UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Technological Infrastructures For Data Science

2122-2-F9101Q052

Obiettivi

Il corso ha lo scopo di fornire una solida conoscenza delle piattaforme tecnologiche (Cloud - Containerizzazione) che consentono la raccolta e la gestione di dati, nonché delle piattaforme elaborative (architetture, algoritmi e infrastrutture) utilizzate per analizzarli. Le esercitazioni forniranno allo studente le competenze di base necessarie per interagire con tali piattaforme.

Contenuti sintetici

Piattaforme virtualizzate per la raccolta e gestione di dati di grandi dimensioni e velocità . Architetture per l'elaborazione dei dati, infrastrutture elaborative, piattaforme Big Data per la Data Science, esempi di piattaforme.

Programma esteso

- Architetture di elaborazione dati
- Gestione delle infrastrutture
- Virtualizzazione e containerizzazione
- Ambienti cloud, modelli e costi
- Piattaforme Big Data per la data Science
- Tipi di workload
- MapReduce: l'ecosistema Hadoop/Spark
- Elaborazione di eventi e stream
- Esempi di piattaforma
    - Apache Spark
    - Apache Kafka

        - Google Cloud Platform
    - DevOps - DataOps - MLOps

Prerequisiti

Conoscenza di base di:

- un linguaggio di programmazione (es. Python)
- dell'architettura di un calcolatore.

Modalità didattica

Lezioni ed esercitazioni in aula. Il corso verrà erogato in lingua inglese

Materiale didattico

Dispense e slide del corso fornite dai docenti.

Periodo di erogazione dell'insegnamento

Secondo anno, primo semestre

Modalità di verifica del profitto e valutazione

L'esame sarà costituito da due parti. Le due parti che dovranno essere **svolte nel medesimo appello**.

La prima parte consisterà in un insieme di domande chiuse ed aperte da svolgersi in un'ora (orientativamente si avranno 9 domande chiuse ed 4 aperte, tuttavia si potranno avere piccole variazioni nella struttura dell'esame). La prima parte dell'esame si svolgerà in forma cartacea ed il peso di ciascuna domanda per la formazione del voto verrà esplicitamente indicato nella parte alta del foglio contenente le domande. Le risposte alle domande verranno scritte sul foglio stesso, ma è facoltà dello studente utilizzare ulteriori fogli per fornire una risposta più estesa alle domande aperte.

La seconda parte sarà costituta da un breve progetto da realizzare in gruppi di 2 persone e dalla sua esposizione orale.

Esempi di tipi di progetto:

- Analisi e test di una particolare piattaforma tecnologica (vantaggi, svantaggi, costi, curva di apprendimento, realizzazione di un'applicazione che utilizzi tale tecnologia)
- Progettazione di un'applicazione per il cloud (scelta del provider, del tipo di macchine virtuali, dei servizi, stima dei costi, stima della qualità del servizio, stima delle dimensioni dei dati supportati, attività correttive necessarie in caso di imprevisti)

Una volta che lo studente avrà svolto entrambe le prove, l'esame si considererà superato se si verificheranno entrambe queste condizioni:

1. Per entrambe le parti lo studente avrà ottenuto più della metà dei 15 punti a disposizione.
2. La somma dei punti delle due parti sarà maggiore o uguale a 18.

In tal caso lo studente potrà registrare un voto costituito dalla somme dei punti.

Orario di ricevimento

Martedì 12:30-14:30, chiedere conferma per email