Abhinandan Gayaki

+91-9503091887 | sapireddyvamsi@gmail.com | github.com/Abhi707-bot | linkedin.com/in/abhinandan-gayaki-2a6380170

Summary

Results-driven Data Engineer with expertise in Python, SQL, and cloud services (Azure & AWS), specializing in real-time data pipelines, ETL workflows, and big data processing. Skilled in designing and optimizing scalable data architectures using Azure Event Hubs, AKS, Databricks, and Cosmos DB, integrating structured (CRM) and unstructured (IoT, web logs) data sources. Proficient in Apache Spark for data transformation, Azure Data Lake Storage for cost-effective data management, and SQL performance tuning to optimize data workflows. Experienced in building cloud-based ETL pipelines with AWS (S3, Glue, Athena) and automating data ingestion for improved efficiency and scalability. Passionate about delivering high-performance, cost-efficient, and reliable cloud data solutions to support business growth and analytics-driven decision-making.

Skills

Programming	Python, SQL, PL/SQL, PySpark, Git, Scripting (Bash).
Platforms	AWS (Glue, S3, Redshift), Azure (ADF, AKS, Cosmos DB, Azure Monitor)
Software	Linux, Docker, Kubernetes, Databricks, Kafka, Airflow

Experience

Cognizant Technology Solutions, *Data Engineer* | Pune, India   *Aug 2022 - Present*

- Designed and optimized a real-time data pipeline using **Azure Event Hubs** and **Kafka**, improving data processing efficiency by 25%. Implemented **Azure Kubernetes Service (AKS)** for batch job orchestration, reducing system downtime by 40%.
- Integrated structured data from **CRM systems** (e.g., Salesforce) and unstructured data from SFTP servers (e.g., CSV, JSON files) into Azure Event Hubs for real-time processing, enabling seamless data flow and real-time analytics.
- Leveraged **Azure Databricks** (Apache Spark) for large-scale data transformation, enrichment, and aggregation, enhancing data quality and boosting business insights by 20%.
- Designed a scalable storage architecture using **Azure Cosmos DB** for low-latency transactional workloads and **Azure Data Lake Storage** for structured/unstructured data analytics.
- Implemented **Azure App Insights** for real-time monitoring, anomaly detection, and performance tuning, improving system reliability and reducing failures.

Projects

ETL Data Pipeline Using Apache Airflow   *Nov 2024 - Dec 2024*
ETL Data Pipeline, Personal Project

- Developed an end-to-end ETL pipeline to scrape Amazon book data using a **Dockerized Python script**, process it, and load it into a **PostgreSQL database**, orchestrated with Apache Airflow.
- Utilized Docker for isolated and scalable web scraping, extracting book details such as title, author, price, and link from Amazon's website.
- Automated the entire ETL process with **Apache Airflow**, leveraging DockerOperator, PythonOperator, and PostgresOperator for efficient execution and monitoring.

ETL Pipeline using AWS services   *Sept 2024 - Oct 2024*
AWS ETL Pipeline, Personal Project

- Designed and implemented an ETL pipeline using **AWS S3, Glue, and Athena** to transform and load data into S3 in Parquet format.
- Configured **AWS Glue Crawler** for automatic schema discovery and population of the Glue Data Catalog, streamlining metadata management.
- Created IAM roles with admin access to securely manage **AWS Glue** resources and permissions within the pipeline.
- Set up triggers for automated job scheduling, allowing daily execution and flexible activation/deactivation based on project requirements.

Certificates

2025	**Databricks Certified**,	Udemy
2024	**AZ-900: Microsoft Azure Fundamentals**,	Microsoft
2024	**AZ-204: Developing Solutions for Microsoft Azure**,	Udemy

Education

9/10	**BTech in Chemical Engineering,** *Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded* | Maharashtra, India	*Jul 2018 - Jun 2022*