

Oleh M.

Senior Data Engineer

SUMMARY

- 10+ years of ETL and data integration experience with batch and streaming integrations.
- 7 years of Databricks: Lakehouse, PySpark, Delta Lake, Unity Catalog, Workflows, batch/streaming pipelines, data processing and performance optimization.
- Power BI Pro (Desktop/Service) and Power Query experience, DAX knowledge.
- Skilled in SQL with Azure SQL Database/SQL Server, including data modeling and query optimization.
- Experienced with ADLS, Parquet (partitioning, compression, storage/read optimization), Azure Synapse Analytics/Notebooks, and Azure Data Factory, including building and maintaining data pipelines.

TECHNICAL SKILLS

Main Technical Skills	Python, SQL, Databricks, Bash
AI Tools & Assistants	ChatGPT
Programming Languages	Go, Java, Python, Scala
AI & Machine Learning	AI, AWS SageMaker, Azure AI Search, GenAI, Kubeflow, LangChain, LLM, Machine Learning
Java Frameworks	Apache Spark, Hibernate, Spring
Scala Frameworks	Apache Spark
Python Libraries and Tools	Dask, Pandas, PySpark
Java Libraries and Tools	JSON
Data Analysis and Visualization Technologies	Apache Airflow, Apache Hive, Apache Nifi, Apache Pig, Apache Spark, Apache Spark Streaming, AWS Athena, Azure Data Factory, Big data, Business Intelligence (BI), Databricks, Data Quality, ETL, HBase, Jupyter Notebook, Lakehouse, Microsoft Azure Synapse Analytics, Oozie, Pandas, Power BI, Sqoop, Teradata
Databases & Management Systems / ORM	Apache Hive, Apache Spark, Apache Spark Streaming, AVRO, AWS Redshift, Cassandra, Data Warehouse, Elasticsearch, Hibernate, Oracle Database, Oracle SQL, PL/SQL, SQL
UI Frameworks, Libraries, and Browsers	Phoenix
	AWS, Azure, GCP

Cloud Platforms, Services & Computing	
Amazon Web Services	AWS Athena, AWS Bedrock, AWS Elastic Kubernetes Service, AWS EMR, AWS Lambda, AWS Redshift, AWS S3, AWS SageMaker
Azure Cloud Services	Azure AI Search, Azure Blob Storage, Azure Cloud Functions, Azure Data Factory, Azure Data Lake, Azure Eventhub, Azure Event Hubs, Azure Machine Learning, Azure Search, Azure Services, Databricks, Microsoft Azure Synapse Analytics
QA, Test Automation, Security	Apache Maven
SDK / API and Integrations	API
Collaboration, Task & Issue Tracking	Atlassian Confluence
Scripting and Command Line Interfaces	Bash
Deployment, CI/CD & Administration	CI/CD, Helm Charts
File Systems, Storage	CSV, Parquet, XML
Virtualization, Containers and Orchestration	Docker, Kubernetes, Nomad
Platforms	IOT, Unity
Message/Queue/Task Brokers	Kafka
iOS Libraries and Tools	MDM
Methodologies, Paradigms and Patterns	REST, UML
Other Technical Skills	Delta lake, ESC, Hortonworks, Iceberg, Microsoft Fabric, ORC

WORK EXPERIENCE

Senior Data Engineer/Solution Architect, NDA

Duration: Mar. 2026 - Sep. 2026

Summary: Microsoft Fabric MDM and Data Warehouse Project.

Responsibilities:

- Analysis of customer and location data from multiple enterprise source systems.
- Design of an MDM solution and unified data model in Microsoft Fabric.
- Development and configuration of Fabric Warehouse as an analytical data layer.
- Design of data ingestion, transformation and integration processes.



- Integration of Power BI with Fabric Warehouse for data analysis, validation and reporting.
- Evaluation of MDM platforms and preparation of the target architecture and implementation roadmap.

Achievements:

- Designed the target MDM and Data Warehouse architecture in Microsoft Fabric.
- Prepared recommendations for creating unified golden records and improving data quality.
- Defined the integration approach between source systems, MDM services, Fabric Warehouse, and Power BI.

Technologies: Power BI, MDM, Microsoft Fabric, Data Warehouse.

Senior Data Engineer/Data Architect (Consultant), Edvantis (LLM-Powered Data Platform for Document & Image Processing)

Duration: 2025 - Mar. 2026

Summary: Designed and built a scalable Databricks data platform for processing imagebased documents. Integrated AWS Bedrock LLM services for text and structured data extraction, orchestrated ETL workflows with Databricks Jobs and Tasks, governed data through Unity Catalog, and delivered Gold-layer results to downstream systems.

Responsibilities:

- **Data Platform Architecture:** Designed and built a scalable Databricks platform for processing image-based 4 documents and managing end-to-end data workflows.
- **LLM-Based Document Processing:** Integrated AWS Bedrock LLM services to extract text and structured data from image documents.
- **ETL Orchestration:** Developed ingestion, transformation, enrichment, and publishing pipelines using Databricks Jobs and Tasks.
- **Data Governance and Storage:** Used Unity Catalog and Delta Lake to organize, secure, store, and process data within the Databricks platform.
- **Downstream Data Delivery:** Exported Gold-layer datasets to MS SQL Server and Elasticsearch for operational and search use cases.
- **External Workload Integration:** Integrated LLM-generated outputs produced by the Databricks platform with Nomad and AWS EKS jobs.

Technologies: Databricks, PySpark, Delta Lake, Unity Catalog, AWS Bedrock, LLM, Databricks Jobs and Tasks, MS SQL Server, Elasticsearch, Nomad, AWS EKS, Docker, Kubernetes, REST APIs.

Architect, 1648 Factory (Generative AI)

Duration: 2023 - 2025

Summary: Processed various document types and integrated them into the Azure Search AI database, while developing a modularized ETL framework and managing the Databricks platform.

Responsibilities:

- **Document Processing:** Processed different document types such as PDF, DOCX, PPTX, etc., for integration into the Azure Search AI database.
- **ETL Framework Development:** Developed a modularized framework for processing ETL tasks, ensuring scalability and maintainability.



- **Databricks Platform Management:** Managed the Databricks platform, including creating clusters, setting up access controls, and consulting teams on Spark and Databricks best practices.
- **Data Pipeline Design and Implementation:** Designed and implemented data pipelines to process and transform large datasets efficiently.
- **Azure Data Lake Utilization:** Utilized Azure Data Lake for scalable storage solutions and efficient data retrieval.
- **PySpark Job Development:** Developed and optimized PySpark jobs for data processing and analytics tasks.
- **Documentation:** Created and maintained comprehensive documentation for data processing workflows and best practices.
- **Collaboration:** Collaborated with cross-functional teams to ensure data quality and integrity throughout the project lifecycle.
- **Monitoring and Alerting:** Implemented monitoring and alerting mechanisms for data pipelines to ensure reliability and performance.

Technologies: PySpark, Databricks, Delta Lake, Azure AI Search, Azure Blob Storage, Azure services.

Architect, Opinov8 (Pharmacy event-sourcing ETL)

Duration: 2021 - 2023

Summary: Designed and implemented a robust event-sourcing architecture and data processing for a pharmacy application using PySpark, Databricks, Lakehouse, Delta Lake, Kafka, Azure Blob Storage, and other Azure services.

Responsibilities:

- Designed a robust event-sourcing architecture to handle real-time data processing and analytics for the pharmacy application.
- Developed a scalable and efficient data processing framework using PySpark and Databricks.
- Implemented data pipelines to ingest, process, and store large volumes of data using Delta Lake and Lakehouse architecture.
- Ensured data integrity and consistency across different stages of the data pipeline.
- Integrated various Azure services such as Azure Blob Storage for scalable storage solutions.
- Utilized Azure Data Lake for efficient data retrieval and processing.
- Implemented event streaming using Kafka to handle real-time data ingestion and processing.
- Ensured reliable and low-latency data streaming for the pharmacy application.
- Optimized PySpark jobs for efficient data processing and analytics.
- Tuned the performance of data pipelines to ensure minimal latency and high throughput.
- Implemented security best practices to ensure data privacy and compliance with industry standards.
- Managed access controls and data encryption to protect sensitive information.
- Collaborated with cross-functional teams to gather requirements and ensure alignment with business objectives.
- Provided consultation and guidance on best practices for using Databricks, PySpark, and other Azure services.
- Implemented monitoring and alerting mechanisms to ensure the reliability and performance of data pipelines.
- Conducted regular maintenance and updates to keep the data processing framework up-to-date.



- Created and maintained comprehensive documentation for the event-sourcing architecture and data processing workflows.
- Documented best practices and guidelines for future reference and knowledge sharing.

Technologies: Python (PySpark), Databricks, Delta Lake, Lakehouse, Kafka, Azure Blob Storage, Azure Data Lake, CSV, JSON, Parquet, Delta.

Architect/Big Data Engineer/MLOps, GlobalLogic

Duration: 2017 - 2021

MLOps, Programming advertising

Duration: 2020 - 2021

Summary: Developed and implemented machine learning infrastructure for an Adserver in AWS, focusing on ETL, data analytics, machine learning, and model inference.

Responsibilities:

- Worked closely with the Data Science team to develop infrastructure for ETL, data analytics, machine learning, and model inference.
- Set up MLOps infrastructure in Kubernetes using Helm charts for Airflow, Dask, and JupyterHub.
- Developed REST API for model inference using Python Async (ASGI) Web Framework.
- Developed ETL processes for machine learning.
- Integrated machine learning models into Airflow and Dask on Kubernetes.
- Explored AWS SageMaker and Kubeflow for potential integration and optimization.
- Documented MLOps processes and best practices in Confluence.

Technologies: Scala, Python, Golang, Airflow, Spark, Dask, Kubernetes, SageMaker, AWS, Google Cloud, Jupyter, JupyterHub.

Big Data Engineer, IoT

Duration: 2018 - 2020

Summary: Machine learning for IoT in Azure. Collaborated with two Data Science teams to set up infrastructure for ETL, data analytics, and machine learning using Databricks, Azure Synapse, Azure Machine Learning, and Azure Data Factory.

Responsibilities:

- Set up environment for Data Engineering and Data Science.
- Developed jobs and pipelines using Scala, Python, SQL, Spark, and PySpark.
- Developed and deployed Spark jobs for Data Science teams in Databricks, Azure Synapse, and Azure Data Factory.
- Assisted Data Science teams with Scala, Spark, DevOps, and MLOps.
- Documented MLOps processes in Confluence.

Technologies: Scala, Python, SQL, Parquet, Spark, Databricks, Azure, Delta Lake, Azure DataFactory, Synapse



Architect, Retail

Duration: 2017 - 2018

Summary: Big Data Integration for a retail chain on Azure. Designed Spark Streaming integration for event sourcing with Azure Event Hubs, using Spark Operator on Kubernetes and Delta Lake for data storage.

Responsibilities:

- Designed ETL integration for event sourcing.
- Developed and deployed job managers and jobs for various topics in Azure Event Hubs.
- Documented the solution in Confluence.

Technologies: Python, SQL, Spark Streaming, Delta Lake, Kubernetes, Azure Delta Lake, Azure Function, Azure EventHub, Event sourcing.

Big Data Engineer, Agilelab (Energy)

Duration: 2016 - 2017

Summary: Big Data Integration for finance company in HortonWorks. Migrated data from Oracle database to HortonWorks using Spark and Spring Framework, converting Oracle PL/SQL to Spark jobs and serving them via REST API.

Responsibilities:

- Converted PL/SQL to Spark jobs and integrated jobs in Spring Framework.
- Scheduled Oozie jobs for workflow management.
- Mapped Oracle database with Hive External Table using Oracle Big Data Integration Library.

Technologies: Java, Maven, Spring, Hibernate, Spark, Oracle SQL, PL/SQL, Hortonworks, Hive, Oozie.

Big Data Engineer, EPAM

Duration: 2014 - 2016

Big Data Engineer, Healthcare

Duration: 2015 - 2016

Summary: File processing for non-profit healthcare organization on AWS. Set up file processing with AWS Lambda, Parquet, Amazon Athena, and Redshift. Developed file parsers for different file formats and aligned schemas.

Responsibilities:

- Implemented ETL file processing workflows.
- Developed parsers for various file formats and aligned data schemas.
- Documented processes in Confluence.

Technologies: Python, SQL, CSV, Parquet, Pandas, Spark, AWS S3, Lambda, ESC, Athena.



Big Data Engineer, Retail

Duration: 2014 - 2015

Summary: Migrated Teradata data to Hive in AWS EMR. Worked on data exploration on Teradata, wrote jobs to efficiently read data from Teradata and store in proper schema in S3.

Responsibilities:

- Developed Airflow DAGs and Spark jobs for data migration and processing.

Technologies: Scala, Python, Spark, Airflow, Teradata, AWS S3, EMR, Hive.

EDUCATION

- **Lviv Polytechnic University (LPU)**

Equipment of the electronic industry, Major: Engineering Mechanics

Sep. 2002 - Dec. 2008

- **Lviv Institute of Banking (LBI)**

Banking, Major: Economics and Entrepreneurship

Sep. 2009 - Dec. 2011

