Sharan Manne

Location: CA, USA **| Phone:** +1 806-370-0084 **| Email:** saisharanrm@gmail.com **|** LinkedIn **|** Portfolio | GitHub

Experienced **GCP Data Engineer** with around **5 years of experience** specializing in building and optimizing scalable data pipelines, cloud-native architectures, and AI-integrated workflows. Strong expertise in **Google Cloud Platform (GCP)**, including services like **BigQuery**, **Dataflow**, and **Cloud Storage**, with hands-on experience deploying robust ETL solutions and real-time data applications. Skilled in **Python**, **SQL**, and **PySpark**, with a proven track record of leveraging **LLMs** and tools like **OpenAI**, **LangChain**, and **Vector DBs** to deliver intelligent, automated systems. Collaborative and detail-oriented, with a focus on delivering clean, production-ready code and actionable insights across cross-functional teams.

SKILLS

Data Engineering & ETL: Apache NiFi, Apache Kafka, AWS Glue, Apache Spark, Apache Hive, Google Cloud Dataflow, Cloud Dataproc, Cloud Composer, Pub/Sub
Cloud Platforms & Warehousing: AWS (S3, Glue, Redshift), GCP (BigQuery, GCS), Snowflake, Azure
Programming & Scripting: Python (Pandas, NumPy), SQL, JavaScript, PySpark, Git, Docker, Kubernetes
GenAI / Large Language Models: OpenAI Models (GPT), Llama models, LangChain, LLM integration in data pipelines
Data Analysis & BI: Exploratory Data Analysis, A/B Testing, statistical techniques, data cleaning, Python visualization (Matplotlib, Seaborn)
Data Tools: Airflow, Dagster, Kafka (simulated), CI/CD workflows, API integration, version control (Git)
Visualization: Power BI, Tableau, Looker Studio
DevOps & Containers: Docker, Kubernetes (basic)
Project Management: Jira, Agile/Scrum, Confluence, Notion

EXPERIENCE

The Commons XR, CA USA **Jun 2025 – Present**
Data Engineer

- Designed and implemented secure, scalable data migration pipelines from **PostgreSQL** to **BigQuery** using **Google Cloud Dataflow** (both Builder and Template jobs).
- Built and tested **SQL** schemas for institutional datasets, including table relationships, **primary/foreign keys**, and **indexing strategies** to optimize query performance.
- Created custom visualizations using **Python** libraries (**Matplotlib**, **Seaborn**) to analyze and present data insights from initial experience runs.
- Automated pipeline validations and job runs using **GCS**, **BigQuery**, and **Dataflow UI**, reducing manual intervention and improving reliability.
- Documented **architecture diagrams**, **troubleshooting steps**, and **permission requirements** to support future production rollouts and team onboarding.

Syneos Health, USA **Jan 2025 – May 2025**
Data Engineer

- Engineered real-time data pipelines to ingest patient vitals from hospital **EHRs (Epic)**, **lab systems**, and **wearable devices** using **Apache NiFi**, **Google Cloud Storage**, and **Google Pub/Sub**, ensuring reliable and continuous clinical data flow.
- Standardized patient health records using **Dataflow** and **Python** (**Pandas**, **NumPy**) to clean, harmonize schemas, and eliminate duplicates—preparing data for downstream analytics and **ML models**.
- Designed **HIPAA-compliant** analytics workflows in **Google BigQuery** to support cross-study insights, predictive modeling, and risk flagging for patient outcomes.
- Automated **ETL** scheduling and monitoring via **Cloud Composer (Airflow)** to extract resource utilization and financial variance data from **SAP HANA/SAP BW**, optimizing budgeting across 20+ clinical sites.
- Ingested trial enrollment, compliance, and patient activity data using **Apache Kafka**, **AWS Kinesis**, and **AWS S3**, enabling sub-second data alerts and operational insights.
- Transformed and loaded structured clinical datasets using **AWS Glue** and stored in **Amazon Redshift**, improving data accessibility and HIPAA-compliant reporting for clinical teams.
- Created interactive **Tableau dashboards** visualizing recruitment trends and site performance **KPIs**—reducing data delivery time by 35% for trial managers and stakeholders.

Cognizant, India **Nov 2021 – Jun 2023**
Analyst

- Designed and maintained **cloud-based BI pipelines** integrating operational and financial data from **SAP BW**, **SAP HANA**, and **SAP IS-H** systems into **Google BigQuery** and **SQL Server**, ensuring comprehensive data availability for real-time reporting.

- Extracted, transformed, and validated large-scale datasets from diverse sources including **hospital EHRs**, **billing systems**, and **retail sales databases** using **SQL**, **Apache Hive**, and **Spark** to enhance data accuracy and consistency.
- Automated **ETL workflows** with **Google Cloud Dataflow** and **Google Cloud Storage (GCS)** to streamline data ingestion and cleansing processes, reducing pipeline failures and ensuring timely updates for downstream dashboards and reports.
- Developed and optimized interactive dashboards using **Tableau** and **Power BI**, visualizing key performance indicators such as **sales trends**, **patient treatment costs**, and **resource utilization** to enable informed business decisions.
- Conducted daily **data quality audits** identifying inconsistencies and missing data, applying **Python (Pandas)** to clean and normalize **healthcare** and **financial datasets**, thereby improving overall data integrity.
- Collaborated with **cross-functional teams** to integrate new data dimensions such as **customer behavior** and **patient demographics**, adapting **BI solutions** to evolving business requirements and enhancing analytics depth.
- Supported large-scale **data migration projects** moving hundreds of thousands of records from on-premise **SAP systems** to **Google Cloud Platform (GCP)**, ensuring **data security**, **compliance**, and minimal operational disruption throughout transition.

Hexaware, India **Jan 2020 – Oct 2021**
Data Analyst

- Cleaned and structured customer support ticket data using **Excel** and **Python**, ensuring consistent date formats, removing invalid records, and preparing datasets for analysis.
- Queried ticket databases using **MySQL** to extract issue frequency, agent performance, and customer type metrics that guided automation potential in service workflows.
- Analyzed response and resolution times with **Pandas**, identifying inefficiencies and repeat issues; used **Matplotlib** to visualize agent workloads and issue patterns.
- Built an interactive **Power BI** dashboard displaying issue categories, agent **KPIs**, and monthly ticket volumes to support **data-driven decision-making** for BPS teams.
- Discovered repetitive issue types suitable for **automation**, reducing manual workload and supporting enhancements in the **MobilityFirst dashboard's** reporting capabilities.

EDUCATION

Master of Science in Computer Science | Texas Tech University, USA **May 2025**
Bachelor of technology in Engineering | CMR Institute of Technology, India **May 2021**

CERTIFICATES

- Google Cloud Certified - Associate Cloud Engineer
- Google Data Analytics Professional Certificate

PROJECTS

Amazon E-Commerce Analytics Project

- Cleaned and structured customer support ticket data using Excel and Python, ensuring consistent date formats, removing invalid records, and preparing datasets for analysis.
- Queried ticket databases using MySQL to extract issue frequency, agent performance, and customer type metrics that guided automation potential in service workflows.

COVID-19 Data Exploration using SQL & BigQuery

- Queried and analyzed large-scale COVID-19 datasets using SQL in Google BigQuery, performing data cleaning, type standardization, and merging multiple sources to generate analytics-ready data.
- Derived key insights on global infection trends, mortality, and vaccination impact, enabling real-time monitoring and visualization of regional patterns and population density correlations.

Nashville Housing Data Cleaning with BigQuery

- Cleaned and standardized real estate housing data in BigQuery by removing duplicates, handling missing values, formatting addresses, and splitting location fields for better granularity.
- Transformed categorical data into structured formats, delivering a high-quality dataset optimized for downstream predictive modeling and market analysis.

AI File-Aware Chatbot Assistant

Python, Streamlit, Meta LLaMA-4 API, OCR (PyTesseract/PDFPlumber), dotenv

- Integrated Meta's LLaMA-4 API via an OpenAI-compatible endpoint to enable natural language conversations.
- Developed an interactive AI assistant that answers user queries based on uploaded PDF and image files.
- Built custom preprocessing utilities using OCR and PDF parsers for accurate content extraction.
- Designed a user-friendly Streamlit UI supporting file uploads, chat history, and real-time responses.
- Ensured secure handling of API keys and environment variables with python-dotenv.