**Kumar Kasani**

Data Engineer

Email: [mailmekumark234@gmail.com.](mailto:mailmekumark234@gmail.com)

Phone: (480)-454-9701

**Professional summary:**

* Over 4+ years of experience as a Data Engineer, specializing in a broad array of technologies, Spark for advanced analytics, Snowflake for data warehousing, and Databricks for unified data analytics, all integrated with cloud solutions across various industries.
* Proficient in SQL, PySpark, Python for big data processing, robust application development, and data analysis.
* Demonstrated proficiency in AWS and Azure utilizing services like EC2, S3, Lambda, AWS Glue, Azure Synapse Analytics and more for scalable and cost-efficient data solutions.
* Hands-on experience on Unified Data Analytics with Databricks, Databricks Workspace User Interface, managing Databricks Notebooks, Delta Lake with Python and Spark SQL.
* Hands-on experience on Snowflake with Tasks, SnowPipe and external integrations with DBT for data ingestion pipelines.
* Experience in developing Spark applications using Spark-SQL in Databricks for data extraction, transformation, and aggregation from multiple file formats for Analyzing& transforming the data to uncover insights into customer usage patterns.
* Working experience in technology stack of PySpark, AWS Glue, AWS S3, Timescale DB, Azure Data Factory, Azure Data Lake, Python, DBT and SQL
* Crafted powerful analytical prototypes using Power BI, Grafana and Tableau.
* Expertise in transforming business requirements into analytical models, building models, developing datamining, and reporting solutions that scales across huge volume of structured and semi-structured data.
* Performed data mining in Snowflake using DBT (Open-source command line data transformation tool with SQL).
* Strong experience in ETL data warehousing and implementing all phases of SDLC which includes requirement gap analysis, design, Datawarehouse implementation, development, testing, deployment and production support maintenance.
* Good communication, interpersonal and quick learning skills with proven ability to adapt to different project environments.

**Technical Skills**

* Big Data Technologies: Python, Spark, Pyspark, SparkSQL
* Cloud Technology: Azure Data Factory, Azure Data Lake Service, AWS Glue, AWS S3, Snowflake, Databricks.
* Databases: MySQL, SQL workbench, Snowflake.
* Analysis and reporting services: MS Excel, Power BI, Grafana, Tableau.
* Languages: Python, SQL.
* Architecture: Relational DBMS, OLTP, OLAP.
* Operating Systems: Windows, Linux.
* Development methodologies: Agile/Scrum

**Working Experience**

Employer: Sinfotech

Client: United Health Care (UHG) May 2023 – Present

Role: Data Engineer

Responsibilities:

* Designed and developed highly efficient data pipelines to ingest, store, and process data from Azure Data Lake.
* Implemented integration of Azure Data Lake to Databricks workspace to read the data source files of different formats like Parquet, CSV etc.,
* Designed and developed Data Lakehouse architecture using Databricks.
* Developed ETL pipelines involving different layers in the Lakehouse like Bronze, Silver, and Gold layers.
* Created multiple tables in Delta format.
* Developed multiple notebooks involving business transformations using PySpark and Spark SQL.
* Worked on different optimization techniques like optimize, vacuum etc.
* Created and used different clusters like all-purpose clusters, Job cluster and SQL warehouses.
* Orchestrated complete end-to-end pipelines using data bricks workflows.
* Responsible to defining best practices and standards shared across the team.
* Worked on defining data governance rules and access levels on different layers and tables in Lakehouse.

Employer: Sinfotech

Client: Nikola Motors March 2022 – April 2023

Role: Data Engineer

Responsibilities:

* Engineered data pipeline framework to ingest time series data continuously generating from electric vehicles using AWS, Timescale DB, Python.
* Continuously syncing data existing as parquet from source into AWS S3 using Python using incremental extraction with metadata tables.
* Implementation of AWS Glue Data Catalog crawlers to create raw tables for respective file types.
* Extensive use of AWS Glue with PySpark to create dataframes and implement business requirement transformations on top it with SparkSQL.
* Implemented data history pipelines to maintain the latest 3 months data at any time for reporting and move the rest of the data into S3 as parquet files to reduce space and cost in Timescale DB.
* Designed timeseries tables in Timescale DB using time and space partitions for faster data retrieval performance.
* Created SQL procedures in Timescale DB for data inserts/updates from stage tables to final tables(partitioned).
* Defined AWS Glue workflow schedules to trigger the data pipelines.
* Created Grafana visualizations for timeseries data reporting.
* Followed agile principles during project execution in completion of deliverables.

Environment: AWS Glue, AWS Athena, AWS S3, PySpark, Timescale DB, Python, Grafana.

Employer: Redemptor

Client: Realtor.com Nov 2020 – Feb 2022

Role: Data Engineer

Responsibilities:

* Created dynamic Stored procedures and Views in Snowflake and Scheduled it for daily run for implementing data ingestion and performing Transformations.
* Develop and maintain Transformation and Load processes to copy data from stage files to tables using the COPY command in Snowflake.
* Experience with Snowflake cloud data warehouse for integrating data which includes loading nested JSON formatted data into Snowflake table.
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like SharePoint, API, and Blob storage.
* Collected and analyzed data on established and prospective customers, competitors, and marketing channels and sources.
* Collate the data and create transformations using DBT (data build tool) that serve business requirements and store it in Snowflake data warehouse.
* Work on solutions of data engineering for changing business requirements on Azure Cloud platform using Azure Data Factory.
* Configure the orchestration of datasets using Azure data factory pipelines via triggers to run the data updates.
* Participate in AGILE meetings to understand requirements on the assigned tasks/stories.
* Develop Databricks applications working with Spark, that provides faster analysis results.
* Responded to all incoming questions and inquiries related to JIRA applications.
* Prepared projects, dashboards, reports, and questions for all JIRA related services.
* Used SQL to perform ETL from Oracle and SQL Server databases and created basic Power BI dashboards.

Environment: Azure Data Factory, Azure Data Lake Service, Snowflake, Power BI, Python

Client: Intel Corp Sep 2019 – Oct 2020

Role: Cloud Data Engineer

Responsibilities:

* Worked on migration of developed Big Data applications and respective data into Azure cloud ecosystem.
* Implemented data processing pipelines in Azure data factory to complete end-to-end ingestion of data.
* Usage of Python scripting embedded in Azure data factory to extract data from different sources into Azure data lake.
* Implemented logging using SQL to maintain audit and its history of data pipeline runs.
* Created Linked Services to connect to various data sources like SQL, SAP, Blob and SharePoint.
* Setting up integration run time in Azure to facilitate the data transfer from different sources to Azure storage container.
* Implemented transformation layer using DBT to create the KPIs and load into Snowflake.
* Imported data from various sources into Spark RDDs and data frames for data profiling and processing.
* Reduced the latency of spark jobs by tweaking the spark configurations and following other performance and Optimization techniques.
* Optimization of Hive queries using best practices and right parameters and using technologies like Python, and PySpark.
* Developed custom aggregate functions using Spark SQL and performed interactive querying.
* Developed applications for the entire batch processing by using Pyspark with Shell scripting in Hadoop.
* Wrote spark SQL and spark scripts(PySpark) in the Databricks environment to validate the periodical inbound data.
* Visualized the results using Tableau dashboards and the Python Seaborn libraries were used for Data interpretation in deployment.

Environment: Azure Data Factory, Azure Data Lake Service, Snowflake, Tableau, Python, DBT.

**Education:**

* West Virginia University, Morgantown WV 2019