

Wykorzystanie technologii Hadoop w SASie



Przeznaczenie szkolenia

Szkolenie pokazuje, jak z poziomu kodu SAS odczytywać, modyfikować i zapisywać dane w Hadoopie. W pierwszej części prezentowane są metody dostępne w module SAS Base, pozwalające na odczyt i zapis danych Hadoop z Data Stepu, zarządzanie systemem plików w Hadoopie i wykonywanie kodów PIG za pomocą procedury HADOOP. Druga część pokazuje mechanizmy dostępne w module SAS/ACCESS Interface to Hadoop, które umożliwiają podpięcie biblioteki do Hadoopa i wykorzystanie SQL do odczytu i zapisu danych. Na zakończenie pokazane są dodatkowe technologie pozwalające na współpracę SASa i Hadoopa, w tym wykonywanie kodów DS2 w Hadoopie czy wykorzystanie Hadoopa jako źródła danych na platformie SAS Viya.



Korzyści wynikające z ukończenia szkolenia

Naucz się

- Odczytywać i zapisywać dane do Hadoopa korzystając z wyrażenia FILENAME.
- Wykonywać polecenia Hadoopa z procedury HADOOP.
- Uruchamiać kody PIG w Hadoopie z programu SAS.
- Dostawać się do danych poprzez bibliotekę SAS.
- Tworzyć i przysyłać zapytania SQL do Hadoopa za pomocą procedury SQL.
- Optymalizować współpracę SASa z Hadoopem.
- Łączyć dane za pomocą procedury SQL i Data Stepu.
- Używać podstawowe procedury SAS na danych Hadoop.
- Modyfikować programy DS2, żeby były wykonywane w Hadoopie.
- Wykorzystywać Hadoop do przechowywania tabel SAS Viya.



Oczekiwane przygotowanie słuchaczy

Programiści SAS, którzy będą korzystać z danych Hadoop w SAS



Język szkolenia

polski



Czas trwania

2 dni / 14 godzin

Agenda szkolenia

Wstęp

- Co to jest Hadoop?
- Współpraca SASa z Hadoopem

Dostęp do HDFS i uruchamianie aplikacji Hadoop z SASa

- Przegląd metod dostępnych w module SAS Base do współpracy z Hadoopem
- Czytanie i pisanie plików Hadoop z SAS Base
- Wykonywanie kodów MapReduce
- Wykonywanie kodów Pig za pomocą procedury HADOOP

Wykorzystanie mechanizmu SQL pass-through

- Mechanizm SQL pass-through
- Łączenie z bazą Hadoop Hive
- Zadawanie zapytań do tabel Hive
- Sprawdzanie metadanych tabel Hadoop Hive
- Tworzenie zapytań SQL pass-through
- Tworzenie i ładowanie tabel Hive za pomocą wyrażenia EXECUTE w SQL pass-through.
- Praca z danymi typu STRING

Wykorzystanie silnika SAS/ACCESS LIBNAME

- Definiowanie biblioteki do Hadoopa
- Używanie opcji zbiorów
- Tworzenie widoków
- Łączenie tabel
- Korzyści korzystania z biblioteki

- Wykorzystanie procedury HDMD do dostępu do danych tekstowych z separatorem, plików XML i innych formatów
- Wskazówki na temat wydajności do stosowania bibliotek SAS/ACCESS
- Kopiowanie danych z SASa do biblioteki Hive

Partycjonowanie i klasteryzacja tabel Hive

- Metody partycjonowania i klasteryzacji w Hadoopie
- Jak partycjonowanie i klasteryzacja mogą zwiększyć wydajność
- Tworzenie i ładowanie partycjonowanych i klasteryzowanych tabel Hive

SAS Viya, Hadoop i SAS Code Accelerator for Hadoop

- Ładowanie i zapisywanie tabel CAS w HDFS
- Wykonywanie wątków DS2 w Hadoopie do agregowania danych
- Wykorzystanie procedury HDMD do odczytu plików HDFS

Przegląd SAS In-Memory Analytics i SAS Code Accelerator for Hadoop (Self-Study)

- Wykorzystanie procedur high-performance i silnika SASHDAT
- Tworzenie sesji serwera SAS LASR Analytic Server
- Wykorzystanie silnika SASIOLA
- Wykonywanie wątków DS2 na Hadoopie
- Wykorzystanie procedury HDMD do dostępu do plików HDFS