

PROFESSIONAL SUMMARY

- Data Engineer with 5+ years of experience developing, implementing, and optimizing data pipelines and ETL processes.
- Proficient in managing **Spark and Hadoop, Apache Storm** on-premises and in the cloud; seeking leadership role to drive data initiatives and foster a culture of data-driven decision-making.
- Strong experience in developing various web-based application technologies like **HTML, CSS, JavaScript, Typescript**.
- Good experience in writing stored procedures, functions, and triggers using **PL/SQL**.
- Good understanding of React JS and working with **Redux / Flux architecture**.
- Expertise in working with operating systems like **Windows and Linux**.
- Excellent analytical and communication skills with the capability to handle new technologies.
- Quick learner and able to adopt any kind of work culture and proficient in solving technical issues in the project.
- Experience in working with cloud technologies like **Amazon Web Services, and Azure**.

SKILLS

Programming languages:	C++, Java, Python, R, Spark, Rust, Hadoop, OOP's, Kafka, Scala
Database:	MYSQL, MongoDB
Cloud Services:	GCP, Azure, AWS cloud, S3, Redshift, Glue.
Web Technologies:	HTML, CSS, JavaScript, Node.js.
Bigdata Services:	Airflow, HDFS, Databricks, data lake
Version Tools:	GitHub, Visual Studio, MS Excel,
ML and AI	AI, ML, LLM, NLP, Deep Learning concepts

CERTIFICATIONS

- **2023** Certified as **AWS Associate Developer**.
- **2023** Earned Badges for **Generative AI** on **Google Cloud**.
- **2019** Certified for 'Database Programming with SQL By **ORACLE ACADEMY**'.

EXPERIENCE

Capital One | Dallas

April-2022 – Present

Data Engineer

- Designed and developed robust and scalable data pipelines, leveraging technologies and Large-scale datasets such as **Apache Spark, RDBMS, NoSQL, Spring Boot, Cosmos DB, Cassandra, and GRA orchestration to Snowflake, Kafka, Hadoop, Flume**, and SQL, to ensure seamless data flow and accessibility for analytical purposes.
- Implemented **Kubernetes**-based solutions for resource optimization, resulting in a 25% reduction in infrastructure costs while maintaining high-performance levels.
- Expert in designing and optimizing ETL processes using SSIS, building multidimensional models with SSAS, and creating insightful reports with SSRS.
- Implementing new tools and best practices, streamlining workflows, finding bottlenecks, and streamlining the development and deployment process are all ways to continuously assess and enhance the CI/CD pipeline.
- Experienced in utilizing DevOps and Quality tools like Azure DevOps, GitHub actions, and Jenkins to streamline workflows and ensure code integrity and quality.
- Writing Scala code to perform data analysis tasks such as data cleansing, aggregation, and transformation.
- Successfully managed the scaling of complex data-driven software platforms to accommodate increasing data volumes, enhancing overall system reliability and responsiveness.

- With Snowflake, I've designed and implemented scalable and reliable data pipelines to handle large volumes of data efficiently. I've optimized data systems to improve reliability, efficiency, and quality, collaborating closely with peers, analysts, and product teams.
- Expertise of creating scalable and effective data processing pipelines, ETL procedures, and streaming apps by utilizing Golang's standard library and concurrency paradigm.
- Performed end-to-end Architecture and implementation assessments of various **AWS** services like **Amazon EMR, redshift, S3, Athena, Glue, and Kinesis**.
- competent in API creation, system or application integration, containerization, and the deployment of **LLMs** into production environments.
- Implemented data security measures in **AWS**, including **IAM** roles and policies, to ensure compliance with industry standards and regulations.
- knowledgeable about utilizing Git to work with cross-functional teams, settle disputes, and smoothly integrate code updates.
- On a day-to-day basis, **experienced Agile Scrum methodology and Jira ticket** system for project development.



Ford | Remote

June 2019 -Dec 2021

Data Engineer

- Implemented and managed robust **CI/CD** pipelines for data processing workflows using **Jenkins**, ensuring smooth and automated deployment of ETL processes.
- Experienced data engineer proficient in Java, SSIS, SAAS, and SSRS for seamless ETL processes and insightful reporting.
- proficient about creating idiomatic, effective, and clean Golang code for developing microservices, APIs, and applications for data processing.
- Worked with different Big Data components like Spring, **Hive, Apache Spark, Tableau, PostgreSQL, SQOOP & HBase, Kafka, and Docker** for data processing and storage.
- Implement a Warehouse Management System (WMS) connection to simplify data flows between logistics and inventory systems. improved data accessibility and accuracy, which led to a 20% increase in the effectiveness of inventory management.
- Writing unit tests and documentation for **Scala** code to ensure reliability, maintainability, and ease of understanding for other team members.
- Leveraged Python's extensive libraries including NumPy, Pandas, and **sci-kit-learn** to preprocess and analyze large-scale datasets.
- Proficient in advanced techniques such as Logistic Regression, Generalized Linear Models, and decision trees for accurate predictive modeling.
- ability to create new features and extract pertinent features from text data to improve the **LLMs'** input data and increase the predicted accuracy and strength of the models.
- Leveraged cross-functional expertise to bridge the gap between data engineering and data science teams, facilitating effective communication and knowledge sharing. Experienced in integrating **machine learning models** to optimize data processing workflows.
- competent in cloning repositories, committing changes, branching, merging, rebasing, and marking releases using Git commands.
- Every day, working with the **Jira ticket system** and Agile Scrum technique for the creation of projects.

T-Mobile | Remote

May 2018 -May 2019

Data Engineer

- Proficient in developing and optimizing data pipelines, with expertise in scripting languages like Python, Java, SQL, and Informatica and advanced skills in Windows Batch scripting.
- Implemented efficient ETL processes using IIS DataStage and Informatica, ensuring seamless data integration and transformation.
- Proficiency in optimizing LLMs using domain-specific datasets to enhance model performance and customize it to meet certain business needs.
- incorporating automated testing frameworks (such as **Scala Test and JUnit**) into the **CI/CD** pipeline to verify the accuracy and dependability of data pipelines and applications.



- Implementing Spark java pipelines to do the ETL processing on top of the raw HDFS data and writing into Hive table.
- Combining Scala-based apps with databases, data warehouses, and data visualization tools, as well as other elements of the data ecosystem
- Experienced in Designing and building data processing pipelines using tools and frameworks in the Hadoop ecosystem.
- Proficient in ETL processes using Azure Data Factory, Kubernetes, **Apache Spark SQL**, GraphQL, and **pySpark** for **Azure Data Lake Analytics** using technologies such as **Node.js** and **Spring Boot**. I have expertise in front-end development with **C#**, **UNIX Shell scripting**, **.NET Framework React.js**, and **HTML5** for creating interactive user interfaces, resulting in streamlined data ingestion and processing workflows.
- Expertise in building Databricks notebooks for data extraction, cleansing, loading, testing, and validation to Azure SQL DB. **Cosmos DB**, **Cassandra**, and **CCM** from sources like DB2 and Teradata. Proficient in deploying pipelines in **Azure Data Factory** using **JSON scripts**.

PROJECTS

- The proposed OSN system aims to combat cyberbullying in social media. It utilizes a semantic-enhanced marginalized denoising autoencoder with dropout noise for specialized representation learning in cyberbullying detection.
- **Emotion Recognition Using Genetic Algorithm and XGBoost**: This study uses encephalography, a brain-computer interface, **genetic algorithms**, and **XGBoost** to identify emotions in immobile people.

EDUCATION

Master of Science – Computer Technology

Eastern Illinois University | Charleston, IL

Jan 2022-May 2023.