



XUNTA DE GALICIA

CONSELLERÍA DE ECONOMÍA, EMPREGO E INDUSTRIA

DIRECCIÓN XERAL DE ORIENTACIÓN E PROMOCIÓN LABORAL



Accións formativas para o emprego

O DIRECTOR XERAL

CERTIFICA: Que FRANCISCO JAVIER SANCHEZ TEMPRANO con DNI/NIE 53198101K
asistiu ao curso de formación "ANALISTA DE DATOS CLOUDERA - VIRTUAL"
que se realizou no Centro de Novas Tecnoloxías de Galicia (CNTG)
do 20/06/2019 ao 05/07/2019, co programa que se especifica ao dorso.



CENTRO DE NOVAS TECNOLOXÍAS
DE GALICIA



Pablo Casal Espido
(asinado electrónicamente)



DORSO QUE SE CITA

CÓDIGO DO CURSO: 2019-XTIFCI1802

TOTAL DE HORAS DO CURSO: 53
(Contido específico máis módulos transversais)

PROGRAMA IMPARTIDO - 48 Horas -

Introdución

Hadoop Basics

- Por que Hadoop?
- Aspectos xerais de Hadoop
- Almacenamento de datos: HDFS
- Procesamento de datos distribuído: YARN, MapReduce e Spark
- Procesamento e análise de datos: Hive e Impala
- Integración de datos: Sqoop
- Outras ferramentas de datos de Hadoop
- Explicación do escenario con exercicios

Introdución a Hive e Impala

- Que é Hive?
- Que é Impala?
- Por que usar Hive e Impala?
- Esquema e almacenamento de datos
- Comparación entre as bases de datos Hive e as tradicionais
- Casos de uso

Consultas con Hive e Impala

- Táboas e bases de datos
- Sintaxis básica nas consultas Hive e Impala
- Tipos de datos
- Tocar emprego para realizar consultas
- Emprego de Beeline (Shell of Hive)
- Emprego da Concha de Impala

Operadores comúns e funcións integradas

- Operadores

- Funcións escalares
- Funcións de agregación

Almacenamento de datos

- Almacenamento de datos
- Creación de bases de datos e táboas
- Carga de datos
- Alteración de bases de datos e táboas
- Simplificación de consultas con vistas
- Almacenamento dos resultados das consultas

Almacenamento e rendemento de datos

- Táboas de particionamento
- Carga de datos en táboas particionadas
- Cando empregar o particionamento
- Elección do formato de almacenamento
- Xestión de metadatos
- Control de acceso a datos

Traballar con varios conxuntos de datos

- UNIÓN e Xuntanza
- Xestión de valores nulos en xuntanzas
- Unións avanzadas

Funcións analíticas e funcións de fiestras

- Utilización de funcións analíticas comúns
- Outras funcións analíticas
- Ventás deslizantes

Datos complexos

- Datos complexos con Hive
- Datos complexos con Impala

Análise de texto

- Uso de expresións regulares
- Procesamento de texto con SerDes en Hive
- Análise de sentimentos e n-gramos

Optimización de colmea

- Realización das consultas
- Bucketing
- Indexación de datos
- Hive at Spark

Optimización do impala

- Realización de consultas
- Mellorar o rendemento de Impala

Ampliando Hive e Impala

- Personalizar SerDes e formatos de ficheiro en Hive
- Transformación de datos con scripts personalizados en Hive
- Funcións definidas polo usuario
- Consultas parametrizadas

Elección da mellor opción

- Comparación entre as bases de datos MapReduce, Hive, Impala e relacional
- ¿Cal elixir?

Conclusión

MÓDULO TRANSVERSAL

Módulo FCOXXX21 - Formación para a igualdade. Duración: 5 horas

DIRECCIÓN XERAL DE ORIENTACIÓN E PROMOCIÓN LABORAL

