**JYOTHSNA**

**DATA ENGINEER**

**phone: +1 609-357-4757**

**Email: srinivas.k@stiorg.com**

## SUMMARY:

## Over 6 years of Professional IT experience as Data Engineer in building data pipelines using Big Data Hadoop Ecosystem, Google Cloud Platform, Spark, Hive, HDFS, YARN, Sqoop, Python, SQL, Looker, GitHub, ETL tools and Production Support.

## Hands of Experience in Google Cloud Platform (Big Query, Google Cloud Storage, Cloud Dataflow, Data Proc, G - Cloud Functions, GSUTIL, BQ command line utilities, Cloud Composer (Air Flow as a service), Apache Beam, Cloud Functions, Cloud SQL, and Looker).

## Experience in using and writing SQL queries, Database Creation, and writing Stored Procedures, DDL, DML SQL queries.

* Experience in Importing and Exporting Data using Sqoop from HDFS to Relational Database Systems and vice-versa.
* Expertise with the tools in Hadoop Ecosystem including Hive, HDFS, Sqoop, Spark and Experience in Installation and Configuration, Hadoop Cluster Maintenance, Cluster Monitoring and Troubleshooting.
* Efficient in working with Hive Data Warehouse tool creating tables, distributing data by implementing Partitioning and Bucketing strategy, writing, and optimizing the HiveQL queries.
* Experience in Ingestion, Storage, Querying, Processing, and analysis of Big Data with hands-on experience in Big Data including Apache Spark, Spark SQL, Hive.
* Experience in Designing, Developing, and Deploying projects in GCP suite including GCP Suite such as Big Query, Data Flow, Data Proc, Google Cloud Storage, Composer, Data Studio, Pub sub, Looker etc.

## Designed, Tested, Maintained the Data Management and Processing systems using spark, GCP, Hadoop and Shell Scripting.

## Strong troubleshooting and production support skills and interaction abilities with end users.

## Experience in building Looker dashboards based on the requirements.

## Expertise in Collecting, Exploring, Analyzing, and Visualizing the data by generating Tableau/Looker reports.

## Involved in all stages of Software Development Life Cycle such as requirements analysis, functional & technical Design, testing, Production Support, and Implementation.

|  |  |
| --- | --- |
| Cloud | Google Cloud Platform (Cloud Storage, Big Query, Data proc, Data flow, Cloud Pub/Sub, Cloud Functions, Cloud Composer, Cloud shell, Cloud SQL), Airflow. |
| Big Data | Apache Spark, Hadoop, HDFS, YARN, Hive, Sqoop, MapReduce, Tez, Ambari, Zookeeper, Airflow, Data warehousing. |
| Job Scheduling & Orchestration | Control-M, Apache Airflow |
| Database | MySQL, SQL Server, DB2, Cassandra, Teradata, Big Query, PostgreSQL. |
| Methodologies | Agile(scrum), Waterfall. |
| Languages | Python, PySpark, SQL, HiveQL, and Shell Scripting. |
| Data Visualization Tools | Looker, Power BI, Microsoft Excel (Pivot tables, graphs, charts, Dashboards). |
| Version Control | Git. |
| Tools | Hue, Looker, IntelliJ IDEA, Eclipse, Maven, Zookeeper, VMware, Putty, JIRA, Toad, DB visualizer. |

## TECHNICAL SKILLS:

**PROFESSIONAL EXPERIENCE:**

**Client: Blue Cross Blue Shield, MI Sept 2022 – Present**

**Role: Data Engineer**

**Responsibilities:**

* Experience in building data pipelines, end to end ETL process for Data Ingestion and transformation in GCP and coordinating task among teams.
* Ingest Data into FHIR Store on GCP from GCS Buckets/BQ using Dataflow/Cloud composer.
* Used Cloud functions to trigger jobs on file arrival in GCS bucket and to update the data in FHIR Store.
* Used Apache airflow in GCP composer environment to build data pipelines and used various airflow operators like bash operator, Hadoop operators and python callable and branching operators.
* Designed and deployed multiple dashboards for different use cases based on business requirements.
* Developing the Looker environment and supporting the business decision process.
* Build data pipelines in airflow in GCP for ETL related jobs using different airflow operators.
* Build a program with Python and Apache beam and execute it in cloud Dataflow to run Data Validation between source file and Big Query tables.
* Good knowledge in using cloud shell for various tasks and deploying services.
* Created Big Query authorized views for row level security or exposing the data to other teams.
* Experience working with ServiceNow, Incident management and problem management, join Bridge lines, provide timely updates, troubleshooting production issues.

**Environment:** GCP, Google cloud storage, API, Big Query, Data Proc, Dataflow, Dataflow, Cloud Composer, SQL, GitHub, Airflow, FHIR Store, HealthCare, Apache Beam, Cloud shell, Python, Looker.

**Data Engineer June 2020 – Jul 2022**

**Client: Health partners, Ind.**

**Responsibilities:**

* Responsible to build the ETL Pipelines (Extract, Transform, Load) from data lake to different databases and reflect it to the frontend. We create and build integration patterns that transform raw data into refined data utilizing tools such as Hadoop, HDFS, Hive, Spark, Python, Sqoop, DB2, SQL Server, GCP
* Worked on a migration project to migrate data from Hadoop to GCP buckets and developed Big Query scripts to load data from GCP buckets to Big Query and scheduling the jobs in Cloud Composer.
* Process and load bound and unbound Data from Google pub/subtopic to Big query using cloud Dataflow with Python and build data pipelines in airflow in GCP for ETL related jobs using different airflow operators.
* Experience in GCP Dataproc, GCS, Cloud functions, Big Query, and good knowledge in using cloud shell for various tasks and deploying services.
* Worked on Hive Partition and bucketing concepts and created hive External & Internal tables with Hive Partition and responsible to troubleshoot issues related to data pipeline failures or slowness, built using MapReduce, Tez, hive or Spark to ensure SLA adherence.
* Developed Spark scripts by using Python as per the requirement.
* Improving performance and optimizing existing algorithms in Hadoop using Spark context, Spark- SQL, Data Frames, Pair RDD’s & Spark YARN.
* Built real time dashboards to report store level sales and region level sales for Walmart US and global data using Looker/Data Studio.
* Effectively followed Agile Methodology and participated in Sprints and daily Scrums to deliver the tasks and user JIRA board to manage and update tasks. Experience working with ServiceNow, Incident management and problem management, join Bridge lines, provide timely updates, troubleshooting production issues, vendor engagement and Maintain SLAs to provide quality services to end users.

**Environment:** Hadoop, HDFS, Spark, python, Teradata, Hive, Aorta, Sqoop, API, GCP, Google cloud storage, Big Query, Data Proc, Dataflow, Cloud Composer, Pub sub, SQL, DB2, UDP, GitHub, Tableau, Data Studio, Looker, etc.,

**Data Engineer Feb 2019 – Mar 2020**

**Client: Accenture**

**Responsibilities:**

* Experience in the principles and best practices of Software Configuration Management (SCM) in Agile, scrum, and Waterfall methodologies.
* Designing Oozie workflows for job scheduling and batch processing.
* Expertise in performing investigation, analysis, recommendation, configuration, installation and testing of new hardware and software.
* Worked on verifying and validating Business Requirements Document, Test Plan, & Test Strategy documents.
* Experience in working GIT for branching, tagging, merging, and maintained GIT source code tool.
* Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of Spark using Scala.
* Effectively followed Agile Methodology and participated in Sprints and daily Scrums to deliver software tasks on-time and with good quality in coordination with onsite and offshore teams.
* Used Shell Scripting (Bash and ksh), PowerShell, Ruby and Python based scripts for merging, branching, and automating the processes across the environments.
* Work closely with other data engineers, product managers, analysts to gather & analyze data requirements to support reporting & analytics.

**EDUCATION**

* Bachelor’s degree in computer science from JNTUH university, Hyderabad, India, May 2020.

**CERTIFICATION**

* Microsoft Certified: Power BI Data Analyst Associate
* Azure Fundamentals