P Shiva Kumar

+1 (806) 412-8900 • [pshiva6991@gmail.com](mailto:pshiva6991@gmail.com) • Houston, Texas, United States of America

GitHub: [github.com/shiwhoa](https://github.com/shiwhoa) • LinkedIn: [linkedin.com/in/p-shiva-kumar/](https://www.linkedin.com/in/p-shiva-kumar/)

# SUMMARY

## Seasoned Data Engineer and AI Engineer with deep domain knowledge in the Refineries and Oil and Gas Industry.

## Expertise in Big Data (Batch and Streaming) and AI Engineering with RAG (Vector DB and Knowledge Graphs).

# EXPERIENCE

## BKO AI, Houston, TexasJan 2024 – Current

*Data Engineer/ Knowledge Graph Engineer/ AI Engineer*

* Entasked to create a real-time data pipeline to transfer on-site refinery IOT data to monitor, analyze, visualize and predict.
* Built Data Streaming pipelines using Apache **Kafka** Azure Functions to transfer refinery IOT data to Databases.
* Spearheaded adoption of Apache **Flink** and built Flink cluster on cloud to transform real-time refinery IOT data.
* Deployed Apache **Airflow** Scheduler DAGs to orchestrate IOT aggregation using Apache Spark.
* Established Azure **Databricks** pipeline and independent Apache **Spark** cluster to analyze refinery IOT data.
* Built knowledge graphs around refinery infrastructure in **neo4j** with efficient **cypher** queries for **GenAI** to use as **RAG** tool.
* Developed in-house **GenAI** to hit REST endpoints, **Vector database in Pinecone**, and other RAG tools.
* Developed independent monitoring tools such as **Prometheus-Grafana** on VMs.

## EOG Resources, Houston, TexasDec 2022 – Jan 2024

*Data Engineer*

* Built Data Streaming pipelines using Apache **Pulsar**, **Kafka** and **RabbitMQ** to transfer IOT data from Oil Wells to Databases.
* Built APIs via **FastAPI** managed on Azure and **Kubernetes** on-premises to demonstrate Oil Well Operations on Web.
* Built Apache **Airflow** mechanism to schedule tasks on Oil Well Finances.
* Optimized business logic of source code to retrieve API get calls 20% faster using python **Multiprocessing** and **GraphQL**.

## Coldwell Banker Richard Ellis (CBRE), Richardson, TexasJun 2022 – Aug 2022

*Data Engineer/ NLP Intern*

* Built and implemented **Huggingface transformer NLP models** on **AWS** EC2 Streamlit API.
* **Statistical modeling** on domain data to attain relationships between various variables.
* Performed **Exploratory Data Analysis** (EDA) on data from **PostGres** with parallel computing on **Dask**.

## ExxonMobil Corporation, Singapore, SingaporeDec 2016 – Sep 2020

*Process Engineer/ Optimization Engineer/ Data Engineer/ Technical Lead*

* Established highest refinery reliability using **JMP** and **Tableau** within past 5 years leading to margins of ~1MUSD/Year.
* Executed key optimizations and loss preventions using **Linear Programming** with savings of ~1.4MUSD/Year.

# EDUCATION

## Texas Tech University, Lubbock, Texas Jan 2021 - Dec 2022

*Master of Science in Computer Science (GPA 3.90/4.00)*

**Master’s Thesis**: Metaphor detection using Transformer models (Skills: **Python,** **Transformers, Deep Learning, NLP**)

## National University of Singapore, Singapore, Singapore Aug 2012 – May 2016

*Bachelor of Engineering in Chemical Engineering (Honors)*

# AWARDS AND HONORS

## Texas A&M Hackathon Winner for best NLP Hack. Produced automated Invoice tracker and information extraction. 2022

## ExxonMobil Outstanding Award for excellent contribution during Plant Turnaround. 2020

## ExxonMobil Plant Restart Award for most efficient startup with minimum cost. 2020

# TECHNICAL SKILLS

**Languages**: ***Python, C#, Java | SQL, Cypher***

**Containerization**: **Docker, Kubernetes**

**Big Data: Kafka, Flink, Pulsar, RabbitMQ, Spark, Databricks**

**GenAI DB: Pinecone Vector DB, Neo4j Knowledge Graph DB**

**GenAI Agent tools: Langchain Agents, Llama-Index Agents**

**LLM: Huggingface Tokenization, Chunking and Vector Embedding**

**Machine Learning: Regression, Classification, Reinforcement Learning, Ensemble Learning** *(Bagging, Boosting, Stacking)*

**Deep Learning: CNN, RNN** *(LSTM, GRU)***, Transformers** *(BERT, RoBERTa, GPT-3)*