



PROGRAMA DE ASIGNATURA

Asignatura : Procesamiento de Grandes Volúmenes de Datos
Clave : <Clave de la asignatura>
Créditos :
Semestre :
Profesor : Dr. Wenceslao Palma

1. Descripción

Esta asignatura profundiza en aspectos teóricos y prácticos relacionados al procesamiento de grandes volúmenes de datos. El propósito general de esta asignatura es preparar al estudiante para el desarrollo de aplicaciones que procesan grandes volúmenes de datos usando un modelo de programación y sus herramientas relacionadas.

2. Objetivos

Comprender los problemas asociados al procesamiento de grandes volúmenes de datos.
Conocer y experimentar con modelos y herramientas asociados al procesamiento de grandes volúmenes de datos.

Desarrollar aplicaciones que procesan grandes volúmenes de datos.

3. Contenido de la asignatura

- 3.1 Introducción.
 - 3.1.1. Procesamiento de grandes volúmenes de datos.
 - 3.1.2. Sistemas de Gestión de Flujos de Datos.
 - 3.1.3. MapReduce.
- 3.2 Sistemas de Gestión de Flujos de Datos
 - 3.2.1. Arquitectura.
 - 3.2.2. Procesamiento de consultas. Ejecución. Operadores físicos. Plan de consultas. Optimización. Lenguajes de Consulta.
 - 3.2.3. Operadores de Consulta: Join. Mjoin, Eddies. Skyline. Algoritmos básicos.
 - 3.2.4. Optimización de consultas. Administración de la memoria. Caching, Load-shedding.
 - 3.2.5. Procesamiento distribuido de flujos de datos. Arquitecturas. Cluster, P2P.
- 3.3 MapReduce.
 - 3.3.1. Modelo.
 - 3.3.2. Map y Reduce.
 - 3.3.3. Paralelización, tolerancia a fallas, distribución de datos y balance de carga.
- 3.4 Hadoop.
 - 3.4.1. Hadoop Distributed FileSystem (HDFS). Diseño, conceptos.
 - 3.4.2. Hadoop IO. Integridad, compresión, serialización.
 - 3.4.3. Desarrollo de aplicaciones en Hadoop.
- 3.5 Consultas complejas y MapReduce.
 - 3.5.1. Operadores incorporados. Counter, Sort, Join.
 - 3.5.2. Consultas complejas. Estrategias de join.
- 3.6 Pig
 - 3.6.1. Modelo de datos, expresiones, comandos.
 - 3.6.2. MapReduce y Pig.



PROGRAMA DE ASIGNATURA

4. Metodología

Clases expositivas por parte del profesor, actividades de laboratorio, exposiciones orales e informes escritos a realizar por los alumnos.

5. Instancias de evaluación

Exposición (1): 30%

Certamen (1): 30%

Tarea (2): 40%

6. Bibliografía Obligatoria

Garcia-Molina H., Ullman J., Widom J. Database System Implementation. Prentice-Hall, 2000.

Deshpande A., Zachary I., Vijayshankar R. Adaptive Query Processing. Now Publishers Inc, 2007 .

White, T. *Hadoop: The Definitive Guide*. Ed. O'Reilly, 2010.

Papers conferencias VLDB, ICDE, SIGMOD.

Papers revistas TKDE, VLDBJ.