Gian Course

Basic Concepts and Issues in Big Data Management

GLOBAL INITIATIVE OF ACADEMIC NETWORKS

सत्यमेव जयते
MHRD
Govt. of India

Delivered by:
Dr. Nicolas Spyratos
Laboratory for Research in Informatics
University of Paris-South, France

Course Dates: **5-17 December 2016**
Venue: **IIT Indore @ Simrol Campus, Madhya Pradesh**

MINISTRY OF
HUMAN RESOURCE DEVELOPMENT, GOVT. OF INDIA

Hosted by:

Indian Institute Of Technology Indore

Discipline of Computer Science and Engineering

Course: Basic Concepts and Issues in Big Data Management

Course objective

Today, scientists regularly encounter limitations due to the very large sizes of data sets, in many areas, including meteorology, genomics, complex physics simulations, and biological and environmental research. The limitations also affect Internet search, finance and business informatics. Striking examples from the business world include Facebook, which handles 40 billion photos from its user base; and Walmart, which handles more than 1 million customer transactions every hour, imported into databases estimated to contain more than 2.5 petabytes of data. The potential uses of big data have also been recognized at the highest administration levels in the US as well as in several other countries [1]. As for the private sector, according to a Gartner survey (September 17, 2014), "73% of organizations have invested or plan to invest in Big Data in the next two years". However, the use of big data has drawn considerable criticism as well from consumer privacy advocates concerned about the threat to privacy represented by increasing storage and integration of personally identifiable information. Nevertheless, in spite this criticism, there is no doubt that the collection and processing of big data holds high expectations for a variety of human activities. In fact, a new field centered on big data is emerging, usually referred to as "data science", and several universities and educational institutions already offer degrees in this field [2].

The objective of this course is to present the basic concepts and issues in big data management and open up new perspectives for students pursuing professional careers and/or academic degrees in data science and related fields.

References

[1] House, W.: Big data: Seizing opportunities, preserving values. http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf (May 2014)

[2] http://101.datascience.community/2012/04/09/colleges-with-data-science-degrees/

Course Overview

This course presents the basic concepts and issues in big data management during the evolution from data warehouses introduced in early 1990s to what is called big data today. The following subjects are presented:

Data warehouses: The need for data integration, storage and analysis, star schemas and SQL extensions, indexing techniques.

Query personalization: How can we perceive what is in a "big answer"? How can we explore a big answer? Incorporating user preferences in query processing.

Big Data today: What it is and how it differs from big data in data warehouses.

Big Data Analytics: Analytic queries and their evaluation using the Apache Hadoop/MapReduce Framework

A functional approach to big data analytics: Analysis context and the HiFun functional query language, query evaluation through rewriting, query execution plans, result visualization and exploration, interfacing big data sets.

Mapping to lower level evaluation mechanisms: Mapping analytic queries from HiFun to SQL group-by queries, **r**ewriting SQL group-by queries based on functional dependencies, mapping analytic queries to Hadoop/MapReduce.

Examples of data mining: Classification, Clustering, Association rules.

Course Completion Certificate

A certificate is awarded jointly by IIT Indore and MHRD under the GIAN program.

Course Schedule and Structure

The course is scheduled from 5-December-2016 till 17-December-2016 at IIT Indore. The course is delivered in 3-hour lectures per day and would last for 10-days. The lecture layout is as following:

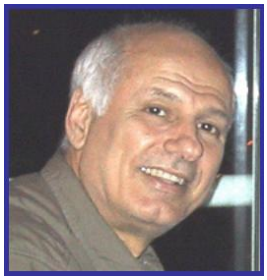
Lecture Set	Date	Covered Topics
1	05/12/16	Data warehouse, the need for data integration, storage and analysis of data, star schemas and SQL extensions, indexing techniques.
2	06/12/16	Incorporating user preferences in query processing, nature of user preference, rewriting preferences as query sequences, designing an interface for preference queries
3	07/12/16	Introduction to big data today, analytic queries and their evaluation, the Apache Hadoop Framework for evaluating analytic queries.
4	08/12/16	The HiFun language: a functional approach to big data analytics, analysis context and query language, result visualization and exploration, interfacing big data sets.
Assignment Day (AD) 1	09/12/16	ASSIGNMENT BREAK (Course Participants Work on the Assignment in this break)
AD 2	10/12/16	
AD 3	11/12/16	
5	12/12/16	Query evaluation through rewriting, query execution plans.
6	13/12/16	From HiFun to SQL group-by queries and to Hadoop/MapReduce jobs.
7	14/12/16	Rewriting SQL group-by queries based on functional dependencies.
8	15/12/16	Examples of data mining: Classification, Clustering, Association rules.
9	16/12/16	Buffer Day
10	17/12/16	**Written Exam of Three Hours**

The course participants would be provided access to computer labs at IIT Indore.

Course Prerequisites: The course participants are expected to have:

- Familiarity with SQL databases (and, ideally, with relational algebra)
- Good knowledge of high-school mathematics (sets, functions, partial orders)

Expert Faculty from France



Dr. Nicolas Spyratos
Professor Emeritus of Computer Science, University of Paris-South, France
Learning and Optimisation Group of Laboratory for Research in Informatics
Affiliated Scientist, Institute of Computer Science (FORTH-ICS), Greece
Scientific Advisor of Japan Science and Technology (JST)

Nicolas Spyratos (spyratos@lri.fr) received his BEng degree from the National Technical University of Athens, Greece, his M.Sc. degree from the University of Ottawa, Canada, his Ph.D. degree from Carleton University, Canada and his "thèse d'état" from the University of Paris South 11, France. He worked as a researcher for Bell-Northern Research in Canada, and for INRIA and the National Research Council (CNRS) in France, prior to joining the University of Paris South as a full professor in 1983, where he was heading the database group from 1985 to 2011. He is currently Professor Emeritus at the University of Paris South, scientific advisor of Japan Science and Technology (JST), and affiliated senior scie ntist, at the Institute of Computer Science of Crete, in Greece (http://www.ics.fort h.gr/). His research interests include databases, big data analytics, conceptual modeling and digital libraries. He is the author of over 200 articles in international journals, books and conferences, and has supervised the work of 24 PhD students. He has also served on the program committee of over 100 international conferences, he has participated in over 25 national, European and international research projects and has served as evaluator for the NSF and the European commission.

Local course coordinator

Dr. Gourinath Banda
Assistant Professor
Computer Science and Engineering
IIT Indore
Simrol, Indore, MP
India

Gourinath completed his PhD degree from the Roskilde University, Denmark and his M.S. Engg. degree from the University of Southern Denmark, Denmark. He received a PhD fellowship from the EU project "Advanced Specialisation and Analysis for Pervasive Systems" and international graduate scholarship jointly granted by the Government of Denmark and Foss Electric A/S Copenhagen. Prior to joining academia, he worked in industry doing research in avionics, formal methods and mobile software systems. His research interests include Internet of Things (IoT), Cyber Physical Systems, Virtual sensing, Semantic modeling of systems, Formal verification, Sensor fusion. He developed the HARTEX realtime kernel with a memory footprint of under 10 KB. Holds a granted patent in the area of mobile software systems. Developed the forward compatible OneIoT protocol.

Practical Information

Who can attend?

- Engineers, Enthusiasts, Professionals and Researchers in computer science, information technology, from commercial and government organizations including R&D laboratories.
- Students at all levels (Bachelors i.e. BSc/BTech/BEngg, Masters i.e. MSc/MTech/MEngg or Doctoral i.e. PhD etc.)
- Faculty from academic institutions and technical institutions who would like to enhance there skills in the upcoming area of Big Data and Analytics.

Course Registration Fees

Student Participants from India: Rs. 2000
Faculty Participants from India: Rs. 5000
Government Research Organization Participants: Rs. 8000
Industry Participants: Rs.15000
Participants from abroad : US $500

The above fee includes all instructional materials, computer usage at IIT Indore labs, free internet access as well as lunch for all lecture days.

Lodging and Boarding

The participants will be provided with accommodation on a nominal payment basis at IIT Indore hostel. That is Rs. 250 per night with two people sharing a room. Please reserve and confirm your room booking by emailing at hostel@iiti.ac.in (mention that course coordinator is Dr. Gourinath Banda). Breakfast and dinner would also be on payment basis (Rs. 100 for both combined).

Steps to follow to enrol for this course

1. It is recommended to pay the applicable fees by either NEFT or Demand Draft (DD) to the IIT Indore account.

a. Details for Electronic fees remittance mode i.e. NEFT or RTGS:

Name of the Beneficiary: **IIT Indore Project and Consultancy Account**
Name of Bank: **Canara Bank**
Branch: **Indore Navlakha**
Beneficiary Account No. **1476101027440**
Bank MICR Code **452015003**
Bank IFS Code **CNRB0001476**

b. Details for Demand Draft: Draw a DD infavour of:

Name of the Beneficiary: **IIT Indore (Project and Consultancy Account)**
Payable at: **Indore**

2. Register online, before the deadline, which is on 1st December 2016 on the http://gian.iiti.ac.in/register.php

3. If you need hostel accommodation then please get in touch with hostel administration mentioned above.