**Kluczowe narzędzia analityczne i Big Data w chmurze:**

**1. Azure**

* **Azure Machine Learning** - Usługa chmurowa do budowania, trenowania i wdrażania modeli AI. Umożliwia integrację z różnymi narzędziami do analizy danych, jak Power BI i SQL.
* **Azure Databricks** - Wykorzystywane do przetwarzania danych Big Data, w tym analizy danych w czasie rzeczywistym, oferujące łatwą integrację z Apache Spark i Python.
* **Azure Synapse Analytics** - Zintegrowana usługa analityczna, która łączy dane w magazynach danych z możliwością uruchamiania analiz w czasie rzeczywistym.
* **Azure IoT Hub** - Umożliwia zbieranie i przetwarzanie danych z urządzeń IoT, w tym wykrywanie anomalii.
* **Azure Stream Analytics** - Platforma do analizy danych strumieniowych, np. z urządzeń IoT, w czasie rzeczywistym.
* **Power BI** - Narzędzie do wizualizacji i analizy danych w formie wykresów i raportów.

**2. AWS**

* **Amazon SageMaker** - Usługa do budowania, trenowania i wdrażania modeli ML, szczególnie przydatna w automatyzacji procesów analitycznych.
* **AWS Lambda** - Narzędzie do uruchamiania kodu w odpowiedzi na zdarzenia, bez potrzeby zarządzania serwerami. Może być użyteczne w analizie danych w czasie rzeczywistym.
* **Amazon Kinesis** - Usługa do przetwarzania danych strumieniowych (np. IoT) w czasie rzeczywistym.
* **Amazon Redshift** - Hurtownia danych do przechowywania i analizowania dużych zbiorów danych.
* **AWS IoT Core** - Usługa do zbierania i analizy danych z urządzeń IoT.

**3. GCP (Google Cloud Platform)**

* **Google BigQuery** - Usługa do przechowywania i analizowania danych w chmurze, wspiera SQL do zapytań analitycznych w Big Data.
* **Google Cloud Machine Learning Engine** - Usługa do budowy i wdrażania modeli AI, w tym wykrywania anomalii i klasyfikacji.
* **Google Cloud IoT Core** - Usługa do zarządzania urządzeniami IoT i przetwarzania danych.
* **Google Cloud Dataproc** - Usługa do uruchamiania klastrów Apache Hadoop i Spark, umożliwiająca przetwarzanie dużych zbiorów danych.
* **Google Dataflow** - Narzędzie do analizowania i przetwarzania strumieniowych danych.

**Przykład 1: Wykrycie anomalii z linii produkcyjnej w Azure**

**Cel: Wykrycie anomalii w danych pochodzących z czujników IoT na linii produkcyjnej (np. zmiany w temperaturze, ciśnieniu, prędkości produkcji).**

**Narzędzia i architektura:**

1. **Azure IoT Hub** - Zbieranie danych z czujników IoT na linii produkcyjnej.
2. **Azure Stream Analytics** - Przetwarzanie danych strumieniowych z urządzeń IoT w czasie rzeczywistym, np. wykrywanie anomalii w danych strumieniowych (np. zmiany przekraczające określone limity).
3. **Azure Machine Learning** - Tworzenie i wdrażanie modelu do detekcji anomalii w danych, np. klasyfikacja wartości "normalnych" vs. "anomalii".
4. **Power BI** - Wizualizacja danych i wyników analizy, np. w formie wykresów, raportów na temat wykrytych anomalii w czasie rzeczywistym.

**PoC:**

* PoC może polegać na zbieraniu danych z urządzeń IoT, które będą przesyłane do **Azure IoT Hub**.
* Następnie dane będą przekazywane do **Azure Stream Analytics**, gdzie zostaną wstępnie przetworzone (np. filtry, agregacje).
* Zastosowanie modelu wykrywania anomalii w **Azure Machine Learning** pozwoli zidentyfikować, kiedy dane wykraczają poza standardowe normy, co może wskazywać na problem w produkcji.
* Z wyników analizy stworzymy raporty w **Power BI**, gdzie zespół produkcyjny będzie mógł monitorować wyniki na bieżąco.

**Przykład 2: Architektura Big Data dla Rockstar Games**

**Cel: Przetwarzanie danych z gry w celu wykrycia i zapobiegania błędom w grze oraz analizy satysfakcji użytkowników (np. analiza crashy, analiza preferencji graczy).**

**Narzędzia i architektura:**

1. **Azure Databricks** - Zbieranie i przetwarzanie danych z gier w czasie rzeczywistym. Umożliwia obróbkę dużych danych w oparciu o Apache Spark.
2. **Azure Synapse Analytics** - Integracja z hurtownią danych, analiza danych pochodzących z gier, takich jak logi błędów, dane o sesjach graczy.
3. **Azure Machine Learning** - Użycie algorytmów do analizy satysfakcji graczy (np. analiza nastrojów z komentarzy) oraz prognozowania potencjalnych błędów (np. crashy).
4. **Power BI** - Wizualizacja wyników analizy, np. raporty o błędach, preferencjach graczy, ocenach.
5. **Azure Blob Storage** - Przechowywanie danych z logów, obrazów, sesji gier.

**Architektura:**

1. Dane z gry (np. logi, wyniki sesji graczy) są przesyłane do **Azure Blob Storage**.
2. Z tych danych, za pomocą **Azure Databricks**, przetwarzamy i analizujemy informacje, takie jak występowanie błędów, crashy czy preferencji graczy.
3. Przeprowadzamy analizę danych w **Azure Synapse Analytics**, która jest magazynem danych i pozwala na analizowanie dużych zbiorów danych.
4. Wykorzystujemy **Azure Machine Learning** do tworzenia modeli predykcyjnych na podstawie danych historycznych, przewidując możliwe problemy (np. błędy w grze).
5. Wyniki analizy prezentujemy w **Power BI**, tworząc interaktywne raporty, które pomagają zespołowi deweloperskiemu identyfikować problemy i optymalizować doświadczenie graczy.

**Przygotowanie do certyfikacji**

Aby przygotować się do certyfikacji takich jak **DP-203** (Data Engineering), **DP-100** (AI Engineer), czy **AI-102** (AI Solution Architect), ważne jest, aby mieć dogłębną wiedzę na temat narzędzi analitycznych i Big Data w wybranej chmurze. Przykłady narzędzi i architektur podane powyżej będą częścią Twojej wiedzy praktycznej, którą możesz zastosować do rozwiązywania problemów w testach certyfikacyjnych. Kluczowe tematy to:

* **Przechowywanie danych i Big Data** (np. Azure Blob Storage, AWS S3, GCP BigQuery).
* **Przetwarzanie danych w czasie rzeczywistym** (np. Stream Analytics, AWS Kinesis, GCP Dataflow).
* **Budowanie i trenowanie modeli AI/ML** (np. Azure Machine Learning, SageMaker, Google ML).

Sukces w tych certyfikacjach będzie wymagał znajomości narzędzi chmurowych oraz umiejętności ich integracji w rozwiązania analityczne.