

Saketh Bantu | Cloud Infrastructure Data Engineer | Agentic AI & MLOps

Chester Springs, PA | (484) 943-1942 | sakethban2@gmail.com | www.linkedin.com/in/sakethbantu/

Professional Summary

Highly experienced **Cloud Data Engineer** (Gen AI, MLOps & Infrastructure Architecture) with 5+ years currently driving data innovation in the CPG (Product: Consumer Packaged Goods) sector. Proven expertise in building **AI Agents** and scalable data products that power commercial analytics. Proven ability to ensure the reliability, performance, and availability of complex distributed systems using **AWS**, **GCP**, and **Kubernetes**. Adept at leveraging observability tools (Prometheus, Grafana) and modern ETL frameworks (Spark, **Kafka**, **Airflow**) for mission-critical enterprise analytics. Passionate about automating incident triage, building self-healing architectures, and driving root cause analysis (RCA) to optimize production environments.

Technical Skills

Category	Skills
<i>Cloud Platforms</i>	Azure (Data Factory, Synapse, Blob, AD), AWS (S3, Glue, SageMaker, Redshift), GCP (BigQuery, Vertex AI, Cloud run)
<i>Data Lake & Warehousing</i>	Microsoft Fabric , Medallion Architecture , Delta Lake, Snowflake, Databricks, Redshift, Azure SQL DB
<i>Generative AI & MLOps</i>	LangChain , RAG , LLMs , Vector Databases (Pinecone) , MLflow, Airflow, Kubeflow, CI/CD, Prometheus, Grafana, TensorFlow,
<i>Data Engineering</i>	PySpark , SQL , ETL/ELT, Kafka, AWS Glue, Azure Logic Apps, Kusto Query Language (KQL), Azure Fabric(Onelake) , Databricks Apache Spark (PySpark) , Apache Kafka , Redshift , Azure SQL DB
<i>Microservices & DevOps</i>	Docker , Kubernetes (EKS) , Prometheus , Flask, FastAPI, Jenkins, Terraform, Git/GitHub, Azure AD, OAuth
<i>Visualization & Reporting</i>	Power BI , Tableau, Grafana , Streamlit, Azure Data Explorer (ADX)
<i>Languages</i>	Java, React.js, Javascript, typescript, Python, SQL, Shell Script, Bash,

Certifications

- **Microsoft Certified: Azure Data Fundamentals (DP-900)**
- **Microsoft Certified: Azure Security Engineer Associate (AZ-500)**
- **AWS Cloud Practitioner – Trained**

Education

- **Master of Science, Cybersecurity and Information Security**, Eastern Illinois University - May 2023
- **Bachelor of Technology, Computer Science and Engineering**, JNTUH - May 2020

Research Publication

- "Galaxies classification using Deep Learning Algorithm in Convolutional Neural Networks". UGC care group. May 6, 2020

Professional Experience

Cloud Data Engineer (AI ML Pipelines) | May 2023 – Present

Church & Dwight; Princeton, NJ

- **AI Agent Development:** Architected and operationalized an intelligent **Retrieval-Augmented Generation (RAG)** pipeline using **LangChain** and OpenAI API. Enabling automated enterprise knowledge retrieval and report summarization for business stakeholders.

- **Commercial Data Scale:** Architected high-throughput ETL/ELT workflows using **PySpark** and **Azure Data Factory** to integrate **100M+ records**, delivering critical data for analyzing customer usage patterns and commercial insights.
- **Pipeline Optimization:** Improved ETL throughput by **60%** by optimizing Databricks Spark-SQL applications, aggregating data from disparate sources to uncover actionable business insights.
- **MLOps Implementation:** Established a robust MLOps platform by deploying low-latency, containerized ML inference services using Docker and Kubernetes for scalable model serving.
- **Data Reliability:** Implemented comprehensive observability (Prometheus/Grafana) and data validation (Great Expectations) to ensure 100% data integrity for critical production reports.
- Experience in developing Spark applications using **Spark-SQL** in **Databricks** for data extraction, transformation, and aggregation from multiple file formats for Analyzing& transforming the data to uncover insights into the customer usage patterns.
- Built scalable, type-safe frontend dashboards using **TypeScript** and React to visualize complex commercial analytics and provide an interactive user interface for custom-built LangChain AI agents.
- **Managed multi-cloud data environments** including **GCP BigQuery** and **Azure SQL Database**, ensuring high reliability, performance, and security for critical training and production data.

Cloud Data & Infrastructure Engineer | July 2022 – December 2022

Toyota Forklift NW, Kent, WA.

- **Real-Time Analytics:** Designed event-driven ETL pipelines utilizing **Kafka** and PySpark for scalable processing of time-series IoT sensor streams, supporting predictive maintenance.
- Developed high-performance data transformation modules using **Java**, integrating seamlessly with Apache Kafka and Spark to accelerate real-time stream processing and enhance backend microservice reliability.
- **Data Lakehouse:** Engineered a Data Lakehouse architecture combining S3, Delta Lake, and Redshift, reducing ETL runtime by **45%** through partition pruning and caching optimizations.
- **Self-Healing & CI/CD:** Built robust orchestration frameworks using Airflow DAGs and AWS Lambda, managing containerized data services on Amazon EKS clusters with automated schema validation to enable self-healing deployments.
- Implemented event-driven serverless functions using Node.js and **JavaScript** within AWS Lambda and Azure Logic Apps to automate API data extraction and trigger continuous MLOps training pipelines.
- **Predictive Modeling:** Developed and deployed predictive maintenance models in SageMaker, integrating outputs with Grafana dashboards for real-time visualization.
- Implemented **CI/CD** for data ingestion with **Airflow DAGs** and **AWS Lambda**, ensuring automated schema validation and data profiling using **Great Expectations**.
- **Established an API layer for data access** via **Fast API** microservices, supporting consumption by downstream applications.

Data Analyst/ BI Engineer | November 2020 – October 2021

Vamstar.io (Lifesciences AI); Hyderabad, India.

- **Business Intelligence:** Developed complex data models and analytics artifacts, culminating in business reports and dashboards in **Tableau** and **Power BI** to track critical KPIs.
- **Data Governance:** Implemented role-based security and authentication protocols for Tableau

Server, ensuring governance and controlled access to sensitive commercial reports. Configured role-based security, authentication protocols, and access controls for Tableau Server, ensuring strict governance for sensitive commercial reports.

- **Strategic Analysis:** Utilized advanced Python and SQL to extract and manipulate large datasets, simplifying complex data for strategic business decision-making.
- **Optimized dashboard performance** and user experience by implementing warm caching techniques and Tableau features for seamless web application integration.