00:00 - xampp was installed first.

00:03 - Created a database named <reddit> in phpmyadmin of xampp to store data.

00:17 - Created the necessary tables for the created database <reddit>.

The table was created by executing the sql that was already used for work and backed up.

The name of the loaded file is reddit.empty.sql.

This sql creates an empty tables with no data records.

00:29 - Displayed the result of executing Sql.

00:32 - Checked the created tables.

There are no records in the tables.

00:48 - Copy the script source for the testing to the xampp/htdocs folder.

Check the script source.

Setting up a connecting informations to connect to mysql in the source.

$servername - is the IP address or hostname of the computer where mysql is installed.

$username - is the user name to connect to mysql.

$password - is the password to connect to mysql.

$dbname - is the name of the database to be used for the test.

This value,$dbname, is the name of the database created in phpmyadmin.

In my case, the datas are as follows.

$servername = “localhost”;

$username = “root”;

$password = “”;

$dbname = “reddit”;

01:33 - Open the browser and enter the url for testing to run.

In my case, the url was as following.

https://localhost/database-bz2.php

01:50 – A form to select the bz2/json data file to be imported is displayed.

The bz2 file to be loaded is RC\_2001-10.bz2 from the reddit server.

In my case, the smallest of the three data files retrieved from the reddit server was loaded in order to reduce the execution time and convenient for creating a video.

01:57 – When I upload a file by clicking the upload button, the script starts analyzing the file and importing the data.

02:03 - RC\_2001-10.bz2 was successfully uploaded and the situation of script execution was printed on the page of the browser while json analysis was performed.

First, the name, capacity, and experiment start time in utc of the imported bz2 file are displayed.

02:08 - In the script, whenever 20000 rows are imported into the comments table, the current time in utc format is output.

02:39 - The number of records imported into the comments table is 150429, and the total execution time is

1639574373-1639574343 = 30 seconds.

02:39 - After the data import for the comments table is finished, the data for the links table and the subreddits table are imported.

The count of records loaded into the links table is 24370, and the count of records in subreddits is 32.

The time to import data into the subreddits table is 2 seconds.

The time to import data into the links table is 5 seconds.

Data importing time by script varies slightly depending on the specification of the computer and the number of executions, but a conclusion could be obtained.

Although it was not written in video, the importing time was measured while adding or removing constraints to the table. The conclusion on this is as explained in the submitted pdf.

*When dealing with small record sets, the difference was very small. However, as the count of records increases, the time difference between importing becomes larger.  
Thus, we can find that importing from a constrained schema takes more time than an unconstrained schema.  
The reason is that each time a row is inserted into a constrained schema, the number of checks to see if the inserted row violates the constraint increases.  
When a NOT NULL constraint is added, other checks are also performed, such as when the value is not null.  
This is a more important consideration as more and more additional time is required as more rows are added.*