

Introducing the Dataset

The dataset is from a hypothetical coding platform. Users come and submit their solutions to existing problems on the platform and improve their programming skills.

The datasets has three tables - **Users, Problems, Submissions**

Unzip tables :-

- Unzip dataset in Itversity -
 - **unzip Hive/Joins_in_Hive/Coding_platform.zip -d Hive/Joins_in_Hive**

Submission table :-

- Create submission table -
 - **create table submissions(user_id string, problem_id string, attempts_range int)**
row format delimited
fields terminated by ' ';
TBLPROPERTIES ("serialization.null.format"="");
- Put file to HDFS -
 - **hdfs dfs -put**
/home/itv180149/Hive/Joins_in_Hive/Coding_platform/submissions.txt
/user/itv180149
- Load data into Hive table -
 - **load data inpath "/user/itv180149/submissions.txt" into table**
submissions;

Users table :-

- Creating users table -
 - **create table users(user_id string, country string, rating int)**
row format delimited
fields terminated by ' '
TBLPROPERTIES ('serialization.null.format'='');
- Put file to HDFS -
 - **hdfs dfs -put**
/home/itv180149/Hive/Joins_in_Hive/Coding_platform/users.txt
/user/itv180149
- Load data into Hive table -
 - **load data inpath "/user/itv180149/users.txt" into table users;**

Problems table :-

- Creating problems table -
 - **create table problems(problem_id string, level_type string, points int,**
tags array<string>)
row format delimited
fields terminated by ' '
collection items terminated by '\$'
TBLPROPERTIES ('serialization.null.format'='');

- Put file to HDFS -
 - **hdfs dfs -put**
/home/itv180149/Hive/Joins_in_Hive/Coding_platform/problems.txt
/user/itv180149
- Load data into Hive table -
 - **load data inpath "/user/itv180149/problems.txt" into table problems;**