DTS Assignment 1 (40 points)

---

1. From your project description, determine the data that will be processed in your system. For this data, design separate data models for the MySQL database. The MySQL database needs to store only domain-related data (e.g. restaurants, items), which provide the context to interpret transactional data (e.g. orders). **(5 points)**
2. Create a new MySQL database with the necessary tables in the MS Azure platform or locally. Import data into tables and generate a CSV file (one or many of them) with aggregated data from all tables, which contains at least 100 rows (note: the CSV file generated for tables, which are not connected to any other tables, can contain less rows, but no less than 10 rows). **(5 points)**
3. Import data from MySQL to HDFS by using Sqoop. **5 points**
4. Implement a Map Reduce job that will perform an analysis on HDFS data by using data filtering and aggregation (**note:** the MR job example cannot be exactly the same as the one from the lab exercises). **(15 points)**

Instructions for submitting the assignment

Assignment 1 must be submitted in eŠtudij as a single ZIP document that **must** contain the following:

1. a CSV file (one or more of them) with prepared data obtained from the MySQL database,
2. a screenshot with clearly visible data in the MySQL database (e.g. a result of the SQL query where at least two tables are joined) and in HDFS,
3. a Word document, which contains:
    - the Java code for Mapper in Reducer classes and a screenshot with a clearly visible result of the executed MR job (use the HDFS *cat* command),

Assignment 1 will be graded based on the submitted documents with respect to the defined number of points for each sub-assignment. During the defense, the general knowledge and understanding of key concepts and used technologies and tools will also be evaluated with additional questions. The quality of the defenses will be graded with additional (max.) 10 points.