

SRINATH KETHAVATH

(341)-356-3550 | srinath.keth@gmail.com | Hayward, CA
[LinkedIn Profile](#) | github.com/srinath-keth

PROFESSIONAL SUMMARY

- Data-driven Business & Data Engineer with expertise in Big Data, Cloud Infrastructure, and Advanced Analytics, driving automation, optimization, and cost savings across industries.
- Designed and deployed scalable ETL pipelines using Apache Spark, Python, SQL, and Airflow, reducing manual data processing by 60% and improving data reliability to 99.9%.
- Built and optimized real-time data processing using Kafka, Spark Streaming, and AWS Glue, enabling fraud detection systems that saved \$2M annually in financial losses.
- Developed predictive models using AWS SageMaker, TensorFlow, and Databricks, enhancing forecasting accuracy, reducing downtime by 25%, and optimizing supply chain efficiency.
- Created enterprise-level BI dashboards in Power BI, Tableau, and Looker, driving data-driven decision-making and increasing reporting efficiency by 50%.
- Integrated business intelligence tools with SAP ERP, Salesforce, and IoT analytics, leading to 20% cost savings in procurement and operational workflows.
- Automated issue tracking & reporting using Jira, Alteryx, and Apache NiFi, streamlining workflows and improving collaboration across cross-functional teams.

TECHNICAL SKILLS

Programming: Python, SQL, Scala, R, Shell Scripting, Linux/Unix.

Data Engineering Tools: Apache Airflow, Apache Kafka, dbt, Talend, Informatica, Apache Spark, Hadoop, Hive, Apache NiFi, Alteryx.

Data Warehousing & Modeling: Snowflake, Amazon Redshift, Google BigQuery, Star and Snowflake Schema Design.

Cloud Platforms: AWS (Glue, Lambda, RDS, S3, SageMaker), Azure (Data Factory, Data Lake, Synapse Analytics), GCP (BigQuery, Dataflow).

Frameworks & BI Tools: Pandas, NumPy, TensorFlow, PyTorch, Scikit-Learn, MLflow, Tableau, Power BI, Looker.

Version Control & CI/CD: Git, Docker, Kubernetes, Terraform.

Database Management: PostgreSQL, MySQL, Oracle, MongoDB, Cassandra, Pinecone, FastAPI, Flask, Django REST.

PROFESSIONAL EXPERIENCE

Data Engineer

Broadridge Financial Solutions

Nov 2023-Dec 2024

Sacramento, CA

- Designed and implemented Apache Spark and Python-based data pipelines to process multi-terabyte datasets, achieving 99.9% data accuracy and improving operational efficiency across business units.
- Deployed Big Data workflows on AWS using Glue, Lambda, and Step Functions, reducing ETL processing time by 40% and enabling faster decision-making for key stakeholders.
- Integrated real-time data from Kafka streams and optimized Spark Streaming workflows, supporting fraud detection and saving the company \$2M annually.
- Built and containerized scalable applications using Docker and Kubernetes, reducing deployment errors by 25% and ensuring system reliability for mission-critical tasks.
- Developed actionable insights with Hive, Power BI, and Tableau, delivering automated reports that cut manual effort by 50% and enhanced revenue forecasting accuracy.

Data Engineer

Softusvista

Sep 2021 – Nov 2022

Remote

- Designed and implemented scalable ETL pipelines using Apache Airflow and AWS Glue, processing 1TB+ of data weekly from ERP and CRM systems.
- Developed real-time streaming workflows using Apache Kafka and Spark, reducing data pipeline latency by 25% for critical applications.

- Built and optimized data warehouses in Snowflake and Redshift, improving query performance by 40% and supporting business intelligence reporting.
- Designed Looker models and Tableau dashboards to visualize KPIs, empowering business teams with actionable insights and reducing manual reporting time by 60%.

Business Data Analyst
Dixon Technology Group

Sep 2020 – Aug 2021
Remote

- Engineered and automated ETL pipelines using SQL, Python, Apache Spark, and Airflow, reducing manual data processing by 60% and enabling real-time manufacturing analytics for improved efficiency.
- Implemented machine learning models using AWS SageMaker and Databricks for predictive maintenance and pricing optimization, cutting production downtime by 25% and improving supply chain resilience.
- Automated reporting and issue tracking using Alteryx, Apache NiFi, Snowflake, and Jira, streamlining workflows for the Head of Business Intelligence and ensuring seamless collaboration across teams.

PROJECTS

AI-Driven Location Feasibility Platform

[Project Link](#)

- Developed a SaaS platform to identify optimal business types and services for commercial properties, solving real-world challenges by leveraging advanced machine learning, analytics, and scalable data pipelines.
- Deployed a two-stage machine learning pipeline in AWS SageMaker to analyze and predict business feasibility, Incorporated ChatGPT-driven sentiment analysis and generated Business Suitability Scores (BSS) with 85% accuracy, recommending business types tailored to market demand and location attributes.
- LLMs (OpenAI & LangChain) for sentiment analysis, improving business type recommendations by 85% accuracy.

Real-Time and Batch Data Pipelines

[Project Link](#)

- Designed and implemented a scalable data pipeline using Apache Kafka, Docker, and PostgreSQL, enabling seamless real-time streaming and batch processing of synthetic stock price and sales trend data.
- Developed ETL workflows with Python, Apache Spark, and Airflow, automating data ingestion, transformation, and reporting, reducing processing latency by 40%.

AI Powered News Research Platform

[Project Link](#)

- Developed an AI-driven news research tool using Python, LangChain, and OpenAI GPT-4, enabling users to extract, process, and query news articles with 90% accuracy in retrieving relevant information.
- Implemented FAISS-based vector search, optimizing semantic retrieval speed by 50%, ensuring fast and precise results from stored embeddings.
- Enhanced AI-powered insights by integrating Sentence Transformers and UnstructuredURLLoader, improving text parsing efficiency for 3x faster processing.

EDUCATION

Master of Science in Business Analytics,
California State University, East Bay | GPA: 3.8

Graduation date: Dec 2024
Hayward, California

CERTIFICATIONS

- Microsoft Certified Azure Data Engineer - Associate
- IBM Certified Professional Data Engineer - Coursera