



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Ingegneria
ANNO ACCADEMICO OFFERTA	2019/2020
ANNO ACCADEMICO EROGAZIONE	2021/2022
CORSO DILAUREA	INGEGNERIA DELL'INNOVAZIONE PER LE IMPRESE DIGITALI
INSEGNAMENTO	TECNOLOGIE PER I BIG DATA
TIPO DI ATTIVITA'	C
AMBITO	10655-Attività formative affini o integrative
CODICE INSEGNAMENTO	20412
SETTORI SCIENTIFICO-DISCIPLINARI	INF/01
DOCENTE RESPONSABILE	ANDOLINA SALVATORE Ricercatore a tempo determinato Univ. di PALERMO
ALTRI DOCENTI	
CFU	6
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	103
NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA	47
PROPEDEUTICITA'	
MUTUAZIONI	
ANNO DI CORSO	3
PERIODO DELLE LEZIONI	2° semestre
MODALITA' DI FREQUENZA	Facoltativa
TIPO DI VALUTAZIONE	Voto in trentesimi
ORARIO DI RICEVIMENTO DEGLI STUDENTI	ANDOLINA SALVATORE Martedì 15:00 16:00 MS Teams su appuntamento

DOCENTE: Prof. SALVATORE ANDOLINA

PREREQUISITI	Calcolatori elettronici e Programmazione.
RISULTATI DI APPRENDIMENTO ATTESI	<p>Conoscenza e capacita' di comprensione Lo studente al termine del Corso avra' conoscenza dei modelli di calcolo, delle architetture e delle infrastrutture necessarie all'elaborazione di grandi moli di dati. Avra' inoltre familiarita' con le problematiche legate all'analisi di grandi moli di dati.</p> <p>Capacita' di applicare conoscenza e comprensione Lo studente sara' in grado di analizzare scenari caratterizzati dalla presenza di grandi moli di data e sara' in grado di fornire opportune soluzioni progettuali per la realizzazione di sistemi in grado di gestire tali dati. Sara' in grado di progettare sistemi software efficienti per l'elaborazione di grandi moli di dati.</p> <p>Autonomia di giudizio Lo studente sara' in grado di valutare la bonta' delle diverse soluzioni algoritmiche e architetture per la gestione di grandi moli di dati e sara' inoltre in grado di valutare le performance dei diversi approcci interpretando opportuni indicatori.</p> <p>Abilita' comunicative Lo studente acquisira' la capacita' di comunicare ed esprimere problematiche inerenti all'oggetto del corso. Sara' in grado di sostenere conversazioni su tematiche relative alla progettazione e gestione di sistemi per la gestione di grandi moli di dati.</p> <p>Capacita' d'apprendimento Lo studente avra' appreso come analizzare differenti applicazioni che richiedono l'uso di Big Data e avra' acquisito gli strumenti necessari ad apprendere autonomamente i principi di funzionamento di nuovi tool per la gestione dei big data.</p>
VALUTAZIONE DELL'APPRENDIMENTO	<p>Le conoscenze e le competenze acquisite dallo studente saranno verificate attraverso una prova scritta (prova in itinere + prova finale o prova complessiva) e un colloquio orale.</p> <p>Valutazione della prova scritta Durante il corso, in accordo con il calendario accademico, sara' possibile sostenere una prova in itinere. Tale prova, a discrezione dell'allievo potra' essere completata con una prova finale da sostenere nel periodo compreso tra la fine delle lezioni ed il primo appello del corso. La media pesata della prova in itinere e di quella finale costituisce il voto della prova scritta. La prova scritta (in itinere + finale o complessiva) e' costituita da almeno tre esercizi volti a verificare le conoscenze dello studente degli argomenti affrontati durante il corso, e di applicare le capacita' e le conoscenze acquisite. Nello svolgimento assume fondamentale importanza il commento teorico dei risultati ottenuti. L'articolazione della soluzione consente di apprezzare tutti i livelli di preparazione. La valutazione e' espressa in trentesimi ed un minimo di 15 e' richiesto per accedere alla prova orale.</p> <p>Valutazione per la prova orale La prova orale consiste in un colloquio, volto ad accertare il possesso delle competenze e delle conoscenze disciplinari previste dal corso; la valutazione viene espressa in trentesimi. Durante il colloquio orale lo studente dovra' essere in grado di discutere le soluzioni proposte durante la prova scritta; inoltre saranno proposte domande di diverso e crescente livello di complessita' al fine di valutare il raggiungimento degli obiettivi formativi e le abilita' comunicative dello studente. Infine, allo scopo di valutare l'autonomia di giudizio, sara' richiesto di analizzare le caratteristiche di specifici scenari applicativi e di proporre le soluzioni piu' adeguate ai problemi individuati. La valutazione finale terra' conto sia del punteggio della prova scritta (50%) che di quello delle prova orale (50%). Eccellente 30-30 e lode. Durante entrambe le prove lo studente dovra' dimostrare padronanza completa degli argomenti del corso. Durante il colloquio orale l'allievo dovra' dimostrare la maturita' di saper collegare i diversi aspetti trattati e la capacita' di saper generalizzare. Dovra' mostrare autonomia nella soluzione dei quesiti e la capacita' di individuare le informazioni necessarie per la soluzione degli stessi. Molto buono 27-29 Buona padronanza degli argomenti, lo studente e' in grado di applicare le conoscenze per risolvere i problemi proposti. Buono 24-26 buona conoscenza dei principali, discreta padronanza e proprieta' di linguaggio, con capacita' di applicare autonomamente le conoscenze alla soluzione dei problemi proposti. Discreto 21-23 Piu' che sufficiente padronanza degli argomenti principali dell'insegnamento, limitata capacita' di applicare autonomamente le conoscenze</p>

	acquisite. Sufficiente 18-20 conoscenza di base degli argomenti principali dell'insegnamento e del linguaggio tecnico. Insufficiente non possiede una conoscenza accettabile dei contenuti degli argomenti trattati nell'insegnamento.
OBIETTIVI FORMATIVI	L'obiettivo del corso e' di formare lo studente con riferimento alla conoscenza dei principi basilari dei sistemi di gestione di grandi moli di dati. In particolare sara' in grado di valutare diversi approcci architetturali e algoritmici allo stato dell'arte e di individuare le soluzioni piu' adatte agli scenari sotto esame.
ORGANIZZAZIONE DELLA DIDATTICA	Lezioni, Esercitazioni.
TESTI CONSIGLIATI	Dispense fornite dal docente.

PROGRAMMA

ORE	Lezioni
2	Introduzione ai concetti di base dei Big Data: terminologia, aspetti principali ed esempi di applicazioni
4	Problematiche principali legate alla gestione dei Big Data: volume dei dati e occupazione di memoria, velocita' dell'elaborazione e complessita' computazionale, presenza di errori e accuratezza dei dati, comprensione dei dati.
4	Infrastrutture per la gestione dei Big Data: Architetture Distribuite e Parallele; Cloud Computing per i Big Data.
4	Memorizzazione dei Big Data: Memorizzazione strutturata; Database non relazionali; Tipologie di database NoSql;
3	Tecniche di pre-elaborazione dei Big Data: Tipi di errori; Gestione degli errori.
6	Tecniche di pre-elaborazione dei Big Data: Filtraggio, Trasformazione, Integrazione. Paradigma MapReduce. Apache Hadoop.
3	Design pattern per i Big Data con il paradigma MapReduce: pattern di aggregazione, inverted index, filtraggio, top-k, eliminazione dei duplicati, partizionamento, join reduce-side e map-side, ordinamento totale dei dati.
3	Tecniche di scheduling per i Big Data. Tecniche di elaborazione dei Big Data ad alto livello: Apache Spark.
3	Tecniche di elaborazione di dataset distribuiti. Dataset distribuiti (RDD) con Apache Spark: trasformazioni, azioni e persistenza dei dati.
3	SQL con i Big Data: Apache Spark SQL. Usi avanzati, funzioni a finestra, User Defined Functions (UDF).
ORE	Esercitazioni
6	Progettazione di sistemi per la gestione distribuita di big data
6	Progettazione di sistemi per l'elaborazione parallela di big data