

Databricks Modern Data Engineering – Przewodnik Referencyjny

Ten dokument stanowi uzupełnienie szkolenia i zawiera zbiór kluczowych linków do dokumentacji technicznej, materiałów wideo oraz kursów.

Fundamenty: Lakehouse i Spark

Podstawy architektury oraz silnika obliczeniowego Apache Spark.

Dokumentacja:

- Co to jest Data Lakehouse? – [Link](#)
- Architektura Apache Spark – [Link](#)
- Słownik pojęć Databricks (Workspace, Cluster, Job) – [Link](#)
- Delta Lake (ACID i Log Transakcyjny) – [Link](#)
- Silnik Photon – [Link](#)
- Architektura Medalionowa – [Link](#)

Wideo (YouTube):

- Architektura Lakehouse – [Link](#)
- Databricks Lakehouse Platform – [Link](#)
- Apache Spark Architecture Deep Dive – [Link](#)

Szkolenia (Databricks Academy):

- Databricks Fundamentals (Kurs darmowy) – [Link](#)

Delta Lake i Transakcyjność (ACID)

Format zapewniający bezpieczeństwo, spójność i wersjonowanie danych.

Dokumentacja:

- Przewodnik po Delta Lake – [Link](#)
- Historia tabel i Time Travel (RESTORE) – [Link](#)
- Walidacja danych (Constraints) – [Link](#)

Wideo:

- Delta Lake Deep Dive (Szczegóły dziennika transakcyjnego) – [Link](#)
- Delta Tables Best Practices – [Link](#)

Databricks Modern Data Engineering – Przewodnik Referencyjny

Ingestion: Ładowanie Danych

Nowoczesne metody wczytywania plików (Auto Loader, COPY INTO).

Dokumentacja:

- Auto Loader (cloudFiles) – [Link](#)
- COPY INTO (SQL Batch Loading) – [Link](#)
- Ewolucja schematu w Auto Loaderze – [Link](#)

Wideo:

- Databricks Auto Loader Demo – [Link](#)
- Ingest by Lakehouse Connection – [Link](#)

Architektura Medallion i Transformacje

Projektowanie potoków danych (Bronze/Silver/Gold) i operacje ETL.

Dokumentacja:

- Architektura Medallion – [Link](#)
- Lakeflow Spark Declarative Pipelines (SDP) – [Link](#)
- Komenda MERGE INTO (Upserty/SCD) – [Link](#)

Arkusze informacyjne (Cheat Sheets):

- PySpark DataFrame API Reference – [Link](#)
- Spark SQL Functions – [Link](#)

Optymalizacja i Wydajność

Techniki przyspieszania zapytań i redukcji kosztów.

Dokumentacja:

- Liquid Clustering (LC) – [Link](#)
- OPTIMIZE i Z-Ordering – [Link](#)
- VACUUM (Czyszczenie starych plików) – [Link](#)
- Adaptive Query Execution (AQE) – [Link](#)

Wideo:

- Performance Tuning on Databricks – [Link](#)
- Optimize Cost of Datarbricks. – [Link](#)

Databricks Modern Data Engineering – Przewodnik Referencyjny

Governance: Unity Catalog

Zarządzanie dostępem, bezpieczeństwem i katalogiem danych.

Dokumentacja:

- Unity Catalog (UC) i Ograniczenia RLS/CLS – [Link](#)
- Unity Catalog: Pierwsze kroki – [Link](#)
- Zarządzanie uprawnieniami (GRANT/REVOKE) – [Link](#)

Szkolenia:

- Data Governance with Unity Catalog (Wymaga logowania) – [Link](#)

AI i Generative BI (Nowe Funkcje)

Narzędzia wspierające pracę inżyniera danych przy użyciu sztucznej inteligencji.

Narzędzia:

- Databricks Assistant (Wsparcie w kodowaniu) – [Link](#)
- AI/BI Genie (Analityka w języku naturalnym) – [Link](#)
- Mosaic AI – [Link](#)
- Mosaic AI Vector Search / RAG – [Link](#)
- AI Functions (Funkcje AI) – [Link](#)

Projekty Praktyczne (End-to-End)

Gotowe przykłady implementacji łączące inżynierię danych z AI.

Zasoby:

- Databricks Solution Accelerators (Gotowe notebooki) – [Link](#)
- Projekt: Retail Orders End-to-End (GitHub) – [Link](#)

Społeczność i Wydarzenia

Miejsca do dalszej nauki i wymiany wiedzy.

- Wirtualny Festiwal Nauki (Październik 2025) – [Link](#)
- Forum Społeczności Databricks – [Link](#)

Certyfikacja

Szkolenie pokrywa materiał wymagany do egzaminu:

- [Databricks Certified Data Engineer Associate](#) – Informacje o egzaminie