

Laboratorio di Ingegneria Informatica

Mini-progetti ("tesine") relativi a Tecnologia delle Basi di Dati Aggiornato al 15/11/2013

1. Data provenance nei DBMS commerciali

Tutor: Domenico Beneventano, Sonia Bergamaschi

Data: Ottobre 2013

Descrizione: La "data provenance" (o "data lineage") viene definita come "la descrizione dell'origine di un dato e del processo tramite il quale esso è derivato". L'importante concetto di "data provenance" è stato molto studiato nell'ultimo decennio nell'ambito della ricerca scientifica; più recentemente, i principali DBMS commerciali, quali Oracle ed SQL SERVER, supportano alcune funzionalità di "data provenance". L'obiettivo della tesina è appunto quello di analizzare e valutare le funzionalità di "data provenance" supportate nei DBMS commerciali.

4. Il sistema Trio
Tutor: Domenico Beneventano, Sonia Bergamaschi
Data: Ottobre 2012
Descrizione: La "data provenance" (o "data lineage") viene definita come "la descrizione dell'origine di un dato e del processo tramite il quale esso è derivato". In un sistema di integrazione dati la "data provenance" è importante per gestire i conflitti tra i dati provenienti dalle diverse sorgenti. Trio è un nuovo tipo di DBMS che gestisce anche la data lineage. Il progetto si propone l'obiettivo di valutare il sistema Trio e la sua applicabilità alla gestione dei conflitti nel sistema di integrazione dati MOMIS. Riferimenti:

- Il progetto Trio (Stanford University): <http://infolab.stanford.edu/trio/>
- Il sistema MOMIS <http://www.dbgroup.unimo.it/Momis/>

2. NOSQL database – comparazione rispetto a sistemi tradizionali

Tutor: Sonia Bergamaschi, Giovanni Simonini

Data: Ottobre 2013

Descrizione: Negli ultimi anni sta crescendo il movimento **NOSQL**, questo movimento propone sistemi di gestione delle basi di dati non più basati sul modello relazionale, che implementano data model specifici per determinate applicazioni.

In particolare **MongoDB** e **CouchDB** sono due sistemi document-oriented altamente scalabili. Il progetto richiede l'analisi dei due sistemi sopra citati e il confronto con un tipico sistema

relazione (per esempio MYSQL) tramite l'implementazione di un benchmark.

3. Creazione e/o modifica delle pagine in Italiano di Wikipedia sui concetti fondamentali del

linguaggio di Tecnologia delle Basi di Dati ed SQL esteso.

Tutor: Sonia Bergamaschi

Data: Ottobre 2013

Descrizione: Il progetto prevede l'individuazione e la modifica alle pagine in italiano di Wikipedia che descrivono i principali concetti legati all' SQL esteso e l'eventuale creazione di nuove pagine (es. Join SQL, Stored Procedure ecc.). I requisiti della tesina sono:

- Studio e individuazione delle metodologie offerte da wikipedia per la modifica e/o creazione di articoli in italiano;
- Definizione delle modifiche alle pagine di it.Wikipedia da aggiornare e/o aggiungere sulla base dei concetti fondamentali legati al linguaggio SQL esteso;
- Creazione e/o aggiornamento delle pagine di Wikipedia precedentemente individuate.

4. Valorizzazione delle Object e data property in RDF

Tutor: Francesco Guerra

Data: Ottobre 2012

Descrizione: In “Xiaoyan Yang, Cecilia M. Procopiuc, and Divesh Srivastava. 2009. Summarizing relational databases. Proc. VLDB Endow. 2, 1 (August 2009), 634-645.” viene proposta una tecnica per riassumere il contenuto di un database relazionale, basata sull’analisi della cardinalità degli attributi e sulle foreign key, per:

1. stabilire l’importanza di una tabella in un db relazionale
2. definire uno spazio metrico sulle tabelle nel db
3. fare il clustering delle tabelle sulla base della metrica definita al punto 2 e stabilire la tabella più importante per ogni cluster (l’elemento riassuntivo) sulla base della tecnica definita al punto 1

In questa tesina si richiede di sviluppare un approccio analogo a quello descritto nel paper, e unicamente per quanto concerne il punto 1, per un documento RDF con schema associato. Il parallelismo con la modellazione relazionale è diretto: una tabella relazionale corrisponde a una classe RDF, gli attributi delle tabelle a elementi connessi attraverso data type property, le foreign key sono object property. Lo scopo è quello di sviluppare una tecnica per conoscere l’importanza delle classi in uno schema RDF.

5. Big Data e Semantic Web

Tutor: Prof. Sonia Bergamaschi, Ing. Francesco Guerra

Data: Ottobre 2012

Con il termine “Big data” si indicano sorgenti dati le cui dimensioni (dell’ordine dei gigabytes o petabytes) diventano problematiche nella gestione dei dati stessi. All’interno di questo scenario le tradizionali tecniche legate alle basi di dati relazionali per la ricerca, l’integrazione, la correlazione, la fusione e l’analisi dei dati non sono sufficienti [Big Data Now: Current Perspectives from O’Reilly Radar, 2001]. Come conseguenza, nuovi paradigmi per la gestione di Big Data, tra cui (1) tecniche innovative per l’archiviazione e il recupero dei dati, (2) algoritmi di analisi e (3) nuovi servizi relativi ai dati, devono essere sviluppati.

Le tecnologie del Semantic Web consentono alle persone di creare archivi di dati sul Web, di

costruire vocabolari, e scrivere regole [<http://www.w3.org/standards/semanticweb/>]. L’obiettivo della tesina è quello di accoppiare gli sforzi in queste due aree di ricerca per la ricerca, l’integrazione e correlazione dei Big Data. In particolare, è necessario studiare e sviluppare tecniche di apprendimento probabilistiche che sfruttano i metadati forniti dal Web Semantic ed estratti direttamente dalle sorgenti

Big Data. Proposed research activity- Study of the state of the art of techniques for managing big data. - Analysis and implementation of techniques for extracting metadata with reference to big data. - Analysis and implementation of techniques to perform integration and searching on large scale data: this includes the management of uncertainty in schema and data matching.

6. Applicazione web in ambiente Google App Engine settore Assicurativo

Tutor: Sonia Bergamaschi, Mirko Orsini

Data: Marzo 2013

Descrizione: sviluppo di una applicazione web tramite il framework Google App Engine per la gestione delle perizie assicurative. L'obiettivo del progetto è la generazione automatica di documenti descrittivi delle perizie assicurative attraverso l'interazione con un sistema gestionale. I documenti di tipo Google Spreadsheet o Google Document dovranno essere generati attraverso appositi template, a partire dai dati presenti all'interno di un database Google Cloud SQL.

Riferimenti:

<https://developers.google.com/appengine/> <https://developers.google.com/cloud-sql/>
<https://developers.google.com/drive/>

7. Integrazione delle Google Apps for Business in applicazioni sviluppate con Google App Engine

Tutor: Sonia Bergamaschi, Mirko Orsini

Data: Marzo 2013

Descrizione: Analisi e studio di tecniche di integrazione delle Google Apps for Business in applicazioni web sviluppate tramite il framework Google App Engine. L'obiettivo è sviluppare dei moduli software che consentano di integrare nelle applicazioni web sviluppate tramite il framework Google App Engine, nel modo più veloce possibile, i Google Spreadsheet, Google Calendar e altri servizi di Google Apps for Business. Riferimenti:<http://www.google.it/intl/it/enterprise/apps/business/>
<https://developers.google.com/google-apps/spreadsheets/> <https://developers.google.com/google-apps/calendar/> <https://developers.google.com/appengine/docs/whatisgoogleappengine>

8. Analisi di strumenti per il Rapid Application Development in ambiente Java

Tutor: Sonia Bergamaschi, Mirko Orsini

Data: Marzo 2013

Descrizione: Il Rapid application development (RAD) è una metodologia per lo sviluppo software che utilizza una pianificazione minima ed ha il vantaggio di arrivare rapidamente allo sviluppo di prototipi. Questa metodologia consente di arrivare ad uno sviluppo del software più rapido anche in progetti

Software complessi grazie all'utilizzo di concetti come il riuso del software. La tesina prevede una analisi dei principali tools per il Rapid Application Development in ambiente Java che consentano di costruire rapidamente l'interfaccia con i Database ed il mapping Object-Relational (ORM). La tesina prevede il confronto dello strumento Spring Roo (utilizzato da Tom Tom) con gli altri principali tools.

Riferimenti:

<http://www.springsource.org/spring-roo> http://it.wikipedia.org/wiki/Rapid_Application_Development

Storico

(tesine già assegnate)

Realizzazione di una Applicazione Web

Tutor: Maurizio Vincini

Data: Ottobre 2012

Descrizione: Il gruppo di lavoro deve realizzare un'applicazione informatica che risponda alle specifiche applicative in seguito descritte e che permetta di gestire in modo opportuno tutte le situazioni che tali specifiche di alto livello non comprendono. L'applicazione deve essere documentata attraverso il progetto concettuale, logico e fisico, deve memorizzare i dati in un database, utilizzando un DBMS opportunamente scelto. L'applicazione database deve essere realizzata secondo una modalità client/server o web- based. Il linguaggio di programmazione per la parte applicativa può essere scelto dal gruppo di lavoro.

Specifiche dei requisiti:

- . 1.1. Valutazione dei sistemi per la gestione della sincronizzazione dei principali DBMS installati su dispositivi palmari
- . 1.2. Analisi e progetto di allineamento delle informazioni tra DBMS server e DBMS su dispositivi palmari
- . 1.3. Accesso dell'utenza ad un sito Web: valutazione di criteri di raccolta e gestione di Statistiche all'interno di applicazioni Web Database tramite software open source

Sperimentazione della libreria SQLite per la realizzazione di applicazioni per smartphone

Tutor: Serena Sorrentino, Sonia Bergamaschi

Data: Ottobre 2012

Descrizione: SQLite è una libreria software scritta in linguaggio C che implementa un DBMS SQL di tipo ACID incorporabile all'interno di applicazioni (per maggiori dettagli visitare il sito <http://www.sqlite.org>). Il candidato dovrà studiare la libreria SQLite al fine di delinearne i vantaggi e gli svantaggi rispetto alle soluzioni alternative proposte in letteratura. Dovrà, inoltre, proporre ed implementare una semplice applicazione per smartphone (iPhone, o altri Smartphone) che faccia uso della libreria.

Applicazione web-database per la gestione delle pubblicazioni di un gruppo di ricerca

Tutor: Marius Octavian Olaru, Sonia Bergamaschi

Data: Ottobre 2012

Descrizione: Il progetto prevede la realizzazione di un'applicazione web-database che gestisca gli articoli di un gruppo di ricerca.

I requisiti che l'applicazione dovrà soddisfare sono:

- inserimento, modifica, cancellazione di articoli;
 - esecuzione di ricerche elementari (in base a titolo e/o autore);
 - memorizzazione persistente dei file allegati agli articoli (su DBMS o filesystem);
 - esportazione dei riferimenti di articoli nel formato BibTeX
<http://www.bibtex.org/Format/>. Requisiti opzionali:
 - implementazione di un metodo di indicizzazione e ricerca avanzata, usando strumenti come Apache Lucene;
 - prevedere un meccanismo di gestione degli accessi;
 - implementazione di un Web Service che permetta l'interrogazione del database degli articoli.
- Strumenti richiesti:
- DBMS (consigliato MySQL)
 - linguaggio web dinamico: consigliato JSP

Sistemi di Business Intelligence : Realizzazione di cubi multidimensionali

Effettuata da: GRISENDI GIULIA, DODA XHENSILA e PACCHIONI MARCO

Tutor: Domenico Beneventano

Data: Ottobre 2013

Descrizione: Il gruppo di lavoro, partendo da un data warehouse già costruito, deve utilizzare un Sistema di Business Intelligence per realizzare cubi OLAP multidimensionali. I requisiti della tesina sono:

- Scelta del Sistema di Business Intelligence, sia tra quelli commerciali (es. Analysis Services-SQL SERVER) che open source (es. Pentaho Analysis Services - Mondrian)
 - Progettazione e realizzazione di cubi multidimensionali
- Analisi dei dati dei cubi OLAP tramite interrogazioni MDX oppure tramite strumenti propri del Sistema di Business Intelligence utilizzato

Record Linkage

Tutor: Domenico Beneventano

Data: Ottobre 2012

Descrizione: Con Record Linkage si intende il problema di identificare – in maniera efficace ed efficiente - record che rappresentano lo stesso oggetto del mondo reale (il problema è anche noto con il termine di Object Identification, Duplicate Detection e Entity Matching).

Il Record linkage svolge un ruolo chiave in tutti i sistemi (tra i principali: data integration, data

warehousing, open data) in cui i dati provenienti da diverse sorgenti devono essere integrati/combinati.

I requisiti della tesina sono: - Analisi di strumenti (prototipi e commerciali) di Record Linkage quali ad esempio

- o “Fuzzy Matching Component” di SQL Server Integration Services
- o Duplicate Detection (DuDe) toolkit

(www.hpi.uni-potsdam.de/naumann/projekte/dude_duplicate_detection.html) - Valutazione degli strumenti individuati in casi reali, quali ad esempio

Gestione delle allerte astronomiche mediante BerkleyDB e visualizzazione dati su dispositivo mobile Apple iOS.

Tutor: Matteo Interlandi, Sonia Bergamaschi

Data: Marzo 2013

Descrizione : AGILE (<http://agile.iasfbo.inaf.it>) è un satellite scientifico lanciato nel 2007 dedicato all'Astrofisica delle Alte Energie progettato per rivelare i raggi gamma provenienti dall'universo. AGILE dispone di un sistema di allerta in grado di rilevare improvvise e violente esplosioni (flare) provenienti da oggetti celesti spesso ignoti.

Ice, un framework per lo sviluppo di applicazione in ambiente distribuito in cui la persistenza degli oggetti è realizzata mediante il DBMS NoSQL BerkleyDB. Inoltre Ice permette di integrare le piattaforme Apple iOS nell'ambiente distribuito.

L'obiettivo del lavoro è quello di studiare l'integrazione tra un database NoSQL come BerkleyDB e i dispositivi mobili. In particolare, bisognerà memorizzare i dati sui flare gamma

per la gestione dei dati

come oggetti Ice nel BerkleyDB, comprese le mappe del cielo, e permetterne l'interrogazione e la visualizzazione mediante interfaccia Apple iOS.

Riferimenti: BerkleyDB: www.oracle.com/technetwork/products/berkeleydb/overview/index.html
Oracle NOSQL: www.oracle.com/technetwork/products/nosqldb/overview/index.html Ice: www.zeroc.com

Main Memory Database System

Effettuata da : Malvasi Enrico, Candini Alessandro, Po Mirko

Tutor: Laura Po, Sonia Bergamaschi

Data: Ottobre 2013 (assegnata ma ancora disponibile)

Descrizione: Un in-memory database (IMDB) o anche main memory database system (MMDB) è un sistema di gestione di database che utilizza principalmente la memoria principale per l'archiviazione di dati.

Questo sistema si pone in contrasto con i sistemi di gestione di database che utilizzano un meccanismo di archiviazione su disco.

I database archiviati in memoria principale sono più veloci di quelli archiviati su disco in quanto gli algoritmi di ottimizzazione interna sono più semplici ed eseguono meno istruzioni di CPU. L'accesso ai dati in memoria riduce l'attività di I/O in fase di interrogazione, questo migliora le performance dei sistemi rispetto a quelli su disco. Nelle applicazioni in cui il tempo di risposta è un fattore critico, i database in memoria principale sono spesso utilizzati.

Alcuni prodotti: – ASE (Adaptative Server Enterprise) - Sybase

– TimesTen - Oracle

– MonetDB Viene richiesto di effettuare un'analisi di vantaggi e svantaggi dei MMDB e dei principali prodotti disponibili. Riferimenti:

- Primo articolo pubblicato <http://www.inf.uni-konstanz.de/dbis/teaching/ws0203/main-memory-dbms/download/MainMemoryDatabaseSystemsAnOverview.pdf>

- Wikibooks -http://en.wikibooks.org/wiki/Design_of_Main_Memory_Database_System

Comparazione tra BerkleyDB e Oracle NOSQL nella gestione di allerte astronomiche

Effettuata da: Mirko Montanari e Alberto Paganelli

Tutor: Sonia Bergamaschi, Giovanni Simonini

Data: Marzo 2013

Descrizione : AGILE (<http://agile.iasfbo.inaf.it>) è un satellite scientifico lanciato nel 2007 dedicato all'Astrofisica delle Alte Energie progettato per rivelare i raggi gamma provenienti dall'universo. AGILE dispone di un sistema di allerta in grado di rilevare improvvise e violente esplosioni (flare) provenienti da oggetti celesti spesso ignoti.

L'obiettivo del lavoro è quello di memorizzare i dati sui flare gamma in un DBMS NoSQL quale BerkleyDB (BDB) e Oracle NOSQL, effettuando un preliminare studio teorico sui due DBMS (pro e contro) e misurando le performance delle due soluzioni (su memorizzazione e interrogazione dei dati).

Opzionalmente, si potrebbe utilizzare Ice, un framework per lo sviluppo di applicazione in ambiente distribuito in cui la persistenza degli oggetti è realizzata mediante BDB. Ogni flare gamma diventerebbe un oggetto da memorizzare nel DBMS NoSQL e si potrebbe studiare l'integrazione di BDB e Oracle NoSQL in Ice e le relative performance.

Riferimenti: BerkleyDB: www.oracle.com/technetwork/products/berkeleydb/overview/index.html

Oracle NOSQL: www.oracle.com/technetwork/products/nosqldb/overview/index.html Ice: www.zeroc.com

Applicazione web in ambiente Google App Engine, settore Estetica

Tutor: Sonia Bergamachi

Data: marzo 2015

Descrizione:

Progetto, sviluppo e realizzazione di una applicazione web per la gestione dei clienti/prodotti/servizi offerti da un centro

estetico. Il progetto verrà sviluppato in ambiente Google App Engine, con tecnologia Python, sfruttando *NDB Datastore*, un database non relazionale.

L'applicazione web dovrà gestire i clienti del centro estetico, i trattamenti che effettuano, i prodotti che acquistano ecc.

Riferimenti:

<https://cloud.google.com/appengine/docs>

<https://cloud.google.com/appengine/docs/python/ndb/>