

Program szkolenia:

Apache Avro - Dane i Struktura

Informacje:

Nazwa:	Apache Avro - Dane i Struktura
Kod:	BigDataML-avro
Kategoria:	BigData, streaming i Machine Learning
Grupa docelowa:	architekci developerzy DevOps
Czas trwania:	1 dzień
Forma:	40% wykłady / 60% warsztaty

Celem warsztatu jest zrozumienie powodów dla których warto mieć zdefiniowaną strukturę danych. Uczestnicy będą mogli na przykładach sprawdzić jakie problemy generuje brak schemy oraz jakie są dobre praktyki jej projektowania. Stosując format Apache Avro będziemy modelować dane w taki sposób aby zachowały forward i backward compatibility.

Zobaczymy jak radzić sobie z olbrzymią ilością danych poprzez użycie odpowiedniego partycjonowania.

Szczegółowy program:

1. Definiowanie struktur danych

1.1. Dlaczego warto mieć zdefiniowaną strukturę danych?

1.2. Przykład problemów przy braku schemy

1.3. Przykład czytania danych ze zdefiniowaną schemą

1.4. Zalety: mniej kodu, szybsze dostarczanie

2. Kompatybilność danych

2.1. Przykład problemów ze zmianą danych

2.2. Walidacja kompatybilności

3. Analiza danych pod kątem odpowiedniego partycjonowania

4. Definiowanie Struktury dla danych

4.1. Tworzenie struktury dla danych

4.2. Generowanie klas ze struktury

4.3. Zapis danych (rezultatu zadania) używając struktury