

Bases de datos NoSQL: HBase

Ejercicios de evaluación

PROGRAMA EXPERTO EN “*Big Data*”

Curso 2014-2015



Objetivo

El objetivo de estos ejercicios es realizar un caso de uso completo: determinación del modelo, importación de datos, desarrollo de un cliente Java para explotación de datos

Evaluación

Puntuación: se indica en los propios apartados

Plazo de entrega: 15 de Abril

Material proporcionado

- Máquina virtual oficial del curso en su versión del primer trimestre del curso 2014/2015
- Material teórico/práctico explicado en clase
- Archivos con el enunciado, modelo de datos propuesto y proyecto Java necesarios para el ejercicio de evaluación
- Archivos de datos NYSE_daily

Ejercicios

Este ejercicio se divide en 3 apartados:

1. Determinación de un modelo de datos alternativo al propuesto (3 puntos)
2. Desarrollo del código Java que importa el modelo de datos (3 puntos)
3. Desarrollo del código Java que implementa el filtrado (4 puntos)

Excepciones

- Si el alumno propone su propio modelo de datos, la puntuación del apartado 1 será de 3 puntos. Si el alumno usa el modelo propuesto, la puntuación del apartado 1 será de 1.5 puntos
- Si el alumno no importa los datos usando Mapreduce, puede optar por crear unos pocos registros (por ejemplo, 10) con un programa tipo “BasicoEmpresa” de los que se han hecho en clase. En este caso el apartado 2 tendrá una puntuación de 1.5 puntos
- Si el alumno no implementa el código Java correspondiente al filtrado puede entregar las mismas queries usando hbase Shell. En este caso la puntuación será de 2 puntos

Datos de entrada:

1. Fichero de datos de entrada en formato CSV o TSV compatible con el programa de importación que desarrolle el alumno
2. Ruta en el PC del alumno del fichero de datos de entrada
3. Fecha en la que el programa tiene que hacer la predicción en el formato “20/03/2000”

Entregas

Normas para la entrega:

- El alumno entregará un proyecto con
 - Un documento describiendo el modelo de datos propuesto
 - El fichero de datos de entrada
 - Las clases Java necesarias para importar el modelo de datos
 - El código Java que hace el filtrado en el método “getMetric” de la clase “EstimateMarketDrop” (sección comentada con “Inicio del código del alumno”)
- El código deben tener líneas de comentario explicando lo que hace
- El código debe funcionar en la máquina virtual que se usó durante las clases ejecutándolo mediante “Run as Java Application” de Eclipse