Spark and Scala

ANALYSIS OF YOUTUBE DATA

Description:

This dataset consists of the information of videos being uploaded, viewed, subscribed etc. Get the dataset attached with the document. (Text file)

Parameters:

- Column 1: Video id of 11 characters.

- Column 2: Uploader of the video.

- Column 3: Interval between day of establishment of YouTube and the date of uploading of the video.

- Column 4: Category of the video.

- Column 5: Length of the video.

- Column 6: Number of views for the video.

- Column 7: Rating on the video.

- Column 8: Number of ratings given for the video

- Column 9: Number of comments on the videos.

- Column 10: Related video ids with the uploaded video.

Problem and Solutions:

1. Find out the 5 User Id who has uploaded maximum number of Videos and highest number of viewer with video-Id.

2. Find the top 20 videos, for which, maximum peoples has subscribed and maximum number of ratting has been given in YouTube.

3. Find out top five video categories which uploaded maximum number of videos in YouTube.

4. Find out those videos which has great rating, max subscription, viewer and comments.

5. Count the number of videos whose size is lesser that 1MB(1024kb)

6. Find out the video ids categorized as "Comedy" and "Film & Animation"

- Highest number of viewer with video-id.
- User-id (uploader name) who has uploaded max number of videos **
- Top 20 video ids of maximum subscription of videos
- Maximum ratings of videos with video-id, category, uploader etc.
- Number of videos, sized lesser than 1 MB.
- Video-ids categorized as "Comedy" & "Film & Animation"