Hive

Hive Documentation:
https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML

Connect with beeline:

```
!connect jdbc:hive2://<host>:<port>/<base> <user> <pass>
!connect jdbc:hive2://<hadoop4>:10000 jpodeszwik jpodeszwik
!connect jdbc:hive2://<hadoop4>:10000/xyz jpodeszwik jpodeszwik
```

Create database:

```
create database xyz;
```

Change database:

```
use xyz;
```

Create table and load data:

```
CREATE TABLE transfers(
                src STRING,
                dst STRING,
                amount INT,
                date STRING
        )
        ROW FORMAT DELIMITED
        FIELDS TERMINATED BY ","
        STORED AS TEXTFILE;

LOAD DATA INPATH "/user/xyz/transfers" INTO TABLE transfers;
```

Create table in specific location (which might already exist):

```
CREATE TABLE transfers2(
                src STRING,
                dst STRING,
                amount INT,
                date STRING
        )
        ROW FORMAT DELIMITED
        FIELDS TERMINATED BY ","
        STORED AS TEXTFILE
        LOCATION "/user/xyz/transfers_table";
```

Create external table:

```
CREATE EXTERNAL TABLE transfers3 <rest_of_create_command>;
```

Change delimiter:

```
create table transfers4
row format delimited
fields terminated by ";"
as select * from transfers;
```

Change data format:

```
create table transfers5
stored as orc
as select * from transfers;
```

Enable compression:

```
set hive.exec.compress.output=true;
set mapreduce.output.fileoutputformat.compress=true;
set mapreduce.output.fileoutputformat.compress.codec=org.apache.hadoop.io.compress.GzipCodec;
set mapreduce.output.fileoutputformat.compress.type=BLOCK;
```

Enable dynamic partitioning:

```
SET hive.exec.dynamic.partition=true;
```

Refresh data about partitions:

```
MSCK REPAIR TABLE <tabela>;
```

```
CREATE TABLE transfers5(
                src STRING,
                dst STRING,
                amount INT,
                date STRING
        )
        partitioned by (log_time string)
        ROW FORMAT DELIMITED
        FIELDS TERMINATED BY ","
        STORED AS TEXTFILE
        LOCATION "/user/xyz/events";
```

Print information about table:

```
show create table transfers;
```

Tasks

1. Create database <login> and use it in next tasks.
2. Create table transfers. Load transfer data into this table.
3. Create ‘external’ table and load the same data into it.
4. Drop both tables and see what happens to data on hdfs.
5. Create table in ORC format. Check if the data is binary.
6. Sum field ’amount’ by source field.
7. Create table ’owners’ from data loaded with sqoop.
8. Create table ’named_transfers’ by joining ’owners’ and ’transfers’. Replace src and dst fields with names.
9. Create table partitiones by source account id.