

Mr. Afroz Ilahi Kurne

Contact No :- +91-9890972090

Email Id :- afrozkurne@gmail.com

Professional Summary -

- Having 6 years of overall IT experience in Big Data Technology.
- Hands-on experience of **Google Cloud Platform** components and Big Data technologies.
- Experience of creating data pipelines using **Apache Beam** with **Cloud Dataflow**.
- Good working knowledge on **Core Java ,Scala** and **Python** Concepts.
- Good working experience in using Apache Hadoop and its ecosystem components like

HDFS, Map Reduce, Sqoop, Hive, Spark.

- Having experience in various Google Cloud Products like **Dataflow with python, Cloud Composer, BigQuery, Cloud Storage, Cloud Compute Instance.**
- Involved in scheduling jobs using CRON scheduler as well as with **Google Cloud Composer.**
- Involved in migrating lots of ETL from Legacy System to Hadoop.
- Experience in data analytics using Hive, Spark.
- Logical, Analytical and Good interpersonal skills committed to perform quality work.
- Very flexible and can work independently as well as in a team environment.
- Willing to update my knowledge and learn new skills according to business requirements.
- Good written and verbal communication skills.

TECHNICAL SKILLS -

- **Hadoop Framework & Eco-Systems :**
 - HDFS, Sqoop,Hive,Hbase, Spark-Core,Spark-SQL,Spark JDBC,Azkaban, Autosys..
- **Hadoop Platform :**
 - Apache Hadoop, Hortonworks, Cloudera.
- **Cloud Platform :**
 - AWS Cloud with EC2, EMR, S3.

- Google Cloud with BigQuery, Dataproc BigTable , Cloud Composer and Cloud Storage.
- **OS Platform :**
 - Linux (RHEL, Centos, Ubuntu), Windows XP/7/8/10, VM Workstation and Virtual box.
- **RDBMS :**
 - MySQL, Oracle.
- **Programming Language :**
 - Java, Scala, Python .
- **Tools/Technology :**
 - JIRA, Confluence, Perforce, SVN, Jenkins, Git , TeamCity , uDeploy, BitBucket.
- **IDE :**
 - Eclipse, IntelliJ IDEA, PyCharm.

EDUCATION -

Degree	School / University Name	Year of Graduation	Any specialization
MCA	Pune University	2016	Computer Application
BCA	Solapur University	2014	Computer Application
H.S.C	Maharashtra Board	2011	Science
S.S.C	Maharashtra Board	2008	NA

Experience -

Synechron India PVT LTD, Pune (March-05-2021-present)

Datametica Solutions PVT LTD,Pune(Aug-19-2019 – Feb-28-2021)

TIBCO Software India PVT LTD Pune(Sep 2016 – Aug -9-2019)

PROJECT - 4

Project Name – Financial Data Repository(FDR)

Domain –Banking and Finance

Role – Bigdata Developer

Team Size -8

Description of the project –

In this project, client wants to migrate their existing RDBMS data to Hive and hbase.The framework is based on Autosys orchestration tool and wrapper written on Shell, Python and Scala.

Responsibilities-

- Migrating data from various sources to HDFS and HBASE using Spark JDBC
- Performing preprocessing and cleaning on data using Spark Dataframe.
- Analyzing the data and migrating file formats from csv to avro.
- Loading the historical data from RDBMS to hive using Spark JDBC
- Migrated and Globalized the data from one region to another region using Spark.(region → NAM , APAC and EMEA)
- Creating a tables and performing the optimization operations on data through Spark.
- Realtime job monitoring and scheduling in hadoop.
- Creating email automation for getting status of job through mail.

PROJECT - 3

Project Name – Datawarehousing and ETL

Environment - BigQuery ,BigTable ,Data Proc Cluster,Cloud Composer and Cloud Storage

Domain –Banking and Finance

Role – Data Engineer

Team Size -7**Description of the project -**

In this project ,Client wants to migrate their Netezza and Data Stage data to GCP.Client wants to design a new model in GCP.

Responsibilities-

- Involved from the analysis phase through the implementation phase of the project.
- Involved in extracting data from Netezza and DataStage to BigQuery and BigTable.
- Involved in discussion, to automate Google cloud job development process and finalizing framework designing.

PROJECT - 2

Project Name – Datawarehousing and ETL

Domain –Banking and Finance

Role – Bigdata Developer

Team Size -8

Description of the project -

Description of the project -

This is an ETL development project, with daily scheduled ingestion and higher layer table population using Spark job, exporting data to Elastic Search, Hbase and Oracle. The framework is based on Azkaban orchestration tool and wrapper written on Shell, Python and Scala.

Responsibilities-

- Using Spark/Scala, transformed the imported data into Gold (transformed data)/Smith(aggregated data) layers.
- Involved in development of various wrapper scripts for providing all configuration details required for invoking the ingestion tool.
- Developed framework to encrypt, mask and tokenize sensitive data on Hive table.
- Solving analytical problem provided by Business team using Spark.

- DAG analysis and memory management to tune Spark program for better performance.
- Ad-hoc analysis of data using Hive and Spark SQL.
- History data load from Oracle to Hive and partitioned the data based on date (Highest handled data size is 1.4 TB).
- Involved in design and implementation of small file corrector tool for correcting the small file issue of history data on hdfs
- Implemented SCD1 and SCD2 on Hive using custom Spark job.
- Implemented a framework to load data from Hbase to Hive using Spark-Scala.
- Involved in design and implementation of spark framework to transfer data from hive to oracle for downstream applications.
- Loaded and transformed large sets of structured and semi structured data.

PROJECT - 1

Project Name – Global Store Repository(Nielsen)

Domain – Media

Role – Hadoop Developer

Team Size –10

Description of the project –

This Project is for Storing data into a repository and data coming from different sources.

This accept data from multiple sources and keep just one copy i.e. golden copy for stores. The functionality, therefore, enables to have just one record for one store, irrespective of how many times or sources the store has been uploaded from.

Responsibilities-

- Migrating data from various sources (Oracle, SQL server) to Hadoop using Sqoop.
- Automated the ingestion of sliding data from Oracle to Hadoop as per the business requirement.
- Involved in development of various wrapper scripts for providing all configuration details required for invoking the ingestion tool.
- Involved in Hive table creation, loading with data and writing Hive queries.
- Loaded and transformed large sets of structured and semi structured data using hive.

PERSONAL DETAILS -

- Gender : Male
- Marital status : Single
- Nationality : Indian
- PAN : YES
- Languages : English, Hindi, Marathi

Mr. Afroz Ilahi Kurne