Akshay Kailasa | Data Engineer

Irving, TEXAS **|** +1 (940)758-1526 | akshay.kailasa@aogjob.com| LinkedIn

SUMMARY

A seasoned Data Engineer with over 4+ years of experience specializing in the design, development, and optimization of complex data systems. Spearheaded initiatives to enhance data architecture, significantly improving data accessibility, storage, and processing times by implementing robust data pipelines using SQL and big data technologies

Previously at Nstrarx.inc in India, I excelled in cloud data engineering, developing and managing large-scale ETL solutions and real-time data processing platforms using a variety of technologies including AWS services, Spark, Flink, Kafka, and Python. Throughout my career, I have demonstrated a strong ability to work collaboratively with data science teams and analysts, providing the infrastructure and clean data sets necessary for advanced analytics and machine learning projects

TECHNICAL SKILLS

Language: Python, SQL, Java, C++, Scala.

Python Packages: Pandas, NumPy, MatplotLib, SciPy, Scikit-Learn, SeaBorn, PyTorch, TensorFlow, ggplot2, Plotly, Keras, LangChain

Data Analytics Skills: Data Manipulation, Predictive Analysis, Data Cleaning, Data Mining, Data Visualization, Statistical Modelling, EDA, ETL, Data Modelling.

Databases: MySQL, PostgreSQL, MongoDB, MSSQL, Oracle SQL, Snowflake, Hive, DynamoDB

Tools: Tableau, Power BI, ArcGIS, MongoDB, Alteryx, Neo4j, Django, OpenAI, AI Advanced Excel (Pivot Tables, VLOOKUP), Talend, Microsoft office Suite (Word, PowerPoint, Excel, Outlook), Jira, Looker

Cloud Technologies: Azure (Data Bricks), AWS (EC2, S3, Redshift, Step function, AWS Athena, EMR, Lambda function, RDS, SQS, ECS, Glue, DynamoDB, AWS API, Airflow), DBT

Data Components: HDFS, Hue, MapReduce, PIG, Hive, HCatalog, HBase, Sqoop, Impala, Zookeeper, Flume, Kafka, Yarn, Cloudera Manager, Kerberos, Pyspark Airflow, Kafka Snowflake

Spark Components: Apache Spark, Data Frames, Spark SQL, Spark, YARN, Pair RDDS

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 in Information Technology	

PROFESSIONAL EXPERIENCE

Accurate Healthcare, USA | Data Engineer   **Jan 2023 - Present**

- Developed and maintained robust data management systems adhering to HIPAA regulations, focusing on the integrity and security of sensitive patient data. Implemented comprehensive data encryption and access controls, significantly reducing the risk of data breaches using Snowflake's secure data sharing capabilities.
- Engineered seamless integrations between clinical healthcare systems and billing platforms using API-driven approaches with Azure Databricks, optimizing existing data pipelines and reducing data processing times by 30%, thereby increasing efficiency in billing operations and patient data management.
- Leveraged Azure Databricks for developing advanced data analytics tools that utilize statistical models and machine learning algorithms to facilitate data-driven decision-making, enhancing patient care and operational efficiency by providing insights into treatment outcomes and resource allocation.
- Utilized Snowflake's scalable architecture to prepare clean, well-structured datasets for use in machine learning models, focusing on predicting patient outcomes and identifying billing anomalies. Collaborated with data scientists to refine data features and improve model accuracy.
- Developed complex SQL queries on Snowflake to extract, transform, and load (ETL) data, leading to a 40% improvement in data retrieval speed and accuracy for business intelligence reporting.
- Authored detailed technical documentation for all data processes and systems, ensuring clarity and usability across multiple departments. Established guidelines and practices that comply with healthcare regulations, enhancing organizational adherence to legal and ethical standards.
- Worked closely with IT specialists and healthcare professionals to develop and customize data solutions that meet specific organizational needs using Azure Databricks and Snowflake. Improved patient service delivery by 25% through tailored data applications that enhance user experience and operational workflows.

Nstarx.inc, India | Cloud Data Engineer   **Aug 2018 - Dec 2021**

- Developed Spark jobs in Spark to perform ETL from SQL Server to Hadoop.
- Experienced in integrating various data sources and systems to enable seamless data flow and processing across the organization.
- Developing a data normalization and consolidation process using AWS Glue, reducing data redundancy by 40% and improving data quality scores by 50%.
- Leveraged Spark to accelerate data processing, achieving a 25% increase in processing speed for daily batches of up to 50GB, enhancing the team's data analysis capability.
- Collaborated on projects requiring a deep understanding of structured and unstructured data.
- Worked on a Migration project to migrate SQL Server to AWS cloud using AWS (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 job.
- Implemented a real-time data processing platform using Flink and Kafka that processed millions of data per day.
- Normalized and aggregated data from various sources and implemented complex data transformations using Scala and Flink APIs.
- Created AWS API using AWS lambda (Python) to retrieve cloud data and use it in Reports.
- Implemented and maintained AWS infrastructure using Terraform, resulting in significant cost savings and improved performance.
- Created SSIS packages to extract data from OLTP to OLAP systems with transformations and scheduled Jobs to run the packages daily.
- Developed and maintained SSIS jobs, SSRS reports, data modelling & visualization using Power BI
- Credential My Doc is a one-stop Medical Staff and Provider Enrolment pre-populated form creation and process management tool.
- Used Python Django Framework for Data scraping and saving the data in POSTGRESQL Database.
- Worked on POSTGRESQL, Python (Django framework) selenium web driver. Implemented XPATHS according to business logic.