Università degli Studi di Roma “Tor Vergata”

**Facsimile Scheda Insegnamento**

Docente responsabile dell’insegnamento/attività formativa

Nome: Roberto

Cognome: BASILI

Denominazione insegnamento/attività formativa

Italiano: **Basi di Dati e Conoscenza**

Inglese: Database Management Systems

Informazioni insegnamento/attività formativa

A.A.: 2021/22 Laurea: L

CdS: Informatica

Codice

Canale: unico

CFU: 12

Lingua: Italiana

Docente del modulo didattico (compilare solo per attività formative articolate in moduli)

NA

Denominazione modulo didattico (compilare solo per attività formative articolate in moduli)

NA

Obiettivi formativi e risultati di apprendimento attesi

* OBIETTIVI FORMATIVI:
* CONOSCENZA E CAPACITÀ DI COMPRENSIONE:
* CAPACITÀ DI APPLICARE CONOSCENZA E COMPRENSIONE:
* AUTONOMIA DI GIUDIZIO:
* ABILITÀ COMUNICATIVE:
* CAPACITÀ DI APPRENDIMENTO:

Il Corso ha l’obbiettivo di introdurre i formalismi, le metodologie e le tecnologie per la gestione dei Dati, con una particolare enfasi per gli standard industriali legate alle Basi di Dati Relazionali. Nel Corso vengono anche introdotte le prassi principali in ambito industriale, relative ai linguaggi di progettazione (ER) e definizione dei dati (ER, SQL) sino agli standard di progettazione applicativa in ambito Web, come JDBC e PHP. Tra gli argomenti avanzati si presentano i limiti formali delle Basi di Dati relazionali e si introducono le nozioni di base sui sistemi basati su conoscenza. Si introducono le evoluzioni sulle tecnologie non relazionali (noSQL) utilizzate negli scenari di Big Data Analytics e gli elementi di base del *Data Mining*.

Al termine del corso, lo studente avrà acquisito le competenze necessarie per comprendere la progettazione dei modelli logici dei dati e il loro impatto nello sviluppo delle applicazioni (\*conoscenza e capacità di comprensione\*). In particolare, lo studente avrà avuto modo di apprendere gli strumenti e le tecnologie per progettare basi di dati di media complessità negli scenari di applicazione industriale delle basi di dati relazionali (\*capacità di applicare conoscenza e comprensione\*). Il riferimento a contesti applicativi e la necessità di individuare gli elementi essenziali delle informazioni presenti nei dati e la capacità di progettarne le opportune interrelazioni nella analisi di un dominio applicativo o di un processo di business realizzano una forte \*autonomia di giudizio\* nello studente, che è un obbiettivo rilevante del Corso. Osserviamo che la analisi richiesta nella progettazione logica dei dati agisce su documentazione tecnica (requisiti) o su dati di fonti aperte (ad es. pagine Web) che corrispondono a dati tipici della comunicazione organizzativa o mediatica. Queste competenze dunque stimolano in modo sistematico le \*abilità comunicative\* dello studente e le capacità di agire in modo consistente da consumatore o produttore di informazione mediata dalla tecnologia. La \*capacità di apprendimento\* in questo Corso è dunque stimolata in modo significativo sia nei processi interpretativi che nei processi di progettazione, entrambi, pur agenti su testi liberi, orientati alla creazione di una versione formalizzata delle informazioni e conoscenze dei domini di *business* considerati.

(Inglese)

* LEARNING OUTCOMES:
* KNOWLEDGE AND UNDERSTANDING:
* APPLYING KNOWLEDGE AND UNDERSTANDING:
* MAKING JUDGEMENTS:
* COMMUNICATION SKILLS:
* LEARNING SKILLS:

The aim of the Course is to introduce the student to formalisms, methodologies, technologies and best practices on data management processes, with a specific emphasis on the industrial standards of Relational Database Management systems (RDBMs). The course presents an overview on the role of DBs on the programming and design practices, by introducing data design languages (ER, SQL) and programming paradigms, such as JDBC e PHP. An introduction to the notion of automated reasoning and knowledge based system is also presented as a relevant connection between the course focus and the current industry Web practices. Further advanced topics (such as noSQL, Multimedia Databases, and Data Mining) are used as a motivating introduction to Big Data Analytics scenarios.

At the end of the course, the student is expected to own the knowledge needed in the understanding of data design activities and their impact in the development of digitally mediated business applications (\*knowledge and acquisition of skills\*). In particular, the student will master industry standard tools and systems for the design of small or medium sized databases in application scenarios based on relational DBs (\*comprehension and application capabilities\*). The data analysis methods that the student will master allow him to proceed in autonomy in the data interpretation and design processes enforcing his own strong \*autonomy of judgment\*, that is a relevant target of the Course. Notice that data interpretation acts on technical or Web documents that correspond to typical on-line organisational and communication. These capabilities thus trigger in a systematic way the student’s \*communication abilities\*. Finally, \*learning skills\* are triggered in the Course as the data analysis and formalisation processes on the one side are fed by text (i.e. vague and unstructured) data but are also strongly rooted in technological settings, thus enforcing an integrated use of generalization and technological practices.

Prerequisiti

Non esistono propedeuticità obbligatorie da rispettare. E' tuttavia vivamente consigliato come prerequisito avere sostenuto gli esami legati alla introduzioni di base all’Algebra ed alla Logica e l’insegnamento di: *Fondamenti di Informatica*.

Inglese

There is no mandatory prerequisite to meet. However, it is highly recommended to have passed the exams of courses about Algebra and Logic (such as Calculus I (Analisi Matematica 1), Geometry (Geometria)) and in particular, Foundations of Computer Science (Fondamenti di Informatica).

Programma

**Introduzione**. Introduzione ai sistemi di basi di dati. Gestione dei dati. Una prospettiva storica. *File system* e DBSM. Vantaggi di un DBMS. Interrogazioni in un DBMS. Gestione delle transazioni. Struttura di un DBMS.

**Il modello relazionale.** Introduzione al modello relazionale. Chiavi e Vincoli di integrità sulle relazioni. Applicazione dei vincoli di integrità. Interrogazione di basi di dati relazionali. Introduzione alle viste.

**Algebra relazionale.** Introduzione. Algebra relazionale. Selezione e proiezione. Operazioni sugli insiemi. Join. Esempi di interrogazioni algebriche

**SQL.** Introduzione ad SQL. Interrogazioni SQL di base. Operatori di UNION, INTERSECT e EXCEPT. Interrogazioni annidate. Interrogazioni annidate correlate. Operatori di confronto tra insiemi. Operatori di aggregazione: clausole GROUP BY e HAVING. Valori *null*. Join esterni. Vincoli di integrità complessi in SQL. Asserzioni su più tabelle. *Trigger* e basi di dati attive. Vincoli e *trigger*

**Il modello Entità-Relazione.** Progettazione di basi di dati e diagrammi ER.

Entità, attributi e insiemi di entità. Relazioni e insiemi di relazioni. Estensioni del modello ER. Vincoli di chiave. Vincoli di partecipazione. Entità deboli. Gerarchie di classi. Aggregazione. Progettazione concettuale con il modello ER. Progettazione concettuale per grandi organizzazioni

**La progettazione logica: dallo schema ER al relazionale.** Da insiemi di entità e relazioni a tabelle. Traduzione di insiemi di relazioni con vincoli di chiave. Traduzione di insiemi di relazioni con vincoli di partecipazione. La traduzione di insiemi di entità deboli. Traduzione delle gerarchie di classi. Traduzione di diagrammi ER con aggregazione.

**Basi di Dati e Applicazioni.** L’accesso alle basi di dati da parte delle applicazioni. SQL incapsulato. Cursori. SQL dinamico. Introduzione a JDBC. Classi e interfacce JDBC. Esecuzione dei comandi SQL. SQLJ. *Stored procedure.* Applicazioni Web:Introduzione alle reti di calcolatori e Web. Documenti HTML. Documenti XML. Introduzione all’XML. L’architettura delle applicazioni *three-tier.* Il livello di presentazione. Il livello intermedio.

**File e indici.** Memorizzazione esterna dei dati. Organizzazioni dei *ﬁle* e indicizzazione. Strutture di dati per gli indici. Indici *hash.* Indici ad albero. Confronto tra organizzazioni di *ﬁle.* Indici e miglioramento delle prestazioni. Definizione degli indici in SQL:1999.

**Esecuzione delle interrogazioni.** Il catalogo di sistema. Introduzione alla valutazione degli operatori. Percorsi di accesso. Algoritmi per operazioni relazionali. Introduzione all’ottimizzazione delle interrogazioni. Piani di valutazione delle interrogazioni

**Tendenze evolutive delle Basi di Dati.**

Limiti delle basi di dati: introduzione ai sistemi basati su conoscenza. Metadati e DB multimediali. Cenni ai paradigmi NoSQL. Introduzione ai concetti di base di Data Mining.

Nella seconda parte del corso, agli studenti viene richiesto il completamento di un progetto applicativo che utilizza un database. E’ prevista dunque la implementazione di una semplice interfaccia funzionale (per esempio attraverso un Web browser) ad una base di dati relazionale, la cui progettazione logica ed il cui popolamento (a partire da sorgenti informative pubbliche distribuite ed indipendenti) costituiscono il fulcro del progetto. Tipici esempi di dati pubblici disponibili e spesso eterogenei logicamente sono gli Open Data di istituzioni pubbliche o gli *stream* dei social network. La riconciliazione di fonti così diverse è assunta come sfida applicativa di grande valore per la progettazione.

Durante il corso sono previste attività di didattica integrativa (DI) oltre alla didattica erogata (DE) secondo le linee guida ANVUR, principalmente dedicate agli incontri ed ai test validativi legati al progetto. Agli studenti è richiesto in particolare di partecipare attivamente a tali incontri sulla base della formazione dei team di progetto prevista. SI prevede che alla didattica integrativa siano dedicate almeno 6 ore, ovvero minimo 1 ora per ogni CFU del corso, per ogni studente e team.

(Inglese)

Introduction. Introduction to database systems. Data management. Historical overview. File System and DBSM. DBSM advantages. DBSM queries. Transactions management. DBSM structure.

Relational model. Introduction to the Relational model. Relation Keys and Constraints of integrity. Constraints integrity application. Relational database query. Introduction to the Views.

Relational algebra. Introduction. Relational algebra. Selection and projection. Operations on sets. Join. Examples of algebraic queries.

SQL. Introduction to SQL. Basic SQL queries. UNION, INTERSECT and EXCEPT operations. Inserted queries. Related inserted queries. Comparison operators between sets. Aggregation operator: GROUP BY and HAVING conditions. Null value. External Joins. Complex integrity constraints in SQL. Statements on many tables. Trigger and active databases. Constraints and Triggers.

Entity-Relationship model. Databases and ER diagrams design. Entity, attributes and sets. Relation and sets of relations. Extensions of ER model. Key constraints. Participation constraints. Weak entities. Class hierarchies. Aggregations. Conceptual design with ER model. Conceptual design for big groups.

Logical design: from the ER model to the Relational one. From entities and relations sets to tables. Relation sets translation with key constraints. Relation sets translation with participation constraints. Weak entities sets translation. Class hierarchy translation. ER diagrams translation with aggregation.

DB & Applications. Application access to databases. Encapsulated SQL. Pointers. Dynamic SQL. Introduction to JDBC. JDBC class and interfaces. SQL command execution. SQLJ. Stored procedure. Internet applications: Introduction to the computer network and Web. HTML documents. XML documents. Introduction to XML. Three-tier application architecture. Presentation level. Intermediate level.

File and indexes. External data storage. File organization and indexing. Data structures for indexes. Hash indexes. Tree-structured indexes. Comparison between file organizations. Performances improvement and indexes. SQL: 1999 definition of indexes.

Queries execution. Catalog system. Introduction to operator evaluation. Access paths. Relational operations algorithm. Introduction to queries optimization. Queries evaluation plans.

Advanced topics. From relational DBs to Reasoning systems: introduction to Knowledge Based Systems. Metadata and multimedia DB. Introduction to NoSQL paradigms. Introduction to Data Mining.

Modalità di valutazione

* → Prova scritta
* → Prova orale (facoltativa)
* Valutazione in itinere
* → Valutazione di progetto
* Valutazione di tirocinio
* Prova pratica
* Prova di laboratorio

Descrizione delle modalità e dei criteri di verifica dell’apprendimento

Italiano

In due test di midterm (equivalenti ad una singole prove orale), vengono utilizzate domande chiuse per la verifica dei temi dell’intero programma, ed una prova di progettazione di un DB. Il progetto verte sullo sviluppo di una piccola applicazione software che usa il DB e quindi richiede l’apprendimento delle tecnologie per l’interfacciamento tra i programmi applicativi e un DBMS, che richiedono la competenza nei protocolli di interoperabilità dati e applicazioni più comuni (PHP o JDBC).

La verifica delle competenze tende ad studiare le conoscenze acquisite in ampiezza sui temi del programma e la capacità modellistica del dominio (nella prova di progettazione).

Infine le competenze tecnologiche SQL, Java e JDBC (o PHP) sono legate allo sviluppo del progetto dove il contributo su questi temi fornito dal singolo studente del team (max 3 persone) viene verificato con una discussione orale.

Inglese

Written and Oral. Two MidTerm can be used as an equivalent of the final full test. A (non mandatory) project on the design of a small application (possibly Web-based) insisting on a RDB is required, and its oral presentation is considered a possible part of the exam.

Testi adottati

* *Slides ed altri materiali digitali messi a disposizione on-line.*
* **Versione Italiana.** "*Sistemi di Basi di Dati*", di Raghu Ramakrishnan e Johannes Gehrke, Edizione Italiana, McGraw Hill, 2004

Inglese

* *Teaching material made available on-line.*
* "*Database Management Systems*", 3rd Edition, Raghu Ramakrishnan e Johannes Gehrke, McGraw Hill, 2000 (ISBN 0-07-232206-3).

Bibliografia di riferimento

* **Versione Italiana.** "*Sistemi di Basi di Dati*", di Raghu Ramakrishnan e Johannes Gehrke, Edizione Italiana, McGraw Hill, 2004
* **Basi di Dati:** "*Database Management Systems*", 3rd Edition, Raghu Ramakrishnan e Johannes Gehrke, McGraw Hill, 2000 (ISBN 0-07-232206-3).
* *BASI DI DATI*, Paolo Atzeni, Stefano Ceri, Piero Fraternali, Stefano Paraboschi e Riccardo Torlone, Edizione 5, 2018, pp. 832. ISBN: 9788838694455.

Inglese

* "*Database Management Systems*", 3rd Edition, Raghu Ramakrishnan e Johannes Gehrke, McGraw Hill, 2000 (ISBN 0-07-232206-3).
* *BASI DI DATI*, Paolo Atzeni, Stefano Ceri, Piero Fraternali, Stefano Paraboschi e Riccardo Torlone, Edizione 5, 2018, pp. 832. ISBN: 9788838694455.

Modalità di svolgimento

* → Modalità in presenza
* Modalità a distanza

Descrizione della modalità di svolgimento e metodi didattici adottati

Italiano

Vengono erogate soprattutto lezioni frontali associate ad esercitazioni su temi selezionati. Si erogano poi alcune esercitazioni simulative della situazione d’esame, con la discussione collegiale delle soluzioni. Alcune delle lezioni frontali sono dedicate alla discussione delle domande chiuse d’esame in preparazione delle prove di esonero e della prova finale.

Inglese

The course consists of lecture style lessons, augmented with sessions for Question answering with students. A specific emphasis is taken by specific lessons dedicated to specific topics and practices (e.g. solving relational algebra problems or complex usage of SQL). Some lessons are collective and devoted to discuss the solution of exam topics, such as the design of the ER of a domain and application. One lesson is always devoted to introduce the methods and types of closed questions of the exam written session.

Modalità di frequenza

* → Frequenza Facoltativa
* Frequenza Obbligatoria

Descrizione della modalità di frequenza

Italiano

Lezioni frontali, esercitazioni su temi selezionati e discussione collegiale delle soluzioni in aula.

Inglese

Lecture style lessons, lab practices discussed in the classroom, and collective discussion about the exam solutions.