



# UNICUSANO

Università degli Studi Niccolò Cusano - Telematica Roma

<b>Insegnamento</b>	Algoritmi e strutture dati per i Big Data
<b>Livello e corso di studio</b>	Laurea Magistrale in Ingegneria Informatica (LM32)
<b>Settore scientifico disciplinare (SSD)</b>	ING-INF/05
<b>Anno di corso</b>	2
<b>Numero totale di crediti</b>	9
<b>Propedeuticità</b>	-
<b>Presentazione</b>	Il Corso di Algoritmi e strutture dati per i Big Data ha lo scopo di far acquisire allo studente una buona conoscenza dei principi che governano la progettazione di sistemi per il trattamento di ingenti moli di dati e permettono di modellarne il comportamento. Il Corso propone i concetti basilari delle architetture distribuite al servizio del processamento ed immagazzinamento di Big Data e li declina nello studio degli algoritmi e, più in generale, delle tecniche di pre-elaborazione, riduzione dimensionale, clustering, classificazione e predizione. Inoltre, obiettivo formativo del Corso è fornire allo studente una conoscenza nel dettaglio sugli approcci alla memorizzazione e strutturazione dei dati, sia relazionali che non-relazionali. Le Etivity associate al Corso sviluppano le competenze necessarie a progettare ed analizzare sistemi ed algoritmi per i Big Data attraverso l'uso del calcolatore.
<b>Obiettivi formativi</b>	Il Corso di Informatica ha i seguenti obiettivi formativi: <ol style="list-style-type: none"> <li>1. Illustrare i principi basilari dei sistemi di gestione di grandi moli di dati</li> <li>2. Illustrare i diversi approcci architetture allo stato dell'arte</li> <li>3. Illustrare i diversi approcci algoritmici allo stato dell'arte</li> </ol>
<b>Risultati di apprendimento attesi</b>	<p><b>Conoscenza e capacità di comprensione</b> Lo studente al termine del Corso avrà conoscenza dei modelli di calcolo, delle architetture e delle infrastrutture necessarie all'elaborazione di grandi moli di dati. Lo studente, inoltre, avrà familiarità con le problematiche legate all'analisi di grandi moli di dati. Inoltre, tramite le Etivity gli studenti acquisiranno la capacità di analizzare infrastrutture e sistemi per i Big Data.</p> <p><b>Applicazione delle conoscenze</b> Lo studente sarà in grado di analizzare scenari caratterizzati dalla presenza di grandi moli di dati; sarà inoltre in grado di fornire opportune soluzioni progettuali per la realizzazione di sistemi in grado di gestire tali dati, e di progettare sistemi software efficienti per l'elaborazione di grandi moli di dati. Le Etivity prevedono l'applicazione delle conoscenze teoriche alla progettazione ed implementazione di applicazioni basate sui Big Data all'interno di ambienti per la programmazione orientata al data analytics.</p> <p><b>Capacità di trarre conclusioni</b> Lo studente sarà in grado di valutare la bontà delle diverse soluzioni algoritmiche e architetture per la gestione di grandi moli di dati e sarà inoltre in grado di valutare le performance dei diversi approcci interpretando opportuni indicatori. Infine, lo studente sarà in grado di effettuare ricerche bibliografiche, di analizzare ed interpretare le fonti rilevanti, al fine di analizzare nuovi paradigmi, approcci, architetture ed algoritmi per il processamento di grandi moli di dati.</p> <p><b>Abilità comunicative</b> Lo studente sarà in grado di descrivere e sostenere conversazioni su tematiche relative alla progettazione e gestione di sistemi per la gestione di grandi moli di dati., ed alla risoluzione di problemi tipici di tali sistemi, adoperando una terminologia adeguata.</p> <p><b>Capacità di apprendere</b> Lo studente al termine del Corso avrà conoscenza delle nozioni fondamentali necessarie per l'analisi e la progettazione di diverse applicazioni che richiedono l'uso di Big Data. Questo gli consentirà di identificare gli strumenti necessari ad apprendere autonomamente i principi di funzionamento di nuovi tool per la gestione dei big data.</p>

<b>Prerequisiti</b>	-
<b>Organizzazione dell'insegnamento</b>	<p>Il Corso è sviluppato attraverso le <b>lezioni preregistrate audio-video</b> che compongono, insieme a slide e dispense, i materiali di studio disponibili in piattaforma.</p> <p>Sono poi proposti dei <b>test di autovalutazione</b>, di tipo asincrono, che corredano le lezioni preregistrate e consentono agli studenti di accertare sia la comprensione, sia il grado di conoscenza acquisita dei contenuti di ognuna delle lezioni.</p> <p>La <b>didattica interattiva</b> è svolta nel forum della “classe virtuale” e comprende <b>4 Etivity</b> che applicano le conoscenze acquisite nelle lezioni di teoria all’analisi ed alla progettazione di applicazioni basate sui Big Data attraverso l’uso di ambienti per la programmazione orientata al data analytics.</p> <p>In particolare, il Corso di Algoritmi e strutture dati per i Big Data prevede 9 Crediti formativi. Il carico totale di studio per questo modulo di insegnamento è di circa 225 ore così suddivise in:</p> <ul style="list-style-type: none"> <li>● circa 170 ore per la visualizzazione e lo studio del materiale videoregistrato;</li> <li>● circa 50 ore di Didattica Interattiva per l’elaborazione e la consegna di 4 Etivity;</li> <li>● circa 5 ore di Didattica Interattiva per l’esecuzione dei test di autovalutazione.</li> </ul>
<b>Contenuti del corso</b>	<p><b>Modulo 1 – Introduzione (impegno di 21 ore)</b> dove sono affrontati i seguenti argomenti: Introduzione ai concetti di base dei Big Data: terminologia, aspetti principali ed esempi di applicazioni. Problematiche principali legate alla gestione dei Big Data: volume dei dati e occupazione di memoria, velocità dell’elaborazione e complessità computazionale, presenza di errori e accuratezza dei dati, comprensione dei dati.</p> <p><b>Modulo 2 – Infrastrutture (impegno di 27,5 ore)</b> dove sono affrontati i seguenti argomenti: Infrastrutture per la gestione dei Big Data: Architetture Distribuite e Parallele; Cloud Computing per i Big Data. Esercitazione su Cloud per il Big Data.</p> <p><b>Etivity 1 (15 ore di carico di studio)</b> – Progettazione di sistemi per la gestione distribuita di big data.</p> <p><b>Modulo 3 – Memorizzazione e pre-elaborazione (impegno di 26 ore)</b> dove sono affrontati i seguenti argomenti: Memorizzazione dei Big Data: Memorizzazione strutturata; Database non relazionali; Tipologie di database NoSql; Tecniche di pre-elaborazione dei Big Data: Tipi di errori; Gestione degli errori. Esercitazione su DB NoSql.</p> <p><b>Modulo 4 – Tecniche di pre-elaborazione (impegno di 29,5 ore)</b> dove sono affrontati i seguenti argomenti: Tecniche di pre-elaborazione dei Big Data - Filtraggio, Trasformazione, Integrazione. Esercitazione sulle tecniche di pre-elaborazione</p> <p><b>Etivity 2 (15 ore di carico di studio)</b> – Progettazione di sistemi per l’elaborazione parallela di big data.</p> <p><b>Modulo 5 – Riduzione dimensionale (impegno di 15,5 ore)</b> dove sono affrontati i seguenti argomenti: Riduzione del numero di dimensioni dei Big Data: Principal-Component Analysis, Singular-Value Decomposition. Esercitazione su PCA ed SVD.</p> <p><b>Modulo 6 – Clustering (impegno di 15,5 ore)</b> dove sono affrontati i seguenti argomenti: Clustering dei Big Data: Partizionamento in Cluster; Fuzzy Clustering; Clustering Relazionale. Esercitazione sul clustering.</p> <p><b>Etivity 3 (15 ore di carico di studio)</b> – Progettazione di sistemi per l’applicazione di algoritmi avanzati di analisi dei big data.</p> <p><b>Modulo 7 – Classificazione (impegno di 15,5 ore)</b> dove sono affrontati i seguenti argomenti: Algoritmi di classificazione dei Big Data: Criteri di classificazione; Classificatori Bayesiani; Support Vector Machines.</p> <p><b>Modulo 8 – Predizione (impegno di 15,5 ore)</b> dove sono affrontati i seguenti argomenti: Algoritmi di Predizione basati sui Big Data: Macchine a stati finiti; Modelli probabilistici; Modelli Ricorrenti.</p> <p><b>Esercitazioni su compiti d’esame (impegno di 5 ore)</b></p> <p><b>Etivity 4 (6 ore di carico di studio)</b> – Simulazione esame.</p>
<b>Materiali di studio</b>	<p>· MATERIALI DIDATTICI A CURA DEL DOCENTE</p> <p>Il materiale didattico presente in piattaforma è suddiviso in 8 moduli. Essi ricoprono interamente il programma e ciascuno di essi contiene dispense, slide e videolezioni in cui il docente commenta le slide. Tale materiale contiene tutti gli elementi necessari per affrontare lo studio della materia.</p>
<b>Modalità di verifica dell’apprendimento</b>	<p>L’esame consiste nello svolgimento di una <b>prova scritta</b> tendente ad accertare le capacità di analisi e rielaborazione dei concetti acquisiti e di una <b>serie di attività (e-tivity)</b> svolte durante il Corso nelle <b>classi virtuali</b>.</p>

	I risultati di apprendimento attesi circa le conoscenze della materia e la capacità di applicarle sono valutate dalla prova scritta, mentre le abilità comunicative, la capacità di trarre conclusioni e la capacità di autoapprendimento sono valutate in itinere attraverso le e-tivity.
<b>Criteri per l'assegnazione dell'elaborato finale</b>	L'assegnazione dell' <b>elaborato finale</b> avverrà sulla base di un colloquio con il docente in cui lo studente manifesterà i propri specifici <b>interessi</b> in relazione a qualche argomento che intende approfondire; non esistono <b>preclusioni</b> alla richiesta di assegnazione della tesi e non è prevista una <b>media particolare</b> per poterla richiedere.