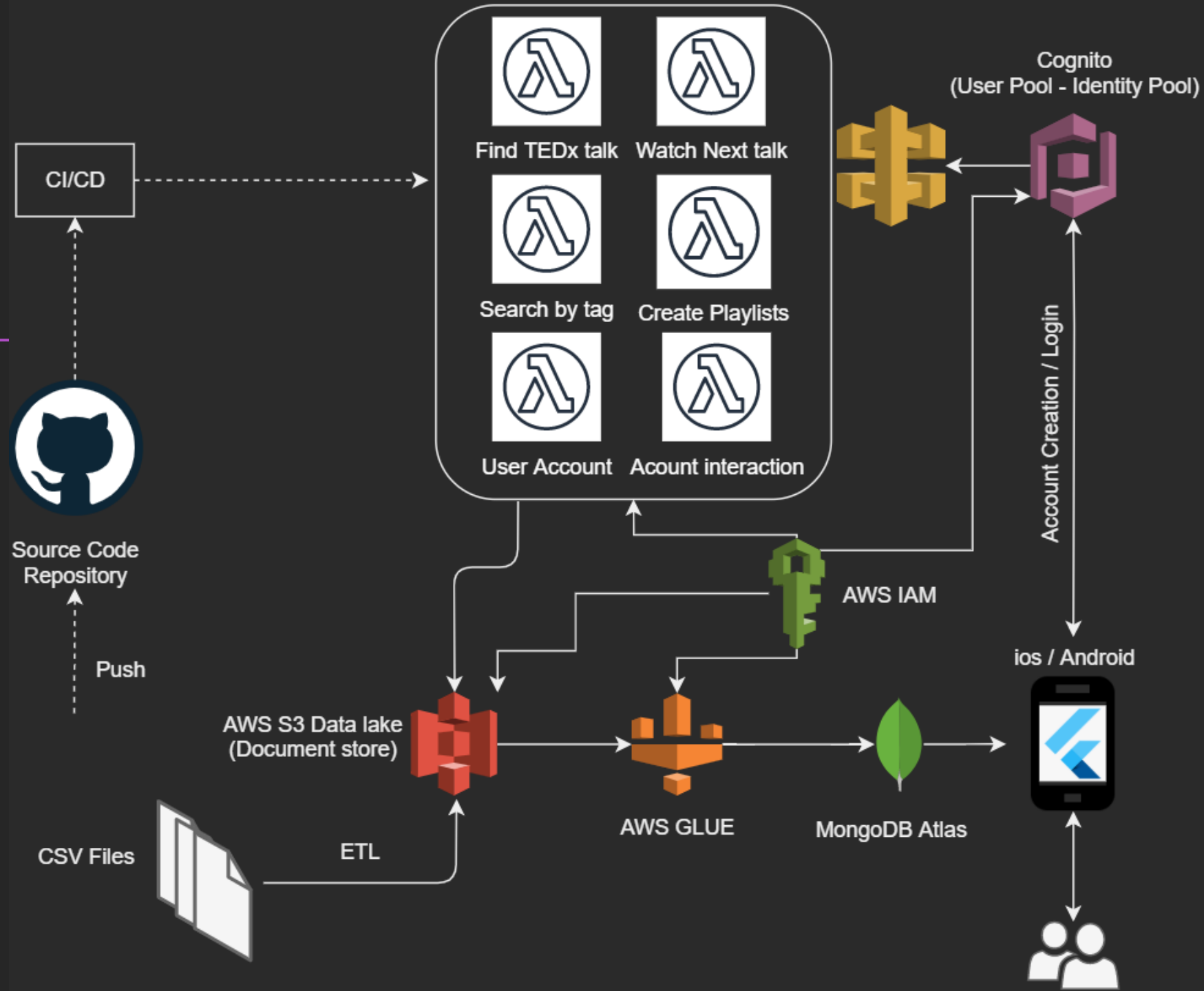


MyTEDx



SCENARIO



RELEASE PLAN

0.1

Define requirements

Design system

+ Aggiungi un'altra scheda

0.2

Build the development environment

test the environment

+ Aggiungi un'altra scheda

0.3

Build infrastructure

Build APIs

0/4

+ Aggiungi un'altra scheda

0.4

Build user interface

+ Aggiungi un'altra scheda

0.9

Testing

+ Aggiungi un'altra scheda

1.0

RELEASE

+ Aggiungi un'altra scheda

OBIETTIVI

- Implementare ricerca video per titolo
- Implementare suggerimento prossimo video attinente
- Implementare ricerca video per tag
- Implementare possibilità di creare playlist
- Implementare interazione tra account
- Tema scuro

UTENTI

VALORE

MyTEDx si propone come alternativa all'app ufficiale di TED Conferences LLC 'TED' (20M di download)

MyTEDx punta sulla migrazione di utenti in cerca di un'alternativa:

- 'easy on the eyes'
- con possibilità di creare più playlists
- con possibilità di interazione tra account

CRITICITA'



L'ALTA REATTIVITA' è una caratteristica garantita da una gestione efficiente degli algoritmi di ricerca e da un'interazione efficace con il database.



INTERFACCIA SEMPLICE ED INTUITIVA: Il design user-friendly dell'interfaccia di MyTEDx la rende intuitiva e di semplice utilizzo. L'aspetto lineare e pulito permette un comodo e piacevole utilizzo e previene potenziali errori da parte dell'utente.



INTEGRAZIONE DATI AGGIORNATI : Il DataBase permette una comoda integrazione dei dati e di conseguenza la facile possibilità di aggiornamento degli stessi.



GESTIONE ACCOUNT: La gestione degli account impedisce il sovraccarico dei profili, inserendo alcune limitazioni nella condivisione e nel salvataggio di video. Gli account inattivi per un periodo di tempo prolungato vengono eliminati.

Parte 2

A thin, vertical white line is positioned to the right of the text 'Parte 2', extending from the top of the text area down to the bottom of the slide.

Aggiunta watch_next_dataset

Lo script del job pyspark è stato aggiornato con le seguenti righe di codice:

READ WATCH_NEXT DATASET

```
ted_watch_next_dataset_path = "s3://pirotta-bucket-mytedx/watch_next_dataset.csv"
watch_next_dataset = spark.read.option("header","true").csv(ted_watch_next_dataset_path)
```

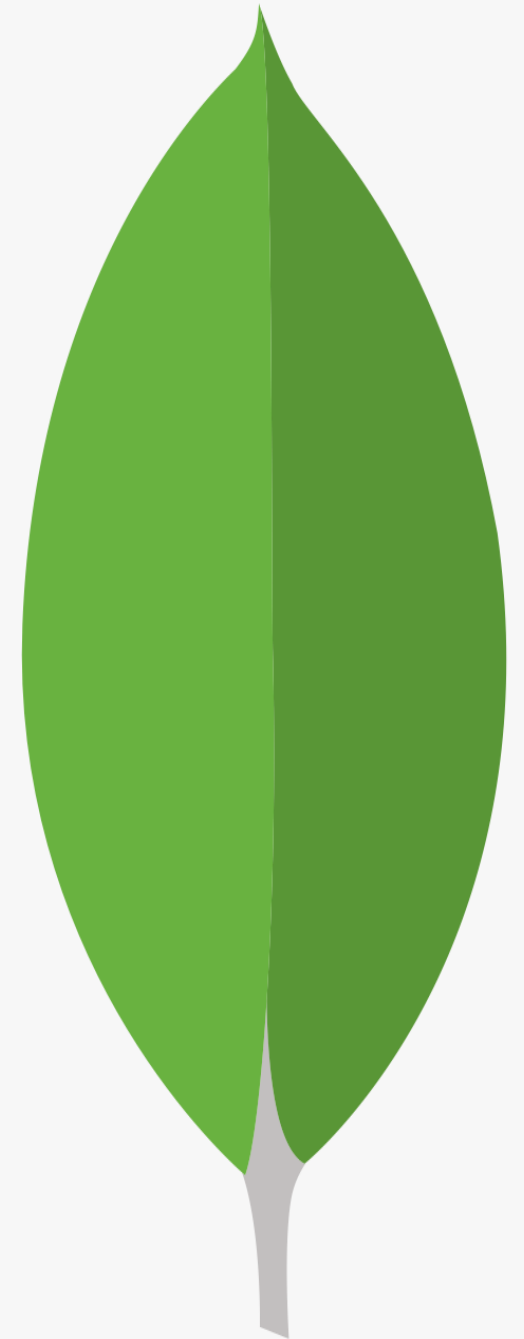
ADD WATCH_NEXT TO TEDX_DATASET

```
watch_next_dataset_agg = watch_next_dataset.groupBy(col("idx").alias("idx_ref2")).agg(collect_list("url").alias("url_next"),collect_list("watch_next_idx").alias("id_next"))
watch_next_dataset_agg.printSchema()
tedx_dataset_agg = tedx_dataset_agg.join(watch_next_dataset_agg, tedx_dataset_agg._id == watch_next_dataset_agg.idx_ref2, "left") \
    .select(col("*")) \
    .drop("idx_ref2") \
tedx_dataset_agg.printSchema()
```


Collection risultante

Lo script è stato modificato al fine di eseguire un ulteriore join con il dataset contenente i video consigliati. Lo schema risultante in MongoDB è il seguente:

<i>_id:</i>	id del video in riproduzione
<i>main_speaker:</i>	relatore del talk in riproduzione
<i>details:</i>	dettagli sul video in riproduzione
<i>posted:</i>	data di pubblicazione del video in riproduzione
<i>url:</i>	url del video in riproduzione
> <i>tags: Array</i>	tags del video in riproduzione
> <i>url_next: Array</i>	url dei video consigliati
> <i>id_next: Array</i>	id dei video consigliati



Criticità tecniche

Sintassi del
linguaggio query
Pyspark

Difficoltà nel testare
lo script causa
tempi di avvio
dell'architettura
cloud prolungati

Possibili evoluzioni



RICERCA VIDEO IN BASE
A DATA PUBBLICAZIONE



RICERCA VIDEO IN BASE
A RELATORE