

## Cloud Architecture

IDs: 314722307, 211406343, 208869222

שאלה 1:

בקובץ ה-ZIP מצורפים שני פרוייקטים: אחד של ה-producer והשני של ה-consumer .

על מנת לייצר את הקלאסטיר השתמשנו ב-

bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties\.

בשביל להרים את הברוקיר

bin\windows\kafka-server-start.bat .\config\server.properties\.

בשביל להרים את השרת

בפרוייקט של ה producer השתמשנו במחלקת KafkaTopicConfig בשביל להגדיר את ה- Topics .

```
1 package com.cloud.kafka.hw3.config;
2 import org.apache.kafka.clients.admin.NewTopic;
3 import org.springframework.context.annotation.Bean;
4 import org.springframework.context.annotation.Configuration;
5 import org.springframework.kafka.config.TopicBuilder;
6
7 @Configuration
8 public class KafkaTopicConfig {
9
10     @Bean
11     public NewTopic wikimediaRecentChanges() {
12         return TopicBuilder.name(name: "wikimedia_recent_changes")
13             .build();
14     }
15
16     @Bean
17     public NewTopic wikimediaPageCreate() {
18         return TopicBuilder.name(name: "wikimedia_page_create")
19             .build();
20     }
21
22     @Bean
23     public NewTopic wikimediaRevisionCreate() {
24         return TopicBuilder.name(name: "wikimedia_revision_create")
25             .build();
26     }
27 }
```

## השמתשנו ב- Kafka producer באופן הבא:

```
package com.cloud.kafka.hw3.kafka;
import java.net.URI;

import org.slf4j.Logger;
import org.springframework.kafka.core.KafkaTemplate;
import org.springframework.stereotype.Service;
import com.cloud.kafka.hw3.WikimediaEventsHandler;

import com.launchdarkly.eventsourcing.EventHandler;
import com.launchdarkly.eventsourcing.EventSource;

@Service
public class KafkaProducer {

    private static final Logger LOGGER = org.slf4j.LoggerFactory.getLogger(clazz: KafkaProducer.class);
    private KafkaTemplate<String, String> kafkaTemplate;

    public KafkaProducer(KafkaTemplate<String, String> kafkaTemplate) {
        this.kafkaTemplate = kafkaTemplate;
    }

    public void sendMessage(){
        String topic = "wikimedia_page_create";
        // String topic = "wikimedia_revision_create";
        // String topic = "wikimedia_recent_changes";
        EventHandler eventHandler = new WikimediaEventsHandler(kafkaTemplate, topic);
        String url = "https://stream.wikimedia.org/v2/stream/page-create";
        // String url = "https://stream.wikimedia.org/v2/stream/revision-create";
        // String url = "https://stream.wikimedia.org/v2/stream/recentchanges";
        EventSource.Builder builder = new EventSource.Builder(eventHandler, URI.create(url));
        EventSource eventSource = builder.build();
        eventSource.start();
    }
}
```

אירועים מתוך page-create נשמרים בתוך Topic בשם Wikimedia\_page\_create  
אירועים מתוך recent changes נשמרים ב Topic בשם Wikimedia\_recent\_changes  
אירועים מתוך revision-create נשמרים ב-Topic בשם Wikimedia\_revision\_create

בשביל לקרוא את האירועים שבקאפקא, השתמשנו ב-

```
Bin\windows\kafka-console-consumer.bat --topic <Topic name> --from-beginning  
--bootstrap-server localhost:9092
```

Example of page create:

```
PS C:\Users\ahmed\Documents\kafka> .\bin\windows\kafka-console-consumer.bat --topic wikimedia_page_create --from-beginning --bootstrap-server localhost:9092  
{"schema":"/mediawiki/revision/create/1.1.0","meta":{"uri":"https://www.wikidata.org/wiki/Q112404861","request_id":"26b94ced-d13c-46f8-9235-86914a093661","id":"bd2127e1-a8ad-4f78-9ebb-7705c5696307","dt":"2022-06-12T12:11:11Z","domain":"www.wikidata.org","stream":"mediawiki.page-create","topic":"eqiad.mediawiki.page-create","partition":0,"offset":234297770},"database":"wikidatawiki","page_id":107296520,"page_title":"Q112404861","page_namespace":0,"rev_id":1658940351,"rev_timestamp":"2022-06-12T12:11:11Z","rev_sha1":"gtqlygv080s59a5izgdr9lw6qnk8t","rev_minor_edit":false,"rev_len":8270,"rev_content_model":"wikibase-item","rev_content_format":"application/json","performer":{"user_text":"Vojtěch Dostál","user_groups":["propertycreator","*", "user","autoconfirmed"],"user_is_bot":false,"user_id":963971,"user_registration_dt":"2014-05-12T20:28:45Z","user_edit_count":2739535},"page_is_redirect":false,"comment":"/* wbeditentity-create-item:0 */ new item from NKC dataset ([[[:toollabs:editgroups/b/OR/1ffffac6c346|details]])","parsedcomment":"<span dir=\"auto\"><span class=\"autocomment\">wbeditentity-create-item:0|: </span> new item from NKC dataset (<a href=\"https://iw.toolforge.org/editgroups/b/OR/1ffffac6c346\" class=\"extiw\" title=\"toollabs:editgroups/b/OR/1ffffac6c346\">details</a></span>","rev_slots":{"main":{"rev_slot_content_model":"wikibase-item","rev_slot_sha1":"gtqlygv080s59a5izgdr9lw6qnk8t","rev_slot_size":8270,"rev_slot_origin_rev_id":1658940351}}}  
{"schema":"/mediawiki/revision/create/1.1.0","meta":{"uri":"https://www.wikidata.org/wiki/Q112404862","request_id":"ddc3281b-ae25-4398-87dd-e5013c0abd95","id":"82a04f64-ed58-41c7-bd40-4cc7b4a30f08","dt":"2022-06-12T12:11:12Z","domain":"www.wikidata.org","stream":"mediawiki.page-create","topic":"eqiad.mediawiki.page-create","partition":0,"offset":234297771},"database":"wikidatawiki","page_id":107296521,"page_title":"Q112404862","page_namespace":0,"rev_id":1658940358,"rev_timestamp":"2022-06-12T12:11:12Z","rev_sha1":"pfiwk76668d2rnap61s3qifewjfnt80","rev_minor_edit":false,"rev_len":563,"rev_content_model":"wikibase-item","rev_content_format":"application/json","performer":{"user_text":"TongcyBot","user_groups":["*", "user","autoconfirmed"],"user_is_bot":false,"user_id":4746516,"user_registration_dt":"2020-11-19T18:23:32Z","user_edit_count":10700},"page_is_redirect":false,"comment":"/* wbeditentity-create-item:0 */","parsedcomment":"<span dir=\"auto\"><span class=\"autocomment\">wbeditentity-create-item:0|</span></span>","rev_slots":{"main":{"rev_slot_content_model":"wikibase-item","rev_slot_sha1":"pfiwk76668d2rnap61s3qifewjfnt80","rev_slot_size":563,"rev_slot_origin_rev_id":16589403
```

השמתשנו ב- MySQL בשביל לשמור אופייקטים עם שדות בשביל להשתמש בהן אחר כך למדדים (אמנם לא מימשנו מימדים בגלל חוסר זמן) אבל הוצאנו מהזרמים את המידע הנדרש.

Schema of recent changes:

Result Grid							
Filter Rows:							
Edit: Export/Import: Wrap Cell Content: Fetch rows:							
_id	bot	wiki_id	request_id	server_url	title	type	user
1	0	271283635	NULL	https://ar.wikipedia.org	تصنيف:جميع المقالات بدون مصدر	categorize	Sgh45
2	0	271283636	NULL	https://ar.wikipedia.org	تصنيف:مقالات بدون مصدر منذ 2018	categorize	Sgh45
3	1	1515237879	NULL	https://en.wikipedia.org	COVID-19 vaccination in Bulgaria	edit	BattyBot
4	1	1710988990	NULL	https://www.wikidata.org	Q84265050	edit	Emijrpbot
5	0	344193	NULL	https://sn.wikipedia.org	Mitauro ye Chinaija-Kongo	edit	Tsvarakadega
6	1	110132363	NULL	https://vi.wikipedia.org	Bulbophyllum tylophorum	edit	ThitxongkhaiAWB
7	1	1945442196	NULL	https://commons.wikimedia.org	Category:Large images	categorize	Crystal-bot
8	1	620340	NULL	https://zh.wikinews.org	2022年6月12日中文報紙頭條	edit	Cewbot
9	0	438967874	NULL	https://ru.wikipedia.org	Осрпей, Утун	edit	HAHapees
10	0	1710988991	NULL	https://www.wikidata.org	Q112418750	new	Vojtěch Dostál
11	0	1515237880	NULL	https://en.wikipedia.org	Church of the Saviour, Thessaloniki	edit	Florificapis
12	0	149285	NULL	https://lld.wikipedia.org	Categoria:Zità de l'USA	categorize	Asenoner
13	1	1945442197	NULL	https://commons.wikimedia.org	File:IGN - Alpes Maritimes - Temp...	edit	Crystal-bot
14	0	149286	NULL	https://lld.wikipedia.org	Categoria:Washington	categorize	Asenoner
15	1	1945442198	NULL	https://commons.wikimedia.org	Commons:Copyright rules by terr...	edit	Ameisenigel

קוד מלא נמצא בפרוייקט consumerdb

## שאלה C-2

בתרגיל זה השמתשנו ב-3 TOPICS בשביל לשמור את הנתונים בקאפקא.  
השתמשנו ב-PARTITION אחד, אמנם ניתן לייצר יותר: במחלקת KafkaTopicConfig.java  
אפשר לעשות פשוט:

```
@Bean
public NewTopic wikimediaRecentChanges() {
    return TopicBuilder.name("wikimedia_recent_changes")
        .partitions(partitionCount: 4)
        .replicas(replicaCount: 3)
        .build();
}
```

למשל פה ניתן לחלק את ה-TOPIC בשם Wikimedia\_recent\_changes לתוך 4  
partition עם 3 ריפלקות.

בסביבה PRODUCTION אמיתית נצטרך יותר פארטישנים המפוזרים על מכונות שונות  
(כאן KAFKA לוקאלית ולכן מספר פארטישנים גדול לא ממש יעיל) ומספיק 3 ריפלקות  
מבחינה תיאוריטית.

## שאלה D-2

גישה OFFLINE בשביל ליצר מדדים היא למשל לשמור את ההודעות בתוך בסיס ניתונים חיצוני (mysql, Cassandra..) ואחרי זה לעבד אותם. זה דומה למה שעשינו בתרגיל.