



EARTH SCIENCES



Università degli Studi di Torino
Scuola di Dottorato in Scienze della Natura e Tecnologie Innovative
Dipartimento di Scienze della Terra

Seminario breve

Database NOSQL, NewSQL e sistemi di memorizzazione dati nell'era del cloud computing e dei big data

Diego Guenzi
Università degli Studi di Torino
diego.guenzi@unito.it

Venerdì 4 marzo 2016
ore 14:30 - 16:30

Aula Ruffini - via Valperga Caluso 35, Torino



Programma

- L'evoluzione tecnologica
 - Dai mainframe al cloud computing
- Il problema della memorizzazione dei dati nel tempo
 - L'approccio a file
 - Le basi di dati
 - Il modello relazionale
 - Il linguaggio SQL
 - Il data warehousing
 - Big data
 - Le basi di dati NOSQL
 - SQL + NOSQL = NewSQL
 - Piattaforme distribuite per l'object storage
- Conclusioni
 - Come scegliere la metodologia più appropriata?

Introduzione

Il problema della memorizzazione e dell'analisi delle informazioni è un tema molto caldo: basti pensare alle moli di dati che devono gestire quotidianamente colossi come Facebook, Google ed Amazon; senza spingerci a questi livelli¹, quantità enormi di dati possono anche essere riscontrate nel settore della ricerca scientifica e, in tali casi, bisogna ricorrere necessariamente all'uso di strumenti ad-hoc².

Questo seminario si pone l'obiettivo di offrire una panoramica sullo stato dell'arte dei principali sistemi di memorizzazione e gestione dei dati, partendo dai più tradizionali sino ad arrivare alle piattaforme per il trattamento dei cosiddetti big data. Tale panoramica può essere il punto di partenza per chi ha a che fare con moli di dati di una certa rilevanza e necessita di metodi relativamente semplici ed efficienti per la loro gestione ed analisi.

¹ Secondo alcuni ingegneri di Google, nel 2008 la società californiana processava approssimativamente 20 petabyte (cioè 20.000 terabyte circa) al giorno

² Per esempio, la decodifica del genoma umano ha richiesto 13 anni di analisi -dal 1990 al 2003- mentre, grazie ai miglioramenti tecnologici e all'adozione di strumenti specifici per il trattamento dei dati, è ora possibile in una settimana circa