

PROFESSIONAL SUMMARY:

- I have 5 years of experience in IT industry as a Data Engineer with Tableau and Business Intelligence tools with strong background in Databases and Data warehousing concepts.
- Good working knowledge of Amazon Web Services (AWS) Cloud Platform which includes services like EC2, S3, VPC, ELB, IAM, DynamoDB, Cloud Front, Cloud Watch, Route 53, autoscaling, Security Groups, EC2 Container Service (ECS), Red shift, CloudWatch, CloudFormation, CloudTrail, Kinesis, SQS, SNS, SES.
- Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse.
- Solid experience in Hadoop distributed file system (HDFS), Sqoop, Hive, HBase, Spark, MapReduce, Ambari, Kafka, Yarn, Airflow, Flume, Oozie, Zookeeper and pig.
- Experience developing Kafka producers and Kafka Consumers for streaming millions of events per second on streaming data.
- Extensively used Terraform in AWS Virtual Private Cloud to automatically setup and modify settings by interfacing with control layer.
- Expertise in AWS Spark, Snowflake, Lambda, RedShift, DMS, EMR, RDS, EC2, and AWS stack, as well as familiarity with cloud computing platforms.
- Good experience in developing ETL pipelines in and out of data warehouse using combination of Python and Snowflakes Snows by writing SQL queries.
- Strong working experience in developing Spark applications using Spark components (Spark Core, Data Frames, Spark SQL, Spark-ML and spark streaming API's).
- Implemented complex analytical solutions using Tableau and PowerBI to access data from numerous sources, including Oracle, SQL server, Teradata, Hive, and flat files.
- Excel in analyzing data using Spark SQL, HiveQL, Pig Latin, Spark/Scala, PySpark and custom MapReduce programs in java.
- Well-versed in project management as well as Scrum and Agile methodologies.
- **AWS Certified Solutions Architect – Professional.**

SKILLS:

Programming	: Python, Java, Scala, Shell Script, SQL, T-SQL, C#, PL/SQL
Databases	: Oracle, MySQL, SQL Server, MongoDB, Cassandra, DynamoDB, PostgreSQL, Teradata, Hbase, Bigtable
Big Data Ecosystem	: Hadoop, MapReduce, HDFS, Sqoop, PIG, Hive, HBase, Flume, Kafka, Apache Spark, Spark Streaming, Tez, Impala.
ETL Tools	: Talend, Informatica, SSIS, NIFI.
Visualization	: Tableau, Power BI, SSRS, Looker.
Cloud Technologies	: AWS, Azure, GCP, Snowflake
Messaging Systems	: Kafka, Kinesis Streaming, Flume, Google Cloud Data Flow, NIFI, Spark Streaming
DevOps Tools	: Docker, Kubernetes, Jenkins, Terraform, Ansible.
Other Tools	: Git, GitHub, Bitbucket, Linux , Jira, Confluence, SAS

PROFESSIONAL EXPERIENCE:

FIS | OH || Database Analyst Specialist

Feb 2024 – Present

- Responsible for requirement gathering, user meetings, discussing the issues to be resolved and translated the user inputs into ETL design documents.
- Created ETL mapping document and ETL design templates for the development team.
- Created external tables on top of S3 data which can be queried using AWS services like Athena.
- A solid experience and understanding of architecting, designing and operationalization of large-scale data and analytics solutions on Snowflake Cloud Data Warehouse
- Have good experience in developing ETL pipelines in and out of data warehouse using combination of Python and Snowflakes Snows by writing SQL queries
- Involved in migration from on prem to Snowflake Cloud.
- Used Analytical function in hive for extracting the required data from complex datasets.
- Extensively used to handle the Equipment Data and Vision Link data by using XML, XSD, JSON files and loaded into Teradata database.
- To Build distributed, reliable, and scalable data pipelines to ingest and process data in real-time.

- Created ETL pipelines using Stream Analytics and Data Factory to ingest data from Event Hubs and Topics into SQL Data Warehouse.
- Designing and implement a fully operational production grade large scale data solution on Snowflake Data Warehouse.
- Experience on Migrating legacy data warehouse and other databases (SQL Server / Oracle Database/ Teradata 15.0, DB2) to Snowflake.
- Data Analysis and Profiling to discover key join relationships, data types, and assess data quality utilizing SQL queries. Data cleansing, Data manipulation and exploratory analysis to identify, analyse and interpret trends and patterns in large data sets.
- Data Modelling/Data Architecture: Organization and arrangement of data in Staging /Intermediate/Final Targets using tables, views, etc.
- Developed Snowflake views to load and unload data from and to an AWS S3 bucket, as well as transferring the code to production.
- Worked on Migrating objects from DB2 to Snowflake.
- Expertise in creating tables using Hadoop File System and executing queries on Hadoop cluster using hive scripts.
- Have experience in Dimensional Modelling using Snowflake schema methodologies of Data Warehouse and Integration projects
- Experience with Snowflake cloud data warehouse and AWS S3 bucket for integrating data from multiple source system which include loading nested JSON formatted data into snowflake table.
- Worked on DataStage 11.7 as a medium to migrate the On Prem DB to Snowflake Cloud

ENVIRONMENT: Snowflake cloud Datawarehouse, AWS S3, AWS IAM, DBT (data build tool), Oracle Database 10g/11g, Teradata 15.0, Python, Tableau, Hadoop, control-M, SQL Server, DB2.DataStage 11.7.

SID GLOBAL SOLUTIONS | | Exton, PA | | API Data Engineer

Sep 2022 – Jan 2024

- Responsible for managing cloud computing tool AWS and the code in Git (version controlling) and deploying and operating AWS, specifically VPC, EC2, S3, EBS, IAM, ELB, Cloud Formation, and Cloud Watch using the AWS console and AWS CLI.
- Primarily Responsible for converting the Manual Report system to a fully automated CI/CD Data Pipeline that ingests data from different Marketing platforms to AWS S3 data lake.
- Utilized AWS services with a focus on big data analytics, enterprise data warehouse, and business intelligence solutions to ensure optimal architecture, scalability, flexibility.
- Designed AWS architecture, Cloud migration, AWS EMR, DynamoDB, Redshift, and event processing using lambda function.
- Gathered data from Google AdWords, Apple search ad, Facebook ad, Bing ad, Snapchat ad, Omniture data, and CSG using their API.
- Developed MapReduce programs to parse the raw data and create intermediate data which would be further used to be loaded into Hive portioned data.
- Involved in creating Hive tables, loading the data into it, and writing Hive queries to analyze the data.
- Involved in data ingestion into HDFS using Sqoop for full load and Flume for the incremental load on a variety of sources like web server, RDBMS, and Data API.
- Experience in custom aggregate functions using Spark SQL and performed interactive querying.
- Created UDFs to calculate the pending payment for the given customer data based on the last day of every month and used in Hive Scripts.
- Created Advance connections, join new tables, create & manage Extract and Monitor Queries using SQL Developer.
- Created Tableau scorecards, dashboards, Heat maps using show me functionality.
- Responsible for data modelling in MongoDB to load data which is coming as structured as well as unstructured data.
- Used data visualization techniques to effectively communicate analytical results and demonstrate our model performance to define ongoing progress and achieved our objectives.
- Demonstrated Key Performance Indicator (KPI) dashboards using Tableau.

ENVIRONMENT: Hadoop, HDFS, Hive, MapReduce, Sqoop, LINUX, MapR, Big Data, UNIX Shell Scripting, TWS, Python, SQL Server, Tableau, PySpark, Cassandra.

SARAG SYSTEMS PVT LTD. | | Hyderabad, India | | Data Engineer

Jan 2019 - Nov 2020

- Build scalable and reliable ETL systems to pull large and complex data together from different systems efficiently.

- Linked data lineage to data quality and business glossary work within the overall data governance program.
- Used HTML, CSS, Bootstrap, JavaScript, JQuery for developing controls and validations in the application
- Building the pipelines to copy the data from source to destination in Azure Data Factory (ADF V1).
- Experienced in designing and deployment of Hadoop cluster and different Big Data analytic tools including Pig, Hive, HBase, Kafka, Spark with Cloudera distribution.
- Transforming data in Azure Data Factory with the ADF Transformations.
- Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics.
- Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure) and processing the data in Azure Databricks. Experience in Moving Data in and out of Windows Azure SQL Databases and Blob Storage.
- Worked on Developing Data Pipeline to Ingest Hive tables and File Feeds and generate Insights into Cassandra DB.
- Worked on AI Processor piece of SOI. Well versed with building CI/CD pipelines with Jenkins, used tech stack like Gitlab, Jenkins, Helm, Kubernetes.
- Developed complex SQL queries for data analysis, and integrated BigQuery with other GCP services for comprehensive data solutions.
- Built and deployed machine learning models on GCP using AI Platform and TensorFlow, enhancing predictive analytics capabilities.
- Developed a disaster recovery strategy using GCP services such as Cloud Storage, Cloud SQL, and Compute Engine, ensuring business continuity.
- Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.
- Worked on installing, configuring, and monitoring Apache Airflow for running both batch and streaming workflows.
- Written Scala and python script notebooks for Azure Data bricks transformation task.
- Involved in converting Hive/aw queries into Spark transformations using Spark RDDs and Scala.
- Analyzed the SQL scripts and designed the solution to implement using Scala.
- Data Ingestion to one or more cloud Azure Services - (Azure Data Lake, Azure Storage, Azure DW) and cloud migration processing data in Azure Databricks.
- Good Understanding of Data ingestion, Airflow Operators for Data Orchestration and other related python libraries.
- Used Spark-Streaming APIs to perform necessary transformations.
- Worked with spark to consume data from Kafka and convert that to common format using Scala.
- Converted existing Map Reduce jobs into Spark transformations and actions using Spark RDDs, Data frames and Spark SQL APIs.

ENVIRONMENT: Hadoop, HDFS, Azure, Hive, Spark, Cloudera, Kafka, PySpark, GCP, Scala, Pig, Cassandra, Agile methods, MySQL, HDFS.

EDUCATION:

- **Master of Science in Computer Engineering** - Wright State University, Dayton, OH
- **Bachelor of Technology in Electronics and Communication Engineering** – Mahatma Gandhi Institute of Technology, India