**Rashmith Reddy Boppidi**

**9455369045**

**rashmithboppidi17@gmail.com**

**Career Objective:**

* To obtain a challenging position in the field of Computer Engineering, utilizing my educational and previous work experience skills in designing, developing and problem solving in the Big Data domain.

**Professional Summary:**

* Overall **8+** years of IT industry experience, working in a progressive and dynamic environment, with emphasis on Data Integrity and Data Quality, Business Intelligence concepts, Database Management system, development and complete Project Life Cycle in Data Warehousing and Client/Server technologies.
* Design Data models using **Erwin**.
* Worked as **Data scientist** for **4+** years with **key programming languages** and their packages. (**R, R-shiny, Python**)
* Strong experience in creating data marts and data warehouse in **Snowflake** ( **snow pipe**, **snowpark**)
* Creating dashboards on **PowerBI** and **Tableau**
* Thorough knowledge of the **SDLC** with hands on experience in the requirements analysis and design and customer acceptance of the Business Intelligence solutions.
* Experience with solid understanding of Business Requirement Gathering, Business Process Flow, Business Process Modeling and Business Analysis.
* Extensive experience in  **Snowflake , Azure** and **GCP (Google Cloud Platform)**.
* Worked extensively with **Apache Hadoop ecosystem** like **HDFS**, **Spark**, **Flume**, **Kafka**, **Airflow** and **Hive**.
* Similarly used **GCP** services like **Cloud storage**, **Big Query**, **Cloud Composer**, **Pub Sub**, **Cloud Monitoring**, **Cloud Functions, Cloud Compute Engine Data Proc** and **Power BI.**
* Extensive experience in Data Warehouse services like **Amazon Redshift**, **Google Big Query** and **Snowflake**.
* Used orchestration tools like **Apache Airflow** and **Step Function**.
* Have knowledge in both the technical and functional aspects of the projects. Well versed in handling SQL Database and writing SQL queries to test the data.
* Experience in Data Warehouse development working with **Data Migration**, Data Conversion and **ETL** using **Microsoft SQL server** Integration services and **SQL Server**.
* Extensive experience in developing complex data extract, applications, and ad- hoc queries as requested by internal and external customers using **SQL**.
* Experience with dashboard / report design with **Tableau**, **Power BI** and **Quick Sight**.
* Hands-on experience on Unix shell scripting.
* Maintain the production Business Intelligence products and solutions, including resolving production issues and returning quickly to priority problems.
* Troubleshoot and resolve data issues impacting extract delivery.
* Created different interactive views and reporting Dashboards by combining multiple views using **Tableau Desktop**.

**Technical Skills:**

|  |  |
| --- | --- |
| Big data/Hadoop Ecosystem | Spark, Kafka, Airflow. |
| Opensource libraries | Scikit-learn, NumPy, SciPy, OpenCV, Deep learning, NLP, Keras, Matplotlib (for visualization), R-shiny |
| NoSQL Databases | MongoDB, Cassandra, HBase |
| Data Analysis Skills | Data Cleaning, Data Visualization, Feature Selection, Pandas, SQL |
| Programming Languages | Python, SQL, R, Scala, Python, PL/SQL, Linux shell scripts. |
| Database | Oracle 11g/10g, DB2, Microsoft SQL, MySQL, Teradata |
| Cloud Ecosystem | Azure, GCP, Snowflake |
| Automation and scheduling | Airflow, Google cloud Composer, Azure data factory |
| Reporting tools | Tableau, POWER BI. |
| Data Lake | Google Cloud Storage, Azure data lake storage |
| Data Warehouse | AWS Redshift, Snowflake, Google Big Query, Snowflake |

**Professional Experience:**

**Client: Molina Health Care Dec 2022 to Present**

**Location: CA (remote)**

**Role: Data engineer**

**Responsibilities:**

* + Conducted data quality assessments and implemented data cleansing and standardization procedures.
  + Created views, store procedure and tables in **BigQuery ( Data-warehouse).**
  + Created and managed Pipelines using **GCP Data Flow.**
  + Migrated cron jobs to **airflow/ composer**  in **GCP.**
  + Have written python **DAGs** in airflow which orchestrate end to end **data pipelines** for multiple applications.
  + Process and load bound and unbound data from google **pub/sub** topic to **big-query** using cloud data flow with python
  + Worked with both Maximized and Auto-scale functionality.
  + Used Temporary and Transient tables on diff datasets. A
  + Cloned Production data for code modifications and testing.
  + Shared sample data using grant access to customer for UAT.
  + Performed ad-hoc analysis and provided data-driven recommendations to improve business operations.
  + Designed and implemented data models (star schema, snowflake schema) to optimize query performance and facilitate data analysis.
  + Conducted exploratory data analysis (EDA) to uncover insights and identify trends.
  + Utilized SQL and Python to perform complex data analysis and generate meaningful insights.

**Environment**: GCP( bigquerry , Data flow), Airflow, Python.

**Client: Walmart Nov 2021 to Oct 2022**

**Location: Bentonville, AR**

**Role: Data Engineer**

**Responsibilities:**

* Created views using **big query** as per business requirements.
* Managed the deployments from sandbox to production in GCP.
* Responsible for optimizing many store procedures.
* Created and maintained pipelines in Data Factory.
* Creating and maintaining pipelines in Azure data factory.
* Creating PR from Azure github repository.
* Modifying existing JSON configs for transformation.
* Tested ETL pipelines before deploying it to production.
* Implemented both ETL and ELT architectures in Azure using Data Factory
* Ingested huge volume and variety of data from disparate source systems into Azure Data Lake Gen2 using Azure Data Factory.
* Created reusable pipelines in Data Factory to extract, transform and load data

**Environment**: Big query- GCP, Azure data factory, Azure repository, Json, ETL

**Client: Flowserve Jan 2021 to Nov 2021**

**Location: Irving, TX**

**Role: Data Engineer**

**Responsibilities:**

* Built Real-time and batch pipelines using **GCP services.**
* Part of **Inventory management team**
* Dealing t**ime-series models & depending on trend, seasonality and cyclicality ARIMA**
* Running predictive model on inventory data and review demand and decision making for requested order size and quantitative information
* Analyzing and **visualizing** using **R** and **R- shiny**
* Modeled and developed a new data warehouse from scratch and then migrated that data warehouse to **Big Query**. Automated data pipelines and quality control checks.
* Architected and implemented a solution to migrate the data platform from **Hadoop** to **Google Cloud Platform.**
* Helped to implement **PowerBI** in the organization. Developed PowerBI reports for every department in the company.
* **Kafka** is a distributed event streaming platform that is often used to handle large volumes of data in real-time.
* Developed **PySpark** scripts for ETL using **DataProc**.
* Experience in developing APIs in cloud.
* Good Knowledge on **IAM** rules and **Cloud Security.**
* Designed **PowerBI** data visualizations using cross tabs, maps, scatter plots, pie charts etc.
* Kafka is designed to be distributed from the ground up. It can be scaled horizontally across multiple servers or nodes to handle high volumes of data and traffic. This makes it suitable for large-scale data streaming applications.

**Environment:** Google Cloud Storage, Dataflow, DataProc, Cloud function, Kubernetes, Big Query, Cloud run, Cloud Composer, Airflow, Hive, Docker, Hive, HDFS, Compute Engine, etc.

**Client: IAAI Apr 2020 to Dec 2020**

**Location: Westchester, NY**

**Role: Data Scientist**

**Responsibilities:**

* Analytical data importation from SQL to HDFS and Hive.
* **Forecast Customers transaction** using time series analysis using **R**  and **Python libraries.**
* Carrying out specified data processing and statistical techniques such as **sampling techniques, estimation , hypothesis testing , times series, correlation and regression analysis** usin**g R.**
* **Applied various data mining techniques: linear regression** and **logistic regression, classification and clustering** using **R and visualized** using **R -shiny.**
* Arrived independent variables with **financial Ratios** for **modeling** and **forecasting** the data using time series analysis**-R and Python**
* Hadoop log file management and review experience.
* Imported data from a variety of sources into the **Azure data lake**.
* In the context of **Agile Methodology**, **testing, and defect reporting**.
* Migrated **SQL database to Azure Data Lake**, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and controlling and granting database access and Migrating On premise databases to Azure Data Lake store using Azure Data factory.
* To Handle **SSL Certificates, Play Frameworks Web services API was utilized**, Slick 3.0 was utilized to create the persistence layer. The Dispatcher was built on AKKA and uses actors to manage transaction reading and dispatching in the cache at the same time.
* Used **Spark RDDs, Python, and Scala** to translate Hive/SQL queries into Spark transformations.
* Using a mix of **Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics**, extract, transform, and load data from sources systems to Azure Data Storage services. Ingestion of data into one or more Azure Services (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing in Azure Databricks.
* Using copy command, **Implemented Snow pipe**, stage and file upload to the **Snowflake database**.
* Worked on **sequence files, ORC files, bucketing, and partitioning** to improve Hive speed and storage
* Worked on **Snowflake Virtual Warehouses.**
* Working knowledge and administrator experience of **Continuous Integration** Strategies and tools like **Jenkins, TeamCity.**
* Working knowledge of source code and configuration management solution like **Git, SVN**
* Writes **Hive Queries** for examining data in the **Hive warehouse** using HQL.

**Environment**: R, R-shiny, Hadoop, Hue, HDFS, Spark, AWS, MapReduce, Hive, Python, NoSQL, Cloudera, Linux, MySQL, SQL

**Client: Cooper Technologies. Feb 2017 to Nov 2019**

**Location: India**

**Role: Data Engineer**

**Responsibilities:**

* Creating a spark Jobs for putting data into **Hive** tables from multiple sources according to criteria.
* Worked on spark core, **Spark Streaming**, and **spark SQL** modules of Spark
* Created various Spark POCs and deployed them on the Yarn cluster, comparing Spark's performance.
* Used Spark RDDs, Python, and Scala to translate Hive/SQL queries into Spark transformations.
* Imported data from a variety of sources into the Azure data lake.
* Using a mix of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics, extract, transform, and load data from sources systems to Azure Data Storage services. Ingestion of data into one or more Azure Services (A**zure Data Lake, Azure Storage, Azure SQL, Azure DW**) and processing in Azure Databricks.
* Working knowledge and administrator experience of Continuous Integration Strategies and tools like Jenkins, TeamCity.
* Working knowledge of source code and configuration management solution like Git, SVN
* To Handle SSL Certificates, Play Frameworks Web services API was utilized, Slick 3.0 was utilized to create the persistence layer. The Dispatcher was built on AKKA and uses actors to manage transaction reading and dispatching in the cache at the same time.
* Designed & implemented migration strategies for traditional systems on Azure (Lift and shift/Azure Migrate, other third-party tools) worked on Azure suite: Azure SQL Database, Azure Data Lake (ADLS), Azure Data Factory (ADF) V2, Azure SQL Data Warehouse, Azure Service Bus, Azure key Vault, Azure Analysis Service (AAS), Azure Blob Storage, Azure Search, Azure App Service, Azure data Platform Services.
* Migrated SQL database to Azure Data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and controlling and granting database access and Migrating On premise databases to Azure Data Lake store using Azure Data factory.
* Worked on sequence files, ORC files, bucketing, and partitioning to improve Hive speed and storage
* Experience working on Snowflake Multi-Cluster Warehouses.
* Worked on Snowflake Virtual Warehouses.
* Using copy command, Implemented Snow pipe, stage and file upload to the Snowflake database.
* Created python and shell scripts to automate tasks on a regular basis.
* As part of the functional needs for Big Data, I developed numerous complex Map Reduce applications.
* In the context of Agile Methodology, testing, and defect reporting.

**Environment**: Spark, Hive, Python, NoSQL, Linux, MySQL, SQL

**Education:**

* MS in Computer Science - Rowan university - 2021