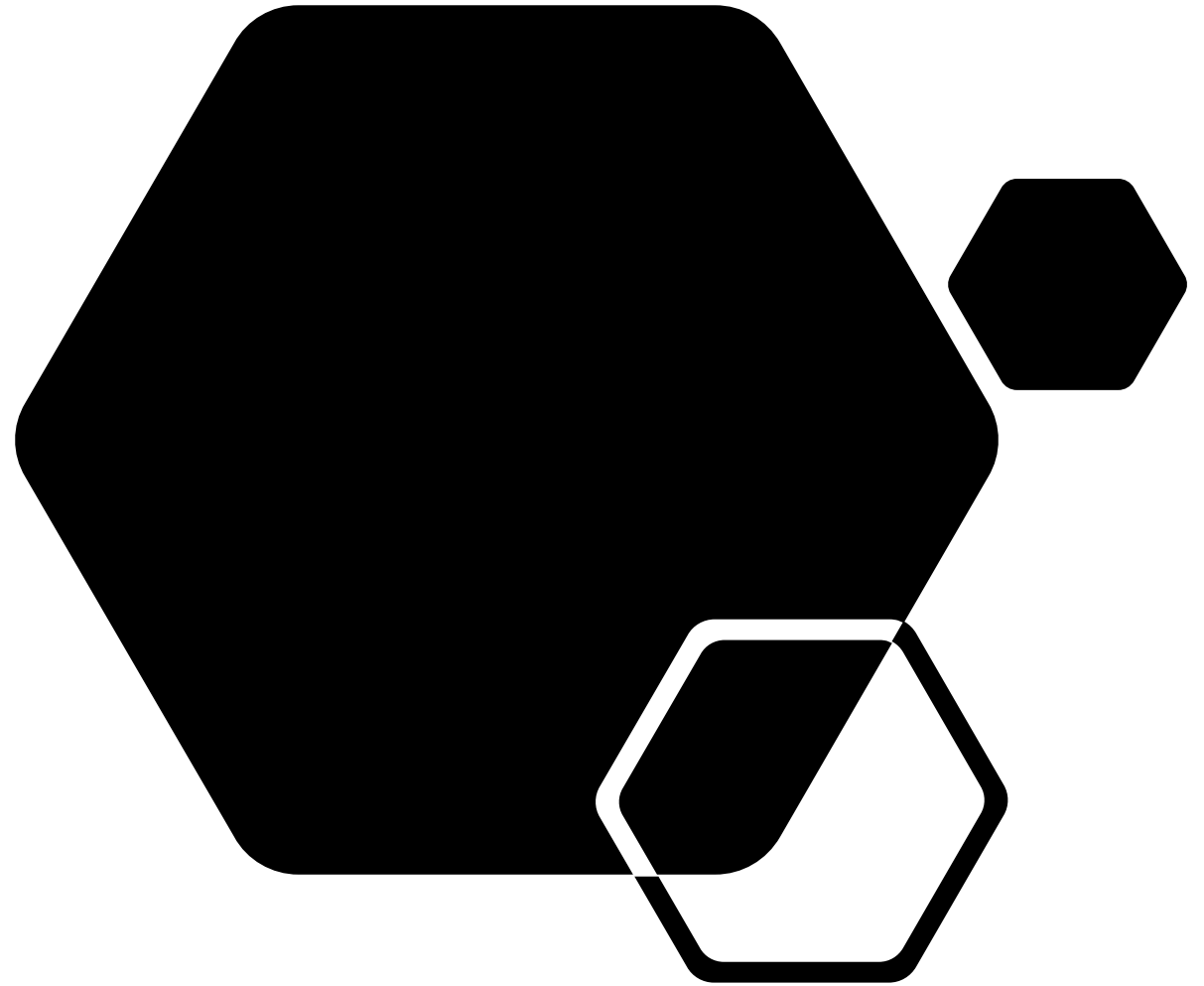
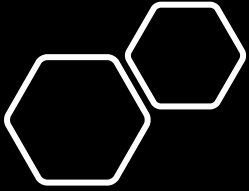


Gerry

AWS Lambda

Carne Federico – 1059865





Watch Next – Lambda

- Dato un `_id`, restituisce il campo `watch_next` del corrispondente talk
- Nella collection `tedx_data`, `watch_next` è un array di `_id` ma sarebbe più interessante restituire una lista di talk con alcuni dettagli, come titolo o speaker.
- L'operatore `$lookup` permette il join ma è poco performante e andrebbe eseguito a ogni richiesta. Visto che i campi `title`, `posted` e `main_speaker` sono ritenuti stabili è più utile, all'inserimento nel DWH, definire `watch_next` non più come un semplice array di id ma come un array di talk, con titolo, url, ecc. . A questo proposito è stato modificato il job Glue incaricato di creare il DWH.
- L'handler filtra in base all'id e restituisce il campo `watch_next`. In questo modo la lambda è più reattiva, al costo di far aumentare di dimensioni la collection

CreateDataLake.py

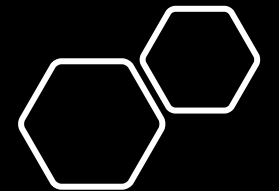
```
# ADD WATCH NEXT DETAILS
watch_next_dataset = watch_next_dataset_filtered.join(tedx_dataset_agg.drop("tags", "details"),
    watch_next_dataset_filtered.watch_next_idx == tedx_dataset_agg._id, "left")

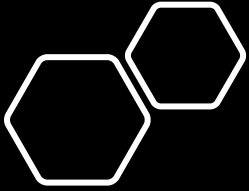
# ADD WATCH NEXT TO AGGREGATE MODEL
watch_next_dataset_agg = watch_next_dataset.groupBy(col("idx").alias("idx_ref_watch_next")) \
    .agg(collect_list(struct("_id", "url", "title", "posted", "main_speaker")) \
    .alias("watch_next"))
```

handler.js

```
talk.findOne({_id: body._id}).then(t => {
    callback(null, {
        statusCode: 200,
        body: JSON.stringify(t.watch_next)
    })
})
```

Watch Next - Job Glue & Handler





Watch Next - API

La lambda function viene esposta tramite API Gateway. Riceve due parametri in ingresso:

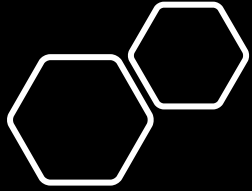
- **_id**: `String`, l'id del talk
- **slice**: `Number`, il numero di watch_next da restituire

Esempio di richiesta:

```
{  
  "_id": "8d2005ec35280deb6a438dc87b225f89",  
  "slice": 3  
}
```

Esempio di risposta:

```
[  
  {  
    "_id": "fe35edd737282ab3a325f2387cf1b50b",  
    "title": "What makes the Great Wall of China so extraordinary",  
    "url": "https://www.ted.com/talks/megan_campisi_and_pen_pen...",  
    "posted": "Posted Sep 2015",  
    "main_speaker": "Megan Campisi and Pen-Pen Chen"  
  },  
  ...  
]
```



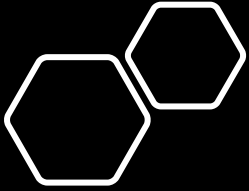
Quiz Generation - Lambda

Generate_Quiz

- Riceve in input una lista di id dei talk, il nome del quiz e alcuni parametri per la generazione dei quesiti
- Controlla che almeno uno tra gli id richiesti esista, crea l'id del quiz e lo restituisce come risposta
- Crea l'oggetto di configurazione per la generazione dei requisiti e invoca in modo asincrono `Generate_Quiz_Async`

Generate_Quiz_Async

- Non è esposta tramite API Gateway ma viene chiamata da `Generate_Quiz`, che le passa l'id del quiz, l'oggetto di configurazione e la lista di id.
- Inserisce nella collection `gerry_quiz_data` il quiz da generare
- Unisce le sentence di ogni talk per crearne la trascrizione e le concatena una all'altra.
- Effettua una richiesta all'API di Quillionz passando un token di autorizzazione, l'oggetto di configurazione e le trascrizioni concatenate, ricevendo come risposta un documento JSON.
- Aggiorna il quiz nella collection aggiungendo i quesiti o i possibili errori generati



Quiz Generation - API

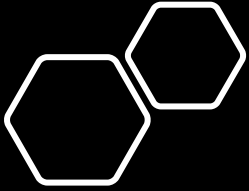
Parametri in ingresso a Generate_Quiz:

- **id_list**: [**String**], la lista di id dei talk
- **title**: **String**, il nome del quiz
- **mcq**: **Boolean**, quesiti in forma di scelta multipla e vero/falso
- **recall**: **Boolean**, quesiti in forma di testo bucato
- **wh_questions**: **Boolean**, quesiti in forma di frasi interrogative
- **short_answer**: **Boolean**, quesiti a cui rispondere con definizioni, spiegazioni, ...

Esempio di richiesta a Generate_Quiz:

```
{  
  "id_list": [  
    "8d2005ec35280deb6a438dc87b225f89"  
  ],  
  "title": "test_api",  
  "mcq": true,  
  "recall": true,  
  "wh_question": true,  
  "short_answer": false  
}
```

Esempio di risposta: { **"_id"**: "d0990e96ca618ac2da9ce152b897ba9a" }



Get Quiz - Lambda & API

Restituisce un quiz dato l'id, così da monitorarne lo stato.

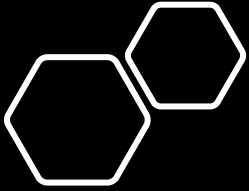
Parametri in ingresso:

- `_id`: `String`, l'id del quiz

Esempio di richiesta: `{ "_id": "d0990e96ca618ac2da9ce152b897ba9a" }`

Esempio di risposta (Generate_Quiz_Async in esecuzione):

```
{
  "_id": "d0990e96ca618ac2da9ce152b897ba9a",
  "status": "generating",
  "talks": [
    {
      "_id": "8d2005ec35280deb6a438dc87b225f89",
      "title": "The intangible effects of walls",
      ...
    },
    ...
  ],
  "title": "test_api",
  "question_params": {
    "shortAnswer": false,
    "recall": true,
    "mcq": true,
    "whQuestions": true
  }
}
```



Get Quiz - Lambda & API

Esempio di risposta (Generate_Quiz_Async terminata):

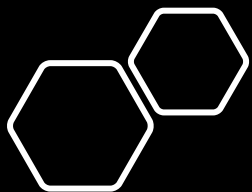
```
{
  "_id": "d0990e96ca618ac2da9ce152b897ba9a",
  "status": "to_validate",
  ...
  "questions": {
    "recall": [
      {
        "Question": "The more a participant was against the  

                     _____ reunification, the further away they  

                     estimated cities to be.",
        "Answer": "German",
        "originalSentence": "The more a participant was  

                           against the German reunification, the further  

                           away they estimated cities to be."
      },
      ...
    ],
    "multipleChoiceQuestions": {
      "mcq": [ ... ],
      "trueFalse": [ ... ]
    },
    "whQuestions": { ... }
  }
}
```

Criticità e possibili evoluzioni

Criticità

- La generazione del quiz potrebbe, per testi lunghi, richiedere più di 15 minuti
- Non è richiesta nessuna autenticazione per generare quiz
- L'API free di Quillionz ha dei limiti:
 1. Massimo 20 richieste al mese
 2. Massimo 3000 parole per richiesta
 3. Non si può definire il numero di quesiti
 4. Per alcuni testi può non generare alcun quesito

Possibili evoluzioni

- Per superare il limite dei 15 minuti si può spostare in AWS Batch la function `Generate_Quiz_Async`
- Per l'autenticazione e il riconoscimento dei moderatori serve integrare Cognito
- Visti i limiti dell'API di Quillionz si potrebbe pensare di appoggiarsi a un altro servizio per la generazione dei quesiti