

Summary:

Results-oriented **Cloud Data Engineer** with **7 years of experience** designing and implementing data solutions using cutting-edge technologies (**Python, SQL, Airflow**) on cloud platforms like (**GCP**) and **on-premises** systems. Proven expertise in **automation** and **optimisation** of **scalable data pipelines** performing **ETL/ELT** processes. Skilled in collaborating with cross-functional teams to deliver high-impact business solutions.

Certifications:

- * **Certified Cloud Architect**: Validating expertise in designing and implementing scalable, secure and resilient cloud architectures.
- * **Certified Cloud Data Engineer**: Demonstrating proficiency in developing and optimizing cloud-based data engineering solutions.
- * **Monitoring in Docker**: Focused on container monitoring using **Grafana** and **Prometheus** including real-time performance alerting.
- * **Python for Data Science**: Focused on data analysis, visualization, and predictive modeling using Python libraries (Pandas, NumPy).
- * **AI/ML Statistical Analysis**: Trained on key statistical methods used in AI/ML models.
- * **Machine Learning with Python**: Implemented ML algorithms in Python, focusing on supervised, unsupervised and reinforcement learning.
- * **Kubernetes**: Learned Kubernetes architecture, deployments, and cluster management.
- * **MySQL**: Hands-on experience with MySQL database design, queries, and optimization.

Work Experiences:

Deutsche Bank

2024-Present

Cloud Data Engineer

- * Optimized composer (file process of airflow), **10x faster file transfer**/downloads with higher reliability under variable networks.
- * Optimized **API authentication** by caching access tokens with expiry, implementing single-flight refresh and interceptor-based retries, eliminating redundant refreshes and stabilizing concurrency while **reducing auth calls by >90%**.
- * **Faster, cheaper** BigQuery-to-file exports by **exploiting default temp tables** and minimizing recompute.
- * Optimized webhook pipeline, p95 reduced from **~90 seconds to ~15 seconds**, p50 under **10 seconds**, with resilient async design.
- * Architected **scalable** and **resilient data pipelines** leveraging GCP services.
- * Created physical and logical data models using SAP PowerDesigner.
- * Managed **real-time** and batch ETL data pipelines, ensuring efficient data handling and real-time insights.
- * Orchestrated complex workflows with **airflow** & optimized end-to-end data processing.
- * **Optimized SQL queries** and **Python scripts** for efficient operations e.g., data retrieval and ingestion with large data in **bigquery** and **GCS**.
- * Spearheaded the implementation of **terraform** and **github actions** for robust management of **airflow DAG, GCS, GSM and Cloud functions**.
- * Conducted multiple POCs for ETL/ELT processes across on-premise and GCP environments.
- * **Migrated legacy** data workflows from on-prem to GCP, reducing infrastructure maintenance overhead and achieving up to **40% cost savings**.

Infosys

2022-2024

Data Engineer

- * Delivered end-to-end batch data processing pipelines on **GCP** utilizing mage, python & pandas for telecommunication clients.
- * Engineered ETL/ELT processes using **dataflow** and **cloud functions** to ensure high data quality and actionable insights.
- * Designed and optimized schemas for efficient storage and retrieval.
- * Implemented **falco** for monitoring and developed VMs via **vagrant** integrated with cloud monitoring & logging for comprehensive monitoring.
- * Devised star schema data models using **lucidchart** for streamlined data processing.
- * Enhanced data pipeline efficiency and reliability using **pubsub** and GCP native services like **GCS & composer**.

Tata Consultancy Services

2019-2022

Software Developer

- * Developed scalable data processing pipelines using **dataflow** and apache beam, ensuring seamless integration with data warehouse.
- * Built **OLAP** systems leveraging bigquery and optimized data models for analytical workloads for **healthcare clients**.
- * Automated daily device registration and reporting **reducing manual intervention by 60%** and significantly improving processing speed.
- * **Mentored a team of 5-6 engineers** offering technical guidance on python programming.
- * Designed cloud storage-based hierarchical storage architecture with lifecycle policies, enabling efficient data retrieval.
- * Developed Python-based backend components handling **API calls** and interaction with csv files.
- * Additionally worked on a POC with Azure Data Factory and **databricks** on interoperable solutions.

Skills:

- * Cloud Platforms: **Google Cloud Platform / GCP, Amazon Web Services / AWS, Microsoft Azure**
- * Google Cloud Platform: **BigQuery, Composer, Cloud Storage, Dataflow, Pub/Sub, Compute engine, Cloud functions**
- * Data Engineering & Warehousing: **ETL/ELT processes, schema design, dimensional modeling, OLTP/OLAP systems**
- * Programming Languages & Libraries: **Python, PySpark, Pandas, NumPy, SQL, Paramiko, Java**
- * Data Orchestration & Integration: **Apache Airflow, Mage, Kafka, SSH & SFTP Connections**
- * Database Technologies & Data Modeling: **SAP Power Designer, MySQL, MongoDB, Google BigQuery**
- * Automation & DevOps: **Terraform, GitHub Actions, Vagrant, Falco**
- * Containerization & Orchestration: **Docker, Kubernetes, Linux**
- * Version Control & Agile Practices: **Git, Bitbucket, Jira, Scrum, Agile Methodologies, REST APIs, Microservices architecture**
- * Soft Skills: **Strong analytical & problem-solving abilities**, excellent communication skills and **readiness to learn new technologies**

Education:

Bachelor of Technology in Computer Science - (B. Tech CS), Honors | KURUKSHETRA UNIVERSITY

2015-2019