**Vayunandan Sanam**

**sanamds538@gmail.com**

**(313) 800-1543**

**SUMMARY**

* Snowflake Certified with over 10 years of professional experience in IT across Telecom, Banking domains using Python and Big data frameworks expertise in big data development using Hadoop and Spark ecosystems.
* Worked with different Hadoop and Spark ecosystem components such as HDFS, Map Reduce, HBase, Spark, Yarn, Kafka, Pig, Hive, Sqoop, Impala, Nifi and Flume.
* Experience in working with Teradata and making the data to be batch processing using distributed computing.
* Expertise in data architecture including data ingestion pipeline design, data modeling, processing and data mining.
* Hands-on experience with Snowflake utilities, SnowSQL, Snow Pipe, and Big Data model techniques using Python / Java.
* Developed multiple real-time producer and consumer applications using Spark and Kafka.
* Development of Jobs in the Matillion ETL tool.
* ETL pipelines in and out of the data warehouse using a combination of Python, Snowflake, and Matillion Writing SQL queries against Snowflake.
* Experience in building Snow pipes.
* Define virtual warehouse sizing for Snowflake for different types of workloads.
* Extensive experience working with and developing PySpark applications to integrate various RDBMS systems like Teradata, Oracle, SQL Server and MySQL database.
* Experience in using Snowflake Clone and Time Travel.
* Expertise in Batch and Real-time data processing on Spark Streaming using Lambda architecture.
* Expertise in job work-flow scheduling using Crontab, Oozie and Airflow.
* Very good understanding of NoSQL databases like HBase and Cassandra.
* Hands-on experience with message brokers such as Apache Kafka and RabbitMQ.
* Expertise in writing custom UDFs in Pig & Hive Core Functionality.
* Hands on experience dealing with AVRO and Parquet file format, following best Practices and improving the performance using Partitioning, Bucketing, and Map side-joins and creating Indexes.
* Experienced in working with AWS EMR and using the EMR cluster and various EC2 instance types based on requirements.
* Developed PySpark Scripts in Spark Streaming to process huge data from data lakes like Kafka, s3, Kinesis and constructed ETL Pipelines to feed databases and dashboards.
* Developed robust ETL pipelines in Azure data factory to move data from on perm to Azure SQL Data warehouse.
* Developed visual reports, dashboards and KPI scorecards using Power BI desktop.
* Perform DAX queries and functions in Power BI.
* Experience in Data modeling techniques like star and snowflake schemas using Erwin tool.
* Experience in working on SSIS, SSAS & SSRS project packages.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Hadoop/Big Data Technologies** | HDFS, Map Reduce, Sqoop, Flume, Pig, Hive, Oozie, Impala, Zookeeper, Ambary, Storm, Spark and Kafka |
| **Databases** | Oracle, MY SQL, MS SQL server, Teradata |
| **Web Technologies** | HTML, XML, JSON, JavaScript |
| **Web Services** | Apache, Cassandra, Tomcat |
| **Programming and Scripting** | JavaScript, Shell Scripting, Python, Scala |
| **Operating Systems** | Linux, Unix, Windows 8, Windows 7, Windows Server 2008/2003 |
| **Cloud Computing** | AWS – EC2, S3, Redshift, Athena, Lambda functions, Cloud Watch  Azure – Azure HDInsight, Data Bricks (ADBX), Data Lake (ADLS), Azure AD, Data Factory, Azure Active Directory, Azure Functions, Azure Stream Analytics. |
| **AWS Components** | EC2, S3, RDS, Redshirt, EMR, DynamoDB, Lambda, RDS, SNS, SQS, Event Bridge, Aurora, API Gateway, AppSync |
| **ELT/ELT Capability** | SSIS, Talend, Informatica |
| **Visualization & Reporting** | Tableau, Power BI |
| **Version Control** | Git, SVN |

**Professional Experience:**

**T-Mobile, Kansas City, MO Aug 2021 - Present**

**Sr. Snowflake Developer**

**Responsibilities:**

* Involved in developing a generic framework using Scala and Spark that can use does develop applications for pre-process, process and reconciliation of transaction data.
* Worked with Data scientists to gather the required information about the data, cleaned, processed, structured and dumped into traditional databases (ETL transformation).
* Implemented advanced procedures of feature engineering for data science team using the in-memory computing capabilities like Apache Spark written in Scala.
* Implemented Spark scripts in Python to perform extraction of required data from the data sets and store it on HDFS.
* Developed spark scripts and python functions that involve performing transformations and actions on data sets.
* Configuring Spark Streaming in Python to receive real time data from Kafka and store it on to HDFS.
* Converted 230 views queries from Sql server🡪snowflake compatibility.
* Developed and implemented data pipelines using AWS services such as S3, EMR, Glue, Athena, and Redshift to process peta byte-scale data in real time.
* Developed SQL scripts to analyze the data that is stored on Redshift.
* Developed Spark scripts by using Scala shell commands as per the requirement.
* Migrated existing python programs to Spark using Scala.
* Experience in setting up row-level security, network policies and user access.
* Have worked on enabling time travel for some permanent tables and have enabled data masking via enabling row level security and dynamic data masking to restrict data access based on the roles assumed
* Experience in configuring Data Sharing and Virtual Private Snowflake.
* Performed daily admin tasks of user registration, security configuration, and usage monitoring
* Snowflake architecture was considered during storage of data in databases.
* Setup full CI/CD pipelines so that each commit a developer makes will go through the standard process of the software lifecycle and gets tested well enough before it can make it to production
* Involved in developing CICD pipelines using Jenkins to automate the code deployment. These Pipeline involved AWS services like triggering lambda by Kafka topics and sending a notification about data dump in Athena.
* Worked on building centralized Data repository Storage such as Data Lake on AWS Cloud utilizing primary services like S3, EMR, Redshift and Athena.
* Worked on ETL Migration services by developing and deploying AWS Lambda functions for generating a server less data pipeline which can be written to AWS Glue Catalog and queried from Athena.

**Environment:** Snowflake, Redshift, SQL server, BI Architect, AWS, AZURE, TALEND, JENKINS, Spark, Spark SQL, Python, PySpark, Databricks, AWS Services, EC2, S3, Cloud Watch, Athena, Step Functions, Snowflake, MySQL, PostgreSQL, Oracle, Git.

**Citi Group, Tampa, FL May 2018 - Jul 2021**

**Snowflake Developer**

**Responsibilities:**

* Created data pipeline for several events of ingestion, aggregation, and load consumer response data from AWS S3 bucket into Hive external tables.
* Worked on Oracle Databases, Redshift, and Snowflake.
* Creating Lambda SQS queues to create event-based triggers for Matillion
* Create conceptual, logical, and physical models for OLTP, Data Warehouse Data Vault and Data Mart Star/Snowflake schema implementations.
* Used Alteryx for Data Preparation and then Tableau for Visualization and Reporting.
* Processed data in Alteryx to create TDE for tableau reporting.
* Working experience with Kimball Methodology and Data Vault Modeling.
* Define virtual warehouse sizing for Snowflake for a different type of workloads.
* Created roles and access level privileges and took care of Snowflake Admin Activity end to end.
* Major challenges of the system were to integrate many systems and access them which are spread across South America; creating a process to involve third-party vendors and suppliers; creating authorization for various department users with different roles.
* Worked in Scaled Agile (SAFE) environment as System/DBT QA, hands-on experience with Rally, JIRA.
* Validated the data load process for Hadoop using the Hive SQL queries.
* Integrated Kafka, Spark and Cassandra for streamline analytics for creating a predictive model.
* Developed UDF's in Python to implement functions according to the specifications.
* Experience in NumPy and Pandas for data manipulation and data preprocessing.
* Involved in carrying out the important design decisions in creating UDFs, partitioning the data in hive tables at two different levels based on the related columns for efficient retrieval and processing of queries.
* Have worked on data sharing of some tables to the downstream consumers (both reader and consumer accounts)
* Worked on importing and exporting data into HDFS and Hive using Sqoop, built analytics on Hive tables using Hive Context.
* Developed the Scalable ETL pipeline using PySpark on AWS Databricks and loaded the enriched data into AWS Data Lake.
* Designed custom-built input adapters using Spark and Hive to ingest and analyze data in Airflow and ingested the enriched data to snowflake.
* Collected and processed large amounts of log data and staging data in HDFS using Kafka for further analysis.
* Experience in Metadata management like Zero copy cloning, Time travel and Data sharing.
* Involved in data migration team from on-premises SQL cluster to AWS and did transformations, standardizations and cleaning in AWS Databricks and stored in AWS SQL DB.
* Developed Spark/Scala, R and Python for regular expression (RegEx) project in the Hadoop/Hive environment with Linux/Windows for big data resources.

**Environment:** PySpark, MySQL, Snowflake, MongoDB, Azure HDInsight, Data Bricks, Azure AD, Blob Storage, Data Factory, Scala, Hadoop 2.x (HDFS, Map Reduce, Yarn), Hive, HBase.

**Deutsche Bank, Cary, NC Feb 2016 - Apr 2018**

**Data Engineer**

**Responsibilities:**

* Performance monitoring and Optimizing Indexes tasks by using Performance Monitor, SQL Profiler, Database Tuning Advisor, and Index tuning wizard.
* Acted as point of contact to resolve locking/blocking and performance issues.
* Wrote scripts and indexing strategy for migration to Confidential Redshift from SQL Server and MySQL databases
* Worked on AWS Data Pipeline to configure data loads from S3 to Redshift
* Used JSON schema to define table and column mapping from S3 data to Redshift
* Helped individual teams to set up their repositories in bit bucket and maintain their code and help them set up jobs which can make use of CI/CD environment.
* Involved in various Transformation and data cleansing activities using various Control flow and data flow tasks in SSIS packages during data migration.
* Develop and Maintain database objects viz. Tables, Views, Cursors, Constraints, User Defined Functions, Clustered and Non-Clustered Indexes, Triggers (implement business logic).
* Developed Mappings, Sessions, and Workflows to extract, validate, and transform data according to the business rules using Informatica.
* Design, develop, test and deploy Power BI scripts and perform detailed analytics.
* Capable of implementing row-level security on data along with an understanding of application security layer models in Power BI.
* Installed and Configured Enterprise gateway and Personal gateway in Power BI service.
* Implementing the use of Hyperlinks, Filters for developing Complex dashboards and created Aggregations, Totals, percentages using Key Performance Measures (KPI).
* Used Direct Query and Import data connections for the reports, scheduling the reports and maintaining extract refresh for the reports on the Power BI Service.
* Reviewed custom SQL queries and tuned for better performance in Power BI Desktop.
* Experience in connecting the MS Data verse with Power BI desktop and creating reports.
* Involved in direct interaction with Power BI Admin for requesting appropriate permissions for various Groups assigned to Projects at workspace level.
* Worked on SAS programs to import and export data from various databases.
* Hands on JIRA application, used to track & update status in JIRA for tickets assigned to me. Also created new JIRA tickets for new requirements, enhancements and bug fixes.

**Environment:** Power BI (Desktop, Report Server, Pro), Informatica, Agile, Oracle, Toad, Netezza, PL/SQL, SQL Server, T-SQL, Confluence, WOW, Power Point, Visio, VBA, MS-Access, Excel, Control M.

**Bank of America, Charlotte, NC Feb 2015 - Jan 2016**

**Tableau Developer**

**Responsibilities:**

* Primarily responsible for design, development and management of BI reporting projects for company clients in MSTR and Tableau.
* Worked closely with business to assess requirements and translate them into technical requirements. Worked with data warehouse teams to build physical data model to meet all data requirements
* Generated Quarterly, Monthly and Weekly Dashboards for Team level and Employee Level data and applied Quick filters, Parameters and sets to handle views more efficiently.
* Generated Context filters and Data source filters while handling huge volumes of data and for better performance.
* Created reports for users using Tableau Desktop by connecting to multiple data sources (MS Excel, MS Access, SQL server, Oracle).
* Designed multiple Dashboards with the help of Tableau features like Calculated Fields, Groups, Sets, Parameters and Actions etc.
* Created Database tables and logical views required for reporting needs.
* Used OLAP Apply functions for complex logic and VLDB properties to alter the SQL queries for better performance
* Designed Micro Strategy Schema objects and Application objects by creating Facts, Attributes, Hierarchies, Reports, Filters, Metrics and Templates using Micro Strategy Desktop and Architect.
* Created and established KPI’s based on BRD and logical data model.

**Environment:** Tableau (Desktop, Server), Python, Micro Strategy 9.x/8.x (Administrator, Desktop, Web, Distribution Services, Command Manager), Oracle, Windows Server, SQL server.

**Info Edge Solutions, India June 2013 – Dec 2014**

**BI Developer**

**Responsibilities:**

* Identifying the Data Sources for various KPIs and designing end-to-end data flow diagrams (dfd's) using Visio to illustrate data flow in various systems involved in a process.
* Performed detailed data analysis and data profiling using complex SQL queries on various sources systems including Oracle and Teradata.
* Involved in performing complex data preparation, data blending and creation of data models
* Blended and combined data using Alteryx from different sources such as Workday, SAP BPC, Salesforce, SAP Fiori, and Oracle Primavera P6 and published combined Data Mart (HANA Models).
* Worked directly with Business Lines in Customer Care, Loans, Management Information Systems, Card Management to gather dashboard requirements for development
* Developed complex database programs like SQL triggers and stored procedures for implementing the business rules
* Worked on designing the database, data relationships and normalizing the database
* Used Tableau and Salesforce Visual analytics for creating data dashboards

**Environment:** Alteryx, SQL Server, SSIS, Python, Tableau, Power BI

**Educational Qualification:**

* Bachelors in Electronics, India, May 2013