



NOTE (da eliminare al termine della compilazione del programma):

- le parti di testo NON evidenziate con colore NON vanno modificate;
- le parti evidenziate in verde devono essere personalizzate da ogni docente e poi tolta l'evidenziazione;
- i campi vuoti vanno riempiti secondo le indicazioni;
- tutto il testo evidenziato in giallo va eliminato nella versione finale da ognuno: costituisce solo indicazioni/istruzioni per la compilazione dei relativi campi.

Principali informazioni sull'insegnamento

Denominazione dell'insegnamento	Gestione dei dati strutturati e non strutturati	
Corso di studio	Laurea Magistrale in Data Science	
Anno Accademico	2024/25	
Crediti formativi universitari (CFU) / European Credit Transfer and Accumulation System (ECTS)	6 CFU	
Settore Scientifico Disciplinare	ING-INF/05	
Lingua di erogazione	Italiano	
Anno di corso	Primo	
Periodo di erogazione	1^ semestre, le date esatte sono riportate nel manifesto/regolamento	
Obbligo di frequenza	No, ma la frequenza è fortemente raccomandata	
Sito web del corso di studio	https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/data-science/data-science	

Docente/i	
Nome e cognome	Corrado Loglisci
Indirizzo mail	corrado.loglisci@uniba.it
Telefono	+39 0805443407
Sede	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. 5^ piano.
Sede virtuale	Piattaforma e-learning UNIBA - https://elearning.uniba.it/
Sito web del docente	https://www.uniba.it/it/docenti/corrado-loglisci
Ricevimento (giorni, orari e modalità, es. su appuntamento)	Giovedì 15:00-17:00, previa prenotazione



Syllabus	
Obiettivi formativi	<ul style="list-style-type: none">• Capacità nello studio di problematiche di organizzazione, memorizzazione e accesso di dati, di qualsivoglia complessità.• Apprendere conoscenze su modelli per rappresentare in forma direttamente computabile dati avente una forma strutturata.• Capacità d'uso di tecnologie per la organizzazione, memorizzazione e ritrovamento di dati strutturati.• Apprendere conoscenze per la definizione di una rappresentazione strutturata di dati originariamente non strutturati (non direttamente computabili)• Capacità d'uso di tecnologie per la creazione di un 'surrogato' strutturato di dati non strutturati
Prerequisiti	Principi di algebra lineare (strutture, insiemi) e programmazione ad oggetti
Contenuti di insegnamento (Programma)	<ol style="list-style-type: none">1. Dati strutturati<ol style="list-style-type: none">a. Analisi dei requisitib. Modellazione concettuale: diagrammi EERc. Modello logico relazionaled. Il linguaggio SQLe. Esercitazione su PostgreSQL2. Dati non-strutturati<ol style="list-style-type: none">a. Modello logico relazionale ad oggetti<ul style="list-style-type: none">▪ Esercitazione su PostgreSQLb. Dati spaziali<ul style="list-style-type: none">▪ Modelli computazionali▪ Oggetti spaziali, geometrie ed operazioni▪ Esercitazione su Postgisc. Dati testuali<ul style="list-style-type: none">▪ Modelli di rappresentazione▪ Pipeline di Natural language processing▪ Librerie python di Natural language processing
Testi di riferimento	<p>R.Elmasri, S.Navathe, Fundamentals of Database Systems 7th edition. Pearson, 2016.</p> <p>C.D. Manning, P. Raghavan and H. Schutze, Introduction to Information Retrieval. Cambridge Univ.Press, 2009</p> <p>Rigaux, Philippe - Scholl, Michel - Voisard, Agnès Spatial database : with application to GIS / Philippe Rigaux, Michel Scholl, Agnès Voisard. - Morgan Kaufmann, 2001</p>



	Gli studenti che lo desiderano possono ottenere i testi in prestito dalla Biblioteca. Può convenire verificarne la disponibilità mediante il Sistema Bibliotecario di Ateneo https://opac.uniba.it/easyweb/w8018/index.php? e contattare la biblioteca per concordare il prestito.			
Note ai testi di riferimento	I libri di testo sono integrati con materiale illustrato e fornito dal docente ed eventuale materiale di approfondimento, resi disponibili sulla piattaforma ADA (vedi sopra "sede virtuale").			
Organizzazione della didattica				
Ore				
Totali	Didattica frontale	Laboratorio/esercitazione	Progetto	Studio individuale
150 ore	32 ore	30 ore	88 ore	150 ore
CFU/ETCS				
6 CFU	4 CFU	2 CFU		

Metodi didattici	
	Le lezioni frontali saranno dedicate all'apprendimento dei modelli teorici e dei concetti di base, coadiuvati da alcuni esempi. Le ore di esercitazione saranno dedicate sia all'esecuzione di esercizi in classe, anche coinvolgendo direttamente gli studenti nella risoluzione degli stessi, sia alla implementazione di database, gestione dei dati e algoritmi per la trasformazione di dati spaziali e testuali. Si prevede l'utilizzo della piattaforma di e-learning del dipartimento per la pubblicazione del materiale didattico, la discussione degli argomenti delle lezioni tra docente/studente e studenti/studenti, la condivisione degli esercizi e la pubblicazione di materiale integrativo e di approfondimento.

Risultati di apprendimento previsti	
Conoscenza e capacità di comprensione	<ul style="list-style-type: none"> Acquisizione di capacità tali da affrontare problematiche di organizzazione, manipolazione e accesso a dati.



	<ul style="list-style-type: none">• Acquisizione di conoscenze su modelli concettuali e logici di rappresentazione dei dati strutturati e tecniche per la gestione di dati non strutturati usando i modelli computazionali per dati strutturati.
Conoscenza e capacità di comprensione applicate	<ul style="list-style-type: none">• Applicazione di modelli e tecniche di gestione di dati strutturati e non strutturati nella progettazione di sistemi software orientati ai dati.• Applicazione delle capacità d'uso di tecnologie per basi di dati e per la trasformazione (pre-elaborazione) di dati non strutturati.
Competenze trasversali	<p>Autonomia di giudizio</p> <ul style="list-style-type: none">• Valutazione di adeguatezza e criticità di un modello di rappresentazione di dati per il problema di gestione dati alla mano.• Valutazione delle possibili soluzioni rispetto alle caratteristiche del dato da rappresentare e alle informazioni di interesse per le applicazioni specifiche. <p>Abilità comunicative</p> <ul style="list-style-type: none">• Capacità di illustrare in modo adeguato concetti, modelli e tecniche per la gestione di dati strutturati e non strutturati.• Capacità di illustrare in modo formale e critico le scelte di modellazione fatte rispetto al problema di gestione dati alla mano. <p>Capacità di apprendere in modo autonomo</p> <ul style="list-style-type: none">• Capacità di analisi e problem solving per la gestione dei dati di interesse per le applicazioni specifiche.• Capacità di sintetizzare aspetti cruciali dei dati di interesse per la applicazione specifica e individuazione di modelli di rappresentazione adeguati.

Valutazione	
Modalità di verifica dell'apprendimento	<p>Discussione orale riguardante:</p> <ul style="list-style-type: none">• Progetto/caso di studio realizzato con le tecnologie viste durante il corso (da concordare col docente, in tempo congruo). Il progetto richiede capacità di applicazione di concetti visti per la gestione di dati strutturati e non strutturati.



	<ul style="list-style-type: none">• Concetti appresi durante il corso. Richiede capacità di esposizione di concetti visti per la gestione di dati strutturati e non strutturati.
Criteria di valutazione	<p>Lo studente sarà valutato sulla base delle</p> <ul style="list-style-type: none">• competenze acquisite su modelli di rappresentazione e strumenti per la definizione, interrogazione e manipolazione di dati strutturati• competenze acquisite su tecniche per la costruzione di rappresentazioni strutturate da dati non-strutturati.• capacità di applicare le suddette competenze a casi reali attraverso le tecnologie presentate durante il corso.
Criteria di misurazione dell'apprendimento e di attribuzione del voto finale	<p>Il voto finale è attribuito in trentesimi. L'esame si intende superato quando il voto finale è maggiore o uguale a 18. L'esame si intende superato quando la valutazione è maggiore o uguale a 18 sia per la discussione riguardante il progetto sia per la esposizione dei concetti appresi.</p>
Altro	<p>Si suggerisce agli studenti di affidarsi esclusivamente alle informazioni/comunicazioni fornite sui siti ufficiali del Dipartimento di Informatica, ovvero sui gruppi social solo se costituiti e amministrati esclusivamente dai docenti dei relativi insegnamenti:</p> <ul style="list-style-type: none">• https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea• https://www.uniba.it/it/ricerca/dipartimenti/informatica• https://elearning.uniba.it/ <p>I programmi di tutti gli insegnamenti sono disponibili al seguente link:</p> <ul style="list-style-type: none">• https://elearning.uniba.it/ <p>Le informazioni che tutti gli studenti dovrebbero conoscere sono scritte nei regolamenti didattici dei Corsi di Studi disponibili nel sito:</p> <ul style="list-style-type: none">• https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea <p>Si suggerisce agli studenti di diffidare delle informazioni e dei materiali circolanti su siti o gruppi social non ufficiali, poiché spesso sono risultati non affidabili, non corretti o incompleti. Per ogni dubbio, chiedere un incontro al docente secondo le modalità previste per il ricevimento.</p> <hr/> <p>https://elearning.uniba.it/course/view.php?id=4964</p>



Main information on the course

Course name	Structured and Unstructured data management	
Degree	Master degree in Data Science	
Academic year	2024/25	
European Credit Transfer and Accumulation System (ECTS), in Italian Crediti Formativi Universitari (CFU)	6 CFU	
Settore Scientifico Disciplinare		
Course language	Italian	
Course year	First	
Course period	First Semester - exact dates can be found in the didactic regulations	
Course attendance requirement	None, but it is highly recommended to attend classes	
Website of the Degree	https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/data-science/data-science	

Teacher(s)

Name and Surname	Corrado Loglisci
email	corrado.loglisci@uniba.it
phone	+39 0805443407
office	Department of Computer science, Via Orabona 4, 70125, Bari. Floor 5 th
e-learning platform	Piattaforma ADA - https://elearning.di.uniba.it/
Teacher's homepage	https://www.uniba.it/it/docenti/corrado-loglisci
Office hours	Thursday 3:00-5:00 pm, by reservation

Syllabus

Course goals	<ul style="list-style-type: none">• Ability to study problems of organization, storage and access of data, of any complexity.• Learn knowledge of models to represent in a directly computable form data having a structured form.• Ability to use technologies for the organization, storage and retrieval of structured data.• Learn knowledge for the definition of a structured representation of originally unstructured data (not directly computable)• Ability to use technologies for the creation of a structured 'surrogate' of unstructured data
Prerequisites/requirements	Principles of linear algebra (structures, sets) and object-oriented programming
Course program	<ol style="list-style-type: none">1. Structured data<ul style="list-style-type: none">• Requirements analysis• Conceptual modeling: EER diagrams• Relational logical model• ANSI SQL language<ul style="list-style-type: none">◦ Practice on PostgreSQL2. Unstructured data



	<ul style="list-style-type: none"> • Object-relational logical model <ul style="list-style-type: none"> ○ Practice on PostgreSQL • Spatial data <ul style="list-style-type: none"> ○ Computational models ○ Spatial objects, geometries and operations ○ Practice on Postgis • Textual data <ul style="list-style-type: none"> ○ Representation models ○ Pipeline of Natural language processing pipeline ○ Python libraries of Natural language processing 			
Books of reference	<p>R.Elmasri, S.Navathe, Fundamentals of Database Systems 7th edition. Pearson, 2016.</p> <p>C.D. Manning, P. Raghavan and H. Schutze, Introduction to Information Retrieval. Cambridge Univ.Press, 2009</p> <p>Rigaux, Philippe - Scholl, Michel - Voisard, Agnès Spatial database : with application to GIS / Philippe Rigaux, Michel Scholl, Agnès Voisard. - Morgan Kaufmann, 2001</p> <p>Students who wish can borrow texts from the Library. It may be advisable to check availability through the University Library System https://opac.uniba.it/easyweb/w8018/index.php? and contact the library to arrange the loan.</p>			
Notes to the books	The textbooks are integrated with illustrated material provided by the teacher and any additional material, made available on the ADA platform (see above “virtual location”).			
Organization of the didactic activities				
Hours				
Total	Lectures	Practice sessions	Project work	Individual study
150 hours	32 hours	30 hours	0 hours	88 hours
CFU/ETCS				
6 CFU	4 CFU	2 CFU	0 CFU	
Teaching methods				
	<p>The lectures will be dedicated to learning theoretical models and basic concepts, supported by some examples. The hours of practice will be dedicated both to the execution of exercises in class, also directly involving students in solving them, and to the implementation of databases, data management and algorithms for the transformation of spatial and textual data. The use of the department's e-learning platform is planned for the publication of teaching material, the discussion of lesson topics between teacher/student and student/student, the sharing of exercises and the publication of supplementary and in-depth material.</p>			



Expected learning outcomes	
Knowledge and understanding	<ul style="list-style-type: none">○ Acquisition of skills to address issues of organization, manipulation and access to data.○ Acquisition of knowledge on conceptual and logical models of representation of structured data and techniques for the management of unstructured data using computational models for structured data.
Applying knowledge and understanding	<ul style="list-style-type: none">○ Application of structured and unstructured data management models and techniques in the design of data-oriented software systems.○ Application of the skills of using database technologies and for the transformation (pre-processing) of unstructured data.
Other skills	<p><i>Making judgements</i></p> <ul style="list-style-type: none">○ Evaluation of the adequacy and criticality of a data representation model for the data management problem at hand.○ Evaluation of possible solutions with respect to the characteristics of the data to be represented and the information of interest for specific applications. <p><i>Communication</i></p> <ul style="list-style-type: none">○ Ability to adequately illustrate concepts, models and techniques for managing structured and unstructured data.○ Ability to formally and critically illustrate the modeling choices made with respect to the data management problem at hand. <p><i>Learning skills</i></p> <ul style="list-style-type: none">○ Analysis and problem solving skills for the management of data of interest for specific applications.○ Ability to summarize crucial aspects of the data of interest for the specific application and identification of adequate representation models.

Assessment	
Assessment methods	<p>Oral discussion regarding:</p> <ul style="list-style-type: none">○ Project/case study created with the technologies seen during the course (to be agreed with the teacher, in a reasonable amount of time). The project requires the ability to apply concepts seen for the management of structured and unstructured data.○ Concepts learned during the course. It requires the ability to present concepts seen for the management of structured and unstructured data.
Evaluation criteria	The student will be assessed on the basis of



	<ul style="list-style-type: none">- skills acquired on representation models and tools for defining, querying and manipulating structured data- skills acquired on techniques for building structured representations from unstructured data.- ability to apply the aforementioned skills to real cases through the technologies presented during the course.
Measurements and final grade	The final grade is awarded in thirtieths. The exam is considered passed when the final grade is greater than or equal to 18. The exam is considered passed when the evaluation is greater than or equal to 18 both for the discussion regarding the project and for the presentation of the concepts learned.
Further information	https://elearning.uniba.it/course/view.php?id=4964