SAI KISHORE CHINTALA

Data Engineer

saikishorechintala01@gmail.com | 206-671-8651 | Redmond, WA | linkedin.com/in/sai-kishore-Chintala

PROFESSIONAL SUMMARY

Data Engineer who builds production **ETL/ELT pipelines** using **SQL** and **PySpark** for querying and **Python** for scripting and automation, on Databricks and **Azure Data Factory**, with hands-on experience in **big data technologies** (**Spark, Hadoop, Hive, Kafka**). Specializes in **data modeling**, **data warehousing** and **data architecture** (medallion, star schema), with automated **data quality** checks and monitoring for enterprise analytics platforms. Improves **reporting and analysis** processes through automated Tableau and Power BI dashboards built with cross-functional stakeholders. Holds an M.S. in Data Science (GPA 3.8/4.0) with academic training in **machine learning, statistical analysis, clustering and hypothesis testing**, and keeps expanding expertise in dbt and Databricks through self-directed projects and certifications.

TECHNICAL SKILLS

Query & Scripting Languages: Advanced SQL, SQL Stored Procedures, PySpark, Python
Big Data Technologies: Apache Spark, Hadoop, Hive, Kafka, AWS EMR, Databricks (Accredited: AI Agent Fundamentals), Delta Lake, Spark Structured Streaming
ETL Tools & Orchestration: Azure Data Factory, dbt, Apache Airflow
Data Modeling & Warehousing: Data Modeling, Data Warehouse Design, Dimensional/Star Schema Design, Slowly Changing Dimensions, Medallion Architecture, Snowflake, SQL Server, PostgreSQL, MySQL, MongoDB (Non-Relational Database)
Academic Skills (M.S. Data Science): Machine Learning, Statistical Analysis, Prediction, Clustering, Hypothesis Testing, HiveQL, Spark SQL, PL/SQL, DDL, MDX, KornShell, Informatica, SSIS, ODI, BODI, DataStage
Data Quality: Data Validation, Data Quality Monitoring, Error Handling, Data Lineage Documentation
Cloud & DevOps: AWS (Amazon S3, EMR, Lambda, SageMaker), Microsoft Azure, Azure DevOps, Docker, Git, GitHub Actions, CI/CD
BI & Reporting: Microsoft Power BI (DAX), Tableau, Data Visualization, KPI Reporting, Trend Analysis
Collaboration: Technical Design Reviews, Cross-Functional Stakeholder Partnership, Mentoring, Agile/Scrum

PROFESSIONAL EXPERIENCE

Novel Technologies | Data Engineer | Remote **October 2026 – Present**

- Build **medallion (bronze-silver-gold) ETL pipelines** in Azure Databricks with **PySpark** that turn raw data into analytics-ready tables.
- Build automated **data quality checks** in PySpark and **Delta Lake** that validate data at every pipeline layer (bronze, silver and gold).
- Integrate HL7 v2.x and CCD/CCDA healthcare data into **real-time Databricks pipelines** using **Spark Streaming** and **Python**.
- Automate deployment and testing of data pipeline jobs through **CI/CD** with GitHub Actions and Azure DevOps.

Aptech Staffing | Data Engineer | Remote **April 2026 – October 2026**

- Designed and maintained scalable **ETL/ELT pipelines** in **Azure Data Factory** and **Databricks (PySpark)**, implementing **data quality frameworks** and **error handling** to keep downstream systems reliable.
- Automated manual data-prep workflows and deployments with **Azure DevOps** and **Git**, supporting **CI/CD** practices and improving pipeline reliability.

Capgemini | Data Engineer | Chennai, India **May 2022 – Aug 2023**

- Designed **dimensional data models** and built **SQL stored procedures** to ingest data from **8+ SAP ERP systems** into **Snowflake and PostgreSQL** data warehouses, improving reporting accuracy by **20%**.
- Automated a reporting process with a **Python** pipeline that ingests unstructured data and generates visual reports, saving the analytics team **15+ hours per week**.
- Designed and deployed **12+ automated reporting dashboards** in **Tableau and Power BI** with custom **DAX** measures, giving stakeholders real-time visibility into logistics KPIs.
- Optimized **SQL queries** and data aggregation strategies, improving dashboard load performance by **40%**.

PROJECT EXPERIENCE

End-to-End Retail Data Pipeline | Databricks, dbt, Apache Airflow, AWS S3, Delta Lake

- Designed a **medallion architecture** on Databricks and Delta Lake with **25+ incremental dbt models**, using AWS S3 as the storage layer.
- Orchestrated the pipeline with **Apache Airflow DAGs**, added automated **data quality tests**, and documented **data lineage** end to end.

Real-Time Streaming Data Pipeline (Azure) | Azure Data Factory, Event Hub, Databricks, PySpark, Kafka

- Processed streaming events with **PySpark Structured Streaming** on Databricks, using Event Hub, Kafka and Azure Data Factory.
- Designed **star schema data models** with **Slowly Changing Dimension (SCD)** logic for the streaming data.

EDUCATION

Lewis University | M.S. in Data Science | GPA: 3.8/4.0 | Aug 2024 – Dec 2025 (transfer from Seattle University, Sep 2023 – Jun 2024)
Dr. MGR University | B.E. in Electronics & Communication Engineering | GPA: 3.5/4.0 | Jul 2018 – May 2022

CERTIFICATIONS

Databricks Accredited AI Agent Fundamentals (Sep 2026) | Google Python Crash Course | Data Science & AI, Naresh Technologies (Dec 2022 – May 2023)