A blue hexagon with white text

AI-generated content may be incorrect.A logo for a company

AI-generated content may be incorrect.A logo for a company

AI-generated content may be incorrect.

# SUMMARY AKSHAY KAILASA

# Texas, USA | +1(940) 758-1526 | [akshaykailasa@outlook.com](mailto:akshaykailasa@outlook.com)

**Data Engineer** with 5+ years of experience designing and optimizing scalable ETL pipelines using Apache Spark, Python, and SQL. Skilled in Azure and AWS ecosystems (ADF, Databricks, Glue, Redshift, DMS) with proven ability to integrate multi-source data, enhance data quality, and improve processing efficiency by up to 40%. Adept at cloud migration, Delta Lake architecture, and predictive analytics for business-critical insights.

# EDUCATION

## University of North Texas, Denton, Texas May 2023

Master of Science in Computer Science

## MVSR Engineering College, Hyderabad, India May 2018

Bachelor of Engineering (B.Tech.) in Information Technology

# PROFESSIONAL EXPERIENCE

## Deloitte, USA | Data Engineer May 2023 – Present

* Designed and maintained scalable ETL pipelines using Apache Spark, Python, and SQL. Conducted performance tuning for SQL databases, and Azure Databricks, cutting query response times by 30%.
* Built **Delta Lake** architecture with ACID compliance for claims and patient data, enabling reproducible analytics and data versioning
* Designed and developed scalable **ETL pipelines in Azure Data Factory and Azure Databricks** to process healthcare claims data from multiple providers and third-party sources, ensuring timely and accurate data availability for analytics and reporting.
* Implemented workflows with Azure Data Factory, and Azure Databricks, optimizing data transformation and reducing processing time by 40%.
* Developed predictive maintenance algorithms using Python and Databricks, resulting in a 20% reduction in equipment downtime and maintenance costs.
* Built robust **Delta Lake architectures on Azure Data Lake Storage (ADLS)** for storing and versioning large volumes of claims and patient data with ACID compliance..
* Participated in data quality checks and validation frameworks, ensuring accuracy and completeness of member eligibility, claims status, and provider payment data.

## Nstarx.inc, India | Data Engineer Nov 2018 - Dec 2021

* Credential My Doc is a one-stop Medical Staff and Provider Enrolment pre-populated form creation and process management tool.
* Worked on Custom Webcrawlers to get the data from different license verification site which led to reduction in time by 40%.
* Orchestrated and developed ETL pipelines utilizing AWS Glue for automated data cataloging. Implemented Python in the backend for data loading downstream, resulting in a 20% reduction in report retrieval time.
* Applied transformations including deduplication and Change Data Capture using Python and Lambda functions on usage data extracted from Athena. Ensured accurate data loads to AWS S3 buckets and AWS Redshift.
* Developed stored procedures, functions, and views in SQL Server to execute complex calculations for healthcare applications, such as total billing per provider, per location, and per payer.
* Enhanced ETL solutions in SSIS, decreasing execution time for over 100+ providers by 20%. Achieved optimization in data retrieval and processing using transformations such as CASE, UNION, complex joins, CTEs, and Derived Columns.
* Experienced in integrating various data sources and systems to enable seamless data flow and processing across the organization.
* Worked on a Migration project to migrate SQL Server to AWS cloud using AWS (DMS, Lambda, Step Function, Glue, S3).
* Completed transfer legacy process of Feed process to the cloud using Python and AWS hence reducing feed process by 30%. Built ETL pipelines for pre-processing and transformations using AWS Glue jobs.

# CERTIFICATION

* [Databricks Certified Data Engineer Professional](https://credentials.databricks.com/cf9e5160-06cb-4699-8503-884f82565aab#gs.11e14i) Academy Accreditation.
* [AWS Certified Solutions Architect](https://www.credly.com/badges/48b6da48-3208-48b6-8164-8c7b0dfb1844/linked_in?t=ry08gv) - Associate.
* [Databricks Data Engineering Associate](https://credentials.databricks.com/6c4abb43-feed-4eb3-a830-823024ac35e2?record_view=true&gs.2p3jcq) Academy Accreditation.
* [Snowflake Hands on Essentials Data Warehousing](https://achieve.snowflake.com/c97ac5ca-16ff-4ab6-9021-3bdf8bb090c2#gs.1ws1w4) - Accreditation.
* [Astronomer Certification for Apache Airflow Fundamentals](https://www.credly.com/badges/12b268f9-6b37-4b73-9683-88d1c9e379c9)

# TECHNICAL SKILLS

**Methodologies:** SDLC, Agile/ Scrum, Waterfall

**Language:** Python, SQL, R, SCALA, YAML

**Python Packages**: Pandas, NumPy, Matplotlib, SciPy, Scikit-Learn, SeaBorn, PyTorch. TensorFlow, ggplot2, Plotly, Keras

**Data Analytics Skills:** Data Manipulation, Data Cleaning, Data Visualization, Exploratory Data Analysis, Data Analysis

**Data Components:** HDFS, Hue, MapReduce, PIG, Hive, HCatalog, HBase, Sqoop, Impala, Kimball Modeling,

DBT(Models, Seeds, Generic and custom tests, Documentation, Analyses, Snapsots, Macros)

**Other Technical Skills:** AWS, AZURE, Databricks, NLP, A/B Testing, Hypothesis testing, ETL, Databricks, Hadoop, Spark, Snowflake, Big Query, Apache Airflow, Critical Thinking, Communication Skills, Presentation Skills, Problem-Solving

**Databases:** MySQL, MS SQL Server, ETL SSRS, SSIS, SSAS.

**Tools:** Tableau, Power BI, Advanced Excel, MS Office Suite (Excel, Visio, Access, Word, PowerPoint), Visual Studio, GIT, Jupyter Notebook

**Version Control:** Git, GitHub