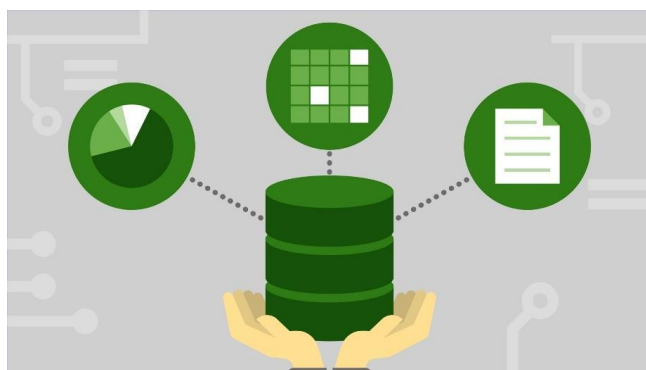


به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



آزمایشگاه پایگاه داده

دستور کار شماره ده (کار با آپاچی دروید و کافکا)

امیرعلی رایگان

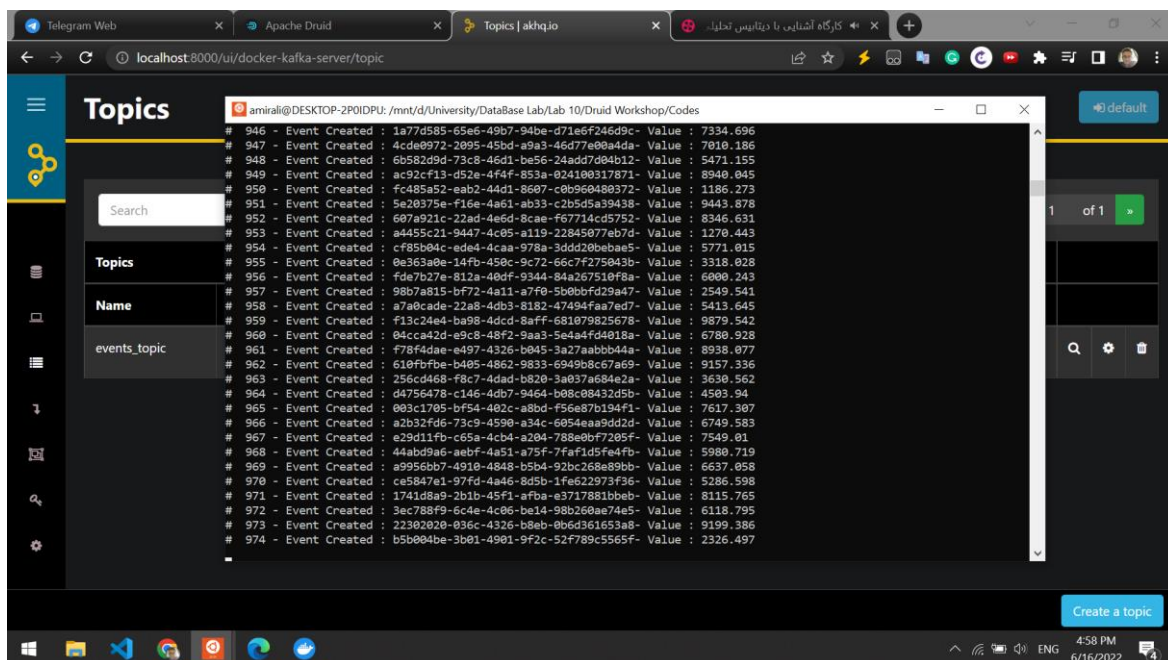
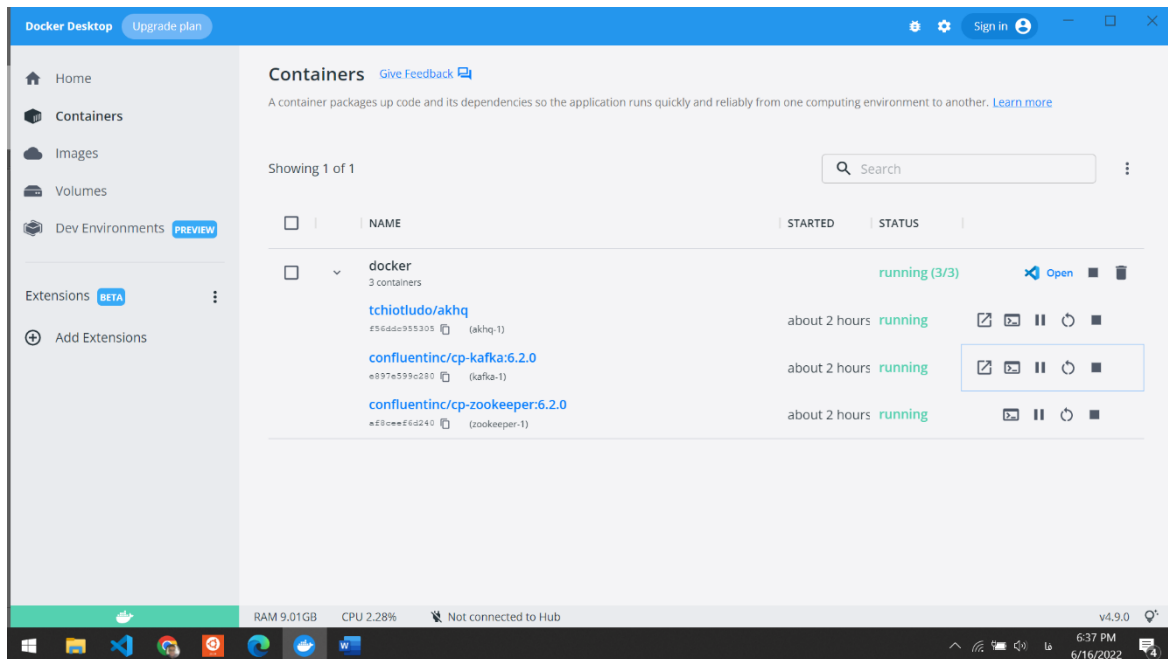
۸۱۰۱۹۷۶۲۳

اسفند ۱۴۰۰

گزارش دستورکار انجام شده

گام اول. راه اندازی آپاچی دروید و داکر و کافکا

در گام اول این آزمایش مطابق ویدیو های آموزشی که در سایت آپارات قرار دارد با سرویس های اشاره شده آشنا شدیم و آنها را در سیستم شخصی بالا آوردیم. عکس های اسکرین شات مربوط به این سرویس ها در ادامه آمده اند.



The screenshot shows the Apache Kafka Topics management interface. The browser address bar indicates the URL: `localhost:8000/ui/docker-kafka-server/topic?search=&topicListView=HIDE_INTERNAL&page=1`. The interface has a dark theme. At the top, there's a search bar and a 'Hide internal topics' dropdown. Below this, a table lists the topics. The table has columns: Topics, Partitions, Replications, and Consumer Groups. The 'events_topic' is listed with 3 partitions, 1 replication factor, and 1 in sync. The bottom right corner has a 'Create a topic' button.

Topics	Partitions	Replications	Consumer Groups
Name	Count	Factor	In Sync
events_topic	3	1	1

The screenshot shows the Apache Druid Unified Console. The browser address bar indicates the URL: `localhost:8888/unified-console.html#ingestion`. The interface is dark-themed. The top navigation bar includes links for Load data, Ingestion, Datasources, Segments, Services, and Query. The main content area is divided into two sections: Supervisors and Tasks. The Supervisors section shows a table with columns: Datasource, Type, Topic/Stream, Status, and Actions. The Tasks section shows a table with columns: Task ID, Group ID, Type, Datasource, Location, Created time, Status, Duration, and Actions.

Datasource	Type	Topic/Stream	Status
events	kafka	events_topic	RUNNING
events_topic	kafka	events_topic	RUNNING
events_topic_daily	kafka	events_topic	RUNNING
events_topic_hourly	kafka	events_topic	RUNNING

Task ID	Group ID	Type	Datasource	Location	Created time	Status
index_kafka_events_85360f66e9155a_fhnmblidj	index_kafka_events	index_kafka	events	localhost:8101	2022-06-16T13:34:47.232Z	RUNNING
index_kafka_events_topic_daily_ffc93d0d83f1cba_fafnmld	index_kafka_events_topic_daily	index_kafka	events_topic_daily	localhost:8100	2022-06-16T13:34:42.968Z	RUNNING
index_kafka_events_topic_hourly_a63874dca1cc738_eceikngn	index_kafka_events_topic_hourly	index_kafka	events_topic_hourly		2022-06-16T14:02:31.039Z	PENDING
index_kafka_events_topic_3a7ca7b0525772c_cpghell	index_kafka_events_topic	index_kafka	events_topic		2022-06-16T13:34:47.484Z	PENDING

```

amirali@DESKTOP-2P0IDPU: ~/apache-druid-0.22.1/bin
amirali@DESKTOP-2P0IDPU:~/apache-druid-0.22.1$ cd bin
amirali@DESKTOP-2P0IDPU:~/apache-druid-0.22.1/bin$ ls
broker.sh      middleManager.sh  start-cluster-data-server
coordinator.sh node.sh           start-cluster-master-no-zk-server
dsq1          overlord.sh      start-cluster-master-with-zk-server
dsq1-main     post-index-task  start-cluster-query-server
generate-example-metrics post-index-task-main start-micro-quickstart
historical.sh run-druid         start-nano-quickstart
java-util     run-zk           start-single-server-large
jconsole.sh   service          start-single-server-medium
amirali@DESKTOP-2P0IDPU:~/apache-druid-0.22.1/bin$ ./start-micro-quickstart
[Thu Jun 16 18:04:11 2022] Running command[zk], logging to[/home/amirali/apache-druid-0.22.1/bin/run-druid coordinator-overlord conf/druid/single-server/micro-quickstart]
[Thu Jun 16 18:04:11 2022] Running command[broker], logging to[/home/amirali/apache-druid-0.22.1/bin/run-druid broker conf/druid/single-server/micro-quickstart]
[Thu Jun 16 18:04:11 2022] Running command[router], logging to[/home/amirali/apache-druid-0.22.1/bin/run-druid router conf/druid/single-server/micro-quickstart]
[Thu Jun 16 18:04:11 2022] Running command[historical], logging to[/home/amirali/apache-druid-0.22.1/bin/run-druid historical conf/druid/single-server/micro-quickstart]
[Thu Jun 16 18:04:11 2022] Running command[middleManager], logging to[/home/amirali/apache-druid-0.22.1/bin/run-druid middleManager conf/druid/single-server/micro-quickstart]
kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 13:54:54,716] DEBUG [Controller id=0] Topics not in preference
() (kafka.controller.KafkaController)
[2022-06-16 13:54:54,717] TRACE [Controller id=0] Leader imbalance ratio
a.controller.KafkaController)
docker-kafka-1 [2022-06-16 13:59:54,718] INFO [Controller id=0] Processing automatic pr
ction (kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 13:59:54,719] TRACE [Controller id=0] Checking need to trig
kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 13:59:54,721] DEBUG [Controller id=0] Topics not in preferre
() (kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 13:59:54,721] TRACE [Controller id=0] Leader imbalance ratio
a.controller.KafkaController)
docker-kafka-1 [2022-06-16 14:04:54,721] INFO [Controller id=0] Processing automatic pr
ction (kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 14:04:54,721] TRACE [Controller id=0] Checking need to trig
kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 14:04:54,722] DEBUG [Controller id=0] Topics not in preferre
() (kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 14:04:54,722] TRACE [Controller id=0] Leader imbalance ratio
a.controller.KafkaController)
docker-kafka-1 [2022-06-16 14:09:54,723] INFO [Controller id=0] Processing automatic pr
ction (kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 14:09:54,724] TRACE [Controller id=0] Checking need to trig
kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 14:09:54,725] DEBUG [Controller id=0] Topics not in preferre
() (kafka.controller.KafkaController)
docker-kafka-1 [2022-06-16 14:09:54,725] TRACE [Controller id=0] Leader imbalance ratio
a.controller.KafkaController)
  
```

گام دوم. اجرای کوئری های دلخواه روی داده های تولید شده

در این گام با استفاده از امکاناتی که آپاچی دروید در اختیار ما قرار داده کوئری هایی روی داده های تولید شده می زنیم و نتیجه ی آنها را می بینیم.

کوئری ها به شکل زیر هستند:

- تمام event های یک روزه ای که در آخرین سال هستند.

__time	count	daily_min	event_id	event_type	page	product_id	sum_event_value	user_name
2021-10-07T00:00:00.000	1	0	12e4617a-b986-4601-801-h0ver	hover	search/blog/list	product05	5051.925	liurarodgers
2021-10-07T00:00:00.000	1	0	d95b71f-76cb-4db7-9bd-h0ver	hover	category/tag	product08	6033.357	athompson
2021-10-08T00:00:00.000	1	0	e46eeba4-6e82-44cd-912-click	click	tag/tags	product06	8283.769	christianyoung
2021-10-08T00:00:00.000	1	0	eb93f5da-71e3-4dba-9ef-idle_5	idle_5	posts	product01	3932.376	thomasguerra
2021-10-10T00:00:00.000	1	0	999324fa-c72e-4599-9d9-idle_5	idle_5	list	product09	6342.494	susandaniel
2021-10-10T00:00:00.000	1	0	9c8bad70-0cb8-4ec7-9eb-idle_5	idle_5	list/categories	product09	7327.352	mthompson
2021-10-11T00:00:00.000	1	0	156843d7-4f25-4baa-809-idle_5	idle_5	explore/wp-content/sear	product04	2498.23	estradasarah
2021-10-11T00:00:00.000	1	0	90661aaf-061d-4485-a52-buy	buy	tag	product09	6322.67	dcrook
2021-10-12T00:00:00.000	1	0	4722c08-41ed-4953-bec-idle_5	idle_5	wp-content	product06	1678.109	sbrown
2021-10-12T00:00:00.000	1	0	496aed52-e25a-4217-ba0-hover	hover	list	product08	3097.768	tblivins
2021-10-12T00:00:00.000	1	0	511f616-65ea-4117-a02-idle_5	idle_5	explore	product04	1860.968	michael13

- تمام event هایی که ارزش آنها بین ۹۵۰۰ و ۹۸۰۰ است.

The screenshot shows the Apache Druid console interface. The query executed is:

```
1 select *
2 from events
3 where event_value > 9500 and event_value < 9800
```

The results show 41 rows in 0.09s. The table columns are: _time, event_id, event_type, event_value, page, product_id, and user_name. The data is filtered to show events with values between 9500 and 9800.

_time	event_id	event_type	event_value	page	product_id	user_name
2021-10-01T00:00:00.000Z	ff253b72-b2ba-4f7e-aeb3-ab94e	idle_5	9687.437	main/tag	product09	kimmichael
2021-10-05T00:00:00.000Z	25d7e8b0-e836-4b73-9135-3fa2	idle_5	9658.711	explore	product04	nmorales
2021-10-05T00:00:00.000Z	a163f6e-3dbb-4d94-976c-c2b3	idle_5	9508.95	category/app/explore	product03	davidmosley
2021-10-06T00:00:00.000Z	413a11ab-e7ce-4cbc-aad9-f875c	idle_5	9502.408	categories	product08	arivas
2021-10-07T00:00:00.000Z	64340989-4739-45f8-bd18-95af	hover	9501.062	main	product05	fagairre
2021-10-07T00:00:00.000Z	8e980079-393d-402a-8ccd-6d1c	hover	9656.116	category/wp-content/main	product04	lindseydavidson
2021-10-10T00:00:00.000Z	9b06c08c-e462-4075-bd8e-243d	buy	9657.089	tag	product10	martinezmatthew
2021-10-16T00:00:00.000Z	1e672c58-db3d-4d61-bd31-08ec	click	9718.795	search/category/app	product09	rdark
2021-10-16T00:00:00.000Z	add49129-f185-4a9b-82a0-e7be	hover	9717.244	tag/search/tag	product01	kristin70
2021-10-19T00:00:00.000Z	80187419-525c-497b-a3c5-8494	hover	9598.766	blog/explore	product05	pittscarrie
2021-10-20T00:00:00.000Z	869af06f-7618-44ce-bee4-1c817	buy	9673.683	tags	product10	renee hunter

- تعداد دفعاتی که هر محصول در کل event ها تکرار شده.

The screenshot shows the Apache Druid console interface. The query executed is:

```
1 SELECT product_id, COUNT(*) as "count"
2 FROM events
3 GROUP BY 1
4 ORDER BY 2 desc
```

The results show 10 rows in 0.35s. The table columns are: product_id and count. The data is grouped by product_id and ordered by count in descending order.

product_id	count
product05	171
product01	158
product03	156
product07	146
product06	145
product10	145
product09	144
product04	142
product08	137
product02	125

- تعداد event هایی که هر روز اتفاق افتاده اند.

The screenshot shows the Apache Druid console interface. The query editor contains the following SQL:

```
1 SELECT TIME_SLONK(_time, 'PT1H') as "Time"
2 FROM event_topic_daily
3 GROUP BY 1
4 ORDER BY 1 ASC
```

The results table shows 66 results in 0.26s. The table has two columns: 'Time' and 'Count'.

Time	Count
2021-10-12T00:00:00.000Z	3
2021-10-13T00:00:00.000Z	5
2021-10-14T00:00:00.000Z	1
2021-10-15T00:00:00.000Z	2
2021-10-16T00:00:00.000Z	2
2021-10-17T00:00:00.000Z	2
2021-10-18T00:00:00.000Z	2
2021-10-19T00:00:00.000Z	1
2021-10-20T00:00:00.000Z	3
2021-10-21T00:00:00.000Z	1
2021-10-22T00:00:00.000Z	2

- تعداد event هایی که به ازای هر محصول در هر صفحه اتفاق افتاده اند.

The screenshot shows the Apache Druid console interface. The query editor contains the following SQL:

```
1 SELECT product_id, page, COUNT(*) as "count"
2 FROM events_topic_hourly
3 GROUP BY 1,2
4 ORDER BY 3 DESC
```

The results table shows 100+ results in 1.08s. The table has three columns: 'product_id', 'page', and 'count'.

product_id	page	count
product03	tags	16
product07	tag	16
product07	wp-content	16
product04	search	15
product05	categories	15
product02	tag	14
product04	main	14
product10	tags	14
product06	blog	13
product01	explore	12
product03	blog	12

مشکلات و توضیحات تکمیلی

در فرآیند اتصال کافکا و داکر مشکلاتی به وجود آمد که با پاسخ استاد حل شد.

آنچه آموختم / پیشنهادات

تجربه بسیار خوب و ملموسی از کار با کافکا و به خصوص داکر به وجود آمد.