

Najważniejsze narzędzia analityczne / Big Data / AI w chmurze Microsoft Azure:

Składowanie danych:

- Azure Data Lake Storage Gen2 – skalowalna hurtownia danych zoptymalizowana pod analizę dużych zbiorów danych.

ETL / ELT:

- Azure Data Factory – umożliwia tworzenie pipeline'ów do przetwarzania i integracji danych z różnych źródeł.

Compute / analityka:

- Azure Synapse Analytics – rozwiązanie łączące hurtownię danych z silnikiem analitycznym (SQL + Spark).
- Azure Databricks – platforma oparta na Apache Spark, idealna do analizy danych, ML i przetwarzania Big Data.

AI / ML:

- Azure Machine Learning – kompletna platforma do budowy, trenowania, testowania i wdrażania modeli AI/ML.

Monitoring:

- Azure Monitor oraz Log Analytics – umożliwiają monitorowanie infrastruktury, alertowanie i zbieranie logów.

IoT / dane strumieniowe:

- Azure IoT Hub – do odbierania danych z urządzeń brzegowych.
- Azure Stream Analytics – analiza danych strumieniowych w czasie rzeczywistym.

Wizualizacja danych:

- Power BI – narzędzie do tworzenia raportów i dashboardów z danych przetworzonych w Azure.

Przykład 1: Proof of Concept – Wykrywanie anomalii na linii produkcyjnej (IoT)

Cel: Wykrywanie anomalii w danych z czujników IoT (np. temperatura, wibracje, ciśnienie) w czasie rzeczywistym, aby zapobiegać awariom.

Architektura rozwiązania:

1. Zbieranie danych z czujników IoT przy pomocy Azure IoT Hub.
2. Przetwarzanie danych strumieniowych:
 - Użycie Azure Stream Analytics do wstępnej analizy i przekształcania danych w czasie rzeczywistym.
 - Alternatywnie – zastosowanie Azure Functions z własną logiką analizy.
3. Składowanie danych:
 - Dane surowe zapisywane w Azure Data Lake Storage Gen2.
 - Metadane i wyniki analizy przechowywane w Azure SQL Database.
4. Budowa i wdrożenie modelu AI:
 - Model trenowany w Azure Machine Learning (np. do wykrywania anomalii).
 - Model udostępniany jako REST API, zintegrowany np. z Azure Function.
5. Monitoring i alertowanie:
 - Użycie Azure Monitor do śledzenia anomalii i generowania alertów (e-mail, SMS).
6. Wizualizacja danych i raportowanie:
 - Stworzenie dashboardów w Power BI prezentujących stan urządzeń i historię anomalii.

Przykład 2: Architektura Big Data dla Rockstar Games

Cel: Zaprojektowanie architektury przetwarzania danych telemetrycznych z gier (np. GTA Online), aby:

- wykrywać błędy i anomalie w działaniu gry,
- analizować zachowania i satysfakcję graczy,
- zapobiegać oszustwom (cheaty, exploits),
- przewidywać odejścia graczy (churn prediction).

Elementy architektury:

1. Zbieranie danych:

- Zdarzenia w czasie rzeczywistym (np. logi serwera, akcje gracza) trafiają do Azure Event Hubs.
- Dane batchowe (np. z baz danych, logów) pobierane przez Azure Data Factory.

2. Składowanie danych:

- Wszystkie dane trafiają do Azure Data Lake Gen2 – podzielone na warstwy: raw, curated, processed.

3. Przetwarzanie danych:

- Azure Synapse Analytics – przetwarzanie danych SQL-owych (analizy historyczne, hurtownia danych).
- Azure Databricks – analiza złożonych danych i trenowanie modeli ML przy użyciu Apache Spark.

4. Uczenie maszynowe:

- Azure Machine Learning służy do trenowania i wdrażania modeli:
 - wykrywania błędów i nieprawidłowych wzorców zachowania graczy (anomaly detection),
 - przewidywania odejścia gracza (model churn prediction),
 - analizy sentymentu wypowiedzi graczy na czacie, forach (NLP).

5. Monitoring i bezpieczeństwo:

- Azure Monitor oraz Application Insights do monitorowania serwerów i działania gry.

- Azure Sentinel do wykrywania nieautoryzowanych działań i ochrony przed oszustwami.

6. Raportowanie i wizualizacja:

- Power BI – tworzenie dashboardów dla zespołów developerskich, QA i marketingu.