DEEKSHITH ALAMPALLY

California, USA | 5156192664 | deekshithalampally@gmail.com
LinkedIn | GitHub | Portfolio

PROFESSIONAL SUMMARY

Data Engineer with 2 years of experience building batch ETL pipelines in retail analytics environments. Experienced in AWS-based data workflows (S3, Redshift), Apache Airflow orchestration, and PySpark transformations. Strong foundation in dimensional data modeling, SQL performance optimization, and data quality validation to support scalable reporting systems.

TECHNICAL SKILLS

Programming: Python, SQL
Data Engineering: ETL/ELT, Incremental Loads, Data Validation, Star Schema Modeling, Batch Processing
Cloud: AWS (S3, Redshift, EC2, Lambda)
Orchestration: Apache Airflow
Big Data: PySpark
Databases: PostgreSQL, MySQL, SQL Server
Visualization: Tableau, Power BI
Tools: Docker, Git, Jira

PROFESSIONAL EXPERIENCE

Cognizant Technology Solutions – Data Engineer | Dec 2022 – Aug 2023
- Contributed to batch ETL pipelines processing 200K–500K daily retail transaction records into AWS-based analytics systems.
- Ingested raw order, inventory, and product datasets into Amazon S3 and supported transformation and loading into Amazon Redshift.
- Designed and maintained star schema models including fact_sales and dimension tables.
- Developed SQL transformation logic for order volume trends, category performance, and inventory analysis.
- Built Python ingestion scripts with schema validation and retry-safe processing.
- Contributed to Apache Airflow DAG scheduling, dependency management, and failure monitoring.
- Used PySpark for large historical sales dataset processing and aggregation.
- Implemented data quality checks and performed root cause analysis on pipeline failures.

Kapi Technologies – Junior Data Engineer | Aug 2021 – Dec 2022
- Wrote advanced SQL queries (joins, CTEs, window functions) for retail sales and inventory analysis.
- Automated KPI reports using Python (pandas), reducing manual effort by ~30%.
- Designed aggregation tables supporting dashboards tracking revenue trends and inventory turnover.
- Improved query performance through indexing and restructuring.
- Built Power BI dashboards for regional and store-level reporting.

PROJECTS

Production-Grade ELT Pipeline
- Built Airflow-orchestrated ELT pipeline ingesting API data into object storage and PostgreSQL.
- Designed dimensional data models (fact and dimension tables).
- Implemented idempotent loading logic and failure recovery handling.
- Containerized using Docker to simulate production-style architecture.

YouTube Analytics Data Warehouse
- Automated extraction of YouTube metrics using Python and REST APIs.
- Designed SQL warehouse schema tracking views, CTR, watch time, and subscriber growth.
- Developed Tableau dashboards for reporting.

EDUCATION

Master of Science in Computer Science – California State University, San Bernardino | GPA: 3.77
Bachelor of Technology in Computer Science & Engineering