PRUDHVI GUDIPATI
DATA ENGINEER
469-614-0323 | prudhvi.g@mycvtalent.com | Plano, Texas

SUMMARY

- 4+ years of data engineering experience with a focus on data ingestion, data validation (structured & unstructured), data modeling, and creating insightful reports and visualizations using various tools.
- Experience in creating interactive dashboards and reports using Data Studio to visualize key business metrics from BigQuery data, enabling data-driven decision-making for stakeholders.
- Expertise in SQL, database design, and ETL pipeline implementation, experience in Azure (Azure DevOps, Azure Data Lake, Azure Data Factory, Azure Databricks), showcasing expertise in leveraging diverse cloud services.
- Engineered enterprise solutions by employing batch processing with data bricks and integrating streaming frameworks, including Spark Streaming, Apache Kafka, and Apache Airflow.

EDUCATION

Master of Science in Computer Science **Dec 2023**
University of Texas at Arlington, TX

Bachelor's in Computer Science **Mar 2020**
Sathyabama Institute of Science and Technology, Tamil Nadu, India

SKILLS

- **Methodologies & Language:** Agile/Scrum, Scala, Python, R, Java, SQL
- **IDE's:** PyCharm, Jupyter Notebook
- **Big Data Ecosystem:** Hadoop, Hive, HDFS, Sqoop, Apache Airflow, Apache Kafka, Apache Spark, Apache Flink
- **ETL and Cloud Technologies:** SSIS, Azure (Azure Data Lake, Azure SQL Datawarehouse, Azure Databricks), GCP (BigQuery, Data Flow, Pub/Sub, Data Studio, Cloud Storage)
- **Visualizations:** Tableau, Power BI, Excel
- **Packages & Data Processing:** NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, PySpark, Data Pipelines, Jenkins
- **Version Control & Database:** GitHub, Gitlab, SQL Server, PostgreSQL, MongoDB, DynamoDB, MySQL, Snowflake
- **Other Skills:** REST API, Node.JS, CRM
- **Operating Systems:** Windows, MacOS

WORK EXPERIENCE

Principal Financial, TX **May 2023 – Current**
Data Engineer

- Automated efficient data pipelines that parsed and stored raw data into partitioned Hive tables, improving data retrieval for reporting and analysis by 20%.
- Built and orchestrated complex data pipelines using Airflow to automate data ingestion, transformation, and loading tasks between various data sources (Kafka, HDFS) and data warehouses, ensuring reliable and scheduled data delivery for business intelligence dashboards.
- Increased data retrieval efficiency by approximately 40% by implementing partitioning and clustering strategies for high-volume tables in BigQuery.
- Create a scalable messaging system using Pub/Sub to ingest real-time data from various sources into Dataflow pipelines.
- Establish ETL workflows using Apache Spark and Python (Pandas, NumPy, BeautifulSoup), resulting in a 30% reduction in data processing time and improved data accuracy.
- Execute data pipelines using Scala and PySpark (including libraries like NumPy and Pandas) to efficiently ingest, clean, and transform large-scale datasets (structured, semi-structured, and unstructured) from various sources.
- Enhanced vision system algorithms, resulting in a 30% improvement in object detection accuracy and faster processing times.
- Implemented a real-time status monitoring dashboard, increasing team awareness and response efficiency by 40%.
- Conducted a self-assessment initiative that identified skill gaps, leading to a 25% increase in targeted training and development.

Exert Infotech, India **Aug 2018 – Dec 2021**
Data Engineer

- Used Spark to accelerate data processing, achieving a 25% increase in processing speed for daily batches of up to 50GB, enhancing the team's data analysis capability.
- Implemented a real-time data streaming pipeline using Kafka to capture and process high-volume website traffic data, enabling real-time fraud detection and personalization of customer experiences.
- Designed and implemented an end-to-end data pipeline using Azure Data Factory (ADF) to orchestrate data movement from various sources (databases, APIs) to ADL.
- Developed and maintained HiveQL queries to analyze user clickstream data stored in HDFS, enabling product managers to identify user behavior patterns and optimize app features.
- Employed Snowflake materialized views, data masking, and optimization to deliver efficient and secure data solutions for seamless data management and advanced big data analytics.
- Orchestrated data pipelines for automated data movement and transformation between various data sources and Azure services, including ADLS, ASDW, and Databricks.
- Worked on CI/CD solutions using Git and Jenkins to set up and configure the big data architecture on the Azure cloud platform.

- Coordinated a streamlined release process, reducing deployment errors by 20% and accelerating time-to-market by 15%.
- Ensured compliance with industry regulations, achieving a 98% compliance rate and reducing audit issues by 50%.
- Developed a classification model that improved data categorization accuracy by 35%, enhancing decision-making processes.

CERTIFICATION

- **Azure Data Scientist**
- **Advance Google Analytics**