

# Business Intelligence I

## Lab 1 – Introduction + Project Overview

Academic Year 2023-2024

Rui Monteiro

Pedro Batista

Bruno Jardim

Duarte Rodrigues

[rmonteiro@novaims.unl.pt](mailto:rmonteiro@novaims.unl.pt)

[ibatista@novaims.unl.pt](mailto:ibatista@novaims.unl.pt)

[bjardim@novaims.unl.pt](mailto:bjardim@novaims.unl.pt)

[duarte.rodrigues@novaims.unl.pt](mailto:duarte.rodrigues@novaims.unl.pt)

# Lab Plan

- Lab calendar
- Lab methodology
- Lab software and resources
- BI I (practical) Project overview

**NOVA**

**IMS**

Information  
Management  
School

# Lab Calendar

Instituto Superior de Estatística e Gestão de Informação  
Universidade Nova de Lisboa

# Lab Calendar

## Lab general topics (subject to small changes):

### Lab 1: Program Overview (12 Sep.)

Overview of BI I Labs, presentation of Project Guidelines and review of course software

### Lab 2: Data Warehousing (10 Oct.)

Business Needs, designing a Data Warehouse, implementing a DW schema

### Lab 3: Practical steps in DW (17 Oct.)

Building a Data Warehouse, implementing SCDs, Date dimensions, further SQL

### Lab 4: Introduction to ETL (31 Oct.)

Understanding ETL, first steps with Data Integration and Transformation tools

### Lab 5: Continuing with ETL (7 Nov.)

Developing a Full Load of the Staging Area, with integral load of dimensions and facts

### Lab 6: Improving the ETL (14 Nov.)

Enhancing the SA load with logs, use of variables, monitoring for errors

### Lab 7: Completing full ETL (21 Nov.)

Additional ETL to load the DW, testing full integral loads to SA and DW with SCDs

### Lab 8: Making ETL dynamic (28 Nov.)

Use of project parameters, adding logic to allow for both integral and incremental loads

### Lab 9: Review of Lab Work (5 Dec.)

Review and consolidation of work covered in all the Labs, demonstration of PowerBI

# Lab Calendar (cont.)

- Further details will be made available with each Lab
- Please keep in mind the following important dates:
  - We expect to make detailed Project Guide (Handout) available on 10<sup>th</sup> Oct. (2<sup>nd</sup> Lab)
  - Last week of October is Trimester break/Exam period (no mid-term exam in BI 😎)
  - First **INTERMEDIATE PROJECT DELIVERY** is due 23h59 on **19<sup>th</sup> of November**
  - Last **FINAL PROJECT DELIVERY** is due 23h59 on **23<sup>rd</sup> of December**
  - Both dates (and times) are non-negotiable; see further details on “Project section”

**NOVA**

**IMS**

Information  
Management  
School

# Lab Methodology

Instituto Superior de Estatística e Gestão de Informação  
Universidade Nova de Lisboa

# Lab Methodology

- All labs (regardless of teacher/class/time) will follow the same methodology:
  - We will explain context of WHAT IS GOING TO BE DONE
  - We will demonstrate DOING IT using a special (customized) use-case (IMSALES)
  - We expect EVERY STUDENT to follow along, in the lab; we help fix any problems
  - We make each Lab solution (solved use-case) available in Moodle, after each Lab
  - We STRONGLY RECOMMEND each student replicates – by themselves – the techniques and skills demonstrated in the Lab, in their own time
  - The PRACTICAL PROJECT basically uses these same techniques, skills and mindset to solve a different challenge (a new data problem, chosen by your Group)

# Lab Methodology (cont.)

- All labs will make use of a fictional company, to demonstrate all of the BI techniques and skills you will be learning:



IMSales is a trading company (business-to-business) that imports and exports **specialty foods from all around the world**. IMSales does not manufacture its own products, these are manufactured by their suppliers and sold to third-party retailers (Clients).

IMSales Headquarters are in Seattle, USA and there is a second office in London, UK that works as a business representation.



# Lab Methodology (cont.)

Each individual Lab teacher will have their own (distinct) Office Hours (2 hours each week).

These will be **published and announced in Moodle**.

We will also have a dedicated **Lab & Project Help forum** in Moodle, where all students are encouraged to place their questions and problems (both for Labs and the Project).

Please note that emails sent directly to teachers – *unless to do with personal (private) questions/issues* – may not be answered (or will be answered quite late...).

The **ONLY OFFICIAL** (monitored) channel is **MOODLE**.

# Lab Methodology (cont.)

☆ Lookups for Loading of Fact Table in DW not working	 28 de dezembro de ;	 Pedro Batista 29 de dezembro de ;	3	⋮
☆ Error CSV file datatype four byte signed integer	 28 de dezembro de ;	 Pedro Batista 28 de dezembro de ;	1	⋮
☆ Incremental Load - Last load date variable	 19 de dezembro de ;	 Pedro Batista 21 de dezembro de ;	1	⋮
☆ Error during Incremental Load	 19 de dezembro de ;	 Pedro Batista 19 de dezembro de ;	1	⋮
☆ SQL_Script_Create_Table	 19 de dezembro de ;	 Pedro Batista 19 de dezembro de ;	1	⋮
☆ Surrogate key and Business Key (STG and DW)	 18 de dezembro de ;	 Pedro Batista 18 de dezembro de ;	1	⋮
☆ Primary key for fact table	 20 de novembro de ;	 Pedro Batista 21 de novembro de ;	3	⋮
☆ 11th Question Unavailable - FAQ	 21 de novembro de ;	 Pedro Batista 21 de novembro de ;	1	⋮
☆ Project language	 20 de novembro de ;	 Pedro Batista 20 de novembro de ;	1	⋮
☆ Project questions - datasets, hierarchies, dimensions	 10 de novembro de ;	 Pedro Batista 10 de novembro de ;	1	⋮
☆ This week's class - recording?	 30 de outubro de 20	 Pedro Batista 7 de novembro de 21	3	⋮
☆ Dataset size limit for the project	 4 de novembro de 21	 Pedro Batista 4 de novembro de 21	1	⋮
☆ Whatsapp group creation for group formation?	 6 de setembro de 20	 20 de outubro de 20	33	⋮
☆ Can we enrich our starting dataset, by bringing in new data?	 Pedro Batista 18 de outubro de 20	 Pedro Batista 18 de outubro de 20	0	⋮
☆ P1 & P4 with prof. Rui Monteiro on Saturday Oct 15	 14 de outubro de 20	 Rui Monteiro 15 de outubro de 20	3	⋮
☆ Exam Mode	 12 de outubro de 20	 Pedro Batista 13 de outubro de 20	2	⋮
☆ Saturday's class - recording?	 7 de outubro de 202	 Pedro Batista 11 de outubro de 20	6	⋮
☆ [FAQ] Getting an error when using diagrams in SSMS? See this post, please	 Pedro Batista 11 de setembro de 2	 Pedro Batista 13 de setembro de 2	4	⋮
☆ Error in Setup	 7 de setembro de 20	 Pedro Batista 8 de setembro de 20	3	⋮

A sample of previous Forum questions

**NOVA**

**IMS**

Information  
Management  
School

# Lab Software and Resources

# Lab Software and Resources

The essence of the BI work we will be executing will be Cloud-based (as explained in Lecture).

As such, there is very little to install, on your part.

The one exception is **Azure Data Studio** (as also used for the “SQL and Databases Crash Course”; only required by the time of the 2<sup>nd</sup> Lab – 10<sup>th</sup> of October; see next slides).

Almost no prior knowledge or skills are required for Business Intelligence... again, the only (minor) exception is a **basic understanding of SQL / Databases**. Obviously, the Crash Course is already sufficient; alternatives are available – see next slides.

# Lab Software and Resources (cont.)



## Azure Data Studio

(available for Windows + MacOS + Linux)

Download and install from Microsoft:

<https://learn.microsoft.com/en-us/sql/azure-data-studio/download-azure-data-studio>

(on Windows, select User Installer)

# Lab Software and Resources (cont.)

The screenshot displays the Azure Data Studio interface in dark theme. On the left, the 'CONNECTIONS' pane shows a tree view of servers and databases. The 'SQLQuery\_1 - localhost:grated' window shows a SQL query: `SELECT TOP (1000) [ClientID], [CompanyName], [ContactName], [ContactTitle], [Address], [City], [Region], [PostalCode], [Country], [Phone], [Fax] FROM [IMSALES].[dbo].[clients]`. The 'Results' pane shows a table with 10 columns and 29 rows of data.

**Azure Data Studio with Dark Theme**

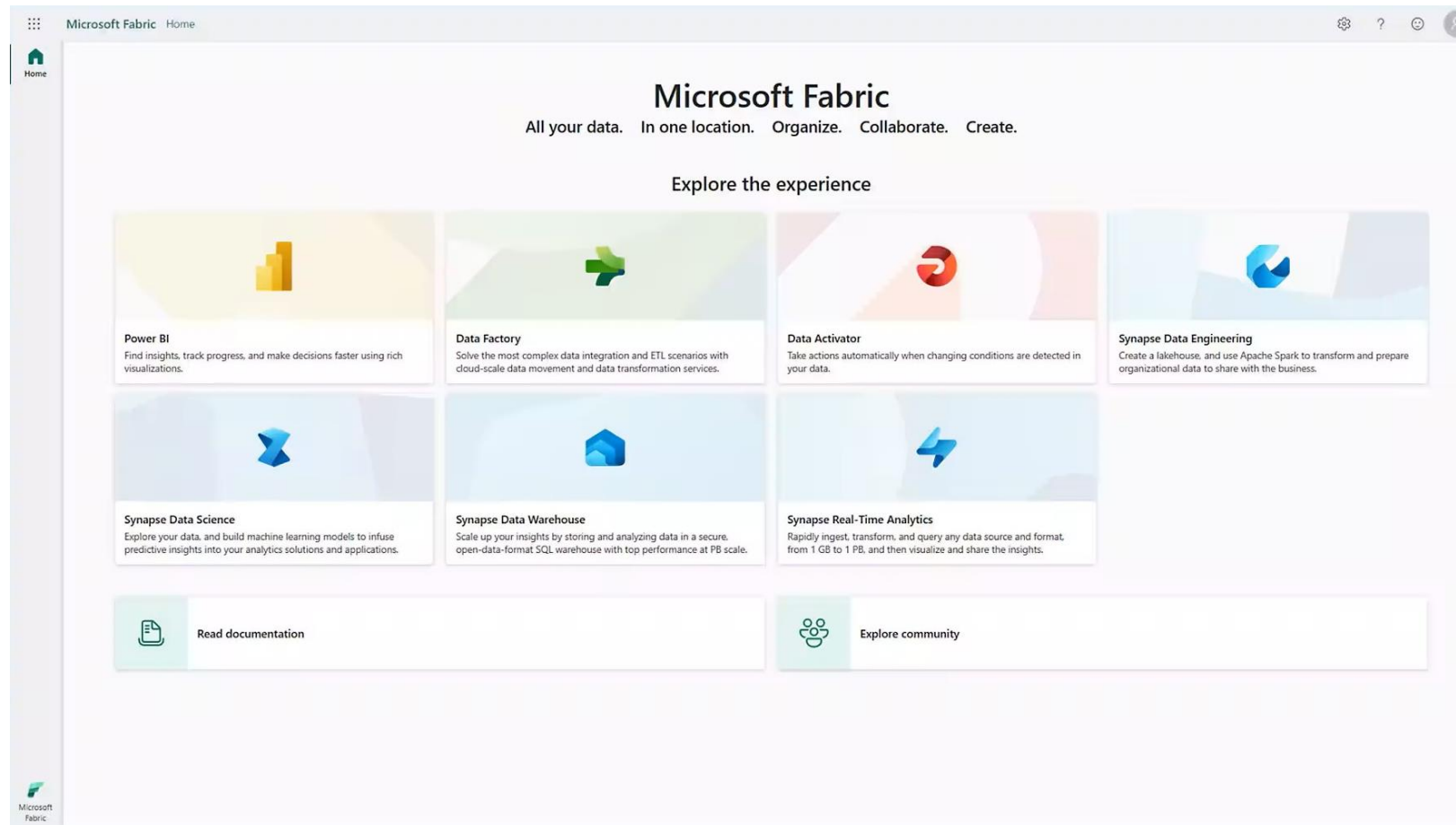
	ClientID	CompanyName	ContactName	ContactTitle	Address	City	Region	PostalCode	Country
1	ALFKI	Alfreds Futterkiste	Maria Anders	Sales Representative	Obere Str. 57	Berlin	NULL	12209	Germany
2	ANATR	Ana Trujillo Emparedados y helados	Ana Trujillo	Owner	Avda. de la Constitución 2222	México D.F.	NULL	05021	Mexico
3	ANTON	Antonio Moreno Taquería	Antonio Moreno	Owner	Mataderos 2312	México D.F.	NULL	05023	Mexico
4	AROUT	Around the Horn	Thomas Hardy	Sales Representative	120 Hanover Sq.	London	NULL	W11 1DP	UK
5	BERGS	Berglunds snabbköp	Christina Berglund	Order Administrator	Berguvsvägen 8	Luleå	NULL	S-950 22	Sweden
6	BLAUS	Blauer See Delikatessen	Hanna Moos	Sales Representative	Forsterstr. 57	Mannheim	NULL	68306	Germany
7	BLONP	Blondesdél père et fils	Frédérique Citeaux	Marketing Manager	24, place Kléber	Strasbourg	NULL	67000	France
8	BOLID	Bólido Comidas preparadas	Martin Sommer	Owner	C/ Araquil, 67	Madrid	NULL	28023	Spain
9	BONAP	Bon app'	Laurence Leblan	Owner	12, rue des Bouchers	Marseille	NULL	13008	France
10	BOTM	Bottom-Dollar Markets	Elizabeth Lincoln	Accounting Manager	23 Tsawassen Blvd.	Tsawassen	BC	T2F 8W4	Canada
11	BSBEV	B's Beverages	Victoria Ashworth	Sales Representative	Fauntleroy Circus	London	NULL	EC2 5MT	UK
12	CACTU	Cactus Comidas para llevar	Patricio Simpson	Sales Agent	Cerrito 333	Buenos Aires	NULL	1010	Argentina
13	CENTC	Centro comercial Moctezuma	Francisco Chang	Marketing Manager	Sierras de Granada 9993	México D.F.	NULL	05022	Mexico
14	CHOPS	Chop-suey Chinese	Yang Wang	Owner	Hauptstr. 29	Bern	NULL	3012	Switzerland
15	COMP	Comércio Mineiro	Pedro Afonso	Sales Associate	Av. dos Lusíadas, 23	Sao Paulo	SP	05432-043	Brazil
16	CONSH	Consolidated Holdings	Elizabeth Brown	Sales Representative	Berkeley Gardens 12 Brewery	London	NULL	W1X 6LT	UK
17	DRACD	Drachenblut Delikatessen	Sven Ottilie	Order Administrator	Walserweg 21	Aachen	NULL	52066	Germany
18	DUMON	Du monde entier	Janine Labrune	Owner	67, rue des Cinquante Otages	Nantes	NULL	44000	France
19	EASTC	Eastern Connection	Ann Devon	Sales Agent	35 King George	London	NULL	W1X 6FW	UK
20	ERNSH	Ernst Handel	Roland Mendel	Sales Manager	Kirchgasse 6	Graz	NULL	8010	Austria
21	FAMIA	Familia Arquibaldo	Aria Cruz	Marketing Assistant	Rua Orós, 92	Sao Paulo	SP	05442-030	Brazil
22	FISSA	FISSA Fabrica Inter. Salchichas S.A.	Diego Roel	Accounting Manager	C/ Moralzarzal, 86	Madrid	NULL	28034	Spain
23	FOLIG	Folies gourmandes	Martine Rancé	Assistant Sales Agent	184, chaussée de Tournai	Lille	NULL	59000	France
24	FOLKO	Folk och få Hb	Maria Larsson	Owner	Åkergatan 24	Bräcke	NULL	S-844 67	Sweden
25	FRANK	Frankenversand	Peter Franken	Marketing Manager	Berliner Platz 43	München	NULL	80805	Germany
26	FRANR	France restauration	Carine Schmitt	Marketing Manager	54, rue Royale	Nantes	NULL	44000	France
27	FRANS	Franchi S.p.A.	Paolo Accorti	Sales Representative	Via Monte Bianco 34	Torino	NULL	10100	Italy
28	FURIB	Furia Bacalhau e Frutos do Mar	Lino Rodriguez	Sales Manager	Jardim das rosas n. 32	Lisboa	NULL	1675	Portugal
29	GALED	Galería del gastrónomo	Eduardo Saavedra	Marketing Manager	Rambla de Cataluña, 23	Barcelona	NULL	08022	Spain

# Lab Software and Resources (cont.)

- Where can I learn more about databases and SQL:
  - Wikipedia:  
<https://en.wikipedia.org/wiki/Database>  
<https://en.wikipedia.org/wiki/SQL>
  - W3Schools | Khan Academy | ... (online, hands-on tutorial on SQL):  
<http://www.w3schools.com/sql/>  
<https://www.khanacademy.org/computing/computerprogramming/sql>  
<http://www.sqlcourse.com/>
  - Microsoft Virtual Academy  
<http://www.microsoftvirtualacademy.com/product-training/sql-server>
  - Online tool for 'quick and dirty' experiments  
<http://sqlfiddle.com/>

# Lab Software and Resources (cont.)

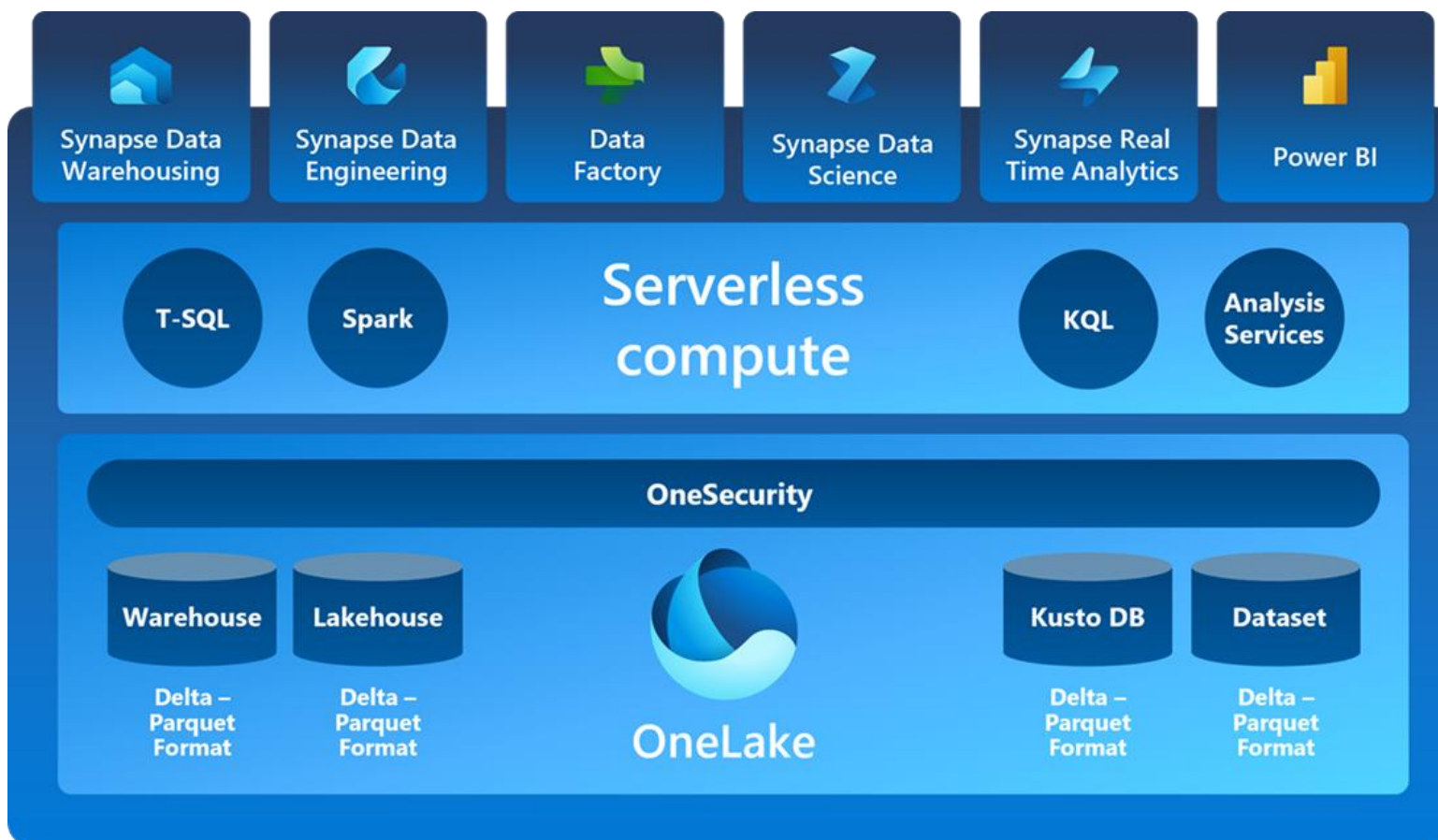
Cloud services used in Lab is (new) Microsoft offer:





# Lab Software and Resources (cont.)

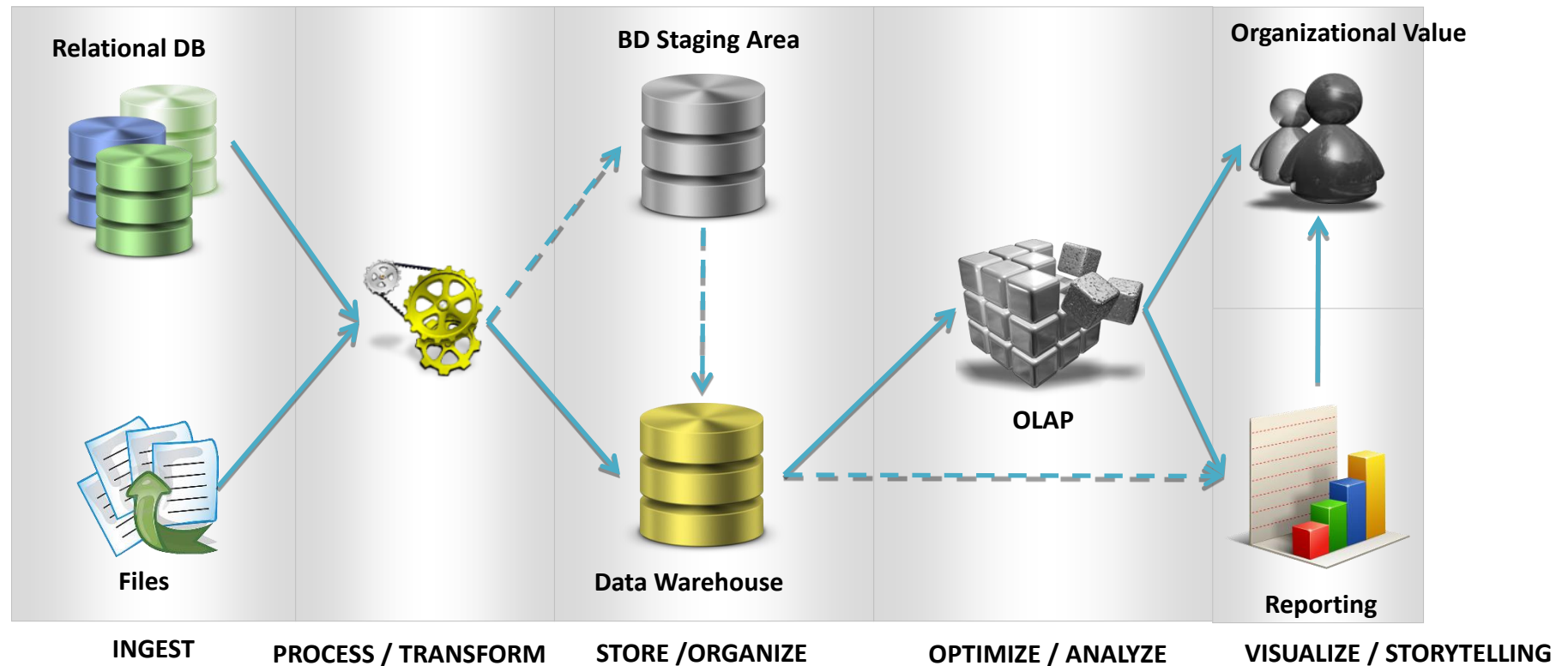
Complete + integrated Data Architecture (more details in later labs):



# BI I Project Overview

# BI I Project Overview

Main Goal of Project: **demonstrate you can build a BI solution!**



# BI I Project Overview (cont.)

- Remember: full ( definitive) details will be made available in the **“BI I Project (Handout) Guide”**, to be made available on date of 2<sup>nd</sup> Lab (10<sup>th</sup> of October), in Moodle.
- Some preliminary indications:
  - Group project (in real-life, BI projects are ALWAYS carried out by teams with several members)
  - Allowed Group size will be between ONE (1) and THREE (3) members
  - In special cases – if Project is based on original and real-life context, and demonstrably complex – Lab teacher may approve larger groups (4 or 5 members)
  - There will be a Moodle resource created to make Group selection easier, but the ONLY Group that officially counts is the list of students on Project Report cover.

# BI I Project Overview (cont.)

- Project Roadmap (in general terms):
  1. Global context and explaining / illustrating the “starting data problem”
  2. Translating and documenting the Business Problem into Business (data) Needs
  3. Data Discovery and Profiling of original (operational) data sources (typically, relational databases)
  4. Data Warehouse planning, design and development
  5. Staging Area and development of transformation (data cleansing) and processing routines – Staging Area ETL
  6. Data Warehouse ETL (loading DW with data from Staging Area)
  7. Conclusions and Lessons Learnt

# BI I Project Overview (cont.)

- Delivered in two (2) distinct but complementary parts:

**First Part :** **midnight 19<sup>th</sup> of November 2023**

*Only the Design and Implementation of Data Warehouse (using SQL script) but must submit initial part of the Report (so we understand what you are thinking and what you have done!)*

**Second Part :** **midnight 23rd of December 2023**

*Project Report (complete version), the Staging and DW databases and all of the ETL work*

# BI I Project Overview (cont.)

- Other things to be aware of (and may be made clearer when Project Guide is published; if needed, teachers will explain in Lab)
  - There will be penalizations if Project guidelines are NOT FOLLOWED (these will include, for example, Group sizes, late deliveries, wrong or missing components in the project delivery, etc.)
  - Further details – as already mentioned – will be detailed in Project Guide
  - Up to the point of planning and designing your Data Warehouse, your Lab teacher will be happy to advise on Project work; after that stage, your teacher is happy to clear or answer technical doubts BUT WILL NOT be able to make decisions (or do actual work) about your project on your behalf – you must be able to demonstrate ETL skills, when creating and loading your Data Warehouse.
  - Delivery methods (how you share your work with teachers) will be set at later date

# BI I Project Overview (cont.)

- On what Data Problem should I work? How to find data?
  - We always recommend that your Group works a real-life problem, if at all possible!
  - We have been recommending this action for many years, and it always seems to bring the best results (typically, Groups working on real-life contexts are more motivated, have a deeper understanding of work to be done and get better grades!)
  - If this is not possible, we will be pointing to online (internet) resources where you can get data that is close to real-life problems (typically, not as good but still quite solid choice – and you still get to choose what business domain to work on)
  - As (last) alternative, we hope to have – around 2<sup>nd</sup> or 3<sup>rd</sup> lab – collections of artificial datasets that represent synthetic “organizations” having a data problem that you can attempt to fix – quite similar to what you can get by searching through online resources, but curated and slightly refined/customized by the teachers



# Questions and doubts



**NOVA**

**IMS**

Information  
Management  
School

**Thank you!**