kod szkolenia: DBX-DE / PL DL 2d

Databricks w pigułce – od danych do AI w dwa dni

Szkolenie Databricks w pigułce to dwudniowy przegląd całej platformy Databricks Data Intelligence Platform: od organizacji danych w Unity Catalog, przez ładowanie i przetwarzanie danych, po dashboardy, Genie i rozwiązania AI. Program pokazuje, jak te elementy łączą się w jeden przepływ pracy, na jednym scenariuszu biznesowym RetailOps (fikcyjna firma handlowa, na której danych realizujemy wszystkie ćwiczenia).

Szkolenie prowadzimy w formule hands-on demo. Trener pokazuje na ekranie każdy etap pracy w Databricks i omawia go na bieżąco, a uczestnicy śledzą pokaz i zadają pytania. Program ma dużo pokazów i mało teorii. Po każdym module jest krótka sesja „Sprawdź się”: quiz albo proste zadanie w narzędziu, które utrwala poznane funkcje i pokazuje, gdzie ich szukać w Databricks. Pełna samodzielna praktyka odbywa się po szkoleniu: uczestnicy dostają zadania domowe do każdego modułu oraz materiały poszkoleniowe z rozwiązaniami. Dzięki takiej formule na szkoleniu odnajdą się zarówno osoby, które dopiero zaczynają pracę z Databricks, jak i te, które już z niego korzystają i chcą uporządkować wiedzę.

Szkolenie to wiedza w pigułce. Każdy temat pokazujemy raz, w skrócie, tak żeby uczestnik zrozumiał, do czego służy i kiedy go użyć. Nie wchodzimy w szczegóły konfiguracji, optymalizacji ani przypadki brzegowe. Do tego służą moduły ścieżki Databricks (DBX-FE, LKH, ETL, LKF, DWH, RAG, AGT): każdy z nich to pełne 7 godzin pracy z jednym obszarem, w którym wchodzimy w temat głęboko, z laboratoriami i warsztatem końcowym. Przy każdym module poniżej wskazujemy szkolenie, na którym temat rozwijamy w całości.

Odbiorcy szkolenia

Szkolenie jest przeznaczone dla osób, które chcą w krótkim czasie poznać całą platformę Databricks i zrozumieć, jak w niej pracować z danymi, niezależnie od obecnego poziomu.
W szczególności kurs sprawdzi się dla:

- analityków, inżynierów danych i osób z IT, które zaczynają pracę z Databricks,
- osób, które już korzystają z Databricks w jednym obszarze i chcą poznać pozostałe,
- liderów zespołów i architektów, którzy potrzebują pełnego obrazu platformy przed decyzjami projektowymi,
- zespołów mieszanych (analitycy, inżynierowie, BI), które chcą zbudować wspólny słownik pojęć.

Szkolenie będzie szczególnie wartościowe dla osób, które chcą świadomie wybrać, w który obszar Databricks wejść głębiej.

Korzyści

- Pełny obraz platformy – poznasz rolę workspace, compute, Unity Catalog, Lakeflow, SQL warehouse, Genie i narzędzi AI oraz to, jak te elementy współpracują.
- Praca z danymi od początku do końca – zobaczysz drogę od pliku do tabeli Delta, pipeline'u i dashboardu na jednym scenariuszu.
- Unity Catalog – zrozumiesz, jak w Databricks są uporządkowane dane i kto ma do nich dostęp.
- Analityka i Genie – zobaczysz, jak z danych powstaje dashboard AI/BI i Genie Agent odpowiadający na pytania w języku naturalnym.
- AI w Databricks – poznasz Genie Code, AI Functions oraz to, czym są RAG i agenci w Agent Bricks.
- Sprawdzenie wiedzy w trakcie – po każdym module krótki quiz albo proste zadanie w narzędziu.
- Praktyka po szkoleniu – dostaniesz zadania domowe i materiały z rozwiązaniami, żeby samodzielnie przećwiczyć każdy etap.
- Mapa dalszej ścieżki – dowiesz się, które szkolenie wybrać dalej, zależnie od roli i potrzeb.

Program szkolenia

Dzień 1 – Platforma i dane
Temat wiodący: „Od pliku do zaplanowanego procesu”
1. Moduł – Databricks Data Intelligence Platform
Cel modułu: uczestnik wie, z jakich elementów składa się platforma i gdzie w niej pracuje.

- Architektura platformy – co działa po stronie Databricks, a co w chmurze klienta
- Workspace – notebooki, SQL editor, Catalog Explorer i asystent AI Genie Code
- Compute – serverless, klastry i SQL warehouse; jak wybór wpływa na koszt

- Demo z trenerem: uruchomienie pierwszego notebooka
- Sprawdź się: quiz – rodzaje compute i gdzie w workspace znaleźć notebooki, SQL editor i Catalog Explorer
- Szczegółowo: DBX-FE

2. Moduł – Unity Catalog w skrócie

Cel modułu: uczestnik rozumie, jak są uporządkowane dane i jak kontroluje się do nich dostęp.

- Hierarchia danych – katalog → schemat → tabela, widok lub Volume
- Uprawnienia – GRANT / REVOKE i właściciel obiektu
- Wyszukiwanie danych i lineage – skąd pochodzi tabela i kto z niej korzysta
- Demo z trenerem: utworzenie schematu i nadanie uprawnień
- Sprawdź się: odszukanie tabeli w Catalog Explorer i sprawdzenie, kto ma do niej dostęp i skąd pochodzą dane
- Zadanie po szkoleniu: uporządkowanie danych RetailOps we własnym schemacie
- Szczegółowo: DBX-FE

3. Moduł – Ładowanie danych i tabele Delta

Cel modułu: uczestnik wie, jak wczytać plik i zapisać wynik jako tabelę Delta.

- Volumes – miejsce na pliki objęte uprawnieniami Unity Catalog
- Odczyt plików w SQL (read_files()) i przyrostowe ładowanie przez Auto Loader
- Tabela Delta – zapis, aktualizacja przez MERGE, historia zmian i powrót do poprzedniej wersji
- Lakeflow Connect – gotowe konektory do systemów źródłowych (omówienie)
- Demo z trenerem: plik zamówień RetailOps → tabela Delta
- Sprawdź się: podgląd historii zmian tabeli Delta i odczyt jej poprzedniej wersji
- Zadanie po szkoleniu: załadowanie drugiego pliku i aktualizacja tabeli przez MERGE
- Szczegółowo: DBX-LKH

4. Moduł – Przekształcenia i warstwy danych

Cel modułu: uczestnik wie, jak z surowych danych powstaje zestaw gotowy do analizy.

- Architektura Medallion – warstwy Bronze, Silver i Gold w praktyce
- Przekształcenia w SQL – łączenie tabel, agregacje, obsługa braków i duplikatów
- Proste reguły jakości danych
- Demo z trenerem: z surowych zamówień do tabeli sprzedaży gotowej do raportu
- Sprawdź się: quiz – do której warstwy Medallion należy dana tabela
- Zadanie po szkoleniu: dodatkowa tabela Gold ze sprzedażą według regionu
- Szczegółowo: DBX-ETL

5. Moduł – Lakeflow Jobs

Cel modułu: uczestnik wie, jak uruchamiać przygotowany proces automatycznie.

- Job z kilku kroków, harmonogram i powiadomienie o błędzie
- Historia uruchomień – jak sprawdzić, co poszło nie tak
- Demo z trenerem: job z harmonogramem dla przepływu z modułów 3–4
- Sprawdź się: odczytanie historii uruchomień joba i znalezienie kroku, który zakończył się błędem
- Zadanie po szkoleniu: dodanie kroku z tabelą z zadania do modułu 4
- Szczegółowo: DBX-LKF

Dzień 2 – Od danych do wartości

Temat wiodący: „Dane, z których korzysta biznes i AI”

6. Moduł – Lakeflow pipelines

Cel modułu: uczestnik wie, czym deklaratywny pipeline różni się od notebooka uruchamianego jako job.

- Lakeflow pipelines (wcześniej Delta Live Tables) – graf zależności i przyrostowe przetwarzanie
- Reguły jakości (expectations) i podgląd przebiegu pipeline'u
- Demo z trenerem: pipeline Bronze → Silver → Gold dla danych RetailOps
- Sprawdź się: quiz – pipeline czy job: co wybrać w podanym scenariuszu
- Zadanie po szkoleniu: dodanie reguły jakości i analiza odrzuconych rekordów
- Szczegółowo: DBX-LKF

7. Moduł – SQL warehouse i dashboardy AI/BI

Cel modułu: uczestnik wie, jak przygotować dashboard na danych z warstwy Gold.

- SQL warehouse – moc obliczeniowa dla zapytań i dashboardów
- Metric views – wskaźniki zdefiniowane raz w Unity Catalog (omówienie)
- Dashboard AI/BI – wizualizacje, filtry i publikacja
- Demo z trenerem: dashboard sprzedaży z filtrami
- Sprawdź się: dodanie filtra i nowej wizualizacji do dashboardu
- Zadanie po szkoleniu: druga strona dashboardu z analizą produktów
- Szczegółowo: DBX-DWH

8. Moduł – Genie: pytania do danych w języku naturalnym

Cel modułu: uczestnik wie, jak przygotować Genie Agent i kiedy jego odpowiedziom można ufać.

- Genie Agent – zestaw tabel, instrukcje i przykładowe zapytania
- Uprawnienia – Genie widzi tylko dane dostępne dla użytkownika
- Demo z trenerem: Genie Agent dla danych sprzedażowych i kilka pytań testowych
- Sprawdź się: zadanie pytania w Genie i sprawdzenie wygenerowanego zapytania SQL
- Zadanie po szkoleniu: uzupełnienie instrukcji i test na 5 pytaniach kontrolnych
- Szczegółowo: DBX-DWH

9. Moduł – AI w Databricks

Cel modułu: uczestnik zna możliwości AI w Databricks i wie, od czego zacząć.

- AI Functions w SQL – klasyfikacja, ekstrakcja i streszczanie tekstu (ai_query, ai_classify)
- RAG z AI Search – odpowiedzi oparte na dokumentach firmy
- Agenci w Agent Bricks – asystent, który korzysta z narzędzi i danych
- Demo z trenerem: klasyfikacja opinii klientów RetailOps przez AI Functions, pokaz RAG i agenta
- Sprawdź się: quiz – AI Functions, RAG czy agent: co wybrać do danego problemu
- Zadanie po szkoleniu: ekstrakcja informacji z opinii klientów przez AI Functions
- Szczegółowo: DBX-RAG, DBX-AGT

10. Moduł – Podsumowanie i mapa dalszej ścieżki

Cel modułu: uczestnik łączy elementy obu dni i wie, który obszar rozwijać dalej.

- Przegląd całego przepływu: plik → tabela Delta → pipeline → dashboard i Genie → AI
- Dobre praktyki na start – koszty compute, minimalne uprawnienia, brak haseł w kodzie
- Sprawdź się: quiz podsumowujący oba dni

- Omówienie zadań domowych i materiałów poszkoleniowych
- Dalsza ścieżka według roli: inżynier danych (DBX-LKH → ETL → LKF), analityk i BI (DBX-DWH), AI (DBX-RAG → AGT), certyfikacja (DBX-DEA, szkolenie dedykowane)

Oczekiwane przygotowanie uczestnika

- Brak wymagań formalnych; szkolenie jest dostępne niezależnie od poziomu uczestnika
- Pomocna jest znajomość podstaw SQL (SELECT, WHERE, JOIN)
- Nie jest wymagana znajomość Python ani wcześniejsza praca z Databricks

Dla zwiększenia komfortu pracy oraz efektywności szkolenia zalecamy skorzystanie z dodatkowego ekranu. Brak dodatkowego ekranu nie jest przeciwwskazaniem do udziału w szkoleniu, ale w znaczący sposób wpływa na komfort pracy podczas zajęć
Informacje oraz wymagania dotyczące uczestniczenia w szkoleniach w formule zdalnej dostępne na: https://www.altkomakademia.pl/jak-dzialamy/o-szkoleniach/?distance-learning

Szkolenie obejmuje

* dostęp do portalu słuchacza Altkom Akademii
Metoda szkolenia:

- przygotowane środowisko Databricks z Unity Catalog, SQL warehouse i danymi scenariusza RetailOps na czas szkolenia
- notebooki z przebiegiem wszystkich pokazów, w formie elektronicznej
- quizy i proste zadania „Sprawdź się” po każdym module
- zadania domowe do każdego modułu oraz materiały poszkoleniowe z rozwiązaniami; zadania można wykonać po szkoleniu w bezpłatnej Databricks Free Edition

Uwaga:W trakcie wybranych szkoleń Akademia zapewnia Uczestnikom dostęp do płatnych narzędzi i usług niezbędnych do realizacji programu szkolenia (w szczególności usług Azure, 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/)

Czas trwania

2 dni / 14 godzin

Język

Wykład: polski
Materiały: angielski