**VINEETH KUMAR SENIOR SNOWFLAKE ENGINEER**

**- 234-738-0297 ∙  - vineethg724@gmail.com**

**SUMMARY**

* Imperative IT experience in all aspects of **BigData**, **ETL/ELT**, **Cloud** **Database** and Migration, business Analytics applications including design, development validation, deployment, reporting & analysis.
* Experience in **snowflake** cloud database,Amazon Redshift and data warehousing application tools like **Informatica**, Oracle Data Integrator (**ODI**) working on all phases of data warehouse development life cycle, from gathering requirements to testing, implementation and support.
* Experience in designing **Dimensional Modeling**, creating Relational Database, **Data warehouse solutions** and understanding Entity Relationship Diagram on several databases and reporting systems.
* Partnered with other engineering teams to help architect and build the data pipeline that ingest hundreds of billions of data points for Field Analytics Platform utilizing AWS.
* Experience on working with DBT CLI Transformation Tool
* Experience on working with Modern Replication Tools like Fivetran and HVR
* Extended capability using various open-source data technologies like Hadoop, Kafka, Spark.
* Experienced in implementation of Cloud technologies like **Oracle**, **Azure,** and **snowflake**.
* Experience in **Apache NIFI** which is a Hadoop technology and Integrating Apache NIFIand Apache Kafka.
* Experience in reading data from S3 bucket **RAW** layer and loading into target S3 bucket **Process** layer with transformed data using python code and then load into Redshift tables.
* Working on **AWS CLI** commands for load Redshift warehouse, **HDFS**, Scripting to handle files Gzip, backup, and transfer to **edge node**... etc. operational requirements.
* Extensive experience in using **Informatica** (Power Center) and Oracle Data Integrator (**ODI**) for implementation of **ETL** methodology in Data Extraction, **Transformation** and **Loading**.
* Extensive **SQL** Development experience in **Health**, **Banking**, **Insurance**, **Financial** industries with a strong understanding of data & analytics.
* Experience in different features in **Snowflake** like data share, **time** **travel**, **failsafe**, **zero** **copy** **clone**, **snowflake** **procedures** and standards for data pipelining and integration with **Snowflake** data warehouses.
* Experience in **designing and developing efficient Error handling methods** for **ETL/ELT** mappings and workflows to load the data from the various sources like **MSSQL**, **DB2**, **Files**, **Teradata, Oracle**, **Postgres** into dimensions and facts.
* Good at **analyzing**, **requirement** **gathering**, documenting, and editing Business/User Requirements. Also worked in distinct Business units, Process units, Review processes etc.
* Used ODI Designer to develop complex interfaces (mappings) to load the data from the various sources like **Oracle**, **DB2,** **and SQL Server** etc. into dimensions and facts.
* Implemented the Change Data Capture (**CDC**) feature of **ODI** to minimize the data load times.
* Experienced in designing, tuning, and leveraging large data warehouses.
* Good knowledge in working with FTP and development tools like **SQL** developer, **TOAD** and **Eclipse**. Experienced in text processing tools such as **AWK.**
* Working experience on **Json, Avro, Parquet, Orc, Csv** file structures based on business requirement to have best-optimized data operations on multiple pipelines.
* Experience in loading into **Hive** stage and target tables from source **HDFS** files and configuring parameters for logs.
* Working on **AWS CLI** commands for BQ jobs, HDFS, Scripting to handle files **Gzip**, **backup**, and transfer to edge node... etc. operational requirements.
* Extensive experience in writing complex **SQL** Queries and **PL/SQL**, which supports Data Integration, processes.
* Performance **tuning** and query **optimization** techniques in transactional and data warehouse environments.
* Experienced in creating Tables, Stored Procedures, Views, Indexes, Cursors, Triggers, User Profiles, Relational Database Models and Data Integrity in observing Business Rules.
* Expert in data warehousing techniques/concepts such as data cleansing, Surrogate key assignment, Slowly Changing Dimensions **SCD TYPE 1**, **SCD TYPE 2**, **Change Data Capture (CDC)**.
* Familiar with **Oracle EBS** concepts of Financial Analytics, Spend Analytics and Supply Chain modules.
* Expertise in configure Reports / Dashboards with different Views (Drill-Down, Pivot Table, Chart and Column Selector, global and local Filters) with developing Dashboards with drill-down &drill-across capabilities, using **Tableau** and **Oracle Business Intelligence Tool** (**OBIEE**).
* Providing remote and onsite support for customers.
* A Self-starter with a positive attitude, willingness to learn new concepts and to accept challenges, as well a very good team player.
* Excellent communication skills, versatile team player with excellent interpersonal skills.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Languages** | PySpark, Scala, Python, SQL, PL/SQL, Shell Scripting. |
| **Cloud Technologies** | AWS, Snowflake, GCP, Azure, Oracle Cloud. |
| **Big data** | Cloudera, HDFS, SQOOP, Kafka, Hive, Impala, Spark, Flume, Glue |
| **ETL Tools** | ODI (Oracle Data Integrator) 12c, 11x, OWB 10x, Informatica. 10.x/9.x, Fivetran, DBT |
| **Databases** | Oracle 11g/12c, MY SQL, Snowflake, HBase, Redshift. |
| **Data Visualization Tools** | OBIEE 11G/12c, Tableau |
| **Version Control** | Git hub, Bitbucket, SVN. |
| **Scheduling Tools** | Crontab, Apache Airflow |
| **Operating System** | Windows, UNIX. |
| **Miscellaneous** | Microsoft office, SQL Developer, WINSCP, PUTTY. |

**Snowflake Developer | Credit Acceptance | Michigan | May 2022 – till date**

**WORK EXPERIENCE**

**Project Description:** is an auto finance company providing automobile loans and other related financial products. The company operates its financial program through a national network of dealer-partners, the automobile dealers participating in the programs.

**Key Contributions:**

* Created various stages like internal/external stages to load data from various sources like AWS S3 and on-premises systems.
* Gather the different volumetrics like Total number of tables, space occupied by different tables, index size, total occupied size, number of partition groups etc.
* ﻿﻿Migrating the tables using Fivetran Tool and enabled the CDC using Fivetran.
* Migration of Zendesk, salesforce, Jira API data using Fivetran replication tool.
* Transformation of Data using DBT CLI Tool.
* Responsible for all activities related to the development, implementation, administration, and support of ETL processes for large scale Snowflake cloud data warehouse.
* Worked on Snowflake Cloud Database and Integrated Automated generic Python Framework to process XML, CSV, JSON, TSV, TXT files.
* Transformation of Data using DBT CLI Tool.
* Implementation of SCD2 using snapshot functionality in DBT.
* Develop stored procedure/views in Snowflake and used Talend for loading Dimensions and Facts.
* Bulk loading from external stage AWS S3 to internal stage Snowflake using COPY command.
* Developed various views using windowing functions like row\_number, rank and dense\_rank.
* Used COPY, LIST, PUT and GET commands for validating internal and external stage files.
* Contributed to roadmaps for Enterprise Cloud Data Lake architecture and worked with Data Architects of various departments to finalize the roadmap for **Data Lake**.
* Writing complex Snowflake SQL scripts in Snowflake cloud data warehouse to Business Analysis and reporting.
* Utilized AWS DataSync to transfer large amounts of data between on-premises storage and S3.
* Designed and implemented S3 bucket policies to ensure proper access controls and security.
* Utilized S3's versioning feature to maintain a historical record of changes to objects.
* Configured DataSync to run in a fully automated and scalable manner.
* Develop framework for converting existing PowerCenter mappings to PySpark jobs and provide guidance to development team working on **PySpark** as ETL platform.
* Design and develop ETL integration patterns using Python on Spark and created PySpark frame to bring data from DB2 to **AWS S3**.
* Develop data warehouse model in Snowflake for over 100 datasets using WhereScape and created reports in Looker based on Snowflake Connections.
* Created **Snowpipe** on S3 bucket to load data into Base tables and have strong understanding on Snowpipe configuration for continuous data ingestion.
* Validating the data from **SQL Server** to Snowflake to make sure it has a perfect match and implemented Change Data Capture technology in Talend to load deltas to a Data Warehouse.
* Design and Develop ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift.
* Data Extraction, aggregations and consolidation of Adobe data within AWS Glue using PySpark.
* Create external tables with partitions using Hive, AWS Athena and Redshift
* Designed External and Managed tables in Hive and processed data to the HDFS using Sqoop
* Create user defined functions UDF in Redshift
* Migrate Adobe Marketing Campaign data from Oracle into HDFS using Hive, Pig, Sqoop Defined virtual warehouse sizing for Snowflake for different type of workloads and built the Logical and Physical data model for Snowflake as per the changes required.
* Used **UNIX** for Automatic Scheduling jobs, Involved in Unit Testing of newly created **PL/SQL** blocks of code.
* Created and maintained ETL jobs using AWS Glue to transform data from various sources.
* Developed custom Glue scripts to handle complex data transformations and business logic.
* Built and deployed serverless applications using AWS Lambda and API Gateway., Utilized Lambda functions to automate data processing and orchestration tasks. Develop SQL queries in **SnowSQL** and coding for Stored Procedures/Triggers, transformation logic using Snow pipe. Optimize and fine tune queries and performance tuning of Big data workloads.
* Strong understanding of various data integration patterns to **Hadoop**, Snowflake Multi-Cluster Warehouse, Size, and credit usage. Utilized Spark SQL API in PySpark to extract and load data and perform SQL queries.
* Worked on developing Pyspark script to encrypting the raw data by using hashing algorithms concepts on client specified columns. Responsible for Design, Development, and testing of the database and Developed Stored Procedures, Views, and Triggers.
* Creating data model that correlates all the metrics and gives a valuable output.
* Performing ETL testing activities like running the Jobs, Extracting the data using necessary queries from database transform, and upload into the Data warehouse servers.
* Pre-processing using Hive and Pig.
* Access, Excel, CSV, Oracle, flat files using connectors, tasks and transformations provided by AWS Data Pipeline.

**Environment:** Python, Pyspark Scala, Snowflake, Amazon Web Services, Hadoop, PySpark, Fivetran, Data Lake.

**Snowflake Developer | Health Care Service Corporation | Chicago, IL | May 2021 – April 2022**

**Project Description:** HCSC is the fifth-largest health insurer in the US overall and employs more than 23,000 people. As of 2019, it was noted to be the third-largest commercial health insurer in the United States. It serves nearly 16 million members. HCSC offers group life, disability, and dental policies as well as a range of other individual policies. The company also provides various care management and wellness resources.

**Key Contributions:**

* Analyze and troubleshoot diverse issues within Hadoop ecosystem tools, identify the performance issues, bottlenecks in production environment and recreate complex customer and production reported issues to determine root cause and validate the solution. ﻿﻿
* Migrating the tables using Fivetran Tool and enabled the CDC using Fivetran.
* Worked closely with executives, senior leadership teams, management and business analysts in business process modeling and mapping, and the understanding of business requirements.
* Transformation of Data using DBT CLI Tool.
* Migrated data from on-premises SQL Database to Azure Synapse Analytics using Azure Data Factory, designed optimized database architecture
* Created Azure Data Factory for copying data from Azure BLOB storage to SQL Server
* Implement ad-hoc analysis solutions using Azure Data Lake Analytics/Store, HDInsight/Databricks
* Work with similar Microsoft on-prem data platforms, specifically SQL Server and SSIS, SSRS, and SSAS
* Created Reusable ADF pipelines to call REST APIs and consume Kafka Events.
* Performed Kafka analysis, feature selection, feature extraction using Apache Spark Machine.
* Created entity relationship diagrams and multidimensional data models for merging Confidential and Whitefence data sets into one single DataWarehouse using Embarcadero ER/Studio
* Created logical and physical data model for Online Campaign Data Management using ERstudio and Visio
* Applied data cleansing/data scrubbing techniques to ensure consistency amongst data sets.
* Lowered processing costs and time for Order processing system through the review and reverse engineering of existing database structures, whigh reduced redundancies and consolidated databases.
* Migrate data into RV Data Pipeline using DataBricks, Spark SQL and Scala.
* Migrate Confidential Callcenter Data into RV data pipeline from Oracle into HDFS using Hive and Sqoop
* Build Data migration processes using SQL Server as database and SSIS as ETL
* Design, Develop, Test and Maintain Allconnects Data Warehouse which is build in Oracle 12c
* Load data into Amazon Redshift and use AWS Cloud Watch to collect and monitor AWS RDS instances within Confidential O Designed and developed ETL/ELT processes to handle data migration from multiple business units and sources including Oracle, Postgres, Informix, MSSQL, Access and others.
* Used AWS Redshift, S3, Spectrum and Athena services to query large amount data stored on S3 to create a Virtual Data Lake without having to go through ETL process. Developed ETL processes to move data between RDBMS and NoSQL data storage.
* Used Control-M for scheduling DataStage jobs and used Logic Apps for scheduling ADF pipelines
* 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. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks.
* ﻿﻿Migrating the tables using Fivetran Tool and enabled the CDC using Fivetran.
* Responsible for writing Hive Queries to analyze the data in Hive warehouse using Hive Query Language (HQL).
* Extracted the data and updated it into HDFS using Sqoop Import from various sources like Oracle, Teradata, SQL server etc.
* Created Hive staging tables and external tables and joined the tables as required, involved in developing Hive DDLs to create, drop and alter tables and dynamic Partitioning, Static Partitioning and Bucketing.
* Installed and configured Hadoop Map Reduce, Hive, HDFS, Pig, Sqoop, Flume and Oozie on Hadoop cluster.
* Worked on Microsoft Azure services like HDInsight Clusters, BLOB, ADLS, Data Factory and Logic Apps and done POC on Azure Data Bricks.
* Implement ad-hoc analysis solutions using Azure Data Lake Analytics/Store, HDInsight/Databricks
* Developing and configuring Build and Release (CI/CD) processes using Azure DevOps, along with managing application code using Azure GIT with required security standards for .Net and java applications.
* Implemented Sqoop jobs for data ingestion from the Oracle to Hive.
* Worked with various formats of files like delimited text files, click stream log files, Apache log files, Avro files, JSON files, XML Files. Mastered in using different columnar file formats like RC, ORC and Parquet formats.
* Developed custom the Unix/BASH SHELL scripts for the purpose of pre and post validations of the master and slave nodes, before and after the configuration of the name node and data nodes respectively.
* Developed job workflows in Oozie for automating the tasks of loading the data into HDFS.
* Implemented compact and efficient file storage of big data by using various file formats like Avro, Parquet, JSON and using compression methods like GZip, Snappy on top of the files.
* Exploring with Spark, improving performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame and Pair RDD's.
* Worked on Spark using Python as well as Scala and Spark SQL for faster testing and processing of data.
* Developed data warehouse model in snowflake for over 100 datasets using whereScape.
* Migrated Map reduce jobs to Spark jobs for achieving a better performance.

**Environment:** Hadoop, HDFS, Microsoft Azure services like HDinsight, BLOB, ADLS, Logic Apps etc., Hive, Sqoop, SnowFlake, Apache Spark, Spark-SQL, ETL, Maven, Fivetran, DBT, Oozie, Java, Python, Unix shell scripting.

**Snowflake Developer | Express Scripts| Austin, TX | Dec 2019 – April 2021**

**Project Description:** Express Scriptsis a pharmacy benefit management organization. It was the 22nd-largest company in the United States by total revenue as well as the largest (PBM) organization in the United States.

**Key Contributions:**

* Worked closely with executives, senior leadership teams, management and business analysts in business process modeling and mapping, and the understanding of business requirements.
* Created the business models from business cases and enterprise architecture requirements for process monitoring, improvement and reporting and led the team in business intelligence solutions development
* Developed and executed a migration strategy to move Data Warehouse from an Oracle platform to AWS Redshift. o Used BI Tools such as ThoughtSpot and SAP Tools to create and maintain Reports
* SAP Data Services Integrator ETL developer with strong ability to write procedures to ETL data into a Data Warehouse from a variety of data sources including flat files and database links (Postgres, MySQL, Oracle).
* Loaded data to STAR schemas (fact, bridge and dimension tables) for use in organization wide OLAP and analytics and written batch Files and Unix scripting to automate data load processes
* Knowledge of extracting data from sources such as Google and Bing AdWords and Analytics using Java API into Datawarehouse
* Migration of Zendesk, salesforce, Jira API data using Fivetran replication tool.
* Experience in performing transformations and actions on **RDD**, **Data frames, Data sets** using **Apache spark**.
* Configured and managed data access policies and permissions using AWS Lake Formation.
* Implemented data lake governance practices to ensure compliance and data security.
* Utilized AWS Glue Data Brew to clean and prepare data for analysis and visualization.
* Created and maintained Data Brew recipes to automate data preparation workflows, Deployed and managed relational databases on Amazon RDS, including MySQL and PostgreSQL.
* Configured RDS instances for high availability, performance, and security.
* Good Knowledge of Spark and Hadoop Architecture and experience in using **Spark** for data processing.
* In-depth knowledge of **PySpark** and experienced in building the **Spark applications** using **Python**.
* Developed pipelines using **python** to **migrate historical data** from On-premises Hadoop Hive tables to **S3** bucket in **AWS** and load into Redshift tables in batch loads by triggering though apache airflow.
* Defining for table’s schema’s, and views definition in **Redshift** for analytical and reporting requirements.
* Experience in creating and loading Stages in **AWS S3**. Have knowledge on different storage class and life cycle policy implementation.
* Hands on experience on AWS cloud services like **EMR, Glue, lambda.**
* Created **Hive Managed** and **External tables** with partitions and bucketing with **ORC** format in Hive and loaded data in to Hive tables.
* Hands on experience in using **AWS Cloud Watch** for monitoring the logs of both **EKS** and **EC2** instances and configured alerts from Cloud watch.
* Developed shell scripts on AWS CLI for compression with Gzip, backup, transfer to edge node with all necessary file operational requirements for **Redshift** jobs.
* Working with **JSON, Avro, ORC and Parquet** file in HDFS systems.
* Working with HDFS config files for application logs Yarn-site.xml, yarn-default.xml. mapred-site.xml and setting up log-aggregation properties in config files.
* Created **Dags** and **scheduling** jobs through Apache **Airflow** triggers.
* Prepared the design document by gathering the requirements from the Business Architect.
* End to End Development of Tableau reports based on user requirement like QA, QC, Data Validation, Functionality testing and Deploy **Tableau** reports to Stagging, Dev to Production.
* Table calculation, Level of Detail (LOD) over different calculated field using parameters.
* Developing Tableau data **visualization** using cross tabs, heat maps, waffle charts, geographic map, pie charts and bar charts, Word cloud, pareto chart etc...

**Environment:** Redshift, AWS S3, AWS CLI, Airflow, Confluence, Hive, HDFS, Py-Spark, Putty, IntelliJ, GitHub, Linux, Tableau.

**Snowflake Developer | American International Group | Forth-Worth, TX | Aug 2017 – July 2019**

**Project Description:** American International Group, Inc., also known as AIG, is an American multinational finance and insurance corporation with operations in more than 80 countries and jurisdictions.

**Key Contributions:**

* Analysis of the specifications provided by the clients.
* Prepared the content for installation and configuration of **Spark** using **Eclipse** and **Jupyter Notebook.**
* Developed the hands-on exercises in **spark** using java for all the transformations.
* Designed exercises tasks on using the **Parquet** File Format with **Impala Tables** along with queries
* Documented few trouble shooting scenarios along with workaround.
* Migrated the existing data to Hadoop from (My SQL and Oracle) using **Sqoop** for transfer the data and there by using, spark for transformation to load data to hive.
* Experience on null handling, decoding feature vectors using hash lookups. Involved on defining **RDD** structures for core spark and Data frame-based **SQL** processing in **Spark SQL.**
* Familiar with Data Ingestion pipeline design and advanced data processing tools.
* Created **Hive** tables and queried enormous amounts of data to apply the transformations done on data samples in **Pig** on a large scale of data distributed across the cluster.
* Automated all the jobs, for pulling data from FTP server to load data into **Hive** tables, using **Oozie** workflows.
* Created a chain for **Map/Reduce** jobs to process multiple inputs, join on Map side and Reduce side.
* Load and transform large datasets of structured and semi structured using various input formats.
* Designed exercises tasks on using the Parquet File Format with **Impala** Tables along with queries.
* Documented few trouble shooting scenarios along with workaround.
* Migrated an existing on-premises application to **AWS.**
* Used AWS services like **EC2** and **S3** for small data sets.
* Used Cloud watch logs to move application logs to S3 and create alarms based on a few exceptions raised by applications.
* 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.
* Primarily involved in Data Migration using SQL, SQL Azure, Azure Storage, and Azure Data Factory, SSIS, PowerShell.
* Architect & implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream.
* Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB).

**Environment:** AWS EC2, S3Hadoop, Spark, HDFS, Hive, Pig, Oozie, Sqoop, MapReduce, Linux, MySQL Workbench, Java**,**Eclipse, SQL.

**Snowflake/ODI Developer | IBM India Pvt Ltd | Hyderabad, India| Apr 2015 – Jul 2017**

**Project Description:** International Business Machines Corporation (IBM) is an American multinational Technology and consulting company headquartered in [Armonk, New York](https://en.wikipedia.org/wiki/Armonk,_New_York), with more than 350,000 employees serving clients in 170 countries.

**Key Contributions:**

* Gathered and analyzed the business requirements and defined their relationships with existing Business entities and provided status reporting of team activities against the plan or schedule, inform task accomplishment, issues, and status.
* Stakeholder matrix mapping to understand needs and project influence for increased technology adoption.
* Interacted with Business Analysts to understand the requirements and the impact of the ETL on the business.
* Used **ODI 12c** for extraction, loading and transformation ETL of data in the data warehouse.
* Led the team in business intelligence solutions development.
* Used Yellow Interfaces Reusable mappings in **ODI 12c** for developing most complex interfaces.
* Worked closely with the **ETL** and Database Team to coordinate the Necessary Data movement
* Worked with multiple sources such as Relational database, Flat files for Extraction, Transforming and Loading data into target warehouse tables.
* Experience with Snowflake **Multi-Cluster** Warehouses and Snowflake Virtual Warehouses.
* Loading and unloading data in Snowflake using Snow pipe, Copy Command.
* Loading and processing **JSON** data in Snowflake using the explicit Variant feature.
* Prepared technical specifications to develop ODI ELT **mappings** to load data into various tables confirming to the business rules.
* Implemented **Performance tuning** for the jobs which are running long in both ODI and DB side.
* Worked on the ETL processes, generating load plans, and running them.
* Customized ODI **Knowledge modules** to support snowflake and achieve requirements. Extracted data from specific technology, transforming the data, checking it, integrating it, etc.
* Creation of best practices and standards for data pipelining and integration with **Snowflake** data warehouses.
* Experience in designing the warehouse architecture in Snowflake, creating, and loading Stages in Snowflake.
* Created and implemented ER models and dimensional models (star schemas).

**Environment:** Snowflake, ODI 12.1.2.0, SQL, WinSCP, Putty, Linux.

Master of Science, Computational Data Science.

**EDUCATION**

**Kent State University | USA.**

Bachelor of Science, Computer Science.

**JNTU | India.**