



# TEDxComments



DWH - AWS Glue



Next



# Table of contents



01 Job PySpark

02 Data & Schema

03 Potential Flaws

04 Future Development



Back



Next

# Job pyspark



```
## READ WATCH NEXT DATASET
wn_dataset_path = "s3://fc-tedx-2024-data/csv/related_videos.csv"
wn_dataset = spark.read.option("header", "true").csv(wn_dataset_path)

wn_real_id = wn_dataset.select(col("internalId").alias("internal"), col("id").alias("real_id")).distinct()

wn_dataset = wn_dataset.join(wn_real_id, wn_dataset.related_id==wn_real_id.internal, "left")

# CREATE THE AGGREGATE MODEL, ADD TAGS TO TEDX_DATASET
wn_dataset_agg = \
    wn_dataset.groupBy(col("id").alias("id_ref")).agg(collect_list("real_id").alias("related_videos"))

tedx_dataset_agg_img_wn = tedx_dataset_agg_img.join(wn_dataset_agg, tedx_dataset_agg_img._id == \
    wn_dataset_agg.id_ref, "left").drop("id_ref")

wn_dataset_views = wn_dataset.select(col("real_id"), col("viewedCount").alias("views")).distinct()

tedx_dataset_agg_img_wn_views = tedx_dataset_agg_img_wn.join(wn_dataset_views, \
    tedx_dataset_agg_img_wn._id == wn_dataset_views.real_id, "left").drop("real_id")
```

Back



Next

# Schema



```
_id: "102008"
slug: "chris_kluwe_how_augmented_reality_will_change_sports_and_build_empathy"
speakers: "Chris Kluwe"
title: "How augmented reality will change sports ... and build empathy"
url: "https://www.ted.com/talks/chris_kluwe_how_augmented_reality_will_chang..."
description: "Chris Kluwe wants to look into the future of sports and think about ho..."
duration: "537"
publishedAt: "2014-05-22T15:00:59Z"
tags: Array (3)
  0: "technology"
  1: "sports"
  2: "virtual reality"
img_url: "https://pe.tedcdn.com/images/ted/9b43acda94d7e3743aa877ca5b492862b860f..."
related_videos: Array (6)
  0: "33592"
  1: "60796"
  2: "77920"
  3: "16891"
  4: "79546"
  5: "117240"
views: "1361406"
```



Back



Next

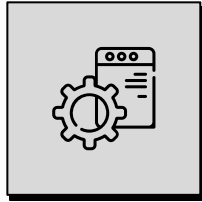


# Potential Flaws



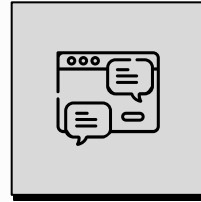
01

02



## Aggiornamento dei Dati

I dati delle visualizzazioni  
devono essere  
costantemente aggiornati per  
essere up-to-date.



## Aggregazione dei Dati

Per evitare ridondanza nei  
dati è stato necessario  
aggiungere una chiave  
referenziale

Back

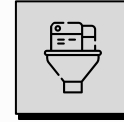


Next





# Future development



Inserimento di commenti	Valutazione dei commenti	Tagging in base al topic
Le persone potranno inserire commenti ad un video	I commenti verranno valutati dagli utenti così da far emergere i più inerenti	I commenti dovranno essere categorizzati in base alla loro tipologia ed argomento



Back



Next



# Thanks!



Colpani Filippo - 1078874  
Foglieni Luca - 1081399



**Credits:** This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Back

