

BigQuery – poziom zaawansowany

Zaawansowane techniki analizy danych i automatyzacji zapytań w środowisku BigQuery

Cele szkolenia

Szkolenie pozwala uczestnikom:

- opanować zaawansowane techniki przetwarzania i analizy danych w BigQuery,
- zrozumieć i stosować funkcje analityczne i okienkowe do tworzenia dynamicznych raportów,
- automatyzować przetwarzanie danych poprzez tworzenie procedur i funkcji użytkownika,
- rozszerzyć swoje umiejętności z zakresu zarządzania błędami zapytań SQL,
- lepiej integrować dane BigQuery z innymi narzędziami, takimi jak Excel.

Umiejętności

Dzięki szkoleniu uczestnik będzie:

- Łączyć dane z różnych źródeł z wykorzystaniem UNION, INTERSECT, EXCEPT.
- Tworzyć i zarządzać widokami w BigQuery.
- Pobierać dane do Excela przy użyciu Power Query.
- Wykorzystywać zaawansowane klauzule grupujące: ROLLUP, CUBE, GROUPING SETS.
- Pisać zagnieżdżone zapytania oraz korzystać z CTE (konstrukcja WITH).
- Tworzyć raporty z użyciem funkcji rankingowych: ROW_NUMBER, RANK, DENSE_RANK.
- Wykonywać porównania okresowe (np. miesiąc do miesiąca) z wykorzystaniem LAG i LEAD.
- Stosować PIVOT i UNPIVOT w celu transformacji danych.
- Obsługiwać funkcje okien (Window Functions), np. NTILE.
- Tworzyć procedury i funkcje użytkownika z wykorzystaniem zmiennych, warunków i pętli.
- Obsługiwać błędy za pomocą funkcji SAFE_* i ERROR().

Profil uczestników

Szkolenie jest przeznaczone dla:

analityków danych, specjalistów BI, data scientistów,
programistów pracujących z BigQuery lub planujących migrację danych do tego środowiska,
specjalistów IT odpowiedzialnych za integrację, raportowanie i analizę danych,
osób odpowiedzialnych za automatyzację przetwarzania danych.

Przygotowanie uczestników

Uczestnicy powinni:

- znać podstawowe konstrukcje SQL (SELECT, JOIN, WHERE, GROUP BY),
- mieć doświadczenie w pracy z bazami danych (preferowane środowisko BigQuery),
- znać podstawowe pojęcia związane z analityką danych i przetwarzaniem zapytań,

UWAGA! Comarch nie dostarcza w przypadku tego szkolenia środowiska dla uczestników. Uczestnicy pracują na środowisku udostępnionym przez ich firmę.

Szczegółowy program szkolenia

1. Powtórzenie/przypomnienie informacji o podstawowych konstrukcjach zapytań SELECT

Wykorzystanie aliasów (AS) dla kolumn i tabel

Grupowanie danych

Instrukcje warunkowe

Złączenia tabel

2. Łączenie zbiorów danych

UNION, UNION ALL, INTERSECT, EXCEPT

3. Przygotowywanie widoków

Tworzenie widoków z gotowych zapytań

Obsługa widoków

4. Pobieranie danych z bazy danych do Excel przy pomocy PowerQuery

Pobieranie danych z tabel, widoków

Wykorzystanie instrukcji SELECT do importu danych

5. Rozszerzone klauzule grupujące

Wykorzystanie klauzul ROLLUP, CUBE, GROUPING, GROUPING SETS

6. Zagnieżdżanie zapytań

Select w zapytaniu SELECT

CTE - Wykorzystanie konstrukcji WITH

7. Funkcje analityczne

Funkcje raportujące, rankingowe np. ROW_NUMBER, RANK, DENSE_RANK

PARTITION BY, ORDER BY w funkcjach okiennych

Zagnieżdżanie agregacji, np. SUM(COUNT)

Poruszanie się po rekordach - LIMIT, OFFSET

Raporty okres do okresu, np. miesiąc do miesiąca – LAG, LEAD

Piwotowanie i de-piwotowanie danych - PIVOT, UNPIVOT

Korzystanie z funkcji okna (Window Functions), np. NTILE

Zamiennik CROSS APPLY: JOIN z UNNEST()

Zamiennik OUTER APPLY: LEFT JOIN z QUALIFY

8. Procedury użytkownika

Tworzenie procedur

Zmienne i stałe - deklarowanie, zakres życia

Konstrukcje sterujące przepływem programu (warunki IF, pętle WHILE)

Instrukcja CASE

Procedury i funkcje - różne tryby przekazywania parametrów, zwracanie wartości

Różnice i podobieństwa w wykorzystaniu zmiennych tabelarycznych oraz tabel tymczasowych

9. Funkcje użytkownika

Funkcje skalarne i użytkownika (Scalar and User-defined Functions)

10. Obsługa błędów

Pozyskiwanie informacji o błędach

Funkcja ERROR()

SAFE_CAST(), SAFE_OFFSET(), SAFE_DIVIDE()

Metoda realizacji szkolenia

Szkolenie realizowane jest w formie naprzemiennie następujących po sobie mini wykładów oraz ćwiczeń praktycznych. Szkolenie łączy w sobie fachową wiedzę merytoryczną z praktycznymi przykładami jej wykorzystania w środowisku pracy.

Liczba dni, liczba godzin szkoleniowych

2 dni, 16 godzin szkoleniowych

Ścieżka rozwoju po szkoleniu

- [Python dla analizy danych](#)
- Google Cloud Platform
- [Google Looker Studio](#), [Tableau](#) lub [Power BI](#).