

Akshat Dev

+918603293842 | dakshat75@gmail.com | [LinkedIn](#) | [GitHub](#)

Profile Summary

Data Engineer with hands-on experience in Ab Initio, Apache Spark, SQL, and Unix/Linux, building and supporting scalable ETL pipelines across banking, telecom, and EdTech domains. Strong in ETL development, performance tuning, data validation, and L2 production support, with growing expertise in Databricks and Snowflake.

Technical Skills

Languages: SQL, Python, PySpark

ETL/ELT: Ab Initio, Apache Spark, Azure Databricks

Visualization Tools: Power Bi

Database: SQL Server, MySQL

Data Warehouse: Snowflake

ITSM & Collaboration Tools: JIRA, Confluence, ServiceNow

Experience

BxData Tech Solutions

April 2025 – December 2025

Junior Data Engineer – Fulltime (On-site)

Gurugram, India

- Designed and developed batch and event-driven ETL pipelines using Ab Initio GDE and Co>Operating System to process data from secure Linux servers into Hive and Hadoop environments.
- Implemented SCD Type 2 logic, data validation, and HDT separation using core Ab Initio components such as FBE, Reformat, Rollup, Scan, Join, Partition, and Redefine to ensure accurate historical tracking and high data quality.
- Built scalable data processing workflows for large-volume batch loads and event-triggered jobs, optimizing performance through effective partitioning and parallel processing techniques.
- Monitored and supported enterprise-scale Ab Initio ETL pipelines using Control Center across multiple environments, ensuring high availability and operational stability.
- Collaborated with upstream and downstream teams to manage source ingestion, troubleshoot job failures, and ensure reliable delivery of processed data to Hive and Hadoop systems.
- Provided L2 production support, including job monitoring, failure analysis, SLA-driven incident resolution, and controlled reruns.
- Executed SQL-based data validation and reconciliation across Oracle databases and Hadoop ecosystems to maintain data integrity.
- Managed Linux/Unix environments, performing system health checks, log analysis, and proactive performance tuning to prevent pipeline disruptions.

Technical Skills: Ab Initio GDE, Ab Initio Control Center, Apache Spark, Yarn, UNIX/Linux, SQL, Hive, Hadoop, JIRA, Confluence, ServiceNow, Grafana

Pass 11 Plus (EdTech Platform) under JMS Provider Ltd

February 2023 – March 2025

Data Analyst – Contractual (Remote)

Birmingham, UK

- Owned end-to-end analytics for 2,000+ students across multiple tuition centres, using SQL, Excel, and Power BI.
- Designed and optimized complex SQL queries (CTEs, joins, window functions) for performance tracking and KPI analysis.
- Developed and maintained SQL stored procedures to upsert student records, ensuring accurate and consistent data management across academic and administrative systems.
- Implemented soft delete logic to manage inactive student records while preserving historical data for reporting and compliance purposes.
- Developed interactive Power BI dashboards for student performance, tutor effectiveness, and centre-level metrics.
- Reduced report delivery time by 40% by automating recurring SQL and Excel-based reports.
- Implemented data quality checks and validation rules, maintaining 99% accuracy across all reporting layers.
- Delivered ad-hoc insights and performance summaries to management and teaching staff.

Technical Skills: SQL, Power BI, Advanced Excel (Power Query, Pivot Tables, XLOOKUP), Data Modeling, KPI Dashboards

Personal Project

Azure Databricks Lakehouse with Star Schema & SCD Implementation

Tech Stack: Azure Databricks, ADLS Gen2, Unity Catalog, PySpark, Spark Structured Streaming, Delta Lake, SQL

GitHub: <https://github.com/AkshatDev2002/azure-databricks-lakehouse-data-engineering>

- Designed and deployed an end-to-end Azure Databricks Lakehouse using ADLS Gen2 and Unity Catalog, implementing Medallion Architecture (Source, Bronze, Silver, Gold) with governed catalogs, schemas, and external locations.
- Modeled curated data using a Star Schema, building fact tables (Orders) and dimension tables (Customers, Products, Regions) to enable analytical querying and BI-ready datasets.
- Built a parameterized bronze ingestion framework using Spark Structured Streaming (Auto Loader) with widgets, checkpointing, and fault-tolerant ingestion from Azure Data Lake.
- Developed Silver-layer transformation pipelines in PySpark, applying data cleansing, enrichment, window functions, OOP-based reusable logic, and custom SQL/Python UDFs.
- Implemented Slowly Changing Dimensions with SCD Type 1 (Customers) and SCD Type 2 (Products) and delivered Delta tables optimized for governed analytics and reporting.

Courses and Certifications

- Fundamentals of Azure Databricks
- PYSPARK End to End Developer Course (Spark with Python)
- Fundamentals of Docker & Kubernetes
- ETL and ELT in PostgreSQL

Education

Central Institute of Technology

Bachelor of Technology in Computer Science and Engineering

September 2020 – June 2023

Kokrajhar, Assam

Central Institute of Technology

Diploma in Computer Science and Engineering

August 2017 – July 2020

Kokrajhar, Assam