



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Matematica e Informatica
ANNO ACCADEMICO OFFERTA	2018/2019
ANNO ACCADEMICO EROGAZIONE	2018/2019
CORSO DILAUREA MAGISTRALE	INFORMATICA
INSEGNAMENTO	ELABORAZIONE DATI
CODICE INSEGNAMENTO	19742
MODULI	Si
NUMERO DI MODULI	2
SETTORI SCIENTIFICO-DISCIPLINARI	INF/01
DOCENTE RESPONSABILE	GIANCARLO RAFFAELE Professore Ordinario Univ. di PALERMO
ALTRI DOCENTI	GIANCARLO RAFFAELE Professore Ordinario Univ. di PALERMO ROMBO SIMONA ESTER Professore Ordinario Univ. di PALERMO
CFU	12
PROPEDEUTICITA'	
MUTUAZIONI	
ANNO DI CORSO	1
PERIODO DELLE LEZIONI	2° semestre
MODALITA' DI FREQUENZA	Facoltativa
TIPO DI VALUTAZIONE	Voto in trentesimi
ORARIO DI RICEVIMENTO DEGLI STUDENTI	GIANCARLO RAFFAELE Lunedì 15:00 17:00 Stanza 106 Dipartimento di Matematica ed Informatica Giovedì 15:00 17:00 Stanza 106 Dipartimento di Matematica ed Informatica ROMBO SIMONA ESTER Lunedì 9:30 13:30 In presenza: Via Archirafi 34, Piano II, Stanza 220 - Telematico: via Microsoft Teams o altro canale - In entrambi i casi, e' consigliabile prenotarsi tramite email alla docente

DOCENTE: Prof. RAFFAELE GIANCARLO

PREREQUISITI	Conoscenze di base su: algebra e geometria, algoritmi e strutture dati, progettazione di basi di dati. Conoscenze avanzate di programmazione.
RISULTATI DI APPRENDIMENTO ATTESI	<p>Conoscenza e capacita' di comprensione Acquisizione di conoscenze sull'organizzazione dei sistemi informativi aziendali, sull'utilizzo e la progettazione dei sistemi di supporto alle decisioni, sulle problematiche correlate al big data management. Acquisizione degli strumenti avanzati per l'analisi ed il progetto di algoritmi. Capacita' di utilizzare il linguaggio specifico proprio di queste discipline specialistiche.</p> <p>Capacita' di applicare conoscenza e comprensione Capacita' di analizzare problematiche relative a contesti caratterizzati da grosse moli di dati e proporre soluzioni progettuali per la gestione e analisi dei dati in tali contesti ed il supporto alle decisioni. Capacita' di sviluppare software basato su algoritmi efficienti per grosse quantita' di dati.</p> <p>Autonomia di giudizio Capacita' di analizzare e valutare soluzioni per la gestione di grosse quantita' di dati. Capacita' di progettare sistemi di supporto alle decisioni mediante l'analisi delle specifiche tecniche fornite. Essere in grado di valutare le implicazioni e i risultati degli studi algoritmici presentati durante il corso e della complessita' computazionale dei problemi ad essi associati.</p> <p>Abilita' comunicative Capacita' di descrivere soluzioni di progettazione di sistemi informativi complessi e sistemi di supporto alle decisioni e analizzarne le prestazioni. Capacita' di cooperare per determinare delle soluzioni di progettazione appropriate in contesti applicativi caratterizzati dalla presenza di grosse quantita' di dati. Capacita' di esporre i risultati degli studi algoritmici, anche ad un pubblico non esperto. Essere in grado di sostenere l'importanza ed evidenziare le ricadute tecnologiche delle teorie studiate.</p> <p>Capacita' d'apprendimento Capacita' di aggiornamento attraverso la consultazione di testi avanzati e di pubblicazioni scientifiche correlate agli argomenti affrontati durante il corso. Capacita' di seguire, utilizzando le conoscenze acquisite, sia master di secondo livello, che corsi di approfondimento e seminari specialistici nell'ambito della Business Intelligence e del progetto di algoritmi e strutture dati.</p>
VALUTAZIONE DELL'APPRENDIMENTO	<p>Prova progettuale, prova orale.</p> <p>Prova progettuale: consistera' nella realizzazione e implementazione di un progetto basato su tecnologie di Big Data. E' obbligatorio che il progetto sia realizzato da un team costituito da un minimo di due ad un massimo di cinque studenti, al fine di favorire lo sviluppo della capacita' di lavorare in team. Verranno effettuate due revisioni di progetto, durante le quali la docente verifichera' il contributo individuale di ciascun componente del team e assegnera' una valutazione alle diverse fasi di progettazione. Alla fine al progetto verra' attribuito un punteggio complessivo, pesato per ciascuno studente in funzione del contributo individuale. Contribuiranno ad una valutazione positiva l'originalita' della soluzione proposta, il rigore metodologico, la dimostrazione di avere acquisito appieno gli strumenti tecnici forniti dal corso. I requisiti per conseguire il punteggio minimo per accedere alla seconda parte dell'esame consistono nella capacita' di realizzare un progetto che, seppure in modo basilare, rispetti le specifiche assegnate e sia corretto.</p> <p>Prova orale: servira' per verificare le conoscenze acquisite durante il corso, l'autonomia nell'approfondire contenuti anche complessi e la capacita' individuale di trovare soluzioni a problemi proposti. La prova orale avra' inizio, per ciascuno dei due moduli, con la presentazione di un argomento a scelta tra quelli trattati durante il corso. In particolare, per il modulo di Big Data Management, lo studente produrra' anche una breve tesina. Quindi, verra' verificata la capacita' critica e l'autonomia di giudizio dello studente attraverso un numero adeguato di domande. La valutazione della prova orale andra' a integrare il punteggio conseguito da ciascuno studente a seguito della valutazione del progetto.</p> <p>Si fara' la media aritmetica tra il voto conseguito per il progetto e quello conseguito per la prova orale. In particolare: 18-21: Conoscenza sufficiente di tutti gli argomenti del programma. 22-24: Conoscenza discreta di tutti gli argomenti del programma. 25-27: Conoscenza buona di tutti gli argomenti del programma. 28-30: Conoscenza ottima di tutti gli argomenti del programma.</p>
ORGANIZZAZIONE DELLA DIDATTICA	Lezioni frontali

**MODULO
SCIENZA E INGEGNERIA DEGLI ALGORITMI**

Prof. RAFFAELE GIANCARLO

TESTI CONSIGLIATI

William J.Cook, William H. Cunningham, William R. Pulleyblank, Alexander Schrijver. Combinatorial Optimization, Wiley 1997

Per le parti che riguardano algoritmi di approssimazione- For the part regarding approximation algorithms Robert Endre Tarjan. Data Structure and Network Algorithms, SIAM 1984

Per le parti che riguardano strutture dati in memoria interna-for the part regarding data structures in internal memory Camil Demetrescu, Irene Finocchi, Giuseppe F. Italiano, Algoritmi e Strutture dati, McGraw Hill, 2005

Per le parti che riguardano l'analisi ammortizzata di algoritmi. For the part regarding amortised analysis of algorithms. H. Cormen. C. Leiserson, R. Rivest, C. Stein Introduzione agli algoritmi e strutture dati, McGraw Hill, 2001

Per le parti che riguardano hashing e schemi di approssimazione-for the parts regarding hashing and approximation schemes

Materiale distribuito dal docente.

Tutto il resto del programma The remaining part of the syllabus.

TIPO DI ATTIVITA'	B
AMBITO	50341-Discipline Informatiche
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	102
NUMERO DI ORE RISERVATE ALLE ATTIVITA' DIDATTICHE ASSISTITE	48

OBIETTIVI FORMATIVI DEL MODULO

Esporre lo studente a tecniche avanzate di progetto ed analisi di algoritmi. In particolare, si copre tutto lo spettro delle strutture dati dinamiche e degli algoritmi, con approfondito studio di complessita' computazionale di problemi intrattabili o che coinvolgono grosse quantita' di dati e di loro soluzioni approssimate.

PROGRAMMA

ORE	Lezioni
4	Analisi ammortizzata: Metodo dei crediti. Metodo del potenziale. Analisi Sperimentale di Algoritmi
6	STRUTTURE DATI AVANZATE IN MEMORIA INTERNA: Bilanciamento Red Black Trees ed analisi delle operazioni su di essi. Linking and Cutting Trees ed analisi delle operazioni su di essi. Self-organizing List.
6	Self-adjusting binary trees ed analisi delle operazioni su di essi. Self-organizing Data Structure. Self-organizing List.
6	Dizionari Succinti in Spazio: Universal e Perfect Hashing.
5	Dizionari Succinti in Spazio: Bloom Filters.
6	MODELLO STREAMING. Motivazione. Il modello di calcolo streaming. Esempi di algoritmi streaming per mining di grosse quantita' di dati.
6	ALGORITMI STREAMING. Frequency Moments. Sketches.
4	ANALISI E SCHEMI DI COMPRESSIONE DATI Schemi di compressione, schemi di compressione adattivi. Ingegneria di compression boosting. Strutture dati efficienti per compressione dati. Benchmark per analisi di compressione.
3	TEORIA DEI PROBLEMI NP COMPLETI E APPROSSIMAZIONI POLINOMIALI Schemi di Approssimazione Polinomiale Inapprossimabilita' di Problemi
2	Metodi approssimati, Metodi euristici. TSP con disuguaglianza triangolare;

**MODULO
BIG DATA MANAGEMENT**

Prof.ssa SIMONA ESTER ROMBO

TESTI CONSIGLIATI

DATA WAREHOUSE: M. Golfarelli, S. Rizzi, "Data Warehouse – Teoria e pratica della progettazione", Seconda Edizione, McGraw Hill, 2005.

DATA MINING: J. Han, M. Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, 2006.

TIPO DI ATTIVITA'	B
AMBITO	50341-Discipline Informatiche
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	102
NUMERO DI ORE RISERVATE ALLE ATTIVITA' DIDATTICHE ASSISTITE	48

OBIETTIVI FORMATIVI DEL MODULO

Obiettivo del corso di Big Data Management e' fornire agli studenti di Informatica conoscenze sull'organizzazione dei sistemi informativi aziendali, sull'utilizzo e la progettazione dei sistemi di supporto alle decisioni, sulle problematiche correlate alla gestione di grosse quantita' di dati. Il corso iniziera' con una panoramica su Business Intelligence e si concentrera' su progettazione di Data Warehouse e OLAP. Il corso affrontera' anche la progettazione di database non relazionali e l'utilizzo di tecnologie quali Apache Hadoop e Spark. Infine, si mostreranno degli scenari applicativi reali quali ad esempio reti sociali, smart cities e dati biologici.

PROGRAMMA

ORE	Lezioni
2	INTRODUZIONE AI BIG DATA Esigenze. Contesti applicativi. Una panoramica sulle tecnologie.
4	BUSINESS INTELLIGENCE E DATA WAREHOUSING Acquisizione dei dati. Elaborazione dei dati al fine di produrre informazioni. Archiviazione dei dati grezzi e delle informazioni. Trasmissione dei dati e delle informazioni. Presentazione dei dati e delle informazioni. Sistemi di supporto alle decisioni. Architetture per il Data Warehousing: architettura ad un livello, architettura a due livelli, architettura a tre livelli. I metadati. Qualita' di un Data Warehouse. Il modello multidimensionale e l'OLAP. Le principali operazioni OLAP. Modelli logici a supporto del Data Warehousing.
4	PROGETTAZIONE DI UN DATA WAREHOUSE Metodologia di progettazione di un Data Warehouse, scelta del processo, scelta della granularita, identificare e rendere conformi le dimensioni, scelta delle misure, memorizzare pre-calcoli nella tabella dei fatti, completare la tabella delle dimensioni, scelta della durata del database, tracciare le "slowly changing dimension".
6	PROGETTAZIONE DI DATA WAREHOUSE IN PENTAHO Studio dei moduli: Kettle, Data Integration, OLAP.
4	ANALISI E VISUALIZZAZIONE Applicazioni e strumenti specifici per l'analisi di big data e la visualizzazione dei risultati di complesse query di tipo decisionale.
6	BIG DATA MINING Richiami di Data Mining. Market Basket Analysis. Classificazione e Predizione. Alberi di decisione. K-Nearest Neighbor. Clustering. Outlier detection. Problematiche di data mining su grandi moli di dati, sfide e soluzioni.
4	MEMORIZZAZIONE STRUTTURATA Database non relazionali. Tipologie di database NoSQL. Analisi di vantaggi e svantaggi dei database non relazionali. Esempi di implementazione di database NoSQL ed esercizi.
6	TECNOLOGIE PER BIG DATA (I) Studio dell'algoritmo Map-Reduce. Studio di strumenti come Apache Hadoop e Spark. Applicazioni ed esercizi usando gli strumenti acquisiti.
6	TECNOLOGIE PER BIG DATA (II) Studio e applicazione di specifiche librerie di Spark.
6	ANALISI DI ALCUNI CONTESTI REALI DI BIG DATA Reti sociali. Reti biologiche. Next-Generation Sequencing (NGS). Smart Cities.