## **BDA MINI-PROJECT**

# **Youtube Data Analysis**

**Guide: Prof. Vasundhara Ghate** 

# **GROUP 3**

**PA 10 ATHARV NIKAM** 

PA 17 KETAKI PATIL

**PA 27 SAISRIJAN GUPTA** 

**PA 31 ANJALI NAIR** 

**PA 51 SHIVRAJ PATIL** 



# **AGENDA**

- Problem statement
- Objectives
- Dataset
- Tools Used(Brief Explanation)
- System Architecture
- Output
- Visualization Screenshots (PowerBi)
- Conclusion



#### PROBLEM STATEMENT

"YouTube has over a billion users and every day people watch hundreds of millions of hours on YouTube and generate billions of views".

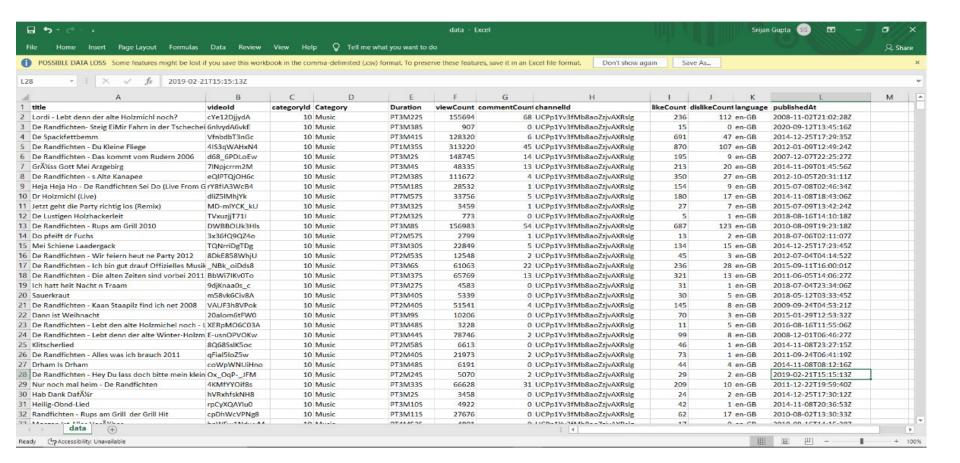
Every day, people across the world are uploading 1.2 million videos to YouTube, or over 100 hours per minute and this number is ever increasing. To analyze and understand the activity occurring on such a massive scale, a relational SQL database is not enough. Such kind of data is well suited to a massively parallel and distributed system like Hadoop.



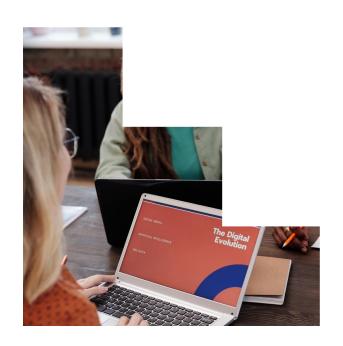
#### **OBJECTIVES**

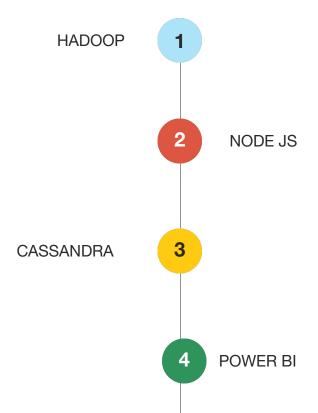
The main objective of this project is to focus on how data generated from YouTube can be mined and utilized by different companies to make targeted, real time and informed decisions about their product that can increase their market share.

#### DATASET

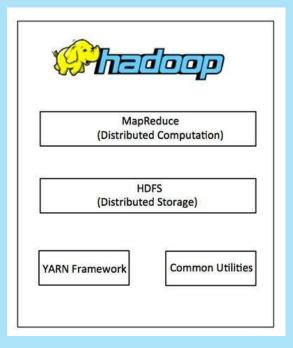


# **Tools Used**





### **HADOOP**



#### INTRODUCTION

- Hadoop is an open-source framework that allows to store and process big data in a distributed environment across clusters of computers using simple programming models.
- It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.

#### **FEATURES**

#### Open Source:

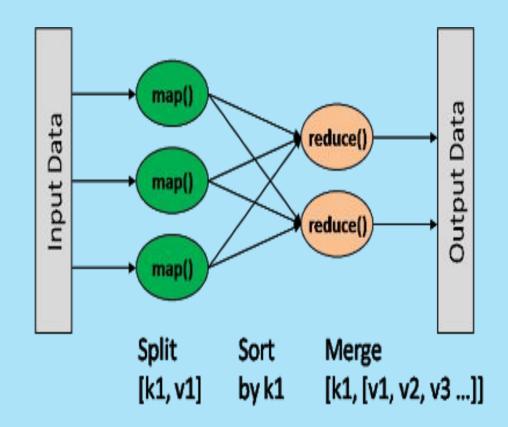
It is an open-source project the source-code is available online for anyone to understand it or make some modifications as per their industry requirement.

#### Highly Scalable Cluster:

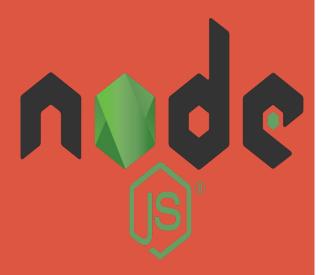
A large amount of data is divided into multiple inexpensive machines in a cluster which is processed in parallel.

#### **MAP REDUCE:**

- MapReduce is a processing technique and a program model for distributed computing based on java.
- The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs).
- Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples.



# NODE JS



#### INTRODUCTION

- Node.js is a server-side platform built on Google Chrome's JavaScript Engine (V8 Engine).
- Node.js is an open source, cross-platform runtime environment for developing server-side and networking applications.

#### **FEATURES**

- Asynchronous and Event Driven All APIs of Node.js library are asynchronous, that is, non-blocking.
- Very Fast Being built on Google Chrome's V8 JavaScript Engine,
   Node.js library is very fast in code execution.
- No Buffering Node.js applications never buffer any data. These applications simply output the data in chunks.

## **CASSANDRA**



#### **INTRODUCTION**

- Apache Cassandra is a highly scalable, high-performance distributed database
- It is a type of NoSQL database.

#### **FEATURES**

#### Highly Fault Tolerant:

If any node in the cluster fails, other nodes will take over and complete the work.

#### Easy Data Distribution:

Data distribution in Cassandra is very easy because it provides the flexibility to distribute data where you need by replicating data across multiple data centers.

#### • Transaction Support:

Cassandra supports properties like Atomicity, Consistency, Isolation, and Durability (ACID).

# Power BI



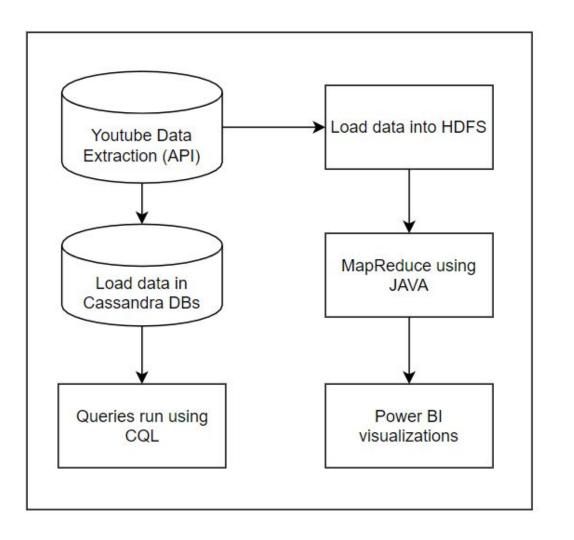
#### INTRODUCTION

 Unify data from many sources to create interactive, immersive dashboards and reports that provide actionable insights and drive business results.

#### **KEY FEATURES**

- Connect to data
- Transform and model the data
- Create charts and graphs
- Create reports and dashboards that are collections of visuals
- Share reports with others using the Power BI service





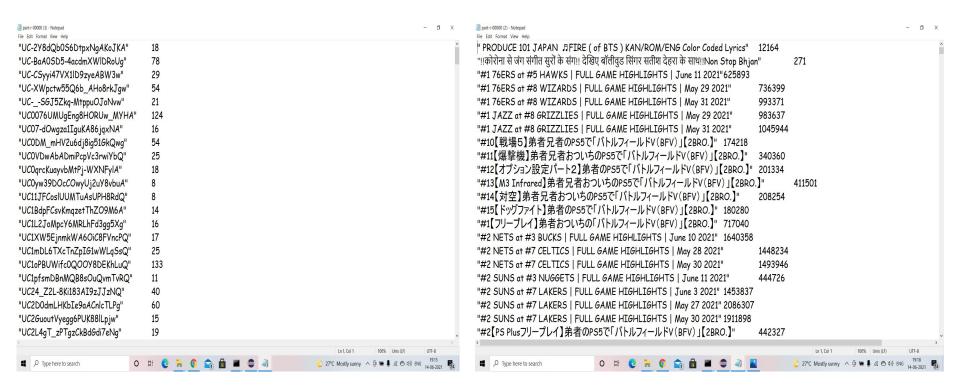
# SYSTEM ARCHITECTURE

# **OUTPUT - MapReduce**

#### 1. CATEGORY

```
part-r-00000 - Notepad
File Edit Format View Help
"Autos & Vehicles"
                          1923
"Comedy"
                 318
"Education"
                 1885
"Entertainment" 3592
"Film & Animation"
                          4025
"Gaming"
                 6596
"Howto & Style" 729
"Music" 8705
"News & Politics"
                          2741
"Nonprofits & Activism" 1096
"People & Blogs"
                          2856
"Pets & Animals"
                          30
"Science & Technology"
                          2121
"Sports"
                 1966
"Travel & Events"
                          301
```

### 2. UPLOADERS 3. VIEWS



## **OUTPUT - CASSANDRA**

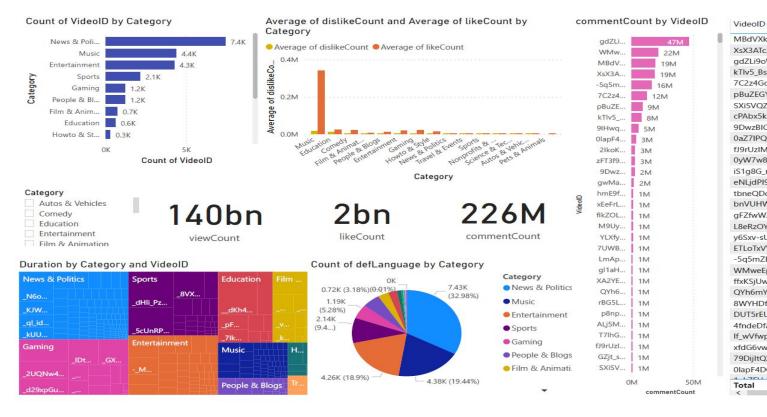
#### TABLE CREATION AND IMPORT OF DATASET

```
Administrator: Command Prompt - cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.4 | CQL spec 3.4.4 | Native protocol v4]
Ise HELP for help.
WARNING: pyreadline dependency missing. Install to enable tab completion.
calsh> use trial
cqlsh:trial> CREATE TABLE trial.trial name (
        ... title text,
        ... videoId text PRIMARY KEY,
        ... categoryId int,
        ... category text,
        ... duration text,
        ... viewCount bigint,
        ... commentCount bigint
        ..., channelId text,
        ... likeCount bigint,
        ... dislikeCount bigint,
        ... language text,
        ... publishedAt text);
cqlsh:trial> COPY trial (title,videoId,categoryId,category.duration,viewCount,commentCount,channelId,likeCount,dislikeCount,language,publishedAt) FROM 'C:\Users\gupta\Downloads\data.csv' WITH HEADER=true;
Column family 'trial' not found
qlsh:trial> COPY trial.trial_name (title,videoId,categoryId,category.duration,viewCount,commentCount,channelId,likeCount,dislikeCount,language,publishedAt) FROM 'C:\Users\gupta\Downloads\data.csv' WITH HEADER=t
Jsing 7 child processes
```

#### **DATASET IN TABULAR FORMAT**

```
Administrator: Command Prompt - cqlsh
ttributeError: 'NoneType' object has no attribute 'add_timer'
  ttributeError: 'NoneType' object has no attribute 'add_timer'
    cls. loop.add timer(timer)
AttributeError: 'NoneType' object has no attribute 'add_timer'
Processed: 22527 rows; Rate: 3551 rows/s; Avg. rate: 3511 rows/s
22527 rows imported from 1 files in 6.417 seconds (0 skipped).
colsh:trial> select * from trial.trial name:
 videoid
                        category
                                                                   | categorvid | channelid
                                                                                                                                             | commentcount | dislikecount | duration | language | likecount | publishedat
                                                                                     25 | UC4oOAvXFuNlKivYiUynOKTA |
                                                                                                                                                                                              16 |
                                                                                                                                                                                                             PT3M52S | es-419 |
                                                                                                                                                                                                                                                                  73 | 2021-05-12T20:45:15Z |
                  Duque viajar i nuevamente a Cali para reunirse con los j venes | Semana Noticias
  WkQBtOtdwU4
                                    News & Politics |
                                                                                     25 | UChDKvK059fYz3J02f10Z6sg |
                                                                                                                                                                                                2 |
                                                                                                                                                                                                                  PT47S |
                                                                                                                                                                                                                                        null |
                                                                                                                                                                                                                                                                214 | 2021-05-25T16:00:12Z |
                                           High School Runner With Cancer Crosses Finish Line With Teammates |
  ujP9ZXGLuQM
                                                                                     10 | UCgcsRIFVt2xIF180uKd0T4A |
                                                                                                                                                                                                              PT3M315 |
                                                                                                                                                                                                                                       null |
                                                                                                                                                                                                                                                            22061 | 2017-11-04T03:00:027 |
                                                                                                                                                                                             848
                                                                                                                                     Chanol - Para Vos
                                                                                                                                                                             2724573
h75JpMHJnJ8
                                                      Music |
                                                                                     10 | UCYXIviXPAaaaU_AOotpXTAw |
                                                                                                                                                                    0
                                                                                                                                                                                              23 I
                                                                                                                                                                                                              PT3M16S |
                                                                                                                                                                                                                                       null |
                                                                                                                                                                                                                                                             1196 | 2021-06-04704:06:572 |
                                                                                                                                                  Still Hood |
                                                                                                                                                                                  78129
                                                                                     25 | UC5wKpLWxAZBZrun1s3mzwEw |
                                                                                                                                                                                                                                                                 41 | 2021-05-27T02:10:58Z
 XDgOR949Zn0
                                    News & Politics |
                                                                                                                                                                                                             PT1M28S
                                                                                                                                                                                                                                        null |
                                                       مراتوم بهر من المراتوم والمراتوم والم
  αງ ຍິαງ ≣ິαງ Öα∮ éαງ ãαງ ∘αງ √αງ ö
  eYgg6Fhbem0
                                    News & Politics |
                                                                                     25 UCwqusr8YDwM-3mEYTDe3Hzw
                                                                                                                                                                   44
                                                                                                                                                                                           null |
                                                                                                                                                                                                                 PT48S
                                                                                                                                                                                                                                                             null | 2021-06-12T03:13:14Z |
                                                                                                                                                                                                                                                                                                                                                                                                 Tela
ngana: Congress Stages Bizzare Protest Over Fuel Prices Throws Bike In Hussain Sagar Lake |
  -6WpTiZvsio
                                    News & Politics |
                                                                                     25 UCeJWZgSMlzqYEDytDnvzHnw |
                                                                                                                                                                    0 |
                                                                                                                                                                                           null |
                                                                                                                                                                                                              PT2M26S
                                                                                                                                                                                                                                                             null | 2021-05-26T09:23:147 |
                                                                                                                                                                                                                                                                                                                                                                                                  Cvc
 one Tauktae : Heavy losses to Lemon and Custard apple farmers of Virpur Rajkot | TV9News |
 WMweEpGlu U
                                                                                     10 | UC9vrvNSL3xcWGSkV86REBSg |
                                                                                                                                                                                                               PT3M3S |
                                                                                                                                                                                                                                           ko | 17077364 | 2021-05-21T03:46:13Z |
                                                       Music
                                                                                                                                                         7228764
                                                                                                                                                                                       230291
                                                                                                   BTS (δω-φâä∞åîδàäδï¿) Butter Official MV | 341473822
                                                                                     24 | UCtBu8Wb2BUoduUXJS9Uss70 |
                                                                                                                                                                                                                                                                838 | 2021-05-30T15:00:19Z |
  aHaA5vpvv3g
                                        Entertainment
                                                                                                                                                                                              31 | PT16M16S |
                                                                                                                                                                                                                                        null |
                                                                                                     α ζαη Καη ύση ς αη Ναη Εση θαη θαη Ν ΕΡ. 40 1/4 | αη όση μαη θαη έση θαη έση μαη ς 8 |
 AiJdGU17rUE
                                        Entertainment |
                                                                                     24 | UC7vRgYRMzpudUE0JIwhaoPg |
                                                                                                                                                                    0 |
                                                                                                                                                                                                0 | PT38M225 |
                                                                                                                                                                                                                                       null |
                                                                                                                                                                                                                                                                   1 | 2021-02-11T13:58:52Z |
                            KHTOnline 11/02/2021 Approaching Sound Design for Plays with Nicola Chang |
  qJWS3faAlgk
                                                       Music
                                                                                     10 UC9vrvNSL3xcWGSkV86REBSg
                                                                                                                                                               1748
                                                                                                                                                                                             358
                                                                                                                                                                                                            PT4M18S
                                                                                                                                                                                                                                                            81318 | 2017-11-30T23:10:26Z |
                                          BTS (δω-φâāωåîδàäδī¿) ΓÇô HIP HOP LOVER (Color Coded Han Rom Eng Lyrics)
                                                                                                                                                                                          2492651
  nYeONwvvrrg |
                                                     Gaming
                                                                                     20 UCIdFoP9qN1ROZ6jVLFx9vhw |
                                                                                                                                                               8081
                                                                                                                                                                                            159
                                                                                                                                                                                                          PT12M17S |
                                                                                                                                                                                                                                        null |
                                                                                                                                                                                                                                                            13858 | 2021-03-16T08:15:17Z |
                             THEM PETLER DE-R- RT- !? ADOPT ME PLAYFULL PET UPDATE !! ( Roblox Adopt Me )
                                                                                     25 | UCwgusr8YDwM-3mEYTDeJHzw |
                                                                                                                                                                                           null |
                                                                                                                                                                                                                                                             null | 2021-06-12T08:35:43Z |
  AJnM100prvI
                                    News & Politics |
                                                                                                                                                                   37 I
                                                                                                                                                                                                          PT14M295
                                                                                                                                                                                                                                           en
       BJPs Sambit Patra Slams Digvijaya Singh On Clubhouse Row; Calls It A Part Of Toolkit
                                                                                                                                                                                    1140
                                                                                    25 | UC5wKpLWxAZBZrun1s3mzwEw |
  OaccGzs1NaU
                                    News & Politics |
                                                                                                                                                                                               31 |
                                                                                                                                                                                                         PT14M47S
                                                                                                                                                                                                                                        null |
                                                                                                                                                                                                                                                                512 | 2021-05-14T00:16:42Z
                                             a Carúar 1 a fear iar ca Carúar 1 a fear iar xar a far iar xar a far iar xar a far iar xar a far iar a far iar xar a far xar a far iar xar a far xar a far iar xar a far iar xar a far iar xar a far iar xar a far xar a far iar xar a far x
 ὄαν έας ίαν Cας ὄας μας ύαν ῖας ὄας Ε
 ZnOCQ7k-ZgI
                                                                                     10 UCzCedBCSSltI1TFd3bKyN6g
                                                                                                                                                              21657
                                                                                                                                                                                           1025 |
                                                                                                                                                                                                                 PT275 |
                                                                                                                                                                                                                                                           309110 | 2021-06-05T15:00:26Z |
                                                       Music
                                                            EXO 엑소 Dont fight the feeling Character Clip #CHANYEOL
bL3W4Zramtw |
                                                                                     10 | UCdopecNiaAWuw1gabUMSVVA |
                                                                                                                                                                                                              PT1M59S |
                                                                                                                                                                                                                                       null |
                                                                                                                                                                                                                                                                263 | 2021-05-29T18:12:56Z |
                                                      Music
                                                                                                           Dale O Va - Confesi n (En Vivo)
                                                                                                                                                                                   11260
 FEo4_S6pWaQ
                                                                                     25 | UCaXkIU10idiPwiAYu6GcHig |
                                                                                                                                                                                                              PT5M35S |
                                                                                                                                                                                                                                        null |
                                                                                                                                                                                                                                                                273 | 2021-06-05T20:34:09Z |
                                    News & Politics |
                                                                                                                                                                                               18
                                                                              Who Got Next? One-on-One with Dr. Jason Johnson
                                                                                                                                                                                    9817
  717msh3xWbM
                                     News & Politics
                                                                                     25 | UC4oOAvXFuNlKivYjUynOKTA |
                                                                                                                                                                                                              PT45M6S
                                                                                                                                                                                                                                   es-419 |
                                                                                                                                                                                                                                                                 56 | 2021-05-30T23:52:27Z |
      Juanes presenta su nuevo lílbum y se refiere a la crisis en Colombia | Semana Noticias
```

# VISUALIZATION



VideoID	viewCount	1
MBdVXkSdhwU	3862719132	
XsX3ATc3FbA	3789931877	
gdZLi9oWNZg	3359005674	
kTlv5_Bs8aw	2881045169	
7C2z4GqqS5E	2839058268	
pBuZEGYXA6E	2814994137	
SXiSVQZLje8	1910461807	
cPAbx5kgCJo	1826818238	
9DwzBICPhdM	1452662446	
0aZ7IPQ5EXs	1386619757	
fJ9rUzIMcZQ	1336263107	
0yW7w8F2TVA	1293916851	
iS1g8G_njx8	1290257049	
eNLjdPl9zdE	1231721221	
tbneQDc2H3I	1201792742	
bnVUHWCynig	1167034373	
gFZfwWZV074	1165243768	
L8eRzOYhLuw	1101248717	
y6Sxv-sUYtM	1098794763	
ETLoTxVVvjM	1083755184	
-5q5mZbe3V8	1045112971	
WMweEpGlu_U	1024442146	
ffxKSjUwKdU	1018944870	
QYh6mYIJG2Y	1005482052	
8WYHDfJDPDc	983746455	
DUT5rEU6pqM	943716007	
4fndeDfaWCg	938561686	
If_wVfwpfp8	934791272	
xfdG6vwlGGU	930510733	
79DijItQXMM	876150401	
0lapF4DQPKQ	875933541	
4 1.7EV V 1.	070245050	

140322844296

<

vious ount

# Thank you!

**GROUP 3**