

## Experience Summary

---

- ◆ **9+ years of IT experience** as Data Engineer, Developing, Support, Implementation of software applications which includes 7 years of experience in **Big Data using Spark, Scala, Python, Java, Hadoop, Hive, Snowflake, Apache Airflow**.
- ◆ Working as certified as **Quantexa Developer** with Scala, mainly involved in **Scoring, ETL, compounds, Pruning, Task Loading and Entity Resolution**.
- ◆ Have worked in Quantexa version **1.0, 1.6 and Latest on 2.3**
- ◆ Worked in **Scala** language using **RDD, Dataframes** and **Datasets**, using combine functional and oops programming.
- ◆ Have experience in **AWS** services like **S3, EMR, Glue, Athena, IAM, RDS, EKS etc.**
- ◆ Have developed a desktop application using **Java, Derby DB and SQL** on NetBeans.
- ◆ Written **Hive** queries for data analysis to meet the business requirements.
- ◆ Involved in writing data transformations, data cleansing using Spark.
- ◆ Working experience on the **Agile and Waterfall models** and have worked closely with clients.
- ◆ Proficient in IBM mainframes **PLI, JCL** and **DB2**.

## Profile Summary

---

- Currently working in **Barclays India Pvt Ltd** organization since March.2021 as **Data Engineer & Quantexa Developer** in Pune.
- Joined **Xoriant Solutions** in **JAN.2020** and worked for 1Year 3 Months as senior **Data Engineer** in Pune.
- Joined in **IBM PVT Ltd** in **April.2014** as fresher and worked for 5 years and 9 months as **Application Developer Big Data** for in Pune.

## Technical Skills

---

**Languages/Technologies:** Scala, Spark, Hive, Python, HDFS, YARN, Java, SQL, Linux

**Database:** Postgress, MySQL, DB2, Derby

**Defect Tracking Tools:** Rally Dev, Service Now, Jira

**Remote Connection Tools:** Putty, WinSCP

**IDE:** IntelliJ IDEA, NetBeans, RFT

**Tools:** AWS, BigData BigInsight, Cloudera

## Project Experience

---

**APR 2021– Till Date**

**Data Engineer**

**Project # -** Quantexa Application for Trades Data

**Description:** This is the financial crime project which is implemented in Qunatexa Framework. The raw data of customer, accounts , trades and external starts with CSV filed and get transformed and joined together for set structure in parquet file format, from this point by using Qunatexa features, the entities are resolved to identify the same business with distinct information's and connect then together on the Network view to visualise for the investigators and also generate the alerts(Froud trades) using the set of rules.

### Responsibilities:

- Understand the business and the Qunatexa application with fast pace.

- Created the ETLs for converting raw data and transformed, connected in parquet format.
- Written the scoring like document, Network, Fact and custom to implement to rules which are considered as fraud.
- Loading the Alerts generated for investigation on UI, along with its pruned network for investigators.
- Created Compounds and Attributes for Entity resolutions by understanding the data strongly with BA team.
- Developed the ETL for full and incremental loads and data transformation of data using Scala Spark.
- Lead the development team with full application support.
- Support team in technical aspects of project.
- Coordinated with BA and testing team to understand and gather the requirement.
- Report to senior management and sync understanding with Investigation team

**Environment:**

- Quantexa, Spark, Scala, AWS glue, S3, EMR, EKS, EC2, Unix

**JAN 2020– MAR 2021**

**Data Engineer**

**Project # - Data migration from SQL server to AWS, Snowflake on Apache Airflow (Survey Monkey)**

**Description:** The Project is to migrate the whole ETL which is created and schedule on Apache Airflow, which was writing results to SQL server. Job was to create the same ETL in Airflow and load the data Snowflake. In order to achieve this , we created our data Lake and staging in AWS s3 and then loaded to Snowflake. Also implemented new delta process to perform the incremental data load.

**Responsibilities:**

- Understand the business and current application end to end for Migrations
- Created and schedules new Airflow DAGs in python for ETL process.
- Developed the ETL for full and incremental loads and data transformation of data in snowflake using pySpark.
- Used Athena table for data visualizations
- Support team in technical aspects of project.
- Coordinated with BA and testing team to understand and gather the requirement.

**Environment:**

- Python, PySpark, Snowflake, SQLserver, Apache Airflow, AWS Athena, S3, Hive,

**DEC 2017– DEC 2019**

**Data Engineer**

**Project # - Genesis**

**Description:** The Project is to pull the data from different source systems to Big Data and apply data quality checks along with reports generation to client. It uses Sqoop, Hive for data ingestion and RDD & Spark SQL components to query and make some analysis based on customer requirement. This is done using spark jobs which runs daily, weekly monthly & quarterly.

**Responsibilities:**

- Developed Spark scripts by using Scala as per the requirement.
- Involved in the requirement gathering and team meetings.

- Used both Data frames and RDD in **Spark**
- Worked with Hive data & used Virtual box to run on local using Vagrant tool
- Passing data to BI team for converting it to reports using Cognos
- 

**Environment:**

- Scala, Spark, Hive, Sqoop, HDFS, Cloudera

**DEC 2016– DEC 2017**

**Data Engineer**

**Project # - Migration of DB2 tables to Hive tables**

**Description:** This project is to load the data from DB2 tables to Hive Tables using Scoop & do the reconciliation once data is loaded correctly.

**Responsibilities:**

- Import data into HDFS and Hive using Sqoop.
- Involve in create Hive tables, loading with data and writing Hive queries which will run internally in MapReduce way.
- Load data into Hive tables.
- Work closely with the business and analytics team in gathering the system requirements.
- Reconciliation of tables i.e. verifying whether the data is loaded correctly or not.

**Environment:**

- Hive, Sqoop, HDFS, Putty, BigInsight

**Jan 2015– Nov 2016**

**Application Programmer**

**IBM Internal Account**

**Project # - Automation for Mainframe application using JAVA RFT tool.**

**Description:** This project is to automate the lengthy manual process of deployment activity from dev region to production and Weekly report generation automatically and faster.

**Responsibilities:**

- Develop the JAVA code in RFT to Automate the process on terminal-based applications.
- Understand and learn the whole process of deployment in and out to provide best solution
- Developed the REXX code as well to get the required support component faster.
- Work closely with the deployment team to gathering the system requirements.
- Created test cases and tested the application.

**Environment:**

- Java, Rexx, DB2, RFT

**Apr 2014– JAN 2015**

**Application Programmer**

**IBM Internal Account**

**Project # - CEDS[Customer Entitlement Data Server]**

**Description:** Project is for data collection application which holds Customer Entitled information and acts as a repository for information which is enriched, reformatted and passed to other applications.

**Responsibilities:**

- Worked on PLI, JCL to develop and enhance the mainframes application
- Analyzed the requirements for defect and Change management
- Implemented the code changes and unit tested the code
- Performed User Acceptance test (UAT) for some of the TREC's and PCR's

## Qualification

---

- **Bachelor of Engineering** from **SVCE** (Rajiv Gandhi Technical University), Bhopal in 2013 with 70.25%.
- **HSSC** from M.P. Board with 71.6% in 2009.
- **SSC** from M.P. Board with 86.4% in 2007.

## Personal Details

---

**Date of Birth** : 09<sup>th</sup> July 1991  
**Current Location** : Pune, India  
**Gender** : Male  
**Languages Known** : English, Hindi