

Testi del Syllabus

Resp. Did. **CANNATARO MARIO** **Matricola: 000037**

Docenti **CANNATARO MARIO, 5 CFU**
ZUCCO CHIARA, 4 CFU

Anno offerta: **2026/2027**
Insegnamento: **A004704 - CALCOLO PARALLELO E DISTRIBUITO PER L'ELABORAZIONE DI DATI BIOMEDICI**
Corso di studio: **7166R - INGEGNERIA BIOMEDICA**
Anno regolamento: **2026**
CFU: **9**
Settore: **IINF-05/A**
Tipo Attività: **C - Affine/Integrativa**
Anno corso: **1**
Periodo: **Primo Ciclo Semestrale**



Testi in italiano

Lingua insegnamento ITALIANO

Contenuti

Il corso presenta le principali problematiche, metodologie, infrastrutture di calcolo, ambienti di programmazione, algoritmi e strumenti software per la memorizzazione, gestione ed analisi efficiente dei principali dati biomedici, con particolare riferimento ai dati sperimentali generati in biologia (es. dati omici quali dati genomici, proteomici, interattomici) ed ai dati utilizzati in medicina (es. biosegnali, bioimmagini, dati clinici e sanitari). Inoltre, il corso fornisce le basi per progettare e sviluppare applicazioni parallele, distribuite, web-based o cloud-based, per la gestione ed analisi di grossi quantitativi di dati biomedici. Infine, il corso presenta le principali tecniche per l'estrazione di conoscenza dai dati con particolare riferimento alle tecniche di preprocessing e agli algoritmi di data mining.

Testi di riferimento

PARTE I

- Stefano Ceri, Piero Fraternali, Aldo Bongio, Marco Brambilla, Sara Comai, Maristella Matera, Progettazione di dati e applicazioni per il web, McGraw Hill, 2003, ISBN: 8838661389.
http://www.ateneonline.it/LibroAteneo.asp?item_id=1588
- WebML: www.webml.org
- WebRatio: www.webratio.com
- Core Web Programming: <http://www.corewebprogramming.com/>

PARTE II

- Joel Grus, Data Science con Python: Dai Fondamenti al Machine Learning, Egea
- Sebastian Raschka, Machine Learning con Python: costruire algoritmi per generare conoscenza, Apogeo
- Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (Morgan Kaufmann Series in Data Management Systems) 3rd Edition (Slides del libro)

PARTE III

- A. Grama, A. Gupta, G. Karypis, V. Kumar. Introduction to Parallel Computing, Addison-Wesley, 2003.

- Bruce Eckel – Thinking in JAVA, 3rd Edition, Prentice Hall, Capitolo 13 sui thread. Scaricabile qui: <http://www.mindviewinc.com/Books/downloads.html>
- I. Foster. Designing and Building Parallel Programs. Addison-Wesley, 1995, Versione online disponibile qui: <http://www.mcs.anl.gov/dbpp>.

PARTE IV

- Christian Baun, Marcel Kunze, Jens Nimis, Stefan Tai. Cloud Computing. Web-Based Dynamic IT Services, Springer Science & Business Media, 2011. ISBN: 978-3-642-20916-1. <http://www.springer.com/gp/book/9783642209161>
- Rajkumar Buyya, Christian Vecchiola, S.Thamarai Selvi: Mastering Cloud Computing - Foundations and Applications Programming, Morgan Kaufmann, 468 Pages, May 2013, ISBN: 978-0-12-411454-8. <http://www.buyya.com/MasteringClouds/>

Ulteriori letture consigliate per approfondimento

Altro materiale didattico

(es. diapositive o dispense scaricabili dal sito)

- Articoli scientifici forniti dal docente;
- Diapositive fornite dal docente.

Obiettivi formativi

Conoscenza architetture di calcolo parallele e distribuite, modelli ed ambienti di programmazione parallela e distribuita; metodologie per la progettazione e sviluppo di applicazioni parallele, distribuite e web-based; preprocessing dei dati e di analisi data mining; Capacità per analisi requisiti, progettazione di applicazioni parallele o web-based; condurre esperimenti di analisi dati, utilizzo di vari linguaggi/librerie ((MPI, WebML, Python), valutazione delle prestazioni del software.

Il corso ha la finalità di implementare l'empatia, l'ascolto, la comunicazione e la capacità di lavorare in un team.

Prerequisiti

Conoscenza dei linguaggi di programmazione e della basi di dati.

Metodi didattici

Didattica frontale ed esercitazioni al calcolatore.

Altre informazioni

Si riceve previo appuntamento via e-mail.

Modalità di verifica dell'apprendimento

L'esame finale sarà svolto attraverso: (1) l'elaborato di un progetto individuale o di gruppo, (2) la discussione in seduta pubblica del progetto individuale mediante dimostrazione al calcolatore e presentazione mediante slides; (3) una prova orale sugli argomenti trattati a lezione.

Programma esteso

Analisi e progettazione di applicazioni web-based

- Richiami ad Internet, World Wide Web, protocollo HTTP
- Evoluzione Web: distribuzione di contenuti vs erogazione di applicazioni
- Architettura multi-tier delle applicazioni Web-based
 - o Elaborazione lato client: Client-side scripting & programming (JavaScript, Java Applet)
 - o Elaborazione lato server: Server-side scripting and programming (JSP, Java Servlets)
 - o Accesso ai dati: Data access layer (JDBC)
- Analisi e progettazione delle applicazioni Web
 - o Analisi dei requisiti, Specifica, Progettazione, Realizzazione e rilascio, Verifica e manutenzione
 - o Uso del linguaggio di specifica WebML: modellazione dei dati, dell'ipertesto, della navigazione
 - o Uso dello strumento WebRatio per la progettazione e sviluppo di applicazioni web-based

Data Mining e Data Science

- Il processo di scoperta di conoscenza dai dati

- Tecniche di preprocessing (data cleaning, data reduction, normalization, data transformation, distribuzione statistica dei dati)
- Feature selection
- Ensemble learning
- Basi di Machine Learning
- Ambienti software per il data mining
- Esecuzione di un esperimento completo di data mining
- Principali misure di qualità nella classificazione
- Valutazione dei risultati
- Analisi dati in Python

Calcolo Parallelo

- Architetture parallele (Tassonomia di Flynn, Tassonomia estesa di Tanenbaum)
- Richiami di Sistemi Operativi (Processi, Scheduling dei processi, Multiprogrammazione, Comunicazione e Interazione tra processi)
- Modelli di calcolo parallelo: ambiente globale (shared-memory) e ambiente locale (message-passing)
- Cenni a Semafori e Monitor
- Paradigmi a memoria condivisa: OpenMP
- Paradigmi a scambio messaggi: MPI
- Data Parallelism e Task Parallelism
- Thread: Programmazione Multithreaded in Java
- Metriche ed indici per la valutazione delle prestazioni (legge di Amhdal, speed-up, scale-up)

Cloud Computing

- Introduzione al Cloud Computing
- Virtualizzazione
- La Definizione NIST di Cloud Computing
- Caratteristiche: on demand self service, broad network access, resource pooling, rapid elasticity, measured services
- Modelli di Cloud Computing: IaaS, PaaS, SaaS
- Tipi di Cloud: Public, Private, Community, Hybrid Cloud
- Principali Piattaforme Cloud Industriali: AWS, Google AppEngine, Microsoft Azure
- Principali Piattaforme Cloud Open Source: OpenNebula, Eucalyptus
- Sviluppo di applicazioni Cloud su Google AppEngine

Al fine di promuovere i lavori di gruppo, il problem solving e la multidisciplinarietà, saranno effettuati casi-studio di progettazione e/o analisi sulle tematiche maggiormente comuni nella pratica professionale.

Obiettivi Agenda 2030 per lo sviluppo sostenibile

4. Istruzione di qualità

Link docentev

Prof. Mario Cannataro
<https://dsmc.unicz.it/personale/docente/mariocannataro>
 Chiara Zucco <https://dsmc.unicz.it/personale/docente/chiarazucco>

Obiettivi per lo sviluppo sostenibile

Codice	Descrizione
--------	-------------



Testi in inglese

Italian

The course presents the main issues, methodologies, computing infrastructures, programming environments, algorithms, and software tools for efficiently storing, managing, and analyzing key biomedical data, with particular emphasis on experimental data generated in biology (e.g., omics data such as genomics, proteomics, and interactomics) and data used in medicine (e.g., biosignals, bioimages, clinical, and healthcare data). Furthermore, the course provides the foundation for designing and developing parallel, distributed, web-based, or cloud-based applications for managing and analyzing large amounts of biomedical data. Finally, the course presents the main techniques for extracting knowledge from data, with particular emphasis on preprocessing techniques and data mining algorithms.

PART I

- Stefano Ceri, Piero Fraternali, Aldo Bongio, Marco Brambilla, Sara Comai, Maristella Matera, Progettazione di dati e applicazioni per il web, McGraw Hill, 2003, ISBN: 8838661389.
http://www.ateneonline.it/LibroAteneo.asp?item_id=1588
- WebML: www.webml.org
- WebRatio: www.webratio.com
- Core Web Programming: <http://www.corewebprogramming.com/>

PART II

- Joel Grus, Data Science con Python: Dai Fondamenti al Machine Learning, Egea
- Sebastian Raschka, Machine Learning con Python: costruire algoritmi per generare conoscenza, Apogeo
- Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (Morgan Kaufmann Series in Data Management Systems) 3rd Edition (Slides del libro)

PART III

- A. Grama, A. Gupta, G. Karypis, V. Kumar. Introduction to Parallel Computing, Addison-Wesley, 2003.
- Bruce Eckel - Thinking in JAVA, 3rd Edition, Prentice Hall, Capitolo 13 sui thread. Scaricabile qui: <http://www.mindviewinc.com/Books/downloads.html>
- I. Foster. Designing and Building Parallel Programs. Addison-Wesley, 1995, Versione online disponibile qui: <http://www.mcs.anl.gov/dbpp>.

PART IV

- Christian Baun, Marcel Kunze, Jens Nimis, Stefan Tai. Cloud Computing. Web-Based Dynamic IT Services, Springer Science & Business Media, 2011. ISBN: 978-3-642-20916-1.
<http://www.springer.com/gp/book/9783642209161>
- Rajkumar Buyya, Christian Vecchiola, S.Thamarai Selvi: Mastering Cloud Computing - Foundations and Applications Programming, Morgan Kaufmann, 468 Pages, May 2013, ISBN: 978-0-12-411454-8.
<http://www.buyya.com/MasteringClouds/>

Further recommended readings

Other teaching materials

(e.g., slides or handouts downloadable from the website)

- Scientific articles provided by the instructor;
- Slides

Knowledge of parallel and distributed computing architectures, parallel and distributed programming models and environments; methodologies for designing and developing parallel, distributed, and web-based applications; data preprocessing and data mining analysis; skills in requirements analysis, designing parallel or web-based applications; conducting data analysis experiments, using various languages/libraries (MPI, WebML, Python), and evaluating software performance.

The course aims to develop empathy, listening, communication, and teamwork skills.

	Knowledge of programming languages and databases.
	Frontal teaching and project-based learning.
	Meetings are held by appointment via email.
	The final exam will be carried out through: (1) the preparation of an individual or group project, (2) the discussion in a public session of the individual project through a computer demonstration and presentation using slides; (3) an oral exam on the topics covered in class.
	Analysis and Design of Web-Based Applications • Overview of the Internet, the World Wide Web, and the HTTP protocol • Evolution of the Web: content distribution vs. application delivery • Multi-tier architecture of Web-based applications – Client-side processing: client-side scripting & programming (JavaScript, Java Applets) – Server-side processing: server-side scripting & programming (JSP, Java Servlets) – Data access layer: data access through JDBC • Analysis and design of Web applications – Requirements analysis, specification, design, implementation and release, testing and maintenance – Use of the WebML specification language: data, hypertext, and navigation modeling – Use of the WebRatio tool for the design and development of web-based applications Data Mining and Data Science • The process of knowledge discovery from data • Preprocessing techniques (data cleaning, data reduction, normalization, data transformation, statistical data distribution) • Feature selection • Ensemble learning • Basics of machine learning • Software environments for data mining • Execution of a complete data mining experiment • Main quality measures in classification • Evaluation of results • Data analysis in Python Parallel Computing • Parallel architectures (Flynn's taxonomy, Tanenbaum's extended taxonomy) • Operating system concepts (processes, process scheduling, multiprogramming, inter-process communication and synchronization) • Parallel computation models: global environment (shared-memory) and local environment (message-passing) • Introduction to semaphores and monitors • Shared-memory paradigms: OpenMP • Message-passing paradigms: MPI • Data parallelism and task parallelism • Threads: multithreaded programming in Java • Metrics and indices for performance evaluation (Amdahl's law, speed-up, scale-up) Cloud Computing • Introduction to Cloud Computing • Virtualization • NIST definition of Cloud Computing • Characteristics: on-demand self-service, broad network access,

resource pooling, rapid elasticity, measured service

- Cloud computing models: IaaS, PaaS, SaaS
- Types of Cloud: public, private, community, hybrid
- Main industrial Cloud platforms: AWS, Google App Engine, Microsoft Azure
- Main open-source Cloud platforms: OpenNebula, Eucalyptus
- Development of Cloud applications on Google App Engine

To foster teamwork, problem solving, and multidisciplinary, case studies on design and/or analysis will be conducted on topics most relevant to professional practice.

4. Quality Education

Link

Prof. Mario Cannataro
<https://dsmc.unicz.it/personale/docente/mariocannataro>

Chiara Zucco <https://dsmc.unicz.it/personale/docente/chiarazucco>

Obiettivi per lo sviluppo sostenibile

Codice	Descrizione
--------	-------------