

Apoorv Tyagi

myselfapoorvtyagi@gmail.com | +91-7983200401

Summary:

Results-oriented **Cloud Data Engineer** with **six years of experience** designing and implementing data solutions using **cutting-edge technologies** on cloud platforms (**GCP, AWS, Azure**) and **on-premises systems**. Proven expertise in **automation, scalable data pipelines, data warehousing, data modeling, and ETL/ELT processes**, delivering **actionable insights** from raw data. 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 and Docker environment optimization.
- * **Python for Data Science**: Focused on data analysis, visualization, and predictive modeling using Python libraries (Pandas, NumPy).
- * **SQL for Data Science**: Mastered complex SQL queries for data retrieval, aggregation, and optimization.
- * **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

- * Architected **scalable and resilient data pipelines** leveraging GCP services.
- * Managed **real-time and batch** data pipelines, ensuring efficient data handling and real-time insights.
- * Orchestrated complex workflows with **airflow & optimized** end-to-end data processing.
- * Automated ETL processes using custom **python scripts** and libraries (**pandas, paramiko, PGP, requests**) reducing manual effort.
- * Optimized **SQL queries** for efficient data retrieval from **large datasets** in **bigquery**.
- * Spearheaded the implementation of **terraform** and **github actions** for robust **composer** DAG management.
- * Conducted multiple **POCs** for **ETL/ELT** processes across **on-premise** and **GCP** environments.

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.
- * Collaborated with stakeholders to deliver tailored data solutions that meet business requirements.
- * Implemented **falco** for monitoring and developed **vagrant** VMs integrated with **cloud monitoring & logging** for comprehensive monitoring.
- * Devised **star schema** data models using **lucidchart** for streamlined data processing.
- * Designed optimized data storage solutions using **google bigquery**, enhancing data retrieval performance and operational efficiency.
- * 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 **azure data factory** and **azure databricks**, ensuring seamless integration across Azure services.
- * Built **OLAP batch processing systems** with **AWS redshift**, improving data analytics efficiency for healthcare clients.
- * Automated **daily device registration reporting** using **AWS lambda**, eliminating manual effort and reducing processing time.
- * Led a team of **5-6 engineers**, providing mentorship and technical guidance on **cloud-based data engineering solutions**.
- * Designed **AWS s3-based storage solutions** for efficient data retrieval and cost optimization.
- * Developed **python-based backend components** and integrated them with **AWS glue** for scalable ETL processing.
- * Improved **CI/CD pipelines** by developing **automated scripts**, enhancing deployment efficiency.

Skills:

- * Cloud Platforms: **Google Cloud Platform, Amazon Web Services / AWS, 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: **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**
- * Frameworks & APIs: **Django, REST APIs, Microservices architecture**
- * Soft Skills: **Strong analytical & problem-solving abilities**, excellent communication skills and **readiness to learn new technologies**
- * Tools & Technologies: MS Office (Word, Excel, PowerPoint), Outlook, **Lucidchart**

Education:

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

2015-2019