

Deekshith Alampally

 +1(515)-619-2664 ◇  deekshithalampally@gmail.com ◇ California, United States ◇  LinkedIn ◇ GitHub

SUMMARY

Data Engineer with a year of experience designing and maintaining automated ETL/ELT pipelines, SQL-based data warehouses, and analytics-ready data models. Strong in Python, SQL, Airflow, and Docker, with hands-on experience building reliable, scalable data platforms and production-grade pipelines supporting business analytics and reporting.

SKILLS

Programming Languages : Python, SQL, Java, NoSQL, JavaScript, C

Data Engineering & Cloud Platforms AWS (S3, Glue, SageMaker, Lambda, SNS, SQS, EC2, EMR, Redshift), Apache Spark (PySpark), Hadoop, Apache Airflow, Data Warehousing (Star Schema), ETL Pipelines, Data Validation

Data Science & Machine Learning : LangChain, OPEN AI API, CNN, RNN, Supervised & Unsupervised Learning, Retrieval-Augmented Generation (RAG)

Databases SQL Server, MySQL, Oracle, Firebase, supabase, MinIO (S3-compatible), MongoDB

Tools & Collaboration Power BI, Tableau, Excel, Elasticsearch, Jupyter Notebook, Google Colab, Matplotlib, Seaborn, Pandas, NumPy, PyCharm, Visual Studio, Git, GitHub, Jira, Idempotent Loads, GitHub Copilot.

WORK EXPERIENCE

Data Engineer

Cognizant

Aug '22 — Aug '23

Hyderabad, India

- Designed and maintained end-to-end ETL pipelines using Python and SQL to ingest, transform, and load structured business data into centralized SQL databases, enabling reliable downstream data consumption and analytics.
- Implemented robust data quality and validation frameworks using SQL and Python, including schema checks, null thresholds, and reconciliation logic, significantly reducing downstream data inconsistencies and rework.
- Built reusable SQL transformation layers and reporting schemas, improving data accuracy by approximately **25%** and ensuring consistent metric definitions across teams.
- Automated recurring data workflows, reducing manual reporting effort and improving delivery reliability for business stakeholders.

PROJECTS

Production-Grade ETL Pipeline [Link](#)

- Designed and deployed a containerized data platform on Docker, integrating PostgreSQL for warehousing and MinIO (S3) for object storage to simulate a modern cloud-native architecture.
- Implemented a dimensional data model (Star Schema) in Postgres, transforming raw unstructured JSONB logs into optimized Fact and Dimension tables to support high-performance analytical queries.
- Developed idempotent data loading scripts ensuring zero data duplication during historical backfills and recovery from failures, enhancing overall pipeline resilience and data quality.
- Engineered a scalable ELT pipeline using Apache Airflow and Python to ingest data from public APIs, handling pagination and rate limits to ensure reliable daily delivery of 100% of source records.

You Tube Analytics Data Warehouse

- Extracted channel performance metrics using the YouTube API and automated data ingestion with Python.
- Designed a structured SQL data warehouse for views, impressions, watch time, CTR, and subscribers.
- Optimized transformation pipelines to create analytics-ready tables that support trend and content performance analysis.
- Built a Tableau-based creator growth dashboard, improving decision accuracy up to 70%.

CERTIFICATIONS

Python for Data Science and Machine Learning, Udemy

Cloud Computing Fundamentals, Google Cloud Skills Boost

Data, ML, and AI in Google Cloud, Google Cloud Skills Boost

Programming for Everybody Getting Started with Python, Coursera

EDUCATION

Master's degree in Computer Science, California State University- San Bernardino (GPA: 3.77)

Aug '23 — Dec '25

San Bernardino, USA

Bachelor of Technology in Computer science and engineering, Sri Indu College Of Engineering & Technology (GPA: 8.60 CGPA)

Jul '18 — Jul '22

India, India