

Tarea 6

Programación avanzada

Postgrado IIMAS

7 de Noviembre de 2013

Fecha de entrega 14 de Noviembre de 2013.

Tarea hadoop.

Para entrar al servidor

ssh -X -p 2222 usuario@132.248.51.117

El directorio de trabajo que se utilizará para la tarea es /progavanzada. Dentro de ese directorio se encuentran los directorios que contienen los datos de prueba como por ejemplo /progavanzada/datos_cols, /progavanzada/datos_grupo, etc.

Los directorios de salida también deberán apuntar a /progavanzada. Ejemplo /progavanzada/salida.

Como el sistema de archivos distribuidos es compartido por todos los usuarios su directorio de salida se sugiere sea /progavanzada/usuario_sal o algo por el estilo.

Problema 1.- Un archivo de texto (utilizar el directorio datos_cols) tiene n columnas de números double al azar con errores de lectura. Hacer un programa en hadoop que calcule el máximo de la tercera columna.

Problema 2.- Un archivo (en el directorio datos_grupo) tiene dos columnas la primera un entero que identifica a un grupo y la segunda un valor de tipo double de ese grupo. Haga un programa que tenga como salida archivos el identificador del grupo con todos sus valores, después haga un programa que calcule el máximo con esta nueva estructura.

Problema 3.- Un archivo de texto (en el directorio datos_texto) tiene texto. Haga un programa que calcule la frecuencia de las palabras.

Problema 4.- Realice un programa para generar los datos de una gráfica (directorio datos_grafica).

El archivo tiene la siguiente estructura: grupo, subgrupo, coordenada x y coordenada y.

Obtenga todos los datos que tengan el mismo subgrupo y promedie los valores, también entregue la gráfica.

Los problemas tienen que estar hechos con lenguaje java en un proyecto maven, las clases tienen que estar probadas en junit y tener el control de versiones git (se restarán 2 puntos de no tenerlos).

La tarea tiene que ser mandada al ayudante José Angel Hermosillo al correo joseangel.hermosillo@gmail.com en un archivo tar.gz que contenga su nombre y con su respectiva documentación en un archivo en formato pdf.