Sharan Manne

Data Engineer

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

SUMMARY

Experienced Data Engineer with 4 years of expertise in designing and optimizing real-time data pipelines, ETL processes, and cloud data warehousing solutions. Proficient in tools like Apache NiFi, Kafka, AWS services, Snowflake, SQL, SAP systems, and data visualization platforms such as Tableau and Power BI. Skilled in data integration, cleansing, validation, and ensuring data quality and compliance in healthcare and business environments. Strong collaborator with cross-functional teams to deliver actionable insights and drive data-driven decision-making.

TECHNICAL SKILLS

Data Engineering & ETL: Apache NiFi, Apache Kafka, AWS Glue, Apache Spark, Apache Hive, Python (Pandas, NumPy), Google Cloud Dataflow, Cloud Dataproc, Cloud Composer, Pub/Sub
Cloud Platforms: AWS (S3, Glue, Redshift), GCP (BigQuery, GCS, Looker, Dataflow, Dataproc, Cloud Composer, Pub/Sub)
Data Warehousing: Snowflake, AWS Redshift, SAP BW, SAP HANA
Databases & Querying: SQL (Microsoft SQL Server, AWS Redshift), SAP ERP
Data Integration: SAP Data Extraction, SAP BW, SAP HANA, SAP IS-H
Data Visualization & BI: Tableau, Power BI, Looker
Data Formats & Standards: FHIR (Fast Healthcare Interoperability Resources), Healthcare Data Standards
Data Quality & Governance: Data Cleansing, Data Validation, Data Auditing, HIPAA Compliance

EXPERIENCE

Syneos Health, USA **Jan 2025 – Present**

Data Engineer

- Engineered real-time data pipelines to ingest patient health metrics from hospital EHRs (Epic), laboratory systems, and wearable devices (Apple Watch, Fitbit) using Apache NiFi, AWS S3, and Google Cloud Storage (GCS), ensuring seamless and continuous data flow.
- Developed Apache Kafka and Google Pub/Sub streaming topics to facilitate low-latency data exchange across distributed systems for immediate clinical trial monitoring.
- Applied Python (Pandas, NumPy), AWS Glue, and Google Cloud Dataflow to clean, transform, and harmonize heterogeneous data formats—standardizing units and removing duplicates—preparing datasets for downstream analytics.
- Architected scalable data storage solutions in Snowflake, Google BigQuery, and Redshift, consolidating structured patient and clinical trial data to enable efficient querying and HIPAA/FHIR-compliant reporting.
- Integrated SAP HANA and SAP BW systems to extract clinical site staffing, resource usage, and budgeting data, correlating financials with patient health metrics to optimize trial cost management by 40%, orchestrated via Cloud Composer.
- Designed and deployed interactive dashboards using Tableau and Looker to visualize trial progress, patient status, and adverse event trends—providing near real-time insights to doctors and trial managers.
- Ensured end-to-end data governance by enforcing FHIR healthcare data standards and HIPAA compliance, enabling secure, standardized patient data exchange across multiple trial locations.

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.

COVID-19 Visualization Dashboard

- Created an interactive Looker Studio dashboard displaying COVID-19 cases, deaths, and vaccination rates with global and country-level map visualizations, key metrics, and time-series trends.
- Enabled real-time data exploration and drill-down analysis, supporting policymakers and businesses in tracking the pandemic's impact and informing timely decisions.

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.

Movie Data Correlation Analysis

- Cleaned and preprocessed movie datasets by handling NaN values, correcting data types, and visualizing relationships using scatter plots, heatmaps, and regression plots.
- Performed correlation analysis using Pearson coefficients, revealing a strong positive relationship ($r > 0.85$) between budget and gross revenue, offering valuable insights for film industry financial planning.

Data Analytics and Visualization Job Simulation

- Cleaned and analyzed social media engagement data, identifying content trends and top-performing posts to inform strategic decision-making.
- Built interactive Power BI dashboards and presented insights through PowerPoint and video summaries, simulating professional client reporting scenarios.