\*\*1)\*\*

Teraz porozmawiamy o zastosowaniu spersonalizowanej reklamy na przykładzie jednego z największych sklepów w sektorze e-commerce – ASOS. Omówię mechanizm pracy z danymi na tej platformie. W rzeczywistości większość innych firm, od największych, takich jak Amazon czy eBay, po mniejsze, lokalne sklepy, korzysta z podobnych rozwiązań do budowania swoich potoków danych. Potok danych (pipeline) to zautomatyzowany proces przenoszenia danych między różnymi systemami i etapami przetwarzania. Potok obejmuje kroki związane z ekstrakcją, transformacją i ładowaniem danych (ETL – Extract, Transform, Load).

\*\*2)\*\*

Zacznijmy od Extract – skąd ASOS pobiera dane o swoich użytkownikach:

- \*\*Dane o zakupach\*\*: Wewnętrzne bazy danych, CRM (format: dane strukturalne).

- \*\*Dane o zachowaniu użytkowników\*\*: Analityka internetowa, dzienniki serwera (format: surowe logi, JSON, CSV).

- \*\*Dane z aplikacji mobilnej\*\*: Interakcje z aplikacją (format: JSON, zdarzenia).

- \*\*Dane z mediów społecznościowych\*\*: Integracje z Facebookiem, Instagramem (format: dane tekstowe).

- \*\*Dane o interakcji z marketingiem\*\*: Kampanie e-mailowe, reklama płatna (format: dane strukturalne).

\*\*3)\*\*

Jakie narzędzia wykorzystuje zespół inżynierów danych do budowy potoku danych? Korzystają z rozwiązań chmurowych firmy Microsoft, czyli z usług Azure. Nie wszystkie firmy używają Azure – istnieją również rozwiązania od Amazon (AWS) oraz Google. Te usługi są dość podobne pod względem logiki i funkcji, różnią się jednak nazwami i strukturą.

Omówmy teraz usługi, których używa ASOS:

- \*\*Azure Event Hubs\*\*

- Do czego jest używana: Azure Event Hubs działa jako platforma do przechwytywania i przetwarzania danych w czasie rzeczywistym. ASOS wykorzystuje ją do zbierania zdarzeń, takich jak działania użytkowników na stronie i w aplikacji mobilnej (kliknięcia, przeglądanie stron, dodawanie produktów do koszyka, interakcje z reklamami).

- Jak pomaga: Umożliwia przetwarzanie danych zaraz po ich otrzymaniu, zapewniając natychmiastowe dostosowanie ofert reklamowych i personalizację na podstawie aktualnych działań użytkownika.

- \*\*Azure Data Lake\*\*

- Do czego jest używana: Azure Data Lake to centralne repozytorium do przechowywania danych strukturalnych i niestrukturalnych. ASOS przechowuje tam duże ilości danych, takich jak informacje o zakupach, działania użytkowników, dane z aplikacji mobilnych i inne źródła.

- Jak pomaga: Zapewnia wygodne i skalowalne przechowywanie danych do analizy, z którego dane można łatwo pobierać i wykorzystywać do analityki i uczenia maszynowego.

- \*\*Azure Synapse Analytics\*\*

- Do czego jest używana: Azure Synapse Analytics pozwala ASOS integrować dane z różnych źródeł i wykonywać zaawansowane zadania analityczne. System wspiera procesy ETL (ekstrakcja, transformacja i ładowanie danych) i zapewnia możliwości tworzenia skalowalnych rozwiązań analitycznych.

- Jak pomaga: Pozwala łączyć dane o zachowaniach użytkowników, zakupach, informacje demograficzne i inne metryki, co pomaga tworzyć dokładniejsze rekomendacje i analizować efektywność kampanii reklamowych.

- \*\*Azure Databricks\*\*

- Do czego jest używana: Azure Databricks jest wykorzystywana do przetwarzania dużych ilości danych i opracowywania modeli uczenia maszynowego. ASOS używa jej do oczyszczania danych, tworzenia cech (features) oraz testowania i optymalizacji modeli.

- Jak pomaga: Przyspiesza przetwarzanie danych i opracowywanie modeli analitycznych dzięki integracji z Apache Spark, co umożliwia szybkie analizowanie dużych zbiorów danych i tworzenie spersonalizowanych rekomendacji.

- \*\*Azure Machine Learning\*\*

- Do czego jest używana: Azure Machine Learning pozwala ASOS na opracowywanie, trenowanie i wdrażanie modeli uczenia maszynowego do przewidywania zachowań użytkowników i tworzenia spersonalizowanych ofert.

- Jak pomaga: Automatyzuje proces tworzenia modeli do rekomendacji, takich jak kolaboratywna filtracja i przewidywanie preferencji klientów na podstawie danych o zakupach i interakcjach.

Warto dodać, że ASOS przetwarza dane w czasie rzeczywistym, co pozwala na tworzenie rekomendacji również w czasie rzeczywistym – to nazywa się przetwarzaniem strumieniowym (stream-processing). Istnieje także przetwarzanie wsadowe (batch-processing). Jaka jest różnica i jak działają? Mój ulubiony przykład to droga i samochody. Jeśli mamy za zadanie policzyć wszystkie czerwone samochody na drodze, w przypadku stream-processing postawimy osobę z notesem, która będzie liczyć te samochody na bieżąco, gdy będą przejeżdżać – tak robi ASOS. Takie rozwiązanie wymaga jednak większych nakładów. Mniejsze sklepy częściej korzystają z batch-processing. W tym przypadku stawiamy znak stop na drodze, czekamy np. godzinę, a następnie liczymy wszystkie czerwone samochody na drodze, po czym powtarzamy ten proces. Batch-processing może być uruchamiany za pomocą wyzwalaczy, np. o określonej porze lub na podstawie działania użytkownika na stronie, np. dodania produktu do koszyka.

\*\*4)\*\*

Tutaj dodałem wizualizację, aby lepiej zrozumieć usługi Azure. Na schemacie nie są pokazane wszystkie elementy, ale mam nadzieję, że pomoże on trochę w zrozumieniu przepływu danych – od początku, czyli zbierania danych, ich przetwarzania i zapisywania w data lake, aż do końca.

\*\*(omawiam i przechodzę na różne slajdy)\*\*

\*\*8)\*\*

Model wytrenowany na danych można załadować do Azure Machine Learning i wdrożyć jako usługę sieciową. Umożliwia to korzystanie z modelu do prognozowania w czasie rzeczywistym.

\*\*9)\*\*

Gdzie użytkownik będzie mieć kontakt z rekomendacjami? Przede wszystkim na stronie internetowej i w aplikacji, ale także w reklamach na innych stronach, takich jak Facebook, Instagram itd., oraz w reklamowych e-mailach.

\*\*10)\*\*

Jak wdrożenie rekomendacji wpłynęło na ASOS? Przede wszystkim wzrosła satysfakcja klientów, sprzedaż i popularność strony.