



SUDHAKAR BOMMISSETTY

sudhakarcse505@yahoo.in

Ph. +91-850-034-3487/893-914-2520

Technology Specialist – **Bigdata, ETL, Cloud (AWS, Azure, Snowflake)**

GitHub - <https://github.com/sudhakarcse505>

LinkedIn - www.linkedin.com/in/sudhakarcse505

OBJECTIVE:

Actively looking for **Data Engineer** opportunity in Data-Lake, EDW, BI, Big data Analytics with **Challenging Environment** and can have a Chance to **learn and implement** optimized solutions using Cloud services

CERTIFICATIONS:

- Cloudera Certified **CCA Apache Spark and Hadoop developer (CCA-175)**

SUMMARY:

- Passionate about **Data Engineering** with an experience of **8+ years** in building **Enterprise data lake & data warehouse** Applications.
- Extensive Experience in **Hadoop Eco System-HDFS, Apache Spark (Scala & Python), Apache Hive/Impala, Snowflake, Apache NiFi, Apache Kafka, Apache Oozie & Apache Airflow, Rundeck**
- Extensive Experience in building **ETL Batch & Near- Real time Streaming** data Pipelines using **EMR, AWS Glue, ADF** and **Apache NiFi, Python based Frameworks**
- Strong Experience in **building Enterprise Data Warehouses** using **Teradata, Snowflake, Hive/Impala data warehouses.**
- Good Experience in working with **AWS & Azure** Cloud platforms related to Data Engineering services like **S3, EMR, EC2, Glue, Athena, Lambda, snowflake, ADF, Azure Data bricks**
- Strong Experience in building **Data lakes** using **HDFS, S3, MongoDB.**
- Extensive Experience in **Performance Tuning of Spark batch & SQL Streaming jobs** and can Stabilize long running Jobs
- Strong Experience in Automating manual tasks using **Shell/Python Scripting.**
- Working in **Agile** methodology with **Jira, Azure DevOps** as tracking tools

TECHNICAL SKILLS:

Data Ingestion

Apache NiFi, Apache kafka,

Data Processing

HDFS, Yarn, Hive, Apache Spark-SQL (Batch, Streaming)

Workflow Scheduler

Apache Oozie , Run deck, Apache Airflow

Hadoop Distribution

Cloudera, Hortonworks, EMR

No SQL Database

MongoDB

SQL Database

SQL Server, Postgres

Programing Language

Scala, Python

Scripting Language

SQL, Shell Scripting

ETL Tools

AWS Glue, ADF, Incorta, Informatica

Data Warehouses

Teradata, Hive, Snowflake

MY CURRENT ASSIGNMENTS:

Employer : Fractal Analytics Ind Pvt Limited

Role : Senior Data Engineer

TECHNOLOGY: S3, Impala, AWS EMR, Pyspark, Athena, Snowflake

DURATION : 2021-Sep to Present

Cargill Data Asset Solutions –

CDAS is an end-to-end Analytics solution that builds data products from various datasets and can build product layers which can consume by platform team to send data to business clients.

Responsibilities-

- Extracted data from different sources like Impala, APIs, Excel files and S3 files
- Designed and implemented end-to-end data pipeline with Transform, cleaning activities
- Loaded data to snowflake data warehouses for Reporting
- Optimization of End-to-End data Pipeline
- Working with complex queries which includes CTE, Joins, Subqueries to build product layers
- Analysis and exploration of new data sets
- Implemented Python based Data Quality Checks on the and data Availability Checks
- Used Azure DevOps for sprint boards and Git Hub for version control

MY PREVIOUS ASSIGNMENTS:

Employer : Scientific Games Ind Pvt Limited

Role : Data Engineer/ DBA / Product Developer

TECHNOLOGY: Apache Spark, Apache NiFi, Apache kafka, MongoDB, Dockers, AWS, Azure

DURATION : 2018-Jun to 2021-Sep

Enterprise Data Platform- we have developed ETL Product named EDP (Enterprise Data Platform) for our customers to build Common Data Lake for all microservices, I have involved in this Product building from beginning to deployment, maintenance & Upgrades. It includes below phases

- Data Acquisition –
 1. Have used Open-Source Dataflow web-based tool **Apache NiFi**, we have developed a Common Template to Load/Extract data from variety of sources like CSV, Text, Json, RDBMS and Kafka.
 2. Can support Streaming, Bulk & Incremental Loads
- Data Transformation –
 3. Have used Open source Distributed In-Memory Computing engine **Apache Spark**, we have built the common Spark Code to read datasets from variety of sources like RDBMS, NoSQL, HDFS, Kafka, Event-Bus, flat files .etc
 4. Perform various Transformations like Select, Filter, Join, Group, Aggregate, Sort, Data type Conversion, Timestamp format change, drop, distinct...etc.
 5. allow Users to write SQL query Transformations in both Batch and Stream Jobs
- Data Presentation –
 6. using **PowerBI/Grafana** tools to Present the data from our MongoDB Data lake
 7. Users must configure the metrics or Dashboards which they want to see
- EDP Engine –
 8. Have used **Play Framework** to build this Engine, this can be Responsible for validating, Submit, status and Scheduling of EDP jobs
- Docker-Deployments –
 9. Have configure the Apache Hadoop cluster using **Dockers**
 10. Automated the EDP Installation using Shell Scripts
 11. Can Support Linux Ubuntu 16.04 LTS on wards
 12. Have deployed EDP in On-Cloud & On-premise systems

MY PREVIOUS ASSIGNMENTS:

Employer : Cognizant Technology Solutions Pvt Ltd

Client : Blue Cross Blue Shield of Minnesota, U.S.

Role : Team Lead/ETL Developer

TECHNOLOGY: Hadoop Eco –system Tools, Apache Spark-sql, Hive, Hdfs, Shell Scripts

DURATION : 2014-April to Jun-2018

MOP – Member Out of Pocket (MOP) is a vendor to BCBSMN. As Part of the Project, High Mark can send Weekly files to us which has High Mark migrated Members information and can merge with BCBSMN data extracted by Traditional Tools to a Pipe delimited Flat files and Zip the files and send to Vendors for their Further Analysis.