Nandakishor Reddy Pulagam

Senior Data Engineer| Dallas, Texas

215.452.9969| [pnkr.492@gmail.com](mailto:pnkr.492@gmail.com) | [LinkedIn](https://www.linkedin.com/in/kishor-pulagam/)

# Professional Summary

Senior Data Engineer with 10+ years of experience designing end-to-end data solutions for established corporations. Proven ability to deliver value by designing scalable data pipelines, crafting predictive models, and unlocking insights that empower businesses to succeed in today's data-driven world.

# Skills & abilities

* **Programming Languages**: Python, PySpark, SQL, Spark-SQL, Scala, R, PL-SQL, T-SQL
* **Frameworks**: Spark, Spark Structured Streaming, Kafka
* **Cloud Skills**: Azure, Databricks, Azure Synapse, AWS, Microsoft Fabric
* **Databases**: Mongo DB, Azure SQL DB, SQL Server, Oracle, Mongo DB, Data Lakes, Data warehouses
* **Tools**: ADF, Datadog, Airflow, Power BI, Tableau, Power Automate
* **Version Control:** Azure DevOps, Git, Confluence, Jira
* **Professional Experiences**: Team Lead, Data Engineer, Data Scientist, Data Architect, Data Analyst, Scrum Master.

# Experience

## DXC Technology | Senior Professional Data Engineer

## Client: Ahold | Lead Data Engineer Feb 2023 – Present

* Designed and implemented robust data pipelines using Azure Data Factory (ADF) to streamline the ingestion, transformation, and loading (ETL) of prices data, reducing processing time by 30%.
* Developed and optimized data transformation processes using Databricks, ensuring high performance and scalability, resulting in a 20% improvement in processing efficiency.
* Collaborated with cross-functional teams during design meetings to propose innovative solutions for new data requirements, enhancing overall project efficiency and effectiveness.
* Led the operations team by providing technical guidance and support for production environments, ensuring timely resolution of issues and maintaining 99.9% system uptime.
* Implemented best practices for data governance, monitoring, and troubleshooting within the Azure ecosystem, ensuring data integrity and reliability.
* Developed and implemented a data ingestion strategy to support real-time and batch data ingestion using Spark structured streaming, improving timeliness and accuracy of data reporting.
* Implemented the CI-CD pipelines in Azure DevOps to integrate seamlessly with Azure Databricks and ADF, ensuring efficient and reliable deployment of data processing pipelines.
* Explored Microsoft Fabric’s capabilities for end-to-end data pipeline orchestration, integrating with Azure services to assess performance, scalability, and cost-effectiveness.
* Implemented a POC to test Microsoft Fabric’s medallion architecture using Delta Lake, improving data ingestion, processing, and storage efficiency.

## Client: P&G | Senior Data Engineer Mar 2022 – Jan 2023

* Designed and created data ingestion pipelines from various sources, improving data processing efficiency by 30%.
* Developed automated data pipelines using Databricks, enhancing data preparation for AI modeling and reducing manual effort.
* Created and maintained delta tables, scheduling jobs in Databricks to automate data extraction from multiple sources, leading to an improvement in data accuracy and timeliness.
* Implemented the Medallion Architecture to effectively store and manage data from multiple sources, resulting in an increase in data retrieval speed.
* Engaged in comprehensive analysis of existing AI models to understand e-commerce shopper journeys, influencing investments and optimizing ROI for all P&G category products.

## Client: SkyWest Airlines | Senior Data Consultant Jun 2021 – Apr 2022

* Led the development of a predictive maintenance model for an Aviation client, reducing unplanned aircraft downtime by 20%.
* Developed and implemented predictive models to forecast aircraft hydraulic part failures using flight operational and maintenance transaction data, leveraging Azure Databricks.
* Created and optimized machine learning pipelines in Azure Machine Learning Studio, enhancing model training efficiency.
* Designed and maintained data pipelines using Azure Data Factory, streamlining data preparation from multiple sources for predictive modeling.
* Generated insightful dashboards in Databricks for data-driven decision-making, aligning with business needs.
* Collaborated closely with clients to gather requirements and showcased deliverables on time, ensuring project success and client satisfaction.

## DXC Technology | Senior Data Science & Engineering Consultant

## Client: Sanofi | Senior Data Scientist Feb 2021 – Jun 2021

* Implemented data augmentation techniques to generate diverse sample data for effective image labeling in PDF documents using AWS Lambda for scalable and efficient processing.
* Manually labeled images using Labelbox to ensure accurate identification of diagrams, images, graphs, signs, and symbols in PDFs.
* Developed an image recognition model using AWS SageMaker to classify images, significantly reducing manual validation efforts.
* Designed and built data pipelines with AWS Glue to preprocess and integrate image data, facilitating the model training process in AWS SageMaker.
* Utilized AWS S3 for secure and scalable storage of large image datasets, ensuring smooth data flow between preprocessing and model training stages.

## Client: Lufthansa Technik | Senior Data Consultant Jun 2020 – Apr 2021

* Engineered and optimized complex data pipelines using Azure Data Factory to preprocess and integrate maintenance data for predictive modeling, ensuring data integrity and efficiency.
* Designed and led the development of a machine learning model to predict Turnaround Time (TAT) and Defect Estimation (DE) for aircraft maintenance, significantly enhancing the automated quotation process.
* Successfully built and fine-tuned models that outperformed the client's in-house models in terms of Mean Absolute Error, driving substantial improvements in predictive accuracy and reliability.
* Leveraged Azure Machine Learning Service to deploy, run, and monitor machine learning models, ensuring robust, scalable, and maintainable solutions.
* Utilized Azure Databricks for parallel computation and data transformation, enhancing the efficiency and accuracy of data processing and analytics.

## Client: NatWest | Data Engineer Sep 2019 – Jun 2020

* Designed and optimized data pipelines using Azure Data Factory to preprocess and integrate transaction data, ensuring seamless data flow for the AI-powered financial guide application.
* Developed and implemented Natural Language Processing (NLP) algorithms to classify user transaction categories (e.g., shopping, restaurants) based on transaction descriptions.
* Engineered data solutions to analyze user spending patterns and support the AI system in providing personalized financial advice.
* Integrated APIs for real-time data retrieval and processing, enhancing the accuracy and responsiveness of the financial guide application.

## Computer Sciences Corporation | Data Engineer

## Client: Intermountain Healthcare| Database Engineer Jan 2016 – Aug 2019

* Created multiple Oracle Apex applications for a healthcare client and provided database support for different applications.
* Developed and maintained multiple applications using Oracle Apex, streamlining various processes related to hospitals, doctors, and patients.
* Created and optimized PL-SQL packages and triggers, improving the performance of queries for Oracle Apex applications.
* Designed and implemented user interfaces through Oracle Forms and Reports, enhancing user experience and functionality.
* Collaborated with business analysts and functional experts to understand project scope, identify process improvements, and resolve issues, ensuring smooth operation and user satisfaction.
* Provided ongoing support for existing applications, addressing user issues and making necessary enhancements.

## Client: Truven Health Analytics | Database Engineer Aug 2015 – Dec 2015

* Designed, developed, and maintained an Enterprise Data Warehouse using Microsoft SQL Server to support healthcare data management and analytics.
* Created and managed staging (STG) and Operational Data Store (ODS) environments to ensure data integrity, consistency, and accessibility.
* Optimized database performance by implementing indexing, partitioning, and query optimization techniques, resulting in faster data retrieval and processing times.
* Experience in driving architectural discussions, making design decisions, and providing technical guidance to ensure alignment with business goals and industry best practices.
* Collaborated with cross-functional teams to design the databases and enhance the overall system performance and data retrieval efficiency.

# Education

## Drexel University, Philadelphia | MS Data Science | CGPA-3.9 June 2021

* Student leader in Drexel Graduate IEEE and Pragati Associations.
* Member of Drexel AI, Graduate Student Association, Drexel Blockchain, Drexel International Student Association, and Dragons Teach Students Teaching Engaged Minds.

# Certifications

* Databricks Certified Data Engineer Professional.
* Microsoft Certified: Azure Solutions Architect Expert
* Azure Data Engineer Associate certified by Microsoft.
* Certified in Databricks Fundamentals
* Certified in Kafka Fundamentals
* Certified in Azure AI, Data Fundamentals
* Google-certified Data Analytics Professional
* Certified in MongoDB Basis and MongoDB for SQL Pros from MongoDB
* IBM Certified in Machine Learning with Python and Big Data

# Achievements

* Selected by DXC Technology to participate in their prestigious master's sponsorship program in the USA, earning a master's degree in data science debt-free.
* Instructed and developed the skills of over 250 employees in data engineering and data science through immersive boot camp training.