Sai Venkata Krishna Mammahi

Associate IT Consultant
+91 9666475301
svk041994@gmail.com
Bangalore, IN

E X P R I E N C E   S U M M A R Y

- Worked on Retail and Ecom, Airlines Data solutions, and Bigdata Data Engineering product development of data warehousing solutions using Data warehousing Technologies i.e., Snowflake and BigQuery, Bigdata Hadoop Ecosystem Components (Hadoop, Spark).
- Hands-on experience in developing the solution for Ecom and Airlines domain.
- Proficient in analyzing and translating business requirements to technical requirements.
- Built the data pipelines between various cross platforms such as SFMC to S3, S3 to GCP-BigQuery, etc. using data bricks.
- Experience in writing Spark jobs to process and analyze a large amount of data using spark RDD transformations and Actions.
- Experience in Bigdata technologies and RDBMS technologies i.e., Spark Core, Spark SQL, Scala, Hive, Impala, and SQLserver & MySQL.
- Migrated ETL jobs from Cron to Apache Airflow scheduler, managed job dependency easily and handled failing jobs automatically.
- In-depth understanding/knowledge of Hadoop and spark Architecture and Hadoop Ecosystem components.
- Good understanding of Azure cloud services components i.e., Azure Synapse, ADF, Storage accounts.
- Experience in handling different Hadoop-supported file formats and serdes such as JSON, Text files, AVRO, parquet, etc.
- Worked with version control systems i.e., SVN, and git
- Work experience on UNIX and Linux Platform.

E D U C A T A T I O N

- **M.S. (Computer Science) |** Middlesex University, The Burroughs, London (2016-2017).
- **B.Tech. (IT) |** Gurunanak Technical Institutions, Hyderabad, Telangana (2011-2015).

T E C H N I C A L   S K I L L S

- *Cloud Technologies:* AWS Ec2, AWS S3, AWS EMR, Azure BLOB, and ADLS2, Azure Synapse analytics.
- *Spark Skills:* Spark and SparkSQL
- *Data warehouse tools:* Snowflake, GCP-BigQuery
- *orchestration tool and DevOps skills:* Apache Airflow and Docker
- *Big Data Ecosystem*: HDFS, Yarn, Hive, HBase, Sqoop, hue, Impala.
- *Languages*: python and Scala
- *SQL Server:* MSSQL and MySQL
- *Hadoop and Spark Distributions*: Databricks, HDP, and CDH
- *IDE and Other Skills*: Eclipse, IntelliJ, and PyCharm.

E X P E R I E N C E

Boston Consulting Group (BCG)   APAC
Data Engineer   May.2021 – till date

- Working on Retail and Ecom cases and provided solutions for segmentation and retention issues.
- Analyzed the customer behavior and provided and built the solutions to retain the client using GCP-BigQuery, Snowflake, python, and AWS Ec2, S3, and data bricks.
- Built the data pipelines and ETL operations using SFMC, AWS S3, Python and Apache Airflow, AWS SFTP, MSSQL, GCP-BigQuery.
- Working autonomously on three-dimension cube creation, customer behavior analysis, clickstream data analysis.

ITC Infotech   Bangalore, In
Associate IT Consultant   Aug.2019 –
May.2021

- Built the big data solutions for Airline domain real-time issues.
- Provided solutions for complex problems in the airline's domain using Bigdata technologies.
- Worked independently on cabin crews managing system and Feedback management system.
- Involved in data ingestion, data enrichment, and dashboard building of Airlines data using Spark and Scala.

Temenos   Bangalore, In
Software Engineer   Sep.2018 – Jul.2019

- Worked in a product development team of Data Engineering Tool.
- Designed and developed data warehouse application using Java, Spark & Scala**.**
- Design and developed Batch data applications using Spark.
- Involved in code review and Unit Testing of the product/code
- Developed Data Engineering Components for ETL operations (Hive and Spark SQL).

P R O J E C T S

Project 1: UA Retention Model
Client: Under Armour – Northern America
Tech Stack: GCP-BigQuery, Snowflake, Aws Ec2, Aws S3, Ms. SQL, data bricks, python and Apache Airflow

Project Description:

The client is an American sports company that manufactures sports, casual apparel, other accessories, and equipment. The company has a diverse range of products being sold across multiple channels – Wholesale, Factory House, Brand House, and E-commerce.

In the module under discussion, BCG identified opportunities in the Retention module for the E-commerce channel. The aim was to analyze the customer behavior on campaigns, whom to target for the specified campaigns, send the offers to the most relevant customers by looking at their purchase and

browsing history from various channels and sources, and plan to build a product recommendations platform for clients and users.

Role and Responsibilities:

- Built the Ecom solutions using GCP-BigQuery, Snowflake, MSSQL, and Python technologies.
- Integrated three Datawarehouse databases and created a master data hub.
- Analyzed customer behavior using legacy and historical data.
- Built the data pipelines from multiple sources such as SFMC, SFTP, Datawarehouse, and S3 Bucket.
- Working with classic BCG consultants and Under Armour analytical team to provide the solutions for the retention module.
- Scheduled the tasks using Apache Airflow to simplify the task monitoring.

Project 2: SpiceJet RM Consulting (Data Ingestion & Data Enrichment) (SGDE & SGDL)
Client: SpiceJet
Tech Stack: CDH stack (Cloudera 5.14), Spark 2.2, Scala 2.11, Python, Hadoop, SQL Server, and Eclipse

Project Description:

SpiceJet is the second largest domestic and budget airline based in Noida, India, and a wholly owned subsidiary of space express. The Strategic Hadoop ecosystem is now part of SGDE architecture, it is now the default solution for most data requirements. Therefore, an end-to-end, effective, and cost-efficient data model is necessary to ensure the needs of the Business stakeholders, and Passengers are covered appropriately with the levels of the exigency of a digital organization. The enrichment team will get the Passenger feedback data, Cabin crew roster data, fuel prices data, cargo shipment data (Domestic and International) from the ingestion team from various sources (SFTP, SQL Server, Salesforce, ARMS ODS, Navitire ODS, and Merlot).
The enrichment team performs transformations, recycling data backup of Feedback, productivity Analysis data, NH data, and cargo shipment data using Spark and Scala. The enrichment team performs transformations, actions, and stores into a destination, works on different file formats, and optimization techniques to improve the performance of spark jobs.

Role and Responsibilities:

- Working on data fetching from SpiceJet data lake using spark
- Working standalone and Single point of contact for ARMS-related issues.
- Doing data ingestion in between Navitire and ARMS data lakes.
- Involved in data backup and recycling process.
- Working closely with IFSD and IT team to provide the solutions for Cabin crew and Golden Records-related queries.
- Written the Spark Jobs in Scala to process a variety of data formats and performed the operations using Spark transformations and Actions.

Project 2: Temenos Data Engineering Product (TDE)
Client: Temenos
Tech Stack: Hadoop, Hive, Tez, Spark, JDBC

Project Description:

The product is built from the ground to sophisticate Hadoop application development without writing or maintaining complicated Apache Hadoop code. To Meet the Enterprise needs by tackling the challenges of big data application development. TDE provides a suite of components to deliver lower cost, higher capacity infrastructure. It has 4 Modules such as Admin, Designer, Deployer, and Scheduler, Admin module enables administrators to change TDE server configuration without having to manually edit the configuration files, control projects, and users. The Administration module features a tabbed interface that groups related configuration settings within each tab. The Designer Module is an intuitive graphical drag and drops application development environment with prebuilt ETL components. Finally, the Scheduler module enables application schedulers to build a pipeline or a workflow from a sequence of jobs developed in designers and provides greater control over jobs and makes it easier to repeat the jobs at predetermined intervals.

Role and Responsibilities:
- Involved in Requirement analysis and design/model of Data warehouse product development (Hive Components)
- Developed Data warehouse component (Hive) in Designer module (Internal Tool) to make easy ETL operations
- Written Java-JDBC connection code to get access to the Hive metastore to retrieve the user-selected fields
- Written Scala-match in spark and Scala for data warehouse components i.e. (Hive operations)
- Written Spark data frame/datasets application to process, perform ETL operations using SparkSQL.
- Temenos jobs (ETL operation) execution done in Spark execution engine in Temenos data engineering tool
- Designed and implemented, view sample data functionality to view first 150 records in designer module
- Worked with testing team to fix the bugs and implementation team