Chandan Reddy Nandyala

Data Engineer

**Location:** GA**| Email:** chandanr9707@gmail.com |**Contact No**: [(913) 358-9595](tel:(913)%20358-9595)

**SUMMARY**

* Having around 3 years of extensive experience in Hadoop/Bigdata processing.
* Comprehensive working experience in implementing Big Data projects using Hive, Spark, and Sqoop.
* Excellent working knowledge of HDFS Filesystem and Hadoop Demons such as Resource Manager, Node Manager, Name Node, Data Node, Secondary Name Node, Containers etc.
* Provided access to tables and ingested them to Snowflake using lambda functions.
* Experience working on Spark and Spark Streaming.
* Knowledge in installing, configuring, and using Hadoop ecosystem components like Hadoop Map Reduce, HDFS, HBase, Hive, kafka.
* Implemented AWS Step Functions to orchestrate the AWS Data Pipelines and Lambda functions.
* Applied Transformations on the data using PySpark Data frames.
* Developed AWS Lambda function to trigger glue jobs, S3 file transfer and start/stop EMR.
* Worked on ECS services, Docker images and containers.
* Sound Involvement with AWS services like S3, Glue, EMR, IAM, Auto Scaling and Lambda to trigger resources.
* Experience with building data pipelines using Azure Data Factory, Azure Databricks, and stacking data to Azure Data Lake, Azure SQL Database to control and concede database access.
* Strong working experience with SQL and NoSQL databases (Cosmos DB, MongoDB), tuning, backup and creating data pipelines.
* Experienced in scripting with Python (PySpark), Scala and Spark-SQL for development, aggregation from various file formats such as XML, JSON, CSV, Avro, Parquet.
* Developing End to End ETL Data pipeline that take the data from surge and loading it into the RDBMS using the Spark.
* Experience developing Pig Latin and HiveQL scripts for Data Analysis and ETL purposes and extended the default functionality by writing User Defined Functions (UDFs), (UDAFs) for custom data specific processing.

**SKILLS**

|  |  |
| --- | --- |
| **Languages** | Python, Java, SQL, Shell script. |
| **AWS** | S3, Lambda Functions, EC2, Redshift, EMR, Dynamo DB, Cloud Watch, IAM, Azure Databricks, Azure SQL, Azure Blob, ADLS, Azure Key Vault, ADF |
| **Databases** | MySQL, Postgres, Snowflake, Oracle, DB2, Hive, Redshift, Azure SQL DW, Databricks Database, HBase, Cosmos DB, Mongo DB |
| **On-Prem** | Hadoop (CDH), Hive, HDFS, MapReduce, Sqoop, Spark. |
| **Tools** | Power BI, Tableau, Jira, Confluence |
| **CI/CD & Version Control** | Jenkins, Gitlab, Git, Bitbucket, Gitlab, Kafka |
| **Scheduler** | Airflow, ADF |

|  |  |  |
| --- | --- | --- |
| **EDUCATION** |  | |
| **University of Central Missouri, Warrensburg, MO Jan 2021-Dec 2022** Masters in Computer System Analysis/Analyst, GPA: 3.4/4.0  **Relevant courses**: Cloud Computing, Advanced programming in Java, Big Data Architecture, Big Data Solutions, Business Intelligence Analysis, Data Resource Management, and Information Security. |  |  |
|  |  | **May2017** |
|  |  |  |
| **PROFESSIONALEXPERIENCE**  **Epsilon, TX Jan 2022 – Present**  **Data Engineer**   * Developing ETL Pipelines in and out of data warehouse, develop reports using advanced SQL queries. * Performed S3 buckets creation, and policies on the IAM role based polices. * Developed the PySpark code for AWS Glue jobs and for EMR. * Strong experience in migrating other databases to Snowflake. * Professional knowledge of AWS Redshift. * Formulated procedures for installation and configuration of Tableau desktop. * Developed pipelines using Kafka to load the data into HDFS from external systems, RDBMS. * Implemented Apache Airflow as the orchestration tool for authoring, scheduling and monitoring the Data Pipelines. * Implementing and Managing ETL solutions and automating operational processes. * Designed Tableau bar graphs, scattered plots, and geographical maps to create detailed level summary reports and dashboards * Enforced best security practices to secure containerized environment using Seccomp, Se Linux, and security features of Kubernetes * Involved in building ETL pipelines using AWS Data Pipelines. * Utilizing python and Kafka built data pipelines for pulling data from multiple sources into AWS S3. * Writing Spark SQL jobs to enrich (Join), aggregate and build KPI tables. * Writing lambda functions using python for different use cases like creating dynamic DAG, Event Triggers. * Experience in Kubernetes to deploy scale, load balance and manage Docker containers. * Created an Automated Databricks workflow notebook to run multiple data loads (Databricks notebooks) in parallel using Python. * Involved in publishing of various kinds of live, interactive data visualizations, dashboards, reports and workbooks from Tableau Desktop to Tableau servers * Hosting MongoDB database on cloud worked on managing MongoDB cluster, designed its structure, and written some MongoDB queries. * Worked on scheduling all the jobs using Airflow scripts using python. * Used GIT for code review during stages of branching, merging, and staging. * Developed, implemented, and maintained data analytics protocols, standards, and documentation.   **Dell, India**  **Data Engineer Jul 2018 – Nov 2020**     * Gathering all the required data from multiple data sources and creating datasets that will be used in analysis along with ETL tasks. * Created the test environment for the Staging area, loading the Staging area with data from multiple sources. * Collecting, aggregating, and moving data from servers to HDFS using Apache Sqoop. * Created Hive tables, loaded them with data and wrote Hive Queries. * Worked on migrating the data from HDFS to Azure Data Lake. * Data ingestion to one or more Azure services (Azure Data Lake, Azure Storage, Azure SQL) and processing the data in Azure Databricks. * Worked on Azure Data Factory (ADF), Integration Run Time (IR), File System Data Ingestion, and Relational Data Ingestion. * Wrote and executed various MYSQL database queries from Python using Python-MySQL connector and MySQL dB package. * Have good experience working with Azure BLOB and Data Lake storage and loading data into Azure Synapse analytics (DW). * Experience in GitHub and Bitbucket. * Built complex logics in SQL to get consolidated view of the data. * Performed Data Analysis and Data profiling using complex SQL on various sources systems including MS SQL Server. * Create multiple kinds of Power BI Reports and Dashboards. Reporting, delivery of automated reports & Power Bl reports customization. * Working on building visuals and dashboards using Power BI reporting tool. * Developed Pipelines and Workflows which transforms the raw data from several formats to parquet files for consumption by downstream systems. * Wrote, executed, performance tuned SQL queries for data analysis and profiling and wrote complex SQL queries using joins and sub queries |  |  |
|  |  |  |