

Azure Data Lake Storage Gen2

To skalowalny magazyn danych Big Data zgodny z Hadoopem, wykorzystywany do przechowywania surowych, nieustrukturalnych i półstrukturalnych danych. Sprawdza się wszędzie tam, gdzie potrzebujesz przechować ogromne zbiory danych (logi, pliki, JSON, CSV, audio) z możliwością integracji ze Spark, Hive czy Databricks.

Zalety:

- Działa z narzędziami Big Data (np. Spark, Hive)
 - Daje dokładną kontrolę dostępu do plików
 - Przechowuje różne typy danych (CSV, JSON, logi, audio)
 - Szybki dostęp do danych
-

Azure Data Factory (ADF)

To narzędzie do bezkodowego budowania i orkiestracji potoków danych (ETL/ELT), szczególnie przydatne gdy integrujesz wiele źródeł i potrzebujesz przetwarzać dane regularnie (np. harmonogramy, triggerowane przepływy). Idealne do łączenia danych z REST API, SQL, Data Lake i ładowania ich do hurtowni lub BI.

Zalety:

- Tworzenie procesów ETL bez pisanie kodu
 - Łatwe łączenie z wieloma źródłami danych
 - Można ustawić harmonogramy i automatyzację
 - Pokazuje błędy i postęp działania
-

Azure Databricks

To platforma analityczna z Apache Spark, wykorzystywana do zaawansowanego przetwarzania danych i tworzenia modeli ML. Sprawdza się w projektach opartych na notebookach, analizie danych w batch/stream oraz we współpracy zespołowej.

Zalety:

- Obsługuje wiele języków (Python, SQL, R...)
 - Świetne środowisko do analizy i ML
 - Automatycznie dostosowuje moc obliczeniową
 - Działa zespołowo w notatnikach
 - Używa Delta Lake (łatwe wersjonowanie danych)
-

Azure Stream Analytics

To serverless'owa usługa do przetwarzania danych strumieniowych w czasie rzeczywistym, świetna do analizowania danych IoT, telemetrycznych, logów i wykrywania anomalii. Wykorzystywana w systemach monitorowania i alertowania.

Zalety:

- Proste zapytania w SQL
 - Pracuje w czasie rzeczywistym (streaming)
 - Integruje się z IoT i Power BI
 - Nie trzeba nic konfigurować – działa od razu
-

Azure Synapse Analytics

To nowoczesna hurtownia danych łącząca moc SQL i Spark z możliwością przetwarzania danych na żądanie lub z wykorzystaniem klastrów. Idealna do budowy raportów, integracji z BI oraz analizy dużych zbiorów danych z Data Lake.

Zalety:

- Można wybrać: stałe zasoby lub zapytania na żądanie
 - Łączy SQL, Spark i BI
 - Wspiera szybkie ładowanie danych (COPY INTO)
 - Przetwarza dane równolegle – szybciej przy dużych zbiorach
-

Microsoft Fabric (Azure MS Fabric)

To zintegrowana platforma do analityki danych, która łączy Power BI, Synapse i Data Factory w jednym środowisku. Świetnie sprawdza się w projektach end-to-end – od ładowania danych, przez modelowanie, po wizualizację.

Zalety:

- Wszystko w jednym miejscu (Power BI, dane, modele)
 - Można pracować bez kodowania
 - Obsługuje SQL, DAX, PySpark
 - Działa z Git i automatyzacją
-

Azure Monitor

To usługa do zbierania metryk i logów z zasobów Azure oraz tworzenia alertów i dashboardów. Pomaga w monitorowaniu infrastruktury, aplikacji, zasobów oraz całych środowisk chmurowych, identyfikując problemy i reagując automatycznie.

Zalety:

- Zbiera dane z aplikacji i serwerów
 - Automatycznie wysyła alerty
 - Pokazuje wykresy i błędy
 - Można analizować dane historyczne
-

Azure Security Center

To centrum bezpieczeństwa chmurowego, które wykrywa zagrożenia, ocenia poziom zabezpieczeń i proponuje działania naprawcze. Szczególnie przydatne do ochrony VM, kontenerów i usług PaaS przed atakami.

Zalety:

- Pokazuje, jak dobrze zabezpieczone są Twoje zasoby
 - Chroni przed zagrożeniami (np. brute force)
 - Daje zalecenia jak poprawić bezpieczeństwo
 - Działa z Defenderem i PaaS
-

Azure RBAC + POSIX ACLs (Data Lake Access)

To systemy zarządzania dostępem do danych i zasobów – RBAC działa na poziomie ról użytkowników, a POSIX ACL pozwala ustawić szczegółowy dostęp do plików i folderów. Stosowane w Data Lake i Storage Accounts, gdy potrzebna jest kontrola kto widzi i modyfikuje dane.

Zalety:

- Nadajesz uprawnienia użytkownikom i aplikacjom
 - Możesz określić kto widzi i edytuje konkretne pliki
 - Działa na wielu poziomach (plik, folder, kontener)
-

PolyBase

To technologie do szybkiego ładowania danych z plików (CSV, Parquet) do Synapse Analytics. COPY INTO to prostsza i szybsza alternatywa dla PolyBase, przydatna w hurtowniach i lakehouse.

Zalety:

- Szybko ładuje dane do Synapse
 - Obsługuje różne formaty plików (CSV, Parquet)
 - COPY INTO jest prostsze niż PolyBase
 - Przetwarza dane równolegle – szybciej
-

Azure Databricks

To warstwowa architektura danych w stylu Lakehouse, która zapewnia transakcyjność i wersjonowanie danych. Przydatna gdy przetwarzasz dane batch + stream w jednym miejscu i chcesz mieć kontrolę nad jakością.

Zalety:

- Dane są uporządkowane warstwami (raw → czyste → agregaty)
- Można cofnąć zmiany i kontrolować wersje
- Obsługuje jednocześnie batch i streaming
- Działa z popularnymi magazynami danych