Sankeerth Data Engineer

Mountainview, California, 94041 | 669-248-2551 | [sankeerth.danalytics@gmail.com](mailto:sankeerth.danalytics@gmail.com)

# Professional Summary

* Data Engineer with 10+ years of experience in Data Engineering, ETL, ELT, Data mining with large data sets of Structured and Unstructured data, and implementing production grade data warehousing solutions on large scale data technologies
* Strong experience in migrating other databases to Snowflake.
* 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.
* Experience in building Snowpipe and using Snowflake Clone and Time Travel.
* Experience im designing star schema and snowflake schema for Data Warehouse.
* Participates in the development improvement and maintenance of snowflake database applications
* Build the Logical and Physical data model for snowflake as per the changes required
* Define roles, privileges required to access different database objects.
* In-depth knowledge of Snowflake Database, Schema and Table structures.
* Define virtual warehouse sizing for Snowflake for different type of workloads.
* Hands-on experience with Snowflake utilities, SnowSQL, SnowPipe, Connectors, Data sharing, Cloning and creating tasks.
* Develop transformation logic using snowpipeline.
* Strong knowledge in Data Warehousing concepts such as Schemas (Star, Snowflake), ETL & ELT process and dimensional modeling.
* Participate in design meetings for creation of the Data Model and provide guidance on best data architecture practices Experience with Snowflake Multi-Cluster Warehouses.
* Work with domain experts, engineers, and other data scientists to develop, implement, and improve upon existing systems.
* Optimize and use Performance tuning of Big Data workloads.
* Understand current Production state of application and determine the impact of new implementation on existing business processes.
* Experience on AWS Redshift, S3, Glue Catalog, Athena, Dynamo DB and Kinesis.
* Experience on migrating on-premise SQL database to Azure Data Lake, Azure SQL Database, Databricks and Azure SQL Data Warehouse using Azure Data factory & Controlling and granting database access.
* Operationalize data ingestion, data transformation and data visualization for enterprise use.
* Mentor and train junior team members and ensure coding standard is followed across the project.
* Help talent acquisition team in hiring quality engineers.

# Education

**MASTER’S IN COMPUTER SCIENCE |MAY 2017 | UNIVERSITY OF NEVADA, RENO**

* Minor: Statistics GPA: 3.9

**BACHELORS IN COMPUTER SCIENCE| AUG 2014 | PANINEEYA INSTITUTE OF TECH & SCIENCE**

* Minor: Math GPA: 3.7

# Skills & Abilities

**Programming Languages:** Python, Java, Scala, C, C++, Shell Scripting

**Databases:** SQL, Oracle, MySQL, PostgreSQL, Cassandra, Informatica, Hive, Presto, MS SQL Server

**Datawarehouse:** Snowflake, Amazon Redshift, Azure Data Warehouse, Teradata

**Big Data:** Hadoop, HDInsight, Apache Spark, Apache Kafka, Airflow, Mongo DB

**Azure Cloud:** Databricks, Data Factory, Data Lake, Delta Lake, Lakehouse, Cosmos DB, Synapse

**AWS Cloud:** Amazon S3, Redshift, Glue, Lambda, Athena, Dynamo DB, Kinesis **GCP Cloud:** Big Query, Cloud Storage, Data Studio, Big Table, Cloud Pub&Sub **ETL/ELT:** Data Warehouse, Data Marts, Data Model, Data Pipelines, ETL, ELT, DBT **BI Tools:** Tableau, PowerBI, Alteryx, DOMO, Looker.

**Soft Skills:** Communication, Data Management, Mentorship

# Experience

**SR. DATA ENGINEER MAY 2022 - PRESENT**

**AMERICAN AIRLINES MOUNTAINVIEW, CA**

* Extract Transform and Load data from on-premise sources systems to Azure Data Storage services using a combination of Azure Sertvice (Data Factory, T-SQL, Spark SQL and U-SQL Azure Data Lake Analytics. to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) for data ingestion and processing in Azure Databricks.
* 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.
* Use Azure Data Factory to extract, transform, and load data from various sources into Azure data storage services.
* Ingest data into one or more Azure services, including Azure Data Lake, Azure Storage, Azure SQL, and Azure DW, and process the data in Azure Databricks.
* Develop JSON scripts for deploying pipelines in Azure Data Factory (ADF) that process data using SQL activities.
* Developed Spark applications using Pyspark and Spark-SQL for data extraction, transformation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
* Analyze, design and build Modern data solutions using Azure PaaS service to support visualization of data.
* Understand current Production state of application and determine the impact of new implementation on existing business processes.
* Responsible for estimating cluster size, monitoring and troubleshooting of the Spark databricks cluster.
* Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
* Wrote UDFs in Scala and Pyspark to meet specific business requirements
* Developed JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data using the Sql Activity. Hands-on experience on developing SQL Scripts for automation purpose.
* Creating Databricks notebooks using SQL, Python and automated notebooks using jobs.
* Creating Spark clusters and configuring high concurrency clusters using Azure Databricks to speed up the preparation of high-quality data.
* Worked extensively on the migration of different data products from Oracle to Azure
* Spun up HDInsight clusters and used Hadoop ecosystem tools like Kafka, Spark and data bricks for real-time analytics streaming, Sqoop, pig, hive and Cosmos DB for batch jobs
* Deployed the developed pipelines using Azure DevOps for CI/CD.
* Database creation (DDL, DML, DCL), database tuning, SQL tuning, performance planning.
* Worked on SPARK for real time streaming of data into the cluster and loaded all datasets into Hive from Source files using Spark
* Developed SQL queries /scripts to validate the data such as checking duplicates, null values, truncated values and ensuring correct data aggregations.
* Performed data quality analysis using advanced SQL skills. Performed Count Validation, Dimensional Analysis, Statistical Analysis and Data Quality Validation in Data Migration.
* Extensive SQL querying for ETL pipelines on Staging, Data warehouse and Data Mart.
* Strategic expertise in design of experiments, data collection, analysis and visualization
* Explored data in a variety of ways and across multiple visualizations using Power BI.
* Developed Power BI reports and dashboards from multiple data sources using data blending.
* Designed Power BI data visualization utilizing cross tabs, maps, scatter plots, pie, bar and density charts.
* Extensive use of DAX (Data Analysis Expressions) functions for the reports and for the tabular models.
* Created Power BI reports by using Joins in multiple tables from multiples database using complex SQL queries.
* Designed a Power BI data model with multiple fact tables and dimensions depending on the business requirements.
* Experience in custom visuals and groups creation in Power BI.
* Experience in scheduling refresh of Power BI reports, hourly and on-demand.
* **Environment:** Azure Data Bricks, Data Factory, Spark SQL, Azure Data Lake Storage, Azure Data Explorer, Azure Synapse, Delta Lake, Lakehouse, Cosmos DB, Power BI Desktop, Power BI Pro, Tableau Desktop, Tableau Server

**SR. DATA ENGINEER APR 2021 – MAY 2022**

**FEDEX DALLAS, TEXAS**

* Worked on Apache Spark with Python (PySpark, Spark SQL) to develop and execute Big Data Analytics and Machine learning applications, executed machine learning use cases under Spark ML and MLlib
* Architect & Implement medium to large scale BI solutions on Azure using Azure Data Services (Data Lake, Data Factory, Stream Analytics, Azure SQL DW, Azure Databricks & Cosmos DB)
* Design, Setup & Maintain Administrator for Azure SQL Database, Azure Data Warehouse and Data Factory
* Development of Spark structured streaming to read the data from Kafka in real time and batch modes, apply different mode of Change data captures (CDCs) and then load the data into Hive
* Developed & maintained end-to-end ETL data pipelines on large datasets in Azure Data Factory.
* Developed and Configured Kafka brokers to pipeline server logs data into spark streaming
* Created Stream sets pipeline for event logs using Kafka, Stream sets Data Collector and Spark Streaming in cluster mode by customizing with mask plugins, filters and distributed existing Kafka topics across applications using Stream sets Control Hub.
* Develop and maintain data pipelines using Azure Data Factory and Azure Databricks.
* Automated monthly data purge processes through Azure Data Lake and Azure Data Factory, resulting in decreased storage costs by 25%.
* Developed spark code and Spark SQL/streaming for faster testing and processing of data.
* Developed Python scripts to automate the ETL process using Apache Airflow
* Deployed the pipelines using Azure DevOps for CI/CD.
* Implemented complex DAX functions for various fact calculations for efficient data visualization in Power BI on Healthcare shipments.
* Automated Power Query refresh using power shell script and windows task scheduler.
* Scheduled Automatic refresh and scheduling refresh in power bi service.
* Worked on timely reports such as Yearly, Quarterly, Monthly, and Daily for multiple teams.
* Created Parameterized Report, Dashboard Report, linked report and Sub Report by Year, Quarter, Month, and Week. Created Drill Down Reports, Drill Through Report by Region.
* Designed and developed Power BI graphical and visualization solutions with business requirement documents and plans for creating interactive dashboards.
* Gathering the requirements and writing the ad-hoc SQL query batches to update the data and metadata for delta and history loads on Healthcare shipments.
* Developed Tableau dashboards to design multiple scorecards and dashboards to display information required by different departments and upper-level management.
* Created calculated measures & dimension members using multi-dimensional expression (MDX).
* Worked closely with the team in meeting the deadlines pertaining to design and development deliverables.
* **Environment:** Azure Data Factory, Spark SQL, Azure Data Lake Storage, Azure Data Explorer, Azure Synapse, Azure Databricks, Power BI Desktop, Power BI Pro, Tableau Desktop, Tableau Server

**SR. DATA ENGINEER MAR 2019 – APR 2021**

**FACEBOOK MENLO PARK, CA**

* Migrated data from on-premise databases (Oracle, My SQL) to Snowflake through data files using Amazon S3 as external stage
* Converted Business logic into respective Snowflake functions
* Bulk loaded from external stage (AWS S3) to internal stage (Snowflake) using COPY command.
* Used import and export from internal stage (Snowflake) vs external stage (S3 Bucket).
* Worked with Snowflake utilities SnowSQL, Snow Pipe, Time Travel and Zero copy Clone, Copy command, Tasks etc.
* Worked on all activities related to the development, implementation, administration and support for large scale Snowflake cloud data warehouse.
* Created database objects like Snowflake tables, file formats, sequences, internal & external stages.
* Modifed existing queries and Writing Snowflake SQL queries.
* Extracted Real time feed using Kafka and Spark Streaming and convert it to RDD and process data in the form of Data Frame and save the data as Parquet format in HDFS.
* Developed real-time processing and core jobs using Spark Streaming with Kafka as a data pipeline system.
* Used Spark and Spark-SQL to read the parquet data and create the tables in Hive.
* Monitored Spark jobs and capture the logs for in Spark applications.
* Developed multiple Kafka Producers and Consumers as per the requirement specifications.
* Used Spark Streaming APIs to built Real-time Data Pipelines with Kafka Connect from Kafka in for building common learner data model which gets the data and persist it to Cassandra.
* Used Kafka and Kafka brokers, initiated the spark context and processed live streaming information with RDD and used Kafka to load data into HDFS and NoSQL databases.
* Staging the Kafka Data into Snowflake DB by flattening the same for different functional service.
* Developed stored procedures/views in Snowflake and Have Extracted and loaded data from AWS S3 to Snowflake Cloud Data Warehouse.
* Used Kafka functionalities like distribution, partition, replicated commit log service for messaging systems by maintaining feeds and
* Created applications using Kafka, which monitors consumer lag within Apache Kafka clusters.
* Created Pattern Recognition from source data for accounts, vendors and transactions.
* Used COPY, LIST, PUT and GET commands for validating internal and external stage files.
* Perform troubleshooting analysis and resolution of critical issues.
* Worked on streams and tasks to identify Change Data Captures.
* Worked with Spark for improving performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's.
* **Environment:** Snowflake, Amazon S3, Spark, Python, Spark SQL, PySpark, Spark Streaming, Hive, Apache Kafka, Oracle, My SQL, Power BI Desktop, Power BI Pro, Tableau Desktop, Tableau Server

**SR. DATA ENGINEER MAY 2017 – MAR 2019**

**DESIGNER SHOE WAREHOUSE COLUMBUS, OHIO**

* Developed the data model design and database migration deployment production releases to endure our database objects and corresponding metadata were successfully implemented to the production platform environments like Dev, Test and Prod using Snowflake.
* Performed day-to-day integration with the Database Administrators from IBM DB2, SQL Server, Oracle DB to ensure the insertion of database tables, columns and its metadata have been successfully implemented in all environments in AWS Aurora and Snowflake.
* Performed ETL data translation using informatica of functional requirements using Source to Target Mapping documents to support large datasets to Snowflake DB.
* Performed logical and physical data structure designs and DDL generation to facilitate the implementation of database tables and columns based on schema using ERwin Data Modeler.
* Assisted Project Managers and Developers in performing ETL solution design and development to produce reporting, dashboarding and data analytics deliverables.
* Created Data Dictionaries, supporting the four Business Units: Finances, Marketing, Retail, and Supply Chain.
* Worked on Architecture Design for Multistate implementation or deployment.
* Implemented one time Data Migration of Multistate level data from SQL server to Snowflake by using Python and SnowSQL.
* Developed ETL Pipelines in and out of data warehouse and generated regulatory and financial reports using advanced SQL queries in Snowflake and Tableau.
* Analyze requirements provided by various business users.
* Create Source and Target Definitions in the repository using Informatica
* Identify source systems, connectivity, tables, and fields to ensure data lineage for mapping.
* Created different levels of stored displays to help different groups, users and executives to easily understand the data model per their requirements.
* Designed, developed and documented Informatica mappings as per the requirements.
* Created mapplets, common functions, reusable transformations, look-ups.
* Implemented the error handling and exception reporting as per the user requirements.
* Tested the data and data integrity among various sources and targets.
* Implemented entire ETL job plan (50 jobs) through Informatica Scheduler and worked on a total of 70 mappings, 10 stored procedures.
* Developed reports and dashboards from multiple data sources such as on-premises and cloud using Power BI desktop.
* Created reports from importing, direct query and live data from several data sources.
* Created different measures, remove rows, replace values, split column, date & time column.
* **Environment:** Snowflake, Oracle, MS SQL Server, My SQL, Informatica Power Center 9.x, Tera Data 13.x, Oracle 10g/9i, Power BI Desktop, Power BI Pro, Tableau Desktop, Tableau Server

**DATA ENGINEER II JUN 2015 – MAY 2017**

**MILLENNIUM INFOTECH PRINCETON, NEW JERSEY**

* Developed ETL pipelines using Informatica to implement the business requirements.
* Communicated with business customers to discuss the issues and requirements.
* Created shell scripts to fine tune the ETL flow of the Informatica workflows.
* Used Informatica file watch events to pole the FTP sites for the external mainframe files.
* Performance tuning was done at the functional level and map level.
* Used relational SQL wherever possible to minimize the data transfer over the network.
* Effectively used Informatica parameter files for defining mapping variables, workflow variables, FTP connections and relational connections.
* Performed enhancements and maintenance activities of the data warehouse including tuning, modifying of stored procedures for code enhancements.
* Effectively worked in Informatica version based environment and used deployment groups to migrate the objects.
* Used debugger in identifying bugs in existing mappings by analyzing data flow, evaluating transformations.
* Designed workflows with many sessions with decision, assignment task, event wait, and event raise tasks, used Informatica Scheduler to schedule jobs.
* Reviewed and analyzed functional requirements, mapping documents, problem solving and trouble shooting.
* Performed unit testing at various levels of the ETL and actively involved in team code reviews.
* Created logical data model from the conceptual model and its conversion into the physical database design using Erwin Data Modeler.
* Developed Data mapping, Transformation and Cleansing rules for the Data Management
* Created Power BI reports from various data sources such as SQL Server, Oracle & Teradata.
* Using a query editor in Power BI to perform complex transformations to create Business KPIs.
* Working with import and direct query, creating custom table in the Power BI.
* Developed and published reports in Power BI using DAX formulas and expressions.
* Created several user roles and groups to the end-user and provided row level security to them.
* Developed Tableau dashboards using bars, lines, scatter plots, bubbles and histograms.
* Combined views and reports into interactive dashboards in Tableau Desktop that were presented to Business Users, Program Managers, and End Users.
* Reviewed basic SQL queries and edited inner, left, and right joins in Tableau Desktop by connecting live/dynamic and static datasets.
* Developed story telling dashboards in Tableau Desktop and published to Tableau Server.
* Scheduled data refresh on Tableau Server for weekly and monthly increments based on business change to ensure changed data is displayed accurately.
* **Environment:** Informatica Power Center, Oracle, MS SQL Server, My SQL, Tera Data 13.x, Power BI, Erwin Data Modeler, Power BI Desktop, Power BI Pro, Tableau Desktop, Tableau Server

**ETL ENGINEER AUG 2012 – DEC 2014**

**HONEYPOT LLC HYDERABAD, INDIA**

* Implemented action filters, parameters, calculated filed and set for preparing dashboards and worksheets in Tableau.
* Created Tableau scorecards, dashboards using Stack bars, bar graphs, scattered plots, geographical maps, Gantt charts using show me functionality.
* Worked on Tableau Desktop, by creating Complex reports & Dashboards.
* Created procedures to ensure the quality, usability and accuracy of reports.
* Validated tableau reports by writing SQL queries and comparing to results from DB.
* Scheduling the monthly/weekly/daily reports to run automatically.
* Collaborated with business stakeholders to gather report requirements, design report layouts, and deliver high-quality and visually appealing reports.
* Developed complex store procedures, efficient triggers, required functions, and creating indexes and indexed views for performance.
* Designed ETL data flows using SSIS, creating mappings/workflows to extract data from SQL Server, and Performed Data Migration and Transformation from Access using SQL Server SSIS.
* Created SSIS jobs to automate the reports generation, cube refresh packages.
* Developed stored procedures and triggers to facilitate consistent data entry into the database.
* Monitored performance using SQL Server Performance tuning in SQL Server
* Involved in database design, data modeling, T-SQL development; developed and implemented ETL packages using SSIS and complex reports
* Created and maintained SSIS package (ETL) to extract and transform data, and to load into staging database.
* Worked on all activities related to the development, implementation, administration and support of ETL processes for large-scale data warehouses using SSIS.
* Participated in the creation of data standards, policies and procedure to ensure data quality.
* Performed Error & Event Handling: Precedence Constraints, Break Points, Check Points, Logging.
* Implemented Features, Attributes, Hierarchies, Star & Snowflake Schemas for Data Marts.
* Developed SSAS Cubes, Aggregation, KPIs, Measures, Partitioning Cube, Data Mining Models, and Deploying and Processing SSAS objects.
* Created Ad hoc reports and reports with complex formulas and querying the database for Business Intelligence.
* Designed, Developed, and managed both paper-based and interactive Web-based reports with SQL Server Reporting Services (SSRS).
* Developed Parameterized, Charts, Graph, Linked, Dashboard, Scorecards, and Report on SSAS Cube using Drill-down, Drill-through, and Cascading reports using SSRS.
* **Environment:** Tableau Desktop, Tableau Server, Oracle, My SQL, MS Access, MS SQL Server, SQL Server Integration Services (SSIS), SQL Server Analysis Services (SSAS), SQL Server Reporting Services (SSRS), SQL, Tableau, MS Excel, Python, UNIX/Linux.