

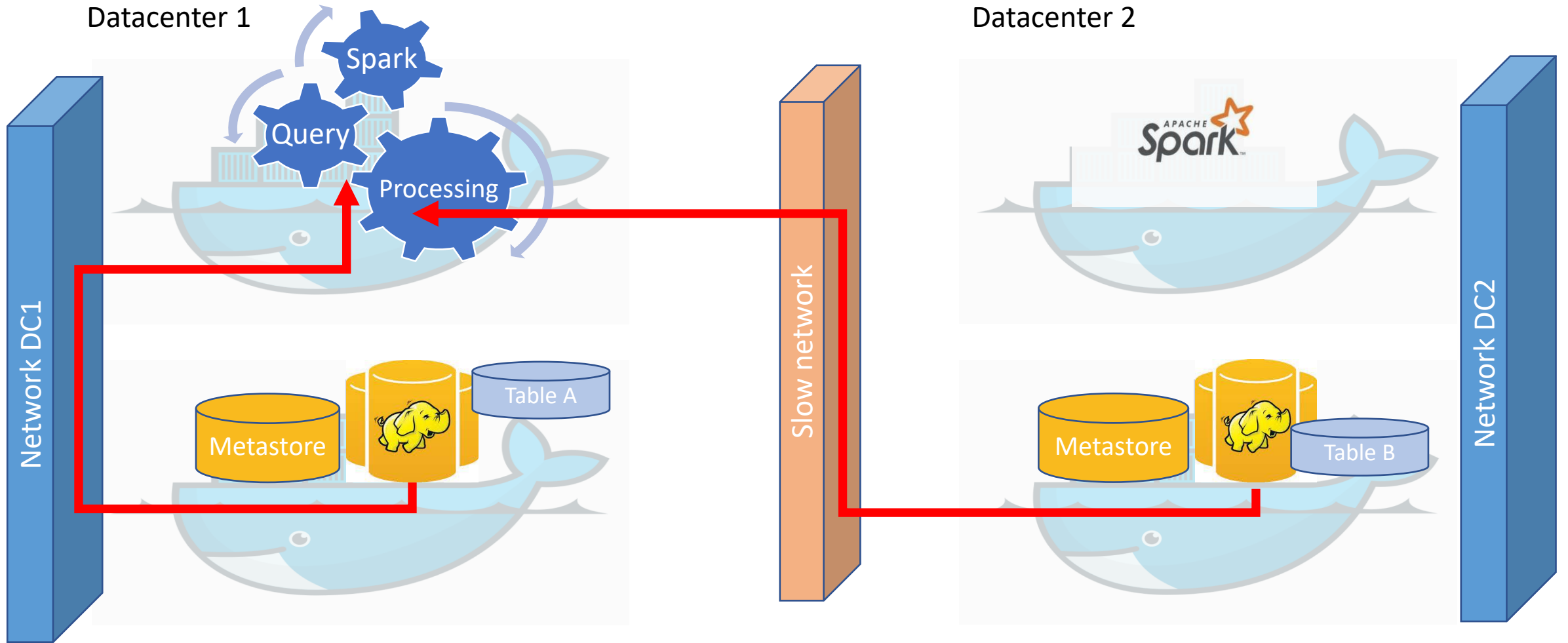


Qflock

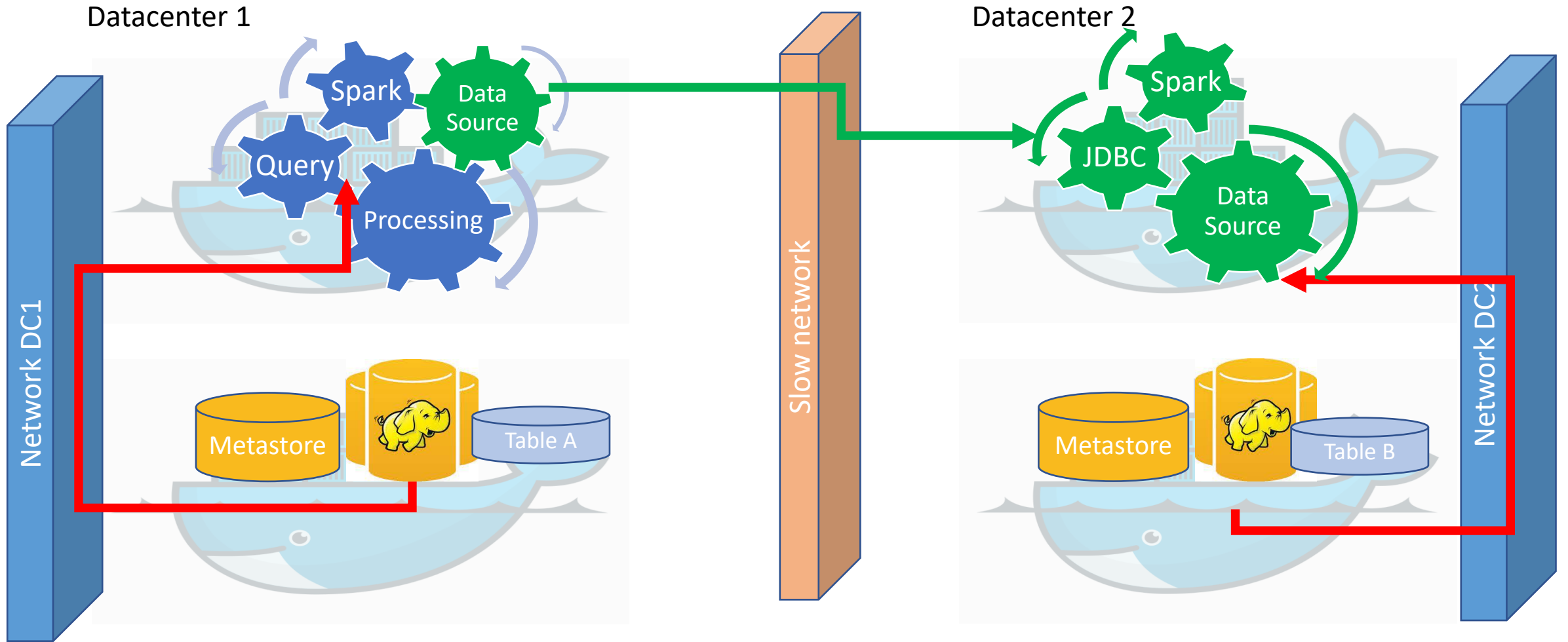
May 2022



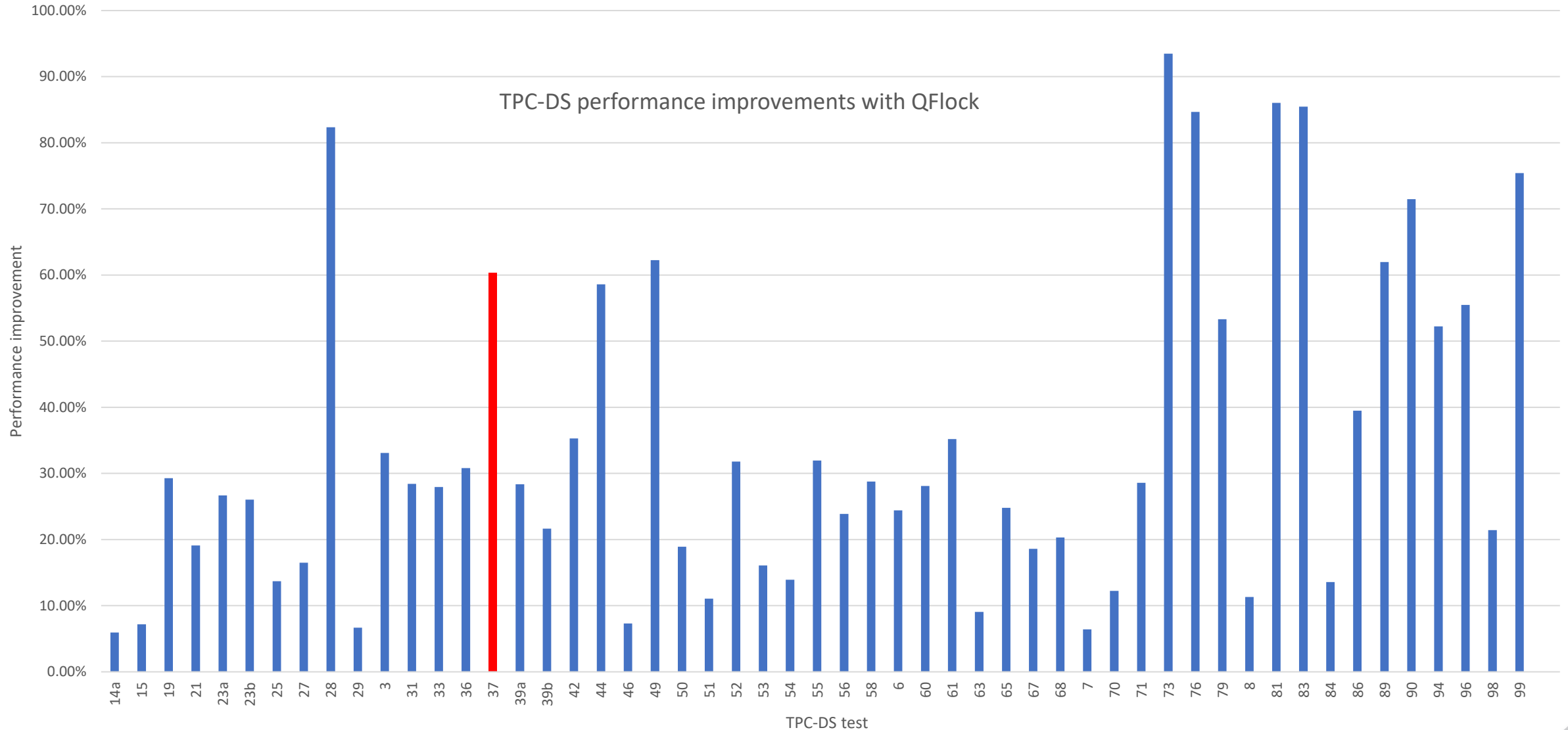
Perform TPC-DS query (baseline)



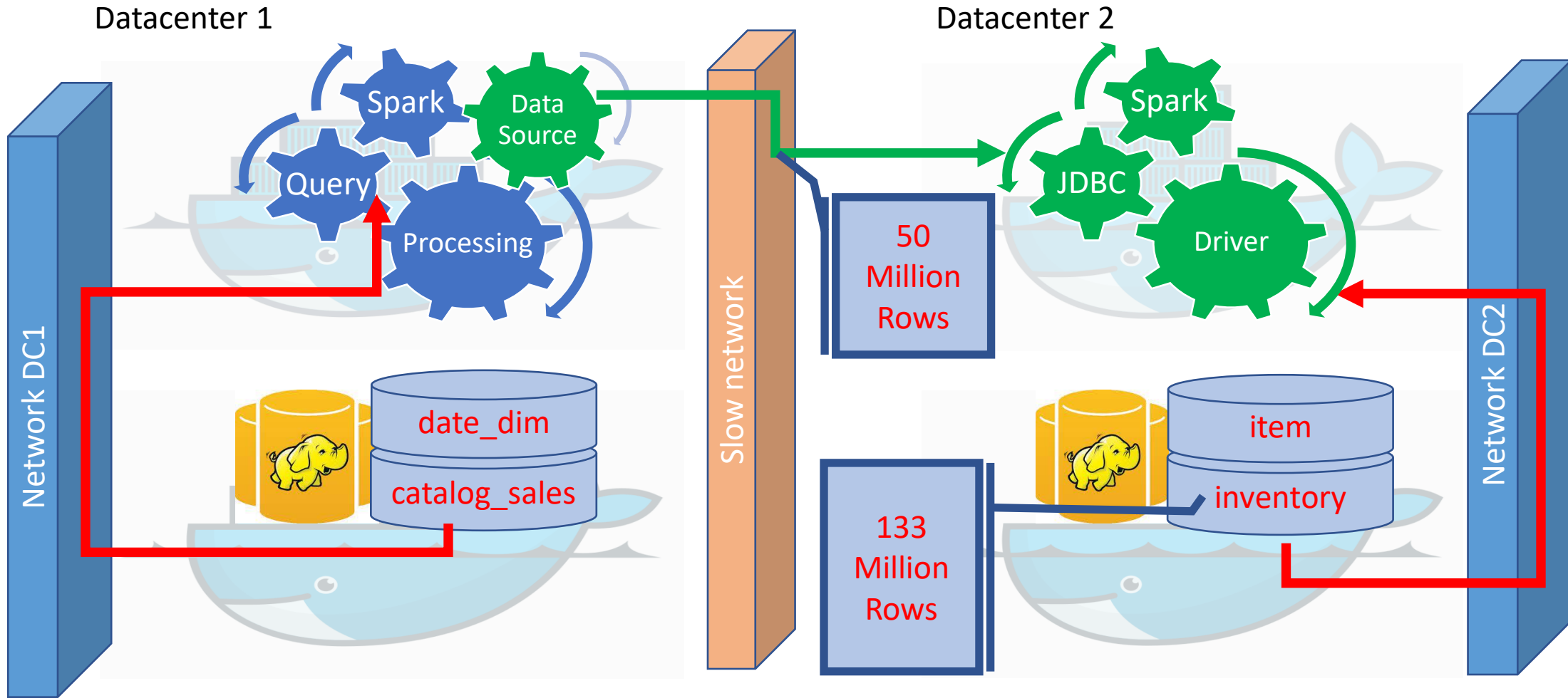
Perform TPC-DS query (QFlock)



QFlock performance improvements



QFlock data reduction explanation (Test 37)



Test 37 Logical Plan

Spark Logical Plan (baseline)

```
Sort [i_item_id#146 ASC NULLS FIRST], true, Statistics(sizeInBytes=314.0 B
+- Aggregate [i_item_id#146, i_item_desc#149, i_current_price#150], [i_ite
  +- Project [i_item_id#146, i_item_desc#149, i_current_price#150], Stati
    +- Join Inner, (cs_item_sk#220L = i_item_sk#145L), Statistics(sizeIn
      :- Project [i_item_sk#145L, i_item_id#146, i_item_desc#149, i_cur
        : +- Join Inner, (d_date_sk#29L = inv_date_sk#4741L), Statistics
          :   :- Project [i_item_sk#145L, i_item_id#146, i_item_desc#149,
            :     : +- Join Inner, (inv_item_sk#4742L = i_item_sk#145L), Sta
              :       :- Project [i_item_sk#145L, i_item_id#146, i_item_des
                :         +- Filter ((isnotnull(i_current_price#150) AND (((
                  :           +- Relation tpcds.item[i_item_sk#145L,i_item_id
                :       +- Project [inv_date_sk#4741L, inv_item_sk#4742L], St
                  :         +- Filter ((isnotnull(inv_quantity_on_hand#4744L)
                    :           +- Relation tpcds.inventory[inv_date_sk#4741L,i
                :     +- Project [d_date_sk#29L], Statistics(sizeInBytes=1141.4 K
                  :       +- Filter ((isnotnull(d_date#31) AND ((cast(d_date#31 as
                    :         +- Relation tpcds.date_dim[d_date_sk#29L,d_date_id#30
                +- Project [cs_item_sk#220L], Statistics(sizeInBytes=219.7 MiB, r
                  +- Filter isnotnull(cs_item_sk#220L), Statistics(sizeInBytes=3
                    +- Relation tpcds.catalog_sales[cs_sold_date_sk#205L,cs_sol
```

Spark Logical Plan (QFlock)

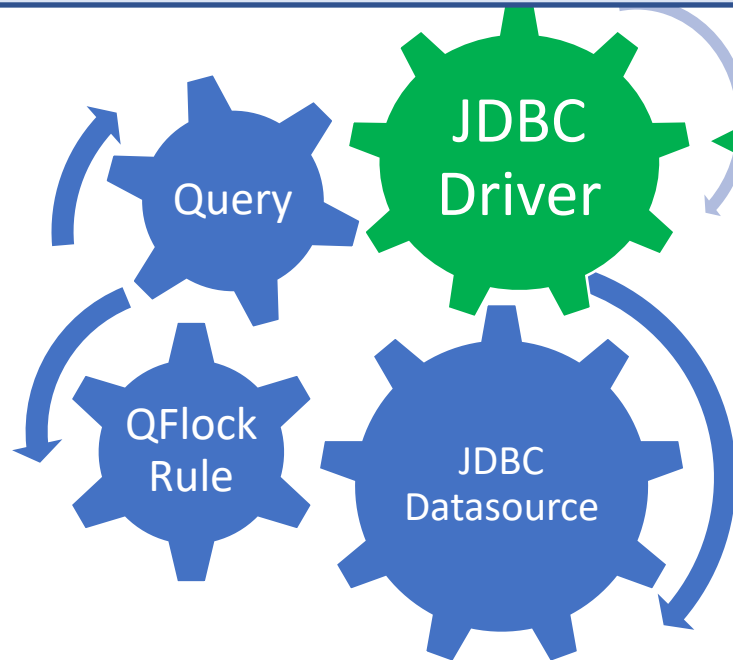
```
Sort [i_item_id#285 ASC NULLS FIRST], true, Statistics(sizeInBytes=4.17E+2
+- Aggregate [i_item_id#285, i_item_desc#288, i_current_price#289], [i_ite
  +- Project [i_item_id#285, i_item_desc#288, i_current_price#289], Stati
    +- Join Inner, (cs_item_sk#450L = i_item_sk#284L), Statistics(sizeIn
      :- Project [i_item_sk#284L, i_item_id#285, i_item_desc#288, i_cur
        : +- Join Inner, (d_date_sk#29L = inv_date_sk#8715L), Statistics
          :   :- Project [i_item_sk#284L, i_item_id#285, i_item_desc#288,
            :     : +- Join Inner, (inv_item_sk#8716L = i_item_sk#284L), Sta
              :       :- RelationV2[i_item_sk#284L, i_item_id#285, i_item_d
                :         +- RelationV2[inv_date_sk#8715L, inv_item_sk#8716L] c
                :       +- Project [d_date_sk#29L], Statistics(sizeInBytes=1141.4 K
                  :         +- Filter ((isnotnull(d_date#31) AND ((cast(d_date#31 as
                    :           +- Relation tpcds.date_dim[d_date_sk#29L,d_date_id#30
                +- Project [cs_item_sk#450L], Statistics(sizeInBytes=219.7 MiB, r
                  +- Filter isnotnull(cs_item_sk#450L), Statistics(sizeInBytes=3
                    +- Relation tpcds.catalog_sales[cs_sold_date_sk#435L,cs_sol
```



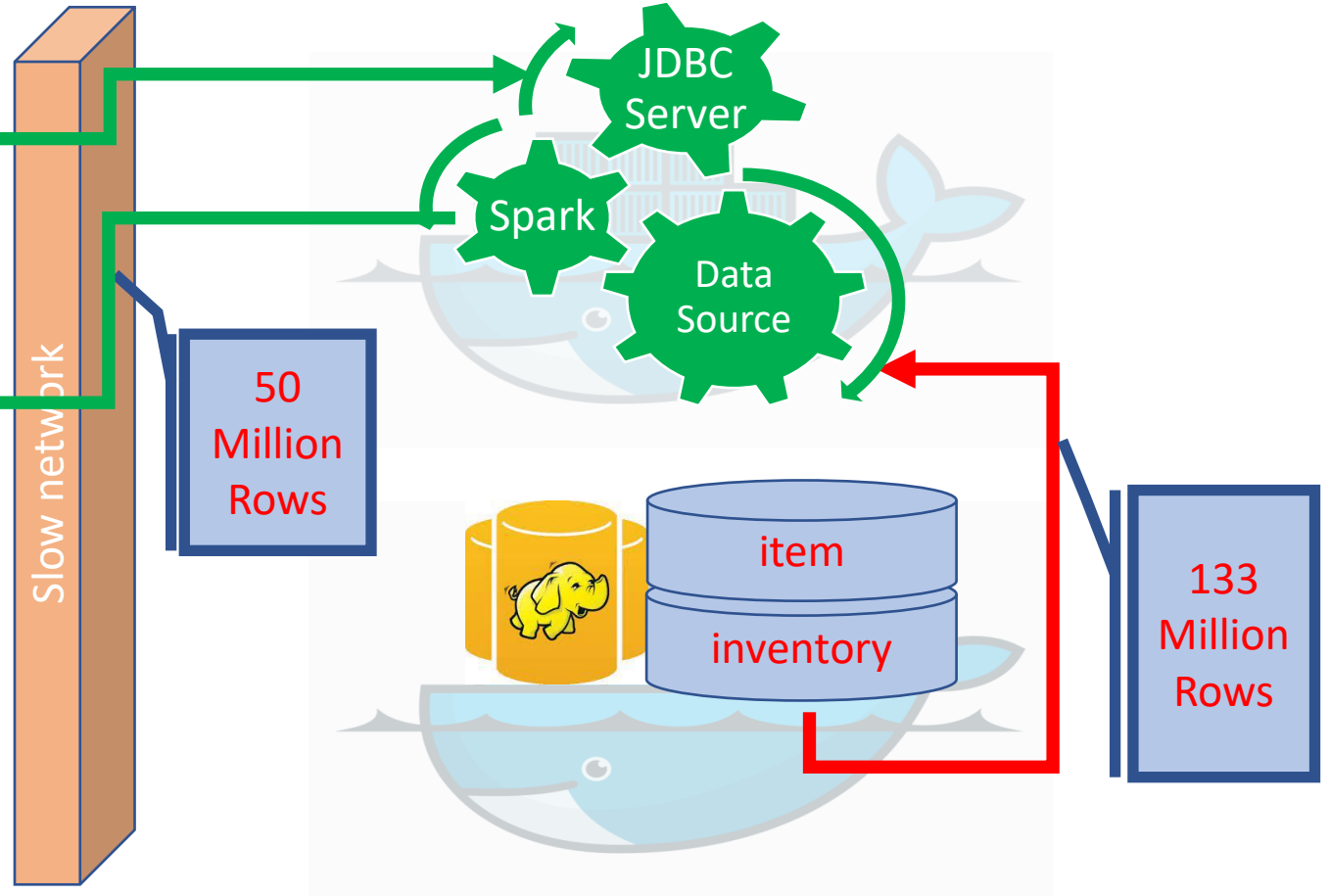
Test 37 JDBC query

Datacenter 1

```
SELECT inv_date_sk,inv_item_sk  
FROM inventory_0  
WHERE inv_quantity_on_hand >= '100' AND  
inv_quantity_on_hand <= '500'
```



Datacenter 2



Happy hacking!!!

- If you have any questions, problems or suggestions
- Please do not hesitate to create an issue at:
- <https://github.com/open-infrastructure-labs/QFlock/issues>

