

Lupanch

Email id: lupanch.gupta.tech@gmail.com

Ph#: 858-375-9303

Professional Summary:

- Over **7+ years** of IT experience in **AI/ML and Data Engineer** across diverse industries, specializing in designing and developing scalable **big data pipelines, data lakes, and ETL/ELT workflows** on cloud platforms like **AWS, Azure, and GCP**.
- Excellent understanding of all stages in a typical SDLC like Requirement Analysis, Design, Programming, Project Status Review (PSR), Unit Testing, Integration Testing, Support.
- Proficient in **PyTorch, TensorFlow, FastAPI**, and scalable model deployment with **MLOps** tools across cloud environments.
- Experienced in applying LLMops and safety frameworks (PromptLayer, Guardrails.ai, TruLens) to ensure.
- Expertise reliability in and building compliance and deploying of **GenAI LLM-applications**. Powered applications using LangChain, Llama, **OpenAI APIs**, and transformer models for tasks like document summarization, retrieval, and customer service automation.
- Built scalable ML pipelines with **Vertex AI, Kubeflow, and Docker**, ensuring smooth model training, deployment, and monitoring in production environments.
- Strong hands-on expertise in Large Language Models (LLMs) including designing and deploying LLM-driven systems for document parsing, recommendation engines, and generative **AI applications**.
- Expert in the full ML lifecycle—from data ingestion, processing, model development, **MLOps**, to real-time deployment—leveraging tools such as **TensorFlow, PyTorch**, Apache Spark, Azure Synapse, and **Databricks**.
- Experienced in **Databricks & Big Data** Technologies mainly on Data and Delta Lake Implementation.
- Good understanding of **Spark Architecture** including **Spark Core, Spark SQL, Data** Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
- Skilled in **data** ingestion, modeling (Star/Snowflake), performance tuning, and processing structured and semi-structured **data** using Spark SQL, MapReduce, Spark Streaming, and Spark MLlib.
- Well-versed in cloud-native **data** services like **AWS Glue, Redshift, Azure Data Factory, GCP BigQuery, Dataflow**, Pub/Sub, and orchestration tools like **Apache Airflow, SSIS, and Flume**.
- Good understanding of business requirements, Warehouse schemas and mapping rules for the implementation of simple to complex **ETL designs**.
- Expert in implementing various business rules for Data Extraction, Transforming and Loading (ETL) between Homogenous and Heterogeneous Systems using **Azure Data Factory (ADF)**.
- Good Understanding of **Azure Big data technologies** like **Azure Data Lake Analytics, Azure Data Lake Store, Azure Data Factory**.
- Experience in Migrating SQL database to Azure **Data** Lake, Azure **Data** Lake Analytics, Azure SQL Database, **Data** Bricks and Azure SQL **Data** Warehouse and controlling and granting database access and Migrating On-premises databases to Azure **Data** Lake store using Azure **Data** Factory.
- Experience in Developing Spark applications using Spark - SQL in Databricks for **data** extraction, transformation, and aggregation from multiple file formats for analyzing & transforming the **data** to uncover insights into customer usage patterns.
- Extensive experience in developing tabular and multidimensional **SSAS Cubes**, Aggregation, KPIs, Measures, Partitioning Cube, Data Mining Models, deploying and Processing **SSAS objects**.
- Scheduled reports for Daily, Weekly, and Monthly reports on sales and marketing information for various categories and regions based on business needs using the **Power BI** reporting tool.
- Good knowledge of Data Visualization including producing tables, graphs, and listings using various procedures and tools such as **Tableau**.
- Experienced in Data Migration from RDBMS to **Snowflake cloud data warehouse**.
- Exposure to extracting raw data from different sources, establishing database connections for various databases and preparing analysis reports on different parameters. Also, possess knowledge of **ETL**.
- Proven expertise in optimizing **Hive SQL queries**, Spark Jobs, distributed transactions, high availability setups, and **AWS utilization** for large dataset **ETL processing**, system reliability.
- Experienced in development and support, demonstrating in-depth knowledge of Python, SQL, Oracle, SQL, PL/SQL, and T-SQL queries.

- Ability to work independently and as part of a team to accomplish critical business objectives as well as good decision-making skills under high-pressure complex scenarios.
- Excellent communication skills, with an ability to understand the concepts and technical and non-technical requirements.
- Experienced in using agile approaches, including Extreme Programming, Test-Driven Development and **Agile Scrum**.

Technical Skills:

Programming & Scripting	Python, Scala, PySpark, SQL, Java, Bash
AI/ML	RAG, MCP, LLMS (LARGE LANGUAGE MODELS), AI/ML (ARTIFICIAL INTELLIGENCE), AWS Sagemaker, Jupyter, Azure Machine learning, Databricks notebooks, TensorFlow, PyTorch, FastAPI, scikit-learn.
Spark Framework	Spark API, Spark Streaming, Spark Structured Streaming, Spark SQL
ETL Data Pipelines	Apache Airflow, Sqoop, Flume, Apache Kafka, SSIS
Big Data Systems	Amazon Web Services (AWS), Azure, Google Cloud Platform (GCP), Apache Spark, Spark Streaming, Apache Kafka, Hive, Amazon S3, AWS Kinesis
Visualization	Tableau, Power BI
Cloud Platforms	AWS, GCP, Azure
Scheduler Tools	Apache Airflow, Azure Data Factory, AWS Glue, Step functions
Databases	HBase, DynamoDB, MongoDB, BigQuery, SQL, Hive, MySQL, Oracle, PL/SQL, RDBMS, AWS Redshift, Amazon RDS, Teradata, Snowflake
CI/CD Tools	Jenkins, GitHub, GitLab
Operating Systems	Windows, Linux, Unix, Mac OS X

Professional Experience:

GenAI/ML/Data Engineer | Client: Broadway Bank, San Antonio, TX | Duration: Mar 2025 – Till Date

Responsibilities:

- Assist in building **Data** pipelines to improve **data** quality and facilitate iterations for accommodating new user requirements.
- Integrated **Generative AI tools (Hugging Face Transformers, OpenAI APIs)** to support demand forecasting and supply chain risk analysis, enhancing decision-making processes with LLM-generated insights.
- Involved in developing end-to-end machine learning pipelines, fine-tuning Large Language Models (LLMs), and applying **GenAI** for real-world automation and insight generation.
- Worked with **Chat GPT and Chat GPT API** to leverage its capabilities with several applications.
- Prototyped lightweight **RAG-like** context chaining with BERT embeddings and NER output before full-stack GenAI frameworks became standardized.
- Developed and deployed enterprise-grade **RAG** assistants using **GPT-4, LangChain, and Vertex AI**, enabling high-accuracy Q&A and document intelligence.
- Built **GenAI** models using LangChain, prompt engineering, and Azure Data Lake, supporting optimized database access and full data pipelines for critical strategic projects.
- Created RESTful services using **FastAPI**, integrated with **PostgreSQL, DynamoDB, and S3** for robust data handling.
- Designed MLOps pipelines using **Vertex AI** for **CI/CD**, streamlining deployment and model integration in production.
- Proactively working on building **data** lineage which would display incoming and outgoing reference of an asset by extensive use of **data** pipelines and other Meta tools such as Uni dash, Dai query, Diffs and Sev.
- Performed **data** transformation and processing tasks within **Snowflake** by leveraging **Snowflake's** built-in SQL capabilities. Also, **data** quality and consistency by implementing **data** validation checks, **data** cleansing, and normalization processes.
- Architected and implemented end-to-end **ETL pipelines** using **Azure Data Factory, SSIS, and HDInsight** for ingesting data from source systems to **Azure SQL DB, Azure Data Lake, and Azure Synapse**.

- Led one-time large-scale data migration projects from on-premise Oracle to Azure SQL Data Warehouse, including schema design, star schema modeling, and optimization.
- Developed data profiling, cleansing, and rollback frameworks with automated batch pipeline restarts, implemented data masking, encryption, and SSIS IR for secure and reliable workflows.
- Used **Azure Databricks and PySpark** for data transformation, curation, and building scalable ELT jobs; created **Databricks** notebooks and managed Spark clusters with high concurrency.
- Built real-time streaming pipelines using **Kafka and Spark** Structured Streaming; handled ingestion from Azure Event Hubs and batch processing with Hive bucketing and partitioning.
- Developed and managed ADF pipelines with features like lookup, stored procedures, data flows, Azure Functions, copy activity, and data ingestion using **Azure Blob Storage**.
- Design the **data** bricks notebooks using **Pyspark** for **data** extraction, transformation and landing the **data** into the **data** bricks Delta Table.
- Performed preliminary data analysis using descriptive statistics and handled anomalies such as removing duplicates and imputing missing values.
- Performed data mining on large datasets (mostly structured) and raw text data using different Data Exploration techniques.
- Developed **Spark applications** using **Pyspark and Spark-SQL** for data extraction, transformation, and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
- Working extensively with Databricks to build scalable and efficient **data** processing and transform **data** pipelines to handle high-volume **data** ingestion, transformation, and aggregation.
- Leveraged Databricks notebooks and **Spark clusters** to perform **data** ingestion, ETL (Extract, Transform, Load), and **data** wrangling tasks.
- Dump **data** into Azure **Data** Lake Storage and analyzing raw **data** using Azure **Data** Lake Analytics jobs.
- Implement complex **data** transformations using **Pyspark, SQL, and Data** Frame operations within the Databricks environment.
- Design, build and launch efficient & reliable **ETL** pipelines to move and transform **data** (both large and small amounts).
- Created **Azure Databricks notebooks in Pyspark** for transforming raw JSON **data** into structured **data**.
- Developed and optimized data warehouse for datasets using **Azure Data Lake Storage** Explorer and Kusto Data Explorer.
- Data Ingestion to one or more **Azure Services** - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In **Azure Databricks**.
- Perform bash operations, using **GIT** for version control.
- Run SQL DDL/DML scripts to set up database objects in **Azure SQL Database**.
- Collaborate with **Data** Scientists and ML Engineers to prepare and curate large-scale structured and unstructured datasets for **AI/ML** model training and deployment using **Vertex AI**.
- Collaborated with cross-functional teams to gather and analyse business requirements, translating them into technical specifications and data models.

Environment: GenAI, ML, LangChain, FastAPI, Data Warehouse, AirFlow, Kafka, Spark, MapReduce, Hadoop, Snowflake, Hive, Azure, PySpark, Docker, Kubernetes, AWS, MangoDB, CI/CD, Tableau, Redshift, Power BI, Rest APIS, Teradata, GCP, Windows.

AI/ML/Data Engineer | Client: Ameriprise Financial, Minneapolis, MN | Duration: Jan 2023 – Feb 2025

Responsibilities:

- Monitored incoming **data** analytics requests and distributed results to support IoT hub and streaming analytics.
- Prepared documentation and analytic reports, delivering summarized results, analysis, and conclusions to stakeholders.
- Applied core **NLP techniques**—tokenization, stemming, lemmatization, sentiment scoring, and POS tagging—to evaluate rhetorical structure, sentiment intensity, and public reaction trends.
- Integrated **ChatGPT API** into client-facing platforms, automating 86% of manual tasks and reducing support load through contextual AI interactions.

- Built production-grade, dynamic-response **chatbots using the LangChain** Framework, leveraging LLMs for real-time, context-aware answers across high-volume environments.
- Developed LLM-based assistants using **GPT-3 (Azure OpenAI) and LangChain**, integrated into Microsoft Teams for HR and IT support automation.
- Analyzed AI-generated responses and iteratively improved prompt structures, enhancing model accuracy and reducing hallucinations by 30%.
- Implemented continuous monitoring frameworks for **GenAI** applications, ensuring regulatory compliance and minimizing bias in AI outputs.
- Developed an **AI-powered migration** tool that transforms Java Struts applications into FastAPI and React using OpenAI's real-time API and Hugging Face models.
- Developed reusable ML/AI models for forecasting and operational risk mitigation across nodes in the supply chain, deployed via Vertex AI Pipelines and Kubeflow.
- Integrated **LLM-powered** recommendation systems that increased user retention by 25% through context-aware product suggestions.
- Developed and maintained **data** pipelines to extract, transform, and load (ETL) **data** from various sources into the company's **data** warehouse.
- Created Python and Spark **data** pipelines for ETL parsing and analytics, resulting in a structured **data** model.
- Conducted comprehensive performance tuning and optimization of complex SQL queries, resulting in an improvement in database response time.
- Deployed the **Cloudera Hadoop** framework for distributed computing, increasing **data** processing throughput over a cluster of up to seventeen nodes.
- Involved in creating Hive tables, loading with **data**, and writing Hive queries on top of **data** present in HDFS.
- Created **data** visualizations using Tableau and other Power BI tools to support **data** analysis and decision-making.
- Developed **CI/CD pipelines** to automate deployment processes and ensure **data** solutions are continuously integrated and delivered.
- Participated in transforming **Hive/SQL queries** into transforms using Python.
- Created pipelines in **ADF** using linked services, datasets, and pipelines to extract, transform, and load **data** from different sources like **Azure SQL**, blob storage, the **Azure SQL Data Warehouse**, the write-back tool, and backwards.
- Extract, transform, and load **data** from source systems to Azure **Data** Storage services using a combination of Azure **Data** Factory and **Data** Lake Analytics.
- **Data** ingestion to one or more Azure services (Azure **Data** Lake, Azure Storage, and Azure SQL) and processing the **data** in Azure **Data** bricks.
- Managed version control of ETL/ELT scripts and **data** science models within **CI/CD** workflows.
- Wrote **Spark-Streaming** applications to consume the data from **Kafka topics** and write the processed streams to **HBase**.
- Configured, assembled, and dispatched new information pipelines underway utilizing **Apache Spark**.
- Performing **Sqoop tasks** to load data into HDFS and conduct validations.
- Designed an **Apache Airflow** Data Pipeline to automate data ingestion and retrieval.
- Integrated automated testing into **CI/CD** pipelines to ensure data quality and reliability.
- Developed solutions for detecting lost history records, hard deletes, and missing data in **ETL**.
- Built data validation checks to validate the data integrity and consistency between the source and target databases.
- Utilized **SQL** for querying and analyzing large datasets.
- Developed a Databricks script to load data from Redshift into Snowflake and perform data quality tests.

Environment: GenAI, ML, NLP, LongChain, Kafka, Spark, AWS, Azure, Python, Scala, Airflow, ETL, SSIS, Redshift, Data Factory, Data Bricks, Jira, SQL, Snowflake, Power BI, Data Cleaning, Data Profiling, Data Mining and Windows.

Data Engineer | Client: Gentiva Healthcare Services, Tampa, FL | Duration: Feb 2019 – Dec 2022

Responsibilities:

- Applying the Azure PaaS service, analyze, create, and develop modern **data** solutions that facilitate **data** visualization. Recognize the application state in production at the moment and assess how a new implementation will affect ongoing business procedures.

- Engineered near real-time pipelines using Apache Spark and Cloud Dataproc, enabling event-driven analytics on high-volume datasets.
- Developed **Python scripts** for **data** cleansing, transformation, integration, and migration, incorporating complex SQL logic and PL/SQL stored procedures.
- Using a combination of **Azure Data Factory**, T-SQL, Spark SQL, and U-SQL Azure **Data** Lake Analytics, extract, transform, and load **data** from source systems to Azure **Data** Storage services. **Data** processing in Azure Databricks following **data** ingestion to one or more Azure Services (Azure **Data** Lake, Azure Storage, Azure SQL, and Azure DW).
- Written multiple python/API scripts using python to connect to Tableau to refresh/extract the **data**.
- To extract, transform, and load **data** from many sources, including Azure SQL, Blob storage, Azure SQL **Data** Warehouse, write-back tools, and rearward, pipelines were created in ADF utilizing linked services/datasets/pipeline.
- Built Spark apps for **data** extraction, transformation, and aggregation from various file formats using PySpark and Spark-SQL. The **data** was then analyzed and transformed to reveal insights into client usage trends.
- Experience in Object Oriented Programming (Scala, Python), Unix scripting or related programming languages and exposure to some of Python's ML ecosystem (numpy, panda, sklearn, TensorFlow, etc.).
- Participated in gathering team input for some product business and functional requirements, updated user comments in JIRA 6.4, and updated Confluence documentation.
- Practical expertise in **data** cleansing, which involves eliminating mistakes, inconsistencies, and duplication to prepare **data** for analysis. To comprehend the properties of the **data** and spot patterns and trends, use exploratory **data** analysis, or EDA.
- Good Experience with ODBC, JDBC, and Other Connection Properties for **Data** Extraction from Various **Data** Sources, Including Databases, Flat Files, and Web Scraping
- Worked together to guarantee **data** availability and quality with the infrastructure, network, database, application, and business intelligence teams.

Environment: Azure Data Factory, Databricks, Azure Synapse, Datalake, Blob Storage, ETL, Python, PySpark, TensorFlow.

References: Will be provided upon request.