

Chmury Obliczeniowe

Prezentacja projektu

Cel projektu

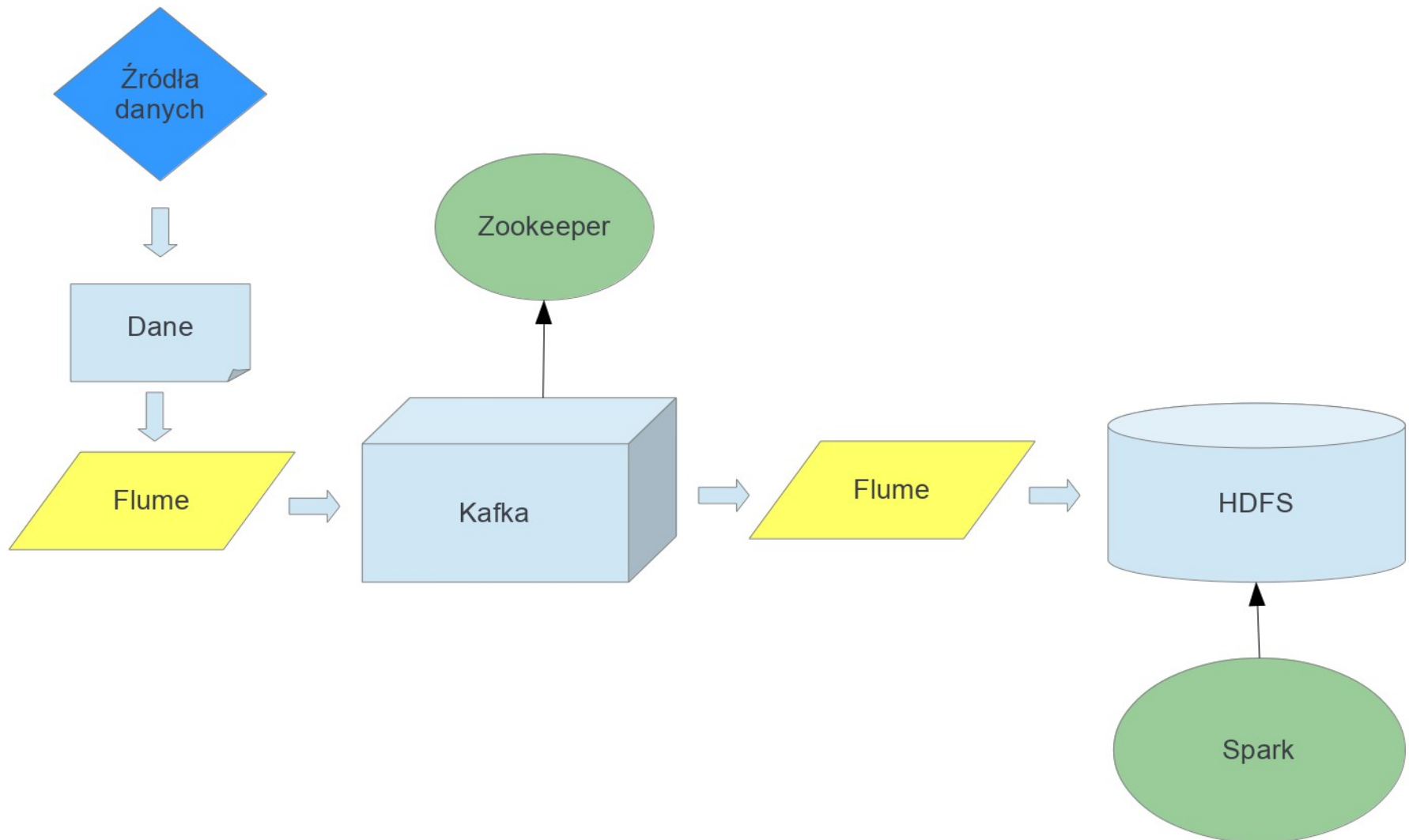
- Stworzyć rozproszony system do przesyłania, przetwarzania i przechowywania danych
- System powinien zapewniać dużą skalowalność i wysoką dostępność

Użyte technologie

- Storage – Hadoop
- Computing – Spark
- Streaming – Kafka, Flume

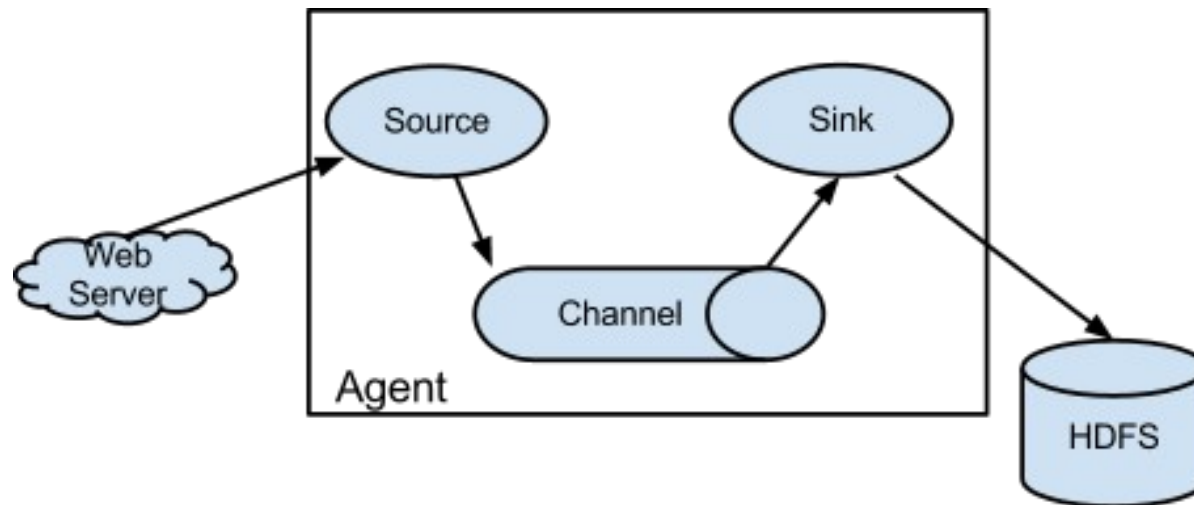


Realizacja



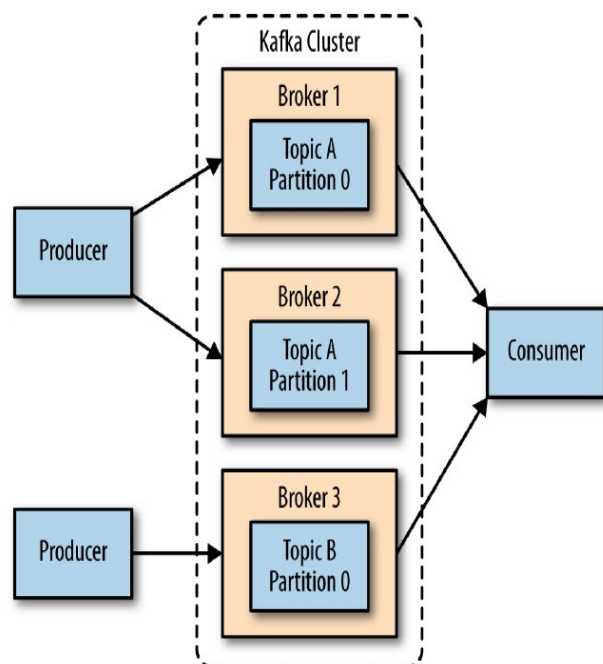
Flume

- Agenci umożliwiające przesyłanie różnych danych od źródła do ujścia
- Łatwy w konfiguracji i wystarczająco niezawodny

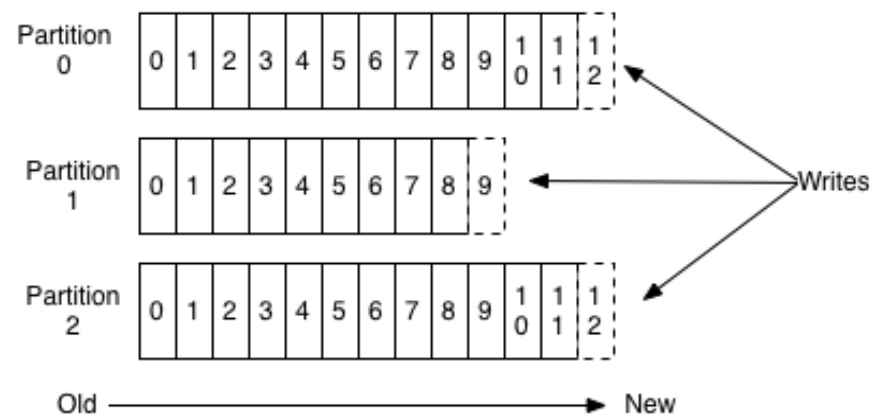


Kafka

- Niezawodny i skalowalny system przesyłania komunikatów
- Używany we wszystkich dużych rozwiązaniach produkcyjnych

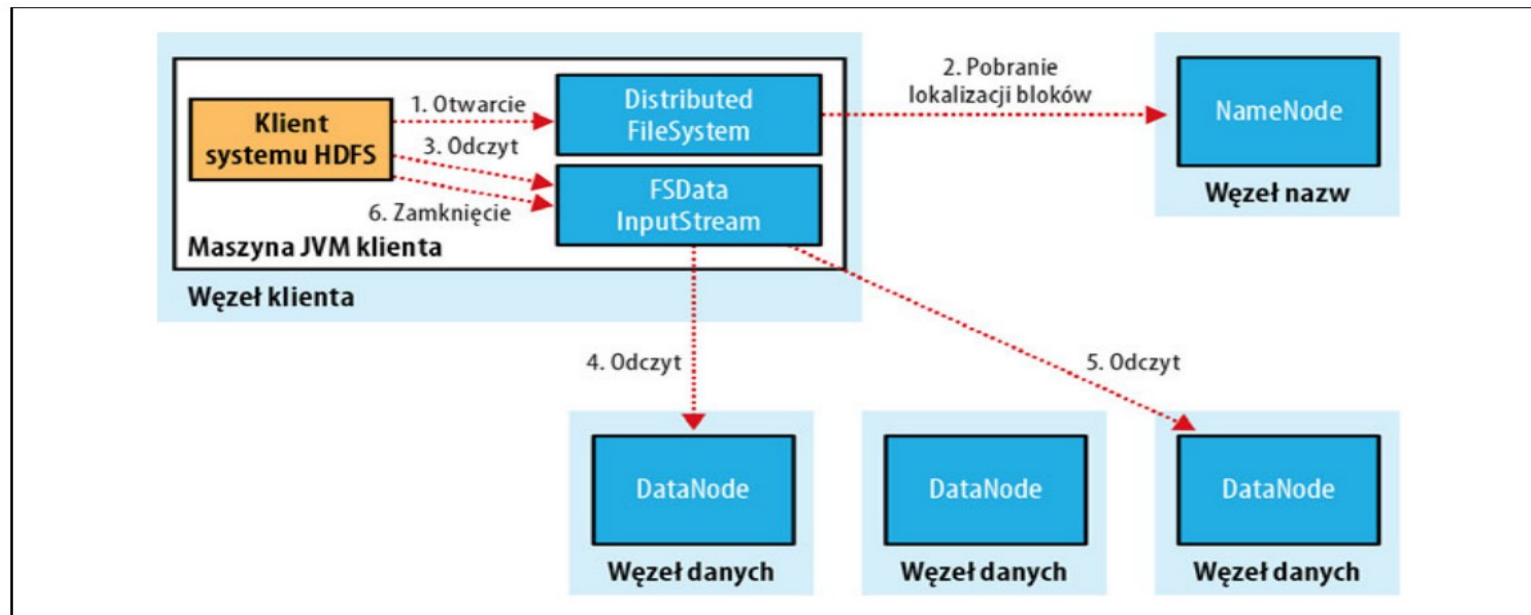


Anatomy of a Topic



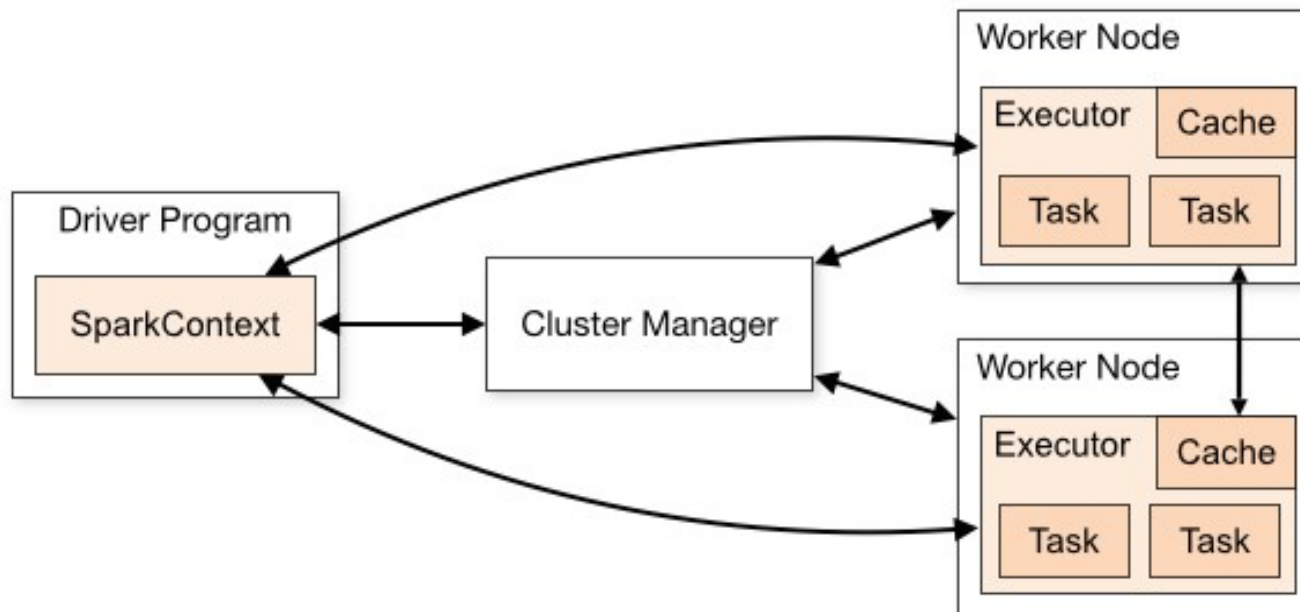
HDFS

- Skalowalny
- Wysoko dostępny
- Replikacja danych



Spark

- Rozproszony system przetwarzania danych w systemie wsadowym
- Posiada węzły zarządzające i węzły robocze.



Wyniki

Prezentacja projektu na żywo

Podsumowanie i wnioski

- Przy pomocy tych narzędzi można stworzyć dowolnie uniwersalny system przetwarzania danych charakteryzujący się wysoką dostępnością i skalowalnością
- Tego typu systemy występują we wszystkich komercyjnych rozwiązaniach produkcyjnych (LinkedIn)
- System jest skomplikowany i wymaga dużej ilości pracy w dziedzinie konfiguracji i optymalizacji