

# Rishabh Chaudha

**Data Engineer** (August 2016 – Present)

**Email:** [chaudha.rishabh@gmail.com](mailto:chaudha.rishabh@gmail.com) **Phone:** +91-8074282973  
**LinkedIn:** <https://www.linkedin.com/in/rishabh-chaudha-48957150/>  
**Github:** <https://github.com/rishabhchaudha>

## Profile Summary

- 4+ years of experience in analysis, design, development, customization, and Implementation of Big Data Applications.
- Worked on a number of strategically important technology projects in various business domains.
- Demonstrated expertise in Spark & Hadoop .
- Expertise in developing Real-Time Streaming Solutions using Spark Streaming & Kafka.
- Competence in Collecting, aggregating and loading log data from multiple sources into HDFS.

## Skills

**Big Data Ecosystems:** Spark, Hadoop  
Spark Streaming, Kafka, MapReduce,  
HDFS, Hive, Sqoop

**Cloud :** AWS

**IDE :** Eclipse, IntelliJ

**Tools:** SyncSort DMX Express

**Schedulers :** Autosys, Oozie

**Language:** Python, Scala, JAVA

**Databases :** Teradata , SQL

## Experience

### Data Engineer

**Cognizant Technology Solutions (October 2019 – Present)**

#### Responsibilities

- Developed Streaming Pipelines using D-Streams and Structured streaming in Scala & Python.
- Used Kafka Streams to Configure Spark streaming to get information and then store it in HDFS.
- Work experience with cloud infrastructure like Amazon Web Services (AWS).
- Created Data Pipelines as per the business requirements and scheduled it using Oozie Coordinators.
- Involved in Migration of existing Legacy systems to Spark Distributed System Environment.
- Conversion of Oracle procedures into Spark SQL.

### Hadoop Developer

**Tata Consultancy Services (August 2016 – September 2019)**

#### Responsibilities

- Involved in creating an enterprise level consolidated environment of deposit data that presents a robust history of deposit behaviour.
- Development of ETL application as per client requirement using Spark, MapReduce, Hive, DMX and Unix script.
- Created Snapshot process for fast changing daily account data and CDC process for slow changing account data.
- Developed the ETL application for third party vendor data using Python and Spark.
- Defining job flows in Autosys and Load and transform large sets of structured data in Teradata.
- Migrated the existing process of applications into Pyspark and implemented new modules.
- Handled a team of 6 members and successfully delivered a project.
- Knowledge Transfer to new joiners at Hyderabad and Chennai locations for relevant business application logic.

## Achievements

- Certified as Data Scientist at Organisation level Data Science Summit 2020.
- Received star of the month and service commitment awards in previous organization.
- Received numerous client appreciations for flawless project implementations.
- Participated in social activities such as purpose for life, which motivates each individual to contribute towards societal well-being.

## Education

<b>B.E.(Bachelor of Engineering) in Computer Science</b>	2012-2016	Rajiv Gandhi Technical University	7.13
<b>H.S.C</b>	2012	C.B.S.E	82 %
<b>S.S.C</b>	2010	C.B.S.E	9.4

## **Project Details**

### **AIOPS – Data Lake**

Ingestion of heterogeneous data from many different sources about all components of the IT environment — networks, applications, infrastructure, cloud instances, storage and more to hive and Building Anomaly models to predict the outages.

### **MLF - Merrill Lynch Foreign**

Creating an enterprise level consolidated environment of deposit data related with Foreign Accounts. Coding the CDC, Snapshot process, generation of surrogate key to uniquely identify each foreign deposit account and processing the data to Downstreams and Teradata for further analysis.

### **FDIC 370 Federal Deposit Insurance Corporation**

According to the new rules of FDIC, Designing of the ETL application that identifies the customers from an institution that are liable for the pre-defined insurance amount in case of failed institution. Handling calculations on party,account level and also on the basis of Ownership Right and Capacity

### **Ranger**

Profiling the consumption and invoice details data received from source systems with smart meter and IDR usage and aggregating it on brands, plans & contracts. Comparisoin of the data generated with the actual invoices generated by the vendors.

### **Third Party Sourcing**

Incorporation of the third party vendor data for the bank in the warehouse and handling calculations of the insurance amount in accordance with the FDIC 370 rules using Pyspark for ETL process.