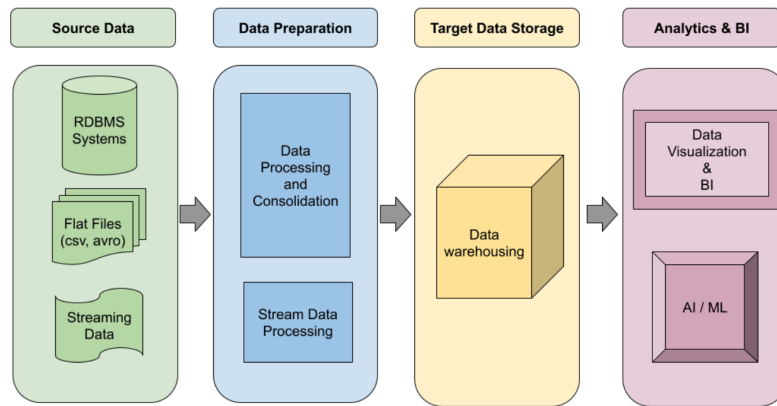


# A One-Pager Software Requirements Specification (SRS) Format

**Project Name:** 5G Data Analytics - Data Pipelining

## **Project Overview:**

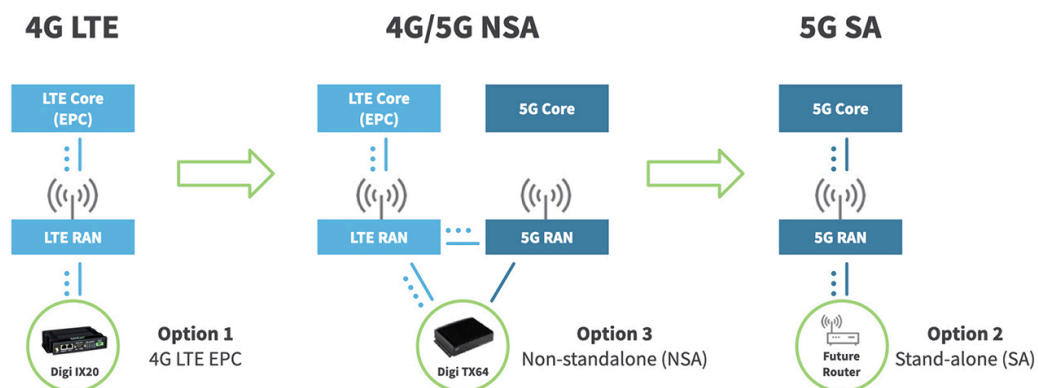
- **Purpose:** Incoming raw data that will be used for the project comes in unorganized pieces that must be organized using data pipelining methods in order for the application to successfully read and operate on the data it is given. Data pipelines allow for the automation, movement, processing, and preparation of raw data so it's ready for use in applications like reports, dashboards, machine learning or decision-making. These uses will play a pivotal role in the development of the application as a whole.
- **Target Audience:** Data pipelines are primarily used by developers (specifically data engineers) and their applications in order to provide users or customers a viable experience when using an application. Fetching raw data, organizing it, and presenting it all requires the use of data pipelines - making it a crucial function in all applications dealing with raw data inputs.
- **Scope:**
  - **Included:**
    - Real-time and batch data pipelines
    - Key performance indicator (KPI) calculations
    - Alert generation
    - Data visualization dashboards
  - **Excluded:**
    - Physical network management or hardware



## Functional Requirements:

- **Data Ingestion:**

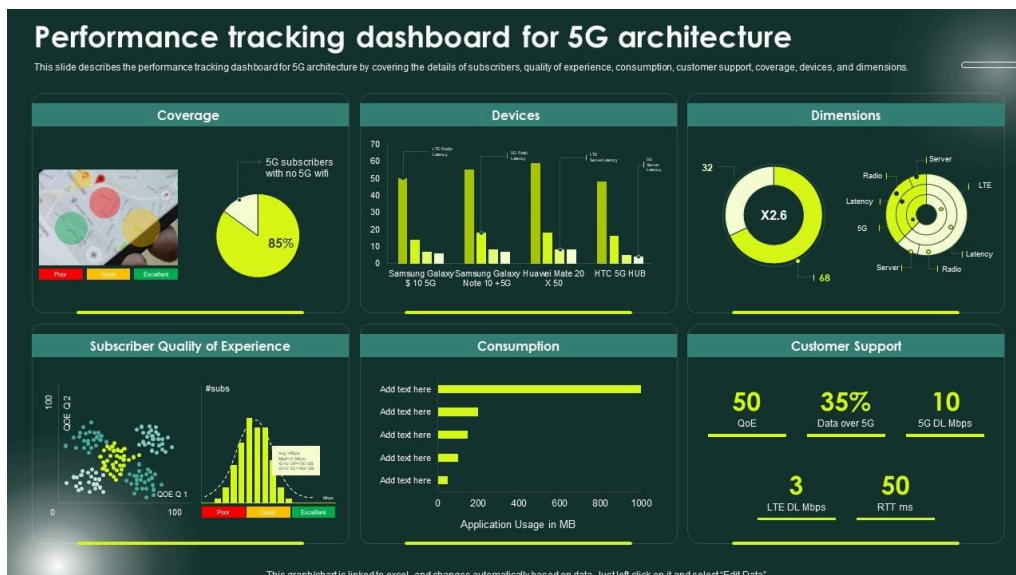
- Collect performance metrics from APIs using real-time data streaming platforms like Amazon Kinesis or Apache Kafka.
- Support both real-time and batch data ingestion methods.
- Implement a batch data input procedure using Apache Spark or AWS Glue to manage old data or data that cannot be collected in real-time.
- Ensure that the ingestion layer can handle several data types like JSON and CSV and support batch and real-time data sources.



- **Data Transformation:**

- To clean, aggregate, and filter incoming data in real time, use real-time data processing frameworks like Amazon Kinesis Data Analytics or Apache Flink.

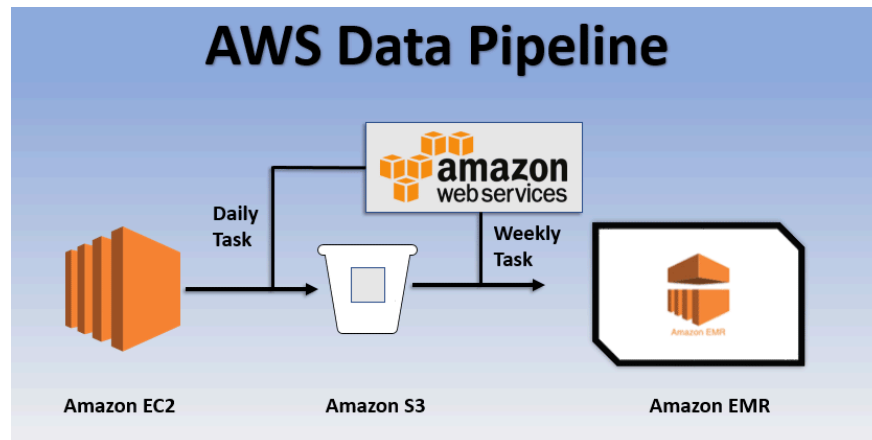
- Use a batch data processing framework like Apache Spark or AWS Glue to manage old data and calculate key performance indicators (KPIs) like latency, throughput, and dependability.
- Use data transformation logic to clean, aggregate, and filter the network data according to the requirements,
- Calculate the KPIs and save the findings for analysis and visualization in a data storage.
- **Alerting:**
  - Monitor the calculated KPIs continuously and send out trigger alerts when the values exceed the specified thresholds.
  - Use a notification system like Slack, email, or SMS to notify the appropriate parties.
- **Data Visualization:**
  - Create a web-based dashboard to track and view the alerts in real-time and the KPIs.
  - Use data visualization programs like Grafana or Kibana to create a personalized dashboard.
  - Make sure the dashboard can be accessed and used on a variety of platforms, including desktop and mobile.



## Non-Functional Requirements:

- **Performance:**

- Response time should be a maximum latency of 100ms to ensure real-time analytics - including extraction, transformation, and loading times.
- Handle high volumes of incoming 5G data to be analyzed & processed.
- Utilize AWS EC2 services for auto-scaling capabilities for resource monitoring and allocation from high-volume data.



- **Security:**
  - Ensure data encryption in transit and at rest
  - Implement role-based access control (RBAC) for authorized users
- **Usability:**
  - The web-based dashboard should have an easy-to-use user interface for selecting time ranges, key performance indicators (KPIs) and alert levels
- **Reliability:**
  - Achieve 99.9% system uptime with fault-tolerant architecture
- **Maintainability:**
  - Ensure the system follows modular design practices, allowing easy updates and scaling

#### **Assumptions and Dependencies:**

- The pipeline must be able to process real-time data with a delay of fewer than 100ms

#### **Acceptance Criteria:**

- Dashboard displays all three major KPIs (latency, throughput, and reliability) upon authentication
- Pipeline maintains an uptime of 99.9% after deployment

- Ingestion of real-time data must be reflected by a change in displayed KPIs within 100ms
- Data is saved in a long term storage solution and accessible upon authentication
- JSON and CSV data types are supported for ingestion