NAGA SAI BALARAM YEDIDA

Data Engineer

✉ ynsbalaram@gmail.com
Andhra Pradesh
www.linkedin.com/in/balaram-yedida
https://github.com/ynsbalaram/Projects

EDUCATION

Btech

Computer Science and engineering

Bharath university

June 2017 - June 2021

Chennai

8.9/10

SKILLS

- **Programming** : Java, Scala, SQL
- **Big Data** : Hadoop, Hive, Sqoop, Spark, Hbase, My-SQL
- **Cloud Services** : AWS Redshift, Glue, Athena,EMR
- **Scheduler** : Informatica,Airflow
- **CI/CD** : Git, Perforce,Maven
- **IDE** : Eclipse, Pycharm IDE
- **Project Planning**: Jira, Agile

CERTIFICATIONS

- Completed the BIg Data Masters program Conducted by Trendytech
  Jan 2023

AWARDS

- Received "Wall Of Fame" award 2 times

CAREER OBJECTIVE

As a skilled data engineer with 2 years of experience in building and optimizing end-to-end data pipelines, I am seeking a challenging role in a dynamic organization where I can leverage my expertise in designing scalable, reliable, and performant data infrastructures. I am passionate about exploring emerging technologies and collaborating with cross-functional teams to drive business insights and deliver innovative solutions.

WORK EXPERIENCE

Data Engineering Analyst

Accenture

April 2021 - current   Hyderabad

- Ingested data from 12 different data sources upstream, including data logs, using **Sqoop** and Marlin.
- Designed and implemented a real-time data pipeline that integrates 50 million raw records from an **S3** data source, processing semi-structured data using **Spark**, **Python**, and **Sqoop**.
- Optimized existing pipelines to handle growing data requirements, resulting in a **50%** reduction in resources and a **10x** faster processing time.
- Led the migration from GCP to Tesseract using Presto, resulting in annual cost savings of **$15,000** for the company.
- Collaborated with the client to gather and analyze requirements, ensuring that the data pipeline was built with solid business value in mind.
- Became an expert in operational excellence, proactively resolving job failures before they could impact the SLA.

PROJECTS

Credit Card Fraud Detection

***Technologies Used*:-**
Sqoop,HDFS,Hive,HBase,MySQL,Amazon RDS,Spark Structured Streaming,Kafka
***Project Architecture*:-**

- Imported data from Amazon RDS to HDFS using Sqoop and scheduled the job in Airflow to run every 8 hours.
- Created external and bucketed tables to efficiently load Sqoop data into Hive tables.
- Established a lookup table in HBase and a structured table in Hive to accelerate data retrieval while processing streaming data.
- Configured a Kafka producer and consumer to integrate with HBase and process data through Spark Streaming.
- Scheduled the job at regular intervals to prevent data loss, leveraging Apache Airflow's scheduling capabilities.

Customer 360 Pipeline

Technologies Used:-S3, Sqoop, Hive, Spark, HBase, Airflow
Project Architecture:-

- Created an S3 bucket to store the text files that the order processing team put every day between 5 pm to 6 pm. Used Sqoop to fetch customer information from the MySQL/Oracle database and dump it into Hive.
- Developed a Spark job to process closed orders against the customer information using the output from the Hive table. Created a Hive table from the output path of the Spark job and uploaded it into HBase.
- Configured an Airflow DAG to orchestrate the entire data pipeline, including running Sqoop, Spark, and HBase jobs.
- Configured the Airflow DAG to send email notifications using Gmail SMTP server on pipeline completion or failure.