KENYA REVENUE AUTHORITY

ISO 9001:2015 CERTIFIED

TERMS OF REFERENCE FOR DATA ANALYTICS CENTER OF EXCELLENCE

Tulipe Ushuru, Tujitegemee!

TABLE OF CONTENTS

DATA ANALYTICS CENTER OF EXCELLENCE

1 Executive Summary

The Kenya Revenue Authority (KRA) is implementing a strategic initiative to establish a Data Analytics Center of Excellence (CoE). This centralized hub will leverage advanced data analytics, Machine Learning (ML) and Artificial Intelligence (AI) to transform tax administration, significantly boosting operational efficiency and revenue growth. The primary objective is to harness data and leverage technology for real-time insights to drive tax administration, achieving self reliance for the country through adequate domestic revenue mobilization. The CoE aims to enable data-driven tax policy modelling, streamline revenue collection, improve tax compliance, combat tax fraud, and mainstream data analytics by developing technology enablers for data analytics and a highly skilled internal workforce augmented by industry experts to actualize data driven tax policies, revenue mobilization and tax administration locally and regionally.

The implementation of the Data Analytics CoE is structured into the following pillars:

i) A Unified Data Analytics & AI Platform that includes the design and implementation of a scalable, secure platform to support real-time and batch processing, AI/ML model developments, simulation and deployment, data governance and quality automation; with integrated tax intelligence, including unified fiscal data & process flow intelligence capabilities to integrate and standardize fiscal and compliance data, enable secure inter-agency and private sector data exchange where applicable, operationalize analytics embedded into business decisioning systems through configurable automation engines (Compliance, Case Management, Investigations & Enforcement, and finOps/embedded revenue assurance workflows, API actions, alerting/notifications and Robotic Process Automation) with end-to-end audit trails and real-time oversight dashboards; thus driving coordinated fiscal visibility, insights-driven policy changes, trade facilitation, internal operational efficiency and sustained revenue growth.

ii) Capacity Building focusing on developing and delivering structured training and mentorship for staff up to various certification levels.

iii) Augmentation of internal staff capacity with experts from various industry verticals and Data analytics and engineering experts in various domains for hands-on knowledge transfer; provision of strategic guidance, establishment of a robust governance and operational framework, and developing high-impact analytics solutions and use cases, including support to tax modelling and simulation *(e.g. tax modelling, sector-based risk profile modelling, digital assets transaction risk profiling, etc)*

The CoE will follow a phased, iterative implementation process over 18 months to accelerate its time-to-value. Success will be measured based on clear criteria, including the tangible Return on Investment (ROI) from analytics, sustained improvement in tax compliance and revenue collection and a general increase in data literacy across the organization.

1.1 Introduction

The Kenya Revenue Authority (KRA) plays a critical role in revenue collection, ensuring compliance with tax laws, facilitating trade and border control. To enhance its operational efficiency, revenue collection and adaptability to emerging business models and technologies, KRA is establishing a Data analytics Center of Excellence (CoE) as a strategic initiative anchored on continuous learning, best practices, professional development, and knowledge management in data analytics, and other emerging technologies, and developing highly professional tax and customs practitioners. In parallel, the CoE will act as the delivery and institutionalization vehicle for a unified fiscal data and process-flow intelligence capability that connects revenue and compliance data across collection, reconciliation, settlement, and reporting—translating analytics into action through automation and auditable end-to-end workflows across both domestic taxes and customs and border control.

The Data Analytics CoE will bridge identified gaps by leveraging real-time ecosystem integration, modernized and unified analytics platforms, and the use of machine learning, AI and blockchain to enhance revenue mobilization, real-time compliance management, personalized taxpayer experiences, detect and deter tax fraud, facilitate trade and use verifiable insights to inform policy changes; aligning KRA with global best practices. In addition to strengthening KRA's analytics capacity, the CoE will unify fiscal and compliance data across internal systems and external stakeholders, automate data governance and quality processes and provide near real-time fiscal visibility and auditable end-to-end process traceability. These strategic objectives aim to significantly modernize tax administration and enhance overall efficiency.

Key goals include **improved tax compliance**, targeting a substantial reduction in tax gaps through data-driven compliance management; **revenue growth**, bolstered by AI for integrated risk assessment and taxpayer profiling embedded in domestic taxes and customs & border control to increase collection efficiency; **faster fraud detection and prevention** via real-time monitoring and automated alerts to prevent revenue leakages; **informed tax policy decisions** through reliable fiscal intelligence; **personalized taxpayer experiences** powered by insights; and a **sustainable innovation culture** to continuously explore and integrate emerging technologies into tax administration processes.

1.2 Objectives of the CoE

The overarching aim of the CoE is to leverage advanced technologies and real-time data analytics to enhance tax compliance, optimize revenue collection, and foster a culture of continuous improvement within KRA. The following are the main objectives:

1. Utilize data analytics and emerging technologies to improve tax compliance, taxpayer experience and reduce tax fraud
2. Implement advanced technologies to enhance and streamline revenue collection processes
3. Institutionalize a unified fiscal data & process flow intelligence capability to integrate and standardize fiscal and compliance data across internal systems and external stakeholders, enabling secure interoperability and a single trusted view of revenue performance and customer compliance status.
4. Automate critical revenue process flows (validation, reconciliation, settlement, exception handling, and related compliance workflows) to reduce manual intervention, improve turnaround time, and strengthen auditability across the revenue lifecycle.

5. Promote institutional coordination and data sharing by enabling inter-agency fiscal visibility and collaboration frameworks where applicable for unified fiscal oversight and reporting.
6. Mainstream data analytics in KRA by developing and maintaining a highly skilled data analytics workforce, including capability transfer across data engineering, analytics, AI/ML, and automation disciplines.
7. Augment internal staff capacity with industry expertise in automating various aspects of advanced analytics, tax modelling and sector/industry specific risk profiling and compliance monitoring, while doing knowledge transfer in the process.
8. Foster a culture of innovation and continuous improvement within KRA to continuously modernize tax administration and integrate emerging technologies responsibly.

The CoE will enhance revenue collection and provide valuable benefits to KRA, including:

1. **Data Science and Analytics Products:** Develop various analytics use cases and models leveraging analytics and tax knowledge and deploy AI/ML models, conduct advanced analytics, and provide expertise in data mining and tax modeling.
2. **Data Visualization and Reporting:** Create interactive dashboards, visualizations, and reports to communicate data insights effectively to stakeholders.
3. **Data Strategy and Governance:** Enhance and maintain KRA's overall data strategy, data governance framework, and data quality standards.
4. **Data Architecture and Engineering:** Establish and operationalize a Unified data, analytics & AI architecture including data integration, data governance and quality process automation, data warehousing/lake housing and embedded Generative AI capabilities.
5. **Fiscal Data Integration & Interoperability:** Establish and operationalize a unified fiscal data framework, standardize critical fiscal/compliance data structures, and enable secure interoperability across internal KRA systems and external stakeholders, including inter-agency data exchange where applicable.
6. **Process Flow Automation, Revenue Enhancement & Real-Time Fiscal Visibility:** Embed intelligence into end-to-end revenue and compliance workflows (validation, reconciliation, settlement, exception handling, compliance case management, investigations/enforcement, and FinOps/revenue enhancement), delivering auditable traceability, automated alerts, and real-time oversight dashboards to support faster interventions and reduced leakages, leveraging existing revenue assurance capabilities.
7. **Training and Enablement:** Develop and deliver training programs to enhance data literacy, data skills, and the adoption of data-driven decision-making across KRA.

1.3 Key Initiatives

The CoE will prioritize the following initiatives:

- A Unified Data Analytics & AI Platform that includes the design and implementation of a scalable, secure platform to support real-time and batch processing, AI/ML model developments, simulation and deployment, data governance and quality automation; with integrated tax intelligence, including unified fiscal data & process flow intelligence capabilities to integrate and standardize fiscal and compliance data, enable secure inter-agency and private sector data exchange where applicable, operationalize analytics embedded into business decisioning systems through configurable automation engines (Compliance, Case Management, Investigations & Enforcement, and finOps/embedded revenue assurance workflows, API actions,

alerting/notifications and Robotic Process Automation) with end-to-end audit trails and real-time oversight dashboards; thus driving coordinated fiscal visibility, insights-driven policy changes, trade facilitation, internal operational efficiency and sustained revenue growth.

- Capacity Building focusing on developing and delivering structured training and mentorship for staff up to various certification levels.
- Augmentation of internal staff capacity with experts from various industry verticals and Data analytics and engineering experts in various domains for hands-on knowledge transfer; provision of strategic guidance, establishment of a robust governance and operational framework, and developing high-impact analytics solutions and use cases, including support to tax modelling and simulation *(e.g. tax modelling, sector-based risk profile modelling, digital assets transaction risk profiling, etc)*

2 Key Deliverables and Expectations

The Data Analytics Center of Excellence (CoE) is expected to enhance tax administration through data-driven decision-making, fraud detection, and revenue optimization. The following are the key deliverables:

i) **Deployment of a Unified Data Analytics and AI Platform**

a. A **multi-tier architecture** supporting both cloud-native and hybrid deployment models with auto-scaling capabilities to handle varying workloads. Enabling unified storage for structured, semi-structured, and unstructured data; Using **microservices-based design** and containerized deployment to allow independent scaling and deployment of platform components; Designed for **high availability and disaster recovery** with 99.99% uptime SLA, automated failovers, and geo-redundant backups. Supporting **real-time & batch processing;** with **event-driven architecture** enabling immediate response to critical business events and anomalies and **data pipeline orchestration** with automated scheduling, dependency management, and error handling. Supporting AI/**ML Model Development, Simulation & Deployment;**

b. Deployment of AI-driven models, and predictive analytics solutions to improve tax compliance, revenue forecasting and customer experience. AI-powered chatbots/assistants, and self-service analytics dashboards.

c. Automated Fraud Detection & Integrated Risk Profiling – AI-powered fraud detection models to identify tax evasion, underreporting, and suspicious transactions on the go across customs and domestic taxes.

d. Deployment of other functionalities outlined in the detailed scope.

e. Unified Fiscal Data & Process Automation – Standardize priority fiscal data and automate key revenue workflows for auditability and real-time visibility

f. Note: All platform deliverables shall be modular, interoperable, and capable of phased deployment and independent extension.

ii) **Data Governance & Quality Automation**
   a. **Data Governance Process Automation** with meta data management, data catalogue, policy enforcement capabilities, data access governance with approval workflows and regular review enforcement framework.
   b. **Data Quality Process Automation** with automated data profiling, rule-based and self-learning validation engine with configurable business rules, anomaly detection, real-time data quality score cards and dashboards to show data health by various dimensions; Automated data cleansing and enrichment pipelines with SLAs, data lineage tracking, data quality incident management with auto-alerting and assignment and resolution workflows as well as self-healing data pipelines with automated retry logic, error quarantine and reconciliation processes.

iii) **Capacity Building & Training Programs** – Development and delivery of trainings as outlined in the training and capacity building scope.

iv) **Experts Augmentation** – Provision of experts as outlined in the detailed expert augmentation scope to co-create various analytics products with KRA and perform hands-on knowledge transfer in the process.

3 Scope of Work

3.1 Summary

The implementation of the Data Analytics Centre of Excellence (CoE) involves a comprehensive scope of work designed to establish a modern, agile, and data-driven environment that enhances tax compliance, revenue growth and customer experience through advanced analytics and AI. The implementation of the Data Analytics Center of Excellence (CoE) is structured into four key areas, with the Data Analytics Platform pillar incorporating a unified fiscal data & process flow intelligence capability to standardize fiscal and compliance data across priority systems, automate critical revenue workflows, and enable real-time fiscal visibility and auditability.

a) Data Analytics Platform

The initiative will focus on designing and implementing a scalable, secure, and high-performance infrastructure capable of supporting both real-time and batch data processing. The platform is anchored by the establishment of a centralized data lakehouse architecture that integrates diverse structured and unstructured data sources, utilizing hybrid computing resources (cloud and on-premise) to optimize analytics workloads. Core functionalities include capability to deploy advanced AI/ML models and analytics capabilities for pattern identification, fraud detection, and compliance monitoring, ensuring seamless integration with existing internal systems and external data sources for comprehensive analysis. The platform will incorporate action engines, and robotic process automation to automate the delivery of analytics outputs, supported by extensive training and knowledge transfer for relevant personnel on all platform components. The platform will also provide unified fiscal data integration, automated revenue workflow controls (validation, reconciliation, settlement, and exception handling), end-to-end audit trails, and role-based real-time monitoring dashboards for operational and oversight users

b) Experts Augmentation Services

The expert augmentation services will focus on practical application and operational agility which includes defining and developing key, high-impact analytics use cases specifically designed to enhance revenue collection, optimize tax compliance and deliver policy changes informed by verifiable

insights. This will entail provision of technical experts—including data engineers, data scientists, and cloud architects and various tax modelling and vertical industry experts to augment internal delivery capacity. Hands-on delivery and knowledge transfer will be used, supplemented by collaborations with academia and industry partners, to ensure continuous learning and enhance decision-making and operational efficiency across all departments. To drive successful adoption and ensure the sustainability of these initiatives, a focused approach to change management and stakeholder engagement will be conducted. Finally, to optimize efficiency and responsiveness in delivering data analytics products and services, agile ways of working will be established for both delivery and operation.

c) Training and Capacity Building

The CoE will develop and foster a robust data-driven culture and ensure self-sufficiency in analytics operations. This involves development and delivery of structured training and mentorship programs to upskill existing staff in various skillsets around data and analytics, AI, tax modelling and other related areas as outlined in the detailed training scope.

4 SECTION 1: EXPERT AUGMENTATION FOR THE DATA ANALYTICS CENTER OF EXCELLENCE

4.1 Introduction

KRA seeks the services of a highly qualified firm/ consortium to enhance internal capacity by augmenting the internal team with high-quality technical expertise, such as data scientists, big data engineers and business intelligence experts. Augmenting the analytics team with specialized talent focused on Key Result Areas will present KRA with a strategic opportunity to enhance internal data analytics capabilities and drive business performance. Augmented experts will be expected to work hands-on with KRA teams to operationalize analytics into day-to-day revenue and compliance processes, supporting integration, automation, and institutional capability transfer. Bidders should provide detailed Curriculum Vitae (CVs) for proposed specialists. The CV shall detail:

- Education background
- Certifications
- Area of specialization
- Years of experience
- Brief on previous projects and institutions

4.2 Objectives of Expert Augmentation Services

The overall objective of this service is to **strengthen KRA's internal data analytics capacity** by augmenting existing teams with specialized technical expertise to support data-driven policy changes, tax compliance management, decision-making and improved business performance. The consultancy firm shall;

- Augment KRA's analytics function with highly skilled data scientists, tax modelling and vertical industry experts, big data engineers, and business intelligence experts to complement and strengthen existing internal capacity.
- Facilitate structured knowledge transfer to KRA staff through mentoring, coaching, and documentation to ensure sustainability of analytics capabilities beyond the expert augmentation period.
- Support the adoption of industry best practices, standards, and governance frameworks in data analytics, engineering, and visualization.
- Reduce time-to-insight by accelerating the development and deployment of analytics solutions using modern tools, methodologies, and agile practices.
- Provide analytical insights that inform policy formulation, risk profiling, operational planning, performance monitoring, and executive decision-making across KRA.

Scope of Work
The firm shall provide analytics experts' augmentation services to support and strengthen KRA's internal analytics capacity. The key areas of focus shall include:

a) Provision of Specialized Analytics Expertise

- Provision of qualified personnel, including data engineers, data scientists, tax modelling and vertical industry experts, AI/ML specialists, data governance experts, and cloud engineers to strengthen the existing workforce and drive advanced data analytics initiatives
- Design and implement a continuous training and upskilling program to enhance technical expertise in analytics, machine learning, artificial intelligence and related emerging technologies.

b) Embedded technical support

Integrate the proposed specialists within KRA teams to provide hands-on technical support in data engineering, advanced analytics, tax modelling and business intelligence initiatives *(e.g. the creation of tax models, sector-based risk profiles, digital assets transactions risk profiles, etc.)*

c) Data Engineering and Platform Support

Build, optimize, and maintain data pipelines, data models, and analytics platforms to support scalable and reliable data processing and analysis.

d) Knowledge Transfer and Capacity Building

- Facilitate on-the-job training, mentoring, documentation, and skills transfer to ensure long-term sustainability of analytics capabilities within KRA.
- Conduct a skills gap analysis to assess KRA's training needs.
- Develop a training curriculum covering data analytics, AI, cybersecurity, digital assets and data governance.
- Establish an e-learning platform for continuous self-paced professional development with all the training outlined in Section 3 under the Staff Training Focus Area."

e) Performance Monitoring and Reporting

Provide regular progress updates, activity reports, and performance metrics aligned to agreed deliverables and timelines.

4.3 Expert Roles & Qualifications

Bidders are required to submit verifiable evidence of previous projects, including **letters of award, certificates of completion, reference letters, and documented case studies**. The evaluation criteria will prioritize firms that have successfully executed large-scale analytics initiatives in the public sector, particularly in tax administration and revenue collection.

The following table outlines the specific requirements and evaluation criteria for the consultant firm:

- Demonstrated technical expertise and experience in analytics-driven transformation initiatives.

- Relevant educational background and industry-recognized professional certifications.
- Proven track record of successful project implementation with measurable and verifiable outcomes.
- Alignment with recognized best practices in data governance, data security, and regulatory compliance.

Evaluation Criteria	Qualifications	Weight	Bidders Response
Consultant Firm Experience	The firm shall demonstrate: • A clear portfolio of similar data analytics projects awarded and successfully undertaken within the last ten (10) years, showing relevance to large-scale analytics, AI, or enterprise data platforms. (5 Marks) • Documented proof of satisfactory execution and completion of comparable projects, including signed project completion reports, acceptance letters, or implementation sign-offs. (5 Marks) • Verifiable client references, including Local Service Orders (LSOs) or formal reference letters directly linked to the cited projects. (5 Marks) • Certificates of completion issued by previous clients for the referenced projects, confirming successful delivery and contractual closure.	10	
Three (3) Data Scientists	The CVs shall detail: • At least 3 years' experience implementing enterprise analytics solutions within tax administration or financial services. (2 Marks) • Demonstrated specialization in advanced analytics, including predictive modelling, statistical analysis, anomaly detection, segmentation, forecasting, and complex data modelling, tax modelling. (2 Marks) • Relevant Data Science certification, such as in Machine	5	

Evaluation Criteria	Qualifications	Weight	Bidders Response
	Learning, AI, Advanced Analytics, or equivalent. (1 Mark)		
Two (2) Business Intelligence (BI) Developers/Analysts:	The CVs shall detail: • At least 3 years' experience designing and delivering enterprise BI solutions in tax administration, government, or financial services. (2 Marks) • Proven expertise in data querying, modelling, and visualization, including advanced SQL, batch and streaming data processing, ETL/ELT processes, dimensional modelling, and dashboard development using tools such as Power BI, Tableau, Apache Superset, or equivalent. (2 Marks) • Relevant BI/Data Analytics certification (e.g., Microsoft Power BI, Tableau, Data Analytics, or equivalent). (1 Mark)	5	
Two (2) Machine Learning Engineer/ MLOps Specialists	The CVs shall detail: • At least 3 years' experience designing, deploying, and operationalizing machine learning models in enterprise environments, preferably in tax administration, government, or financial services. (2 Marks) • Proven capability in ML engineering and MLOps, including automated model pipelines, CI/CD for ML, containerization, orchestration, model performance monitoring, and optimization. (2 Marks) • Relevant ML or MLOps certification, such as cloud ML engineering, DevOps, or equivalent. (1 Mark)	5	
One (1) Cybersecurity Expert	The CV shall detail: • Demonstrated expertise in big data analytics, artificial intelligence, and machine learning applications within cybersecurity contexts. (2 Marks) • At least one cybersecurity strategy successfully developed and implemented	5	

Evaluation Criteria	Qualifications	Weight	Bidders Response
	in an enterprise environment, preferably in government or financial services. (2 Marks) • Evidence of compliance with data security and protection standards, including adherence to Data Protection Laws and proven experience implementing security best practices. (1 Mark)		
One (1) Change management and capacity building Expert	The CV shall detail: • Proven experience in leading change initiatives within large organizations or government institutions, with at least two change management and training programs successfully implemented. (2 Marks) • Documented training frameworks and delivery capability, including training plans, methodologies, and participant impact assessments demonstrating measurable effectiveness. (2 Marks) • Verifiable client references, including reference letters or testimonials from beneficiary organizations of the training and change programs. (1 Mark)	5	
One (1) Data Governance Analyst	The CV shall detail: • At least 5 years' experience implementing enterprise data governance frameworks, including metadata management, data lineage, access control workflows, and policy enforcement. (2 Marks) • Demonstrated hands-on experience with data governance platforms, such as Collibra, Alation, Informatica, Microsoft Purview, or equivalent. (2 Marks) • Working knowledge of the Data Protection Act and ethical AI standards, with evidence of practical application in regulated environments. (1 Mark)	5	

Evaluation Criteria	Qualifications	Weight	Bidders Response
Two (2) Data Quality Engineers	The CVs shall detail: • At least 5 years' experience supporting enterprise data quality programs within tax administration, financial services, or other regulated environments. (2 Marks) • Demonstrated hands-on expertise in data quality operations, including data profiling, validation rule design, anomaly detection, reconciliation workflows, and end-to-end data quality monitoring. (2 Marks) • Practical experience with recognized data quality tools, such as Great Expectations, Informatica Data Quality, Talend Data Quality, or equivalent. (1 Mark)	5	
Functional Domain Experts – One (1) for each domain	The CVs shall detail: • At least 7 years' demonstrable experience in sector-specific analytics, compliance, operational processes, and risk intelligence within tax administration, financial services, regulatory agencies, or similarly complex environments. (2 Marks) • Proven domain expertise in at least one priority sector, with direct application to data-driven decision-making and compliance analytics in any of the following: Mining; Cross-Border Transactions and Customs Supply Chains; Telecommunications; Digital Economy and Platform Ecosystems; Manufacturing; MSMEs; Petroleum and Energy Value Chains; NGOs; Blockchain and Crypto-Asset Ecosystems. (2 Marks) • Evidence of impact in applied domain analytics, such as sector-specific projects, policy support, risk models, compliance frameworks, or reference letters from client institutions. (1 Mark)	5	
Three (3) Tax Modelling Experts	The CVs shall detail: • At least 7 years' demonstrable experience in tax modelling within tax	5	

Evaluation Criteria	Qualifications	Weight	Bidders Response
	administration, government (national or county), financial services, regulatory agencies, or similarly complex environments (3 Marks). • Evidence of actual impact achieved by applying tax models developed in real world. (2 Marks)		
Two (2) AI Specialists (LLMs)	The CV shall detail: • Demonstrated experience building, fine-tuning, and deploying large language models, including RAG architectures and enterprise-grade NLP pipelines. (2 Marks) • Proven capability in designing and delivering AI Intelligent Assistants for use cases such as audit automation, taxpayer support, knowledge retrieval, or policy research. (2 Marks) • Relevant AI or LLM-focused certification, or documented advanced training in machine learning, NLP, or generative AI. (1 Mark)	5	
Two (2) Knowledge Graph Engineers	The CVs shall detail: • At least 3 years' experience building and deploying enterprise knowledge graphs using platforms such as Neo4j, TigerGraph, or equivalent. (2 Marks) • Demonstrated expertise in graph data engineering, including schema design, ontology development, entity resolution, and complex relationship modelling to support fraud networks and taxpayer intelligence. (2 Marks) • Relevant certification or advanced training in graph databases, semantic technologies, or applied data engineering. (1 Mark)	5	
Two (2) Vector Database / Semantic Search Engineer	The CV shall detail: • At least 3 years' experience designing and implementing enterprise-grade vector search systems using platforms	5	

Evaluation Criteria	Qualifications	Weight	Bidders Response
	such as Pinecone, Weaviate, Milvus, or equivalent. (2 Marks) • Demonstrated expertise in semantic search engineering, including embeddings generation, similarity search, retrieval-augmented generation (RAG), and large-scale semantic data indexing to support AI-driven tax intelligence. (2 Marks) • Relevant certification or advanced training in AI search, NLP, vector databases, or applied machine learning. (1 Mark)		
Two (2) Spatial Intelligence Analyst / GIS Specialist	The CV shall detail: • At least 3 years' experience in GIS, remote sensing, or spatial analytics within government, regulatory, security, or financial services environments. (2 Marks) • Demonstrated expertise in applied spatial intelligence, including geospatial risk mapping, hotspot detection, movement analysis, and spatial-economic modelling relevant to tax, customs, and excise operations. (2 Marks) • Relevant GIS or spatial analytics certification, such as ESRI, QGIS, Remote Sensing, or equivalent professional training. (1 Mark)	5	
One (1) Analytics Project/Product Manager	The CV shall detail: • At least 3–5 years' experience managing analytics, AI, or technology projects within tax administration, financial services, or similarly complex institutions. (2 Marks) • Demonstrated capability in full project lifecycle leadership, including defining goals, scope, timelines, deliverables, risk management, and coordinating cross-functional technical and business teams. (2 Marks)	5	

Evaluation Criteria	Qualifications	Weight	Bidders Response
	• Working knowledge of AI concepts and data analytics methodologies, evidenced through training or applied project work. (1 Mark)		
Two (2) Data Architects	The CVs shall detail: • At least 3 years' experience designing enterprise data architectures, including data models, integration frameworks, and scalable storage solutions for analytics environments within tax administration or financial services. (2 Marks) • Demonstrated expertise in architectural design and governance, including development of documentation, standards, and reference architectures to guide enterprise analytics initiatives. (2 Marks) • Relevant certification or advanced training in data architecture, cloud architecture, or enterprise data management. (1 Mark)	5	
Two (2) Statisticians	The CVs shall detail: • At least 4 years' experience applying statistical techniques in analytics projects within tax administration, financial services, or other regulated environments. (2 Marks) • Demonstrated capability in statistical modelling and experimental design, including hypothesis testing, sampling design, model validation, and ensuring methodological rigor for analytical and predictive work. (2 Marks) • Relevant qualification or certification in Statistics, Applied Mathematics, Econometrics, or a closely related field. (1 Mark)	5	
Two (2) Big Data Engineers	The CVs shall detail: • At least 3 years' experience designing, developing, and maintaining big data architectures within tax administration,	5	

Evaluation Criteria	Qualifications	Weight	Bidders Response
	financial services, or other large-scale data environments. (2 Marks) • Demonstrated proficiency in building scalable data pipelines and managing distributed data platforms, including streaming, batch processing, and cluster-based environments. (2 Marks) • Relevant certification or advanced training in big data engineering, cloud data platforms, or distributed systems. (1 Mark)		
		100	
	Cut Off Score is	80	

5 SECTION 2: UNIFIED DATA ANALYTICS & AI PLATFORM

5.1 Introduction

Kenya Revenue Authority (KRA) seeks to implement a Unified Data Analytics & AI Platform to enhance decision-making, Tax Intelligence, fraud detection, compliance monitoring, and operational efficiency. The platform will support real-time and batch data processing, advanced analytics, and AI-driven insights. This document outlines the terms of reference (TOR) for the development, implementation, and maintenance of the Unified Data Analytics & AI Platform. The platform will support unified fiscal and compliance data integration, automation of critical revenue and compliance workflows, and role-based real-time operational visibility with end-to-end audit trails.

5.2 Objectives

1. Establish a comprehensive data analytics platform to support business intelligence and decision-making.
2. Implement capabilities to enhance tax intelligence, fraud detection, and compliance monitoring through AI and ML-based analytics.
3. Improve operational efficiency through automated reconciliation, rule-based processing and self-learning, exception handling, and real-time operational monitoring with audit trails.
4. Provide evidence-based analytics and simulations on the impact of tax policies and economic trends.
5. Provide a scalable, secure, and integrated platform for data storage, retrieval, and processing.
6. Integrate and standardize fiscal and compliance data across priority internal systems and approved external data sources.
7. Automate critical revenue and compliance workflows, including validation, reconciliation, settlement processing, and exception resolution.
8. Enable self-service analytics for business users

5.3 Scope of work

The Unified Data Analytics & AI Platform comprises five interlinked layers which work together to deliver high-quality insights, strengthen tax intelligence, improve compliance monitoring, and enhance operational efficiency—while also enabling automated reconciliation and exception handling workflows, end-to-end auditability, and role-based real-time operational visibility. The layers must support the full operational loop from data ingestion and processing through decisioning, workflow execution, case handling, and monitoring dashboards.

a) Analytical Layer – Insights & Decision Support

This user-facing layer provides advanced insights through interactive dashboards, detailed reports, and robust self-service analytics. Key capabilities include:

- **Dynamic Dashboards & Reports:** Visualizations for monitoring Key Performance Indicators (KPIs), analyzing revenue performance, and deriving actionable customer insights.
- **Exploratory Data Analysis (EDA):** Tools to identify emerging trends, patterns, and anomalies.
- **AI-Driven Insights:** Embedded functionalities for fraud detection and Tax Intelligence taking advantage of any applicable existing capabilities.

- **Taxpayer Profiling:** A self-enhancing integrated view of taxpayers by integrating data from multiple internal and external sources, leveraging the existing capabilities for 360 degree view of taxpayers.
- **Case Management**: Integrated system to streamline compliance monitoring, exception handling, investigations tracking, and resolution workflows with full traceability leveraging existing case management capabilities.

b) Application Layer – Operational Analytics

This layer focuses on enhancing tax intelligence, fraud detection, and contract compliance through sophisticated analytical tools. Key applications include:

- **AI/ML Models:** To reconcile services, electronic slips and payments, preventing revenue leakage.
- **Fraud Management Tools:** fraud detection capabilities and models enabled by real-time and batch analysis to detect suspicious activities.
- **Contract Analytics:** Natural Language Processing (NLP) to monitor compliance and mitigate financial risks.

c) Data Processing Components – Data Integration & AI Automation

This layer ensures data consistency and reliability through automated processing. Core components include:

- **Reconciliation Engine:** Validates data across multiple sources and implements data quality process automation.
- **Exception Handling & Escalation**: Flags mismatches, routes exceptions for resolution, and tracks closure with audit trails.
- **Rule Engine:** Automates decision-making based on business logic. Proposes new rules / rule enhancements through self-learning.
- **KPI Monitor/Evaluator:** Tracks key performance metrics.
- **ML Core Engine:** Manages machine learning workflows.
- **Aggregators:** Consolidate data for unified analytics.
- **EDA Modules:** Facilitate in-depth statistical analysis and visualization.

d) Data Storage Layer – Storage & Querying

This foundational layer provides a structured storage system using a data Lakehouse architecture. It supports:

- **Real-time and Batch Queries:** Seamless access to diverse data sources.
- **Metadata Management:** Ensures automated data governance.
- **Analytical Database & Big Data Query Engine:** Enable complex reporting and insights generation.
- **System Configuration & Automation:** Maintain security and regulatory compliance.

e) Data Integration Layer – Integration & Ingestion

This layer ensures seamless data collection and integration from multiple sources. Key technologies include:

- **Transformation**: Standardized fiscal and compliance data structures and interoperability across connected systems
- **Streaming:** For real-time event streaming.
- **Batch ingestion:** For batch processing.

- **Logs & metrics parser:** For collecting and analyzing system metrics and logs.
- **Optical Character Recognition (OCR):** To extract and process text from scanned documents.

NB: Bidders who attain score of 80% and above in the technical evaluation will be required to give a demo of the platform to showcase the functionality

5.4 Data Analytics Platform Layers Specifications

Instructions to Bidders:

1. Bidders MUST complete the Table below in the format provided.
2. Bids MUST meet all mandatory (MUST) requirements in the Tables below in order to be considered for further evaluation.
3. Bidders MUST provide a substantial response or clear commitment to meeting the requirements for all features irrespective of any attached technical documents in the table format (bidders Response) below. Use of Yes, No, tick, compliant, blank spaces etc. will be considered non-responsive.
4. Bidders who do not comply with any of the below requirements will NOT be considered for further evaluation.

Specification	Goals/ Responsibilities	Bidders Response
a) Analytics Layer (User-Facing Insights & Decision Support) This layer is responsible for data visualization, business intelligence, and analytics to support decision-making across KRA.		
1. Dashboards & Reports	The platform must include a modern, self-service Business Intelligence (BI) tool for creating interactive dashboards and reports. • **Visualization:** Must offer a wide range of types (maps, charts, tables, heatmaps) and support for geographic data. • **Interactivity:** Dashboards must allow users to drill-down, filter, and explore data dynamically. • **Self-Service:** Business users can explore data, create own reports, and derive insights with basic technical expertise. • **Customization:** Reports must offer customization options to tailor analytics to specific needs. • **Distribution:** Must support scheduled report generation and distribution via email, subscription, or export to formats like PDF and Excel including exception-based alerts and scheduled operational packs for oversight users.	

Specification	Goals/ Responsibilities	Bidders Response
2. Key Performance Indicators (KPIs) Viewer/Monitor	The customizable KPI Viewer must monitor and measure performance against KRA's Corporate Plan at strategic, functional, and individual levels with ability to amend or define new KPIs on demand. • **Real-time Streaming:** Must support real-time data streaming for all applicable critical KPIs.	
3. Business Widgets	The platform must provide User Interfacce components that offer real-time insights and operational controls. • **Revenue Performance Widget:** Shows revenue performance, growth, and performance against targets. • **Customer Insights Widget:** Displays taxpayer behaviour, churn prediction, and engagement analytics**.** • **User-Centric Design**: The platform will cater to diverse users (Tax Data Analyst, Data Scientist, Tax Auditor, Management, Platform Administrator). • **Reconciliation & Exceptions Monitoring (Operational Widget):** Must show outstanding mismatches, exception ageing, SLA status, and closure rate; allow drill-down to transaction/case level; support export to PDF/Excel • **Case & Workflow SLA Dashboard:** Must provide case volumes, assignment queues, time-to-resolution, SLA breaches, escalations, and closure outcomes by unit/region/officer; support filters and drill-down. • **Audit Trail & Traceability Viewer:** Must provide end-to-end traceability from data ingestion through reconciliation/decisioning to case actions and closure; must export immutable logs for audit review.	
4. Exploratory Data Analysis	The platform must provide tools for EDA to analyze and visualize raw data, uncovering patterns, detecting anomalies, and identifying relationships between KRA and third-party datasets.	
5. Enhanced Integrated Taxpayer View	KRA is developing a 360 degree view supported by open source technologies. The platform shall leverage the existing360 degree view if taxpayer to	

Specification	Goals/ Responsibilities	Bidders Response
	provide a self-enhancing Integrated Taxpayer view that consolidates information across internal and 3rd party sources. This holistic view will be embedded in investigations/Audit/Compliance/Debt and customer support processes to enhance customer experience, improving revenue collection, and optimizing decision-making. The view must include all datasets known about the taxpayer, included forecasts.	
6. Multi-pattern recognition	The platform must include multi-pattern recognition capabilities to detect patterns, anomalies, and trends across large datasets. This is crucial for Tax Intelligence, Fraud Management, Data Quality Assurance (DQA), and Reconciliation. Detected anomalies must be configurable to trigger alerts and/or create investigation/compliance cases	
7. Audit Checks	The platform must facilitate systematic reviews (Audit Checks) of processes, systems, and records to ensure compliance with policies and regulations. It must help identify risks, inefficiencies, and areas for improvement (e.g., Compliance, Operational, and Quality Audit Checks).	
8. Investigation Cart	The platform must provide an Investigation Cart: a digital toolkit for auditors and investigators to efficiently collect, analyze, and document evidence.	
9. Case & Incident Management	The platform must include a Case & Incident Management system to provide a structured approach for identifying, tracking, investigating, and resolving issues related to Tax Intelligence, Fraud Management, and DQA. It must ensure efficient resolution through automation, AI-driven workflows, and real-time analytics. It must support end-to-end traceability from detection through assignment, action, and closure. **Note**: Orchestration of cross departmental/enterprise-wide workflows will use an enterprise workflow management solution supporting integration through APIs.	
10. Data Products & Data Science Environment	Data assets or tools developed to provide value to users through data. The platform must support the creation and management of Data Products that transform raw data into valuable insights.	

Specification	Goals/ Responsibilities	Bidders Response
	• **Data Science IDE:** Must provide an integrated development environment (IDE), such as Jupyter Notebooks, with managed compute resources. • **Language Support:** Must support Python and R. • **Libraries:** Must provide pre-installed and managed libraries for ML (e.g., Scikit-learn, TensorFlow) and statistical analysis. • **Graph Analytics:** Must support graph analytics to identify complex networks and relationships between taxpayers and transactions. • **Simulation Engine:** Must support ability to perform simulation and what-if analysis of analytics products e.g. models to establish expected results before rollout.	
11. Business Process Management/ Robotic Process Automation	The platform must incorporate BPM and RPA to optimize workflows and improve efficiency. BPM/RPA must support automation of revenue and compliance workflows including validation, attribution, settlement processing, and exception handling. **Business Process Management (BPM):** • **Process Designer:** Must provide a visual, drag-and-drop designer based on industry standards, supporting all key BPM elements (events, tasks, gateways). • **Business Rules Engine (BRE):** Must allow business users to define, manage, and execute complex rules. • **Process Simulation:** Must support simulation to model the impact of changes on process metrics before deployment. • **Work Queue:** Must provide a unified task list interface for users to view, claim, and complete assigned tasks. • **SLA Management:** Must manage process escalations and Service Level Agreements (SLAs). **Robotic Process Automation (RPA):** • **Bot Types:** Must support both attended (desktop-assisted) and unattended (background) bots.	

Specification	Goals/ Responsibilities	Bidders Response
	• **Application Interaction:** Bots must interact with web apps, desktop apps, legacy systems, and APIs. • **Bot Controller:** Must provide a centralized controller for deploying, scheduling, and monitoring the bot fleet. • **Security:** Must include a secure credential vault for storing bot login credentials.	

b) Applications Layer

This layer encompasses business-specific applications that leverage analytics and AI to support Tax Intelligence, fraud detection, financial analysis, and the operationalization of insights into day-to-day revenue and compliance processes. Applications in this layer must support configurable business rules, workflow initiation, and integration with case management and automation components to ensure insights translate into measurable action.

1. Enhanced Tax Intelligence System

A comprehensive module that identifies non-filers, late filers, under-reporting, incorrect tax positions, sudden revenue drops, high-risk sectors, and ensures actual revenue matches potential revenue among others. The vendor shall evaluate existing capabilities for re-use as applicable to avoid any duplications.

Specification	Detailed Requirements	Bidders Response
1.1 Non-Filer Detection	System identifies non-filers, nil-filers, late filers using ML and data cross-matching (banks, customs, telecom, POS, e-invoicing). Must track trends over time and automatically flag risk.	
1.2 Under-Reporting Detection	Detects under-declaration of income, VAT, turnover using sector-based peer benchmarking, industry margins, machine learning regression deviation models, POS vs filed turnover comparison, and sector norms.	
1.3 Revenue Leakage Early Warning System	Real-time detection of sudden drops in revenue at taxpayer, sector, or national level (daily/weekly). Must show contributing taxpayers and root-cause signals.	
1.4 Tax payer sectorization	• The system shall provide advanced taxpayer segmentation capabilities that: • Automatically group taxpayers based on behavioural, transactional, demographic, sectoral, and compliance characteristics. • Support dynamic, multi-dimensional segmentation.	

	• Refresh segments continuously as new data becomes available. • Support segmentation for compliance interventions, audit prioritisation, service personalization, and early warning detection. • Integrate segmentation outputs with the Tax Intelligence module to prioritise taxpayers with high leakage risk. • Leverage AI/ML to classify tax payers into sectors based on their business activities/transactions using data from 3rd party sources and e-Invoicing transaction records.	
1.5 Sector-Based Compliance Scoring	Sector-level behaviour scoring for manufacturing, retail, wholesale, finance, hospitality, professional services, etc. Includes industry-specific anomalies.	
1.6 ML/AI Modelling for Fraud Detection	The system shall provide AI-driven tax evasion modelling capabilities that: • Identify high-probability evasion scenarios using ML, graph analytics, anomaly detection, and rule-based models. • Detect patterns including under-reporting, non-filing, missing trader fraud, smuggling, carousel fraud, identity manipulation, and leakage. • Generate predictive scores for taxpayers, transactions, sectors, and supply-chain nodes. • Support tax gap estimation models and real-time evasion indicators. • Provide evidence packs for audit, enforcement, and legal processing. • Embed risk modelling into the Tax Intelligence module to score taxpayers and transactions for potential leakage or non-payment. • Model filing habits, chronic late filers, payment delays, and compliance fatigue. Predicts future compliance risks. • AI-driven tax preparation software and Optical Character Recognition (OCR).	

	• AI forecasting models and scenario modelling. • Enhanced graph analysis to detect patterns and uncover hidden insights.	
1.7 Payment Compliance Risk Modelling	• Predicts likelihood of delayed/non-payment (PRWPs), partial payments, and intentional avoidance using financial and transactional data. • Validate all generated e-slips against actual payments and flag unmatched items for Tax Intelligence review.	
1.8 Turnover Mismatch & POS Reconciliation	POS/e-invoice/sales receipt totals reconciled with tax filings. Must detect suppression, split transactions, ghost transactions.	
1.9 Bank Transaction vs Filing Mismatch	Detect mismatch between declared income and bank inflows/outflows using banking, mobile money, fintech data.	
1.10 Customs & Trade vs Domestic Filing Mismatch	Compare import/export data with VAT & income declarations; detect smuggling signals, undervalued imports, missing traders.	
1.11 Mobile Money & Fintech Tax Intelligence	Detects hidden revenue using mobile wallet flows, merchant payments, agent networks, mobile remittances.	
1.12 E-Commerce Seller Oversight	Detects unregistered online merchants, under-declaration by marketplace sellers, platform-level turnover analytics.	
1.13 Refund Under-Declaration Risk Checks	Detects suspicious refund claims that reduce revenue (false input claims, inflated purchases) by matching VAT transactions, import/exports declarations and returns filing data.	
1.14 Revenue Projection & Forecasting Models	The system shall provide forecasting capabilities that: • Generate short-, medium-, and long-term revenue forecasts using statistical and ML models. • Support scenario modelling for policy, sectoral shocks, behavioural changes, and macroeconomic factors.	

	• Provide granular forecasts by taxpayer segment, station, sector, tax head, or transaction type. • Supply predictive indicators for Tax Intelligence operations. • Integrate with KPI dashboards. • Provide predictive indicators for expected vs actual revenue to identify emerging gaps. • Forecasts tax revenue using advanced ML (time-series, gradient boosting). Includes scenario modelling & sensitivity analysis.	
1.15 Regional/Geo-Spatial Revenue Insights	Geo-spatial intelligence on regional performance: identify weak performing regions, compliance hotspots.	
1.16 High-Risk Taxpayer Heatmaps	Automated heatmaps for high-risk taxpayers by sector, region, tax type.	
1.17 Automated Alerts & Escalation	Alerts for non-filing, under-reporting, anomalies, and potential evasion with configurable thresholds.	
1.18 Integration with Case Management	Automatic creation of cases based on risk scores and anomaly detections to streamline follow-up actions.	
1.19 Behavioural Risk Scoring – Risk Analytics	Identifies taxpayers whose behaviour shows early signs of non-compliance (irregular filings, sudden changes).	
1.20 Explainability	System shows why a taxpayer was flagged (features, comparators, reasons). Essential for audit and legal defensibility.	
1.21 Simulation Engine for Compliance Impact, Revenue collection and Tax Gap reduction	Ability to simulate enforcement actions and their impact on revenue (e.g., "What if we run 1,000 audits in sector X?").	
1.22 Historic Trends & Seasonality Analysis	Automatically identifies seasonal filing/payment patterns and historic compliance trends.	
1.23 Role-Based Views for Tax Intelligence Officers	Different dashboards for Commissioner, Directors, Regional Managers, Auditors, Field Officers.	
1.24 Performance KPIs for Assurance Teams	KPIs: recovered revenue, compliance improvement, case aging, ROI of interventions.	

1.25 Provide a feedback management solution	The system should provide a feedback functionality for cases generated from the Tax Intelligence application to the business case management system.	
1.26 Automated Notice Generation	Auto-generate reminders, notices, and follow-up communications for detected non-compliance.	
1.27 Sentiment Analytics	The system shall provide sentiment analytics capabilities that: • Analyse taxpayer sentiment across social media, emails, call logs, disputes, and interactions. • Detect negative sentiment trends that may indicate compliance risks or misinformation. • Integrate sentiment scores into the Risk Analytics. • Provide early warning for sectors or regions under compliance pressure. • Support targeted engagement strategies.	
G(AI) Services KRA seeks to implement (GenAI services) a suite of artificial intelligence capabilities that leverage Large Language Models (LLMs), Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), and vector-based semantic search. This will enhance how KRA's stakeholders interpret laws, process information, as well as how KRA assists taxpayers and internal staff, to improve compliance, enforcement, and operational efficiency. These services combine conversational intelligence, document understanding, predictive capabilities, and domain adaptation to deliver more accurate, faster, and context-aware insights. Below are the required capabilities for G(AI) services:		
2.1 Large Language Model (LLM) Support	• The platform must support deployment and use of enterprise-grade LLMs, including open-source, proprietary, and locally hosted sovereign models. • It should allow fine-tuning with taxation-specific datasets (laws, rulings, taxpayer interactions, audit reports, compliance guidelines). • It should offer the ability to run LLMs on-prem or in a sovereign cloud environment to ensure data privacy, compliance, and confidentiality. • It must support model optimization techniques such as	

	quantization, distillation, and GPU acceleration. • It must host multiple LLMs simultaneously for different tax use cases (audit, risk scoring, taxpayer support).	
2.2 Retrieval-Augmented Generation (RAG)	• The platform must support RAG frameworks to enhance LLM responses using information from tax laws, rulings, circulars, court judgments, and KRA internal policy documents. • The platform must embed documents using vector databases and semantic indexing for accurate retrieval of tax-domain knowledge. • It should support frequent re-indexing and version control for changes in tax laws. • Must provide APIs to integrate RAG outputs into internal systems such as audit tools, domestic tax systems, Customs systems, e-invoicing system, case management tools, analytics dashboards, CRM, and taxpayer portals.	
2.3 Vector Database Capabilities	The platform must include a vector database capable of storing millions of high-dimensional embeddings, supporting search, semantic retrieval, similarity search, and classification.	
2.4 Long-Context Document Processing	• The platform must enable processing of long documents such as the taxpayer Act, regulations, tribunal rulings, audit manuals, and legal interpretations. • It must be able to summarize, interpret, compare, and extract key insights from lengthy documents. • Should support multi-document reasoning (cross-referencing laws, finance bills, and practice notes). • Must flag inconsistencies, outdated laws, or contradictions in policy texts.	
2.5 NLP-Driven Document Q&A	• The system must provide an interface for users to ask natural-language questions and receive accurate responses grounded in tax-domain content like (FAQs, guides, rulings, laws and procedures).	

	• Ability to generate explanations, breakdowns, and step-by-step guidance for taxpayers and internal staff. • Provide multi-language support where applicable (e.g., English, Kiswahili). • Must use guardrails to prevent hallucinations and ensure legally-compliant responses.	
2.6 GenAI Chatbots (Internal & External)	• The platform shall provide an internal chatbot for staff to assist with tax policy interpretation, audit preparation, compliance checks, internal procedures, and system troubleshooting. • The platform shall provide a public-facing chatbot for taxpayers, offering guidance on filing, registration, payments, refunds, and understanding tax obligations. The chatbot shall facilitate escalation of complex issues to a human agent. • The platform must support omnichannel delivery: web portals, APIs, WhatsApp, mobile apps, and SMS gateways for user interaction. • The platform must provide analytics on chatbot(s) usage, accuracy, escalation patterns, and interaction logging.	
2.7 Integration with External LLM APIs	• The platform must support secure integration with external LLM APIs through API gateways. • The platform MUST ensure taxpayer or sensitive data never leaves sovereign control unless explicitly approved through governance controls. • The platform shall mask, tokenize, or anonymize data before sending it to external models. • Must include configurable network security controls, audit trails, and logging for all API calls.	
2.8 Governance, Security & Compliance for GenAI	• Must provide a GenAI governance framework defining roles, access levels, data handling rules, and model validation processes. • Support for content filtering, toxicity detection, bias testing, and correctness validation.	

	• All model responses must be logged for auditability and regulatory compliance. • Support human-in-the-loop workflows for sensitive decisions e.g. (audit flags, risk scoring, taxpayer classification).	
2.9 Model Monitoring & Lifecycle Management	• Must provide continuous monitoring of LLM performance, accuracy, drift, and hallucination rates. • Provide version control, model rollback, scheduled retraining, and lineage tracing. • Must support model A/B testing for safe deployment.	
2.10 Training, Change Management & Capacity Building	• Must provide comprehensive training to tax officers, analysts, auditors, policy experts, and IT teams on GenAI products usage. • Must deliver manuals, curriculum, trainer programs, and user adoption support. • Must include guidelines on responsible AI, data privacy, and ethical usage.	

3. INTEGRATED RISK ANALYTICS

KRA seeks to develop and implement a robust and Integrated Risk Engine (IRE) to systematically identify, assess, and prioritize taxpayer compliance risks across domestic taxes and customs and border control and hence aggregates capabilities of the various existing risk engins. The Risk Analytics Engine must align with the administration's broader Compliance Risk Management (CRM) Framework and strategic objectives, such as promoting voluntary compliance and maximizing revenue. The vendor shall evaluate existing capabilities for re-use/integration where applicable to avoid any duplications.

Specification	**Detailed Requirements**	**Bidders Response**
3.1 Risk Assessment and Prioritization	Calculate a quantifiable risk score for individual taxpayers, taxpayer segments, and specific issues/schemes.	
3.2 Case Selection and Workload Management	Generate prioritized case lists for various compliance interventions (e.g., audits, investigations, education, collection) to ensure efficient resource allocation.	
3.3 Policy and Strategy Feedback	Provide data-driven insights on emerging compliance risks and the effectiveness of current compliance treatments to inform future policy and strategic planning.	

3.4 Real-time Processing	Provide real-time or near-real-time risk assessment for high-volume transactions, such as e-invoicing or imports/exports.	
3.5 User Interface (UI)	A dedicated Risk Management Dashboard for analysts and management.	
3.6 Data Model	A defined, normalized Risk Data Mart separate from the core TAS for performance and analytical flexibility.	
3.7 Data Quality (DQ) Checks	Automated checks applied during ingestion to validate completeness, consistency, and accuracy of data feeding the models.	
3.8 Rule Engine (Business Rules)	A module that allows business users (risk analysts) to define, test, and activate deterministic rules without IT intervention. Rules must be managed in a standardized format (e.g., Decision Tables or a Domain Specific Language).	
3.9 Model Repository	A central library to store, version control, and manage all predictive models (e.g., ML models, statistical models). Use of MLflow, Google Vertex AI, or similar for model tracking and lifecycle management. Must support A/B testing of different model versions concurrently.	
3.10 Explainability and Auditability (XAI)	All risk scores, model weights, and rule applications must be transparent and auditable (i.e., explainable) to justify case selection decisions.	
3.11 Audit Log	The system must log every scoring event, including input features, model version, and the final score.	
3.12 Scoring Pipeline	The engine must be able to calculate sub-scores (e.g., Filing Risk Score, Payment Risk Score) and combine them using weighted averages or other configurable methodologies to produce a Final Composite Risk Score	

3.13 Core TAS Integration	Seamless data exchange with the existing Tax Administration System (TAS) for retrieving historical data and pushing case assignments. The RE must automatically push the prioritized list of Taxpayer IDs, Risk Scores, and the suggested Intervention Type (e.g., Audit, Collection, Inquiry) to the Case Management Module.	
3.14 Third-Party Data Integration	Integration with external data sources (e.g., Customs, Corporate Registry, Banks, E-Invoicing platforms). All external connections must use VPN/Private Links and be encrypted with TLS 1.2+. Data Standards: Must handle various data formats (e.g., XML, JSON, CSV) and support standard protocols (e.g., SFTP, SOAP).	
3.15 Alerts and Notifications	System must issue immediate alerts for critical, high-volume fraud schemes (e.g., missing traders).	
3.16 Virtual Assistant	Incorporate a virtual assistant that: • Provides a comprehensive summary and analysis of a selected risk case/taxpayer ID. • Translates the technical Risk Score into natural language, explaining *why* the taxpayer was flagged and the key indicators that drove the score. • Suggests the most appropriate compliance treatment (e.g., full audit, desk review, educational letter, collection action) based on risk type and score. • Answers specific questions about tax law, procedures, or past audit precedents relevant to the case.	
3.17 Performance	Must be capable of processing large volumes of data and generating risk scores and case lists within defined performance thresholds (e.g., daily/weekly processing).	
3.18 Core Architecture	The IRE must be a modular, service-oriented architecture (SOA) or microservices-based system to ensure scalability, resilience, and independent	

	deployment of components (e.g., Data Ingestion, Rule Engine, Scoring Service).	
3.19 Containerization	Must utilize Docker and be orchestrated via Kubernetes (K8s) for automated deployment and scaling. -	
3.20 API Gateway	All external service calls must pass through a secure API Gateway (e.g., NGINX, AWS API Gateway) for traffic management and security.	
3.21 Backend	Python (for data science/ML models) and/or Java/Spring Boot (for core microservices). -	
3.22 Database	High-performance, fault-tolerant database	
4. Virtual Digital Assets KRA seeks to procure and implement a comprehensive Virtual Digital Asset Tax Management Solution to enhance its ability to track VDA transactions, accurately assess tax liabilities, and enforce compliance. The solution will leverage blockchain analytics and real-time monitoring and integrate with existing KRA systems to improve oversight, efficiency, and revenue mobilization within the digital asset ecosystem.		
4.1 Transaction Data Collection	The system must collect transaction data from multiple sources, including but not limited to: a) Cryptocurrency exchanges (via APIs or CSV uploads). b) Cryptocurrency wallets (hot and cold wallets). c) Blockchain networks (via blockchain explorers or nodes). d) DeFi platforms and NFT marketplaces.	
4.2 Data Integration	The system shall: • Integrate crypto exchanges with KRA systems (e.g., iTax) and external data sources (e.g., AML/CFT systems) • Establish secure API connections with banks and Payment Service Providers (PSPs).	

4.3 Real-Time Transaction Tracking	The system shall: • Monitor VDA transactions in real time across multiple blockchains (e.g., Bitcoin, Ethereum). • Capture real-time transaction data from cryptocurrency/virtual digital, e-commerce platforms, iGaming services, assets exchanges, and other digital payment channels	
4.4 Entity Identification	The solution shall: 1. Identify entities executing trading in the exchanges and executing transfers. 2. Identify merchants' executing payments through digital assets. 3. Link wallet addresses (clustering) to real-world identities using: 4. KYC data from exchanges and block chains. 5. Use Clustering algorithms to group related wallets across multiple blockchains. Integration with Third-party blockchain analytics tools (e.g., Chainalysis, Elliptic).	
4.5 Risk Scoring and Alerts	The solution shall assign risk scores to transactions and entities based on predefined risk indicators (e.g., transaction size, frequency, counterparties). The Risk scoring shall help prioritize compliance, audits and investigations.	
4.6 Tax Calculation	The solution shall Automate the calculation of taxes (Digital Asset Tax and Capital gains/losses) involving: • Sales of VDAs for fiat currency. • Transfer/Exchanges of one VDA for another. • Spending VDAs on goods or services. • Supporting Multiple cost basis methods (e.g., FIFO, LIFO, HIFO). The solution shall:	

	• Identify and classify taxable transactions based on predefined rules and tax categories. • Compute tax in real-time at the point of transaction authorization. • Automatically deduct and remit tax to the designated tax authority accounts. • Identify and classify cryptocurrency transactions based on predefined rules and tax categories. • Compute Transfer or exchange revenues on applicable transactions. • Automatically deduct Tax on transactions and remit tax to the designated tax authority accounts. • Provide Tax Intelligence functions on crypto currency transactions. • Compute revenue on other digital platform services transactions. • Automatically deduct tax on digital platform services to the designated tax authority accounts • Provide Tax Intelligence on digital platform services transactions. • Remit all taxes collected in real-time to the revenue authority.	
4.7 Transaction Visibility and Reporting	• Provide a centralized dashboard for tax authorities with real-time insights into taxable transactions. • Provide functionality to query blockchain transactions e.g by wallet address or another relevant identifier. • Generate detailed reports on taxes collected per merchant, industry, and transaction type. • Implement analytics and trend monitoring to support policy-making.	

5. Unified Fiscal Data Intelligence The Unified Fiscal Data Intelligence Layer establishes the single trusted fiscal and compliance data foundation required for a modern revenue authority to operate with accuracy, speed, and accountability. It is the layer where disparate records—spanning taxpayer obligations, assessments, billing events, payment attempts, receipts, settlement confirmations, refunds, reversals, and downstream accounting postings—are consolidated into standardized, reconciliation-grade data structures that can be reliably used across analytics, operational workflows, and oversight reporting. In practical terms, this layer ensures that fiscal facts are not fragmented across systems, not interpreted differently by different teams, and not vulnerable to silent mismatches that only surface during audit, end-month close, or revenue shortfalls.		
5.1. Source System Onboarding & Connector Framework	Provide a reusable connector framework to onboard priority internal systems and approved external sources using secure integration patterns (API-based, event/stream, batch/file, and controlled DB replication where permitted). Must support configuration-driven onboarding (minimizing bespoke code), connection health checks, rate limiting, retries, and idempotent ingestion controls.	
5.2. Canonical Fiscal & Compliance Data Model	Define and implement a standardized canonical model with versioning and governance. Must include key entities needed for fiscal integrity (e.g., taxpayer/party references, obligations, invoices/demands, payment attempts, receipts/confirmations, settlements, refunds, ledger/accounting events, reconciliation records, exceptions/case references, and audit events). Must include clear keys, relationships, and mapping guidance for onboarding additional sources.	
5.3. Schema Mapping, Transformation & Standardization	Provide transformation services that map source fields into canonical structures with documented mapping specs, version control, and controlled promotion across environments. Must support validation rules at ingest, normalization, enrichment (where authorized), and reversible transforms for audit evidence. Must include lineage from source → transform → canonical entity/attribute.	

5.4. Fiscal Data Quality Controls & Validation Engine	Implement automated data quality controls for completeness, accuracy, consistency, timeliness, and referential integrity. Must support configurable validation rules, thresholds, and tolerance bands; generate data quality scores; and route failed records to governed remediation processes (quarantine, reprocess, correction request). Must maintain evidence of validation outcomes and changes applied.	
5.5. Reconciliation-Grade Data Readiness & Matching Key Management	Define and operationalize reconciliation-grade readiness rules and matching key strategies (exact, fuzzy, window-based time matching, tolerance-based amount matching). Must maintain match keys, correlation IDs, and deduplication logic for high-volume ingestion. Must support backfills, replay, and reprocessing without double-counting.	
5.6. Automated Multi-Source Attribution Engine	Provide automated reconciliation across priority fiscal objects (e.g., invoice, payment attempt, receipt, settlement, ledger posting; refund request, approval, payout). Must support configurable matching rules, exception generation, variance quantification, and reconciliation evidence capture (what matched, why, what didn't). Must produce reconciliation status outputs suitable for downstream workflow handling.	
5.7. Master Data & Reference Data Management	Establish controlled reference data and master data governance needed for consistent fiscal reporting (e.g., tax heads, revenue codes, periods, channels, payment instruments, organizational units). Must support versioning, stewardship workflows (approve/change), and synchronization with consuming services.	
5.8. Metadata, Lineage & Cataloging	Provide an enterprise-grade metadata layer capturing dataset definitions, owners, classification, lineage, freshness/SLAs, and usage. Must support discoverability for analytics teams and auditors, including lineage views from source systems to canonical datasets and reconciliation outputs.	

5.9. Security, Access Control & Data Protection for Fiscal Data	Enforce RBAC (and where required ABAC), encryption in transit/at rest, sensitive field masking/tokenization, and secure secrets/key management. Must support least privilege, segregation-of-duties for data/admin functions, and auditable access to datasets and exports.	
5.10. Auditability & Data Integrity Logging	Maintain immutable audit logs and integrity evidence for ingestion, transformations, reconciliation decisions, and dataset publishing. Must support exportable audit trails for oversight and investigations, including correlation IDs linking source records to canonical outputs and reconciliation results.	
6. Unified Fiscal Data Process Flow Intelligence The Unified Fiscal Data Process Flow Intelligence Layer converts trusted fiscal data into controlled execution across the revenue lifecycle. Whereas the Fiscal Data Intelligence Layer ensures records are standardized and reconciliation-grade, this layer ensures the organization can act on mismatches, risks, and operational signals through governed workflows that are measurable, auditable, and enforceable. It is the layer where fiscal integrity becomes operational reality: exceptions are not just detected—they are routed, assigned, escalated, resolved, and closed with evidence; reconciliations are not just run—they drive downstream actions; and operational states (e.g., pending settlement, failed posting, suspicious reversal, unresolved variance, overdue refund) are managed through structured processes with SLAs and accountability. This layer is responsible for orchestrating revenue and compliance process flows end-to-end, including validation, reconciliation resolution, settlement processing, refund processes, enforcement escalations, and other critical fiscal control workflows .		
6.1. Workflow Orchestration Engine (Configurable Process Definitions)	Provide configurable workflow orchestration capable of modelling and executing revenue and compliance process flows (e.g., validation, reconciliation resolution, settlement handling, refund handling, escalation workflows). Must support state transitions, branching, timers, approvals, and version-controlled workflow definitions with controlled promotion across environments.	
6.2. Business Rules & Decisioning (Rule Versioning + Governance)	Provide rule-based decisioning services used to drive workflow routing and actions (e.g., match thresholds, tolerance bands, severity classification, escalation rules). Must support rule versioning, approval workflows, audit trails for rule changes, and ability to simulate/test rule impacts prior to production deployment.	

6.3. Exception Lifecycle Management (Queues, Ageing, Closure Codes)	Implement end-to-end exception lifecycle management for mismatches and operational failures (e.g., reconciliation mismatches, settlement failures, posting failures, refund anomalies, reversals). Must support categorization, severity scoring, queue assignment, ageing, SLA tracking, resolution codes, evidence attachment, and controlled closure with approvals based on severity.	
6.4. Task Routing, Assignment & Work Queues (Human-in-the-Loop)	Provide task management capabilities including assignment, reassignment, routing by unit/region/officer/skill, workload balancing, and escalation pathways. Must support delegated authority, segregation-of-duties, and role-based access so that actions are performed by authorized users only.	
6.5. SLA Engine & Escalation Management	Provide SLA timers, breach detection, escalation policies, and automated notifications for workflow tasks and exception queues. Must support configurable SLAs by exception type, severity, revenue impact, and organizational unit, and provide metrics for compliance and operational performance monitoring.	
6.6. Case Management Integration (Operational + Compliance Cases)	Integrate workflow execution with case management so that workflow events can create/update cases and cases can drive workflow actions. Must support case creation triggers, case state synchronization, action logging, and linkage to underlying transactions, exceptions, and reconciliation evidence.	
6.7. Settlement / Posting / Refund Process Controls (Operational Controls)	Provide process controls to manage operational fiscal states, including settlement status tracking, posting confirmations, reversals handling, refund stage tracking, and failure recovery. Must support automated routing of failures into exception workflows and ensure end-to-end traceability from detection to closure.	
6.8. Operational Event Trail & Audit Logging (Process-Level Auditability)	Maintain a complete, immutable operational event trail of workflow state changes, user actions, approvals, escalations, and closures. Must support correlation IDs across transactions, workflow instances, exceptions, and cases, and enable audit export for oversight and investigations.	

6.9. Controls: Idempotency, Retries, Replay & Recovery	Ensure safe execution under failure conditions through idempotent processing, controlled retries, dead-letter handling (where applicable), and replay/reprocessing controls to prevent duplication or double-resolution. Must provide runbook-grade operational procedures for recovery and reprocessing.	
6.7. Security & Governance for Operational Workflows	Enforce RBAC/ABAC, segregation-of-duties, privileged access controls, and governance over workflow and rule changes. Must ensure secure handling of sensitive fiscal actions (approvals, reversals, refunds) and provide evidence of access and change control.	
7. Fiscal Data Revenue Enhancement The fiscal data revenue enhancement layer ensures that the unified fiscal data foundation and operational process controls translate into measurable improvements in revenue outcomes. While the fiscal data intelligence layer establishes trusted, reconciliation-grade data and the process flow intelligence layer ensures exceptions and operational breakdowns are systematically resolved, this layer focuses on turning detected leakage points into recoveries, optimizing compliance interventions, and continuously improving revenue performance through closed-loop measurement. It provides the mechanisms to quantify revenue-at-risk, prioritize recovery actions, measure intervention effectiveness, and institutionalize learning so that rules, workflows, and operational playbooks improve over time. This layer therefore emphasizes outcomes such as: (i) continuous identification and quantification of leakages, under-collections, delayed collections, and systemic bottlenecks; (ii) targeted, trackable revenue recovery actions and compliance interventions linked to accountable owners; (iii) measurement frameworks that estimate uplift attributable to interventions (not just gross collections); and (iv) feedback loops that refine decision rules, workflow routing, and operational practices to prevent repeat leakages and maximize sustainable revenue growth.		
7.1. Revenue Leakage Detection & Quantification (Revenue-at-Risk)	Provide capability to identify, classify, and quantify revenue leakages and under-collections across the fiscal lifecycle, including mismatches, delayed settlements, failed postings, reversals, unpaid demands, and abnormal refund patterns. Must compute revenue-at-risk estimates, ageing, exposure bands, and impact by tax head, channel, region, system, and period, with drill-down to transaction level evidence.	

7.2. Revenue Recovery & Intervention Management (Action Tracking)	Provide mechanisms to register, manage, and track recovery actions and compliance interventions (manual or automated), linked to specific leakage events/exceptions/cases, with ownership assignment, due dates, escalation, and closure outcomes. Must integrate with workflow and case components to ensure interventions are executed and closed with evidence.	
7.3. Intervention Prioritization & Targeting (Optimization)	Provide prioritization logic and decision support to target the highest-value interventions based on revenue-at-risk, probability of recovery, time sensitivity, compliance risk, and operational capacity. Must support configurable prioritization rules and "what-to-work-next" views for operations teams.	
7.4. Uplift Measurement & Attribution Framework	Implement an uplift measurement framework to estimate incremental collections attributable to interventions (e.g., before/after, cohort comparisons, matched controls, or other accepted methods depending on data availability). Must distinguish gross collections from incremental uplift, include confidence/assumption disclosure, and support reporting by intervention type and organizational unit.	
7.5. Compliance Outcome Measurement & Effectiveness Analytics	Provide analytics to measure outcomes such as compliance improvement rates, repeat exception rates, resolution cycle time reduction, settlement success rates, refund integrity improvements, and reduction in leakage recurrence. Must enable trend analysis and benchmarking across regions/units and over time.	
7.6. Closed-Loop Feedback into Rules & Workflows	Provide mechanisms to institutionalize learning by feeding outcomes back into business rules, workflow routing, severity classification, SLA policies, and operational playbooks. Must support governance over rule/workflow changes and demonstrate how outcome evidence informs adjustments to prevent recurrence and improve yield.	

7.7. Revenue Performance Operational Packs & Executive Reporting	Provide standardized operational and executive packs (daily/weekly/monthly) summarizing revenue-at-risk, recoveries, uplift, bottlenecks, exception backlogs, and key drivers. Must support role-based views, distribution controls, and auditable versions of reports used for management and oversight.	
7.8. Channel & Partner Performance Analytics (PSPs/Banks/Collections Channels)	Provide performance analytics by collection channel and partner (where applicable), including settlement timeliness, failure rates, exception rates, reversals, and fee/charge impacts. Must support identification of systemic issues and evidence-based engagement with partners to improve performance.	
7.9. Controls for Revenue Enhancement Actions (Governance & SoD)	Ensure governance and segregation-of-duties for revenue enhancement actions (e.g., write-offs, reversals handling, refund approvals, intervention overrides). Must maintain audit trails for actions taken, approvals granted, and changes to prioritization policies or intervention rules.	
7.10. Sustainability & Operationalization (Playbooks, Adoption, Handover)	Provide operational playbooks, standard operating procedures, and adoption mechanisms to embed revenue enhancement practices into day-to-day operations. Must include training/knowledge transfer to ensure sustained usage and continuous improvement beyond implementation.	

8. **Data Components** (Data Processing Engines)

This layer provides data processing modules that process raw data into actionable insights.

i. Reconciliation Engine	The platform must include a Reconciliation Engine to compare, validate, and reconcile data across multiple systems (e.g., iTax, iCMS, ERP), ensuring data integrity and accuracy.	
ii. Rules Engine	The platform must provide a Rule Engine to automate decision-making based on predefined business logic, used in Tax Intelligence, Fraud Management, and Reconciliation.	
iii. KPI Evaluator/Monitor	The platform must feature a KPI Evaluator to measure, track, and analyze KPIs for business processes, ensuring operational, financial, and compliance targets are met.	

iv.	ML Core Engine	The platform must have an ML Core Engine that manages the entire machine learning model lifecycle, from data ingestion and training to deployment, monitoring, and retraining.	
v.	Aggregators	The platform must include Aggregators to collect, summarize, and process data from multiple sources to generate meaningful metrics for the KPI Evaluator and other analytics functions.	
vi.	EDA Module	The platform must provide an EDA Module to analyze, summarize, and visualize data to understand its structure, detect patterns, and identify anomalies before further processing.	
vii.	DQA (Data Quality Assurance)	The platform must feature a Data Governance and Quality Assurance (DQA) module to ensure data integrity, accuracy, and reliability. **A) Data Governance and Cataloguing:** • **Data Governance and Stewardship Workflows:** Must provide stewardship dashboards, social governance enabling users to rate/like, etc; policy documentation management, with meta data management; data catalogue, policy enforcement capabilities, data access governance with approval workflows and regular review enforcement framework • **Automated Data Cataloging & Metadata Harvesting:** Must provide storage of both active and passive meta data, automated scanning of meta data and semantic inferencing to reduce manual effort by data stewards • **Business Glossary & Semantic Layer** Must provide standardized taxonomy of business terms, contextual linking between technical meta data assets to business terms, and versioning.	

	• **Must Provide end to end data lineage** i.e. horizontal and verticle lineages, automated parsing and impact analysis • **Access Control & Security Orchestration** - must be able to integrate to data analytics platforms to enforce access restriction or masking of PII data fields from unauthorized users • The solution must provide Automated Personally Identifiable and Sensitive Personal Information **PII/SPI Discovery** with an heatmap showing where such data resides. • Must provide **Data Sovereignty and Residency** capabilities to track movement of data across jurisdictions to restrict unauthorized jurisdictions, and monitor the "age of data" and trigger automated alerts/scripts to delete data that has exceeded its legal retention period **Data Subject Access Request Automation** – leveraging the data catalog and lineage, the solution must be able to generate a report of "Everywhere" the organisation stores data for any specific user for further action. **B) Data Quality Automation:** • Must perform **automated data profiling -** deep profiling, rule-based and relationship discoveries. • The solution must automate the **monitoring of the six core Data Quality dimensions**, i.e. Accuracy, Completeness, Consistency, Timelines, Validity, and Uniqueness • Must be able to detect the "unknown unknowns" in data quality by providing automated **ML-Based Anomaly Detection** including	

	Volume anomalies, Schema drift and value drift anomalies. • Provide **automated Remediation & Incident Management** with auto-alerting and assignment and resolution workflows – must provide Root Cause analysis leveraging the data lineage to trace errors to their origins, perform closed-loop remediation, and automatically quarantine "bad" data to prevent it from production use. • Must **provide real-time data quality score cards and dashboards** to show data health by various dimensions; • Must provide ability for **automated data cleansing, enrichment pipelines and reconciliation** processes with SLAs **C) Data Quality Process Automation Documentation:** • **Define the Target Operating Model (TOM):** Develop a comprehensive document defining how the automated solution will be embedded into the daily workflow. This includes the "New Way of Working" for data requests, quality resolution, and metadata updates • **Define a Data Governance Charter & Playbook**: Develop the codified set of rules, standards, and procedures for data governance to serve as the "Instruction Manual" for the organisation, outlining who makes decisions and how the automation tool supports those decisions	

9. Data Layer

This layer handles data storage, retrieval, and management, ensuring efficient storage, integration, and accessibility of data. This layer will provide a modern cloud native data platform that is fully compatible with open-source data analytics tools. The platform will host the following components.

Specification	Requirements	Bidders Response
i. Lakehouse	The Lakehouse platform will be used for complex analytics and storing historical data. The solution must provide a unified	

		distributed storage for structured and unstructured data while supporting ACID features and providing S3 object storage compatibility.	
ii.	Query Processing	Must enable interactive querying of large datasets using distributed query engines and other diverse tools.	
iii.	Data Federation	Must allow querying across multiple heterogeneous data sources.	
iv.	Metadata Management	The platform must provide capabilities for metadata management • Automate metadata harvesting from databases, storage, BI tools and ML pipelines • Business metadata definition • Provide a global data catalog that is searchable and filterable for all data assets.	
v.	Database Federation Layer	Must provide unified querying across multiple relational and non-relational databases.	
vi.	Storage Solutions	Must support a variety of storage solutions: • **Analytical Databases:** Optimized for complex analytical queries (OLAP). • **SQL** For highly relational and structured data. • **NoSQL:** For high-velocity and schema-flexible data ingestion. This will also be used for event-storage and key-value access patterns. • **Open Table Format:** Must implement and support open table formats with full Rest Catalog support	
vii.	Real-Time (RT) Data Handler	Must manage and process real-time data streams, ensuring low-latency ingestion and processing for event-driven analytics and anomaly detection.	
viii.	Analytical Database	A database optimized for complex analytical queries rather than transactional processing. It shall support OLAP (Online Analytical Processing), allowing efficient data aggregation, trend analysis, and reporting.	

ix.	Governance & Data Quality Automation Platform	The solution must provide a data governance capability across the 5 layers of the platform with each functionality available as an API. The following capabilities must be provided: ○ End-to-end data lineage • Data pipeline linage • BI dashboards and reports lineage ○ Security & Access Control – • RBAC/ABAC – Fine grained role and attribute-based access • Column/Row level restrictions • AD/IAM/LDAP integration • Traceability auditing for all access and policy changes in the platform ○ Data classification – Must provide capabilities for data masking, encryption and tagging for data sensitivity labeling. ○ Compliance & Policy – Provide an automated way for setting data retention policies, consent tracking as well as regulatory compliance with Kenya's Data Protection Act of 2019 ○ Data quality rules – Provide capability for creating custom / no code rules for data quality ○ Automated Profiling ○ Schema Drift Detection – Ensure schema change alerts and time travel ○ Observability - pipeline health and data freshness ○ Automated Retry jobs for data pipelines ○ Stewardship & Ownership of data sets and dashboards ○ Change request workflows & version control ○ Issue Management – Data quality ticketing ○ AI/ML Governance – model lineage	

	○ Track usage metrics	
10. Data Mediation Layer (Data Collection & Integration) This will handle ingestion, transformation, and schema standardization of raw data from multiple sources**.**		
i. Streaming	Must provide a real-time event streaming platform.	
ii. Log & metrics parsing	Must collect and process metrics and logs from the various components of the analytics platform.	
iii. OCR (Optical Character Recognition)	Must provide capabilities for extracting text from images and scanned documents.	
iv. Data Extraction & Upload	Must provide scalable tools to extract and load structured and unstructured data into the platform.	
11. Data Sources & CDC (Change Data Capture) Connects to external and internal data sources, ensuring continuous updates with CDC techniques.		
i. CDC (Change Data Capture)	Must capture and replicate changes from KRA's source databases in real-time to keep the platform synchronized.	
ii. Digital Data Sources	Must be capable of collecting external data from digital platforms (e.g., social media, IoT devices).	
iii. Enterprise Systems (e.g. Payment Gateway)	The solution must capture and replicate changes from the Payment Gateway.	
12. **Data Analytics Platform Infrastructure requirements** The Data Analytics Platform must be deployed as a hybrid, multi-site clustered setup spanning a primary on-premise cloud, an off-premise cloud environment, and a disaster recovery site. This architecture will ensure continuous availability, seamless failover, and business continuity. The platform must support real-time and data replication across all sites to guarantee data resilience and business continuity. All platform components and tools must be cloud-native, modular, and fully compatible with open-source technologies for data analytics. The platform must decouple storage and compute to enable elastic scaling and optimize resource utilization. The set up must ensure:		

Unified Data Analytics & AI Platform Infrastructure Requirements			
Specification		**Requirements**	**Bidders Response**
1.	Hybrid Cloud Deployment	The platform must be deployed as a hybrid cloud, with an on premise cloud, an off-premise cloud, and a dedicated disaster recovery (DR) site for business continuity and fail over. • Must support elastic scaling. • Must support compute-storage decoupling for independent scaling.	
2.	High Availability (HA)	• All components (compute, storage, catalog services, and orchestration layers) must run in redundant multi-node clusters. • A single point of failure (SPOF) is not permitted. • Automatic failover must be supported. • Must support real-time replication of data across all sites with metadata consistency	
3.	Horizontal & Vertical scalability	• Compute and storage layers must scale horizontally by adding nodes, allowing the platform to grow as data volume and workload demand increase. • Must support vertical scalability for different workload needs	
4.	Cluster-Oriented Resource Management	• Workloads must be distributed across clusters for optimal resource utilization. • Must integrate with a cluster scheduler/orchestrator for: ➢ Containerized execution ➢ Job scheduling ➢ Autoscaling policies	
5.	Distributed Storage and Processing	• The platform must support distributed storage engines and analytical query engines capable of: ➢ Parallel data processing ➢ Distributed caching	

6.	Secure Multi-Node Communication	• All intra-cluster communication must be encrypted. • Node-to-node communication policies must enforce: • Mutual TLS authentication • Role-based access control • Network segmentation • IAM systems	
7.	Observability Across the Cluster & Reliability	The cluster must be fully monitored with: • Metrics collection (CPU, memory, storage, throughput) • Centralized logging • Distributed tracing for analytics workloads • Alerting and automated remediation Platform reliability: • All components must auto-recover after restart. • Must support multi-data-center cluster deployment. • All components must support high availability. • Must support active/standby clusters for DR.	
8.	Support for Mixed Workloads & Scheduling	The cluster must efficiently run both: • Batch analytics (ETL/ELT, large-scale processing) • Real-time analytics (streaming, event processing) Workload isolation via namespace, queue, or resource quota is required. • Must provide a job scheduling utility	
9.	Compatibility & Usability	• Must be compatible with x86 cloud computing platforms. • Must fully support open-source component interfaces. • Must provide API documentation, development guides, samples, and demos as well as markdown documentation of code and platform features.	
10.	Heterogene	• The platform must support	

	ous Cluster Support	heterogeneous nodes with variable CPU types, disk capacity, disk types, and memory sizes. • Must allow arbitrary combinations of node specifications.	
11.	Cluster Deployment & Autoscaling	• Must provide GUI-based installation, deployment, capacity expansion, and auto-scaling. • Auto-scaling must support time-based and load-based rules.	
12.	Multi-Instance Capability	• Must support deploying multiple services of the same component within a single cluster. • Must permit multiple component instances per node to improve resource usage.	
13.	Rolling Operations	• Must support rolling upgrades without service interruption. • Must support rolling restarts with zero service downtime.	
14.	Health & Monitoring	• Must support automatic health checks, preventive diagnostics, and system auditing. • Must provide a GUI-based cluster health checker with monitoring and observability dashboards.	
15.	Maintenance	• Must support remote and local maintenance for all clusters.	
16.	Storage & Distributed File system	**Storage Backends** • Must support a distributed file system and S3-compatible object storage. **REST Access Layer** • The distributed file system must provide REST APIs supporting file creation, deletion, upload, and download. **Balanced Data Distribution** • Must support balanced node scheduling and balanced disk I/O scheduling within a node. **Tiered Storage** • Must support tiered storage using SSD, SAS, SATA, or other disk types on the same node.	
17.	Data exploration	The platform must avail the following capabilities	

	, processing and analysis	• Must provide a unified JDBC portal for seamless data exploration.
		• Must provide full support for open formats: ORC, Parquet.
		• Must support table formats: Iceberg, Delta Lake, Hudi.
		• Must support object storage (S3, MinIO, HDFS)
		• Must support distributed compute clusters like (Spark, Trino, Flink, Kubernetes workloads) that scale elastically
		• Resilient distributed query processing with automatic workload balancing
		• Must provide multi-dimensional SQL execution metrics at both tenant and workload level:
		  ○ submission time
		  ○ execution time
		  ○ wait time
		  ○ CPU/memory usage
		  ○ I/O scan size
		  ○ result count
		• Metadata-Driven Optimization to enhance query planning and execution
		• Must support materialized views, automatic refresh, and intelligent optimization recommendations.
		• Must support ability to execute federated SQL queries
		• Must provide metadata caching to improve read performance.
		• Must integrate with mainstream BI tools like (Tableau, PowerBI, Superset, Looker) with no vendor lock-in.	

18.	Graph Management & Graph Search	• Must support analytics on complex relationships structured as graphs. • Must support building custom knowledge graphs using entities,	

		relationships, and attributes. • Must support graph search via graph type and keyword-based queries.	
19.	Security Requirements	**General Security** • Must control service ports and provide documentation of all ports. • Must provide a security and compliance framework • Must support unified authentication for cluster users and components. • Authentication & Authorization • Must support LDAP, Kerberos, and table/column/data encryption. • Must support unified SSO for management platform and component UIs. • Must provide policy and role-based permission control. • Must provide a data governance mechanism across the entire platform **Access Controls** ➢ Must support GUI-based switching between policy-based and role-based permission modes. ➢ Must support fine-grained control at row-level and time-level. ➢ Must support delegation/federation to specialized security services. ➢ Must support dynamic data masking. Must log all operations with: ➢ username ➢ client IP ➢ start & end time ➢ operation name/activities ➢ operation result	
20.	Multi-Tenancy	• Must support multi-tenant management with unified resource controls to support isolated Production, Quality Assurance and Development environments. • Must support resource isolation, usage statistics, and configurable resource caps. • Must support static isolation, dynamic preemption, and time-based dynamic adjustment of resources.	

		• Must allow deployment of multiple services of the same component in the same cluster, with physical resource isolation.	
21.	Distributed In-Memory Computing	• Must provide a distributed in-memory compute engine with SQL syntax support. • Must support container orchestration-based deployment	
22.	Separation of Roles/Profiles	• The platform must ensure separation of operational roles. It must include roles for: o Data Engineers o Data Scientists o BI developers & Analysts o Data Governance o Quality assurance & testing o BI Platform Engineers o Security & Audit o Infrastructure Engineers o Administrators o BI Platform Users	
23.	Comprehensive knowledge transfer requirement (user guides, technical training and technical support)	• Deployment & Configuration Documentation ➢ Deployment Guides o Installation and setup guides o Environment configuration (DEV, QA, PROD, DR) ➢ Infrastructure as Code Documentation o CI/CD pipeline documentation o Automated deployment ➢ Configuration Standards o Network configuration o Cluster configuration o Data retention & tiering settings o Resource allocation templates (CPU, memory, partitions) • Operational and Engineering Knowledge ➢ Deployment processes, automation & CI/CD pipelines	

	➢ Infrastructure performance tuning and workload tuning ➢ Monitoring, observability & Alerting ➢ Diagnostic procedures and troubleshooting ➢ Deliver operationalization and handover documentation ➢ Site/DR and replication configuration documentation • Platform Management & Governance Documentation ➢ Platform Governance o Platform administration roles o Tenant management ➢ Change Management o Change request process o Release management o Migration procedures • User Enablement Documentation ➢ User Guides o Using BI tools o Running analytics and ML workloads ➢ Training Material o Slides, videos, and walkthroughs o Hands-on labs o FAQ & troubleshooting tips ➢ Developer Documentation o API documentation	
24. Required Platform Documentation	• Architecture Documentation ➢ High-Level Architecture o Hybrid deployment architecture (on-prem, off-prem, DR) o Cluster topology diagrams o Data flow & workload flow diagrams o Integration architecture ➢ Detailed Technical Architecture o Storage architecture o Compute architecture o Ingestion architecture	

	○ Metadata & catalog architecture ○ Security architecture ○ Networking & connectivity architecture ○ Multi-site replication architecture ○ Failover & DR design ➢ Design Documents ○ Component design specifications ○ Sizing & capacity planning ○ HA & scaling design (horizontal + vertical) ○ Reference data architectures • Operations & Administration Documentation ➢ Runbooks ○ Service restarts ○ Cluster scaling ○ Node replacement ○ Backup & restore ○ DR failover & failback ○ Security patching ➢ Standard Operating Procedures (SOPs) ○ Upgrading the platform ○ Managing secrets and credentials ○ Onboarding new users ➢ Monitoring & Alerting Documentation ○ Alert definitions and thresholds ○ Incident escalation procedures ○ Dashboards ➢ Troubleshooting Guides ○ Common cluster issues ○ Query performance troubleshooting ○ Pipeline/job failure diagnostics ○ Networking & connectivity issues ○ Storage bottleneck	

	troubleshooting • Security, Governance & Compliance Documentation ➢ Security Documentation o IAM model (RBAC, ABAC) o Encryption policies (in transit and at rest) o Audit logging configuration ➢ Data Governance Documentation o Data Security Controls & Access Management and compliance o Data Quality Management & Lineage tracing o Data Retention, archival & tiered Storage management o Metadata Management and cataloging practices o Business glossary & data dictionary o Data classification and tagging o Data privacy policy ➢ Compliance Documentation o Data access and usage logs o Retention and archival policies o • Data Pipeline & Integration Documentation ➢ Ingestion Documentation ➢ Transformation & Processing Documentation ➢ Output & Consumption Documentation	

6 SECTION 3: STAFF TRAINING AND CAPACITY BUILDING

6.1 Introduction

Investing in analytics staff training is essential for KRA to strengthen evidence-based decision making enhance its capacity to operate as a modern, data-driven tax administration. As the operating environment becomes increasingly digital and intelligence-led, staff require advanced competencies in analytics, artificial intelligence, machine learning, data engineering, and cloud technologies. Equipping staff with these skills will not only improve individual performance but also drive institutional transformation. This section outlines the mandatory training programs to be delivered as part of the capacity-building initiative.

6.2 Scope of work

6.2.1 Summary

All the training courses should be provided up to certification level. For the training areas identified, a detailed training proposal should be provided outlining:

- **Training objectives:** Clearly define the goals of the training program.
- **Target Audience:** Outline the intended participants of the training. This includes their roles, experience levels, and any specific characteristics that make them suitable for the training program such as minimum requirements.
- **Course content:** Including topics to be covered, methodologies to be used, and any resources or materials that will be provided. This section should give a clear picture of what participants can expect to learn.
- **Training Methodology**: Explain the approach to be used in delivering the training, such as workshops, seminars, online courses, or hands-on sessions.
    - Bidders shall propose appropriate training venues, noting that KRA will not provide venues.
- **Duration and Schedule:** Outline the length of the training program as well as any breaks or milestones throughout the training. Bidders should also specify number of cohorts for each training based on the optimal class size and KRA's requirements.
- **Proposed Trainer(s):** Detailed trainer bio, outlining their qualifications and experience to deliver the proposed training.
- **Evaluation and Assessment:** Describe how the effectiveness of the training will be measured. This could include pre- and post-training assessments, feedback surveys, or performance metrics to evaluate participant learning and engagement.
- **Training Costs**: Detailed breakdown of all the costs associated with the training program. Including training venue, materials, cloud credits, assessments, certification fees and administration. Please note that KRA will not provide training venue and as such this should be included in the costs.

  All training deliverables must include course materials, datasets (where applicable), certificates, and documented evaluation reports.

6.2.2 Staff Training Focus Areas

Focus Area	Specific Course	Delivery Methodology	Bidder Response
1. Data Analytics & AI Foundations	Understanding Data, Data Types & Sources	Online/Instructor-led	
	Excel for Data Analysis	Hands-on Workshop	
	Introduction to SQL	Hands-on Training	
	Data Visualization (Power BI/Tableau)	Practical Workshop	
	Introduction to AI Concepts for Tax Administration	Instructor-led / E-learning	
	Foundations of Data Quality, Metadata & Lineage for Analytics	Instructor-led	
	Fundamentals of Python for Data Analytics	Hands-on Workshop	
	Introduction to R for Statistical Computing	Instructor-led / Online	
2. Advanced Data Analytics & BI	Advanced SQL	Hands-on Training	
	Business Intelligence Dashboards	Practical Workshop	
	Data Warehousing Concepts (ETL, Iceberg)	Instructor-led	
	Revenue Analysis (Tax Evasion Detection)	Case Study-Based Training	
	Spatial Analytics & GIS for Tax Intelligence	Practical Workshop	
	Behavioural Analytics & Taxpayer Segmentation Models	Applied Case Studies	
	Advanced KPI Frameworks for Tax Intelligence & Compliance Monitoring	Workshop / Simulation	
3. Data Science & Machine Learning	Python for Data Analysis	Hands-on Training	
	Machine Learning Basics	Practical Workshop	
	Predictive Analytics for Tax	Case Study-Based Training	
	AI for Tax Compliance (NLP, Fraud Detection)	Instructor-led	
	Deep Learning for Anomaly Detection & Fraud Networks	Hands-on Lab	
	Introduction to LLMs and Generative AI for Tax Use Cases	Instructor-led / Sandbox Sessions	
	Building AI Intelligent Assistant for Audit, Enforcement & Policy	Practical Workshop	

Focus Area	Specific Course	Delivery Methodology	Bidder Response
	Model Monitoring, Drift Detection & Responsible AI Validation.	Hands-on Lab	
	Supervised and Unsupervised Machine Learning Techniques. Deep learning and natural language analysis and application in research.	Hands-on Workshop	
	Predictive Modelling for Fiscal and Economic Policy	Hands-on Training	
	Machine Learning Applications in Economic and Tax Research	Online / Hands-on Workshop	
4. Data Engineering & Big Data	Big Data Tools	Hands-on Workshop	
	Data Pipelines & ETL	Practical Training	
	Cloud Data Platforms (Azure/AWS)	Instructor-led	
	Streaming Data & Real-Time Tax Intelligence	Applied Engineering Lab	
	Knowledge Graph Engineering for Taxpayer Networks	Practical Workshop	
	Vector Databases & Semantic Search for Tax Investigations	Hands-on Lab	
	API Engineering & Secure System Integration for Tax Platforms	Instructor-led	
5. Soft Skills & Strategic Thinking	Data Storytelling	Workshop	
	Critical Thinking in Data	Interactive Sessions	
	Agile & Scrum for Analytics	Instructor-led	
	Change Leadership for AI Transformation	Coaching Sessions	
	Ethics, AI Governance & Responsible Use of Models in Tax Administration	Instructor-led	
	Communication for Decision Intelligence (Presenting AI-Driven Insights to Executives)	Workshop	
	Communicating Research Insights to Non-Technical Audiences	Instructor-led	
6.Specialized Training for Executives	Data-Driven Decision Making	Executive Workshop	
	AI & Emerging Technologies in Tax Admin	Thought Leadership Sessions	
	Policy Intelligence & AI-Based Revenue Forecasting	Simulation Lab	

Focus Area	Specific Course	Delivery Methodology	Bidder Response
	Digital Economy & Crypto Taxation for Executives	Masterclass	
	Executive Training on Autonomous Compliance, Real-Time Risk Detection & Strategic Use of Intelligent Assistants	Executive Briefing	
Statistical Analysis & Econometrics	Econometrics for Policy and Tax Impact Analysis	Instructor-led	
	Time Series Analysis for Revenue and Economic Forecasting	Hands-on Workshop	
	Causal Inference Methods (e.g., Difference-in-Differences, RCTs)	Instructor-led / Case Study-Based	
	Regression Analysis Using Statistical Software (Stata/R/SPSS)	Hands-on Training	
Emerging Technologies & AI	Introduction to Artificial Intelligence for Public Sector Research	Instructor-led / E-learning	
	Ethical Use of AI and Data in Government Statistics	Seminar / Instructor-led	
Research Methods & Design	Survey Design, Sampling Techniques & Data Collection Methods	Instructor-led	
	Introduction to Mixed Methods Research	Online / Instructor-led	
	Writing Research Reports & Policy Briefs	Practical Workshop	

6.2.3 Detailed Training Focus

No.	Key Role	Training Focus	Course / Certification	Pax	Priority	CoE Stage
1	**Data Analysts & BI Specialists**	i. Business Intelligence & Reporting	- Certified Business Intelligence Professional (CBIP)	200	High	Foundation
			- BI & Reporting with Power BI, Tableau, Oracle Analytics			
			- Hands-on Dashboard Design & Data Storytelling			
			- Semantic Reporting using Vector Search–Enabled BI Tools			
			- KPI Automation for Real-Time Tax Intelligence			
			- Standards-Based Dashboard Design for Enterprise Governance			
		ii. Data Analytics & Decision Support	- Fundamentals of Data Analytics	200	High	Foundation
			- Descriptive & Diagnostic Analytics using Excel, SQL, and BI tools			
			- Tax Revenue Data Analysis			
			- Behavioural & Compliance Pattern Analytics			
			- Applied Analytics for Compliance Risk Prioritization			
		iii. Advanced Analytics & Forecasting	- Time Series Forecasting & Predictive Analytics	200	High	Foundation
			- Data-Driven Decision Making for Revenue Management			
			- Advanced Excel, Python & R for Tax Analysis			
			- Early Warning Systems for Tax Gap Detection			
			- Forecast Accuracy Evaluation & Benchmarking Techniques			
		iv. Interactive Dashboards &	- Data Visualization Masterclass	200	High	Foundation

		Data Visualization	- Advanced Dashboard Development in Power BI & Tableau			
			- Automating Reports & Alerts with BI tools			
			- Spatial Dashboards for Excise & Customs Risk Mapping			
			- Designing Dashboards for Executive Decision Intelligence			
		v. Data Storytelling & Communication	- Storytelling with Data	200	Medium	Expansion
			- Presentation Techniques for Decision-Makers			
			- Communicating Insights Effectively			
			- Narrative Design for AI-Generated Insights			
			- Evidence-Based Communication for Policy & Compliance Teams			
2	**Data Scientists**	i. Fundamentals of Data Science	- Introduction to Data Science.	200	High	Foundation
			- Exploratory Data Analysis (EDA) & Feature Engineering			
			- Statistics for Data Science			
			- Taxpayer Behaviour Modelling Foundations			
			- Data Ethics & Model Accountability Foundations			
		ii. Statistical & Predictive Modeling	- Regression, Classification & Clustering Techniques	200	High	Foundation
			- Predictive Models for Tax Evasion & Revenue Forecasting			
			- Machine Learning in Python			
			- Sector-Based Risk Scoring Modelling			
			- Bias Detection & Model Explainability Techniques			
		iii. AI & Data Science for	- AI for Tax Fraud Detection & Risk Assessment	200	High	Expansion

		Revenue Optimization	- Deep Learning for Revenue Optimization			
			- Building AI-Powered Tax Compliance Systems			
			- Neural Networks for Network-Based Fraud Rings			
			- AI for Real-Time Tax Intelligence			
			- Applied AI for High-Risk Sector Monitoring			
		iv. Applied Machine Learning	- Machine Learning Model Deployment (MLOps)	200	High	Foundation
			- Hyperparameter Tuning & Model Optimization			
			- Hands-on AI Model Building.			
			- Automated Model Monitoring & Drift Detection			
			- ML Lifecycle Governance & Auditability Standards			
		v. AI for Tax Compliance & Risk Management	- AI for Tax Gap Analysis & Fraud Detection	200	Medium	Expansion
			- NLP for Automated Compliance Checks			
			- AI-Driven Decision Support Systems			
			- Retrieval-Augmented Generation (RAG) for Tax Investigations			
			- LLM Finetuning for Audit & Enforcement			
			- Secure Deployment of AI Intelligent Assistants			
3	**Data Engineers**	i. Data Engineering & Pipeline Development	- Data Engineering with Python, SQL & Apache Spark	200	High	Foundation
			- Building ETL Pipelines for Large-Scale Data Processing			
			- Data Modelling & Warehouse Optimization			
			- Data Contracts & Tax Data Standardisation			

			- Enterprise Data Integration Standards for Tax Systems			
		ii. Distributed Data Processing & Storage	- Big Data Processing with Hadoop, Spark, and Kafka	200	High	Expansion
			- Optimizing Cloud Data Warehouses (Snowflake, BigQuery, Redshift)			
			- Streaming Data Processing & Real-Time Analytics			
			- Event-Driven Architectures for Compliance Triggers			
			- Resilient Architecture Design for High-Volume Tax Pipelines			
		iii. Cloud-Based Data Engineering	- Google Cloud, AWS, and Azure Data Engineering	200	Medium	Maturity
			- Serverless & Containerized Data Processing (Kubernetes, Docker)			
			- Data Lakehouse Architecture & Implementation			
			- Hybrid Cloud Designs for Sensitive Tax Data			
			- Secure Cloud Patterns for AI & Analytics Workloads			
4	**Platform Administrators & Database Experts**	i. Enterprise Database Administration	- Oracle Database Administration Certification	200	High	Foundation
			- SQL Server & PostgreSQL Administration			
			- Database Performance Tuning & Optimization			
			- Elastic Scaling for Real-Time Analytics Workloads			
			- High-Availability Design for Mission-Critical Tax Systems			
		ii. High-Performance Data Storage & Optimization	- Oracle Exadata Specialist Certification	200	Medium	Expansion
			- Indexing & Partitioning for Large Datasets			

			- Hybrid Cloud Database Management			
			- Vector Indexing & Embedding Storage Optimization			
			- Architecture for Low-Latency AI Retrieval Workloads			
5	**Data Governance & Compliance Experts**	i. Data Privacy & Compliance	- Certified Data Protection Practitioner (CDPP)	200	High	Foundation
			- GDPR, Data Protection & Compliance Laws			
			- Ethical AI & Bias Mitigation			
			- AI Governance & Responsible Automation in Tax			
			- Data Stewardship Practices for National Tax Systems			
		ii. Enterprise Data Management & Governance	- Certified Data Management Professional (CDMP)	200	High	Foundation
			- Data Governance Frameworks & Policies			
			- Data Security & Access Control Best Practices			
			- Metadata Management for AI & Semantic Search			
			- Enterprise Data Cataloging for Intelligent Tax Platforms			
6	**Project & Change Managers**	i. Process Optimization & Operational Efficiency	- Lean Six Sigma Green Belt & Black Belt Certification	200	High	Expansion
			- Process Automation & Optimization with AI			
			- Operational Analytics for Decision Making			
			- Automation Opportunity Mapping for KRA			
			- Value Realisation Tracking for Data & AI Initiatives			
			- Project Management Professional (PMP)	200	High	Foundation

		ii. Agile Project & Change Management	- Agile & Scrum for Data Teams			
			- Stakeholder Engagement & Change Management			
			- Leading AI-Driven Organizational Transformation			
			- Governance for Multi-Disciplinary AI Programs			
7	**Cloud & Infrastructure Experts**	i. Cloud Data Warehousing	- Snowflake, BigQuery, and Azure Synapse Analytics	200	High	Expansion
			- Optimizing Serverless Data Processing			
			- Multi-Cloud Data Architecture			
			- Cloud Cost Optimization for AI Workloads			
			- High-Security Cloud Designs for Tax Data Sovereignty			
		ii. Cloud Security & Compliance	- AWS Certified Security – Specialty	200	Medium	Maturity
			- Zero Trust Security Models for Data Protection			
			- Cloud Data Encryption & Secure Access			
			- Secure AI Deployment & Model Isolation Techniques			
			- Incident Response for AI-Integrated Cloud Environments			
8	**Data Security & Risk Management**	i. Cybersecurity for Data Analytics	- Certified Ethical Hacker (CEH) Training	200	Medium	Expansion
			- Threat Intelligence & Cyber Risk Mitigation			
			- Data Encryption & Secure Data Sharing			
			- Cyber Risk Modelling for Revenue Systems			
			- Security Hardening for AI and Analytics Platforms			
		ii. Risk & Fraud Analytics	- Fraud Detection & Anomaly Detection using AI	200	High	Expansion
			- Behavioural Analytics for Taxpayer Risk Assessment			

			- Automated Risk Scoring Models			
			- Graph Analytics for Fraud Networks			
			- Designing Proactive Fraud Intelligence Frameworks			
9	**Big Data & Real-Time Analytics Specialists**	i. Streaming Analytics & Real-Time Data Processing	- Kafka, Apache Flink, and Spark Streaming	200	High	Expansion
			- Developing Real-Time Dashboards & Alerts			
			- Tax-Specific Big Data Processing			
			- Real-Time Tax Intelligence Frameworks			
			- Real-Time Event Correlation for Compliance Monitoring			
10	**Product Owners & Business Leads**	i. Data-Driven Decision Making for Executives	- Executive Data Analytics & AI for Leadership	50	High	Foundation
			- AI & Digital Transformation in Tax Administration			
			- Strategic Decision Making with Data			
			- AI-Enabled Policy Simulation for EXCO			
			- Executive Oversight of National Tax Intelligence Platforms			

7 SECTION 4: DATA ANALYTICS CENTER OF EXCELLENCE FACILITY

7.1 Background

The facility will provide a collaborative and agile workspace equipped with modern infrastructure to support the effective functioning of the CoE.

7.2 Objective

The primary objective of this project is to set up a state-of-the-art Data Analytics CoE facility that fosters innovation, collaboration, and efficiency. The facility will cater to 300 members of staff and adopt an agile scrum working environment.

7.3 Scope of Work

The scope of work includes but is not limited to:

No.	Features	Minimum Requirements	Bidder's Response
1.	Ergonomic and agile workspace.	• Modular workstations with adjustable height and foot rest. • Ergonomic mesh office chair with an Adjustable height and foot rest. • Integrated green-spaces for improved air quality and cognitive comfort.	
2.	Movable screens for team collaboration.	• 30 Electronic Interactive Whiteboard 55" - 65" Touchscreen Smart Board with camera and mobile stand. • Wireless casting support. • Lockable wheels and sturdy bases • Minimum 4K resolution for displaying analytics dashboards and data visualizations	
3.	Working Pods	• 10 Enclosed working pods with smart glass technology fitted with sliding doors and glass walls made from smart film technology. • Motorized height adjustable desk and chair for each pod.	

No.	Features	Minimum Requirements	Bidder's Response
4.	Break-out areas.	• 4 Smart TV 65" to 110" with gaming capability. • Indoor games; pool table (1), dart board, foosball table (1), board games (10), stress balls. • Relaxation space with ergonomic lounge seating • Wellness corners with mats or ergonomic stretching aids • Informal collaboration furniture (low tables, soft seating)	
5.	Exhibition and Presentation room.	• Six 65"-110" interactive wall mounted multi-screen video array for high impact digital presentations. • 1 Central boardroom style table with executive chairs (15). • Soft seating and refreshment nook with couches (4) and a minimalist coffee table (2).	
6.	Momcube	• Freezer (3-6 cubic foot) for milk storage and hygiene equipment • Comfortable nursing rocking chairs and table	
7.	Kitchen area.	• Refrigerator (400L+), microwave, toaster, kettle, coffee machine. • Adequate seating and countertop space. • Water dispenser (10)	
8.	Boardroom	• Portable (10) teleconferencing speaker for remote collaboration. • Motorized drop-down screen. • Modular conference table that can be split/combined.	

No.	Features	Minimum Requirements	Bidder's Response
		• Executive orthopaedic leather boardroom chairs (20). • Integrated audio system and wireless presentation capability.	

8 SECTION 5: DATA GOVERNANCE AND QUALITY AUTOMATION

8.1 Background

KRA seeks to automate data governance and quality monitoring to ensure alignment with recognized best practices in data governance, data security, and regulatory compliance.

8.2 Objective

The primary objective of this establish data governance frameworks, automated data cleansing, and validation mechanisms. KRA seeks to establish a high trust data foundation by implementing "**always-on quality monitoring"** that identifies anomalies (volume drift, schema changes, null spikes) in real-time before they reach downstream BI dashboards or ML models. KRA also seek to automate the Metadata Catalog and Data Lineage so analysts can easily find, understand, and verify the provenance of data assets.

8.3 Scope of Work

The solution must have the following minimum features:

No.	Features	Minimum Requirements	Bidder's Response
1.	Governance	Active Metadata: Auto-harvesting from warehouses (Snowflake/BigQuery), BI tools (Tableau/PowerBI), and ETL pipelines.	
		Lineage: Automated, column-level visualization showing data provenance and downstream impact.	
		Cataloging: Business glossary with the ability to link technical assets to business definitions.	
2.	Data Quality Automation	Out of the box Profiling: Automatic calculation of nulls, uniqueness, and distribution upon connection.	
		ML Anomaly Detection: AI-driven alerts for volume drift, schema changes, and unexpected value shifts.	
		Self-Healing/ Auto Remediation: Suggested fixes or automated workflows to quarantine non-compliant data.	
		Custom Rule Engine: Allow users towrite custom DQ rules using SQL or a No-Code interface	
3	Technical	Connectivity: Native, low-latency connectors for your specific data stack and cloud providers.	
		Integration: Robust APIs to push alerts into Slack, Jira, or PagerDuty.	
		Deployment: Options for SaaS, VPC, on premise or hybrid to satisfy data residency requirements.	

No.	Features	Minimum Requirements	Bidder's Response
		API- First Design: Robust REST APIs to allow for integration	
		Scalability: Ability to handle millions of tables/assets without performance degradation.	
4.	Usability	Search and Discovery: Intuitive, "Google-like" search for data discovery and no-code rule building.	
		Collaboration: Capability for data owners to "certify" datasets and users to leave ratings/comments.	
		No-Code Interface: Allow business users set up quality checks or update the glossary without writing code	
		Alerting Workflows: Route quality incidents to the specific data owner via integrated ticketing	
5.	Security	PII Discovery: Automated tagging of sensitive data (DPA) using pattern matching to automatically tag sensitive data	
		Access Control: Granular RBAC/ABAC and full audit logs of all metadata changes.	
		Audit Trails: full history of changes to metadata or policies	

9 METHODOLOGY (DEMONSTRATE IN-DEPTH UNDERSTANDING OF THE TOR /REQUIREMENT)

The bidders are required to demonstrate in-depth understanding of this project, their capability to undertake it, and how they propose to do so by giving detailed clause-by-clause responses to these service requirements, and how they propose to deliver the respective service components as stipulated under the scope of work.

The bidders are also required to explain any business and/or technical benefits and implications of their proposed approach.

9.1 Vendor Evaluation/ Work plan/ Implementation Schedule/SLA

The bidders are required to provide a project work plan composed of the following:

a) A log frame illustrating project stages and a high-level work plan.
b) A detailed project implementation plan indicating expected project component activities to accomplish the assignment, the party responsible for each, indicative duration for each activity, milestones, and expected deliverables.

The bidders are notified that this work plan, after necessary review with KRA, will form part of the contract with the successful bidder.

9.2 Vendor Evaluation

#	Requirement	Evaluation Criteria	Max Score	Bidders' response
1.	Vendor Experience Demonstrate experience through previous execution of preferably three (3) projects in deployment of a big data platform of similar magnitude within the last five (7) years.	**2 Marks for each project** Bidder MUST submit: a) Recommendation letters/certificate of completion for each relevant project cited which should be supported by copies of signed. **(1 mark for each project)** b) Executed Contracts and copies of signed LSOs. **(1 mark for each project)** c) Contacts: postal address, telephone and email of the contact person. d) A brief description of the project delivered	6	
2.	Technical staff qualifications. Five (5) staff serving as technical leads in the major	**15 Marks** (1 Mark for degree for all 5 staff and 2 Marks for valid professional certification for	15	

#	Requirement	Evaluation Criteria	Max Score	Bidders' response
	project roles, with the following academic and professional qualifications: a) Academic Qualifications: A minimum of Relevant University Degree or Diploma. (Computer Science, IT, electronics or related fields) b) Professional qualifications for the following roles: • Data Engineer • Data Architect • DevOps Engineer • Data Analyst • Project Manager	data engineer, dev-ops engineer and project manager) Bidders MUST attach the CV of each staff supported by copies of Academic and professional certificates.		
3.	Staff Relevant experience: Minimum 2 years Each qualified staff (refer to clause 2 above) should have experience in implementation, support and maintenance of a big data platform of similar scale and magnitude and serving as a lead in their respective roles.	Staff Relevant experience • Over 3 years– 3 Marks for each qualified staff • 2-3 years – 2 Marks for each qualified staff • Less than 1 Year - 0 Marks Note: Bidders MUST submit a copy of the CV for each staff clearly indicating the years of experience in implementing and supporting the specific solution and the sites supported.	15	

#	Requirement	Evaluation Criteria	Max Score	Bidders' response
4.	Technical Approach and Methodology Bidder MUST demonstrate a good and clear understanding of KRA's Requirements. They MUST propose an approach/methodology and a work plan to capture the requirements and ensure they are comprehensively addressed in the proposed solution	Bidders to demonstrate/provide evidence of a clear and detailed understanding of the solution, including: a) Project delivery Approach/Methodology for implementation and support of the solution – 1.5 Marks b) Work plan (Bidder MUST provide a three (3) year work plan for implementation and support for the solution - 1.5 Marks	3	
	Total Score		39	
	Cut-off score is 33 marks			

Financial Requirement

- **N/B: Bidders to provide a breakdown of how they have arrived at the total cost**
- **Grand Total Cost –To be carried Forward to the FORM FIN 2 Summary of Costs**

9.3 Scoring Criteria Summary

No	Description	Marks	Cut Off Score	Bidders Score
1	**Mandatory Requirements**	P/F		
2.	**Vendor 1**	100	80	
3.	**Vendor 2**	39	33	
4.	**Platform Demo**	10		

9.4 Risks and Mitigation Plan

Introduction

The implementation of the Data Analytics Center of Excellence (CoE) involves multiple dimensions, including technical, financial, and managerial aspects. Identifying potential risks early and establishing a mitigation plan ensures that challenges are proactively managed, reducing disruptions and ensuring the project's success. The following table outlines key risks, their potential impact, and mitigation strategies:

Risk Category	Potential Risk	Impact	Mitigation Plan
Technical Risks	Delays in setting up the data analytics platform due to integration challenges with existing systems.	Project timeline extension, reduced system functionality.	Conduct a thorough system compatibility assessment before implementation. Engage technical experts for seamless integration.
	Data security breaches and compliance issues.	Exposure of sensitive data, regulatory penalties, reputational damage.	Implement robust security policies, access controls, and encryption measures. Conduct regular security audits and compliance checks.
	Poor data quality and inconsistency across sources.	Unreliable analytics, poor decision-making.	Establish data governance frameworks, automated data cleansing, and validation mechanisms.
	Performance issues in analytics processing due to inadequate infrastructure.	Slow insights generation, inefficient operations.	Conduct capacity planning before deployment. Optimize infrastructure and adopt scalable solutions.
Financial Risks	Budget overruns due to unforeseen costs.	Project delays, inability to complete all planned activities.	Conduct detailed cost-benefit analysis, include contingency funds, and monitor expenditures regularly.
	Inadequate funding for long-term sustainability.	Difficulty in maintaining and upgrading analytics capabilities.	Develop a long-term financial plan, secure multi-year funding commitments, and explore alternative funding sources.
Managerial Risks	Resistance to change from internal stakeholders.	Delayed adoption of analytics-driven decision-making.	Implement a structured change management strategy, provide awareness training, and engage key stakeholders from the beginning.
	Skills gap among staff leading to underutilization of the CoE.	Reduced effectiveness of the CoE, poor return on investment.	Develop comprehensive capacity-building programs, mentorship initiatives, and continuous training sessions.
	Lack of clear governance and accountability structures.	Inefficient decision-making, misalignment of goals.	Establish a strong governance framework with clear roles, responsibilities, and performance monitoring mechanisms.

Risk Category	Potential Risk	Impact	Mitigation Plan
	Difficulty in recruiting and retaining top analytics talent.	Inconsistent project execution, knowledge loss.	Offer competitive compensation, career development opportunities, and a dynamic work environment.
Operational Risks	Disruptions in facility setup and IT infrastructure deployment.	Delayed project implementation, increased costs.	Conduct risk assessments before procurement and installation, have backup plans, and engage reliable service providers.
	Inconsistent monitoring and evaluation of project milestones.	Delayed identification of challenges, inefficiencies.	Implement a robust project monitoring framework with periodic evaluations and feedback mechanisms.