

Mallikarjuna Reddy Kummetha

Mobile 📞: +91 – 8885411775

Email: kummetha.arjun@gmail.com

Career Objective:

To obtain a challenging position in a dynamic organization and be an integral part of a growth-oriented company. To utilize and grow the analytical, research, and technical skills that provide me the job satisfaction and make me work with wholehearted commitment.

Professional Summary:

- 8.4 + years of professional experience in information technology with an expert hand in the areas of big data technologies, AWS, and Azure cloud services and 4+ years of experience in Data warehouse implementations including Snowflake.
- Currently working as Senior Big Data Engineer in phData, Bengaluru.
- Certified **Snowflake Snowpro core** and **AWS Solutions Architect – Associate**.
- Having work experience in ingestion, storage, querying, processing, and analysis of big data with hands-on experience in Hadoop Ecosystem development including Map Reduce, HDFS, Hive, Impala, beeline, Sqoop, Spark-core, Spark-Sql, pySpark, Azkaban, Zena, and Hue.
- Skilled in the art of devising roles, shaping views, and implementing techniques for tuning performance, all geared towards elevating the efficiency of the Snowflake system.
- Experienced in utilizing AWS Glue for ETL workflows, enabling efficient data extraction, transformation, and loading.
- Strong knowledge of AWS CloudWatch for monitoring and managing AWS resources, setting up alarms, and collecting metrics.
- Proficient in version control systems like Git, GitLab, and VS for code repository management and collaboration.
- Developed ELT/ETL pipelines to move data to and from Snowflake data store using combination of Python and Snowflake Snow SQL.
- Experienced in using Snowflake Clone and Time Travel.
- Participated in the development, improvement, and maintenance of snowflake database applications.
- In-depth knowledge of Snowflake Database, Schema and Table structures.
- Optimized Spark jobs and workflows by tuning Spark configurations, partitioning and memory allocation settings.
- Strong expertise in optimizing Spark jobs and leveraging Azure Synapse Analytics for big data processing and analytics.

- Expertise in migrations projects from on-premise Hadoop/spark cluster to Snowflake or AWS using Data source toolkit.
- Worked on distributed frameworks such as Apache Spark and Presto in Amazon EMR, Redshift and interact with data in other AWS data stores such as Amazon S3 and Amazon Dynamo DB.
- Strong experience in migrating other databases to Snowflake. In-depth knowledge of Snowflake Database, Schema, and Table structures.
- Good Knowledge on Hadoop Cluster architecture, monitoring the cluster, managing, and troubleshooting Hadoop-related issues.
- Expertise in Spark Memory optimization on Dynamic allocation.
- Involved in Creating Hive Tables and loading processed data into tables and fine-tuned performance using different optimization techniques.
- Developed Hive scripts for client requirements to perform analysis on top of the large datasets and generated different file format outputs by applying performance tuning.
- Expertise in scheduling Applications/Jobs on scheduling tools like Azkaban, Zena, Autosys, Oozie, and Airflow.
- Created a POC server to configure Hadoop and Spark installation using Vagrant virtual software development environment and created a lightweight server for LDAP access.
- Experience with Hadoop distributions like Cloudera CDH and Hortonworks.
- Expertise in RCA for failed jobs using application logs in Azkaban, Zena, and Autosys.
- Knowledge of NoSQL database HBase and stream-processing like Kafka architecture.
- Strong teamwork skills with the ability to work effectively within cross-functional teams.
- Very proactive in learning and leveraging emerging technologies.

Awards and Knowledge Transfer: -

- Led freshers and ETL teams during the training on Hadoop and Spark with live project explanation.
- Got SME award for Bigdata Hadoop technology.
- Received H&PS - Zenith Award for the month of May 2019 in the Young Champion.
- Received best Super Squad team award for 3rd quarter of the year 2022

Technical Skills: -

| | |
|----------------------------------|--|
| Big-data Ecosystem components | Snowflake, Databricks, Hdfs, Yarn, Hive, Sqoop, PySpark, Spark SQL, Azkaban, Zena, Autosys, Kafka, Hbase, Anaconda, JupiterLab, Fivetran and DBT |
|----------------------------------|--|

| | |
|--------------------|---|
| Languages | Python, Scala and Shell scripting |
| IDEs | Eclipse, IntelliJ, PyCharm and Visual Studio |
| Databases | MySQL and HBase, SQL Server |
| Cloud Technologies | Snowflake, AWS, Azure cloud |
| Operating Systems | MacOS, Windows Family, Ubuntu, and Linux |
| Ticketing tool | Service now, Kanban and Jira |
| Others | Putty, MobaXterm, ServiceNow, WinSCP, and GIT |

Project Details:

Client : Freedompay

Environment : Azure Devops, ADLS Gen2, DataFlows, Snowflake, Hive, Impala, Azure Devops, SparkSQL, SSIS, Sql Server, and Shell scripting and Data Source toolkit

Project Duration : Apr-2023 to Present

Organization : phData

Project Description:

FreedomPay platform is a commerce solution that includes mobile payment capabilities, high-level security, incentive technologies, and business intelligence that integrates with live POS systems. Migrating the existing on-premises data SSIS packages, and SQL Server to Snowflake data warehouse as Azure as a cloud storage medium.

Responsibilities:

- Designed and implemented scalable data ingestion pipelines using Azure Data Factory, ingesting data from various sources such as SQL databases, CSV files, and REST APIs.
- Ensured data quality and integrity by performing data validation, cleansing, and transformation operations using Azure Data Factory and Databricks.
- Designed and implemented a cloud-based data warehouse solution using Snowflake on Azure, leveraging its scalability and performance capabilities.
- Collaborated with data analysts and business stakeholders to understand their requirements and implemented appropriate data models and structures in Snowflake.
- Developed and optimized Spark jobs to perform data transformations, aggregations, and machine learning tasks on big data sets.
- Collaborated with DevOps engineers to develop automated CI/CD and test-driven development pipeline using Azure as per the client requirement.
- Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL) and processing the data in Azure Databricks.
- Involved in running all the Hive scripts through Hive on Spark and some through Spark SQL.

- Collaborated on ETL tasks, maintaining data integrity, and verifying pipeline stability.

Project Details:

Client : Cargill

Environment : AWS, Snowflake, Hive, Impala, Liquibase, SparkSQL, ScalaSpark, Oozie, and Shell scripting and Data Source toolkit

Project Duration : Nov-2022 to Mar-2023

Organization : phData

Project Description:

Every day, Cargill connect farmers with markets, customers with ingredients, and people and animals with the food they need to thrive. As part of the RACE migration team needs to migrate CDP2(Hadoop) to CDP3(Snowflake) only FIBI Mexico data.

Responsibilities:

- Developed Shell scripts to extract the DDL from CDH to SF
- Written Spark Scala code to write data frame outputs to S3 buckets with parquet file format
- Implemented Storage integrations with S3 to Snowflake
- Created **Snowflake** Stored procedures, Snow pipes, Streams, and tasks as per the requirement.
- Involved in the creation of **Snowflake** tables, and views using **Snow SQL** and **Python** connector.
- Written AWS scripts to integrate S3 with Snowflake
- Integrating phData tools for managing Snowflake access (using Tram) and data migration through CDC mechanism (using Streamliner).
- Optimized some of the data pipelines using Snowflake utilities.
- Deployment process version control using Liquibase.
- Used Data Source toolkit for converting thousands of queries.

Project Details:

Client : Ingersoll Rand

Environment : AWS, Snowflake, Hive, Sqoop, Yarn, SparkSQL, Pyspark, Rundeck, Flyway, and Shell scripting

Project Duration : Feb-2022 to Oct-2022

Organization : phData

Project Description:

Ingersoll Rand Inc., is a multinational company that provides flow creation and

industrial products. Its products are sold under more than 40 brands across all major global markets. Migrate IR's existing data warehouse and data lake platform hosted on Cloudera to Snowflake to achieve goals.

Responsibilities:

- Developed Shell scripts to extract the DDI from CDH to SF
- Automated whole deployment process version control using Flyway
- Used SQLMorph for converting thousands of queries.
- Implemented Automation for custom needs
- Written AWS scripts to integrate S3 with Snowflake
- Created **Snowflake** Stored procedures, Snow pipes, Streams and tasks as per the requirement.
- Integrating phData tools for managing Snowflake access (using Tram) and data migration through CDC mechanism (using Streamliner).
- Optimized some of the data pipelines using Snowflake utilities.

Project Details:

Client : Marvin windows(Industrail)

Environment : Fivetran and Snowflake

Project Duration : Dec-2021 to Jan-2022

Organization : phData

Project Description:

Marvin creates windows and doors inspired by how people live. Explore our product lines and enhance your new construction, remodel or replacement project.

Responsibilities:

- Using the Fivetran application to extract the data from multiple sources.
- Worked on SQL transformations.
- Using the Postman API tool to handle all snowflake administration.
-

Project Details:

Client : Traveler's

Project Title : Webpd Analytics

Environment : Hdfs, Hive, Sqoop, Yarn, SparkSQL, Pyspark, Autosys and Shell scripting

Project Duration : Mar-2020 to Sep-2021

Organization : Accenture

Project Description:

Travelers is an American insurance company. The company offers insurance through three segments those are Personal Insurance, Business Insurance and Bond and Specialty Insurance. Webpd Analytics project data needs to migrate from existed on-premises servers to Hadoop and Spark layers using multiple bigdata applications. Analyzing the data sets, loading data to tables and storing on other layers for downstream applications and other users, Business Intelligence Analysts, Reporting tools and Data Scientists etc.

Responsibilities:

- Reviewing the business functionality &Analyzing business requirements from RDEM documents.
- Involved in handling of **CSV, XML, ORC** and **Parquet** files using **PySpark SQL** and performed transformations and performance tuning options
- Involved in writing shell scripts, **HQL** scripts and **PySpark** programs per business requirements.
- Using **Sqoop** need to ingest the data from DB2, Teradata and Sql Server to Hadoop layer. Based on requirement needs to load data to direct hive partitioned tables.
- Based on RDEM design and Mapping document needs to create different types of tables and then load the datasets into hive tables with **ORC/Avro/Parquet** file formats.
- Need to load processed data into tables and fine-tuned performance using different optimization techniques
- Implemented **AWS** services for migrating the data from On-primse to cloud using S3, **EMR** and **EC2** services.
- Performed loading of data from **S3 to Pyspark** and applied multiple transformation on data frmaes.
- Performed RCA on failed jobs, upon analysis needs to restart, hold or kill the jobs using Azkaban and Autosys.

Project Details:

Client : HCSC

Project Duration : Sep 2018 – Feb 2020

Role : Big Data Developer

Environment : Hadoop, Hive, Pig, Zena, Spark and Shell Script

Organization : Accenture

1. MA Part B Phase I

Description:

Enterprise Services Portfolio Consumption team is responsible for creating and sending MA Part B & Dual Bill Historical Pharmacy Encounter Claims data extracts to Edifices.

Responsibilities:

- Having good understanding of design document and involved in development as a key resource in implementing the project.
- Developed PIG scripts to perform ETL operations and generate the output extracts and transferring the files to client server.
- Created the Shell Scripts to automate the whole process.
- Developed the Test Plan that Included the scope of the release. Created detailed test cases and executed them manually and documented the results.

2. MedSupp

Description:

The Medsupp lines of business are premium underwritten products that cover the gap for hospital and medical expenses that Medicare doesn't pay, primarily the Medicare approved deductibles and coinsurance. Each of the five states has Medsupp product: IL, NM, OK, TX, MT MedSupp. Although many groups offer their Medicare eligible retirees Medicare supplemental benefits, those members and related claims are not part of this product line. Only individuals are included in this product. Claims for this product line are processed at the FSU in Albuquerque, NM using the Blue-Chip system.

Responsibilities:

- Having good understanding of Design, mapping documents and involved in development as a key resource in implementing the project.
- Developed multiple extracts like ACS Application, ACS Premium, Callidus Base Premium, AON HRA and Extend Member Extracts.
- Developed HQL files by applying transformations to generate the output extracts.
- Created Hive tables and loaded test data to validate that output results.
- Developed Unix shell scripts for high level automation of executing the HQL files and transferring the files to client server.
- Performed unit testing by executing test cases and test plans in compliance with established procedures and methodologies.

3. MA Part B Phase II

Responsibilities:

- Implemented Audit Framework for duplicate and zero-byte file check using existing framework.
- Created Zena workflow to automate the execution flow.

4.CMS UNIVERSE

Description:

On a monthly basis, data from various external and internal sources will be loaded to the data Lake Ingestion layer and subsequent Data Lake Consumption layers. Schema (on data from external sources) and Data Quality checks will be performed to ensure data quality prior to the consumption of the data. Validated CMS universe data will be transmitted to business from the Data Lake outgoing layer.

Responsibilities:

- Created the shell script for the Talend jobs execution
- Created Zena workflow to automate the execution flow.

Project Details:

| | |
|-------------------------|--|
| Client | : Kohl's |
| Project Title | : GCP - Migration |
| Environment | : Hadoop, Hdfs, Hive, MySQL, Azkaban, Shell scripting, Jira and Tonomi |
| Project Duration | : Sep-2017 to July-2018 |
| Organization | : KMRSOFT |

Project Description:

KOHL'S is one of the top 10 retail marketing chain in United States since 55 years, which based out in 1,155 locations. Core marketing is done on Clothing, footwear, bedding, furniture, jewelry, beauty products, electronic, housewares. Introduced DW/BI for analytics and reporting which helps KOHL'S to improve the business strategy and efficiency. They have self-hosted clusters with Hadoop. Recently KOHL'S aimed to migrate all the analytics data storage to **GCP - Google Cloud Platform** using Big data Technologies.

Responsibilities:

- Analyzing business requirements and functionalities.
- Used Sqoop to ingest the data into HDFS and Hive.
- Developed Hive scripts for table creations and Data loading
- Pig Scripting to read, process, Load the hive/HDFS data
- Created workflow using shell script for job scheduling on Azkaban.

- Performed Unit Testing on developed application process.

Project Details:

Client : Kohl's
Project Title : Onprimse DDH
Environment : Hadoop, Hdfs, Hive, MySQL, Azkaban, Service now, and Jenkins.
Project Duration. : May-2015 to Aug-2017
Organization : KMRSOft

Responsibilities:

- Used Azkaban to schedule the flows as per the timings and monitored all running jobs on production layer.
- Worked as L2 engineer on production environment to look after running jobs.
- Performed data cleansing activities using shell script.
- Developed Hive queries for various business requirements and Shell script to implement the workflow for job scheduling.
- Analyzed Azkaban failed flow logs and Applications logs to find RCA.
- Performed data validations using Hue and Hive.
- Deployed Azkaban flows on production using Jenkins.
- Performed data catch up activities using Hive and Sqoop.

Qualifications:

- Qualified in Bachelor of Engineering with specialization in Computer Science with A Grade in 2015 under JNTUK University.

Personal Details: -

Name : Mallikarjuna Reddy Kummetha
Father's Name : China Mala Konda Reddy Kummetha
Date of Birth : Jan 09, 1994
Gender : Male
Languages Known : English, Hindi, and Telugu

Declaration: -

I hereby declare that all the information mentioned above are true and correct to the best of my knowledge.

[Signature]

Mallikarjuna Kummetha