A L KAVINDRA SARMA

(6825499919 | aleelakavin04@gmail.com)

Summary:

- I am an Experienced Data Engineer with over 5 years of expertise in designing, building, and optimizing big data solutions, including 3 years specializing in Azure Databricks and 1 year in Microsoft Fabric.
- Developed and optimized end-to-end ETL pipelines to process large-scale data from retail, healthcare, insurance, and power generation systems, ensuring seamless ingestion and transformation.
- Automated data workflows using Azure Data Factory and Databricks, improving data availability, reliability, and scalability across various industries.
- Designed star and snowflake schemas in Snowflake and Azure Synapse Analytics to store and query policy, claim, and sales data, improving performance.
- Built real-time data ingestion systems, enabling up-to-date information for compensation, policy renewals, and turbine operations to prevent inconsistencies.
- Developed advanced data models to support compensation metrics, patient outcomes, fraud detection, and predictive maintenance use cases.
- Implemented CI/CD pipelines for Databricks notebooks, enhancing deployment speed and code quality through automated testing and version control.
- Certified Azure Data Engineer with expertise in designing and executing data engineering solutions using Azure technologies, managing workflows, optimizing performance, and ensuring secure and compliant data handling.
- Optimized Spark jobs and SQL transformations, reducing data latency by up to 40% and achieving significant cost savings on cloud resources.
- Developed real-time dashboards using Power BI and Tableau to track KPIs, compensation payouts, turbine performance, and patient care metrics.
- Conducted root cause analysis to resolve data discrepancies and improve the accuracy of real-time reporting and analytics.

Technical Skills:

Programming Language	Python, SQL, R, C, Java, Scala
Data Visualization	Tableau, Power BI, Advanced Excel
NoSQL Databases	Azure Cosmos DB
Cloud Platforms	Azure Cloud (Data Lake Storage, Blob Storage), AWS Cloud (EC2, S3, Redshift)
Databases	Azure SQL Database, Oracle, MySQL, PostgreSQL, SQL Server
Technologies	Apache Spark ,Azure Databricks ,Azure Data Factory, Azure Functions, Azure Synapse Analytics, Azure Stream Analytics, Azure Purview, Kubernetes
Methodology/Workflow	Agile, Data Quality and Governance, Advance Analytics, Data Mining, Data Visualization, Data Warehousing, Data Transformation
Development Tools	Git, Azure DevOps ,CI/CD workflows
Operating Systems	Windows, Linux, macOS

Education

- **Master of Science in Business Analytics** | University Of New Haven
- **Bachelors of Technology in Electrical and Electronics Engineering |** K L University

Projects:-
Microsoft Fabric-Driven Sentiment Analysis of Bing News Data:

- Developed an end-to-end data pipeline using Microsoft Fabric to ingest and process real-time news data from Bing API, storing raw JSON in OneLake and transforming it into structured Delta tables.
- Implemented sentiment analysis on news articles using Synapse Data Science, leveraging pre-trained machine learning models to classify sentiment as positive, negative, or neutral.
- Created an automated daily pipeline in Data Factory to orchestrate data ingestion, transformation, and analysis, with results visualized in a Power BI dashboard for timely insights on news sentiment trends.

Super Store Analysis using RFM Segmentation:

- Preprocessed and cleaned Superstore dataset using R to ensure data quality.
- Applied RFM (Recency, Frequency, Monetary) segmentation to classify customers and uncover valuable insights.
- Integrated RFM analysis into Power BI, creating interactive dashboards to visualize customer segments and support targeted marketing strategies.

Port Authority Passengers and Bus Departures Prediction Project:

- Forecasted passenger and bus departures using historical travel records, weather data, and economic indicators in Power BI.
- Enhanced prediction accuracy by incorporating temperature and gas price data.
- Developed interactive dashboards for real-time tracking and operational planning, optimizing bus schedules based on predicted demand.

Work Experience:
Client : ProBPM Inc
Role : Sr. Data Engineer
Dec 2023 - Present

- Designed and developed ETL pipelines to integrate and process large-scale sales and compensation data from retail stores across the U.S., ensuring accurate and timely ingestion of both historical and real-time data.
- Implemented data quality checks and cleansing processes to identify and correct anomalies, ensuring data accuracy, consistency, and reliability.
- Utilized Snowflake and Teradata databases to perform complex data processing tasks, enabling efficient querying and reporting for year-over-year sales and monthly targets.
- Built a real-time data ingestion system to capture live sales data, ensuring up-to-date information for compensation calculations and preventing inconsistencies in tracking.
- Created a dynamic target assignment algorithm, automating monthly target calculations for 5,000+ stores, reducing manual effort by 90%.
- Developed advanced data models to compute compensation metrics based on product-level commissions, location-specific variables, and target achievements.
- Designed Power BI dashboards to provide real-time visibility into store performance, target achievements, and sales trends, enabling business teams to track compensation payouts and bonus eligibility.
- Implemented automated bonus tracking mechanisms for stores exceeding 100% of sales targets, ensuring accurate reward calculations and timely disbursements.
- Collaborated with business and operations teams to refine KPIs and optimize bonus and commission structures, improving reporting processes and aligning with business goals.
- Maintained detailed documentation of ETL processes and database schemas to support seamless knowledge transfer and efficient system maintenance.
- Assisted in the migration of on-premises databases to cloud platforms like AWS RDS and Azure SQL, enabling scalable, cost-effective data storage solutions and improving accessibility.

Client: University Of New Haven
Role: Research Assistant
Nov 2022 – Nov 2023

- Developed and maintained data pipelines to automate the ingestion, processing, and integration of student data within the UIS Portal, ensuring seamless operations for international student services.
- Collaborated with educators, administrators, and immigration officers to design data solutions aligned with key requirements, improving the handling of student requests and decision-making processes.
- Implemented a ticketing system within the UIS portal, enabling streamlined tracking and resolution of student issues, including I-20 travel requests, STEM I-20 extensions, visa status updates, and compliance checks.
- Reduced student wait time and improved visibility into case status by building dynamic dashboards that allowed the administration to track issue resolution based on ticket numbering.
- Optimized data workflows to ensure the real-time availability of case status updates, enhancing the student experience and operational efficiency for international student services.
- Ensured data quality and integrity by developing validation checks and monitoring tools, minimizing errors in processing student records and immigration-related requests.
- Presented complex data findings to non-technical stakeholders, such as university administrators and educators, ensuring insights were actionable and easy to understand.
- Enhanced compliance tracking by building automated reports for visa and immigration status, supporting the administration in meeting government regulations and deadlines.
- Collaborated closely with the administration team to improve key processes related to student immigration services, enabling faster resolutions and informed decision-maki

Client: Cognizant Technology Solutions
Role: Data Engineer
Aug 2019 – July 2022

- Designed, developed, and maintained end-to-end ETL pipelines to ingest and process large volumes of structured and unstructured data from policy, claims, and underwriting systems.
- Deployed and managed scalable data workflows using Azure Data Factory and AWS Glue, ensuring high availability and seamless data integration for downstream analytics.
- Designed optimized star and snowflake schemas in Snowflake and Azure Synapse Analytics to store policy, claim, and customer data, improving query performance.
- Implemented incremental data ingestion with Python and SQL, reducing load times and enabling real-time policy renewals and premium calculations.
- Developed automated data quality checks with PySpark to ensure data accuracy, completeness, and compliance with insurance standards.
- Prepared feature-rich datasets for fraud detection ML models, enhancing claim fraud detection efficiency by 20%.
- Collaborated with data governance teams to track metadata and ensure compliance with HIPAA, GDPR, and CCPA regulations.
- Delivered real-time datasets for Power BI dashboards focused on claims processing, renewals, and agent KPIs, enabling real-time performance tracking.
- Optimized SQL queries and transformation jobs, reducing runtimes by 30% to meet operational deadlines and enhance pipeline efficiency.

Client: Lanco Kondapalli Power Ltd, Andhra Pradesh, India
Role: Associate Data Engineer
May 2018 – Aug 2018

- Developed data pipelines for real-time ingestion, processing, and storage of operational data from gas and steam turbines, ensuring seamless data flow for efficient analytics and reporting.
- Automated data ingestion using ETL frameworks, improving the availability and reliability of operational data while reducing manual intervention and operational overhead.
- Optimized data transformation workflows to handle large volumes of turbine data, enhancing system performance and reducing processing latency.
- Implemented stringent data validation checks to ensure the accuracy and reliability of turbine performance

data, enabling better decision-making and operational outcomes.

- Collaborated with engineers and IT teams to integrate operational data from diverse sources, enhancing accessibility and usability for stakeholders across the power generation process.
- Designed and developed interactive Power BI dashboards, providing real-time insights into turbine operations and enabling management to monitor key performance metrics and make data-driven decisions.
- Ensured seamless integration between turbine control systems and analytics platforms, facilitating better tracking, analysis, and optimization of power generation processes.
- Conducted root cause analysis on data discrepancies, improving the accuracy and reliability of real-time reporting and analytics for turbine operations.
- Developed automated data quality monitoring tools to continuously track data integrity, ensuring early detection of issues and maintaining high data standards.
- Integrated real-time data feeds with machine learning models to enhance predictive maintenance and improve performance forecasting for gas and steam turbines.

Achievements:

- Winner of Fall 2023 Academic Project Competition for developing a predictive model for Port Authority Passengers and Bus Departures.
- Published a paper titled “Development of Numerical Relay for Power System Protection Laboratory” in JCR (Journal of Critical Review).