoring & Certification

Req. ID	Functional Requirement	Priority
DQ-012	*The Platform shall compute a composite Data Quality Score (DQS) for each dataset, calculated as the weighted average of all applicable DAMA dimension scores. Dimension weights shall be configurable by data domain and dataset type.*	Mandatory
DQ-013	The Platform shall enforce a Data Quality Gate at the publication layer: datasets with a DQS below a configurable minimum threshold (default: 95/100) shall be withheld from the analytical consumption layer until remediated and re-certified by an authorized data steward.	Mandatory
DQ-014	Certified datasets shall be assigned a cryptographically signed Data Quality Certificate, recording: DQS at certification, certifying steward identity, certification timestamp, rules evaluated, and expiry date. Certificates shall be stored in the Metadata Repository (see MM-003).	Mandatory

Data Quality Monitoring & Alerting

Req. ID	Functional Requirement	Priority
DQ-015	The Platform shall provide a real-time Data Quality Operations Dashboard displaying live DQS per active dataset, open quarantine record counts, active rule breach alerts, dimension-level trend charts (30/60/90-day rolling), and pipeline health indicators.	Mandatory

DQ-016	The Platform shall support configurable alerting channels (email, webhook, SNMP trap, SIEM integration) for quality breach events, with a minimum alert delivery latency of less than two (2) minutes from breach detection to notification.	Mandatory
DQ-017	The Platform shall maintain a longitudinal Data Quality History Store, retaining DQS time-series data for all datasets for a minimum of five (5) years to support regulatory audit and trend analysis.	Mandatory

MDM Architecture & Golden Record Management

Req. ID	Functional Requirement	Priority
MDM-001	The Platform shall implement a Registry-style and Consolidation-style MDM architecture, capable of operating in either mode or hybrid mode per entity domain, selectable during implementation configuration without requiring code changes.	Mandatory
MDM-002	The Platform shall generate and maintain a Golden Record for each master entity class (minimum: Taxpayer, Legal Entity, Financial Institution, Tax Product, Filing Period, Geographic Administrative Unit). The Golden Record shall represent the single, authoritative, deduplicated, and enriched version of each entity.	Mandatory
MDM-003	The Platform shall implement a configurable survivorship rule engine, determining which source system attribute value is promoted to the Golden Record based on rules including source system trust ranking, recency, completeness score, and frequency of occurrence.	Mandatory
MDM-004	All Golden Record updates shall generate an immutable change event, recording: prior value, new value, survivorship rule applied, source system of change, actor identity, and UTC timestamp. Change events shall be retained for a minimum of five (5) years.	Mandatory
MDM-005	The Platform shall expose Golden Record data via a versioned Master Data API (REST/GraphQL) enabling downstream systems, AI feature stores, and reporting layers to consume authoritative master data in real time without direct database coupling.	Mandatory

Reference Data Management (RDM)

Req. ID	Functional Requirement	Priority

RDM-001	The Platform shall provide a centralized Reference Data Registry maintaining all fiscal and administrative code sets, value domains, and lookup tables (e.g., tax type codes, country codes, currency codes, industry classification codes, filing status codes) as managed, versioned reference data objects.	Mandatory
RDM-002	Each reference data set shall support full lifecycle management: Draft → Review → Approved → Published → Deprecated → Retired - with mandatory steward approval at each transition gate and full version history retention.	Mandatory
RDM-003	The Platform shall enforce referential integrity constraints throughout the MPP warehouse and ELT transformation layer by referencing only Published reference data values. Transformation jobs referencing Deprecated or Retired values shall generate a blocking data quality alert.	Mandatory
RDM-004	The Platform shall support synchronization of reference data with designated authoritative external registries (e.g., FBR national standards body, international tax code authorities) on a configurable schedule. All externally synchronized values shall be human-reviewed before publication. (These specifications will be provided to vendor).	Mandatory
RDM-005	The Platform shall expose all Published reference data sets via the Master Data API (see MDM-005) as a unified, query able domain values service, enabling consistent code resolution across all Platform consumers.	Mandatory

5.11 BI & Analytics

The proposed Tool and Solution shall establish a centralized, scalable, enterprise-grade, Business Intelligence (BI) & Analytics ecosystem capable of supporting enterprise reporting and analytical requirements for operational users, field formations, analysts, management, and executive leadership. The platform shall support near real-time transactional reporting, pre-built/canned reports and dashboards, interactive reporting, self-service analytics, and high-performance aggregated and executive dashboards, enabling operational, strategic, and executive decision-making through a unified enterprise analytics environment.

The proposed solution shall support both **analytical and transaction-level** reporting, with the ability to navigate from summarized and aggregated information down to the lowest available grain of transactional data, subject to defined business requirements and access controls. The platform shall support multiple data consumption horizons, ranging from near real-time operational data to summarized, aggregated, and historical datasets, based on business use cases and performance requirements.

For ease of use and mass-scale adoption, the solution shall enable the development, publishing, organization, and maintenance of pre-built/canned reports and dashboards for standardized and recurring business requirements. Authorized end users shall be able to readily access and consume these reports without requiring technical or report-development expertise. Interactive capabilities shall include, where applicable, prompts, filters, sorting, drill-down, drill-through, navigation, and aggregation.

The proposed solution shall provide a centralized, governed, and scalable enterprise BI and semantic layer, making relevant subject areas, dimensions, measures, KPIs, hierarchies, business attributes, and analytical metrics available to authorized users within the BI tool and integration with the data governance tool. These analytical objects shall be centrally defined, reusable, and consistently available across canned reports, dashboards, ad-hoc reporting, and self-service analytics, thereby ensuring consistent business definitions and a Single Version of Truth (SVOT) across PRAL/FBR.

The platform shall support cross-functional and cross-subject-area analysis, enabling authorized users to analyze and correlate information across enterprise business domains through common/conformed dimensions, measures, and integrated enterprise data models.

The solution shall be designed for enterprise-scale performance, availability, and scalability, capable of supporting mass users and concurrent workloads while accommodating increasing data volumes and evolving analytical requirements. The architecture shall allow the addition and extension of data sources, subject areas, dimensions, measures, KPIs, reports, dashboards, analytical workloads, and users without requiring fundamental redesign of the platform.

The scope of this initiative includes the modernization, migration, enablement, and operationalization of the enterprise BI & Analytics ecosystem, including the migration and rationalization of existing reports, dashboards, analytical workloads, business rules, and associated semantic/data models to the proposed target platform.

Current Baseline

Currently, PRAL/FBR utilizes **Oracle Analytics Server (OAS)** to support enterprise-wide reporting requirements for mass users and field formations. The platform provides dashboards and detailed transactional reporting, including drill-down capabilities to the **lowest grain level of data**.

In parallel, PRAL/FBR has recently acquired **Tableau** as its strategic visualization and analytics platform for **senior management, headquarters, and executive-level reporting**

Metric	Current State (Existing Environment)
Total Dashboards & Reports	35+ dashboards with drill-down capabilities to 1000+ reports *(estimated figures)*
Current BI Platform	Oracle Analytics Server (2022)
Visualization Platform	Tableau 2026.2
End Users	1500+ Users

Requirements and Specification of Enterprise Analytics/BI Tool

Requirement	Specification
Self-Service Analytics	The proposed platform shall Support governed self-service analytics capabilities enabling business users and field officers to create, modify, and consume reports and dashboards with minimal IT dependency.
Transaction-Level Reporting	The platform shall support transaction-level reporting with drill-down and drill-through capabilities up to taxpayer-level detail while maintaining optimal performance and low response times.
Interactive Canned Reports & Analysis	The platform shall support interactive Reports with filtering, slicing, prompts, cascading prompts, tooltips, custom calculations, pivoting, and cross-domain analytical capabilities.
Executive & Operational Dashboards	The platform shall support development of executive, strategic, operational, reports and dashboards with real-time and near real-time KPI monitoring capabilities.
Real-Time & Aggregated Reporting	The platform shall support near real-time, summarized, aggregated, historical, and yearly reporting across operational, analytical, and executive reporting use cases.
Pixel-Perfect Reporting	The platform shall support pixel-perfect printable reports in IRIS-like reporting formats with support for PDF generation and standardized layouts.
Export & Distribution	The platform shall support export of reports and datasets to Excel, PDF, and standard formats along with scheduling, subscriptions, bursting, and automated email-based distribution.
Security & Access Control	The platform shall enforce role-based and jurisdiction-based access control (RBAC), row-level security, and secure access management aligned with organizational and geographic responsibilities.
Mobile Analytics	The platform shall support secure mobile access to dashboards, reports, alerts, and KPIs with responsive design and role-based access controls for executives and field officers.
Query Performance & Optimization	The platform shall support low-latency analytical and transactional query performance with response times in seconds using cubes, in-memory caching, and query acceleration mechanisms.

Audit Logs & Monitoring	The platform shall maintain tamper-evident audit logs for report access, exports, administrative activities, and user interactions with comprehensive monitoring and observability capabilities.
High Availability & Reliability	The platform shall support high availability (HA), failover, workload balancing, and job monitoring to ensure uninterrupted analytics and reporting services.
Data Security & Governance	The platform shall operate in an environment with strong governance controls, encryption, auditability, and data leakage prevention mechanisms.
AI-Assisted Analytics	The platform shall provide AI-assisted analytics including natural language query, automated insight generation, anomaly highlighting, and narrative summarization of dashboards and reports. Integration with the broader AI/ML platform shall be supported as an added advantage.
Cross-Subject Area Analysis	The platform shall enable cross-subject area analysis within a unified reporting interface to support integrated analytics and Taxpayer 360° profiles that consolidate a taxpayer's complete lifecycle across Income Tax, Sales Tax, Federal Excise Duty, Customs Tax, etc. This includes registration, filing, payment, audit, and enforcement history that could generate actionable insights. Base locking and cross-domain analysis shall be supported.
Embedded Analytics	The platform shall support embedding of dashboards, visualizations, and KPI widgets into other FBR/PRAL applications such as Faceless Assessment Center interface via secure embed tokens or APIs, without requiring users to switch platforms
BI Governance	The platform shall provide content usage analytics including report utilization, user adoption, execution frequency, and dormant content identification to support rationalization of the reporting portfolio.
Oracle Analytics Migration Framework	The licensed platform shall natively provide, or be delivered with, migration utilities and connectors capable of importing Oracle Analytics metadata repositories, analyses, dashboards, agents, security roles, and report schedules. Execution of the migration itself is covered under the separately contracted Data Platform & Application Implementation Services engagement.
OAS Migration Scope	The licensed platform shall support comprehensive migration tooling for the existing Oracle Analytics Server (OAS) environment, covering dashboards and reports, the underlying metadata repositories (including the RPD/semantic model layer, subject areas, and metric definitions), security configurations and role-based access mappings, and report schedules and subscriptions. Execution of the migration, including discovery, mapping, conversion, validation, and sign-off for each object type, is covered under the separately contracted Data Platform & Application Implementation Services engagement.

MIS and Ad-hoc Reporting	The proposed BI platform shall be licensed to cater to existing and future ad-hoc reporting requirements currently supported by the MIS and Ad-hoc Information Delivery (AID) teams, including standard MIS reports, recurring management reports, and one-off ad-hoc data requests. The platform shall enable governed self-service analytics capable of progressively reducing dependency on manual report development by MIS/AID teams, while retaining the ability for these teams to build, certify, and publish reports for broader consumption where self-service is not suitable. Migration of MIS and ad-hoc reporting workloads currently served through existing tools to the proposed platform, with no loss of functionality or reporting continuity, is covered under the separately contracted Data Platform & Application Implementation Services engagement.
Geospatial Analytics	The platform must support geospatial analytics including map-based visualization, geofencing, spatial joins, and GIS integration which should cover use cases such as POS terminal mapping, Track & Trace monitoring, and tax office jurisdictional analysis.
Large-Volume & Granular Data Analysis	The platform shall support large-volume, transaction-level and taxpayer-level data analysis through interactive tabular/grid views with filtering, sorting, drill-down and drill-through capabilities. Authorized users, including MIS and Ad-hoc reporting teams, shall be able to extract large-volume datasets in Excel, CSV, PDF, and other standard formats, The bidder shall specify applicable data, row, file-size, and export limitations.
Extensibility & Third-Party Libraries	The proposed BI platform shall provide an extensible framework supporting integration with third-party visualization libraries, JavaScript frameworks/libraries, Python libraries, APIs, SDKs, and custom analytical components, where required.

6. Sizing Estimation for Reference

IMPORTANT NOTICE: REFERENCE SIZING ESTIMATE

All hardware specifications, node counts, and storage capacities presented in this section constitute reference sizing estimates only. These estimates are derived from analysis of current on-premises data footprints and projected growth trajectories. The Vendor shall conduct a formal capacity planning exercise based on their certified technology stack and licensing architecture. The Vendor's validated sizing shall supersede all reference figures contained herein. The procuring authority shall not be bound by these reference estimates during contract negotiation.

This section presents reference hardware sizing estimates for all major platform components of the Central Data Hub. The estimates are derived from analysis of the existing on-premises data footprint, schema-level storage measurements, and projected data growth trajectories over a five-year horizon. The New-Technology components that are missing in the existing ecosystem are benchmarked based on Software tool Specification or Entry Grade Enterprise Production specification.

The section is organized into two primary groups. Group 1 covers the core analytical platform components, comprising the Massively Parallel Processing (MPP) Enterprise Warehouse, the ELT Processing Nodes, and the Object Store. Group 2 covers supporting infrastructure components, comprising Streaming and AI/ML Workloads, Data Governance, Security and Monitoring, and the In-Memory Cache Cluster.

6.1 Core Analytical Platform Components

This group encompasses the primary data processing and storage infrastructure. Sizing within this group is driven directly by measured on-premises database footprints and the storage requirements of the Medallion (Bronze / Silver / Gold) architecture to be implemented on the new platform.

6.1.1 MPP Enterprise Warehouse Sizing

Metric	Day 0 Baseline	Year 1	Year 2	Year 3	Year 4	Year 5
MPP Compressed Storage (TB)	144.0	172.8	198.7	218.6	240.5	264.5
Uncompressed Equivalent (3.5x, TB)	504.0	604.8	695.5	765.1	841.6	925.7

Note: The 3.5x industry-standard decompression multiplier is applied to derive uncompressed equivalents for capacity planning purposes. Actual compression ratios are dependent on the MPP vendor technology and shall be validated during the Vendor's formal sizing exercise.

6.1.2 Object Store Sizing

Metric	Day 0 Baseline	Year 1	Year 2	Year 3	Year 4	Year 5
Total Object Store Capacity (TB)	181.5	199.7	229.6	275.5	330.6	396.7

Note: These figures represent available provisioned capacity requirements. Erasure coding or replication overhead applicable to the selected object storage solution shall be factored into final capacity by the Vendor.

6.1.3 ELT Processing Node Sizing

Parameter	Environment	Nodes	CPU (Physical Cores)	RAM	Local Storage / NIC

ELT Processing Node	Production	3	16 Cores/Node	256 GB/Node	1 TB NVME SSD / 25 GbE
ELT Processing Node	Development	1	16 Cores/Node	256 GB/Node	1 TB NVME SSD / 25 GbE
TOTAL NODES	**Prod + Dev**	**4**	-	-	-

Note: 16 physical cores per node is the reference specification. The Vendor shall adjust core count based on the core-based licensing model applicable to their ELT product.

6.2 Supporting Infrastructure Components

This group covers the infrastructure components that enable AI/ML workloads, real-time data processing, enterprise data governance, platform-wide security enforcement, and high-performance in-memory caching.

6.2.1 Streaming and AI/ML Workload Sizing

Production

Type	Role	Count (Prod)	Hardware Specification Per Node
A	Master / Controller Nodes (HA + Failover)	3	16 physical cores, 256 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 8x1 TB NVME SSD, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU
B	Worker Nodes (Data Processing Engines) - Compute and Spool Intensive	9	32 physical cores, 512 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 12x2 TB NVME SSD, NIC: 25 GbE bonded 2x, Redundant Hot-Swap PDU
C	Streaming Platform Nodes (Event Brokers + Stream Processing)	3	32 physical cores, 512 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 4x4 TB NVME SSD, NIC: 25 GbE bonded 2x, Redundant Hot-Swap PDU
D	Utilities and Gateway Nodes	3	16 physical cores, 256 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 4x1 TB NVME SSD, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU

Development and Test Environment Node Sizing

Type	Role	Count (Prod)	Hardware Specification Per Node
A	Master (No HA)	1	16 cores, 256 GB RAM, 2x500 GB RAID1 + 8x1 TB NVME SSD, 10 GbE
B	Utility + Worker Nodes	3	32 cores, 512 GB RAM, 2x500 GB RAID1 + 12x2 TB NVME SSD, 25 GbE

Five-Year Node and Storage Growth Projection (Production)

Metric	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Total Production Nodes	18	20	22	24	27	30
Worker Node Count	9	11	13	15	18	21
Raw Data Storage - Worker Nodes (TB)	216	264	312	360	432	504

Post 3x Replication (Usable TB)	72	88	104	120	144	168

6.2.2 Data Governance Sizing

Production

Type	Role	Count (Prod)	Hardware Specification Per Node
A	Data Governance Platform and Search Node	1	8 physical cores, 128 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 2x2 TB NVME SSD RAID1, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU
B	Metadata Repository Node (Relational DB Backend)	1	8 physical cores, 128 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 2x4 TB NVME SSD RAID1, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU
C	Monitoring Node	1	4 physical cores, 64 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 2x2 TB NVME SSD RAID1, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU
D	Console and Web UI Node	1	4 physical cores, 64 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 2x1 TB NVME SSD RAID1, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU
E	Job Server / Lineage Processing Node	1	16 physical cores, 256 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 2x2 TB NVME SSD RAID1, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU

Development and Test Environment Node Sizing

Type	Role	Count (Prod)	Hardware Specification Per Node
E	All Services - Single Node	1	All services consolidated on a single node; no HA. Dev/Test footprint shall not expand over the 5-year period.

Five-Year Node Growth Projection (Production)

Metric	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Total Production Nodes	5	5	6	6	7	7
Expansion Trigger		One Job Server Node added every second year (Year 2 and Year 4) to accommodate increased lineage and profiling workloads.				

Note: The sizing baseline for this component was established using an enterprise data governance platform reference. The actual node count, memory, and storage shall be validated by the Vendor against the specific governance product selected and its licensing topology.

6.2.3 Security and Monitoring Sizing

Production

Type	Role	Count	Hardware Specification Per Node
A	Directory Services, Kerberos KDC, Key Management Service, and Platform Monitoring - Deployed in Active-Passive HA Pair	2 (HA Pair)	16 physical cores, 256 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 2x2 TB NVME SSD RAID1, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU

Development and Test Environment Node Sizing

Type	Role	Count	Hardware Specification Per Node
A	Directory Services, Kerberos KDC, Key Management Service, and Platform Monitoring - Deployed in Active-Passive HA Pair	1 Single Node	16 physical cores, 256 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 2x2 TB NVME SSD RAID1, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU

6.2.4 In-Memory Cache Cluster Sizing

Production

Type	Role	Count	Hardware Specification Per Node
A	In-Memory Cache - Clustered Deployment (3-Node Minimum for HA and Data Partitioning)	3 (Prod Initial)	8 physical cores, 128 GB RAM, OS: 2x500 GB NVME SSD RAID1, Data: 2x2 TB NVME SSD RAID1, NIC: 10 GbE bonded 2x, Redundant Hot-Swap PDU

Development and Test Environment Node Sizing

Type	Role	Count	Hardware Specification Per Node
A	Single Node - All Services	1	Standalone non-clustered node. Dev/Test footprint shall not expand over the 5-year period.

Five-Year Cache Node Growth Projection (Production)

Metric	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Total Production Cache Nodes	3	3	4	4	5	5
Expansion Trigger		One Cache Node added every second year (Year 2 and Year 4) to accommodate increased session volumes, AI feature serving loads, and real-time tax lookup concurrency.				

Sizing Assumptions:

- No software licensing costs, database engine overhead, middleware footprint, or operating system overhead has been included in any of the sizing figures. These are hardware-only reference estimates.
- The sizing does not account for rack topology, Top-of-Rack network switches, Spine and Core network switches (uplink and downlink), rack cabling, DNS infrastructure, NTP, or out-of-band management networks for any component group. These elements are entirely dependent on the Vendor's solution architecture and shall be provided in the Vendor's detailed bill of materials in collaboration with the procuring authority's data center team.
- Data & Compute Growth (Approximate 10% yearly growth – estimate, vendor to validate).

7. Infrastructure Implementation & Commissioning

This section formalizes the Vendor's responsibility to physically install, configure, secure, and commission all infrastructure shown in the architecture diagram (Section 5.2 and Annexure), across both Production and Development/Test environments, up to the point where the environment is ready for data platform implementation to begin

7.1 Scope of Work

- Site preparation coordination: verification of rack space, power, cooling, and network cross-connect readiness in coordination with PRAL's facilities team.
- Physical installation (rack-and-stack) of all compute, storage, and networking hardware per the approved Bill of Materials and the sizing specifications defined in Section 6.
- Cabling, network configuration, and connectivity validation across all environments, including VLAN segmentation, firewall rules, and inter-environment isolation consistent with Section 09 (Environments).
- Operating system installation, hardening, and patching for all nodes, aligned to FBR/PRAL's security baseline requirements.
- Installation and base configuration of all licensed platform software components (Lakehouse/object storage engine, MPP Enterprise Data Warehouse, compute cluster runtime, in-memory cache cluster, and supporting infrastructure software) up to the point of readiness for the data platform Implementation Vendor to begin pipeline configuration.
- High-availability and redundancy configuration for all infrastructure components per the architecture diagram and Section 10 (HA, Backup & Disaster Recovery).
- Infrastructure-level backup and disaster recovery configuration (storage replication, backup agents, DR site connectivity) as a physical/platform-layer capability, distinct from the application-level DR procedures
- Infrastructure security hardening: network segmentation, encryption-at-rest configuration, access control at the OS/hypervisor level, and integration with FBR/PRAL's directory services.
- DR failover testing to validate that backup/replication/failover actually works as configured.
- Infrastructure-level performance testing (throughput, latency, IOPS) to confirm the environment meets sizing specifications in Section 6, prior to formal handoff.
- Walkthrough or training sessions for PRAL's internal team and/or the Data Platform Implementation Vendor on the deployed infrastructure.

7.2 Infrastructure Acceptance

"Infrastructure Acceptance" is defined as a formal milestone, marking the point at which all infrastructure is installed, configured, secured, and validated as ready for data platform implementation to commence.

- Infrastructure Acceptance Testing shall include hardware burn-in tests, network throughput and latency validation, HA failover testing, and a security baseline scan.
- Testing shall be jointly witnessed by PRAL and, where applicable, the separately contracted data platform Implementation Vendor.
- Written sign-off from PRAL is mandatory before infrastructure is handed over for data platform implementation.

7.3 Deliverables

- As-built infrastructure **diagrams** and configuration documentation.
- **Network and security configuration documentation:** VLAN mapping, firewall rule sets, IP addressing scheme, and access control matrix.
- **Asset inventory register:** complete list of hardware/software assets including serial numbers, warranty start/end dates, and license keys.
- **HA/DR configuration and failover test report**: documentation of the high-availability setup and results of the DR failover test/drill.
- **Infrastructure Acceptance** Test Report for the entire solution.
- **Environment readiness certificate:** formal sign-off document confirming the infrastructure is ready for the Data Platform Implementation Vendor to begin pipeline configuration.
- **Infrastructure runbooks**, including power-on/power-off procedures, hardware maintenance procedures, and escalation contacts.

7.4 Implementation Timelines & Milestones

This timeline covers infrastructure readiness only, scoped exclusively to the hardware, software licensing, and infrastructure commissioning activities defined in this RFP. The total duration for this scope is 5 months (20 weeks) from contract signing to Infrastructure Acceptance. Data platform and application implementation timelines are defined separately under the Data Platform & Application Implementation Services engagement and commence only after Infrastructure Acceptance is achieved, consistent with Section 19 (Vendor Coordination & Interface Management).

Milestone	Timeline	Exit Criteria/ Deliverables
Contract Signing, Kickoff & Site Readiness Assessment	Week 1-2	Joint kickoff completed; site survey confirming rack space, power, cooling, and network cross-connect readiness.

Hardware Delivery & Software License Provisioning	Week 3-14	All hardware delivered to site per approved Bill of Materials; all software licenses provisioned and activated.
Physical Installation, Cabling & Network Configuration	Week 15	Rack-and-stack complete; VLAN segmentation, firewall rules, and inter-environment isolation configured and validated.
OS Installation, Hardening & Base Platform Software Installation	Week 16-17	All nodes hardened and patched; licensed platform software installed and base-configured across Production and Development/Test environments.
HA, DR & Security Configuration	Week 18-19	High-availability and redundancy configured per the architecture diagram; infrastructure-level backup, DR connectivity, and security hardening in place.
Infrastructure Acceptance Testing & Sign-Off	Week 20	Hardware burn-in, network throughput/latency validation, HA failover testing, and security baseline scan completed; written Infrastructure Acceptance sign-off issued by PRAL.

Milestone-linked payment percentages, where applicable, shall be tied to the exit criteria defined above. Any delay to Infrastructure Acceptance beyond the committed 5-month duration shall be assessed under Section 18 (Liquidated Damages)

8. Performance & Scalability

Purpose & Strategic Context

This section defines how infrastructure performance targets are set, how compute, storage, and network components are validated against these targets before go-live, and how the infrastructure is confirmed to meet the sizing, high-availability, and capacity requirements defined in Section 6 and Section 11, prior to handoff for data platform implementation.

Scope of Work

The Performance and Scalability workstream should address nine interdependent domains, each contributing to a platform that is fast, elastic, cost-efficient, and operationally transparent:

- **Compute Performance Benchmarking**: CPU/core performance validation (e.g., SPEC benchmarks or equivalent) confirming compute nodes meet the sizing specifications defined in Section 6.
- **Storage Throughput & IOPS Testing**: validating read/write throughput, latency, and IOPS of storage systems against vendor-specified and RFP-required performance thresholds.
- **Network Bandwidth & Latency Validation**: testing inter-node, inter-environment, and cross-connect network performance (throughput, latency, packet loss) against architecture requirements.

- **Memory Performance Validation**: validating in-memory cache cluster and node-level memory throughput/latency meet specified performance targets.
- **HA Failover Performance Testing**: measuring failover time and performance impact during simulated node/component failure, confirming it meets RTO targets defined in Section 11.
- **Backup & Recovery Performance Testing**: validating backup completion times and recovery/restore times against defined RPO/RTO targets.
- **Resource Headroom Validation**: confirming available CPU, memory, storage, and network capacity headroom at go-live against projected infrastructure sizing (not workload/application sizing, which is the Data Platform Vendor's responsibility).
- **Infrastructure Stress Testing**: validating stability and performance of compute, storage, and network components under sustained peak load conditions, prior to formal handoff.
- **HTAP Performance:** HTAP performance report showcasing metrics and usage.

09. Environments

The vendor shall deliver a three-tier environment architecture comprising Development, Test, and Production, deployed across two physically segregated infrastructure footprints.

Environment Structure

- Production shall reside on a dedicated physical environment, fully isolated at the hardware, network, and storage levels from all non-production workloads. Under no circumstances shall Production share physical compute, storage, or network fabric with Development or Test.
- Development and Test shall be co-located on a single, capacity-reduced secondary physical environment, logically segregated from one another through independent namespaces, isolated execution contexts, dedicated service accounts, and logically segregated storage.
- The Development and Test environment shall not constitute a fully running replica of Production. It shall be deliberately scaled down in node count and per-node capacity to minimize total cost of ownership while preserving full functional coverage of every Production component. The vendor shall propose and justify all sizing and capacity reduction factors.

Production Environment

- Production shall be provisioned as a complete, fully operational, end-to-end instantiation of the target platform.

- All components including the ELT orchestration layer, compute cluster, MPP Enterprise Data Warehouse, object storage, and in-memory caching tier shall be deployed at full production capacity.
- All performance, concurrency, throughput, and high-availability SLAs specified in this RFP shall be met exclusively against the Production environment.

Security and Isolation

- Each tier shall be governed by formally enforced security boundaries preventing unauthorized lateral movement of data, credentials, configurations, or workloads.
- Development and Test shall have zero direct network reachability to Production data assets, compute resources, credential stores, or metadata repositories.
- RBAC policies, service accounts, secrets management, and audit logging shall be independently maintained per environment.
- All cross-environment activity shall traverse a controlled, auditable pipeline aligned with governance principles.

Environment Parity

- Development and Test shall remain structurally and functionally identical to Production with respect to component versions, patch levels, schema definitions, data models, pipeline logic, and configuration parameters at all times, notwithstanding their reduced capacity.
- The vendor shall provide tooling and documented procedures to detect, report, and remediate parity drift on a continuous basis.

Data Masking and Synthetic Data

- The vendor shall implement a data masking and synthetic data management capability enabling policy-governed promotion of representative datasets from Production into lower tiers.
- Static masking shall be applied at promotion time. Dynamic masking shall be applied at query execution time.
- Masking shall cover all PII and taxpayer-sensitive attributes per the platform's data classification taxonomy.
- Masked datasets shall preserve referential integrity, statistical distribution, and structural fidelity sufficient to support meaningful development, testing, and model validation.

10. HA, Backup & Disaster Recovery

High Availability

All critical components including the ELT pipeline engine, compute clusters, MPP Enterprise Data Warehouse and in-memory caching cluster shall be deployed in an active-active or active-passive configuration appropriate to their workload profile.

No single point of failure shall be permitted at any tier including compute, storage, and network layers.

Automatic failover mechanisms shall be implemented across all layers with formally defined and contractually binding Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO).

Health monitoring, heartbeat detection, and self-healing capabilities shall be implemented to minimize manual intervention during fault conditions.

Sr no.	Requirements
1	As per the business criticality of the platform system availability of 99.99% (4 minutes and 38 seconds unplanned outage) has to be ensured.
2	The system must be designed so that hardware and software failures are handled without the entire system going offline.
3	The back-up/archive/restore (BAR) hardware system shall be supported.
4	The system shall have proper migration strategy from TEST to PROD environments.
5	The system shall not have a “single point of failure” and shall provide complete fault-tolerance with full redundant hardware components.
6	The product must have built-in backup and recovery, and it must support important backup features like full, partial or incremental.
7	The product must support restore of a single table or group of tables from a database level or backup.
8	Vendor should also propose a cost-effective DR Solution; product should have a built in HA/DR solution at application level.
9	System shall possess automated recovery mechanisms with minimum downtime impact on end-users. Describe how faults or failures in each of your system components are handled, the DBA involvement and the downtime impact on the end-users.
10	Failure of any storage should not have an impact on the ongoing activity/query/loss of data. Please describe the procedure of maximum availability feature of your proposed solution.
11	System shall possess automatic failover capabilities i.e. System shall continue to run in the event of failure of a hardware processing unit. Describe also any impact

	on overall performance while the failed hardware processing unit is being replaced.

Backup

The platform shall incorporate a SAN-based backup framework protecting all ecosystem components without reliance on tape-based media.

The platform shall expose all necessary APIs, connectors, and integration hooks to enable application-aware backup operations across the ELT tool, compute cluster state, MPP Warehouse data and metadata, object storage buckets, and the in-memory cache layer.

Sr No.	Requirements
1	All backup data shall be encrypted at rest using AES-256 or equivalent, with encryption keys managed independently from backup media and stored in a dedicated, access-controlled key management facility.
2	Backup storage shall enforce write-once/read-many (WORM) immutability to prevent unauthorized modification or deletion of backup artifacts, including protection against ransomware-driven corruption.
3	The platform shall support full, incremental, and differential backup modes.
4	The platform shall enforce a tiered retention schedule, aligned with statutory tax record retention obligations.
5	The platform shall support granular restore operations at the level of individual tables, schemas, database objects, pipeline job definitions, and object storage prefixes, without requiring a full environment restore.
6	All backup and restore operations shall be executable without taking the production environment offline. Online, non-disruptive backup shall be a mandatory capability across all platform components.
7	Backup job execution, completion status, duration, data volume protected, and any failures shall be logged in a centralized, tamper-evident audit trail accessible to platform administrators.
8	The vendor shall define and document a formally tested, step-by-step restore runbook for each platform component. Restore procedures shall be validated during the implementation phase with results formally signed off by the platform owner.
9	Backup infrastructure capacity shall be sized to retain the full retention schedule defined above without requiring manual capacity intervention for a minimum of 5 years from the date of go-live.
10	Backup infrastructure shall be added in the BoQ.

Disaster Recovery

The vendor to propose, supply, install, and commission a dedicated Disaster Recovery (DR) environment at a geographically separate location from the primary production site.

The DR environment shall be architecturally identical to production in terms of software stack, component versions, configuration baselines, and functional capability.

To optimize total cost of ownership, the DR environment needs to be provisioned at a reduced capacity footprint during normal operations, including:

Sr No.	Requirements
1	Reduced compute cluster node count.
2	Reduced MPP Warehouse appliance node count.
3	Proportionally scaled-down object storage capacity.
4	Smaller in-memory caching cluster sized to sustain critical workloads during a declared disaster event.
5	The DR environment will support rapid horizontal scaling to match full production-level performance if an extended failover scenario requires sustained full-scale operations.
6	Continuous, direct synchronization with production will be maintained through a documented replication mechanism covering MPP Warehouse data, object storage contents, ELT configurations and metadata, and job definitions, within contractually agreed RPO thresholds.
7	The vendor will provide automated, repeatable procedures for planned failover, unplanned failover, and failback operations at the infrastructure layer, and shall conduct full-scale infrastructure DR tests as part of the infrastructure implementation scope
8	The platform will achieve an RTO of no greater than 30 minutes and an RPO of no greater than 15 minutes for Tier-1 critical workloads. Secondary workloads shall observe an RTO of no greater than 4 hours and RPO of no greater than 1 hour.
9	The vendor will document and execute bi-annual DR failover drills post go-live, with results reported formally to the platform owner within 5 business days. Any RTO/RPO breach observed during testing shall trigger a formal remediation plan within 30 days.
10	The DR environment will be subject to automated configuration drift detection against the production baseline, with alerts generated for any deviation in software versions, configuration parameters, or schema definitions. Monthly drift reports shall be produced.
11	The replication channel between the primary production site and the DR environment will traverse a minimum of two independent network paths with automatic path failover. Replication link health will be continuously monitored with alerting triggered within 60 seconds of link degradation.

11. Data Security, Risk & Compliance

FBR has strict Data security and confidentiality policy which supplier is required to comply during all the phases of the project. User related data for development/testing purposes should not be copied on personal laptops/storage used by the project team.

Sr No.	Requirements
1	The system will support integration with AD/LDAP authentication
2	The system can provide single sign-on capabilities.
3	The system will have different levels of access for users including administrator, normal user and power users.
4	Protection from unauthorized user access needs to be ensured. Please describe how reports and data are protected from unauthorized users.
5	Users Groups (i.e. Roles Definition) segregation is required.
6	Passwords stored in an encrypted form.
7	Password can be reset by administrator.
8	Industry standard encryption methodology should be used on data, reports, and logins.
9	Security at different layers needs to be ensured. Please describe the following levels of security in the proposed solution: • User Layer Security • Transmission Layer Security • Application Layer Security • Data Layer Security
10	The system can support both row and column level security. Data in transit should be encrypted and also at rest.
12	The solution can support custom user groups and roles definition and configuration.
13	Integration with FBR's SIEM should be supported.

12. Service Specification and QA requirements

Technical Support and Maintenance

The Vendor shall provide 24/7 support for all infrastructure components throughout the contract term. Qualified technical personnel shall be available for preventive and corrective maintenance, first-level diagnostics, and urgent issue resolution across all supplied hardware and infrastructure software components. Post-warranty support shall include hardware/software patches, security updates, and infrastructure-level performance tuning for the duration specified in the contract.

13. Software and Hardware Services and Warranty

Software and Infrastructure Warranty

The Vendor shall provide warranty coverage for all hardware and infrastructure software components, including operating systems, virtualization/platform software, and supporting infrastructure tools, delivered under this contract for a minimum period of five years from the date of Infrastructure Acceptance. During this period, the Vendor shall rectify, at no additional cost, any defects, bugs, performance issues, or non-conformities identified in the delivered infrastructure and its deliverables.

Hardware Warranty

All hardware components procured and commissioned under this contract, including compute nodes, storage appliances, networking equipment, and associated infrastructure, shall carry a minimum five-year warranty from the date of Infrastructure Acceptance. The Vendor shall ensure next-business-day on-site replacement for any failed hardware component. No single hardware failure shall result in platform downtime exceeding the RTO defined in Section 10 of this RFP.

Software Maintenance and Updates

Throughout the warranty period, the Vendor shall provide all software updates, security patches, and version upgrades for every licensed platform component at no additional cost. All updates shall be tested in the Development/Test environment before promotion to Production. The Vendor shall notify FBR/PRAL of all critical security patches within 24 hours of release and shall apply them within a mutually agreed maintenance window.

Support SLAs

The Vendor shall maintain the following minimum response and resolution commitments throughout the warranty period:

Severity	Description	Response Time	Resolution Time
Critical	Platform down or data loss risk	1 Hour	4 Hours
Major	Significant function impaired	4 Hours	24 Hours
Minor	Non-critical defect or degradation	1 Business Day	5 Business Days

Post-Warranty Support

The Vendor shall propose post-warranty support options in its commercial proposal, including available support tiers, SLAs, annual maintenance charges, and provisions for ongoing platform enhancements and technology refresh.

14. Payment Terms

The following are the payment terms for CDH:

- Responsibility to include all and correct taxes lies on the bidders.
- Price should be in Pak Rupees only.
- The quantity of items or services may be increased or decreased at PRAL's sole discretion based on operational requirements at the time of contract award.
- Bidders must submit bid for all items and services listed in the scope of work / BOQ. Partial bidding is not permitted.
- CAPEX (estimated 60% of total contract value) covers the procurement and supply of all hardware, software licenses, and physical infrastructure components as per the approved Bill of Materials.
- OPEX (estimated 40% of total contract value) covers Hardware/Software Warranty Support for the deployed infrastructure for five years following production go-live.
- Yearly OPEX payments shall be invoiced annually, at the end of each contract year, and are subject to the Vendor meeting the agreed Service Levels (SLAs/KPIs) for that year.
- Payment shall be made as per the milestone-based schedule defined below, tied to the deliverables and acceptance criteria for each phase. No separate CAPEX/OPEX split applies to this engagement

CAPEX (Estimated 60% of the total contract value):

S.No.	Milestone	Deliverables	Payment %	Acceptance Criteria
1	Contract Signing, Kickoff & Site Readiness Assessment	Contract signing, project initiation, mobilization of resources, and commencement of project activities.	20%	Payable upon contract signing and project commencement.
2	System Configuration	All hardware delivered to site per approved Bill of Materials; all software licenses provisioned and activated. Rack-and-stack complete; VLAN segmentation, firewall rules, and inter-environment isolation configured and validated. All nodes hardened and patched; licensed platform software installed and base-configured across Production and Development/Test environments. High-availability and redundancy configured per the architecture diagram; infrastructure-level backup, DR connectivity, and security hardening in place	30%	Hardware delivered per BOM; rack-and-stack complete; network segmentation validated; nodes hardened/patched; HA/DR and security hardening configured per architecture diagram.
3	Sign-off & Acceptance	Hardware burn-in, network throughput/latency validation, HA failover testing, and security baseline scan completed; written Infrastructure Acceptance sign-off issued by PRAL.	50%	Infrastructure Acceptance sign-off issued by PRAL

OPEX (Estimated 40% of the total contract value):

S.No.	Milestone	Deliverables	Payment %	Acceptance Criteria
1	Hardware/ Software Warranty	Vendor shall provide hardware and software support, along with support for infrastructure	20% each year over five years	Subject to the vendor meeting

	Support and expansion	expansion, over a period of five years following production go-live.	following production go-live.	the agreed service levels (SLAs/KPIs)

15.Bill of Quantity Guidelines for Vendor

The Vendor shall submit a detailed Bill of Quantities (BOQ) outlining all costed elements required for the supply, implementation, and operation of the Central Data Hub. The BOQ shall be submitted as a structured document forming part of the commercial proposal.

Formatting Requirements

- The BOQ shall clearly separate one-time costs from recurring annual costs.
- All unit prices shall be quoted inclusive of all sub-components, associated cabling, mounting hardware, and ancillary parts required for full operational readiness. Bundled or package pricing shall be justified in a short technical note explaining what is included.
- All costs shall be quoted in PKR. Foreign currency components, where applicable, shall be identified separately with the applicable exchange rate basis and date stated.
- Each cost line item shall reference the corresponding technical requirement in Section 5 (Functional and Technical Requirements) or Section 6 (Sizing Estimation) that it fulfills, to allow direct traceability between the technical and commercial proposals.
- Quantities quoted shall match, at minimum, the reference sizing defined in Section 6; any deviation shall be explained in an accompanying technical note.

Mandatory Cost Categories

The Vendor shall provide unit-wise and cost-wise breakdowns covering at minimum the following categories:

1. **Hardware Supply:** A line-item breakdown per physical component: compute nodes, storage appliances (stating usable and raw capacity), networking equipment (switches, cabling, transceivers), rack infrastructure (racks, PDUs, cooling if in scope), and any spares or buffer stock. Each line item shall state specification reference, unit of measure, quantity, unit price, and total price.

2. **Software Licenses:** A line-item breakdown per software product (ELT/ETL tool, Object Storage/Lakehouse engine, MPP Enterprise Data Warehouse engine, In-Memory Caching software, Data Governance & Catalog platform, Operating System per node, and any other licensed component required to operate the complete platform), stating license type (Perpetual, Annual, Core-Based, Capacity-Based, or User-Based), licensing metric, quantity, unit price, and total price. No software component required for full platform operation shall be omitted from this breakdown.
3. **Hardware Warranty and Support**: The annual cost per hardware component (matching the line items declared under Hardware Supply) for each of the five years of the warranty period, together with the applicable SLA tier (response time and resolution time) for each component.
4. **Software Maintenance and Support:** The annual maintenance cost per software component (matching the line items declared under Software Licenses) for each of the five years, inclusive of version upgrades, security patches, and SLA-based technical support, together with the applicable SLA tier for each component.

The Vendor shall also submit a Grand Total Summary Consolidating Categories 1–4 into the total evaluated bid value, with Category 5 shown separately as indicative pricing. Please see the financial proposal format for reference as mentioned in the section (20)

16. Mandatory Clauses for Vendor Eligibility

All requirements listed below are non-negotiable. Any proposal that does not fully comply with each clause shall be disqualified from evaluation.

#	Requirement	Mandatory Specification
1	**Full On-Premises Deployment**	No component, pipeline, data asset, or process may reside on or transit through any public cloud infrastructure or external network at any point. The complete CDH shall operate exclusively within the designated on-premises network perimeter. All software and operating system licenses shall be included in the vendor's proposal.
2	**Complete Software Stack for the Full Reference Architecture**	The Vendor must provide a software stack covering every component of the Reference Architecture in Section 5.2, including Data Ingestion, MPP, Object Storage, Governance, AI/ML Processing and BI. Each component must be named with its product, license model, and vendor support. Any proposal missing or leaving a component unaddressed shall be disqualified.
3	**Data Ingestion Tool Capabilities**	The Vendor's proposed data ingestion tool stack must natively support Change Data Capture (CDC), real-time streaming, and API-based ingestion as first-class capabilities. CDC must be log-based. Real-time streaming must support continuous, low-latency event

		processing. Any proposal that does not demonstrate all three capabilities natively within the proposed ingestion tool stack shall be disqualified.
4	**MPP Enterprise Data Warehouse**	The Vendor must propose an enterprise grade OEM Engineered machine that supports the deployment of Enterprise data warehouse capable of scale-out parallel processing, or an equivalent architecture, to support high-volume, high-concurrency analytical workloads with predictable performance, high availability, security, and operational manageability. Software only MPP engines on generic commodity hardware are not acceptable. Proposals not meeting this shall be disqualified
5	**Object Storage**	The Vendor must propose a purpose-built object storage system. Commodity hardware and software-defined storage solutions are not acceptable. The proposed object storage must be natively compatible with the proposed MPP platform, with direct, native integration and without reliance on third-party connectors or middleware. Proposals not meeting this shall be disqualified.
6	**BI Tool**	The Vendor must propose an enterprise grade BI tool with native connectivity to the proposed Object Store and MPP platform. The tool must support self-service reporting and dashboarding for business users, in addition to standard scheduled and operational reporting. Proposals not meeting this shall be disqualified.
7	**Annual turnover**	Minimum average annual turnover of **PKR 2.0 billion** over the last three (3) fiscal years, certified by a Chartered Accountant. In the case of a **Joint Venture (JV)**, this turnover may be met through the **combined average annual turnover of all JV partners**. Additionally, where applicable, the **turnover/financial standing of the OEM** (in support of the Bidder/JV through an OEM authorization or support letter) **may be combined with the Bidder's/JV's turnover** to meet the minimum eligibility threshold.
8	**Enterprise-scale Hardware/Software Readiness**	Minimum of two (2) completed enterprise-scale Hardware/Software Readiness or Infrastructure Deployment engagements (covering procurement, installation, configuration, and commissioning of hardware and/or software systems), each with a contract value of USD 1,000,000 or above (or PKR equivalent), operational/live in production for at least twelve (12) months.
9	**Continuous Operation**	Minimum of (3-4) years of continuous operation as a registered entity providing enterprise hardware/software procurement, integration, and support services, with demonstrated experience as an authorized reseller/partner of relevant OEMs.
10	**Data Governance Tool**	The Vendor must propose an enterprise grade data governance tool backed by vendor support with a defined SLA. Community supported or open-source tools without enterprise support are not acceptable. The tool must align with latest DAMA-DMBOK principles, covering Business Glossary, Data Quality, Data Classification, Data Ownership & Stewardship, Metadata management, Data Lineage, and Master data management. Proposals not meeting this shall be disqualified.

11	**Native Multi-Model & HTAP Processing**	The proposed platform engine shall natively support both Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) within a single unified engine, combined with genuine Hybrid Transactional and Analytical Processing (HTAP), delivered natively by the platform engine itself rather than via federation of separate specialized systems. Proposals that do not demonstrate native Multi-Model / HTAP capability, evidenced by vendor-published global reference deployment data, shall be disqualified from evaluation.

17. Evaluation Criteria

The evaluation shall follow the Single Stage, Two Envelope Bidding procedure in accordance with Rule 36(b) of the Public Procurement Rules 2004. Technical and financial proposals shall be submitted simultaneously in separate sealed envelopes. The Technical Envelope shall be opened and evaluated first; the Financial Envelope shall remain sealed and shall be opened only for bidders securing a minimum of 70% of the total technical score. The contract will be awarded on a turnkey basis to the bidder with the highest Combined Evaluation Score as per the following criteria:

a. Technical Component (TC) = Total technical evaluation score
b. Financial Component (FC) = (Lowest Bid / Respective Bid) * 100
c. Combined Evaluation Score = (TC * 70%) + (FC * 30%)

Technical Evaluation (Total: 100 Marks, 70% Required to Pass)

A. Technical Solution (65 Marks)

Criteria	Description	Marks	Documents Required
Solution Cohesiveness & Enterprise Support	Vendor shall demonstrate an integrated, enterprise-supported infrastructure solution covering all required hardware and infrastructure software components (Section 6), using the minimum number of distinct products necessary to avoid unnecessary architectural complexity. Where multiple components serve overlapping functions, the Vendor shall justify why a single consolidated component was not used instead.	0-8	End-to-end infrastructure architecture diagram showing all proposed hardware/software components and integration points; enterprise support agreement or OEM partnership evidence for each proposed component.
Solution Architecture ELT/Data Integration	Data integration and ingestion covering source connectivity that covers schema enforcement, CDC methodology, and data quality	0-5	Network and connectivity architecture diagram, infrastructure design

	controls. Assessed on completeness of source-to-target mapping approach and robustness of the proposed ingestion methodology.		document, and source system connectivity matrix.
Solution Architecture MPP Enterprise Data Warehouse (Database Engine)	Enterprise Data Warehouse engine covering parallel query processing, workload management, and query optimization capability, in line with Section 5.7. Assessed on architectural soundness and alignment with sizing requirements in Section 6.	0-7	Design of the compute and storage infrastructure supporting the MPP Enterprise Data Warehouse, covering node sizing, cluster architecture, and storage performance characteristics, in line with Section 6
Solution Architecture Object Storage/ Lakehouse	Object storage and Lakehouse layer covering medallion architecture (bronze/silver/gold), open table format support, and native object storage connectivity from the MPP engine. Assessed on architectural completeness and open-format compliance, also covering capacity planning, redundancy, and performance characteristics, in line with Section 6.	0-5	Design of the storage infrastructure layer covering capacity planning, redundancy, performance characteristics, and connectivity architecture to support object storage workloads, in line with Section 6
Solution Architecture Compute Cluster/AI-ML Platform	Distributed compute cluster supporting AI/ML workloads and micro-batch/stream processing, covering compute-storage separation and workload isolation from the MPP engine. Assessed on scalability and workload isolation approach.	0-5	Compute cluster architecture diagram and AI/ML platform design.
BI Tools for Analytics	The Vendor shall specify the proposed BI/semantic layer tool and describe its integration with other platform components (e.g., MPP Enterprise Data Warehouse, Lakehouse, Data Governance and Catalog). Assessed on completeness of the proposed self-service and reporting architecture, and the robustness of tool selection and cross-tool integration	0-6	BI/semantic layer architecture diagram and tool selection justification.
Real-Time/ Streaming Data Ingestion	Compute layer (real-time streaming compute clusters) covering node architecture, workload isolation, and alignment with sizing requirements in Section 6.	0-5	Streaming compute architecture diagram, node/cluster sizing document, and evidence of infrastructure throughput/latency benchmarks for comparable deployments.

HTAP/Multi-Model Processing Capability	The proposed Solution Architecture shall demonstrate native multi-model processing capability, supporting both Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) within a single unified engine, combined with genuine Hybrid Transactional and Analytical Processing (HTAP), rather than federation of separate specialized systems. Assessed on evidence of native (non-federated) HTAP capability and alignment with the mixed transactional/analytical profile	0-7	HTAP/multi-model architecture diagram, and vendor-published benchmark or reference deployment evidence demonstrating concurrent OLTP/OLAP processing without ETL-driven replication.
Data Governance and Security	Demonstrated approach to RBAC, field-level encryption, dynamic data masking, PII protection, audit logging, and latest DAMA-DMBOK alignment across all platform layers. Catalog layer covering metadata management, business glossary, data lineage, and catalog integration across all platform layers. Assessed on completeness of the proposed governance architecture	0-8	security controls note referencing encryption protocols, access control design
High Availability, Backup & Disaster Recovery	High-availability and redundancy mechanisms across compute, storage, and network components, combined with infrastructure-level backup and disaster recovery architecture, including storage replication, backup encryption, DR site connectivity, and RTO/RPO alignment, per Section 10 (HA, Backup & Disaster Recovery).	0-5	HA/redundancy design, DR architecture diagram, backup/replication design document, and RTO/RPO commitment note.
Vendor Presentation and Solution Walkthrough	Vendor shall present the proposed infrastructure solution covering architecture, installation and commissioning approach, and infrastructure readiness handover plan before the evaluation committee	0-4	Presentation delivered in person before the evaluation committee.

Note: For each criterion above, the Vendor shall provide documented justification covering suitability, scalability, licensing model, and vendor/community support maturity for the proposed tool/technology, consistent with FBR's on-premises.

B. Hardware Specifications and Delivery (15 Marks)

Criteria	Description	Marks	Documents Required
Hardware Specifications for Entire Solution	Proposed hardware shall meet or exceed all reference sizing specifications defined in Section 6, covering compute nodes, storage appliances, networking, and supporting infrastructure across Production and Development/Test environments. Assessed on specification compliance, redundancy design, and scalability provisions.	0-10	Detailed BOQ with per-component technical datasheets, sizing justification document, and five-year capacity expansion plan.
Delivery and Commissioning Commitment	Vendor shall provide a binding hardware delivery and commissioning schedule consistent with the timeline defined in Section 7.4 (Implementation Timelines and Milestones). Assessed on delivery lead time commitment, installation plan, and risk mitigation for procurement delays.	0-5	Signed hardware delivery schedule, OEM supply chain confirmation, and contingency plan for delivery risks.

C. Historical Experience (15 Marks)

Criteria	Description	Max Marks	Marking	Documents Required
Experience in Enterprise Hardware & Infrastructure Deployment	Demonstrated experience delivering large-scale, on-premises hardware and infrastructure (compute, storage, and networking) for enterprise analytics platforms, data warehouses, or equivalent large-scale analytical environments, over the past seven years. Minimum of two (2) referenced projects with a contract value of USD 1,000,000 or above (or PKR equivalent), each live	10	2.5 marks per project, minimum 2 projects	Signed contract or agreement, hardware delivery/commissioning certificate, or infrastructure completion report referencing scope and scale.

	in production for a minimum of twelve (12) months.			
Experience in Enterprise Software Licensing & Infrastructure Commissioning	Demonstrated experience in supplying, licensing, and commissioning enterprise software and infrastructure components (e.g., analytical database/data warehouse platforms, object storage, compute clusters) across large-scale on-premises analytics environments, with evidence of infrastructure scale comparable to the CDH scope	65	1 mark per project, minimum 2 projects	Architecture/commissioning diagrams, license procurement documentation, or client attestations confirming infrastructure scale and commissioning scope.

D. Company Profile (05 Marks)

Criteria	Description	Max Marks	Marking	Documents Required
Local Presence	Vendor shall demonstrate sustained local presence in Pakistan with on-ground offices and technical staff across infrastructure, hardware support, and software licensing functions, sufficient to support infrastructure commissioning, warranty service, and post-acceptance operational support.	5	0-5 marks based on adequacy, relevance, and depth of local technical capacity.	HR certificate or affidavit confirming technical staff strength with domain-wise distribution. Lease agreement, utility bills, or incorporation certificate evidencing operating presence in Pakistan.

18. Liquidated Damages

1. **Scope of Penalty Application:**

   - The Vendor must deliver the agreed hardware, software, and infrastructure scope within the specified timeline as outlined in Section 7.4 (Implementation Timelines and Milestones).
   - Failure to implement the system within the agreed timeframe will result in the following penalties.

2. **Liquidated Damages:**

   - **Rate:** A penalty of 0.02% of the total project value per day of non-compliance.
   - **Maximum Deduction:** The cumulative penalty shall not exceed the value of the performance guarantee submitted by the Vendor.

3. **Escalation of non-compliance:**

   - If the delay or failure extends beyond 40 days, PRAL reserves the right to:
     - Invoke the performance guarantee to recover damages and initiate reallocation of the project to alternate suppliers.

4. **Grace Period and Exceptions:**

   - A grace period of 15 working days may be provided under exceptional circumstances upon prior written approval from PRAL.
   - No penalties shall be applied in cases where the delay is caused by factors beyond the Vendor's control, provided the Vendor notifies PRAL in writing within 48 hours of identifying the issue.

5. **Enforcement Mechanism:**

   - The penalty will be calculated and deducted from the Vendor’s subsequent payments.

19. Vendor Coordination & Interface Management

This section defines the coordination, dependency management, and interface responsibilities associated with the Central Data Hub, which is being delivered through two separate engagements. This document constitutes the Data Platform & Application Implementation Services RFP.

- Infrastructure Acceptance (as defined in the Hardware, Software & Infrastructure Implementation RFP) is a formal prerequisite milestone before the Implementation Vendor's platform-build activities may commence on the corresponding environment.
- A joint site-readiness walkthrough shall be conducted between PRAL, the Infrastructure Vendor, and the Implementation Vendor prior to Infrastructure Acceptance sign-off.
- PRAL shall act as the single coordinating authority for schedule alignment between both Vendors; neither Vendor shall be entitled to claim relief for delay caused by the other party without PRAL's documented concurrence.
- Any Infrastructure Vendor-caused delay to Infrastructure Acceptance that demonstrably impacts the Implementation Vendor's committed timeline shall be assessed by PRAL and may result in a corresponding schedule adjustment for the Implementation Vendor, without prejudice to Liquidated Damages applicable to the Infrastructure Vendor under its own contract.
- Shared environment access, credential provisioning, and change control coordination between both Vendors shall follow a joint Change Management process, with PRAL as final approval authority for any change affecting both scopes.
- Both Vendors shall participate in a joint steering committee, or a joint standing item within each Vendor's own steering committee, to track cross-dependency risks and resolve interface issues.
- In the event both scopes are awarded to the same Vendor (single-vendor award), the provisions of this section shall be read as internal workstream coordination requirements rather than inter-Vendor contractual provisions.