Digamber Kalivemula
DATA ENGINEER

AL | 659-910-9622 | kdigamber1729@gmail.com | LinkedIn

SUMMARY

- Data Engineer with 4 years of experience in Data Extraction, Data Modelling, Statistical Modeling, Data Mining, and Data Visualization.
- Working experience with the entire Software Development Life Cycle (SDLC) process and good experience with Agile and Waterfall Methodologies.
- Extensive experience in SQL and Python, including NumPy, SciPy, Scikit-learn, Pandas, Matplotlib, and TensorFlow.
- Proficiency in the development and design of various scalable systems using Hadoop technologies in various environments.
- Ability to analyze data using Hadoop Ecosystems, including HDFS, MapReduce, Hive, and Pig.
- Working knowledge of Amazon's Elastic Cloud Compute (EC2) infrastructure for computational tasks and Simple Storage Service (S3) as a storage mechanism.
- Proven experience designing, developing, and deploying SSIS packages in an MS SQL server environment using SSIS.
- Well-versed in analyzing data on various IDEs like Jupyter Notebook and Visual Studio Code.
- Capable of working with Tableau reports with complex calculations and working on Ad-hoc reporting using Power BI.
- Extensive experience in developing and deploying data pipelines in the cloud, such as AWS and GCP.
- Expertise in developing and publishing reports of various types using Power BI, SSRS, and Business Objects.
- Good Working knowledge of different database systems such as MS SQL Server, MySQL, and PostgreSQL.

EDUCATION

Masters of Science in Data Science   **Aug 2022 – Dec 2023**

University of Alabama at Birmingham   **GPA:3.9**

WORK EXPERIENCE

Data Engineer   **Jun 2023 - Current**
Cisco, AL

- Collaborated with stakeholders to design and develop a unified architecture for Retail data, resulting in the creation of a centralized Data Lake on Azure Cloud.
- Engineered Spark applications using Azure Databricks for efficient data extraction, transformation, and aggregation from diverse systems, significantly improving data processing capabilities and storage on Azure Data Lake.
- Designed and implemented PySpark applications to integrate and process data from various sources including FTP and CSV, utilizing Azure Databricks for streamlined data flows into Snowflake.
- Developed functions in Spark with Scala to unzip, decode, and parse XML files, ensuring accurate data storage into Azure Blob Storage.
- Authored PySpark scripts to manage data ingestion from Azure Event Hub into Delta tables in Databricks, supporting various data management strategies including reload, append, and merge.
- Led the design and development of ETL processes in DataBricks to efficiently migrate Campaign data from Azure DataLake into optimized formats (ORC, Parquet, Text Files).
- Optimized PySpark applications on Databricks, achieving significant cost savings through enhanced performance and reduced resource utilization.
- Constructed pipelines in Azure Data Factory for seamless data transfer of parquet files from ADLS Gen2 to Azure Synapse Analytics, facilitating advanced data warehousing solutions.
- Replaced legacy Hive scripts with Spark DataFrame transformations, accelerating data analysis and enhancing processing efficiency.
- Spearheaded the development of PySpark scripts that reduced organizational costs by 30% through the strategic migration of customer data from SQL Server to Hadoop.
- Managed complex JSON datasets and crafted custom Python functions for effective data parsing using Spark, enhancing data usability and precision.
- Utilized Spark for processing streaming data, conducting interactive queries, and integrating with NoSQL databases, accommodating large-scale data handling needs.
- Oversaw the loading of data pipelines from web servers leveraging Kafka and Spark Streaming API, ensuring robust data flow and timely data availability.
- Implemented Airflow for orchestration and scheduling of data ingestion scripts, optimizing workflow efficiency and eliability.
- Generated critical operational and customer analytics reports using Power BI, providing insights into mobile payments and customer engagement metrics, supporting strategic decision-making.

Data Engineer   **July 2020 - Jun 2022**
314e India Corporation, India

- Collaborated with a team of four to develop an ETL project, successfully migrating data from RDBMS to AWS cloud storage. Led efforts in database architecture, data ingestion, and transformation processes.
- Optimized data storage and retrieval in the data warehouse along with ETL pipeline performance through the implementation of data compression techniques, query optimization, and in-memory processing, reducing cloud resource costs by 30%.
- Enhanced data retrieval efficiency in Databricks by optimizing queries and tuning, which improved resource utilization and processing speed by 68%.
- Implemented various data compression methods and in-memory processing techniques using PyArrow. Worked with multiple file formats, including Parquet and Avro, to develop faster data access and manipulation modules, achieving a 50% reduction in storage costs and improving data access times by 45%.
- Designed complex SQL queries and Python scripts using NumPy to analyze batch and real-time data, creating comprehensive dashboards in Power BI for data visualization. Additionally, performed advanced SQL query performance tuning to enhance the efficiency of real-time application access.

ETL Developer   **July 2019 - Jun 2020**
314e India Corporation, India

- Collaborated with a 4 member team to migrate data from relational database systems to Amazon Web Services, overseeing database architecture, data flow management, and transformation processes.
- Developed a Java-based ETL process for ingesting and converting healthcare data formats such as FHIR, SQL tables, and CCDA into the HL7V2 standard.
- Assisted in the deployment and API development for the IBM FHIR server within a Docker environment, enhancing functionality and supporting large-scale operations.
- Implemented Python-based systems to monitor data access and developed a network-based intrusion detection system to identify and report security breaches.
- Managed RESTful API implementation and testing using FastAPI and Django, led API development in a Databricks environment, and automated system updates to reduce manual workload.

PROJECT

Find My Castle

- Written a web application from scratch to provide nearby available vacant homes using users' location.
- Utilized Python, Flask, Lambda, S3, and SQL databases to create feature-rich APIs, and collaborated cross-functionally to ensure integrated front-end and back-end components.

Emailing - Bot

- Implemented a bot that will take care of sending emails to the users from a database. This bot will have an interface where you can upload a document and choose from the available customer list. This bot will not only send the mail but also track the user document viewing activity and will remove the user access once it's been accessed.

TECHNICAL SKILLS

Programming Languages: Python, Scala, Java, R, Shell Scripting.
Big Data Ecosystem: HDFS, Pig, Hive, Oozie, Sqoop, Kafka, Spark Streaming.
ETL Tools: Informatica, Talend
Cloud Ecosystem: Azure (Azure Data Factory, Azure Data Bricks, Azure Synapse, ADLS Gen2), AWS (EC2, EMR, Lambda, Kinesis, Glue, Redshift and S3), Google Cloud.
Databases: MySQL Server, Oracle DB, HiveQL, Spark SQL, HBase, Redshift, Snowflake.
Orchestration/Tools: Airflow, Oozie, Jira, Git, Matilion, Postman, Power BI, Docker, Jenkins.
Streaming: Spark Streaming, Kafka.
IDE: Pycharm, Anaconda, IntelliJ, Jupyter Notebook, Visual Studio Code.

CERTIFICATIONS

- Amazon Web Service Solution Architect – Associate | AWS
- Cyber Security