SAI CHAITANYA SAMA

+1 (407) 984-2956| saichaitanya.sama@gmail.com | Data & Analytics Engineer | LinkedIn | Portfolio

EDUCATION

Master of Science in Computer Science — **Aug 2024 - May 2026**
University of Central Florida | Orlando, Florida, US — **CGPA: 4.0/4.0**

Bachelor of Technology in Computer Science — **Jun 2018 - Jul 2022**
DVR and Dr. HS MIC College of Technology | Kanchikacherla**,** Andhra Pradesh, India — **CGPA:3.7/4.0**

TECHNICAL SKILLS

- **Programming & Development:** Python, Java, SQL, Scala, PySpark
- **Databases:** MySQL, PostgreSQL, SQL Server, T-SQL, MongoDB, Hive
- **Data Warehousing & Cloud:** Snowflake, Azure Synapse Analytics, Amazon Redshift, Microsoft Azure (Data Factory, Synapse and Fabric), AWS (S3, Glue, Athena, Redshift), GCP (BigQuery, Looker Studio), PalantiR
- **Data & Analytics Tools:** Power BI, Tableau, Databricks, Apache Airflow
- **Methodologies:** Data Visualization, Data Engineering, Machine Learning (Foundational), Agile & Scrum, Data Modelling and Data Warehousing.
- **Productivity & Collaboration:** GitHub, JIRA, Miro, Microsoft Office Suite (Excel, PowerPoint, Word, Outlook)
- **Functional Competencies:** Data Analysis & Visualization, Data Ingestion & Integration, Analytical Thinking & Problem Solving, Research & Business Issue Identification, Communication (Verbal & Written), Time Management & Prioritization, Adaptability in Dynamic Environments, Accountability, Integrity & Professionalism.

PROFESSIONAL WORK EXPERIENCE

Data Engineer | EPAM Systems | Telangana, India — **Nov 2022 - Aug 2024**

- **Enhanced system performance** by identifying and resolving infrastructure bottlenecks, **improving responsiveness by 38%** and ensuring **seamless interaction between front-end user interfaces and backend systems**.
- **Developed and optimized real-time ETL pipelines** across **Snowflake**, **Azure Data Lake**, and **AWS S3**, increasing **data flow efficiency by 57%** and **processing speed by 35%** through effective **data ingestion, transformation, and architecture improvements.**
- **Designed scalable data workflows** leveraging **Azure Data Factory** and **Azure Functions,** implementing metadata-driven triggers that **reduced batch latency by 42%** and enhanced **cross-platform data integration accuracy by 33%.**
- **Automated analytics workflows** using **PySpark** and **Foundry dashboards**, decreasing manual effort by **84**% and delivering **real-time business insights** that improved **decision-making effectiveness by 52%.**
- Collaborated with cross-functional teams to **analyze business requirements**, apply data visualization and modeling techniques, and **ensure data consistency across cloud data platforms** (AWS, Azure, Snowflake).
- **Conducted root cause analysis** on **ETL process failures** and implemented 10+ changes that reduced by 38%.

Data Engineer Intern | EPAM Systems | Telangana, India — **Jan 2022 - Oct 2022**

- **Configured and managed Azure Data Lake Storage Gen2**, designing scalable data storage solutions across **15+ storage accounts** and processing **500+ CSV datasets**, ensuring efficient data ingestion and accessibility for analytics.
- **Engineered and refined end-to-end data pipelines** using **Azure Databricks, Python (Pandas), and SQL**, transforming large datasets into **Parquet format for Snowflake integration**, which improved **data processing speed by 63%** and enhanced **data architecture efficiency**.
- **Analyzed and interpreted large data sets** through **Python and SQL**, uncovering key insights that **informed strategic business decisions**, leading to a **25% improvement in decision accuracy and operational efficiency**.
- **Built and maintained a robust data warehouse structure**, streamlining **data integration** and reducing **retrieval time by 20%**, while improving **data quality and validation accuracy** across systems.
- **Designed and automated Power BI dashboards and reports**, providing **real-time data visualization** and saving **10+ team hours per week**, improving visibility into business performance metrics.
- **Collaborated with cross-functional teams** (analysts, data engineers, and business stakeholders) to optimize data processes, increasing **team productivity by 15%** and ensuring **accurate, timely reporting**.
- **Optimized SQL queries and ETL processes**, reducing **execution times by 38%** and improving **code coverage and performance reliability by 15%** across reporting systems.
- **Demonstrated adaptability, accountability, and analytical thinking** in a dynamic, cloud-based environment aligning technical outcomes with business objectives and maintaining data integrity throughout the process.

PROJECTS

Lakehouse-Based Modern Data Warehouse Architecture **Jul 2025 – Oct 2025**

- Designed and implemented a **modern data warehouse (Lakehouse architecture)** using **Azure Data Factory**, **Azure Databricks**, and **Apache Airflow** to ingest and transform GitHub source data into analytics-ready datasets.
- Built multi-layer **data pipelines (Bronze, Silver, Gold)** on **Delta Lake** with incremental loads, data cleansing, normalization, enrichment, and **star schema modeling** (facts and dimensions with **SCD Type 1**).
- Developed and orchestrated **end-to-end ETL workflows** by combining **ADF pipelines**, **Databricks notebooks**, and **Airflow DAGs**, enabling automated and scalable data processing.
- Integrated **Microsoft SQL Server**, **Unity Catalog**, and GitHub data sources to ensure strong **data governance**, **metadata management**, and optimized **analytics performance**.

Automated Sales Data Processing Pipeline **Jan 2025 – Mar 2025**

- **Developed and deployed a** modular data workflow **using** Apache Airflow**,** PostgreSQL**, and** Docker**, eliminating manual Excel-based steps and cutting** daily sales data processing time by 60%**.**
- Structured a **multi-layer data architecture** (staging, refined, curated) to manage ingestion, cleansing, validation, and aggregation, resulting in well-defined **analytical data models** for reporting.
- Automated **secure data delivery** to **Power BI** dashboards through Airflow-managed triggers, ensuring **timely insights** for Sales and Marketing teams.
- Created **reusable Airflow components** for schema setup, data loading, and permission handling, improving **pipeline transparency, auditability, and maintenance efficiency** across the system.

Azure Retail Data Pipeline Project **Sep 2024 – Dec 2024**

- **Built a** cloud-based data workflow **leveraging** Azure Data Factory**,** Databricks (PySpark)**, and** Data Lake Storage Gen2 **to consolidate and prepare diverse data sources for analytical use.**
- Collected and refined data from **Azure SQL Database** and **REST APIs**, converting raw JSON feeds into structured datasets ready for downstream analysis.
- Implemented **data transformation logic** in Databricks notebooks, performing filtering, joins, and summarizations to produce **high-quality analytical tables** optimized for scalability.
- Designed **interactive Power BI dashboards** connected to curated data layers, delivering **actionable insights** into retail sales trends, customer behaviour, and store performance.

CERTIFICATIONS

- Microsoft Certified: Fabric Data Engineer Associate| **Microsoft (DP-700)**
- Microsoft Certified: Power BI Data Analyst Associate| **Microsoft (PL-300)** **Oct 2025**
- Databricks Certified Data Engineer Associate | **Databricks Academy** **May 2024**
- Academy Accreditation - Databricks Lakehouse Fundamentals | **Databricks Academy** **Feb 2024**
- Azure Data Engineer Associate | **Microsoft (DP-203)** **May 2023**
- Data Fundamentals | **Microsoft (DP-900)** **Nov 2022**
- Azure Fundamentals | **Microsoft (AZ-900)** **Sep 2022**