1. Lista najważniejszych narzędzi analitycznych (7) i dodaj linki do dokumentacji dla Azure i AWS:

<https://www.dataversity.net/eight-big-data-analytics-options-on-microsoft-azure/#>

Microsoft Azure:

#### Azure Synapse Analytics

<https://azure.microsoft.com/en-us/services/synapse-analytics/>

Umożliwia ładowanie dowolnej liczby źródeł danych — zarówno relacyjnych, jak i nierelacyjnych baz danych, zarówno lokalnie, jak i w chmurze Azure. Ujednolica wszystkie dane i pozwala je przetwarzać i analizować za pomocą języka SQL. Ponadto udostępnia Azure Synapse Studio, które oferuje przestrzeń roboczą do analizy dużych zbiorów danych i zadań AI oraz tworzy atrakcyjne wizualizacje danych.

#### Azure Databricks

<https://azure.microsoft.com/en-us/services/databricks/>

Databricks to usługa analityczna oparta na Apache Spark. Apache Spark to tradycyjne narzędzie wykorzystywane do przetwarzania ogromnych ilości nieustrukturyzowanych danych z dużą prędkością. Databricks obsługuje języki takie jak Python, Scala, Java, SQL i R, a także biblioteki AI/ML, takie jak TensorFlow i PyTorch, umożliwiając pracę z danymi Spark przy użyciu dowolnego z tych języków i platform. Ponadto Databricks integruje się z Azure Machine Learning, zapewniając dostęp do dużej liczby wstępnie wytrenowanych algorytmów uczenia maszynowego.

Databricks umożliwia konfigurowanie zarządzanych klastrów Apache Spark z automatycznym skalowaniem i automatycznym zakończeniem, eliminując złożoność konfigurowania Spark w lokalnym centrum danych.

#### Azure HDInsight

<https://azure.microsoft.com/en-us/services/hdinsight/>

Umożliwia wykonywanie złożonych, rozproszonych zadań analitycznych na praktycznie dowolnej objętości danych. Usługa HDInsight umożliwia szybkie tworzenie klastrów danych big data przy użyciu usługi Hadoop i skalowanie ich w górę lub w dół w zależności od potrzeb. Integruje się z innymi usługami platformy Azure, takimi jak Data Factory i Data Lake Storage, umożliwiając zastosowanie analizy Hadoop do danych, które już tam masz.

Usługa HDInsight zawiera pełny zestaw popularnych narzędzi Hadoop, w tym Apache Spark, Apache Kafka, HBase, Hive i Storm. Zapewnia infrastrukturę na skalę korporacyjną w postaci monitorowania, bezpieczeństwa, zgodności i wysokiej dostępności.

#### 

#### Azure Data Factory

<https://azure.microsoft.com/en-us/services/data-factory/>

Azure Data Factory to usługa Extract Transform Load (ETL). ETL to termin oznaczający narzędzia do przetwarzania danych strukturalnych na dużą skalę (pozyskanie danych do baz danych). Proces ETL pobiera ustrukturyzowaną bazę danych, czyści ją i konwertuje dane do formatu odpowiedniego do analizy. Data Factory pomaga tworzyć strategie ETL bez kodu i konfiguracji, za pomocą edytora wizualnego.

Data Factory udostępnia wbudowane łączniki z ponad 90 źródłami danych, w tym Amazon S3, Google BigQuery i wieloma lokalnymi źródłami danych. Można również skopiować dane z Data Factory do usługi Azure File Storage.

#### Azure Machine Learning

<https://azure.microsoft.com/en-us/services/machine-learning/>

Jest to ogromna biblioteka gotowych, przeszkolonych algorytmów uczenia maszynowego. Zapewnia również środowisko do korzystania z tych algorytmów i stosowania ich do zadań w świecie rzeczywistym. Azure ML przyspiesza tworzenie modeli dzięki wygodnemu interfejsowi użytkownika uczenia maszynowego, który umożliwia tworzenie potoków uczenia maszynowego łączących wiele algorytmów z krokami, takimi jak uczenie modeli, testowanie i ocena.

Ponadto usługa Azure ML zapewnia rozwiązania umożliwiające interpretację sztucznej inteligencji. Obejmuje wizualizację i inne dane, które mogą pomóc zrozumieć zachowanie modelu, zastosować metryki sprawiedliwości i dokonać porównań między algorytmami w celu zrozumienia najlepszego wariantu do wyboru.

#### Azure Stream Analytics

<https://azure.microsoft.com/en-us/services/stream-analytics/>

Azure Stream Analytics umożliwia tworzenie potoku end-to-end do przesyłania strumieniowego danych. Stream Analytics umożliwia zdefiniowanie potoku analizy do przesyłania strumieniowego danych, zdefiniowanego przy użyciu składni SQL i przejściem do produkcji w ciągu kilku minut. Skaluje się elastycznie w zależności od objętości przesyłanych strumieniowo danych.

Ponieważ przesyłanie strumieniowe danych często wymaga bardzo wydajnego przetwarzania i odpowiedzi w czasie rzeczywistym, usługa Azure Stream Analytics oferuje opóźnienie poniżej sekundy z gwarantowanym przetwarzaniem zdarzeń „dokładnie raz”. Oferuje również dostępność na poziomie 99,9%.

#### Data Lake Analytics

<https://azure.microsoft.com/en-us/services/data-lake-analytics/>

Azure Data Lake Analytics umożliwia tworzenie programów do transformacji danych przy użyciu różnych języków, w tym U-SQL (specjalny język dostarczany przez firmę Microsoft, który łączy zalety SQL i C#), Python, .NET i R. Może przetwarzać petabajty danych .

Usługa Data Lake Analytics różni się od usługi Azure Synapse Analytics tym, że nie pobiera wszystkich danych do usługi Data Lake, a następnie nie przetwarza ich. Zamiast tego łączy się ze źródłami danych opartymi na platformie Azure, takimi jak Azure Data Lake Storage, i przeprowadza analizę w locie przy użyciu dostarczonego przez Ciebie kodu.

Aws:

* Interactive analytics. Amazon Athena.

<https://aws.amazon.com/athena/?c=a&sec=srv&whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc>

Amazon Athena to interaktywna usługa zapytań, która ułatwia analizowanie danych w Amazon S3 przy użyciu standardowego SQL. Athena jest bezserwerowa, więc nie ma infrastruktury do zarządzania i płacisz tylko za uruchomione zapytania.

Athena jest łatwa w użyciu.Wskaż swoje dane w Amazon S3, zdefiniuj schemat i rozpocznij wykonywanie zapytań przy użyciu standardowego SQL. Dzięki Athena nie ma potrzeby wykonywania złożonych zadań ETL w celu przygotowania danych do analizy (ułatwienie dla osób nie znających SQL).

Athena jest zintegrowana z katalogiem danych AWS Glue, co pozwala na tworzenie ujednoliconego repozytorium metadanych w różnych usługach, przeszukiwanie w celu wykrywania schematów i zapełnianie katalogu nowymi i zmodyfikowanymi definicjami tabel i partycji oraz utrzymywanie schematu, wersjonowanie.

* Big data processing. Amazon EMR.

<https://aws.amazon.com/emr/?c=a&sec=srv>

Amazon EMR to platforma Big Data w chmurze do uruchamiania zadań rozproszonego przetwarzania danych na dużą skalę, interaktywnych zapytań SQL i aplikacji uczenia maszynowego (ML) przy użyciu platform analitycznych typu open source, takich jak Apache Spark, Apache Hive i Presto.

* Data warehousing. Amazon Redshift.

<https://aws.amazon.com/redshift/?c=a&sec=srv>

Amazon Redshift wykorzystuje SQL do analizowania ustrukturyzowanych i częściowo ustrukturyzowanych danych w hurtowniach danych, operacyjnych bazach danych i jeziorach danych.

* Interactive analytics. Amazon Kinesis.

<https://aws.amazon.com/kinesis/?c=a&sec=srv>

Amazon Kinesis ułatwia zbieranie, przetwarzanie i analizowanie danych przesyłanych strumieniowo w czasie rzeczywistym, dzięki czemu możesz uzyskać aktualne informacje i szybko reagować na nowe informacje. Dzięki Amazon Kinesis możesz pozyskiwać dane w czasie rzeczywistym, takie jak wideo, audio, dzienniki aplikacji, strumienie kliknięć witryn internetowych i dane telemetryczne IoT na potrzeby uczenia maszynowego, analiz i innych aplikacji. Umożliwia przetwarzanie i analizowanie danych zaraz po ich otrzymaniu oraz natychmiastowe reagowanie, zamiast czekać na zebranie wszystkich danych przed rozpoczęciem przetwarzania.

* Operational analytics. Amazon OpenSearch Service.

<https://aws.amazon.com/opensearch-service/?c=a&sec=srv>

Amazon OpenSearch Service ułatwia wykonywanie interaktywnych analiz logów, monitorowanie aplikacji w czasie rzeczywistym, wyszukiwanie witryn internetowych i nie tylko. OpenSearch to pakiet rozproszonego wyszukiwania i analizy typu open source, wywodzący się z Elasticsearch.

* Dashboards and visualizations. Amazon Quicksight.

<https://aws.amazon.com/quicksight/?c=a&sec=srv>

Amazon QuickSight pozwala wszystkim w Twojej organizacji zrozumieć Twoje dane, zadając pytania w języku naturalnym, eksplorując interaktywne pulpity nawigacyjne lub automatycznie szukając wzorców i wartości odstających opartych na uczeniu maszynowym.

Umożliwia zadawanie pytań konwersacyjnych dotyczących danych i korzystanie z silnika opartego na ML, aby otrzymywać odpowiednie wizualizacje bez czasochłonnego przygotowywania danych od autorów i administratorów.

* Visual data preparation. AWS Glue DataBrew.

<https://aws.amazon.com/glue/features/databrew/?c=a&sec=srv>

AWS Glue DataBrew tonarzędzie do wizualnego przygotowywania danych, które ułatwia analitykom czyszczenie i normalizowanie danych w celu przygotowania ich do analizy i uczenia maszynowego. Ponad 250 gotowych przekształceń, aby zautomatyzować zadania związane z przygotowywaniem danych, a wszystko to bez konieczności pisania kodu. Zautomatyzowane filtrowanie anomalii, konwertowanie danych do standardowych formatów oraz poprawianie nieprawidłowych wartości i inne zadania.