Venkata Pavithra Chaganti E-mail: [pavithrachaganti02@gmail.com](mailto:pavithrachaganti02@gmail.com)

**Phone: 856.244.4197|Irving, Tx, 75063** [**http://www.linkedin.com/in/cvpavithra**](http://www.linkedin.com/in/cvpavithra)

**Summary**

Results-driven **Data Engineer** with **2+ years of experience** specializing in **Google Cloud Platform (GCP), BigQuery, Cloud Dataflow, Dataproc (PySpark), and Apache Airflow**. Expertise in **designing, optimizing, and scaling end-to-end data pipelines** for **real-time and batch processing**. Strong proficiency in **ETL, SQL, Python, Terraform, Kubernetes, and ML Ops**, with hands-on experience **automating workflows, optimizing cloud infrastructure, and reducing operational costs**. Passionate about **building efficient, high-performance, and secure cloud-native data engineering solutions**.

**KEY SKILLS & TECHNOLOGIES**

✅ **Programming & Query Languages:** Python (Pandas, PySpark), SQL (BigQuery, PostgreSQL), Scala (Basic), Java (Basic)  
✅ **Cloud Platforms & Tools:** Google Cloud (BigQuery, Cloud Dataflow, Cloud Dataproc, Cloud Composer, Cloud Functions, Vertex AI, Cloud Storage)  
✅ **Data Engineering & Big Data:** ETL Pipelines, Data Warehousing, Data Modeling, Batch & Streaming Processing  
✅ **DevOps & Infrastructure:** Terraform (IaC), Kubernetes (GKE), CI/CD Pipelines, Git, Jenkins  
✅ **Data Visualization & Analytics:** Looker Studio, Power BI, Tableau  
✅ **Streaming Technologies:** Apache Beam, Pub/Sub, Kafka (Basic)

# Professional Work Experience

### **Graduate Teaching Assistant | Cloud & Data Engineering | University of Cincinnati |** **Cincinnati, Ohio | 08/2023 – 12/ 2024**

* **Taught Advanced Algorithms, Distributed Computing, and Cloud Data Engineering concepts** focusing on **BigQuery, Dataflow, and Pub/Sub**.
* **Guided students** in **cloud-based data pipeline design** using **Google Cloud Functions, Kubernetes, and Terraform**.
* **Led real-time data processing projects**, optimizing **BigQuery storage & compute efficiency by 30%**.
* **Implemented PySpark-based transformations on Cloud Dataproc**, reducing **processing time by 40%**.
* **Designed and optimized auto-scaling Dataproc clusters**, improving **resource utilization and reducing costs by 30%**.

**Tech Stack:** GCP (BigQuery, Cloud Dataflow, Pub/Sub, Cloud Dataproc, Terraform, Kubernetes), Python (PySpark), SQL

### **Associate Data Engineer | Tata Consultancy Services |Hyderabad, Telangana, India | 07/2022 – 07/2023**

* **Developed and optimized ETL pipelines** in **Google Cloud** for structured & semi-structured data processing.
* **Built streaming data pipelines using** **Pub/Sub → Dataflow → BigQuery**, reducing latency by **40%**.
* **Designed partitioned & clustered BigQuery tables**, improving **query performance by 35% while lowering costs**.
* **Implemented PySpark jobs on Cloud Dataproc**, optimizing **Spark execution time by 50% and memory usage**.
* **Configured high-availability Dataproc clusters**, resulting in **25% cost savings & 40% faster data transformations**.
* **Automated infrastructure provisioning** with **Terraform & Cloud Composer (Apache Airflow)**.

**Tech Stack:** GCP (BigQuery, Dataflow, Cloud Dataproc, Cloud Composer), Terraform, Apache Airflow, PySpark, SQL, Python

### **Data Engineer Intern | Cloud & ML Pipelines| CIOS | Hyderabad, Telangana, India | 07/2021 – 06/2022**

* Developed **real-time streaming solutions** using **scala** with **Kafka**, Spark Streaming, and BigQuery.
* **Developed ETL workflows from SAP BW to BigQuery**, optimizing **real-time analytics**.
* **Built Apache Beam pipelines in Cloud Dataflow**, processing **terabytes of data daily**.
* **Optimized BigQuery storage costs** using **clustering & partitioning**, reducing **monthly expenses by 30%**.
* **Integrated ML workflows using Vertex AI & Dataproc** for **model training**.
* **Implemented real-time monitoring & alerting** pipeline failures using **Cloud Logging & Monitoring**.

**Tech Stack:** GCP (BigQuery, Dataflow, Vertex AI, Dataproc), SQL, Python (PySpark), Apache Beam

### **Software Developer Intern | API & Data Security| Exposys Data Labs** **| Bengaluru, Karnataka, India | 04/2021 – 05/2021**

* **Developed secure API layers** for data processing systems, reducing **security vulnerabilities by 40%**.
* **Optimized backend performance**, reducing **API response time by 20%**.
* **Built data encryption & security measures** for **cloud data storage**.

**Tech Stack:** Python, API Security, Cloud Functions, SQL

**EDUCATION**

**Master’s Degree** – Information Technology  
University of Cincinnati | Aug 2023 – Dec 2024

* Relevant Courses: Advanced Algorithms, System Administration, Cloud Computing, Machine Learning, Cybersecurity

**Bachelor’s Degree** – Information Technology  
Bapatla Engineering College | Jul 2018 – Apr 2022

**CERTIFICATIONS**

**🏆 Google Cloud Professional Data Engineer (In Progress)**

**🏆 Google Associate Cloud Engineer**

**📌 Projects (Featured on GitHub/LinkedIn)**

**🚦 Real-Time Traffic Data Processing System (GCP & PySpark)**

* **Built a real-time pipeline** using **Pub/Sub → Dataflow → BigQuery** to analyze traffic data.
* **Optimized PySpark jobs in Dataproc**, reducing **execution time by 50%**.
* **Stored and queried data in BigQuery**, improving cost efficiency with **clustering & partitioning**.

**🎓 Student Performance Prediction System (ML & GCP)**

* **Developed an ML model using Vertex AI** to predict student performance based on multi-source data.
* **Built ETL pipelines in Cloud Dataflow & BigQuery** for structured student datasets.

**🔧 Additional Skills**

✅ **Version Control & CI/CD:** Git, Jenkins, Terraform  
✅ **Streaming & Batch Processing:** Apache Beam, Kafka (Basic), Cloud Dataflow  
✅ **Databases:** BigQuery, PostgreSQL, Redis, Cassandra  
✅ **Data Visualization:** Looker Studio, Power BI, Tableau  
✅ **Programming:** Python, SQL, Java (Basic), Scala (Basic)