Khaja mujahiddin Mohammed

khajamujahiddin@gmail.com | +1 347-736-5812 | LinkedIn | Portfolio

PROFESSIONAL SUMMARY:

- **Data Engineer | Data Analyst** with **4+ years of experience** in building scalable data pipelines, real-time streaming solutions, and cloud-native analytics platforms across **AWS, Azure, and Hadoop ecosystems**.
- Skilled in **designing ETL workflows** with **AWS Glue, Lambda, SSIS, PySpark, and SQL Server**, integrating structured/unstructured data from databases, APIs, and flat files into centralized data lakes and warehouses.
- Proficient in **machine learning and NLP model development** (sentiment analysis, time-series forecasting, recommendation systems) using **Python, R, spaCy, NLTK, ARIMA, Prophet, and LSTM**, driving predictive insights and personalization strategies.
- Delivered **20+ interactive dashboards and KPIs** using **Power BI, DAX, Power Query, and R Shiny**, enabling 500+ stakeholders across **finance, operations, and retail** to make faster and more accurate data-driven decisions.
- Hands-on with **real-time data processing** using **Kafka, Kinesis, and Spark**, optimizing data workflows for scalability, reducing manual reporting by 40%, and achieving **99.9% cloud uptime**.
- Strong expertise in **data modeling (OLTP, OLAP, star schema, denormalized models)**, **Row-Level Security (RLS)**, and **cost optimization** strategies, saving ~$5K annually on cloud infrastructure.

TECHNICAL SKILLS:

Programming and Scripting	Python (Pandas, NumPy, SciPy, Matplotlib), PySpark, SQL (T-SQL, PL/SQL, MySQL, PostgreSQL), Bash Shell, R, PowerShell
Big Data & Data Engineering	Apache Spark, Kafka, Hadoop (HDFS, Hive, MapReduce), Databricks, Snowflake, Redshift, ETL/ELT Pipelines (AWS Glue, Azure Data Factory, SSIS, Talend, Informatica)
Cloud Platforms & Services	AWS (EC2, S3, Lambda, Glue, RDS, Kinesis, SNS, IAM, CloudWatch, CodePipeline, SageMaker), Data Factory, Databricks, Synapse Analytics, SQL, Blob Storage, Power BI Gateway, BigQuery.
Databases	SQL Server, Oracle, MySQL, PostgreSQL, MongoDB, Teradata, Cassandra
Visualization & BI	Power BI (DAX, Power Query), Tableau, Excel (Power Pivot, Power View, Pivot Tables)
Machine Learning & AI	Scikit-learn, TensorFlow, PyTorch, ARIMA, Prophet, LSTM, Predictive Modeling, NLP (spaCy, NLTK, Transformers)
GenAI & LLMs	OpenAI (GPT-4), LangChain, Hugging Face Transformers, Retrieval-Augmented Generation (RAG), Semantic Search, Llama Index, Chatbot Development
DevOps & Orchestration	Docker, Kubernetes, Apache Airflow, Jenkins CI/CD, Terraform, Control-M
Version Control & Collaboration	Git, GitHub, JIRA, Agile/Scrum methodologies
Operating Systems	Windows, Linux (Ubuntu, CentOS), UNIX

PROFESSIONAL EXPERIENCE:

AWS Data Engineer (AI/ML) | Staples | Boston, Massachusetts **Aug 2024 - Present**

Description: At Staples, I worked on building scalable cloud-based analytics and machine learning solutions to support both B2B and e-commerce operations. My role focused on creating real-time data pipelines, predictive models, and interactive dashboards on AWS, helping business teams make faster, smarter, and data-driven decisions.

Responsibilities:

- Built and deployed **machine learning and NLP models** (sentiment analysis, topic modeling, classification) using **Python, spaCy, and NLTK** to extract insights from unstructured customer data.
- Designed and implemented **real-time streaming pipelines** with **AWS Kinesis & SNS** and automated **ETL workflows** using **AWS Glue & Lambda**, boosting processing efficiency by ~25%.

- Migrated and optimized **ML workloads** to **AWS Cloud (EC2, S3, RDS)**, delivering **99.9% uptime** and improving scalability for 5+ internal analytics applications.
- Performed **time series forecasting** using **ARIMA, Prophet, and LSTM models** to predict demand trends and optimize resource allocation.
- Developed **Spark pipelines** for preprocessing and recommendation engines, enhancing personalization capabilities for **Staples Business Advantage (B2B solutions)**.
- Conducted **customer segmentation and churn analysis**, supporting Staples' strategy to improve retention and loyalty initiatives.
- Reduced **cloud infrastructure costs by $5K annually)** through lifecycle policies, resource utilization monitoring, and workload optimization.
- Deployed **CI/CD pipelines** with **AWS CodePipeline & CodeBuild**, reducing deployment cycles from 2 days to ~4 hours.
- Strengthened **data security and governance** by applying **AWS IAM roles, VPC configurations, and CloudWatch monitoring**, ensuring compliance and secure access control.

Environment: AWS (EC2, S3, RDS, Glue, Lambda, Kinesis, SNS, IAM, VPC, CloudWatch, Code Pipeline, Code Build), Spark, Hadoop, Python, R, spaCy, NLTK, ARIMA, Prophet, LSTM, Django, Power BI, R Shiny, SQL, Data Cleansing, Time-Series Forecasting, ETL Pipelines, Real-Time Streaming, Predictive Analytics, Machine Learning, Cloud Optimization & Security.

Data Analyst (Azure) | PowerBI Developer | SRIK Consulting Services | Hyderabad, India **Sep 2020 – July 2023**

Description: At SRIK Consulting, I worked on building real-time data pipelines and interactive Power BI dashboards for clients across healthcare, finance, and retail. My role was all about turning raw data into clear insights—whether through ETL workflows, predictive models, or automation scripts—to help teams make faster, smarter decisions.

Responsibilities:

- Designed and developed **20+ interactive Power BI dashboards** integrating **SQL Server, MySQL, and Excel** to support **Finance, Operations, and HR teams**, enabling 50+ stakeholders to make data-driven decisions.
- Created advanced **DAX measures** (forecasting, YoY/MoM trends, variance analysis) and optimized data models, improving decision-making accuracy and query performance by **35%**.
- Applied **Row-Level Security (RLS)** and managed **Power BI Server dashboards**, ensuring secure and seamless access for 500+ users across departments.
- Built KPI dashboards, drill-down reports, and automated refresh pipelines using **Power Query, SQL, and Power BI Gateway**, cutting reporting cycle time from 5 to 2 days and reducing manual reporting efforts by **40%** (~10 hours per analyst per week).
- Extracted, transformed, and loaded (ETL) data from diverse sources (databases, APIs, flat files) using **SSIS, Python, and R**, ensuring high-quality, consistent, and reliable datasets for analytics.
- Developed **real-time streaming applications** with **PySpark and Kafka** on distributed Hadoop clusters, enabling high-speed processing of large-scale datasets for clients in healthcare, finance, and retail.
- Designed and implemented **SSIS packages** (VLOOKUPs, Derived Columns, Conditional Splits, Error Handlers) to streamline ETL workflows and improve data integration efficiency.
- Conducted **customer profitability and journey path analyses** to identify high-value users, optimize retention, and enhance conversion rates for key clients.
- Applied **predictive modeling and machine learning techniques** to identify at-risk customers, forecast resource allocation, and improve marketing budget effectiveness.
- Collaborated closely with **business stakeholders, analysts, and IT teams** to define KPIs, validate insights, and deliver analytical solutions with **80%+ adoption rates** across departments.
- Cleaned, transformed, and visualized data using **Python, R, Microsoft Excel (Power Query, Pivot Tables)**, improving reporting accuracy and operational efficiency.
- Automated repetitive tasks with **PowerShell scripts** for storage management, URL monitoring, and application development processes, reducing manual intervention and increasing efficiency.

Environment: Windows Server, SQL Server, MySQL, Hadoop Distributed File System (HDFS), PySpark, Apache Kafka, SSIS, Power BI (Desktop, Gateway, Server), Power Query, DAX, Python, R, PowerShell, Microsoft Excel (Power Pivot, Power View, Pivot Tables), Salesforce, Git, JIRA, and Azure Cloud.

CERTIFICATIONS

- AWS Certified Solutions Architect – Associate
- Google Cloud Professional Data Engineer
- Certified Data Management Professional (CDMP)

EDUCATION

- **Master of Science** (M.S), Data Science – **University of New Haven** Aug2023 – May 2025
- **Bachelor of Technology**, Mechanical Engineer – **Sreenidhi Institute of Science and Technology** Aug 2017 – July 2021