

Databricks: Data Warehousing, AI/BI i Genie

Szkolenie Databricks: Data Warehousing, AI/BI i Genie to praktyczny kurs zamiany danych z warstwy Gold w produkt analityczny dla biznesu. Program koncentruje się na modelu analitycznym, wspólnych definicjach wskaźników w Unity Catalog (metric views), pracy na SQL warehouse oraz budowie dashboardów AI/BI.

Uczestnicy przechodzą przez cały proces: od zaprojektowania tabel faktów i wymiarów, przez zdefiniowanie wskaźników jako metric view, dobór i koszt SQL warehouse, po opublikowanie dashboardu i przygotowanie kontekstu dla Genie, czyli zadawania pytań do danych w języku naturalnym. Dzięki temu szkolenie kończy się dashboardem sprzedażowym opartym na wspólnych, opisanych definicjach wskaźników, z których korzystają wszyscy odbiorcy. Ćwiczenia opierają się na danych RetailOps (fikcyjna firma handlowa, na której danych realizujemy wszystkie ćwiczenia).



Odbiorcy szkolenia

Szkolenie jest przeznaczone dla osób, które przygotowują dane dla raportowania i analiz w Databricks oraz dla osób udostępniających dane odbiorcom biznesowym.

W szczególności kurs sprawdzi się dla:

- analityków danych i specjalistów BI pracujących w SQL,
- inżynierów analityki (analytics engineers) projektujących warstwę Gold,
- twórców raportów, którzy chcą korzystać z dashboardów AI/BI lub łączyć Databricks z Power BI,
- zespołów, które chcą udostępnić użytkownikom biznesowym Genie w kontrolowany sposób.

Szkolenie będzie szczególnie wartościowe dla organizacji, w których te same wskaźniki są liczone inaczej w różnych raportach.



Korzyści

- Model analityczny – nauczysz się projektować tabele faktów i wymiarów z właściwą ziarnistością.
- Wspólne definicje wskaźników – poznasz Unity Catalog metric views i dowiesz się, jak zdefiniować KPI raz i używać ich w SQL, dashboardach AI/BI, Genie i alertach.
- SQL warehouse – dowiesz się, jak dobrać typ i rozmiar warehouse oraz jak kontrolować wydajność i koszt.
- Dashboardy AI/BI – zbudujesz dashboard z filtrami i przechodzeniem do szczegółów (drill-through) oraz poznasz zasady jego publikacji.
- Integracja z narzędziami BI – poznasz wzorce połączenia z Power BI i sposób zachowania spójności wskaźników.
- Genie – dowiesz się, jak przygotować kontekst, przetestować pytania i ustawić uprawnienia, żeby odpowiedzi w języku naturalnym były wiarygodne.



Program szkolenia

1. Moduł – Od warstwy Gold do modelu analitycznego

Cel modułu: uczestnik projektuje tabele tak, żeby analitycy i narzędzia BI mogli z nich łatwo korzystać.

- Ziarnistość danych – na jakim poziomie szczegółu przechowujemy dane (np. pozycja zamówienia czy całe zamówienie)
- Tabele faktów i wymiarów (model gwiazdy) – sprzedaż w jednej tabeli, klienci, produkty i daty w osobnych
- Klucze główne i obce w Unity Catalog – opis powiązań między tabelami, z którego korzystają narzędzia BI i Genie
- Opisy tabel i kolumn, tagi i oznaczanie zaufanych tabel – dokumentacja dla ludzi i dla Genie
- Lab: uczestnik uzupełnia przygotowany model gwiazdy o opisy i klucze

2. Moduł – Unity Catalog metric views

Cel modułu: uczestnik definiuje wskaźniki biznesowe w jednym miejscu, tak żeby wszystkie raporty liczyły je tak samo.

- Problem – ten sam wskaźnik (np. przychód) liczony inaczej w różnych raportach i zespołach
- Metric view – definicja wskaźników i wymiarów zapisana raz w Unity Catalog
- Budowa definicji: źródło danych, powiązania tabel, wymiary (np. region, miesiąc) i miary (np. przychód, marża, a także sumy narastające i średnie kroczące)
- Korzystanie z metric view w zapytaniach SQL, dashboardach i Genie
- Uprawnienia – kto może korzystać z definicji wskaźników
- Lab: uczestnik definiuje przychód, marżę i średnią wartość zamówienia jako metric view

3. Moduł – SQL warehouse: wydajność i koszt

Cel modułu: uczestnik potrafi dobrać i skonfigurować SQL warehouse oraz znaleźć przyczynę wolnych

zapytań.

- SQL warehouse – moc obliczeniowa dla zapytań SQL i dashboardów; rodzaje: serverless, pro, classic
- Rozmiar i skalowanie – jak warehouse reaguje na wielu użytkowników naraz
- Historia i profil zapytania – jak znaleźć zapytanie, które działa wolno, i zrozumieć dlaczego
- Koszt – automatyczne wyłączanie, cache wyników i zużycie warehouse w system tables
- Lab: uczestnik analizuje wolne zapytanie i proponuje lepszą konfigurację warehouse

4. Moduł – Dashboardy AI/BI

Cel modułu: uczestnik buduje i publikuje dashboard dla odbiorców biznesowych.

- Źródła danych dashboardu – zapytania SQL i metric views
- Wizualizacje, filtry, parametry, filtrowanie krzyżowe i przechodzenie do szczegółów na innej stronie dashboardu (drill-through)
- Asystent AI (Genie Code) – tworzenie wykresów, filtrów i stron dashboardu na podstawie opisu w języku naturalnym
- Publikacja – kto widzi dashboard i czy widzi dane na swoich uprawnieniach, czy na uprawnieniach autora
- Harmonogram odświeżania i subskrypcje – wysyłka dashboardu e-mailem, na Slack lub Microsoft Teams
- Pokaz: budowa dashboardu (uczestnicy budują go samodzielnie w warsztacie, moduł 7)

5. Moduł – Integracja z Power BI i innymi narzędziami BI

Cel modułu: uczestnik wie, jak połączyć Databricks z Power BI bez rozjazdu wskaźników.

- Tryby połączenia Power BI: import danych lub zapytania na żywo (DirectQuery) – zalety i koszty każdego (omówienie i pokaz)
- Publikacja danych z Databricks do Power BI
- Spójność wskaźników z Power BI – konektor Power BI nie obsługuje dziś natywnie metric views; stosujemy widok pośredni z MEASURE() albo zapytanie natywne
- Kiedy wystarczy dashboard AI/BI, a kiedy potrzebne jest zewnętrzne narzędzie BI; przeniesienie istniejącego raportu Power BI lub Tableau do dashboardu AI/BI z definicjami wskaźników jako metric views (import przez Genie Code, /importBI)
- Lab: uczestnik dobiera sposób udostępnienia danych dla trzech typów odbiorców

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

Cel modułu: uczestnik przygotowuje Genie tak, żeby odpowiedzi były poprawne i zgodne z uprawnieniami.

- Genie Agent (wcześniej Genie space) – zestaw tabel i metric views, o które użytkownicy mogą pytać w języku naturalnym; użytkownicy biznesowi korzystają z niego w Genie One
- Instrukcje i przykładowe zapytania – jak nauczyć Genie słownictwa firmy i poprawnych obliczeń
- Zaufane zapytania – gotowe, sprawdzone odpowiedzi na najczęstsze pytania
- Testowanie jakości – zestaw pytań kontrolnych i porównanie odpowiedzi z oczekiwanymi
- Uprawnienia i ograniczenia – Genie widzi tylko dane dostępne dla użytkownika; kiedy nie udostępniać Genie
- Lab: uczestnik przygotowuje Genie Agent dla danych sprzedażowych i testuje go na 5 pytaniach kontrolnych (benchmark)

7. Moduł – Warsztat zintegrowany

Cel modułu: uczestnik łączy wskaźniki, dashboard i Genie w jeden produkt analityczny.

- Warsztat: metric view → dashboard KPI → Genie Agent na tych samych definicjach wskaźników. Efektem jest opublikowany dashboard sprzedażowy i Genie Agent oparte na wspólnych, opisanych wskaźnikach.
- Lista kontrolna przed udostępnieniem dashboardu i Genie użytkownikom biznesowym



Oczekiwane przygotowanie uczestnika

- Ukończone szkolenie DBX-FND (Fundamenty Databricks) lub równoważna wiedza o Unity Catalog i tabelach Delta
- Praktyczna znajomość SQL (JOIN, GROUP BY, funkcje agregujące)
- Osobom budującym własną warstwę Gold zalecamy wcześniejsze szkolenie DBX-ETL

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 SQL warehouse i danymi Gold scenariusza RetailOps na czas szkolenia
- uprawnienia do tworzenia dashboardów i Genie Agents w środowisku szkoleniowym
- materiały do ćwiczeń w formie elektronicznej

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

1 dni / 7 godzin

Język

Wykład: polski

Materiały: angielski