

Požadavky k zápočtu z předmětu KI/ODM a KI/KODM:

- 1) Vyzkoušet některý z nástrojů DWH (data warehouse), DL (data lake), DLH (data lakehouse)

https://en.wikipedia.org/wiki/Comparison_of_OLAP_servers

<https://www.gooddata.com/blog/data-warehouse-data-lake-and-analytics-lake-a-detailed-comparison/>

<https://www.sprinkldata.com/blogs/10-best-data-lake-tools-2023>

- 2) Nainstalovat si zvolený **DBMS** + analytický **nástroj OLAP**, nebo zajistit si přístup ke cloudové službě OLAP (trial verze zdarma).
- 3) Vybrat si vhodnou **datovou sadu** pro implementaci
- nejlépe s časovou dimenzí (nutné pro DWH)
 - je možné využít vzorovou databázi zvoleného DBMS
 - <https://www.kaggle.com/datasets>
 - <https://datasetsearch.research.google.com/>
 - a další ...
- 4) Ze získaných dat vytvořit **jednu datovou strukturu** (vhodnou k vizualizaci dat).
- DWH - zvolit si vhodnou strukturu datové kostky (min. 3 tabulky dimenzí + 1 tabulka faktů, struktura hvězdy - star, nebo sněhové vločky - snowflake)
 - DL – rozdělit data do zón (Raw zone, Cleansed zone, Transformed zone)
- 5) Vytvořit alespoň **4 různé dotazy nad datovou strukturou** a připravit si je s sebou k zápočtu.
- 6) Vyzkoušet alespoň **1 metodu data miningu** (klasifikace, shlukování, predikce, regrese, asociační pravidla, text mining) podle povahy získaných dat (pokud to použítý DBMS umožňuje). V případě, že zvolený databázový systém neumožňuje datamining, přidat alespoň 2 další dotazy využívající složitější statistické metody.

Kroky 2) až 6) popsat v dokumentu a tento popis vhodně doplnit obrázky, které ukazují jednotlivé kroky. Dokument pojmenovat **ODM25_PrijmeniJmeno.pdf** (v případě předmětu KI/KODM dokument pojmenovat **KODM25_PrijmeniJmeno.pdf**) a poslat na email kvetuse.sykorova@ujep.cz do **30. 6.** (nejpozději den před vaší prezentací).

Zápočet (prezenční / online) proběhne formou prezentace výsledků úkolů 1-6 v zápočtovém termínu (pochlubit se, co všechno se vám povedlo zrealizovat nad vámi vybraným OLAP systémem a nad vybranými daty) – délka cca 10-20 minut.

Nezapomeňte na konec vaší trial verze systému !!!

Součástí dokumentu musí být:

Úvodní stránka:

- název univerzity, fakulty, katedry,
- název práce např. "OLAP a DB2", "OLAP a lcCube" apod.,
- rok vytvoření práce,
- jméno a příjmení studenta,

- studijní obor a studijní program

Poslední stránka:

- seznam použitých zdrojů dle normy ISO 960 (citace.com a další)

Odkazy na elektronické zdroje v textu uvádět v poznámce pod čarou (jen https://...)

Příklady DWH, DL, DLH:

- MS SQL Server (Visual studio + SQL Server Management Studio),
- Olap Cube Writer,
- DB2 (trial),
- CubeJSServer,
- Mondrian,
- IcCube,
- Apache Hadoop + Map Reduce / Apache Spark / Apache Hive,
- Azure Data Lake,
- AWS Data Lake,
- Snowflake,
- Google Cloud Platform,
- Databricks Delta Lake,
- Starburst,
- a další ...

více např.:

[https://www.wikiwand.com/en/Comparison of OLAP servers](https://www.wikiwand.com/en/Comparison_of_OLAP_servers)

<https://docs.microsoft.com/en-us/sql/mdx/multidimensional-expressions-mdx-reference?view=sql-server-ver15>

<https://www.sqlservercentral.com/articles/mdx-multidimensional-expressions-in-148-minutes>

https://www.ibm.com/support/knowledgecenter/SSEPGG_10.5.0/com.ibm.dwe.navigate.doc/welcome_warehouse.html

<https://www.gooddata.com/blog/data-warehouse-data-lake-and-analytics-lake-a-detailed-comparison/>

<https://www.sprinkledata.com/blogs/10-best-data-lake-tools-2023>

https://en.wikipedia.org/wiki/Data_lake

a další ...