

kod szkolenia: DBX-LKH / PL DL 1d

Databricks Lakehouse Architecture

Szkolenie Databricks (Lakehouse) to trzeci krok w spójnej ścieżce szkoleniowej Fundamental → Explorer → Lakehouse → Transformation. Uczestnicy uczą się projektować i wdrażać architekturę Lakehouse, korzystając z Delta Lake, Auto Loader i Structured Streaming. Program obejmuje także praktyczny cost management, optymalizację i podstawy governance.



Odbiorcy szkolenia

Szkolenie przeznaczone jest dla inżynierów danych oraz zespołów DataOps, które chcą poznać zasady budowy architektury Lakehouse i utrzymania procesów przetwarzania danych w Databricks



Korzyści

Korzyści dla uczestników

- rozumieją koncepcję Lakehouse i architekturę Medallion
- potrafią tworzyć, ładować i aktualizować dane w Delta Lake
- znają techniki batch i streaming load w Databricks
- potrafią optymalizować i monitorować tabele Delta
- znają zasady governance i lineage w Unity Catalog
- rozumieją, jak łączyć transformacje, optymalizację i kontrolę jakości w praktycznym przepływie danych
- są przygotowani do kolejnego etapu ścieżki – szkolenia Databricks Transformation



Program szkolenia

1. Wprowadzenie do architektury Lakehouse

- Koncepcja Lakehouse – połączenie Data Lake i Data Warehouse
- Struktura Medallion Architecture (Bronze, Silver, Gold)
- Rola Delta Lake i Unity Catalog w zarządzaniu danymi
- Projektowanie logiki przepływu danych między warstwami

2. Delta Lake w praktyce

- Operacje ACID i schema enforcement
- MERGE, UPDATE, DELETE i INSERT – modyfikacje tabel Delta
- Time travel i historia zmian (DESCRIBE HISTORY)
- Tworzenie tabel managed i external w Unity Catalog

3. Ingest danych – batch i stream

- COPY INTO jako metoda wsadowego ładowania danych
- Auto Loader (cloudFiles) – incremental ingest i schema evolution
- Monitoring strumieni w nowym interfejsie Streaming UI

4. Optymalizacja i zarządzanie danymi

- OPTIMIZE, ZORDER i VACUUM – mechanizmy optymalizacji Delta Lake
- Partitioning i analiza planów zapytań
- Liquid Clustering – automatyczne klastrowanie danych
- Delta Sharing – współdzielenie danych między zespołami i środowiskami

5. Cost management (practical)

- Koszty batch vs streaming (Auto Loader, Structured Streaming)
- Wpływ OPTIMIZE, ZORDER i VACUUM na koszty
- Planowanie zasobów dla dużych tabel i pipeline'ów
- Architektura kosztów w modelu Bronze-Silver-Gold

6. Observability & Monitoring (light)

- Monitoring w Streaming UI i Metrics UI
- Alerty w Workflows i SQL dashboardy
- Best practices obserwowalności w Lakehouse

7. Security fundamentals (light)

- Row-level security i column masking – podstawy
- Token passthrough – świadomość i scenariusze
- Unity Catalog jako warstwa governance

8. Projekt końcowy

- Zaprojektowanie i wdrożenie mini-Lakehouse z danymi ładowanymi batch i stream, optymalizacją Delta, kontrolą jakości, monitoringiem i kosztami



Oczekiwane przygotowanie uczestnika

- Ukończone szkolenie Databricks Explorer lub równoważna wiedza

- Doświadczenie w pracy z SQL i podstawami PySpark
- Znajomość podstawowych koncepcji chmury i architektury danych



Szkolenie obejmuje

* dostęp do portalu słuchacza Altkom Akademii

Szkolenie prowadzone jest w środowisku Databricks w chmurze. Każdy uczestnik otrzymuje własny workspace z dostępem do Unity Catalog, SQL Editor, Notebooków oraz katalogu z danymi testowymi. Uwaga: W trakcie wybranych szkoleń Akademia zapewnia Uczestnikom dostęp do płatnych narzędzi i usług niezbędnych do realizacji programu szkolenia (w szczególności usług Azure, Fabric oraz innych zasobów chmurowych). Uczestnik zobowiązany jest do korzystania z udostępnionych zasobów wyłącznie w zakresie wynikającym z agendy szkolenia oraz zgodnie z poleceniami trenera. Wykorzystywanie tych zasobów w sposób wykraczający poza zakres szkoleniowy, w szczególności do realizacji projektów prywatnych, testowania dodatkowych usług, modyfikowania konfiguracji środowiska lub podejmowania działań generujących koszty niezwiązane z realizacją ćwiczeń – skutkować będzie obciążeniem Uczestnika kosztami przekraczającymi przewidziany limit, na podstawie wydatków poniesionych przez Akademię (<https://www.altkomakademia.pl/ogolne-warunki-uczestnictwa-w-szkoleniach/>)



Język

Wykład: polski

Materiały: angielski

Metoda egzaminacyjna

Szkolenie nie zawiera formalnego egzaminu. Zaleca się jednak podejście do egzaminu Databricks Certified Data Engineer Associate jako kolejny krok rozwojowy.

Czas trwania

1 dni / 7 godzin

Opis egzaminu

Szkolenie nie zawiera formalnego egzaminu. Zaleca się jednak podejście do egzaminu Databricks Certified Data Engineer Associate jako kolejny krok rozwojowy.