

Akshay Kohchade

Data Engineer | Data Analyst

ETL & PySpark | Python | SQL | AWS Glue | Redshift |
Databricks | Power BI | Snowflake |

+91 8459348473

Aurangabad, Maharashtra

akshaykohchade1@gmail.com

WORK EXPERIENCE

Data Engineer Intern – PST IT Solutions | Aurangabad, Maharashtra

May 2025 – Nov 2025

- Worked on healthcare datasets to support both **analytical reporting** and **data pipeline** requirements.
- Performed data cleaning, transformation, and validation using **PySpark** and **SQL** to ensure data quality for analysis.
- Assisted in building and maintaining **ETL workflows** in **AWS Glue** and **Databricks** for multi-source healthcare data.
- Optimized Spark queries by improving joins and partitioning, leading to better performance on large datasets.
- Prepared transformed datasets in analytics-ready format for **reporting and dashboarding**.
- Supported **data analysis activities** by extracting insights from curated datasets using **SQL** and **Power BI** tables.
- Collaborated with senior team members to debug ETL jobs and improve pipeline reliability.
- Maintained documentation for data flows, transformations, and dataset versions using **Delta tables**.
- Gained experience in writing **SQL** queries to extract, transform, and validate data stored in different sources.

SKILLS

- Languages:** Python, SQL
- Big Data & Frameworks:** Apache Spark, PySpark, Databricks
- AWS Services:** AWS Glue, Amazon S3, Amazon Redshift
- Azure Services:** Azure Data Factory, Azure Databricks
- Other:** MySQL, Oracle Form, Oracle Report, Power BI

PROJECT

Cloud-Native Data Orchestration Pipeline for Healthcare Systems (AWS and Databricks)

- Consolidated fragmented healthcare data from multiple hospital systems into a **centralized cloud data platform** to support both analytics and reporting.
- Designed and implemented ETL pipelines using **AWS Glue** and **Databricks** to ingest, transform, and store patient data.
- Performed data cleaning, validation, and standardization using **PySpark** and **SQL** to ensure high data accuracy.
- Built **analytics-ready datasets** enabling faster access for reporting, trend analysis, and decision-making.
- Optimized Spark transformations and joins to improve **pipeline performance and data processing efficiency**.
- Supported **data analysis and reporting needs** by delivering reliable, structured datasets for dashboards and ad-hoc analysis.
- Helped identify **early healthcare trends and operational insights** through curated and validated datasets.
- Improved reporting reliability and turnaround time by maintaining **consistent schemas and trusted data sources**.
- Ensured datasets were compliant and traceable, supporting healthcare **regulatory and operational requirements**.
- Helped improve overall healthcare outcomes by supporting **accurate, timely, and insight-driven reporting**.
- Enabled better data-driven decisions by bridging **data engineering pipelines with analytical use cases**.

CERTIFICATIONS

Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate

- Demonstrated foundational understanding of Oracle Cloud Infrastructure and AI concepts for building data-driven solutions.
- Oracle | Issued November 2025 · [Verify Credential](#)*

Oracle Data Platform 2025 Certified Foundations Associate

- Demonstrated foundational knowledge of Oracle Data Platform and OCI services for efficient and secure data management.
- Oracle | Issued November 2025 · [Verify Credential](#)*

EDUCATION

Bachelor of Science (B.Sc.)

August 2020 – June 2023

Percentage: 64.78%

Dr. Babasaheb Ambedkar Marathwada University

Aurangabad, Maharashtra