

Opis najważniejszych narzędzi analitycznych

1. Azure Machine Learning

Chmurowa platforma do budowania, trenowania i wdrażania modeli uczenia maszynowego (ML).

Do czego służy:

- Tworzenie modeli ML przy użyciu danych z Azure.
- Automatyzacja eksperymentów ML.
- Wdrażanie modeli jako usług API.
- Monitorowanie i wersjonowanie modeli.

Użycie:

Gdy chcesz analizować dane i prognozować np. zachowania klientów, wykrywać oszustwa, czy tworzyć systemy rekomendacji.

2. Azure DevOps

Zestaw narzędzi do zarządzania cyklem życia aplikacji – od planowania, przez kodowanie, testowanie, aż po wdrażanie.

Do czego służy:

- CI/CD (ciągła integracja i dostarczanie).
- Przechowywanie kodu w repozytoriach Git.
- Zarządzanie zadaniami w stylu Agile.
- Automatyzacja testów i buildów.

Użycie:

Kiedy stworzysz oprogramowanie w zespole i chcesz zautomatyzować procesy tworzenia, testowania i publikowania aplikacji.

3. Azure Databricks

Zintegrowane środowisko analityczne oparte na Apache Spark, służące do pracy z dużymi zbiorami danych i uczenia maszynowego.

Do czego służy:

- Przetwarzanie danych w dużej skali.

- Tworzenie zaawansowanych analiz i pipeline'ów ML.
- Kolaboracja między zespołami Data Science, Data Engineering i BI.

Użycie:

Kiedy pracujesz z Big Data i potrzebujesz szybko analizować i przetwarzać dane w skali petabajtów.

4. Azure Data Factory

Narzędzie typu ETL (Extract, Transform, Load) służące do integracji danych z różnych źródeł.

Do czego służy:

- Automatyzacja pobierania, transformowania i ładowania danych.
- Budowanie potoków danych (pipelines).
- Orkiestracja danych z lokalnych i chmurowych źródeł.

Użycie:

Gdy chcesz zbudować proces przetwarzania danych np. z plików CSV, baz danych, API do magazynu danych lub hurtowni danych.

5. Azure Synapse Analytics

Platforma analityczna typu „data warehouse + big data”, która łączy hurtownię danych i analitykę w jednym miejscu.

Do czego służy:

- Agregacja i analiza danych z wielu źródeł.
- Tworzenie raportów i dashboardów.
- Wsparcie SQL, Spark, Pipelines w jednej platformie.

Użycie:

Kiedy chcesz tworzyć zaawansowane analizy biznesowe na dużych zbiorach danych (Business Intelligence, BI).

6. Azure Stream Analytics

Usługa do przetwarzania danych strumieniowych w czasie rzeczywistym.

Do czego służy:

- Analiza danych z urządzeń IoT, czujników, logów w czasie rzeczywistym.
- Reagowanie na zdarzenia natychmiast po ich wystąpieniu.
- Generowanie alertów i wizualizacji danych na bieżąco.

Użycie:

Gdy potrzebujesz przetwarzać dane „na żywo”, np. monitorować ruch, wykrywać anomalie lub zarządzać systemami IoT.

7. Azure Event Hubs

Platforma do przesyłania dużej ilości danych zdarzeń (event streaming) z różnych źródeł.

Do czego służy:

- Gromadzenie i buforowanie danych strumieniowych.
- Integracja z analizą czasu rzeczywistego (np. Stream Analytics).
- Wysoka przepustowość – miliony zdarzeń na sekundę.

Użycie:

Jako „kanał” do zbierania danych z aplikacji, logów, urządzeń IoT, które potem są analizowane lub przechowywane.

8. Azure Data Lake Storage

Rozszerzony system plików oparty na chmurze, przeznaczony do przechowywania danych w dowolnej formie (strukturalnej i niestukturalnej).

Do czego służy:

- Przechowywanie ogromnych ilości danych – zarówno surowych, jak i przetworzonych.
- Integracja z narzędziami typu Spark, Databricks, Synapse.
- Optymalizacja pod kątem analizy danych.

Azure:

- **Azure Synapse Analytics:** Zintegrowana platforma analityczna, łącząca big data, data warehousing i analytics. Wspiera SQL, Apache Spark, i integracje z Power BI.
- **Azure HDInsight:** Usługa Hadoop oparta na chmurze, pozwalająca na przetwarzanie danych z użyciem Hadoop, Spark, Hive, itp.
- **Azure Data Factory:** Usługa ETL, która pozwala na automatyczne zbieranie, czyszczenie i przetwarzanie danych z różnych źródeł.
- **Azure Cognitive Services:** Zestaw gotowych modeli AI, umożliwiających m.in. detekcję anomalii, rozpoznawanie obrazów, przetwarzanie języka naturalnego.

AWS:

- **Amazon Redshift:** Usługa data warehousing do przetwarzania dużych zbiorów danych, zintegrowana z usługami analitycznymi w AWS.
- **AWS S3:** Główna usługa przechowywania danych, idealna do przechowywania dużych ilości danych w chmurze.
- **Amazon Kinesis:** Usługa strumieniowego przetwarzania danych w czasie rzeczywistym, podobna do Azure Event Hubs.
- **AWS Glue:** Usługa do integracji, przetwarzania i czyszczenia danych w chmurze.

GCP (Google Cloud Platform):

- **BigQuery:** Szybka usługa analityczna na dużych zbiorach danych, oparta na SQL. Doskonata do analizy dużych zbiorów danych w chmurze.
- **Google Cloud Pub/Sub:** Usługa do przesyłania strumieniowych danych w czasie rzeczywistym, podobna do Event Hubs w Azure.
- **Google Cloud Dataflow:** Platforma do przetwarzania danych w czasie rzeczywistym i w partiach, oparta na Apache Beam.
- **Google AI Platform:** Platforma do trenowania i uruchamiania modeli AI, integrująca się z innymi usługami Google Cloud.

1. Zbieranie danych:

- a. **Azure IoT Hub:** Użycie IoT Hub do zbierania danych z urządzeń (np. urządzenia graczy, urządzenia serwerowe).

- b. **Azure Event Hubs:** Zbieranie strumieni danych w czasie rzeczywistym z gier online i interakcji graczy.

2. Przechowywanie danych:

- a. **Azure Data Lake Storage Gen2:** Przechowywanie danych w formacie RAW, zarówno strukturalnych, jak i niestukturalnych. Idealne do przechowywania logów, danych IoT, itp.
- b. **Azure SQL Database** lub **Cosmos DB:** Przechowywanie danych strukturalnych (np. dane o graczach, interakcje w grze).

3. Przetwarzanie danych:

- a. **Azure Databricks:** Użycie Apache Spark w Azure Databricks do przetwarzania dużych ilości danych i analizy w czasie rzeczywistym (np. do analizy logów, przetwarzania danych z gier).
- b. **Azure Stream Analytics:** Użycie do przetwarzania strumieniowego w czasie rzeczywistym, detekcji anomalii i wykrywania problemów w czasie rzeczywistym w danych IoT.

4. Analiza i detekcja anomalii:

- a. **Azure Machine Learning:** Użycie algorytmów uczenia maszynowego (np. klasyfikacja, wykrywanie anomalii) do analizy danych i wykrywania nieprawidłowości (np. błędów w grze).
- b. **Azure Cognitive Services (Anomaly Detector):** Użycie gotowego modelu detekcji anomalii do wykrywania nienormalnych wzorców w danych graczy lub systemów.

5. Wizualizacja danych:

- a. **Power BI:** Tworzenie dashboardów z wynikami analizy, monitorowanie satysfakcji graczy, wizualizacja wykrytych anomalii w czasie rzeczywistym.
- b. **Azure Synapse Analytics:** Integracja danych i tworzenie raportów analitycznych na dużą skalę, łączenie danych z różnych źródeł.

6. Compute:

- a. **Azure Kubernetes Service (AKS):** Użycie kontenerów do uruchamiania usług analitycznych, takich jak modele ML, analizy strumieniowe itp.
- b. **Azure Virtual Machines (VMs):** Jeśli wymagane jest większe dostosowanie obliczeń, użycie maszyn wirtualnych do uruchamiania złożonych procesów obliczeniowych.

7. Automatyzacja:

- a. **Azure Logic Apps:** Automatyzowanie reakcji na wykryte anomalie (np. powiadomienie zespołu deweloperskiego lub automatyczne uruchomienie procedur naprawczych).

8. Bezpieczeństwo:

- a. **Azure Security Center:** Monitorowanie i zabezpieczanie danych, ochrony przed nieautoryzowanym dostępem.

- b. **Azure Key Vault:** Przechowywanie kluczy API, certyfikatów i innych danych wrażliwych.