

## Exercise 5.3: Implementing Aggregation In Your Data Model

In this exercise, you will:

- Design a logical data model that will be able to store and retrieve aggregated values

### Background

During a peer review of an intern's work, your team found that the intern created a schema with the assumption that the queries could use aggregate functions.

It's back to the drawing board, but this time with you working with the hapless intern to come up with a proper design.

- Requirement #1: KillrVideo must be able to count the number of views for each video in a particular month and year. The count does not have to be 100% accurate.
- Requirement #2: KillrVideo must be able to keep track of the total number of uploaded videos, the combined duration (in seconds) of all of these videos, and the average duration (in seconds) of an uploaded video. These statistics must be stored per day.

They do not have to be 100% accurate.

A logical table schema for the Video entity has already been created and looks like:

videos	
video_id	K
uploaded_timestamp	
title	
description	
type	
release_date	
{tags}	
<preview_thumbnails>	
{genres}	

## Steps

1. Review the above requirement #1.
2. The intern made a table to log an entry into whenever a user views a video. The original query would have then used the COUNT aggregate to retrieve the number of views for a particular video.
  - As far as Cassandra is concerned, what is inefficient about the intern's design?
3. Come up with an optimal table schema that meets this requirement using Cassandra.
  - What would be the query that can retrieve the daily count of the number of views for a video for a particular year and month?
  - Is it possible to retrieve the all-time number of views for a video? If possible, what would be the query?
  - Is there anything that needs to be done from the application side?
  - How would your design increment a video's view count? Provide the queries and/or write statements needed to do this.
4. The intern, being ever helpful, suggested that it would be useful to also display the top 10 videos for each month, based on the number of views.
  - Would it be possible to query this information in Cassandra? Why or why not?
  - If not, is there a way to do this outside of Cassandra using your schema?
5. Review the above requirement #2.
6. Modify the existing Videos table and/or come up with your own table schema that will meet these requirements.
  - What is the query to retrieve the total number of videos uploaded on a specific day?
  - What is the query to retrieve the combined duration of all videos uploaded on a specific day?
  - What is the query to get the average duration of all videos uploaded on a specific day?
  - Explain how these values are updated whenever a new video is uploaded.