GOKUL BALASUBRAMANIAM

BIGDATA / DATA ENGINEER

CONTACT

+91 9095067960

gokul.balasubramaniam12@gmail.com

http://surl.li/jnwyr

Chennai, India

SKILLS

- Hadoop
- Hive (HQL)
- Apache Spark
- Yarn
- Oozie
- Sqoop
- Python
- Unix
- Scala
- SQL

GCP

- GCS(Cloud Storage)
- Biquery
- Apache Airflow
- Dataproc

EDUCATION

Bachelor Of Engineering

K S Rangasamy College of Technology

2017-2021

Specialization : Electronics and Communication Engineering

CGPA: 8.83

HSC

The Little Flower Higher secondary School, Salem

2016-2017

Score: 85.33%

LANGUAGES

- English
- Tamil
- Kannada(Average)

ABOUT ME

An experienced Data/Bigdata Engineer who has well exposure in the bigdata stack with various tools and technologies like ETL(Extract, Transform, Load), Data warehousing, Datalake, Hadoop, spark, sql(Hive). Familiar with GCP platforms like Google Cloud Storage(GCS), Dataproc, Cloud Composer(Airflow), Bigquery.
Ambitious to learn and gain knowledge to the greatest extend. Believer of "Always there is a way to improve".

WORK EXPERIENCE

Data/Bigdata Engineer

Tata Consultancy Services Since Jun-2021

- **Design and building ETL(Extract, Transform and Load) pipelines to deliver the data to stakeholders as per requirements.**
- **Processed and transformed the data using the Spark and SparkSQL to load into the Data Warehouse like Hive or Bigquery.**
- **Performance tuning on the spark jobs to make the batch job to run on its most efficient level.**
- **Extracting the data from database like Oracle using sqoop to the Hadoop cluster.**
- **Orchestrating the unix, sqoop, spark tasks with proper alignment using the Apache OOZIE and Apache Airflow(GCP).**
- **Creating tables or Views and developing Hive(SQL) queries to do certain joining, selection, filtration and transformation to extract the desired information as per customer requirements.**
- **Writing unix shell scripts to do initial checks and validations on the Hadoop and Google Cloud Storage buckets(Cloud Composer on GCS).**
- **Used Dataproc to run the Virtual(GCP) cluster to run spark applications.**
- **Automating the pipeline by setting up the upstream and downstream dependency for the GCP service to make it run in a sequence, time dependency for the pipeline to trigger by Airflow DAGs.**
- **Used Mainframe and TWSD(Tiwoli workload Scheduler) to schedule the batches.**
- **Extracting the logs from Logs Explorer and Spark UI.**
- **Communicating with the clients to present our deliveries and obstacles.**
- **ServiceNow and Jira used as ticketing tools for the tracking activities.**
- **Also have some exposure in the Machine Learning and Deep learning.**

ACHECIVEMENTS

- Star of the Month
- Best Teammate Award
- On the Spot Award

CERTIFICATIONS

- NPTEL - Joy Of Python Computing

KEY COMPENTENCIES

- **BIGDATA**
- **Data Warehousing**
- **Managing Stakeholders**
- **GCP**
- **ETL/ELT**
- **Consistent Learner**