ITMO University

Course- Big Data

“Clickhouse”

BY:-

Abizer Safdari

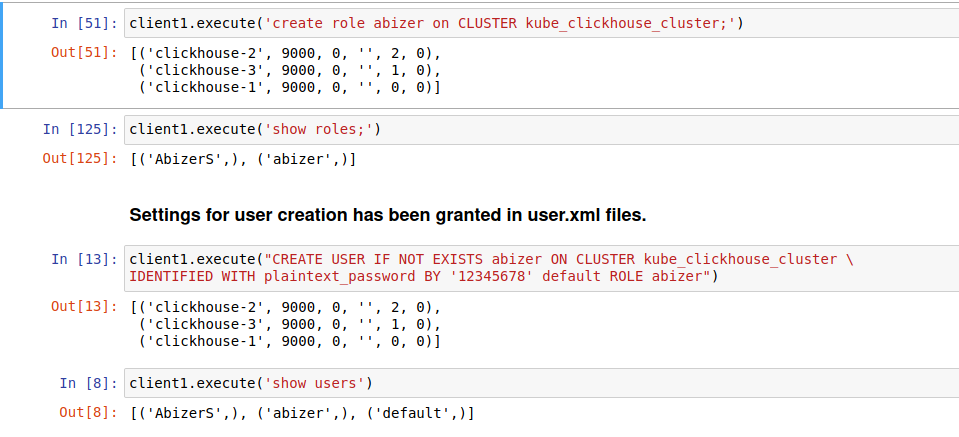
J41321c/J4134c

Saint-Petersburg

2021

User and Role

1. Un-commended authentication permission from nfs\_home/app\_data/clickhouse/commom/users.xml
2. Create user and role

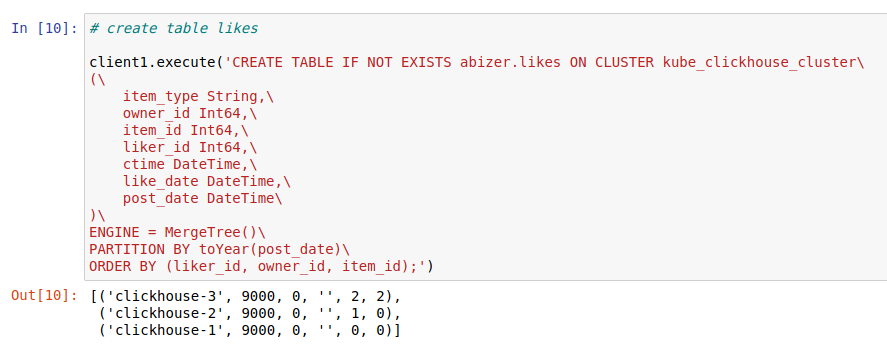


Database:



Table schema:

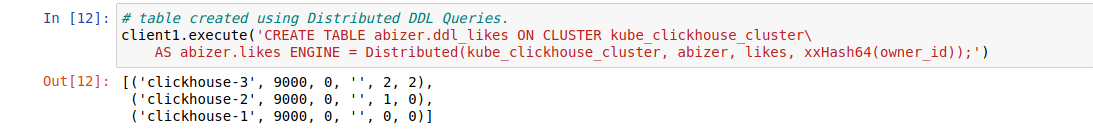
All tables were created using MergeTree Engine



Distributed table:

As cluster is in three shards its better to distribute the data among server.

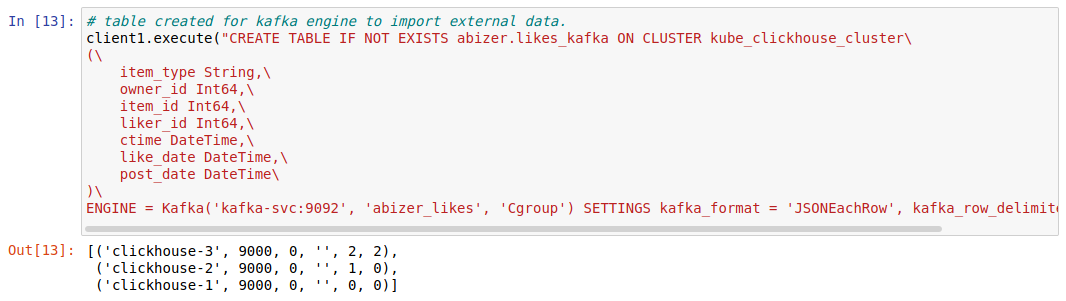
distributed table are created using distributed engine



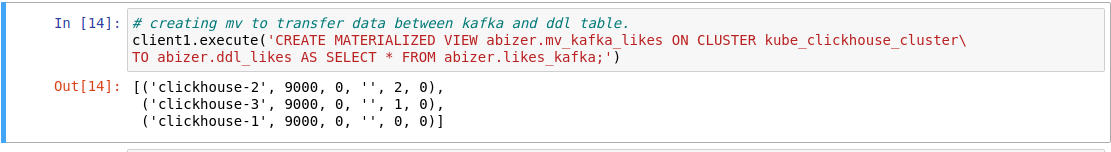
Shared key is used to keep data locality as much as possible.

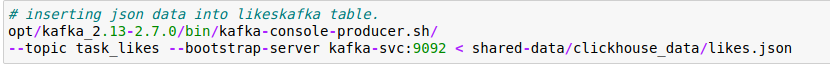
Kafka engine

Kafka engine is used to retrieve data.



After this a materialized view is created to send data from kafka tabke to distributed table.





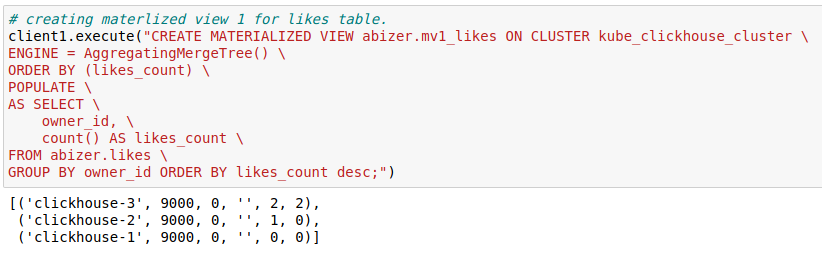
To send data into Kafka table

This can also be used to send data directly into distributed table

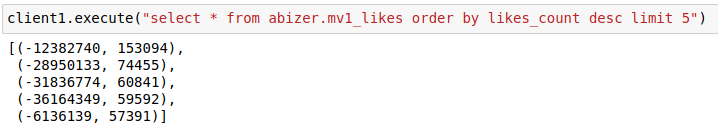
{cat /shared-data/clickhouse\_data/likes.json | clickhouse-client --query="INSERT INTO abizer.ddl\_likes FORMAT JSONEachRow"}

\*Note: Till here all steps are repeated for all tables (friends ,followers ,post, user\_profile)

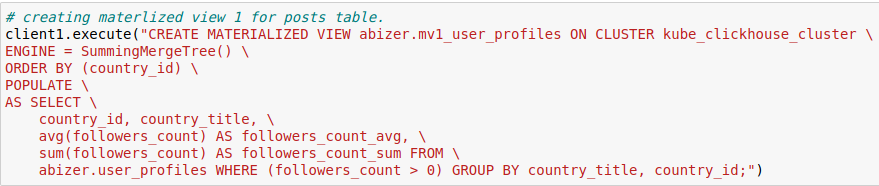
Materialized view:



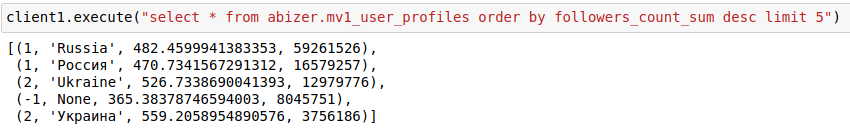
1st view created to show owner id and there total number of likes.



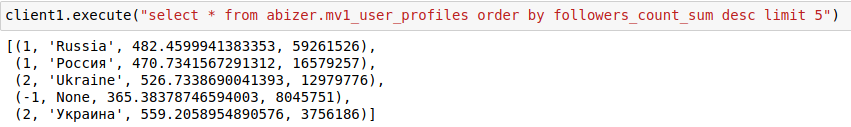
1st output



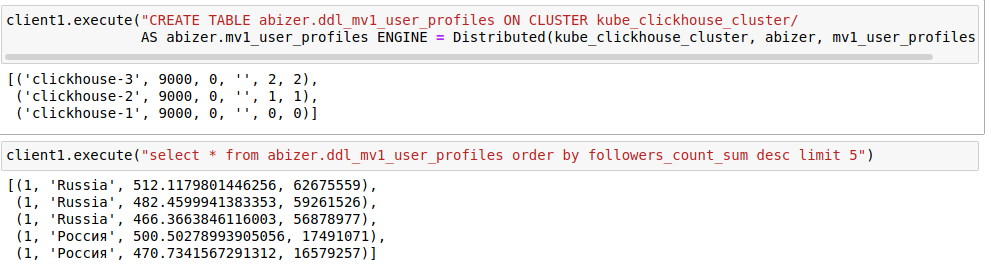
2nd view created to show country id,title, average followers count, total followers count



2nd output



Distributed table of each MV’s are created using distributed engine



Output of distributed table differs from main table

Conclusion

* All tables are created by following json schema.
* Tables are distributed and data locality is kept
* Created one 2 role and 2 user and given access to cluster
* 3 MV’s created(2 shown in report)