BIG DATA PLATFORM FOR TELECOMMUNICATIONS TRAFFIC CONTROL IMPROVEMENT

Terms of reference

Maputo, November 2023

Contents

1. Introduction

The Regulatory Authority for Communications in Mozambique (INCM) aims to regulate, supervise, monitor, sanction, and represent the postal and telecommunications sectors.

INCM intends to implement a new Big Data solution to address current challenges. This implementation will be a large-scale project with the goal of improving efficiency in its operations, including (1) Data Collection, (2) Data Storage, (3) Data Processing and Transformation, (4) Data Analysis, (5) Visualization and Reporting, (6) Anomaly Detection, (7) Compliance and Regulatory Reports, (8) Security, (9) Scalability and Performance, and (10) Compliance Auditing.

The purpose of this terms of reference is to establish the requirements and expectations of INCM for the hiring of a specialized company to implement modern Big Data platforms in telecommunications. Therefore, the proposal should include technical specifications for hardware (servers, switches, SANs), system and database installation and configuration, separately, as well as maintenance and support.

2. Objectives

2.1 General Objective

INCM intends to hire a specialized company to implement Big Data solutions in Telecommunications to efficiently monitor the generated traffic, obtain valuable insights from the collected data for informed decision-making, and ensure regulatory compliance under the new Regulation for Telecommunications Traffic Control, Decree 38/2023 of July 3, 2023.

2.2 Specific Objectives

The specific objectives of this terms of reference are as follows:

1) **Data Collection**:
   - 1.1 Collect data from various sources, such as network elements (NE), call detail records (CDRs), Internet Protocol Detail Records (IPDR), signaling data, and subscriber equipment or devices;
   - 1.2 Use real-time data streaming technologies and batch files to ingest data as it is generated.

2) **Data Storage**:
   - 2.1 Store the collected data in distributed, scalable, and fault-tolerant storage systems, such as Hadoop HDFS, Amazon S3, or Azure Data Lake Storage.
   - 2.2 Use NoSQL databases like Apache Cassandra to store and query large volumes of data.

3) **Data Processing**:
   - 3.1 Apply batch and real-time (stream) processing techniques to clean, transform, and enrich raw data.

3.2 Use technologies such as Apache Spark for data processing and Apache Kafka for real-time processing.

4) **Data Analysis**:

4.1 Implement advanced analytics and machine learning algorithms to detect anomalies, predict network issues, and identify trends.

4.2 Use tools like Apache Flink or TensorFlow for real-time analytics.

5**) Data Visualization and Reporting**:

5.1 Develop dashboards and reporting tools to provide regulators with a user-friendly interface to monitor network performance and generated traffic.

5.2 Use data visualization tools or libraries such as D3.js, Tableau, or Power BI to create informative tables and charts.

6) **Fraud Detection**:

6.1 Apply machine learning models for anomaly detection to identify unusual patterns or behaviors in network traffic that may indicate fraud, security breaches, or network congestion.

6.2 Explore techniques like clustering, classification, and outlier detection.

7) **Compliance and Regulatory Reports**:

7.1 Automate the generation of compliance reports to assess whether telecommunications operators are complying with regulatory standards, which may include verifying applied tariffs and adherence to service level agreements (SLAs).

8) **Security**:

8.1 Incorporate security measures to protect call detail records and other confidential data and ensure regulatory compliance, including encryption, access controls, and data protection regulations.

9) **Scalability and Performance**:

9.1 Ensure that the Big Data solution can scale horizontally to handle the increasing volume of data generated by telecommunications networks.

9.2 Monitor system performance and optimize as needed.

10) Use of OSINT Tools for Cybersecurity Data Analysis.

3. Technical Proposal

The contracted company should provide a detailed description of the proposed technical solution, including technical specifications and necessary diagrams, considering the following aspects:

3.1 Infrastructure

- Describe the hardware and software infrastructure required to implement the Big Data solution. This will be an on-premises solution;
- Detail the scalability and redundancy features to ensure robust and reliable data processing and storage; and
- Provide information about Data Centers, including considerations regarding data residency and security compliance.

3.2 Data Collection

- Specify data collection mechanisms, including data sources and data ingestion processes;
- Provide details on data capture frequency, such as real-time streaming, batch processing, or both; and
- Explain how data will be validated, cleaned, and transformed for analysis.

3.3 Data Processing and Analysis

- Describe the data processing layer architecture, including technologies like Apache Spark, Hadoop, or streaming platforms like Apache Kafka or Apache Flink.
- Describe the algorithms and techniques to be used for traffic analysis, predictive modeling, and anomaly detection; and
- Highlight how data pipelines will be designed and managed to efficiently handle large volumes of data.

3.4 Data Storage and Management

- Specify the database systems or storage solutions to be used for storing raw and processed data.
- Explain data retention policies, data compression techniques, and data partitioning strategies for efficient storage; and
- Address data security measures, such as encryption, access control, and auditing.

3.5 Data Visualization and Reporting

- Provide details on the tools and frameworks that will be used to create interactive dashboards and reports;
- Explain how INCM technicians will access and interact with the visualization layer to monitor generated traffic; and
- Highlight any real-time reporting features for immediate decision-making.

3.6 Predictive Analysis

- Describe the machine learning models and algorithms to be used for predictive analysis, such as network congestion prediction or fraud detection; and
- Explain how model training and retraining will be managed to ensure accuracy over time.

3.7 Integration

- Detail integration points with the telecommunications network infrastructure, regulatory systems, and existing databases; and
- Specify any APIs or connectors that will be developed to facilitate continuous data exchange.

3.8 Performance and Scalability

- Provide performance benchmarks and scalability test results to demonstrate the solution's ability to handle growing data volumes; and
- Detail any load balancing mechanisms and automatic scaling.

3.9 Disaster Recovery and Redundancy

- Explain the disaster recovery plan, including backup and failover mechanisms; and
- Describe how data will be replicated across geographically dispersed locations for redundancy.

3.10 Maintenance and Support

- Specify ongoing maintenance and support services, including software updates, bug fixes, and issue resolution; and
- Describe the Service Level Agreements (SLAs) for technical support and issue resolution.

3.11 Data Privacy and Compliance

- Explain how the solution will adhere to data privacy regulations, including data anonymization and compliance with telecommunications regulatory standards; and
- Provide documentation regarding compliance with relevant industry standards.

3.12 Training

Detail any training programs or materials that will be provided to regulators (INCM technicians) or the internal team for the use and maintenance of the solution.

In addition, companies interested in bidding should submit a comprehensive proposal that includes the following information:

1) Previous experience in big data projects in telecommunications.

2) Work methodology.

3) Detailed description of the proposed solution.

4) Implementation plan and complete project schedule.

5) Detailed resumes of the technical team that will execute the project.

6) References from similar projects completed.

7) Detailed budget, including all project-related expenses.

4. Data Analysis for Regulatory Compliance

We will highlight some necessary analyses for regulatory compliance purposes, namely:

1. **Telecommunications Market Overview**: Obtain a clear view of the market, including the actual subscriber numbers by operator, province, and district, as well as network usage levels and revenues (usage & revenue) by region, operator, technology, service category, and promotions, including the devices used.
2. **Revenue Overview**: Monitor daily revenues from Voice, SMS, and Data, as well as the purchase of packages and top-ups.
3. **Control National and International Traffic**: Monitor all interconnection, national and international traffic, both voice and SMS, and detect voice and SMS bypass frauds.
4. **Subscriber Proper Identification Control**: Ensure that all operators properly verify their subscribers' records.
5. **Commercial Pressure Control**: Verify if operators adhere to rules regarding commercial pressure (applied tariffs, time intervals, volumes, data privacy compliance, etc.).
6. **Network Coverage and Expansion Control**: Verify whether operators have deployed cells and 3G/4G services as planned and provide quality service to subscribers.
7. **OTT Analysis**: Evaluate the subscriber base of each internet giant (Netflix, Whatsapp, Facebook, etc.) to estimate potential opportunities for tax collection.
8. **Used Device Analysis**:
    a. Analyze the subscriber base by device type, capabilities (2G/3G/4G…).
    b. Detect new and unused devices.
    c. Analyze stolen devices.
    d. Detect uncertified devices and cross-reference them with the equipment import database.

Competing companies are free to propose any other relevant analyses in the telecommunications market based on their own experiences and research. However, it is crucial that they are aware of the relevant regulations and ensure compliance when conducting their analyses.

5. Open-Source Intelligence (OSINT) Tools

The contracted company should also include in its proposal the adoption of Open-Source Intelligence (OSINT) tools for collecting and analyzing information from open sources available on the internet. Some of the OSINT tools that may be useful for our activities include:

1. **Shodan**: It is an internet-connected device search tool that can be used to identify vulnerabilities in devices and network systems. The tool can also be used to identify devices that may have been used by hackers.
2. **Maltego**: It is an open-source intelligence tool that can be used to gather information from open sources, such as social networks, public records, and news websites. The tool can be used to create user behavior profiles and identify potential cyber threats.
3. **SpiderFoot**: It is an open-source reconnaissance tool that can be used to gather information about IP addresses, domains, and usernames. The tool can also be used to identify vulnerabilities in systems and applications.
4. **OSINT Framework**: It is a set of open-source intelligence tools that can be used to collect information from open sources, such as social media, discussion forums, and public directories. The tool can be used to identify potential cyber threats and create user behavior profiles.

These OSINT tools can be used in conjunction with other data analysis techniques, such as network traffic analysis, user behavior analysis, and vulnerability analysis, to enhance cybersecurity in the telecommunications industry. The contracted company may suggest other useful tools for the activities of the Telecommunications Traffic Control Unit.

6. Project Deliverables

The contracted company must provide the following project deliverables:

1) Project Plan:

1.1 A detailed project plan describing key milestones, timelines, and dependencies.

1.2 A project schedule that includes the start and end dates of each project phase.

1.3 Work Breakdown Structure (WBS) showing the task and activity division.

2) Data Collection Results:

2.1 Data collection structure design document.

2.2 Data ingestion scripts or connectors for batch and real-time data collection.

2.3 Data validation and cleaning procedures.

3) Data Processing and Analysis Results:

3.1 Data processing pipeline architecture and design documentation.

3.2 Implementation of data processing and analysis algorithms.

3.3 Performance benchmarking results.

4) Data Storage and Management Products:

4.1 Data storage architecture and schema design.

4.2 Data retention policy documentation.

4.3 Backup, recovery, and data retention procedures.

5) Data Visualization and Reporting Results:

5.1 Interactive dashboards and report templates.

5.2 User manuals for using the visualization tools.

5.3 Examples of generated reports and dashboards.

6) Predictive Analysis and Optimization Results:

6.1 Implemented machine learning models and algorithms for predictive analysis.

6.2 Reports on predictive analysis results, including network congestion predictions or fraud detection outcomes.

6.3 Recommendations and action plans for network optimization.

7) Integration Results:

7.1 Integration documentation detailing how the solution connects to the telecommunications network infrastructure and existing regulatory systems.

7.2 APIs or connectors developed for data exchange.

7.3 Results of successful integration testing and validation.

8) Regulatory Compliance and Security Results:

8.1 Documentation demonstrating compliance with regulatory standards and data privacy laws.

8.2 Audit reports and evidence of data security measures.

8.3 Techniques used for anonymizing sensitive data.

9) Performance and Scalability Results:

9.1 Performance testing results and reports.

9.2 Documentation describing how the solution scales to handle increasing data volumes.

9.3 Recommendations for optimizing performance and scalability.

10) Disaster Recovery and Redundancy Results:

10.1 Disaster recovery plan and documentation.

10.2 Evidence of successful backup and recovery testing.

10.3 Documentation of redundancy measures and their effectiveness.

11) Maintenance and Support Products:

11.1 Ongoing support services, including helpdesk contact information.

11.2 Software updates and patches delivered as needed.

11.3 Bug fixes and issue resolution reports.

11.4 **First 3 Years - Knowledge Transfer**: INCM technicians should be in a learning and development phase.

11.5 **Following 2 Years - Independence with Minimal Supervision**: INCM technicians should be in a position to perform their functions more independently but still with minimal supervision.

12) Training Results:

12.1 Training materials and documentation.

12.2 Training sessions or workshops for regulators or internal staff.

12.3 Certification or acknowledgment of completed training.

13) Documentation and Knowledge Transfer:

13.1 Comprehensive project documentation, including design documents, manuals, and architecture diagrams.

13.2 Knowledge transfer sessions to ensure the INCM team can effectively maintain and operate the solution.

14) Final Project Report:

14.1 A comprehensive final project report summarizing project achievements, challenges, and lessons learned.

14.2 Recommendations for future improvements or enhancements.

7. Financial Proposal

The company to be contracted must demonstrate the financial capability to implement all proposed solutions and meet the objectives outlined in this terms of reference.

Furthermore, competing companies must include in their financial proposal the costs of all licenses required for the project's operation over a period of 3 years.

8. Requirements

The following requirements are valid:

a) The company must demonstrate proven experience in designing and implementing data processing systems in the telecommunications sector;

b) Three (3) endorsement letters proving experience in supplying and implementing similar projects in the last three (3) years;

c) Description of the project team, including CVs of the Team Leader. A team with national consultants will have an advantage;

d) Present an organizational chart of the project team;

e) A statement from the competitor itself, proving the levels of knowledge of the professional and technical team available for the execution of the contract, accompanied by their respective CVs and technical training certificates related to the contract's objectives;

f) Have proven experience in implementing Big Data solutions in telecommunications;

g) Have a highly qualified technical team with expertise in data science, data analysis, software engineering, and telecommunications;

h) Have knowledge of Big Data tools, such as Hadoop, Spark, NoSQL, among others;

i) Have the capacity to manage large-scale projects;

j) Be able to deliver the project within the defined timeframe and budget;

k) Have a track record of successful projects in implementing Big Data in telecommunications;

l) Possess relevant certifications and partnerships with other technology companies; and

m) Have the capability to provide ongoing technical support.

9. Training

For INCM, training is a crucial factor in projects of this nature. Therefore, adequate planning and knowledge transfer (Know-How) to INCM technicians are required to manage the platform and ensure its full functionality.

Hence, it is expected that during the first three years, INCM technicians should be in a phase of learning and development. After the initial three years of learning and development, INCM technicians should be in a position where they can perform their functions more independently, albeit with minimal supervision.

10. Evaluation Criteria

For this project, the evaluation criteria will be based on a combination of the technical component and the financial component, with corresponding weights of 75% and 25%, respectively.

Therefore, the evaluation process will follow two main phases:

1. Technical evaluation; and

2. Financial evaluation.

Companies that meet a minimum of 85% of the technical requirements' score will proceed to the financial evaluation.

Nº	Evaluation Aspect	Score
1	**Technical Proposal (Weight 75%)**	
	• Compliance with the requirements of the presented terms of reference	**100**
	• Experience in the field of consulting for the design and implementation of big data solutions	**25**
	• Presentation of a methodology and work plan that addresses the terms of reference requirements	**20**
	• Experience and CV of the consulting team, including nationals	**40**
	• Presentation of reference/endorsement letters, including 3 completed projects in the last five years	**10**
	• **Consultancy execution timeline** (Shorter execution time corresponds to the maximum score)	**5**
2	**Financial Proposal (Weight 25%)**	
2.1	Presentation of all the requirements	**10**
2.2	Lowest rated price	**90**