Przetwarzanie strumieniowe danych - Kafka, Flink i AI w czasie rzeczywistym (kod: AI-BIGDATA)

Zapytaj o szczegóły
tel. 22 63 64 164
akademia@alx.pl

Opis i cel szkolenia

Szkolenie wprowadza uczestników w kluczowe zagadnienia związane z przetwarzaniem strumieniowym danych oraz jego zastosowaniem w nowoczesnych systemach analitycznych, AI i rozwiązaniach agentycznych. Uczestnicy poznają różnice między przetwarzaniem wsadowym a strumieniowym, podstawowe elementy architektury streamingowej oraz najważniejsze wyzwania, takie jak opóźnienia, zarządzanie stanem, gwarancje dostarczania danych i niezawodność systemów.

Program obejmuje przegląd najważniejszych technologii i narzędzi, w tym Apache Kafka, Apache Flink, Redpanda, Apache Iceberg oraz modeli wdrożenia w środowiskach self-managed, BYOC i PaaS/SaaS. Szczególny nacisk położony jest na praktyczne aspekty projektowania i implementacji aplikacji strumieniowych, integrację z modelami ML/LLM, monitorowanie przepływu danych oraz zapewnienie zgodności, bezpieczeństwa i ciągłości działania.

Szkolenie kończy się omówieniem studium przypadku pokazującego zastosowanie streamingu w projekcie, np. wykrywaniu fraudów, personalizacji lub integracji systemów AI w czasie rzeczywistym. Dzięki temu uczestnicy zdobędą uporządkowaną wiedzę oraz praktyczne wskazówki potrzebne do wyboru technologii, projektowania architektury i wdrażania rozwiązań strumieniowych w organizacji.

Czas trwania

3 dni

Program

1. Wprowadzenie do przetwarzania strumieniowego
   - Podstawowe pojęcia i definicje
   - Różnice między przetwarzaniem wsadowym a strumieniowym
   - Zastosowania w branżach: finanse, telekomunikacja, retail, IoT
   - Streaming jako fundament systemów AI i agentycznych
2. Architektura systemów przetwarzania strumieniowego
   - Komponenty i ich funkcje (producenci, brokery, konsumenci)
   - Modele przetwarzania: event-time vs. processing-time, okna czasowe
   - Wyzwania: exactly-once semantics, zarządzanie stanem, opóźnienia
   - Unifikacja streaming + batch (architektura lakehouse)
   - Diskless Kafka i Apache Iceberg jako warstwa storage
3. Narzędzia i technologie
   - Apache Kafka — standard protokołu i backbone ekosystemu
   - Apache Flink — dominujący framework przetwarzania strumieniowego
   - Redpanda i alternatywy kompatybilne z protokołem Kafka
   - Apache Storm — kontekst historyczny, dlaczego odszedł w cień
   - Modele wdrożenia: self-managed, BYOC, w pełni zarządzane (PaaS/SaaS)
   - Kryteria wyboru narzędzia
4. Implementacja przetwarzania strumieniowego
   - Tworzenie aplikacji w Flink (Java, Python/PyFlink, SQL)
   - Obsługa zdarzeń i zarządzanie stanem aplikacji
   - Exactly-once semantics i gwarancje dostarczania

- — Optymalizacja wydajności i skalowalności
- — Integracja z modelami ML/LLM — real-time inference

5. Monitorowanie i zarządzanie
   - — Narzędzia do monitorowania przepływu danych
   - — Identyfikacja i rozwiązywanie problemów (Dead Letter Queue i inne wzorce)
   - — Zapewnienie niezawodności: zero data loss, failover między regionami
   - — Governance i compliance w architekturach strumieniowych
6. Studium przypadku
   - — Analiza rzeczywistego projektu (np. wykrywanie fraudów, personalizacja)
   - — Omówienie napotkanych wyzwań i zastosowanych rozwiązań
   - — Przykład integracji streaming + AI w produkcji
   - — Wnioski i najlepsze praktyki

Zapytaj o szczegóły
tel. 22 63 64 164
akademia@alx.pl

Przeznaczenie i wymagania

Szkolenie jest przeznaczone dla osób, które chcą zrozumieć, projektować lub wdrażać nowoczesne systemy przetwarzania danych w czasie rzeczywistym. Szczególnie skorzystają z niego specjaliści pracujący z danymi, architekturą IT, integracją systemów oraz rozwiązaniami AI.

Szkolenie rekomendowane jest dla:

architektów IT i architektów danych, inżynierów danych i developerów backendowych, specjalistów Data Engineering, Big Data i DevOps, zespołów odpowiedzialnych za integrację systemów i platformy danych, analityków technicznych oraz liderów zespołów IT, osób wdrażających rozwiązania AI, ML lub LLM wymagające danych w czasie rzeczywistym, menedżerów technicznych i product ownerów odpowiedzialnych za rozwój systemów opartych na danych.

Uczestnicy szkolenia powinni posiadać podstawową wiedzę z zakresu systemów informatycznych, baz danych oraz przetwarzania danych. Nie jest wymagana wcześniejsza praktyczna znajomość Apache Kafka, Apache Flink czy innych narzędzi streamingowych, ponieważ szkolenie rozpoczyna się od wprowadzenia do kluczowych pojęć i architektury.

Szkolenie jest odpowiednie zarówno dla osób technicznych, które chcą rozpocząć pracę z przetwarzaniem strumieniowym, jak i dla bardziej doświadczonych specjalistów, którzy chcą uporządkować wiedzę oraz poznać aktualne narzędzia, architektury i dobre praktyki wdrożeniowe.

Certyfikaty

Uczestnicy szkolenia otrzymują imienne certyfikaty sygnowane przez ALX.

Lokalizacje

- — Warszawa – ul. Jasna 14/16A
- — Zdalnie – zajęcia realizowane poprzez platformę Zoom
- — Kraków – ul. św. Filipa 23
- — Warsaw (English) – Jasna 14/16A
- — Online (English) – your home, office or wherever you want
- — na życzenie dowolne miejsce w Polsce, lub UE (zajęcia prowadzone w języku angielskim)

Cena szkolenia

2790 PLN netto (VAT 23%)

W cenę szkoleń organizowanych w naszej siedzibie wliczone są:

— autorskie materiały szkoleniowe,
— indywidualne stanowisko komputerowe do pracy podczas zajęć,
— certyfikaty ukończenia szkolenia,
— drobny poczęstunek oraz ciepłe i zimne napoje,
— możliwość jednorazowego kontaktu z instruktorem (instruktorami) po szkoleniu i zadawania pytań dotyczących materiału szkolenia.

Cena szkolenia nie zawiera obiadów. Można je dokupić w cenie 35 zł netto za obiad.

Zapytaj o szczegóły
tel. 22 63 64 164
akademia@alx.pl