HARNAIN KAUR SARDARNI

Chicago, Illinois | +1-872-271-0725 | harnainkaursardarni@gmail.com | Linkedin | github
Data Engineer | ML Enthusiast | Distributed Systems Engineer

OBJECTIVE

Data Engineer with 3+ years of experience designing scalable data pipelines, streaming applications, and ETL solutions across cloud and on-premise systems. Proficient in PySpark, Kafka, Hive, AWS, Azure, Python, SQL, and ML model integration for predictive analytics and NLP.

SKILLS

Programming & ML:	Python, R, MATLAB, CUDA, Scikit-learn, TensorFlow, PyTorch, XGBoost, LightGBM, NLP
Languages:	Java, SQL, PL/SQL, Scala, Pig Latin, HiveQL, Regex
Big Data & Distributed Systems:	Spark, Kafka, HDFS, Hive, Sqoop, NiFi, HBase, Cassandra, Zookeeper
Cloud & DevOps:	AWS, Azure, GCP, Docker, CI/CD, Git, GitHub, Model Deployment
Analytics & Visualization:	Excel, Tableau, Power BI, SPSS, SAS, Feature Engineering
AI & NLP:	BERT, GPT, LangChain, HuggingFace, RAG, Knowledge Graphs, Conversational AI
Databases:	Oracle, SQL Server, MySQL, MongoDB, Teradata, HBase, Cassandra
Tools:	Jupyter, Maven, Jenkins, WebLogic, Ant, SBT

PROFESSIONAL EXPERIENCE

JB Hunt Transport, AR **July 2024 – Present**

Data Engineer

- Designed and implemented scalable data pipelines using PySpark, Hive, Kafka, and NiFi for both batch and real-time processing with HDFS and MapReduce.
- Developed optimized ETL workflows in Python and SQL, extracting, transforming, and loading data from various sources (RDBMS, NoSQL, S3) for BI reporting.
- Built real-time data solutions using Spark Streaming and AWS S3, storing transformed data in HDFS and Cassandra for advanced analytics.
- Integrated machine learning models (XGBoost, LightGBM, TensorFlow) for predictive analytics and NLP tasks in cloud environments (AWS, Azure).
- Automated workflows using CI/CD pipelines, Git, and AWS Lambda to enhance operational efficiency and reduce manual intervention.

Ryder Transport, IL **Feb 2023 – June 2024**

Data Engineer

- Designed and implemented scalable data pipelines using Apache Spark (PySpark), Kafka, and Flume, improving data processing efficiency by 40%.
- Developed ETL workflows with NiFi, Sqoop, and Airflow for structured and unstructured data across distributed systems.
- Optimized PySpark jobs with RDDs, DataFrames, and Spark SQL, reducing job execution time by 30%.
- Built real-time streaming data pipelines for sensor and log analytics using Kafka Streams and Spark Streaming.
- Utilized AWS (S3, EC2, Glue, EMR) for cloud-native data processing and containerized applications using Docker and Kubernetes.
- Created BI dashboards in Power BI and Tableau; deployed and monitored CI/CD pipelines with Jenkins and Git.

WNS Global Services - Hyderabad, India **May 2021 – July 2022**

Assistant Data Engineer

- Developed scalable ETL pipelines using Apache Spark (PySpark) and Kafka for processing structured and unstructured data.
- Gained hands-on experience with Hive, HDFS, and Sqoop to process large-scale datasets and optimize data workflows.
- Designed and implemented real-time data pipelines using Kafka and Spark Streaming for low-latency processing.
- Worked with relational (MySQL, PostgreSQL) and NoSQL databases (MongoDB, Cassandra) for data integration and storage.
- Utilized AWS (S3, EC2, Glue, EMR) for cloud-based solutions, enhancing scalability and reducing infrastructure costs.
- Created interactive data visualizations and dashboards with Tableau and Power BI to enable data-driven decision-making.

EDUCATION

DePaul University **Dec 2024**

- *Master of Science, Data Science, Computational Methods*