

## **Módulo I Introducción a Hadoop y Big Data**

Este es un módulo introductorio está dirigido a participantes que no tienen ningún conocimiento previo de Cloudera, o en general, de Big Data

- Introducción a conceptos de Big Data
- Fundamentos de Hadoop
- Introducción a Hive e Impala
- Tablas y bases de datos
- Sintaxis básica de SQL para Hive e Impala
- Consultas básicas usando Hive e Impala
- Problemas comunes en la ejecución de consultas
- Buenas prácticas para la ejecución de consultas
- Almacenamiento de los datos
- Mejorando el desempeño del almacenamiento de datos

## **Módulo II Uso de Hive e Impala**

En este módulo se asume que los participantes están familiarizados con los temas revisados en el Módulo I, por lo que se espera un conocimiento básico de Hadoop y Big Data.

- Funciones comunes al hacer consultas
- Usando varias tablas en una consulta
- Funciones analíticas y windowing
- Optimización de consultas en Hive
- Optimización de consultas en Impala
- Seleccionando la mejor herramienta para cada tarea
- Buenas prácticas en el uso de Hive e Impala

## **Módulo III: Uso Avanzado de Hive e Impala e Introducción a Spark y Kudu**

En este módulo se espera que los participantes tengan un conocimiento intermedio de Hadoop y Big Data. Se asume que están familiarizados con los temas revisados en los Módulos I y II.

- Datos complejos en Hive e Impala
- Extensiones para Hive e Impala
- Análisis básico de texto: Uso de expresiones regulares
- Introducción a Spark

- Abstracción de datos en Spark
- Introducción a Kudu
- Técnicas de manejo de datos en Kudu