Name: SHANMUKA CH
Role: Senior Data Engineer
Email ID: shanmukasrinivasch189@gmail.com
LinkedIn: www.linkedin.com/in/shanmuka-srinivas-ch-5a7394291
Contact: +1 9363373226

Professional Overview

- **Senior Data Engineer** with over **12** years of extensive experience designing, building, and optimizing scalable, high-performance data pipelines and cloud-native data platforms across **healthcare, financial services, insurance, and retail sectors.**
- Proven expertise in leveraging **Google Cloud Platform (GCP)** and **AWS** technologies to enable data-driven decision-making, with strong proficiency in **real-time streaming**, **data governance**, **ETL/ELT frameworks**, and **cloud migrations**.
- Adept at collaborating cross-functionally to translate complex business requirements into reliable, secure, and compliant data solutions. Skilled mentor and team leader, driving best practices in **data modeling**, **CI/CD automation**, **metadata management**, and **observability** to accelerate organizational data maturity.
- Designed and implemented scalable **data pipelines** and **data lakes** using **GCP services** such as **BigQuery**, **Dataflow**, **Cloud Composer (Airflow)**, and **Pub/Sub**, supporting large-scale analytical workloads and real-time data streaming.
- Led multiple successful **cloud migration projects** from on-premise and legacy systems (Teradata, AWS) to **cloud-native GCP environments**, ensuring minimal downtime and data integrity.
- Developed and optimized **ETL/ELT frameworks** using **Apache Beam**, **Cloud Data Fusion**, **Python**, and **dbt**, delivering modular, version-controlled, and reusable transformation pipelines.
- Architected **real-time streaming solutions** utilizing **Apache Kafka**, **GCP Pub/Sub**, and **Apache Beam**, enhancing live analytics for healthcare claims, fraud detection, and customer engagement.
- Spearheaded **data governance initiatives**, implementing **data quality frameworks**, **metadata management**, **data lineage**, and **audit trails** with tools like **Data Catalog**, **Dataplex**, and **Apache Atlas** to ensure regulatory compliance (HIPAA, HEDIS).
- Designed and deployed **CI/CD pipelines** for data engineering workflows using **Cloud Build**, **Terraform**, **GitHub Actions**, and **Jenkins**, streamlining deployment and infrastructure automation.
- Applied advanced **data modeling techniques** including **star schema**, **snowflake schema**, and **dimensional modeling**, tailored to healthcare, financial, and retail analytics needs.
- Built interactive dashboards and **data marts** using **Looker Studio** and **Tableau** for diverse stakeholders, enabling actionable insights and operational reporting.
- Ensured data security and privacy through **encryption**, **role-based access control (RBAC)**, and **HIPAA-compliant protocols** across cloud environments.
- Mentored and led junior engineers, fostering knowledge sharing and adoption of best practices in **Data Mesh architecture**, **DataOps frameworks**, and modern data engineering methodologies.
- Integrated **AI/ML-powered tools** such as **Vertex AI**, **Dataform**, and **Trifacta** for anomaly detection, automated data preparation, and enhanced analytical capabilities.
- Implemented **cost optimization strategies** on cloud platforms using tools like **Cost Explorer** and **Data Studio** to monitor and manage cloud resource utilization efficiently.
- Enhanced data pipeline observability and reliability by deploying monitoring and alerting solutions with **GCP Operations Suite (Stackdriver)**, **Cloud Monitoring**, and **Datadog**.
- Collaborated with cross-functional teams including **Data Science**, **Analytics**, **Product**, and **Compliance** to deliver secure, compliant, and business-driven data solutions.
- Standardized and optimized **SQL queries** and database performance to accelerate report generation and reduce computational costs.
- Led the design and development of **data APIs** enabling seamless integration between healthcare platforms, supporting interoperability standards such as **FHIR** and **HL7**.
- Automated infrastructure provisioning and management using **Infrastructure as Code (IaC)** tools including **Terraform** and **AWS CloudFormation**.
- Conducted comprehensive **code reviews**, documentation, and architecture design sessions to ensure maintainability, scalability, and alignment with organizational goals.

- Supported **DevOps transformations** within data teams, promoting agile methodologies and continuous delivery of data products.
- Maintained up-to-date expertise on emerging data engineering trends (2023–2025), including **Data Contracts**, **Streaming ETL**, and **Observability frameworks** to continuously innovate and improve data platforms.

Technical Expertise

Category	Tools/Technologies/Methodologies
Google Cloud Platform (GCP) Services	BigQuery, Dataflow, Cloud Composer (Airflow), Pub/Sub, Data Catalog, Dataplex, Cloud Build, Cloud Monitoring, GCP Operations Suite (Stackdriver), Vertex AI, Data Studio
Amazon Web Services (AWS)	CloudFormation, Cost Explorer
Real-time Streaming Technologies	Apache Kafka, GCP Pub/Sub, Apache Beam
ETL/ELT Frameworks & Tools	Apache Beam, Cloud Data Fusion, Python, dbt, Dataform, Trifacta
Data Modeling Techniques	Star schema, Snowflake schema, Dimensional modeling
Data Governance & Metadata Mgmt.	Apache Atlas, Data Catalog, Dataplex
CI/CD & Automation Tools	Cloud Build, Terraform, GitHub Actions, Jenkins
Data Visualization & Reporting	Looker Studio, Tableau
Data Security & Compliance	Encryption, Role-Based Access Control (RBAC), HIPAA protocols
Infrastructure as Code (IaC)	Terraform, AWS CloudFormation
Monitoring & Observability	GCP Operations Suite (Stackdriver), Cloud Monitoring, Datadog
Database & SQL	Standardizing and optimizing SQL queries, database performance tuning
API Design & Integration	Data APIs, healthcare interoperability standards (FHIR, HL7)
Agile & DevOps Methodologies	Agile frameworks, continuous delivery, DataOps frameworks
Cloud Migration & Modernization	Migrating from on-premise and legacy systems (Teradata, AWS) to cloud-native environments
Emerging Trends & Methodologies	Data Mesh architecture, Data Contracts, Streaming ETL, Observability frameworks

Education

B. TECH IN COMPUTER SCIENCE ENGINEERING **JUNE 2006 – MAY 2010**

Visvesvaraya Technological UNIVERSITY **(VTU)**

Certified Achievements

- AWS Certified: Data Engineer – Associate
- Google Cloud Certified: Professional Data Engineer

Work Portfolio

Client: AMN Healthcare Services Inc, Dallas, TX **Duration: Sept 2023 - Present**

Role: Senior Data Engineer

Responsibilities:

- Design, build, and maintain **robust, scalable, and high-performance data pipelines** using **Google Cloud Platform (GCP)** services such as **BigQuery, Dataflow, Cloud Composer, and Pub/Sub**.
- Collaborate with cross-functional teams (Data Science, Analytics, Product, and Engineering) to translate business requirements into reliable data solutions.
- Develop and optimize **data lakes and data warehouses** tailored to healthcare plan data needs, ensuring support for large-scale analytical workloads.
- Ensure **data quality, data governance, and metadata management** across all data environments using tools like **Data Catalog** and **Dataplex**.

- Architect solutions with **real-time data streaming** capabilities using **Apache Kafka**, **GCP Pub/Sub**, or **Apache Beam** for processing live healthcare data (e.g., claims, patient engagement).
- Implement and manage **CI/CD pipelines** for data engineering projects using **Cloud Build**, **Terraform**, and **GitHub Actions**.
- Utilize **Python** and **SQL** for data transformation, cleansing, and validation, adhering to industry best practices.
- Leverage **dbt (data build tool)** for modular, version-controlled transformation pipelines.
- Create **dashboards and data marts** for healthcare stakeholders using **Looker (Looker Studio)** integrated with GCP.
- Apply **data encryption, privacy controls, and HIPAA-compliant security protocols**, ensuring patient data protection.
- Monitor and troubleshoot production data workflows using **GCP Operations Suite (formerly Stackdriver)** and set up robust alerting and logging.
- Develop and enforce **data modeling standards**, including **star schema**, **snowflake schema**, and **dimensional modeling** relevant to healthcare plan metrics.
- Lead migration efforts from on-premise or legacy systems to **cloud-native GCP environments**, ensuring minimal downtime and data integrity.
- Mentor junior engineers and guide the team on **data architecture best practices**, including modern frameworks like **Data Mesh**.
- Evaluate and integrate **AI/ML-powered data preparation** and anomaly detection tools (e.g., **Vertex AI**, **Dataform**, or **Trifacta**) for enhanced data insights.
- Participate in **code reviews, documentation**, and solution architecture design to ensure maintainability and scalability.
- Work closely with data stewards and compliance officers to implement **data lineage and audit trails**.
- Stay updated with the **latest trends in data engineering (2023–2025)**, such as **Data Contracts**, **Streaming ETL**, **DataOps**, and **Observability frameworks**.
- Build and maintain **data APIs** to enable integration between healthcare platforms, ensuring interoperability via **FHIR/HL7** standards.
- Optimize cost-performance efficiency across GCP resources, recommending tools like **Cost Explorer** and **Data Studio reports** to track usage.

Environment: Google Cloud Platform (GCP), BigQuery, Dataflow, Cloud Composer, Pub/Sub, Cloud Storage, Cloud Build, Vertex AI, Dataform, Looker Studio, Dataplex, Data Catalog, Terraform, GitHub Actions, Apache Beam, Apache Kafka, Python, SQL, dbt (data build tool), Airflow, FHIR, HL7, Data Studio, GCP Operations Suite (Stackdriver), CI/CD, HIPAA compliance tools, Data Mesh architecture, Dimensional Modeling tools, DataOps frameworks, Data Contracts, Real-time streaming platforms, Metadata management tools, Data encryption solutions, Observability tools.

Client: PNC Financial Services, Pittsburgh, PA **Duration: Feb 2020 – Aug 2023**
Role: Senior Data Engineer
Responsibilities:

- Designed and developed **cloud-native data pipelines** using **Google Cloud Platform (GCP)** components such as **BigQuery**, **Dataflow**, **Cloud Composer**, and **Pub/Sub**.
- Implemented **data lake architectures** to centralize and govern enterprise-wide structured and semi-structured data across various lines of business.
- Developed and maintained **ETL/ELT frameworks** using tools like **Apache Beam**, **Cloud Data Fusion**, and **Python** for scalable data ingestion.
- Collaborated with business units to translate financial reporting requirements into efficient **data models** and transformation logic.
- Worked on **data governance and security** using **IAM**, **VPC**, and **Cloud KMS** in GCP to ensure compliance with internal and regulatory policies.
- Enabled **real-time data streaming** using **Kafka** and **Google Cloud Pub/Sub**, enhancing fraud detection and customer analytics use cases.
- Led the migration of on-premise data solutions to **GCP**, significantly improving performance, scalability, and operational cost.

- Built **CI/CD pipelines** for data workflows using **Cloud Build**, **Terraform**, and **GitHub Actions**, adhering to DevOps best practices.
- Conducted **data quality checks** and profiling using **Great Expectations** and custom **Python** scripts to ensure high data integrity.
- Partnered with Data Scientists to develop and deploy **ML-ready datasets**, facilitating advanced analytics for risk modeling and customer segmentation.
- Standardized and optimized **SQL queries and stored procedures** to accelerate dashboard and reporting layer performance.
- Implemented **monitoring and alerting systems** using **Stackdriver**, **Cloud Monitoring**, and **Datadog** for proactive data pipeline management.
- Acted as a mentor for junior engineers and conducted regular **code reviews**, promoting clean, reusable, and modular code practices.
- Designed **financial data marts** and analytical layers using **dimensional modeling (Star/Snowflake schemas)** for lending and treasury business units.
- Engaged with stakeholders to prioritize backlog items and roadmap features based on market-driven demand and customer-centric objectives.
- Orchestrated complex workflows using **Apache Airflow** (via **Cloud Composer**) to automate dependencies and maximize compute efficiency.
- Ensured cross-functional coordination with Product Managers, QA teams, and Compliance during agile sprint planning and delivery.
- Managed **metadata cataloging and lineage tracking** using **Data Catalog**, improving discoverability and audit readiness.
- Developed reusable templates and components for faster onboarding of new data sources and environments, accelerating time-to-market.

Environment: GCP (Google Cloud Platform), BigQuery, Cloud Dataflow, Cloud Composer, Pub/Sub, Cloud Storage, Cloud Build, Cloud Functions, Cloud Data Fusion, Apache Beam, Apache Airflow, Kafka, Terraform, GitHub Actions, Stackdriver, Cloud Monitoring, Data Catalog, IAM, VPC, Cloud KMS, Python, SQL, Great Expectations, Datadog, Docker, Jenkins, Linux/Unix, Agile/Scrum, JIRA, Confluence.

Client: Guidewire Software Inc, San Mateo, CA **Duration: June 2017 – Jan 2020**

Role: Senior Data Engineer

Responsibilities:

- Led the **migration of enterprise-scale data systems** from **Teradata to Google Cloud Platform (GCP)**, ensuring minimal downtime and data integrity.
- Designed and maintained **ETL/ELT pipelines** using tools such as **Apache Beam**, **Dataflow**, and **Cloud Composer**, optimizing for performance and scalability.
- Developed and implemented **data lake architectures** on **Google Cloud Storage (GCS)** using **BigQuery** for real-time analytics and historical trend analysis.
- Collaborated with cross-functional teams to **define data governance policies**, enabling high-quality, reliable, and secure data access across business units.
- Conducted **code reviews and performance tuning** for SQL queries, Python scripts, and Spark jobs to ensure efficiency and compliance with industry standards.
- **Migrated critical legacy Teradata datasets** to **BigQuery**, leveraging **Data Transfer Services (DTS)** and custom ingestion frameworks.
- Built **data validation and quality assurance frameworks**, integrating **unit tests** and **CI/CD pipelines** using Jenkins and GitLab.
- Created **real-time streaming solutions** using **Apache Kafka**, **Pub/Sub**, and **Dataflow**, supporting live analytics and operational dashboards.
- Developed **dashboards and reporting solutions** using **Looker and Tableau**, aligned with product and customer analytics goals.
- Participated in **architecture review boards**, advising on best practices for data platform design, especially in cloud-native environments.

- **Optimized storage and compute costs** on GCP by analyzing query patterns and implementing partitioned and clustered tables in BigQuery.
- Maintained and scaled **data ingestion frameworks** from APIs, flat files, and database sources using **Python, SQL, and Airflow**.
- Implemented **role-based access control (RBAC)** and **encryption policies** across GCP services to meet security and compliance needs.
- Guided junior data engineers and analysts on **cloud-native development practices**, data modeling, and analytics engineering.
- Supported the **DevOps transformation of data pipelines**, integrating Terraform for IaC (Infrastructure as Code) in GCP environments.
- Coordinated with business stakeholders and product managers to translate **business requirements into technical specifications**.
- Developed and maintained **metadata management and lineage tracking** using tools like **Data Catalog** and **Apache Atlas**.
- Provided **post-migration performance benchmarking** reports and improvement plans, showcasing the value of cloud migration to stakeholders.

Environment: Teradata, Google Cloud Platform (GCP), BigQuery, Cloud Storage (GCS), Dataflow, Apache Beam, Apache Kafka, Pub/Sub, Cloud Composer, Airflow, Looker, Tableau, Jenkins, GitLab, Python, SQL, Spark, Terraform, Data Catalog, Apache Atlas, CI/CD, Role-Based Access Control (RBAC), ETL/ELT, Data Lakes, Real-time Streaming, Metadata Management, Data Governance, Infrastructure as Code (IaC).

Client: Evolent Health Inc, Arlington, VA **Duration: Jan 2015 – May 2017**
Role: Data Engineer
Responsibilities:

- Design and develop **robust data pipelines** to support the ingestion, transformation, and loading of large-scale healthcare datasets.
- Implement **cloud-based data infrastructure**, focusing on **AWS-to-GCP migration** using services like BigQuery, Dataflow, and Pub/Sub.
- Build scalable **ETL frameworks** using **Apache Beam**, **Apache Airflow**, and **Python**, ensuring high availability and low latency.
- Maintain and optimize **data lake architectures** with **GCP Cloud Storage** and **BigQuery**, ensuring efficient access to structured and semi-structured data.
- Collaborate with Data Analysts and Data Scientists to **curate datasets** that enhance predictive modeling and operational reporting.
- Ensure **data quality**, integrity, and compliance with healthcare regulations such as **HIPAA** and **HEDIS** standards.
- Develop and enforce **data governance policies**, including metadata management, lineage tracking, and access controls.
- Create and manage **CI/CD pipelines** using **Cloud Build** and **Terraform**, ensuring smooth deployment of data workflows.
- Apply **data modeling** best practices using tools like **dbt** and **Erwin**, supporting healthcare plan analytics and actuarial reporting.
- Perform **performance tuning** of SQL queries and transformation logic in **BigQuery** to reduce cost and improve efficiency.
- Leverage **Kafka** and **Dataflow** for **real-time data processing** and streaming analytics for population health management.
- Document and maintain **version-controlled data architecture artifacts**, workflows, and procedures using **Confluence** and **GitHub**.
- Lead the integration of **third-party APIs** and claims data into centralized data warehouses for unified healthcare reporting.
- Participate in regular **code reviews**, architectural discussions, and data pipeline optimization sprints.
- Research and implement **market-leading practices** in data engineering prevalent such as **Lambda Architecture** and **Columnar Storage Optimization**.

- Collaborate with cross-functional teams in **Agile** environments to deliver business-critical data features on time and within scope.

Environment: Google Cloud Platform (GCP), BigQuery, Dataflow, Pub/Sub, Cloud Storage, AWS, Redshift, S3, Glue, Apache Beam, Apache Airflow, SQL, Python, Spark, dbt, Terraform, Cloud Build, Kafka, CI/CD pipelines, GitHub, HIPAA compliance, HEDIS standards, Data Governance, Agile methodology.

Client: Kroger, Mason, OH **Duration: July 2012 – Dec 2014**
Role: Data Engineer
Responsibilities:

- Design, develop, and maintain scalable **data pipelines** to support Kroger's retail analytics and business intelligence initiatives.
- Build and optimize **ETL processes** using tools like **Apache Hadoop**, **Apache Spark**, and **AWS Glue** for efficient data ingestion and transformation.
- Manage and maintain **AWS data services** such as **Amazon S3**, **Redshift**, **EMR**, and **Kinesis** for storage, processing, and real-time data streaming.
- Collaborate with data scientists, analysts, and business stakeholders to understand data requirements and deliver high-quality datasets.
- Implement **data quality checks** and **monitoring frameworks** to ensure accuracy and reliability of retail sales and inventory data.
- Develop and enforce **data governance policies** in compliance with retail industry standards and Kroger's internal guidelines.
- Optimize SQL queries and database performance for **Amazon Redshift** and relational databases such as **Amazon RDS**.
- Use **Python** and **Scala** for scripting and automation of data workflows and data processing tasks.
- Design and implement data models that support Kroger's customer insights, product assortment, and supply chain analytics.
- Conduct data profiling and troubleshoot issues related to data discrepancies and pipeline failures.
- Utilize **AWS CloudFormation** and **Terraform** for infrastructure as code to automate the deployment of data infrastructure.
- Participate in cross-functional Agile teams to continuously deliver and improve Kroger's data solutions.
- Stay current with emerging big data technologies and retail market trends to recommend innovative solutions for Kroger's data challenges.
- Support migration of on-premise data infrastructure to the **AWS Cloud**, improving scalability and cost efficiency.

Environment: AWS (S3, Redshift, EMR, Kinesis, Glue, CloudFormation, RDS), Hadoop, Spark, Python, Scala, SQL, ETL tools, Data Governance, Agile, Terraform, Big Data Technologies, Cloud Migration, Data Modelling, Data Quality Tools.