NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

C S   S E M I N A R

Title: **Data-Driven Genomic Computing**

Speaker: Stefano Ceri
               Professor of Database Systems
               at Politecnico di Milano,
               Dipartimento di Elettronica e Informazione
               Milano, Italy

Date/Time: 12 November 2015, Thursday, 02:00 PM to 03:30 PM

Venue: TR11, COM1-02-16

Chaired by: Dr Rosenblum, David S., Professor, School of Computing
               (david@comp.nus.edu.sg)

Abstract:

Next Generation Sequencing (NGS) is a technology for reading the DNA that is changing biological research and will change medical practice; thanks to the availability of millions of whole genome sequences, genomic data management may soon become the biggest and most important "big data" problem of mankind.

In this exciting framework, I will describe a new paradigm for raising the level of abstraction in NGS data management, by introducing a genomic data model (GDM) which supports both genomic regions and metadata, a genomic query language (GMQL) which provides high-level operations upon GDM.

I will next show how GDM and GQL are supported by GenData 2020, a cloud-based system for genomic data management. GenData 2020 uses the Spark and Flink frameworks for big data computations; I will also describe a benchmark of Spark and Flink at work on genomic operations.

Biodata:

Stefano Ceri is Professor of Database Systems at the Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB) of Politecnico di Milano. He was visiting professor at the Computer Science Department of Stanford University (1983-1990). His research work covers about four decades (1976-2015) and has been generally concerned with extending database technology in order to incorporate new features: distribution, object-orientation,

rules, streaming data; with the advent of the Web, his research has been targeted towards the engineering of Web-based applications and search systems.

In 2008 he has been awarded an ERC Advanced Grant on "Search Computing". He is currently leading "GenData 2020", a national project focused on building query and analysis systems for genomic data. He was the director of Alta Scuola Politecnica (2010-2013). He is the recipient of the ACM-SIGMOD "Edward T. Codd Innovation Award" (2013), an ACM Fellow and member of the Academia Europaea.