00

Course Overview

Notice

- **Author**

  - ◆ **João Moura Pires (<u>jmp@fct.unl.pt</u>), Carlos Viegas Damásio (<u>cd@fct.unl.pt</u>) and with contributions of José Júlio Alferes**

- **This material can be freely used for personal or academic purposes without any previous authorization from the author, provided that this notice is maintained/kept.**

- **For commercial purposes the use of any part of this material requires the previous authorization from the author.**

Table of Contents

- **Lecturers**

- **Course Content**

- **Course Organization**

- **Bibliography**

- **Assessment**

- **Web Site**

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

NOVA**LINCS**
LABORATORY FOR COMPUTER
SCIENCE AND INFORMATICS

Lecturers

- **João Moura Pires**

  - ◆ Email: jmp@fct.unl.pt

  - ◆ Office: P3/2

  - ◆ Internal phone: 10746

- **Main Research interests**

  - ◆ Spatio temporal analytics

  - ◆ Data Mining

  - ◆ Data Visualization

Course Content

Goals

- **Data modeling** and **query languages** for semantically enriched data, big data and/or open data applications in the Web (NoSQL models)

- **Graph models** for representing interconnected data, in particular for representing relationships between resources in the Web.

- **Modeling multidimensional data** for analytical processing, in particular for representing time and space dimensions for understanding how data evolves and moves

- Use commercial and academic systems to **address practical problems**.

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

NOVA**LINCS**
LABORATORY FOR COMPUTER SCIENCE AND INFORMATICS

Syllabus – **NoSQL data models**

- Alternative models for storing big volumes of data.

- Column, document and graph models.

- Relational, semi-structured and graph data.

- Data modelling with graphs.

- Querying graph models.

- Graph databases.

- Relationship to NoSQL movement and key-value stores.

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

NOVA**LINCS**
LABORATORY FOR COMPUTER
SCIENCE AND INFORMATICS

Syllabus – **Analytics with (big) data**

- Data Warehouses

- (Conceptual) multidimensional data models

- Typical OLAP operations and OLAP query languages. Metadata

- Spatial and temporal dimensions

- Interaction in the data analysis process.

- Reference architectures of a Data Warehouse

- Data Warehouses and big data

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

NOVA**LINCS**
LABORATORY FOR COMPUTER
SCIENCE AND INFORMATICS

Syllabus – **Semantic Web**

- Motivation.

- Linked Open Data.

- Language and semantics of the Resource Description Framework (RDF) and SPARQL query language.

- Ontologies in the Semantic Web: RDF Schema and Web Ontology Language (OWL)

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

NOVA**LINCS** LABORATORY FOR COMPUTER SCIENCE AND INFORMATICS

In the Labs

- Temporal databases

- Neo4j

- Jena Framework/and or GraphDB

- SPARQL language

- Hadoop based query languages

- **Setting-up Your Project**

Course Organization

Course Organization

- **Lectures on Wednesdays from 11:10 to 13:00**

- **Labs**

  - **P1: Tuesday from 11:10 to 13:00**

  - **P2: Wednesday, from 9:10 to 11:00.**

- **Labs Start next Week. October 6th.**

- **Check the schedule on the website**

- **Check the web site: http://md.ssdi.di.fct.unl.pt/index.html**

| Tue	Wed
	09:00 **MD - P2** Lab 119 - Ed. 2
11:00 **MD - PI** Lab 122 - Ed. 2	11:00 **MD - T** 128 - Ed 2

Bibliography

Bibliography

- The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (Third Edition). Ralph Kimball, Margy Ross.Wiley, 2013, ISBN: 978-1-118-53080-1

- Big Data: Concepts, Warehousing, and Analytics (River Publishers Series in Information Science and Technology), by Maribel Yasmina Santos (Author), Carlos Costa (Author), 2019, ISBN-13: 978-8770221849, ISBN-10: 8770221847

- Papers
  - From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design - from Daniel L. Moody.

- Slides of the course
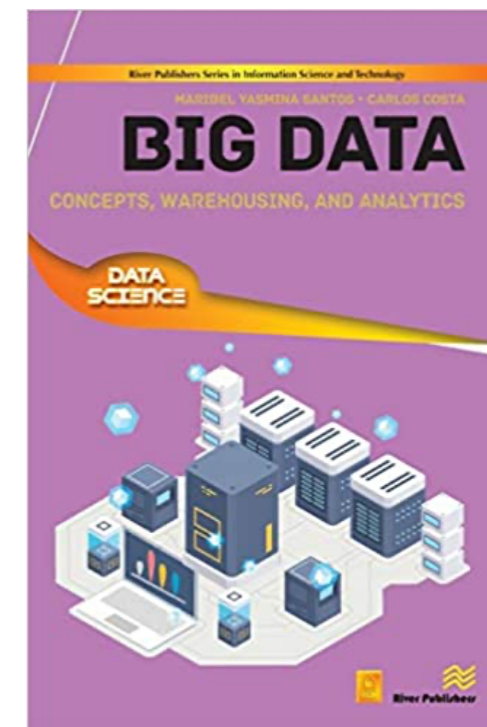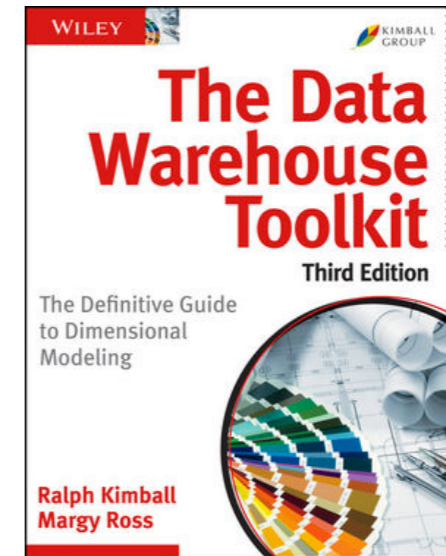
Bibliography

- Graph Databases, Ian Robinson, Jim Webber & Emil Eifrem, O'Reilly, 2015

- Guy Harrison. Next Generation Databases: NoSQL, NewSQL and Big Data. Apress, 2015. ISBN: 978-1484213308.

- Dan Sullivan. NoSQL for Mere Mortals. Addison-Wesley, 2015. ISBN: 978-0134023212

- Ted Hills. NoSQL and SQL data modeling. Technics Publications, 2016. ISBN: 978-1634621090
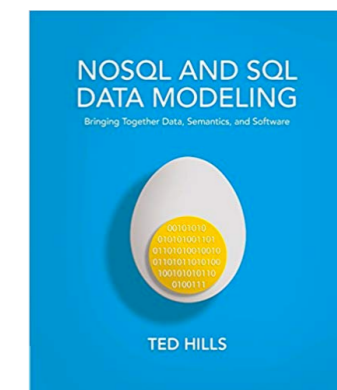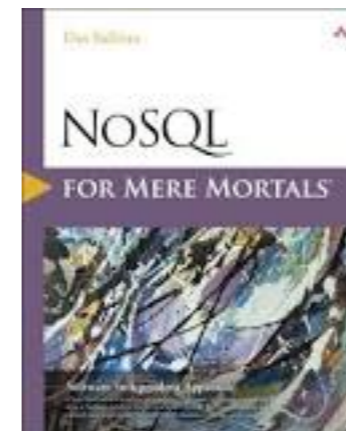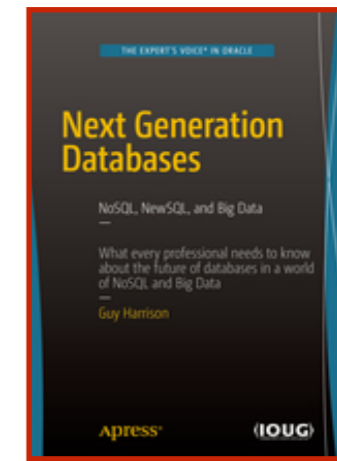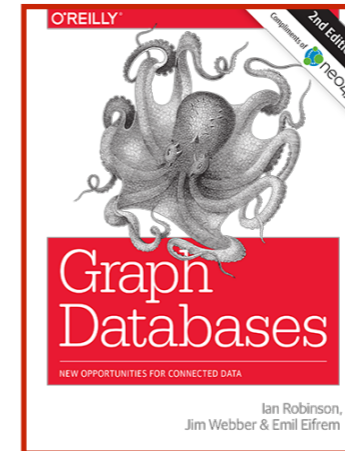
Bibliography

- Foundations of Semantic Web

  Technologies, Pascal Hitzler, Markus

  Krotzsch and Sebastian Rudolph,

  Chapman & Hall/CRC, 2009

- A Semantic Web Primer, Third Edition,

  Grigoris Antoniou, Paul Groth, Frank van

  Harmelen and Rinke Hoekstra, MIT Press,

  2012

- Slides of the course
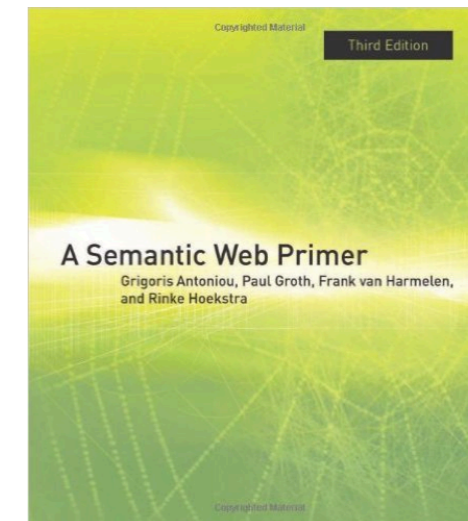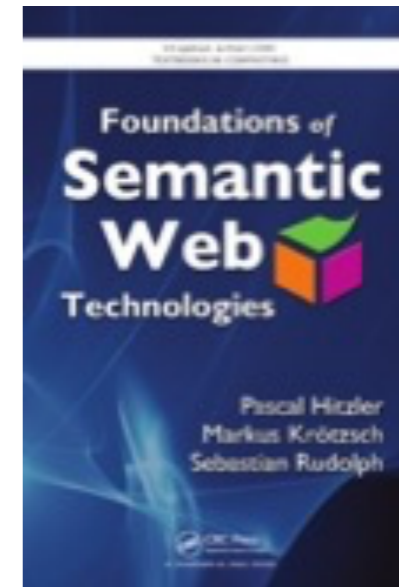
Assessment

Assessment

- **Two mid-term written tests**

  - First week of November, and second week of January

  - Each is worth 25% of the final grade

- One **project** worthing **50%** of the grade,

  - Practical project, done in groups of **2 students**;

  - Each student gets an **individual grade after an oral presentation** and discussion of

    the Project.

- Students approved in the project and not approved in the tests may also do a final exam,

  and the resulting grade replaces the grades of the tests in the final grade formula

Assessment - Requirements

- Course approval requires the following minimal grades:

  - (mean (Test1; Test2) >= 10) AND

  - (Project >= 10) AND

  - (Test1 >= 8) AND

  - (Test2 >= 8)

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

NOVA**LINCS** LABORATORY FOR COMPUTER SCIENCE AND INFORMATICS

Important Dates

- Team Registration - until October 12th

  - ◆ Using a google sheet google
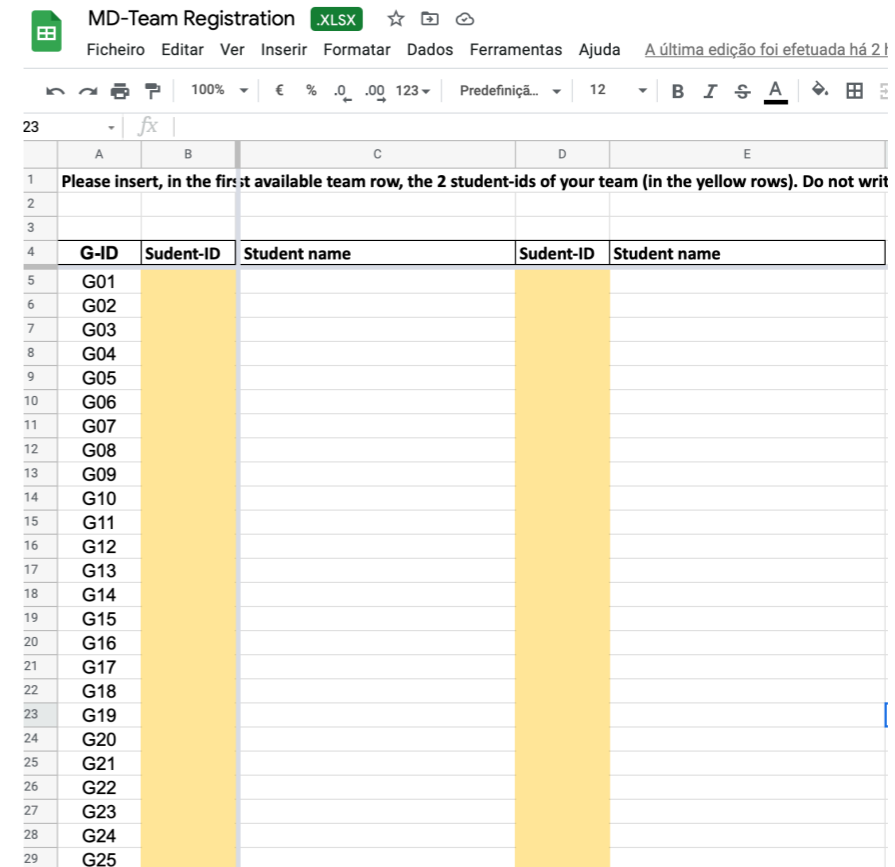
- Subject Registration - until November 9th

  - ◆ As soon as possible

  - ◆ As many iterations as possible

  - ◆ Writing a small document

- Team work delivery - Until January 2th (best date December 29th)

  - ◆ Via google drive - Shared folder (will be shared after the team registration)

- Team work oral presentation and discussion

  - ◆ January 10 - 14

Web Site

Web site: http://md.ssdi.di.fct.unl.pt/

Web site: Information

MD 21/22
(Advanced) Data Modeling

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Home
News
Information
  Bibliography
  Sylabus
  Evaluation Rules
  Schedule
Resources
Summaries
Training
Evaluation

(Advanced) Data Modeling (DM) is a 6 ECTS curricular unit of the set of specialization units offered to the students of 4th or 5th year of Mestrado Integrado em Engenharia Informática (MIEI). This course address data modeling and query languages for semantically enriched data, big data and/or open data applications in the Web. The core of the course will cover graph models, in particular those allowing the representation of relationships between resources in the Web. Besides representing interconnections among data, it is necessary to model and explore multidimensional data for online analytical processing, in particular for representing time and space dimensions for understanding how data evolves and moves.

This course is provided by Departamento de Informática (DI) da Faculdade de Ciências e Tecnologia (FCT) da Universidade Nova de Lisboa (UNL).

Objectives:
Knowledge:

- Graph modeling and query languages
- Linked Open Data principles and Semantic Web concepts
- Languages for representing, reasoning and querying in the Semantic Web
- Concepts, architectures and models of a Data Warehouse
- Multidimensional data modeling for OLAP querying.

Application:

- Identify applications requiring graph modeling
- Model a graph database and query it (e.g. Neo4j with Cypher queries)
- Use a triple store and inference engine (e.g. Apache Jena) for querying with SPARQL data in the Semantic Web
- Analyze, design and query multidimensional models.

Soft-Skills:

- To explore autonomously the recent bibliography of a topic
- To develop critical reasoning regarding recent technology
- To work in a team
- To orally present a survey on a recent topic
- To review a scientific work

Prerequisites:
Knowledge and practice on general programming and relational databases.

Teacher
João Moura Pires (jmp@fct.unl.pt)

Schedule (see at Schedule that will be updated)
Lectures:

- **Portuguese**
- **English** (if required) spoken lectures

Web site: Schedule

Web site: Schedule

Web site: Resources

MD 21/22
(Advanced) Data Modeling

FCt FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

News / Resources /

- Home
- News
- Information
- **Resources**
  - Lectures
  - Papers
  - Links
  - Miscellaneous notes
- Summaries
- Training
- Evaluation

You will find all resources made available for the students, including pdfs of the slides used during the lectures, selected reference resources, short notes and messages produced by the teacher or contributions from students. See also Labs and Links.

This resources do not replace the Bibliography!!

Now the following sections are available:

From Lectures: Material used by the teacher during the lectures.

Papers: Recommended papers

Training: Material used on (individual or group) training sessions.

Miscellaneous notes: Miscellaneous notes like important messages sent by the teacher, summaries from workshops, etc.

Links: This page has links to sites or pages of interest in terms of content and relevance for the students.

Carlos Damásio   João Moura Pires   Contact

Web site: Summaries

MD 21/22
(Advanced) Data Modeling

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

News / Summaries /

Home

News

Information

Resources

Summaries

Training

Evaluation

Lectures

RSS Feed

[T01]: Course overview

29 Sep 2021, 11:10 AM Filed in: Lectures

Course Organization and Overview: Syllabus; Bibliography; Evaluation rules; important dates, etc..
Big data and challenges. The NoSQL movement, CAP and PACELC theorems. Types of NoSQL database systems. Labeled Property Graphs.

Recommended Readings: (i) Consistency Tradeoffs in Modern Distributed Database System Design, from Daniel J. Abadi **(ii)** the page in http://blog.nahurst.com/visual-guide-to-nosql-systems.

Recommended Activities: (i) see the following videos "DT&SC 7-3: What is Big Data?, DT&SC 7-4: Digital Big Data Footprint. from Martin Hilbert **(ii)** Visit the various sections of this site.

To Know:

- A first understand of what is BigData, its main characteristics, and the main challenges
- Understand the examples of digital footprint from the second video Martin Hilbert
- What is the the Semantic Web Vision
- What is about the CAP theorem and its implications
- understand the NoSQL quadrants and the motivation for each one

Web site: Team Work

MD 21/22
(Advanced) Data Modeling

FCt FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

- Home
- News
- Information
- Resources
- Summaries
- Training
- Evaluation
  - Subjects

This area is dedicated in general to the evaluation process and more specifically to the team-work projects, namely detailed instructions about the **registration of teams** and **registration of subjects**, **report delivering** and **team-work discussion**.

Instructions

Team registration

The students have to create **teams with 2 students** and make a **team registration**. The team registration process requires filling the appropriate information in the shared google sheet. After a successful team registration an identification group is like "GNN".

Subject definition

A registered team must register its subject choice for the Practical team-Work. The purpose of subject registration is to provide the teacher with a clear description of the chosen subject, the functional scope of the implementation, the proposed architecture and the technologies used, to get a teacher's approval, and from that point you should keep your subject registration updated as you start the development. By doing that you keep an appropriate live summary of your project for you, for your colleague and for the teacher. You may **update** your subject registration **as many times you want**. Please make your subject registration as clear and complete as possible. Consider the following recommendations:

Abstract; should be a short paragraph which describes the goal of your system and its most important features.

Description: you should describe: (i) the scope of your project; (ii) the proposed architecture and approach.
In order to present the scope you should identify the inputs, the main concerns about the data or the data model, the functionalities and eventually the outputs.
With the respect with the architecture you should describe the main components, how their purpose, how they exchange information and the technology you intend to use and their role.

Tasks: you should keep a list of tasks, as detailed as possible, in order to plan your work and check the feasibility of your approach. By doing this you will have to think in a very concrete way of how to address the all project and consequently you check if the previous description is sound. Also you break your original problem into many small problems that you can distribute the by the team members.

References: the text books, the papers, the sites you use during the design and the implementation of your project

Software: list of software you will use in the development of your project

Development: you can use this section for further details and eventually to indicate a URL where the teacher can check your system.

Shared Google Folder

After the team registration we will share a google folder with each team that will be used to deliver both the subject registration and the final project report and others.

Home

News

Information

Resources

Summaries

Training

Evaluation

Subjects

This area is dedicated in general to the evaluation process and more specifically to the team-work projects, namely detailed instructions about the **registration of teams** and **registration of subjects**, **report delivering** and **team-work discussion**.

Instructions

Team registration

The students have to create **teams with 2 students** and make a **team registration**. The team registration process requires filling the appropriate information in the shared google sheet. After a successful team registration an identification group is like "GNN".

Subject definition

A registered team must register its subject choice for the Practical team-Work. The purpose of subject registration is to provide the teacher with a clear description of the chosen subject, the functional scope of the implementation, the proposed architecture and the technologies used, to get a teacher's approval, and from that point you should keep your subject registration updated as you start the development. By doing that you keep an appropriate live summary of your project for you, for your colleague and for the teacher. You may **update** your subject registration **as many times you want**. Please make your subject registration as clear and complete as possible. Consider the following recommendations:

Abstract; should be a short paragraph which describes the goal of your system and its most important features.

Description: you should describe: (i) the scope of your project; (ii) the proposed architecture and approach.
In order to present the scope you should identify the inputs, the main concerns about the data or the data model, the functionalities and eventually the outputs.
With the respect with the architecture you should describe the main components, how their purpose, how they exchange information and the technology you intend to use and their role.

Tasks: you should keep a list of tasks, as detailed as possible, in order to plan your work and check the feasibility of your approach. By doing this you will have to think in a very concrete way of how to address the all project and consequently you check if the previous description is sound. Also you break your original problem into many small problems that you can distribute the by the team members.

References: the text books, the papers, the sites you use during the design and the implementation of your project

Software: list of software you will use in the development of your project

Development: you can use this section for further details and eventually to indicate a URL where the teacher can check your system.

Shared Google Folder

After the team registration we will share a google folder with each team that will be used to deliver both the subject registration and the final project report and others.

Web site: Team Work

- **One project worthing 50% of the grade**

- **Team Registration - until October 7th**

- **Subject Registration - until November 11th**

  - ◆ **Many iterations**

- **Team work delivery - Until December 18th**

- **Team work oral presentation and discussion - December 21 and 22th,**

Examples of subjects

General plan

	Week	Content	Assessment Dates
1	27/Set- 2/Out	**Course Overview and Introduction**	
2	04/Out - 9/Out	**Introduction to DW**	
3	11/Out - 16/Out	**From Enterprise Models to Dimensional Models**	Team Registration
4	18/Out - 23/Out	**Multidimensional Modeling**	
5	25/Out - 30/0ut	**Multidimensional Modeling**	
6	01/Nov - 06/Nov	**Big Data Warehouse**	
7	08/Nov - 13/Nov	**Temporal Databases**	First Test
8	15/Nov - 20/Nov	**Graph Databases**	Subject Registration
9	22/Nov - 27/Nov	**The Semantic Web Vision**	
10	29/Nov - 04/Dez	Feriado a 1 de Dezembro	
11	06/Dez - 11/Dez	Feriado a 8 de Dezembro	
12	13/Dez - 18/Dez	**RDF and RDFS and RDFS semantics**	
13	20/Dez - 25/dez	**SPARQL query language**	
14	27/Dez - 1/Jan		Team Work Delivery
15	03/Jan - 8/Jan		Second Test
16	10/Jan - 15/Jan		Oral Discussions

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

NOVA**LINCS**
LABORATORY FOR COMPUTER
SCIENCE AND INFORMATICS

Further Reading and Summary

Q&A

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

NOVA**LINCS**
LABORATORY FOR COMPUTER
SCIENCE AND INFORMATICS