

Olena R

Expert Senior Data Engineer with SAP BW and Databricks

SUMMARY

- 15+ years with business analysis and data engineering.
- Data engineer on Databricks + Delta Lake daily: PySpark, Spark SQL, medallion architecture, job tuning, and cost control. Strength is reverse-engineering undocumented legacy warehouses (incl. SAP BW concepts) and reproducing the logic as maintained models and pipelines, with source-to-target reconciliation as the acceptance gate.
- Migration: source cutovers, six-provider schema unification into one canonical model, historical backfills and restatements. Dimensional Gold-Silver-Bronze model and the semantic layer. AWS Flue, MS Fabric, Power BI.
- Domains: Fintech, payments, AML and regulatory reporting; US retirement & wealth; BMW.

TECHNICAL SKILLS

Main Technical Skills	Databricks, SAP BW, PySpark, SQL, Power BI
Programming Languages	Python
Java Frameworks	Apache Spark
Scala Frameworks	Apache Spark
Python Libraries and Tools	Pandas, PySpark
Data Analysis and Visualization Technologies	Apache Airflow, Apache Spark, AWS Athena, Azure Data Factory, Business Analysis, Databricks, Data Quality, ETL, Lakehouse, Pandas, Power BI
Databases & Management Systems / ORM	Apache Spark, Data Warehouse, dbt, Microsoft Azure SQL Server, Microsoft SQL Server, Oracle Database, PostgreSQL, Spark SQL, SQL
Cloud Platforms, Services & Computing	AWS, Azure
Amazon Web Services	AWS Athena, AWS S3
Azure Cloud Services	Azure Data Factory, Databricks, Microsoft Azure SQL Server
Industry Domain Experience	Fintech, ISO
Deployment, CI/CD & Administration	CI/CD

Platforms	CRM
Project Management & Administration	Data Models
Version Control	Git
Hardware	MS
Scripting and Command Line Interfaces	Regex
Methodologies, Paradigms and Patterns	REST
SDK / API and Integrations	RESTful API, Stripe, System integration
Other Technical Skills	DATA ANALYST, Data Model, Delta lake, English, ETL/ELT, Factory, Financial Services, iGaming, Lending, Microsoft Fabric, Monitoring, Product Owner, Provider, Schema, SPSS, Technical Writer

CORE TECHNICAL SKILLS

Databricks & Lakehouse

Production hands-on: PySpark and Spark SQL, Delta Lake medallion architecture (Bronze / Silver / Gold), Databricks Jobs, notebooks and cluster sizing; partitioning, join strategy, caching and broadcast joins tuned against real workloads for both runtime and cost. Microsoft Fabric (OneLake, Lakehouse, Warehouse, Data Factory, Direct Lake) equally in production.

Data Migration

Source-system migration and cutover as delivered work: scoping what a legacy model actually computes, target model design, transfer and transformation of business logic, historical backfill and restatement, and reconciliation batteries (control totals, record-level diffs) that prove source-to-target parity before sign-off.

SAP BW & Source Systems

Familiar with the SAP product line and BW concepts: InfoObjects, InfoProviders / DSO, star-schema extracts and transformation logic, but no hands-on BW development; stated plainly. What transfers is the harder half: reading an undocumented legacy warehouse, mapping its data structures and embedded business rules, and reproducing them provably in a lakehouse. Enterprise sources worked with directly: Oracle, PostgreSQL, MSSQL, HRIS and CRM mirrors, REST APIs and file feeds.

Data Modelling & BI

Dimensional modelling: star and galaxy schemas with conformed customer, product and transaction dimensions, SCD handling, standardised metric definitions. Reusable marts and analytics-ready datasets. Power BI (semantic models, advanced DAX, Direct Lake, RLS), the layer that consumes the migrated model, so target design accounts for how it will be queried.

Cloud



Azure in production: ADLS Gen2, Azure SQL, Key Vault, Data Factory. AWS in working use; S3 and Athena for operational revenue data. Storage, compute separation, partitioning and cost-aware scanning patterns are the same on either side.

LLM-based Tooling

Daily practice: LLM-assisted translation and refactoring of legacy SQL and stored procedure logic into PySpark, transformation scaffolding, reverse-engineering of undocumented models, and generation of test and reconciliation harnesses, every output validated against source data rather than trusted.

SQL & Python SQL

15+ yrs, advanced — complex multi-source joins, CTEs, window functions, incremental and idempotent MERGE logic, execution plans and query optimisation (Spark SQL, PostgreSQL, Athena, T-SQL, Oracle). Python 7+ yrs: PySpark daily, plus pandas, API clients and automation. Airflow DAG design, dbt, Git-based Dev / Test / Prod promotion.

Data Quality

Validation and integrity rules inside the pipeline, reconciliation batteries, coverage and freshness guards, diff checks, failure and SLA alerting; data dictionaries, lineage documentation and data contracts between layers; root-cause investigation with the fix applied at source, not patched downstream.

Delivery & Domains

Requirements discovery and gap analysis directly with business, risk, compliance and finance owners; user stories and acceptance criteria; documentation and handover. Financial services, payments, AML, regulatory reporting, US retirement and wealth, digital lending; enterprise manufacturing (BMW).

WORK EXPERIENCE

Data Engineer / Analytics Engineer - Fintech Platform

Duration: 2025 – August 2026

Summary:

- Own production pipelines end to end on a Delta Lake medallion lakehouse (Databricks / Microsoft Fabric): ingestion from provider REST APIs, internal services and file feeds; PySpark and Spark SQL transformation; curated marts, scheduling and monitoring.
- Migrated and unified chargeback data from six acquirers and PSPs (Stripe, Checkout.com, Revolut, Worldpay and others) into one canonical 31-column schema — each with a different payload shape and dispute lifecycle — then built the dimensional model and reporting layer on top.
- Design the Gold-layer marts the business queries — star and galaxy schemas with conformed customer, product and transaction dimensions and standardised metric definitions — so the same question returns the same number regardless of who asks.
- Validate migrated datasets against source with business and technical stakeholders: reconciliation batteries, control totals, coverage and freshness guards and record-level diffs — run inside the pipeline, not as a one-off sign-off.
- Build ETL/ELT for regulatory and management reporting: AML alerting across fiat (IBAN) and crypto (wallet) domains, risk scoring, transaction monitoring and a regulator-facing migration report.



- Optimise Spark performance and cost — rebuilt a full-year table scan as a rolling 30-day window, cutting runtime sharply and keeping the pipeline scalable as volume grew.
- Trace data defects to source rather than patching downstream: a legal-entity misclassification traced to a case-sensitive regex, a settlement gap caused by an INNER JOIN silently dropping records — each fixed at source and restated.

Technologies: Databricks, Delta Lake, Microsoft Fabric, PySpark, Spark SQL, REST APIs, Azure, Data Factory

Data Engineer — People & Business Analytics - Confidential (NDA) (Multi-brand Consumer & iGaming Group Reporting)

Duration: 2025 – 2026

Summary: Development of weekly and monthly reporting pipelines for a multi-brand consumer and iGaming group, integrating HRIS and operational revenue data to produce consolidated business metrics across multiple brands and projects.

Achievements:

- Built the group's weekly and monthly reporting pipelines in Python against PostgreSQL on Azure — scheduled extraction from an HRIS mirror, transformation, validation and automated delivery.
- Connected operational revenue data in Amazon Athena with headcount denominators to produce per-brand business metrics, including a brand-to-project mapping layer (21 brands into 9 projects) that made cross-source aggregation possible at all

Technologies: Python, PostgreSQL, Azure, Amazon Athena

Product Owner — Data & Platform Delivery (Financial Services, Retirement & Wealth Platform)

Duration: 2025 – Present

Summary: Product ownership and delivery coordination for a US-based financial services platform focused on retirement and wealth management, working closely with engineering teams to ensure data correctness and system integration in a regulated environment.

Responsibilities:

- Collaborated daily with US engineering teams across two delivery teams.
- Managed requirements, data correctness, system cutover, and integration logic.
- Communicated effectively in English within a large regulated enterprise.

Business / Data Analyst - InGain (Digital Lending Platform)

Duration: 01/2025 – 10/2025

Summary: Modeling and integration of lending platform data into the data warehouse, including applications, transactions, and user flows, with development of a reporting layer to support business insights.

Responsibilities:

- Modeled lending platform data and integrated it into the warehouse.
- Built reporting layers on top of integrated data.



Application Analyst — Data & Reporting - BMW – CoStCo (Enterprise Platform)

Duration: 11/2022 – 01/2025

Summary: Analysis and reporting on large Oracle datasets within a global enterprise environment, including data model design and end-user training to support management reporting and decision-making.

Responsibilities:

- Delivered analysis and reporting on large Oracle datasets with SQL, Python and SPSS inside a global enterprise landscape; designed the data models and trained end users on them.
- Found and fixed a production data-integrity issue that had distorted quarterly management reporting for years — a legacy filter excluding whole transaction types; corrected the historical models and presented impact and mitigation to senior stakeholders.

Technologies: Oracle, SQL, Python, SPSS

Earlier Experience

System Analyst / Technical Writer - IORYS (Compliance & Regulatory Documentation)

Duration: 10/2023 – 04/2024

Summary: Mapping data flows and producing ISO-compliant documentation of processes and system requirements to support compliance and regulatory needs.

Responsibilities:

- Mapped data flows and documented system requirements.
- Produced ISO-compliant process documentation.

Senior System Analyst - Via SMS (Financial Services Platform)

Duration: 06/2022 – 10/2022

Summary: Platform and data-flow analysis for a financial services company, supporting business and technical requirements.

Responsibilities:

- Analyzed platform and data flows.

Business Analyst - Intrum (Debt Management Platform)

Duration: 01/2022 – 06/2022

Summary: Data quality remediation and CRM integration analysis for a debt management platform in Switzerland.



Responsibilities:

- Performed platform and data-flow analysis for financial-licence application.
- Remediated CRM integration data quality issues.

Senior Business & Data Analyst (Freelance Engagements)

Duration: 09/2011 – 2022

Summary: Multiple long-term engagements across fintech, insurance, iGaming, and enterprise domains involving source-to-target mapping, SQL modeling, and ETL specification development.

Responsibilities:

- Performed source-to-target data mapping.
- Developed SQL models and ETL specifications.

Technologies: SQL

EDUCATION

- **Master of Engineering, Systems Analytics**
International Solomonov University, Kyiv, 2010

