

Ravi Rajpurohit

+1-682-373-7319 • [Portfolio](#) • ravirajpurohit414@gmail.com • [linkedin.com/in/ravi-rajpurohit/](https://www.linkedin.com/in/ravi-rajpurohit/)

SUMMARY

Data Professional with experience designing robust data platforms and aggregate data layers that power executive reporting and ML inference. Expert in **Python, SQL, and Schema Design**, with a proven track record of managing data stores containing **100M+ records** to ensure availability and scalability. Adept at bridging the gap between business needs and engineering by building secure pipelines and compelling, **intuitive visualizations**.

SKILLS

Data Engineering: Python, SQL, PySpark, Data Modeling, Schema Design, ETL/ELT, Data Warehousing, Kafka

Visualization: Tableau, D3.js, Plotly, QuickSight, Power BI (Creating compelling, intuitive visualizations)

Big Data & Cloud: AWS (Glue, Athena, Lambda, RedShift, Step Functions), Apache Spark, Hadoop, Databricks, Snowflake

Databases: PostgreSQL, DynamoDB, Cassandra, MongoDB, RDS, HBase

DevOps & Tools: Docker, Kubernetes, Jenkins, GitHub Actions, Terraform, Splunk, Prometheus, Airflow, Claude Code

EXPERIENCE

Data Infrastructure Engineer, State of Michigan (Ottawa Area ISD) *January 2025 - Present*

- Designed and implemented a comprehensive **aggregate data layer** on AWS (S3, Glue, Athena), consolidating **15+ siloed data sources** to enable efficient **executive reporting** for leadership.
- Developed robust data pipelines using Python and Spark to process and clean 5M+ daily records, reducing data availability latency by 40% for downstream analytics.
- Created beautiful and intuitive **visualizations** and dashboards to track **user engagement** metrics and **behavioral trends**, cutting report generation time from 3 days to < 1 hour.
- Collaborated closely with business partners and Quality teams to identify key KPIs, designing data schemas that support both analytical solutions and engineering requirements.

Data Engineer - Research Applications, University of Texas at Arlington *August 2024 - January 2025*

- Built scalable data pipelines to support **data science initiatives**, enabling powerful ML inference on high-frequency sensor data streams (**200Hz throughput**).
- Collaborated effectively with data scientists to design and maintain data models that bridged the gap between raw signal collection and complex analytical requirements.
- Automated ETL workflows using **Apache Airflow**, reducing processing latency by 35% and enabling near real-time visualization of experimental metrics.
- Engineered feature extraction tools using Python (Pandas, NumPy) to transform raw signals into compelling data stories.

Data Engineer Intern, Nutanix *May 2023 - August 2023*

- Designed and maintained data pipelines supporting secure visualizations for Sales and Support teams, ensuring the **scalability** of infrastructure handling **500GB+** of daily logs.
- Integrated observability tools (Prometheus, Splunk) to monitor pipeline health, maintaining **99.9% availability** of data services during peak loads.
- Managed backend data stores and optimized SQL query performance by 50%, handling complex data sets to drive cloud cost optimization insights.
- Utilized CI/CD pipelines (Jenkins) to automate testing and deployment, reducing deployment time by 20 minutes per cycle.

Software Data Engineer - MLOps, KaHa Technologies *January 2020 - July 2022*

Context: KaHa powers smart wearables, similar to Apple Watch services, serving 10M+ users

- Designed and managed end-to-end data pipelines using **Python** and **AWS** to support real-time health insights (Sleep, Stress, Vitals), processing **2B+ monthly events**.
- Built a high-throughput messaging architecture using **Kafka** to ingest telemetry data, enabling real-time visualizations and reducing data lag to **<5 seconds**.
- Collaborated with **cross-functional teams** to build a data platform that powered highly customized data applications, contributing to a 25% increase in user retention.
- Maintained NoSQL datastores (MongoDB/DynamoDB) for **high-speed retrieval** of user profile data, ensuring scalability to support 10k concurrent requests.

EDUCATION

Master of Science - Computer Science, University of Texas at Arlington

August 2022 - May 2024