

Shivam Runthala

📍 Navi Mumbai, India

☎ +91 8286746581

✉ Shivam.Runthala@outlook.com

in linkedin.com/in/runthalashivam

📄 PROFESSIONAL SUMMARY

Software Engineer with 3+ years of experience in Big Data. Have hands-on experience with Distributed Systems like Spark, Hive, Kafka & NiFi. Also have worked with Containerization tools like Docker & Kubernetes.

🧠 SKILLS

Big Data

Spark, Hive, Kafka, NiFi, Hadoop, HDFS, SQL

Languages

Java, Scala, Python

Tools

Git, IntelliJ

Other

Docker, Kubernetes, Linux/UNIX, Shell Scripting

📁 PROFESSIONAL EXPERIENCE

Reliance Jio

July 2017 – present | Navi Mumbai

Software Engineer - Big Data Engineer

• Data Ingestion & Analytics

- Designed and developed ingestion and analytics pipelines, majorly for the **FTTX** Business of the organization, leveraging tools like **NiFi**, **Kafka**, **Spark**, and **Hive**.
- Ingestion pipelines were built to handle **real-time streaming event data** coming from Devices sold under Jio's FTTX business like STB(Set Top Box) & also JioPhone. Sources for these data streams were systems like **Solace** and **HiveMQ**.
- Build various **KPIs** on top of these data sources using Spark and Hive to enable the business to get a full overview of User's usage patterns, experience, and the performance of the device itself.
- Worked on defining the **Schema Evolution** strategy for use cases where schema changes are frequent.

• Ingestion Framework

- Built **Data Validation module** for the ingestion framework used within the team, the module logs bad input into an error queue, enabling a feedback loop with the Source for data quality.

• Spark & Kafka Metrics Collection for Data Validation

- Designed and developed a pluggable Metrics Capture module for the Spark Ingestion Workflows running on the cluster. This module captures metrics & errors at different stages of the workflow without adding significant overhead.
- This enabled the Operations team to visualize how the data is flowing within the system and allow end-to-end **comprehensiveness checks & alerts**.

• Internal Tool For Exploring Data

- Designed and developed an Enterprise Application with **AD Integration** for Analysts/Developers to deliver a scalable Notebook interface to Spark in Kerberized Hadoop Cluster through **Livy**.
- Added Airflow support for automated execution of Spark Notebooks. Containerized the application using **Docker** and deployed it in a **Kubernetes** cluster.
- Multiple Business Units within the organization are using the application to analyze on top of Big Data.

• Metrics Capture for Hive REST API

- Designed and developed the component for the Hive REST API to capture **Real-Time App and Client level metrics** to monitor the health and usage of the API. Metrics were captured on **Prometheus** and then visualized using **Grafana**.

• Apache Superset

- Designed and developed the following components for Apache Superset to make it **Enterprise Ready**:
 - Authorization using LDAP Groups and mapping of **LDAP** Groups with internally defined Superset roles.
 - Seamless **Kerberos Integration** for connection with Kerberized Cluster, enabling automatic initialization and renewal of Kerberos ticket.

🎓 EDUCATION

B.E. (Computer Engineering)

M.G.M CET (Mumbai University)

July 2013 – May 2017 | Navi Mumbai, India