CS422 Databese Systems - Project Report

Wentao Feng

March 25, 2019

Instructions:

- For column layout, the materialization is as early as possible.
- For PAX data layout, the number of tuples per page is 100.
- For volcano-style vector-at-a-time engine, the vector size is 1000.
- Data loading time is excluded.
- For each test, the maximum running time is restricted within 60 seconds.
- Other irrelevant variables are controlled as much as possible.

1 Selection, Projection

1.1 SQL Description

```
SELECT L.L_ORDERKEY, L.L_PARTKEY, L.L_SUPPKEY
FROM lineitem L
WHERE L.L_ORDERKEY <= 100
```

1.2 Results

Data Layout	NSM	DSM	DSM	PAX
Execution Unit	Tuple	Column	Vector	Tuple
Execution Time(ms)	2496	>60000	41316	2402

2 Selection, Projection, Join

2.1 SQL Description

```
SELECT L.L_TAX, O.O_TOTALPRICE
FROM lineitem L, orders O
WHERE L.L_ORDERKEY = O.O_ORDERKEY AND O.O_ORDERKEY <= 150
```

2.2 Results

Data Layout	NSM	DSM	DSM	PAX
Execution Unit	Tuple	Column	Vector	Tuple
Execution Time(ms)	2278	>60000	12934	2394

3 Selection, Projection, Join, Aggregation

3.1 SQL Description

```
SELECT SUM(L.L_TAX)
FROM lineitem L, orders O
WHERE L.L_ORDERKEY = O.O_ORDERKEY AND O.O_ORDERKEY <= 80
GROUPBY L.L_SHIPMODE
```

3.2 Results

Data Layout	NSM	DSM	DSM	PAX
Execution Unit	Tuple	Column	Vector	Tuple
Execution Time(ms)	1992	>60000	12905	1892

4 Analysis

4.1 NSM & PAX

The performances of NSM and PAX are critically close. Theoretically, NSM should query faster than PAX, because the PAX layout consumes a bit more materialization time than NSM layout. However, when the query size is not very big, the difference may not be obvious. They both work better than DSM. This result is unconventional, as volcano style engine has calls overhead. Additional time consumed by column-based execution engines may come from repetitive materialization.

4.2 Columnar & Vector engine

The columnar engine should work best among all of them because every execution of it loads and returns all data at once. In our tests, it works worse, as it consumes much time on materialization. Consequently, the vector engine works slightly better than columnar, since a vector is only part of a column.