



# UNIVERSITÀ DI PISA

## DISTRIBUTED ENABLING PLATFORMS

PATRIZIO DAZZI

Academic year	2023/24
Course	INFORMATICA E NETWORKING
Code	534AA
Credits	6

Modules	Area	Type	Hours	Teacher(s)
PIATTAFORME ABILITANTIINF/01 DISTRIBUITE		LEZIONI	48	PATRIZIO DAZZI

### Obiettivi di apprendimento

#### Conoscenze

Al termine del corso, gli studenti avranno acquisito le seguenti conoscenze relative ai Distributed Enabling Platforms:  
Conoscenze di base:

- Fondamenti dei sistemi distribuiti, compresi algoritmi distribuiti, protocolli di consenso e sincronizzazione distribuita.
- Principi di architettura e progettazione di sistemi distribuiti.
- Concetti di scalabilità, affidabilità e tolleranza ai guasti in sistemi distribuiti.

Conoscenze caratterizzanti:

- Piattaforme di storage e processing distribuito (es. Apache Hadoop, Apache Cassandra e Apache Kafka).
- Framework di elaborazione dati in streaming (es. Apache Flink e Apache Beam).
- Modelli di partizionamento, replicazione e coerenza dei dati nei sistemi di storage distribuito.
- Tecniche per l'elaborazione distribuita di grandi volumi di dati in tempo reale e batch.

Conoscenze affini/trasversali:

- Principi di fault tolerance e strategie per la gestione degli errori in sistemi distribuiti.
- Concetti di scalabilità orizzontale e verticalità in ambienti distribuiti.
- Considerazioni sulle prestazioni nella progettazione e implementazione di applicazioni distribuite.
- Metodi di monitoraggio, gestione delle prestazioni e ottimizzazione delle risorse in sistemi distribuiti.

#### Modalità di verifica delle conoscenze

Per accertare l'acquisizione delle conoscenze obiettivo indicate, saranno adottate le seguenti modalità di verifica:

- Valutazione delle prove (facoltative) in itinere: Saranno svolte prove durante il corso per valutare la comprensione degli argomenti trattati. Queste prove potrebbero includere test a risposta multipla o domande aperte che mettono alla prova la conoscenza teorica e l'applicazione pratica delle Distributed Enabling Platforms.
- Valutazione dei progetti: Gli studenti svilupperanno un progetto che richiede l'applicazione delle conoscenze acquisite durante il corso. I progetti verranno valutati sulla base della corretta implementazione, la scelta appropriata delle tecnologie e l'efficacia nel raggiungimento degli obiettivi prefissati.

#### Capacità

Al termine del corso, gli studenti svilupperanno le seguenti capacità/abilità in relazione alle Distributed Enabling Platforms:

- Capacità di progettazione: Gli studenti saranno in grado di progettare soluzioni distribuite utilizzando strumenti appropriati per soddisfare i requisiti specifici. Saranno in grado di identificare le componenti necessarie, definire l'architettura del sistema distribuito e prendere decisioni informate sulla scelta delle tecnologie e dei modelli di progettazione.
- Capacità di implementazione: Gli studenti saranno in grado di implementare sistemi distribuiti utilizzando gli strumenti trattate nel corso. Saranno in grado di utilizzare gli strumenti e le librerie corrispondenti per realizzare le funzionalità richieste, gestire la distribuzione delle risorse e garantire la corretta comunicazione tra i componenti del sistema.
- Capacità di gestione e monitoraggio: Gli studenti acquisiranno la capacità di gestire e monitorare sistemi distribuiti. Saranno in grado di configurare e ottimizzare le risorse di calcolo, monitorare le prestazioni, rilevare e gestire eventuali problemi di errore o degrado delle prestazioni.
- Capacità di valutazione e ottimizzazione delle prestazioni: Gli studenti saranno in grado di valutare le prestazioni dei sistemi distribuiti. Saranno in grado di analizzare i risultati delle misurazioni, identificare eventuali collo di bottiglia e proporre soluzioni per ottimizzare le prestazioni del sistema.



# UNIVERSITÀ DI PISA

---

## Modalità di verifica delle capacità

Lo studente dovrà preparare e presentare una relazione scritta che riporti i risultati dell'attività di progetto, che inquadri quanto realizzato nelle tematiche del corso, sottolineando le problematiche affrontata, i vantaggi e i limiti della soluzione proposta.

## Comportamenti

Al termine del corso lo studente acquisirà la sensibilità necessaria ad identificare le piattaforme abilitanti distribuite più opportune per diverse classi di applicazioni. Lo studente sarà inoltre in grado di approcciare il problema dal punto di vista più opportuno e che gli garantisca di ottimizzare le prestazioni del sistema utilizzato.

## Modalità di verifica dei comportamenti

La congruità della metodologia di indagine per la selezione degli strumenti più opportuni, e per l'inquadramento e definizione del problema saranno determinati dalla valutazione del progetto e della relazione ad esso allegata.

## Prerequisiti (conoscenze iniziali)

Conoscenze e abilità/capacità utili per seguire il corso in modo proficuo.

Conoscenze:

- Conoscenza di base dei concetti fondamentali dei sistemi di elaborazione, compresi i concetti di architettura dei computer, reti di computer, sistemi operativi e basi di dati. Familiarità con almeno un linguaggio di programmazione, possibilmente Java o Scala.
- Conoscenza dei concetti di programmazione, strutture dati e algoritmi.
- Comprendere i concetti di base dei sistemi concorrenti, come la comunicazione e la sincronizzazione tra processi, la gestione delle risorse condivise.
- Comprensione dei principi di base delle reti di computer, inclusi i protocolli di comunicazione, l'architettura client-server e la gestione dei pacchetti.

Abilità/Capacità:

- Capacità di problem-solving: Essere in grado di analizzare i problemi, identificare gli obiettivi e sviluppare soluzioni efficaci.
- Capacità di pensiero critico e di applicazione dei concetti teorici a situazioni pratiche.
- Pensiero analitico: Essere in grado di analizzare e valutare criticamente le informazioni, identificare i collegamenti tra i concetti e cogliere le implicazioni pratiche delle conoscenze acquisite.
- Curiosità e interesse per la tecnologia: Mostrare un interesse attivo per le Distributed Enabling Platforms, mantenendo la curiosità e l'apertura mentale verso le nuove tendenze e gli sviluppi nel campo dei sistemi distribuiti.

## Corequisiti

N/A

## Prerequisiti per studi successivi

N/A

## Indicazioni metodologiche

Il corso si articola nel seguente modo:

- lezioni frontali, con ausilio di slides;
- i materiali didattici, comunicazioni docente-studenti, pubblicazione di test per esercitazioni a casa, formazione di gruppi di lavoro, sono indicati sulla pagina classroom del corso;
- è prevista la presenza di prove intermedie facoltative

## Programma (contenuti dell'insegnamento)

Durante l'insegnamento del corso su Distributed Enabling Platforms, gli argomenti e le attività trattate saranno:

1. Introduzione ai sistemi distribuiti
  - Concetti fondamentali dei sistemi distribuiti
  - Architetture e modelli di distribuzione
  - Scalabilità e affidabilità nei sistemi distribuiti
2. Piattaforme per l'abilitazione distribuita: Panoramica
  - Concetti e ruoli delle piattaforme per l'abilitazione distribuita
  - Esempi di popolari piattaforme per l'elaborazione di stream di dati in tempo reale (es. Apache Kafka) e loro elementi caratterizzanti.



# UNIVERSITÀ DI PISA

## 3. Piattaforme per l'archiviazione distribuita dei dati

- Piattaforme per l'archiviazione distribuita dei dati (es. Apache Hadoop, Apache Cassandra) e le loro caratteristiche distinte.
- Partizionamento dei dati, replica e modelli di coerenza nei sistemi di archiviazione distribuita.

## 4. Piattaforme per l'elaborazione distribuita dei dati

- Piattaforme per l'elaborazione distribuita dei dati (es. Apache Flink, Apache Beam) e le loro capacità uniche.
- Principi di progettazione e casi d'uso di queste piattaforme per l'analisi in tempo reale e l'elaborazione di grandi volumi di dati.

## 5. Tolleranza ai guasti e resilienza

- Tecniche per la gestione dei guasti nei sistemi distribuiti
- Strategie per garantire la tolleranza ai guasti, l'alta disponibilità e il ripristino dei guasti nelle piattaforme per l'abilitazione distribuita.

## 6. Ottimizzazione delle prestazioni

- Strategie per ottimizzare le prestazioni delle piattaforme per l'abilitazione distribuita
- Tecniche per il monitoraggio, il profiling e l'ottimizzazione dei sistemi distribuiti per ottenere prestazioni ottimali.

## 7. Applicazioni pratiche e studi di caso

- Applicazione delle piattaforme per l'abilitazione distribuita in scenari reali
- Analisi di studi di caso e esempi pratici che illustrano l'utilizzo delle piattaforme per l'abilitazione distribuita.

## 8. Progetti pratici e laboratori

- Esercizi pratici e progetti per applicare le conoscenze delle piattaforme per l'abilitazione distribuita in scenari reali.
- Esperienza pratica nella distribuzione, gestione e ottimizzazione di applicazioni distribuite utilizzando le piattaforme selezionate.

## Bibliografia e materiale didattico

- Marteen van Steen, Andrew S. Tanenbaum, Distributed Systems 3rd edition
- Jonas Bone?r, Reactive Microsystems
- Alex Holmes, Hadoop in Practice
- Ben Stopford, Designing Event-Driven Systems
- <https://spark.apache.org/docs/latest/index.html>

## Indicazioni per non frequentanti

Le modalità d'esame per i non frequentanti sono le medesime. Si richiede di contattare il docente con un ragionevole anticipo, che consenta di affrontare la realizzazione del progetto in modo adeguato.

## Modalità d'esame

L'esame è composto da un progetto, concordato con il docente, ed una successiva prova orale, divisa in due parti: una che si focalizzerà sul progetto realizzato, l'altra su tematiche generali affrontate a lezione;

## Stage e tirocini

N/A

## Altri riferimenti web

- <https://hadoop.apache.org/>
- <https://spark.apache.org/>
- <https://kafka.apache.org/>
- <https://flink.apache.org/>
- <https://cassandra.apache.org/>