

BIG DATA ENGINEER

MURALI KRISHNA KOMMI

Email:kmkbigdata@gmail.com

Mobile: +91-6303273304

Professional Summary:

- ✓ 4.2 years of experience in Bigdata and Hadoop ecosystems like Spark Core, SparkSql, YARN, Sqoop, Hive, Snowflake, AWS, Jenkins and Oozie.
- Load and transform large sets of structured, semi-structured and unstructured data from Relational Database Systems to AWS S3 using AWS glue.
- Developed Spark scripts using PySpark API's Dataframes and SparkSql for faster data processing.
- Created automated processes for activities like unit testing using python and databases backup.
- Developed AWS glue scripts to transform data between two different AWS IAM Accounts
- Developed ETL Pipelines to transform data from source to destination based on client requirements.
- Data ingestion and refresh from RDBMS to HDFS using Apache Sqoop and processing data through Spark Core and Spark SQL.
- Designing and creating Hive external tables and creating Partitions and Bucketing techniques in Hive using Hive Query Language (HQL)
- Involved in performance tuning of Spark Application Jobs
- Experience in Lambda Functions, AWS Glue, AWS Athena, AWS S3, Step Functions, CloudFormation, EC2 services in AWS
- Hands on Experience in AWS Management tools like CloudFormation, CloudTrail and CloudWatch architecture, monitoring memory, disk Metrics, logs and Graphs, and setting alarms.
- Hands-on DATABRICKS Data lake platforms and developed python script notebooks.

Working Summary:

- Working as Senior BIG DATA Developer in **C.G.I**, Hyderabad from JUN 2020 to till now.
- Worked as Jr. BIG DATA Developer in **AUTODESK**, Bangalore from Jun 2017 to May 2020

Education:

MASTER OF TECHNOLOGY in V.I.T University (2017 passed out)

Technical Skills Set:

Big Data	: Spark Core, SparkSql, Sqoop , Hive , Snowflake , AWS, Jenkins
Languages	: Python , SQL and
SPARK PLATFORMS	: Qubole, Data Bricks
Tools	: PyCharm, Jira, GitHub , Databricks Notebook
Database	: MYSQL, ORACLE , SNOWFALKE(Cloud Database)

Project Details:

Project : AUTODESK DATA PLATFORM -BUSINESS SYSTEM DATA
Client : AUTODESK

Skills : Hive, S3, DataBricks, Qubole ,Spark

Autodesk Data Platform Projects is to collect structured and semi- structured data from different sources and dumping into ADP. After Collecting the raw data from different source systems running the ETL pipelines to cleansing the data and converting into business needs. Later it is used for Reporting Dashboards and data visualization for internal use to fulfill need of Stakeholders Requirement. In this project we have Salesforce as source system (CRM Tool), which has details of customers and products associated to each customer.

Responsibilities:

- Responsible for Data Ingestion, Data Cleansing, Data Standardization and Data Transformation.
- Worked on creating Hive managed and external tables based on the requirement.
- Implemented Partitioning on Hive tables for better performance.
- Installing, Upgrading and Managing Spark Cluster in Qubole Data lake Collaborated with the infrastructure, network, database.
- Creating the tables in Athena and integrating with looker dashboard. In Looker Stakeholders will create there-own dashboards with processed final output data for business requirements
- Deploying the oozie workflows by using the automated Jenkins Tool.

Project2: Management Operating System (MOS)

Client: CIGNA (Health Insurance)

A Management Operating System (MOS) is the set of tools, meetings, and behaviors used to manage people and processes to deliver results. MOS follows the Plan, Do, Check, Act improvement cycle to control and steadily improve process performance. MOS compares actual accomplishments to expectations, enabling our leaders to effectively plan, manage, and report results through structured communication.

Responsibilities:

- Developed Spark scripts based on PySpark as per requirements
- Implemented Spark by using data frame API's , SparkSql for faster data processing
- Involved in extracting data from various data sources into AWS S3.
- Maintain the day to day data in Datawarehouse without lagging from different source systems.
- Developing the ETL AWS GLUE Jobs to transfer data from source to AWS Athena and then after performing transformations on Athena tables moving data from Athena to Postgres SQL.
- Load the daily injecting data into ATHENA tables and to create Partitioning to improve the query performances
- To Check the Postgres SQL Database for monitoring daily ingestion jobs are successful if not debugging the issue.
- Orchestrating the ETL jobs using AWS Step Functions.
- Running Analytics Pipeline for Aggregating and joining the multiple daily ingestion tables to meet the stakeholder's requirements.
- Scheduled the jobs and transferring final output data to Business team for reporting purpose.
- Creating the cluster group in Databricks for running Daily ingestion, Analytics Pipeline jobs.

- Deploying pipelines using Jenkins which is integrated with GITHUB
- Involved in creating EC2 Instances for RDS .

Environment: AWS GLUE, AWS S3, AWS ATHENA, STEPFUNCTIONS,DATABRICKS, PYSPARK, POSTGRE SQL Database,GITHUB, JENKINS, PyCharm