

Ćwiczenie 12

Cele ćwiczenia

Celem tego ćwiczenia jest porównanie rozwiązań klasy Business Intelligence dostępnych komercyjnie. Zakładam, iż przegląd będę tworzył dla średniej firmy sprzedającej wyroby cukiernicze i pieczywo. Firma posiada kilka oddziałów w całej Polsce oraz aktywnie prowadzi swój własny program lojalnościowy. Firma posiada również własną aplikację mobilną, na której użytkownicy mogą składać zamówienia i zarządzać swoim kontem lojalnościowym (zbierać i wydawać punkty). Analiza porównawcza rozwiązań obejmuje nie tylko narzędzia raportowania, ale również architekturę przechowywania danych.

Na początku zaznaczę, iż rekomenduję oparcie się o jedno środowisko, dzięki czemu łatwiej, bardziej kompatybilnie i szybciej będzie można przeprowadzać różne procesy na danych. Ułatwi to również zarządzanie subskrypcjami.

Analiza rozwiązań

Proces przepływu danych podzieliłem na 6 kluczowych etapów:

1. Pozyskanie danych (dane pozyskiwane są z różnych źródeł: ERP – SQL server, strona webowa, aplikacja mobilna – postgresQL, terminale płatnicze– API).
2. Transformacja i przygotowanie danych (ETL/ELT)
3. Magazynowanie danych (Data Warehouse)
4. Modelowanie danych
5. Raportowanie i analiza
6. Udostępnianie

W kolejnych krokach krótko opiszę każdy z kroków przepływu danych w kontekście danego środowiska. Po uwagę wezmę również takie aspekty jak cenę, skalowalność, jak i dostępność asysty technicznej. Do analizy wybrałem 4 popularne technologie od 4 różnych dostawców rozwiązań BI:

- **Microsoft** (Power BI + Azure),
- **Amazon** (QuickSight + AWS)
- **Google** (Looker + BigQuery)
- **Comarch** (BI Point)

Microsoft

- Cena: Power BI Premium (dla organizacji): 1200 EUR miesięcznie (24,00 USD za licencję Pro × 50 użytkowników)
- Transformacja, przygotowanie i magazynowanie danych: Bogate możliwości ETL/ELT (technologia Azure Data Factory), integracja z MS SQL Server, PostgreSQL, API przez konektory. Synapse umożliwia przechowywanie danych w hurtowniach
- Modelowanie i raportowanie danych: Zaawansowane dashboardy, język DAX, integracja z Excel i innymi technologiami od Microsoft
- Skalowalność: Wysoka – możliwość stopniowego skalowania dzięki chmurze Azure, elastyczny model subskrypcyjny

- Wsparcie: Najpopularniejsze narzędzie, zatem bardzo szeroka społeczność, fora, blogi, masa szkoleń

Amazon

- Cena: QuickSight Standard: 9 USD/użytk./mies., Enterprise: 18 USD/użytkownik, Redshift: ok. 0.25 USD/godz. + storage
- Transformacja, przygotowanie i magazynowanie danych: Integracje przez Glue, Kinesis, Redshift. Obsługa wielu źródeł (API, SQL, JSON). Dobre ETL przez AWS Glue
- Modelowanie i raportowanie danych: Interaktywne dashboardy, ML insights. Raczej słabsze narzędzie od Power BI
- Skalowalność: Wysoka – Redshift i inne komponenty AWS projektowane do dużych środowisk danych
- Wsparcie: płatne, dokumentacja dobra, ale społeczność mniejsza niż Microsoft czy Google

Google

- Cena: Looker: ok. 30–60 USD/użytkownik/miesiąc, BigQuery: płatność za zapytanie i storage
- Transformacja, przygotowanie i magazynowanie danych: Integracja z GCP, BigQuery, Sheets, AppScript. Rozbudowane ETL z Cloud Data Fusion i Dataprep
- Modelowanie i raportowanie danych: Interaktywne dashboardy, LookML. Raczej słabsze narzędzie od Power BI
- Skalowalność: Bardzo wysoka – BigQuery jest bezserwerowe, gotowe do pracy z danymi o ogromnych rozmiarach
- Wsparcie: szkolenia, dokumentacja, mniejsza społeczność niż Power BI

Comarch BI Point

- Cena: Zależna od wdrożenia i modułów, szacunkowo od 2 000–5 000 zł/miesiąc
- Transformacja, przygotowanie i magazynowanie danych: Dobra integracja z Comarch ERP, mniej elastyczna z zewnętrznymi źródłami (np. PostgreSQL, API wymaga dostosowań)
- Modelowanie i raportowanie danych: Interaktywne raporty i dashboardy, ale mniej zaawansowane niż Power BI czy Looker.
- Skalowalność: Ograniczona – bardziej skierowana na średnie firmy, trudniejsza skalowalność w chmurze
- Wsparcie: Silne wsparcie lokalne, dokumentacja po polsku, mniejsza społeczność online

Macierz decyzyjna rozwiązań klasy BI

KRYTERIUM	Waga	Microsoft	Amazon	Google	Comarch
Cena	20%	8	7	6	8
Transformacja, przygotowanie i magazynowanie danych	20%	9	7	9	6
Modelowanie i raportowanie danych	20%	10	7	9	5
Skalowalność	20%	8	10	10	6
Wsparcie	20%	9	6	8	10
Podsumowanie	100%	8.8	7.4	8.4	7.0

Moja ocena rozwiązań i rekomendacje

Na podstawie analizy macierzy decyzyjnej rekomenduję wybór rozwiązania Microsoft Power BI w połączeniu z Azure (Azure Synapse, Data Factory). Mój wybór argumentuję:

- Zdecydowanie najlepszym, najbardziej funkcjonalnym programem do raportowania (Power BI)
- Dużą ilością szkoleń, use case'ów, przykładów użycia w sieci z wykorzystania technologii
- Integracją z innymi narzędziami od Microsoft, które są używane przez naszą firmę (SQL Server, Excel, itp.)
- Możliwością dosyć obszernego skalowania (zważając na dosyć szybki rozwój firmy)
- Tym, że na rynku najłatwiej znaleźć specjalistów BI biegłych w tych rozwiązaniach
- Rozmiarem i doświadczeniem firmy
- Dostosowywaniem przez Microsoft programów do aktualnych potrzeb rynkowych (aktualizacje)

Alternatywne technologie nie wchodzące w skład opisanych wyżej korporacji

- **Tableau** – zaawansowane narzędzie BI do interaktywnej wizualizacji danych

- **Snowflake** – chmurowa hurtownia danych oferująca elastyczne skalowanie, wysoką wydajność i separację zasobów obliczeniowych od przechowywania danych
- **KNIME** – darmowe narzędzie analityczne umożliwiające wizualne projektowanie przepływów danych
- **Dbt** – narzędzie do modelowania danych w architekturze ELT, pozwalające na wersjonowanie, testowanie i dokumentowanie transformacji SQL bezpośrednio w hurtowni danych