

Big Data Infrastructure Final Assignment

Barcelona Technology School

Roberto Arenal
Khaldoun Al-Naierat

Airflow Connections

1) Docker-postgres

Conn Id *	docker-postgres
Conn Type *	Postgres <small>Conn Type missing? Make sure you've installed the right python packages</small>
Description	Docker postgres conn
Host	172.27.240.1
Schema	
Login	postgres
Password	
Port	5432

2) Docker-mysql

Conn Id *	docker-mysql
Conn Type *	MySQL <small>Conn Type missing? Make sure you've installed the right python packages</small>
Description	Docker mysql conn
Host	172.27.240.1
Schema	
Login	root
Password	
Port	3306
Extra	<code>{"charset": "utf8", "local_infile": true}</code>

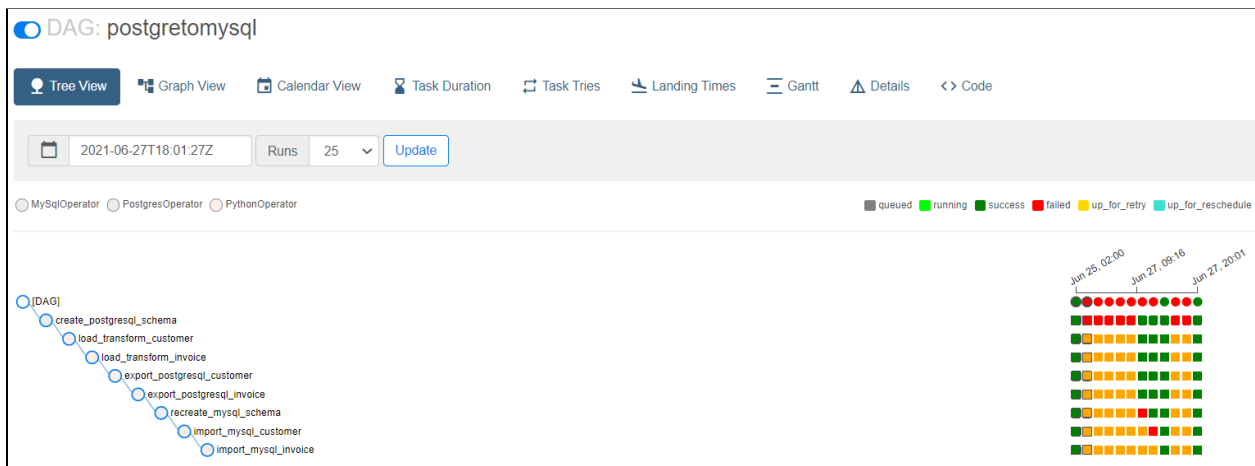
3) Docker-mongo

Conn Id *	docker-mongo
Conn Type *	MongoDB
Conn Type missing? Make sure you've installed the connector	
Description	Docker mongo conn
Host	172.27.240.1
Schema	test
Login	
Password	
Port	27017

DAG 1- Moving from Postgres to MySql

Steps:

- 1) Create postgres schema
- 2) Load customer table to postgres
- 3) Load invoice table to postgres
- 4) Export customer table as csv
- 5) Export invoice table as csv
- 6) Recreate mysql schema
- 7) Import customer table to mysql
- 8) Import invoice table to mysql



Postgress

postgres - 172.27.240.1:5432

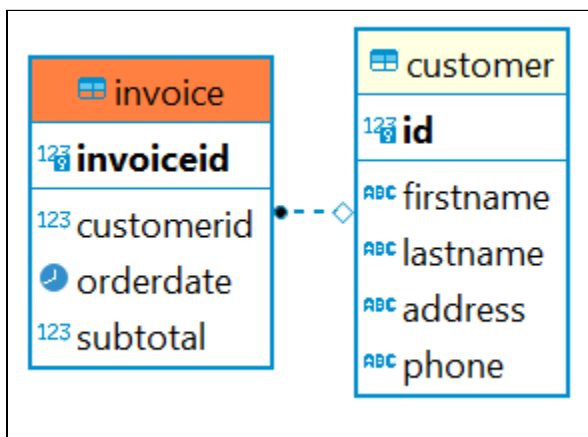
- postgres
 - Schemas
 - public
 - test
 - Tables
 - customer 32K
 - invoice 24K
 - Views
 - Materialized Views
 - Indexes
 - customer.customer_pk
 - invoice.invoice_pk

customer Enter a SQL expression to filter results (use Ctrl+Space)

	id	firstname	lastname	address	phone
1	1	Roberto	Arenal	4 Place Jussieu, 75005 Paris, France	+33144274427
2	2	Donald	Trump	Wellcome Trust Genome Campus, Hin	+441223834244
3	3	Josh	Jacobs	Urbana, IL 61801	+12173332151
4	4	Daniel	Freeman	Pla de Palau, 10	+15417375075

invoice Enter a SQL expression to filter results (use Ctrl+Space)

	invoiceid	customerid	orderdate	subtotal
1	1	1	2019-01-28	5,003.5
2	2	1	2018-03-15	8,657
3	3	3	2019-05-23	10,788
4	4	4	2020-04-08	13,452



MySQL

mysql - localhost:3306

- Databases
 - sys
 - test
 - Tables
 - customer 16K
 - invoice 32K
 - Views
 - Indexes
 - customer.PRIMARY
 - invoice.PRIMARY
 - invoice.invoice_FK

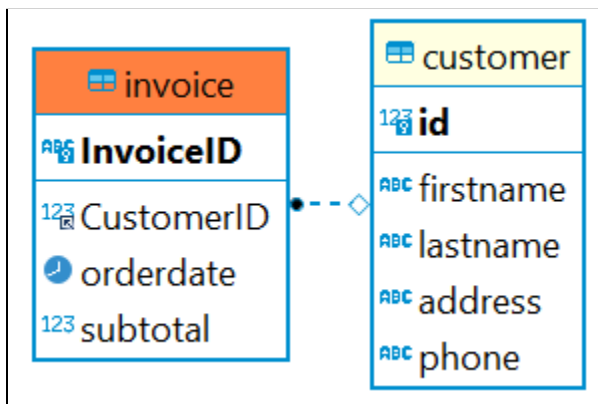
Properties Data ER Diagram

customer Enter a SQL expression to filter results (use Ctrl+Space)

	id	firstname	lastname	address	phone
1	1	Roberto	Arenal	4 Place Jussieu, 75005 Paris, France	+33144274427
2	2	Donald	Trump	Wellcome Trust Genome Campus, Hinx	+441223834244
3	3	Josh	Jacobs	Urbana, IL 61801	+12173332151
4	4	Daniel	Freeman	Pla de Palau, 10	+15417375075

invoice Enter a SQL expression to filter results (use Ctrl+Space)

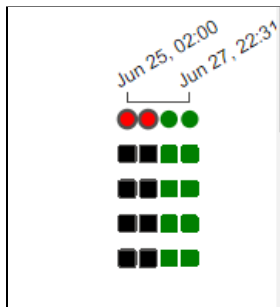
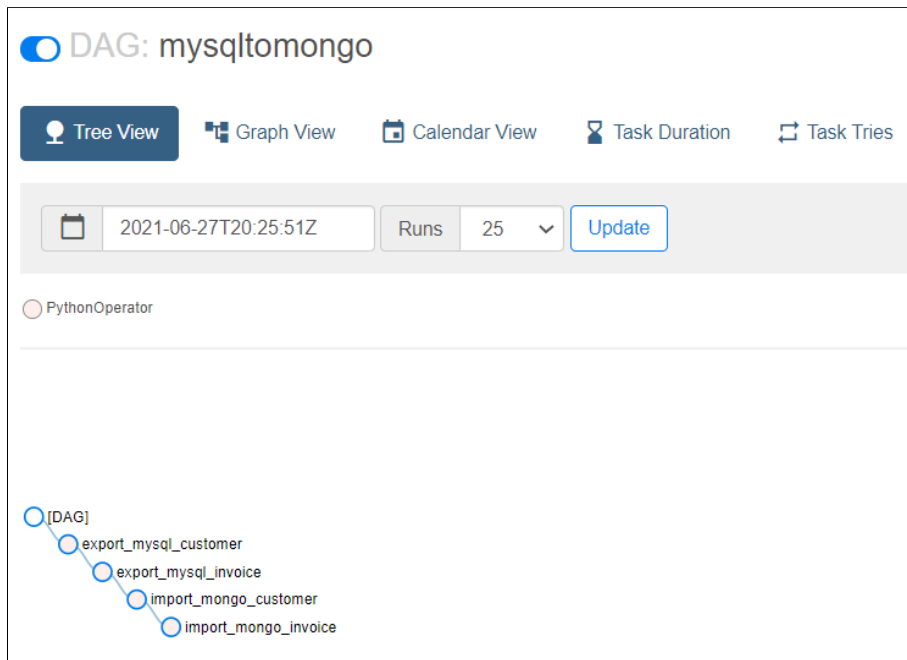
	InvoiceID	CustomerID	orderdate	subtotal
1	1	1	2019-01-28	5,003.5
2	2	1	2018-03-15	8,657
3	3	3	2019-05-23	10,788
4	4	4	2020-04-08	13,452



DAG 2- Moving from MySql to MongoDB

Steps:

- 1) Export customer table from mysql
- 2) Export invoice table from mysql
- 3) Import Customer to mongodb
- 4) Import invoice table to mongodb



MongoDB before running DAG

Database Name ^	Storage Size	Collections	Indexes	
admin	20.0KB	0	1	
config	20.0KB	0	2	
local	20.0KB	1	1	

MongoDB after running DAG

Database Name ^	Storage Size	Collections	Indexes	
admin	20.0KB	0	1	
config	36.0KB	0	2	
local	20.0KB	1	1	
test	40.0KB	2	2	

Collection Name ^	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
customer	4	145.5 B	582.0 B	1	20.0 KB	
invoice	4	104.5 B	418.0 B	1	20.0 KB	

customer				
	_id ObjectId	id String	firstname String	lastname String
1	60d8e0ada31e21aab6bbb269	"1"	"Roberto"	"Arenal"
2	60d8e0ada31e21aab6bbb26a	"2"	"Donald"	"Trump"
3	60d8e0ada31e21aab6bbb26b	"3"	"Josh"	"Jacobs"
4	60d8e0ada31e21aab6bbb26c	"4"	"Daniel"	"Freeman"

🏠 invoice

	_id ObjectId	InvoiceID String	CustomerID String	orderdate String	subtotal
1	60d8e0b01ba7cdc8cb117b90	"1"	"1"	"2019-01-28"	"5003.5"
2	60d8e0b01ba7cdc8cb117b91	"2"	"1"	"2018-03-15"	"8657.0"
3	60d8e0b01ba7cdc8cb117b92	"3"	"3"	"2019-05-23"	"10788.0"
4	60d8e0b01ba7cdc8cb117b93	"4"	"4"	"2020-04-08"	"13452.0"