Zajęcia 5

Zadanie 1

**Cel zadania – przygotować się do interview**,

Notatka z krótkim opisem najważniejszych narzędzi analitycznych/Big Data/AI. Móc opisać **wybrane** przez siebie zagadnienie/narzędzia analityczne **w wybranej** chmurze Azure/AWS/GCP/IBM/Alibaba.

Przykład

1. Robisz PoC na wykrycie anomalii z linii produkcyjnej (jakieś IoT). Jakich narzędzi użyjesz np. w Azure. Jak będzie wyglądało PoC.
2. Zaproponuj architekturę Big Data dla Rockstar Games przetwarzanie danych w celu wykrycia i zapobiegania błędom w grze, satysfakcji użytkowników. Jakie narzędzia, gdzie będzie compute ect…

Podpowiedź - przejrzyj co jest wymagane do certyfikacji np. DP-203, i DP-100, AI-102

<https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3VEHg>

<https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4MbYT>

<https://learn.microsoft.com/pl-pl/credentials/certifications/resources/study-guides/ai-102>

2.

**Azure IOT Hub**

-platforma w chmurze, która umożliwia łączenie, monitorowanie, zarządzanie i kontrolę miliardów urządzeń IoT. Działa ona jako centralny punkt komunikacji między aplikacją IoT a podłączonymi urządzeniami, zapewniając niezawodną i bezpieczną komunikację

-zbieranie danych z logów, opinii czy danych użytkowników

**Azure Data Factory**

-usługa w chmurze, która umożliwia automatyzację przepływu danych i integrację różnorodnych źródeł danych. Działa jako centralna platforma do tworzenia, zarządzania i monitorowania potoków danych, które przenoszą dane między różnymi systemami i magazynami danych

-automatyzacja procesu zbierania danych z różnych źródeł i ich integracji w hurtowni danych

**Azure Data Lake Storage**

-przeznaczony do przechowywania ogromnych ilości danych (petabajty, potencjalnie nawet eksabajty) w ich pierwotnej postaci, organizuje dane w hierarchicznej strukturze plików i folderów, co ułatwia zarządzanie danymi, jest specjalnie zaprojektowany do obsługi procesów analitycznych wielkich danych i uczenia maszynowego. Oferuje wysoce skalowalne przetwarzanie danych i wysoką przepustowość

-skalowalne i bezpieczne przechowywanie dużych ilości danych

**Azure SQL Database**

-usługa zarządzanej bazy danych, która zapewnia klasyczny model relacyjnej bazy danych, podobny do tradycyjnych serwerów SQL. Skupia się na przechowywaniu danych strukturalnych oraz na przetwarzaniu transakcyjnym. Dba o kwestie wysokiej dostępności, tworzenia kopii zapasowych, monitoringu, automatycznego skalowania. Zgodny ze standardami języka SQL, co ułatwia migrację istniejących aplikacji bazodanowych i wykorzystanie narzędzi zgodnych z SQL

-przechowywanie danych strukturyzowanych, które wymagają szybkiego dostępu i przetwarzania transakcyjnego.

**Azure HDInsight**

-w pełni zarządzana usługa w chmurze oferująca klastry takich silników Big Data jak: Apache Hadoop, Spark, Kafka, Storm, HBase. Posiada wsparcie dla różnych zastosowań w obszarze Big Data, w tym ETL (Extract-Transform-Load), analiz z użyciem dużych zbiorów danych, uczenie maszynowe, czy raportowanie. Zintegrowana z usługami Azure takimi jak Azure Data Lake Storage, Azure Blob Storage, Azure Data Factory ułatwia budowę kompleksowych rozwiązań big data

-przetwarzanie i analiza dużych zbiorów danych za pomocą frameworków takich jak Spark

**Azure Databricks**

-usługa oparta o Apache Spark, dostosowana do wykorzystania w środowisku chmury i ukierunkowana na analizę danych, uczenie maszynowe oraz sztuczną inteligencję. Umożliwia prowadzenie prac analitycznych oraz eksperymentów z danymi w jednym środowisku, z wykorzystaniem języków Python, Scala, R, SQL. Bezproblemowa integracja z innymi usługami chmury, w tym Azure Data Lake Storage, Azure Blob Storage, Azure Machine Learning

-przetwarzanie i analiza danych za pomocą języka Python i Spark

**Azure Stream Analytics**

-pozwala na analitykę danych przesyłanych strumieniowo niemalże w czasie rzeczywistym. Oferuje deklaratywny język zapytań podobny do klasycznego SQL-a, dzięki czemu nauka i wykorzystanie są stosunkowo proste. Posiada wbudowane funkcje przystosowane do analizy czasowej, geoprzestrzennej i przetwarzania szeregów czasowych

-przetwarzanie danych strumieniowych w czasie rzeczywistym i wykrywania anomalii

**Azure Machine Learning**

-możliwość wykorzystania interfejsu graficznego lub środowiska Python, aby tworzyć modele uczenia maszynowego bez potrzeby pisania skomplikowanego kodu. Trenowanie modeli na potężnych klastrach obliczeniowych Azure bez konieczności zarządzania infrastrukturą IT. Szybkie wdrażanie modeli oraz gotowe szablony do typowych zadań uczenia maszynowego, takich jak prognozowanie, klasyfikacja i rozpoznawanie obrazów

- tworzenie, trenowanie i wdrażanie modeli uczenia maszynowego do przewidywania błędów w grze, satysfakcji użytkowników i innych celów biznesowych

**Azure Cognitive Services**

-dodawani funkcji AI do aplikacji we względnie przystępny sposób, za pomocą analizy tekstu (emocje, podmioty, pojęcia), rozpoznawania obrazów(wykrywanie obiektów, twrzy), wykrywanie i przetwarzanie mowy

-dodawanie funkcji sztucznej inteligencji do aplikacji, takich jak analiza tekstu, rozpoznawanie obrazu i tłumaczenie maszynowe

**Azure Power BI**

-zapewniająca analitykę samoobsługową typu Business Intelligence. Pozwala łączyć się z wieloma źródłami danych, przekształcać i wizualizować te dane oraz tworzyć atrakcyjne wizualnie i czytelne raporty oraz interaktywne pulpity nawigacyjne (dashboardy), które można z łatwością udostępniać w organizacji

-tworzenie interaktywnych pulpitów nawigacyjnych i raportów do wizualizacji wyników analizy danych