



# UNIVERSITÀ DI PISA

---

## LARGE-SCALE AND MULTI-STRUCTURED DATABASES

**PIETRO DUCANGE**

Anno accademico

2020/21

CdS

ARTIFICIAL INTELLIGENCE AND  
DATA ENGINEERING

Codice

883II

CFU

9

Moduli	Settore/i	Tipo	Ore	Docente/i
LARGE-SCALE AND MULTHNG- STRUCTURED DATABASES	INF/05	LEZIONI	90	PIETRO DUCANGE

### Obiettivi di apprendimento

#### *Conoscenze*

Al termine del corso:

- lo studente avrà acquisito conoscenze in merito agli strumenti e alle metodologie per la progettazione di basi di dati non relazionali. In particolare, lo studente sarà in grado di gestire l'archiviazione, l'aggiornamento ed il recupero di dati complessi e multi-strutturati, anche di dimensioni molto ampie.
- lo studente avrà acquisito conoscenze in merito alle architetture, alle prestazioni ed ai costi delle moderne infrastrutture per la gestione di dati complessi, sia dal punto di vista della quantità di informazione sia dal punto di vista della loro struttura.

Con riferimento all'autonomia di giudizio:

- lo studente sarà in grado di impostare correttamente un progetto per la gestione di dati multi-strutturati e di grosse dimensioni, integrandolo all'interno di una reale applicazione informatica e scegliendo in maniera opportuna le strategie di progetto e di implementazione.

Con riferimento alle abilità comunicativa:

- lo studente sarà in grado di utilizzare del lessico e della terminologia appropriata per presentare, in forma scritta o verbale, le argomentazioni e i risultati del proprio studio e delle attività pratiche realizzate durante il corso

Con riferimento all'abilità ad apprendere:

- Lo studente avrà acquisito un buon livello di autonomia nell'utilizzo del materiale e negli strumenti didattici presentati durante il corso

#### *Modalità di verifica delle conoscenze*

Per l'accertamento delle conoscenze teoriche e pratiche saranno proposte attività di gruppo. Ai gruppi di lavoro saranno proposte attività di:

- approfondimento di questioni teoriche e tecniche
- realizzazione di progetti.
- Saranno previsti confronti periodici in aula tra il docente e il gruppo di studenti che sviluppa le attività di cui sopra.
- I gruppi di lavoro dovranno essere in grado di impostare in maniera ottimale il progetto didattico. Ciò sarà valutato sulla base della documentazione fornita e della discussione delle varie attività durante le ore appositamente dedicate ed in sede di esame orale.
- In sede di esame orale, allo studente sarà data la possibilità di discutere problematiche teorico/pratiche relative agli argomenti trattati nel corso.

#### *Capacità*

Al termine del corso lo studente sarà in grado di:

- Progettare opportunamente una base di dati non relazionale, sulla base dei requisiti (funzionali e non) di una specifica applicazione.
- Utilizzare moderne infrastrutture tecnologiche per la gestione di basi di dati non relazionali (MongoDB, Neo4j, etc)

#### *Modalità di verifica delle capacità*

Durante le lezioni di laboratorio:

- Sarà mostrato allo studente come installare e configurare alcune fra le moderne infrastrutture tecnologiche per la gestione di basi di dati non relazionali
- Saranno proposte attività pratiche per la creazione, la gestione e l'interrogazione di diverse basi di dati non relazionali
- Saranno proposte attività di gruppo per l'approfondimento di questioni tecniche e per la realizzazione di progetti didattici



## UNIVERSITÀ DI PISA

---

- I gruppi di lavoro dovranno presentare relazioni scritte relative alle attività svolte

### Comportamenti

Gli studenti acquisiranno accuratezza e precisione nella progettazione e risoluzione dei problemi relativi alla progettazione di basi di dati non relazionali.

Gli studenti saranno in grado di collaborare con i propri colleghi e svolgere un lavoro di gruppo in modo efficace.

### Prerequisiti (conoscenze iniziali)

- Programmazione in JAVA (incluso l'utilizzo di un IDE)
- Progettazione ed interrogazione di basi di dati relazionali
- Nozioni di base di Ingegneria del Software (incluso realizzazione di Diagrammi UML)
- Nozioni di base di Sistemi Operativi Unix-Based

### Indicazioni metodologiche

Per l'anno accademico 2020/2021 le lezioni saranno svolte in modalità blended. Per le prime 6 settimane le lezioni ed esercitazioni si terranno in modalità a distanza. Successivamente, 4 ore di esercitazioni a settimana si terranno in presenza in laboratorio. In ogni caso, sarà sempre data l'opportunità a tutti gli studenti di seguire anche tali ore di esercitazione in modalità a distanza.

Le esercitazioni in presenza si svolgeranno in laboratorio informatico, si formeranno gruppi di lavoro che dovranno utilizzare i PC personali.

Si utilizzerà la piattaforma di e-learning Google Classroom per fornire il materiale di studio e per il caricamento da parte degli studenti di artefatti (codici, relazioni, presentazioni). Gli studenti potranno accedere a tale piattaforma solo ed esclusivamente utilizzando le credenziali di ateneo.

Il corso sarà svolto interamente in lingua Inglese.

Saranno previsti orari di ricevimento (due ore alla settimana).

### Programma (contenuti dell'insegnamento)

- Introduzione e motivazioni: Introduzione al corso, L'era dei Big Data, Le rivoluzioni delle Basi di Dati
- Fondamenti e proprietà dei database NoSQL: Proprietà ACID vs proprietà BASE, il teorema del CAP, Scalabilità, Sharding, Replicazione, Coerenza.
- Architetture di database NOSQL: Document Databases, Key-values Databases, Column Databases, Graph Databases
- Ripasso di Java, SQL e strumenti per la persistenza dei dati
- Infrastrutture moderne per basi di dati NoSQL: LevelDB, MongoDB, Neo4J (installazione, configurazione, operazioni CRUD, query principali)

### Bibliografia e materiale didattico

Slide fornite dal docente

Articoli scientifici forniti dal docente

Libri Consigliati:

"Guy Harrison, Next Generation Databases, Apress, 2015"

"Dan Sullivan, NoSQL For Mere Mortals, Addison-Wesley, 2015"

"Andreas Meier, Michael Kaufmann, SQL & NoSQL Databases: models, languages, consistency options and architectures for big data management, 2019"

### Indicazioni per non frequentanti

Gli studenti che non frequenteranno il laboratorio, dovranno contattare il docente per stabilire le attività pratiche da svolgere in alternativa.

### Modalità d'esame

Discussione attività di gruppo (in unica volta) e test scritto.

Realizzazione delle attività di gruppo (approfondimenti teorici e progetti didattici). Gli studenti devono fornire al docente opportuna documentazione e presentare i lavori svolti prima della prova orale. Sarà fissata di volta in volta la data per la discussione finale delle attività di gruppo.

Test scritto per la discussione degli argomenti teorici.

Il test scritto e la realizzazione delle attività di gruppo contribuiscono ciascuno al 50% sulla votazione finale.

Ultimo aggiornamento 28/08/2020 18:16