



Facultad de Ingeniería Comisión Académica de Posgrado

598/13

1
LMB

Formulario de Aprobación Curso de Actualización 2013

Asignatura: Administración de datos en la nube (Data management on the cloud) ^{cloud}
(Si el nombre contiene siglas deberán ser aclaradas)

Profesor de la asignatura¹: Geneveva Vargas-Solar, Full time researcher, Centre National de la Recherche Scientifique, France
(título, nombre, grado o cargo, Instituto o Institución)

Profesor Responsable Local¹: Regina Motz, Profesor Titular, Instituto de Computación
(título, nombre, grado, Instituto)

Otros docentes de la Facultad:
(título, nombre, grado, Instituto)

Docentes fuera de Facultad: Javier Alfonso Espinosa Oviedo, Msc., University of Grenoble, France
(título, nombre, cargo, Institución, país)

Instituto ó Unidad: Instituto de Computación
Departamento ó Área: Sistemas de Información

¹ Agregar CV si el curso se dicta por primera vez.
(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

Fecha de inicio y finalización: primer semestre
Horario y Salón: a confirmar

Horas Presenciales: 23
(se deberán discriminar las mismas en el ítem Metodología de enseñanza)

Arancel: \$5700

Publico Objetivo y Cupos: Profesionales y estudiantes de la carrera sistemas de información y tecnología de gestión de datos.

Sin cupo

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción en el Depto. de Posgrado, hasta completar el cupo asignado)

Objetivos: La era de los grandes volúmenes de datos, en la que petabytes de información se acumulan a un ritmo acelerado llama a la utilización masiva de las técnicas para la gestión (almacenar, indexar, duplicar), consulta y análisis. El procesamiento de miles de millones de páginas web, fotos, entradas de archivos de log, etc., llaman al desarrollo de nuevas herramientas y propuestas de nuevos paradigmas de programación.

Cloud computing se está convirtiendo en un enfoque relativamente nuevo para tratar y facilitar el acceso ilimitado a los recursos de computación y almacenamiento para la construcción de nuevas aplicaciones. El principio básico de la computación en la nube es asumir que las aplicaciones, accesibles a través de una red, se basan en una infraestructura orientada a servicios dedicada a brindarles la necesaria (ni más ni menos) cantidad de computación, almacenamiento y recursos de red. En lugar de tener un ordenador o servidor para todo, el contexto de computación está configurando de acuerdo con las características de la aplicación. En lugar de comprar uno para todo, los recursos del equipo / servidor se proporcionan (y se compran) bajo demanda.

Este curso se centrará en la integración y gestión de datos en arquitecturas orientadas a servicios en la nube. Por lo tanto, el curso introducirá brevemente conceptos fundamentales del cloud computing y se ocupará de los datos y la gestión de los servicios en la nube utilizando ejemplos prácticos basados en ambientes existentes en la nube y modelos de ejecución, como (i) ETL y herramientas para la federación de datos, (ii) Map-reduce y su aplicación Hadoop, el más destacado ecosistema de código abierto de herramientas para trabajar con nuevas e interesantes conjuntos de datos a gran escala y, (iii) Bases de datos "NoSQL".

Objetivos específicos:

- Comprender el área emergente de "cloud computing" y cómo se relaciona con los modelos tradicionales de la informática
- Ganar competencias en
 - ETL para recoger grandes colecciones de datos
 - MapReduce como un modelo de programación para el procesamiento distribuido de datos
- Manejo de Bases de datos NoSQL
- Para tener una experiencia en la construcción, el acceso y uso de servicios en la nube mediante la adopción de un enfoque orientado al servicio.

Conocimientos previos exigidos: Fundamentos de Base de Datos
(Detallar el conocimiento exigido)

Conocimientos previos recomendados: Conocimientos básicos sobre construcciones de aplicaciones web.
(Detallar el conocimiento recomendado)

Metodología de enseñanza: El curso se basa en la idea de lograr que los estudiantes comprendan los diversos aspectos de la gestión de datos y servicios a través de un enfoque problema-solución. Por lo tanto, vamos a proponer problemas que puedan surgir en el desarrollo y despliegue de aplicaciones centradas en datos (con grandes conjuntos de datos y servicios) dentro de una nube y se guiará a los estudiantes para proponer y programar soluciones utilizando herramientas específicas.

- Horas clase (teórico/práctico): 10
- Horas clase (práctico): 8
- Horas evaluación: 5
 - Subtotal horas presenciales: 23
- Horas estudio: 16
- Horas resolución ejercicios/prácticos: incluidas durante el curso

- Horas proyecto final/monografía: 6
 - Total de horas de dedicación del estudiante: 45

Forma de evaluación: El objetivo del mini-proyecto es poner a prueba sus habilidades para desarrollar una aplicación en la nube basada en servicios. Confiamos en su capacidad para imaginar una aplicación inicial. La solicitud debe ser simple, pero debe tener las siguientes características:

- Hacer frente a la construcción de una base de datos poliglota, con un proceso de ETL que pueda llenarla mediante la interacción con los servicios de datos existentes (por ejemplo, Facebook, Twitter, Deezer)
- Implementar la base de datos poliglota en una o varias nubes
- Servicios que pueden dar acceso a la misma para la manipulación y consulta de los datos de exportación.
- Desarrollar un servicio que puede analizar los datos mediante un programa de Hadoop.
- Desarrollar una aplicación que utiliza los servicios.
- Implementar la aplicación en una nube

El trabajo se realizará por grupos de 3 ó 4 personas que deben estar seguros de ser capaces de hacer frente a la gestión de datos y la teoría de las cuestiones técnicas (por ejemplo, ETL, almacenamiento, acceso concurrente, consulta, análisis). Cada grupo va a demostrar el mini proyecto durante 1/2 hora, incluyendo preguntas durante el periodo de exámenes. Cada grupo hará entrega de un informe de un máximo de 10 páginas y el código de la aplicación.

Durante la presentación el grupo hará lo siguiente:

Describir el mini-proyecto en concreto, el objetivo y las funciones principales.

- Describir la base de datos poliglota implementado la caracterización de los datos utilizando UML, justificando la elección de los modelos utilizados para el almacenamiento de los datos y el proceso ETL utilizado para su alimentación.
- Explicar que se llevaron a cabo los servicios de acceso a la base de datos poliglota y mostrar que se despliegan en una nube.
- Describir y demostrar el servicio hadoop que analiza la base de datos.
- Describir la arquitectura de la aplicación de base de datos global implementado en la nube.
- Demostrar la aplicación: preparar un kit de prueba e insistir en los puntos fuertes de su aplicación.

Temario:

1. Data, Services and Cloud Computing
 - 1.1 Introduction on massive data, services and Cloud
 - 1.2 Get a first touch in developing and deploying a service for extracting and loading big collections of data in a SQL Server on the Cloud
1. Data Modelling and processing
 - 1.1 Big data models
 - 1.2 Map-reduce programming model and the Hadoop framework
2. Data distribution

-
- 2.1 Replication and sharding
 - 2.2 Management issues
 - 3. Choosing the right data store
 - 3.1 Polyglot Persistence
 - 3.2 NoSQL databases observation: main actors, application types
 - 4. Data service oriented computing
 - 4.1 Building data services for cloud
 - 4.2 Academic and industrial perspectives

Bibliografía:

[1] NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence by Pramod J. Sadalage, Martin Fowler

[2] Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement, by: Eric Redmond and Jim R. Wilson

[3] Thanks to J. Ullmann, P. Valduriez, Cl. Roncancio, Ch. Bobineau, JL. Zechinelli, R. Lozano for the slides provided

<http://deoracle.org/online-pedagogy/teaching-strategies/applying-cloud-computing.html>

<http://blogs.msdn.com/b/brunoterkaly/archive/2010/10/05/how-to-teach-cloud-computing-the-windows-azure-platform-step-1.aspx>

<http://sites.google.com/site/freeonlineteachingtools/cloud>

<http://aws.amazon.com/education/>

<http://www.umiacs.umd.edu/~jimmylin/cloud-2008-Fall/index.html>