Bhanu Prasad   Email:bhanuprasad9331@gmail.com
                                                                        Phone: +917975322102

LinkedIn Profile:
www.linkedin.com/in/bhanuprasad9331

Data Enthusiastic, eager to contribute to team success through hard work and having excellent organizational skills. Motivated to learn, grow and excel on day to day basis.

Having 7.5 years of experience in IT industry. Enriched with the ability to learn new concepts & technologies quickly.

Technical Skills:
✦ Big Data Technologies: Apache Spark, Spark SQL and Kafka.
✦ Cloud Technologies: Azure Databricks, Azure Datafactory, Synapse, Informatica
✦ Databases: MySQL, SQL Server, Hive, Neo4j.
✦ Programming languages: Python, PySpark, Java
✦ Scripting language: Python, Unix and Shell scripting

Project Profile

Role – Data Engineer   **Oct 2021 to Present**

Project Description:

Handling end to end project pipelines to handle terabytes of data in Azure cloud through Batch and Stream processing.

Responsibilities:

- Ingest data from different types of semi structured files to ADLS containers using Pyspark framework.

- Process the data using Fact Dimension modeling and generate reporting tables.

- Implemented Spark transformations like Union, Join, Filter, Coalesce etc.

- Used PySpark for CDC logic and data analysis, improved downstream performance by handing over the data in partition fashion.

- Migrate the data from ADLS to Azure Synapse using Polybase external tables.

- Implement shell and python scripts to call the Airflow and Databricks APIs.

After the completion of initial load and basic validations in production, orchestrate the workflow using Autosys and Apache Airflow

Environment: Apache Spark, Azure Databricks, Azure DataFactory, Synapse, Informatica, Autosys, Airflow, Bamboo.

Role – Data Engineer   **Feb 2017 to Sep 2021**

Project Description:

Collaborated on building a petabyte scale data warehouse solution and data pipeline on AWS for Banking and Retail domain assets and its metadata.

Responsibilities:
- Data migration from RDBMS tables to S3 buckets through HVR data migration tool.
- Read the data from S3 and generate flat files for further processing through EMR.
- Crawl the data in the S3 buckets on a fixed schedule and update the Glue data catalog.
- Implemented multiple stored procedures on AWS Redshift by following core principles as distribution key, sort key etc
- Generate Enriched data through Glue jobs by parsing the original unstructured data.
- Migrate the enriched data to Redshift through Glue scripts.
- Trigger Stored Procedures in the Redshift through AWS Lambda once the data is migrated from S3.

Environment: Apache Spark, AWS Glue, Lambda, Redshift, Step Functions, EMR, AWS SAM, Jenkins.

Role – ETL Developer   **July 2015 to Jan 2017**

Project Description:

Data generated by different Apple Retail applications is stored across multiple shards in RDBMS Tables. Reports to be generated based on this data and display visualizations using Jaspersoft.

Environment: Spring Batch, MySQL, Quartz Scheduler.

Education details:

1) 10th
- Little Angels High School, Tirupati. Board of secondary education (2007-2008)

2) 12th
- Narayana Junior college, Tirupati. Board of Intermediate education (2008-2010)

3) Bachelor Degree
- Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University
- B.E, Electronics and Communication engineering (2010 -2014)

Declaration:

I hereby declare that all the information provided above is correct to the best of my knowledge.

                                                                                  (Bhanu Prasad)