

Optymalizacja raportowania i redukcja kosztów operacyjnych – nowoczesna integracja danych dla europejskiego lidera w branży mediów i rozrywki

Klient

Klient to renomowany europejski operator telewizyjny, oferujący szeroki wybór treści premium, w tym filmy, seriale, wydarzenia sportowe oraz produkcje własne. Działając na wielu rynkach międzynarodowych, dostarcza swoje usługi zarówno poprzez tradycyjne kanały telewizyjne, jak i nowoczesne platformy streamingowe. Firma wyróżnia się innowacyjnym podejściem, łącząc zaawansowane technologie z wysokiej jakości produkcjami. Silny nacisk na tworzenie oryginalnych treści sprawia, że jest jednym z liderów branży rozrywki w Europie.

Wyzwanie biznesowe

Celem klienta było opracowanie mechanizmu umożliwiającego przesyłanie danych analitycznych z systemów on-premises do środowiska S3 Bucket w celu zwiększenia wydajności raportowania oraz obniżenia kosztów operacyjnych.

Istniejące rozwiązanie nie wspierało przesyłania danych w formacie Parquet z bazy danych Oracle (on-premises) do chmury AWS, co ograniczało możliwości optymalizacji procesów analitycznych.

Rozwiązanie

- W ramach realizacji zastosowano konektor Debezium na platformie GIFRÖST, umożliwiający przesyłanie danych z bazy Oracle do środowiska Apache Kafka.
- Transformacje oraz operacje typu join na poszczególnych tabelach (topikach) zostały zrealizowane z wykorzystaniem narzędzia KSQL, odpowiedzialnego za przetwarzanie i strumieniowanie danych. Dzięki temu możliwe było prowadzenie analiz w czasie rzeczywistym.
- W kolejnym etapie dane zostały przekształcone do formatu Parquet, wzbogacone o odpowiednie metadane i przesłane do S3 Buckets za pomocą konektora AWS S3.

Rezultaty

- Ograniczenie konieczności przesyłania pełnych rekordów pozwoliło znacząco obniżyć koszty związane z przetwarzaniem danych w chmurze.
- Wstępne przetwarzanie danych z wykorzystaniem platformy Goldenore oraz KSQL przełożyło się na mniejsze obciążenie infrastruktury AWS i redukcję kosztów operacji typu join.
- Ujednolicenie struktury danych przyrostowych usprawniło proces raportowania oraz weryfikacji poprawności pakietów w aplikacji kontrolnej.

Dlaczego GIFRÖST?

Efektywność kosztowa

Funkcjonalność porównywalna z rynkowym liderem w niższej cenie

2-3 sekundowe opóźnienie

Replikacja danych w czasie rzeczywistym

Wysoka wydajność

Obsługa dużych wolumenów danych (do 2TB logów dziennie)

+ 30 połączeń z bazami danych

Działa bezproblemowo z środowiskami chmurowymi oraz on-premise

Bezproblemowa synchronizacja

Replikacja danych bez wpływu na wydajność baz źródłowych oraz docelowych

* GIFRÖST to nowa nazwa platformy, która podczas wdrożenia u klienta funkcjonowała jako Goldenore CDC

