BlinkDB - Queries with Bounded Errors and Bounded Response Times on Very Large Data

Paweł Kamiński

18 listopada 2015

Problem

Situation

Huge database consists of one table.

Problem

Situation

Huge database consists of one table.

Blockers

Query execution time.

Space taken by database.

Example

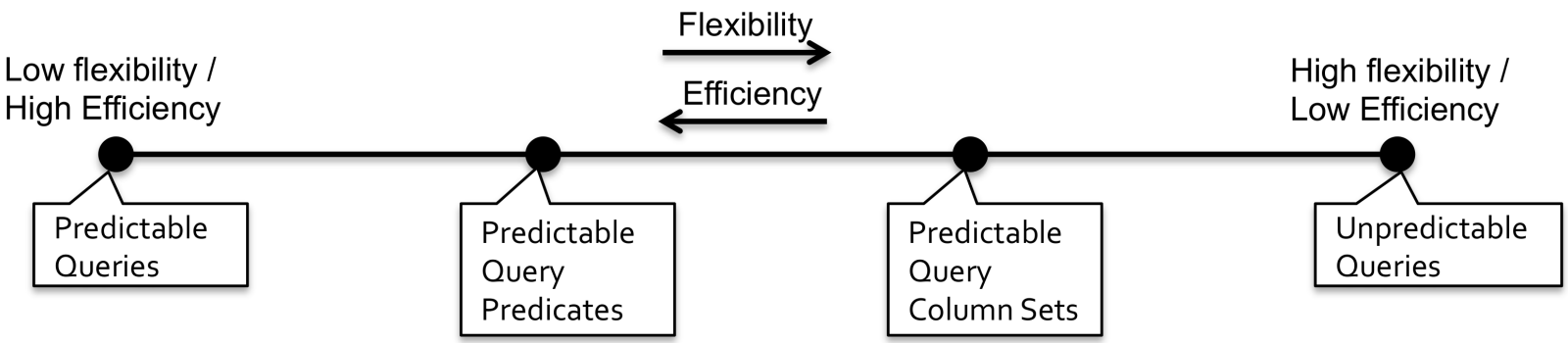
```
SELECT AVG(SessionTime)
FROM Sessions
WHERE city='London'

SELECT AVG(SessionTime)
FROM Sessions
WHERE city='Pontefract'
```

Solution

Average case

In most cases approximated answer is enough. So let's work on samples.



Query Column Set

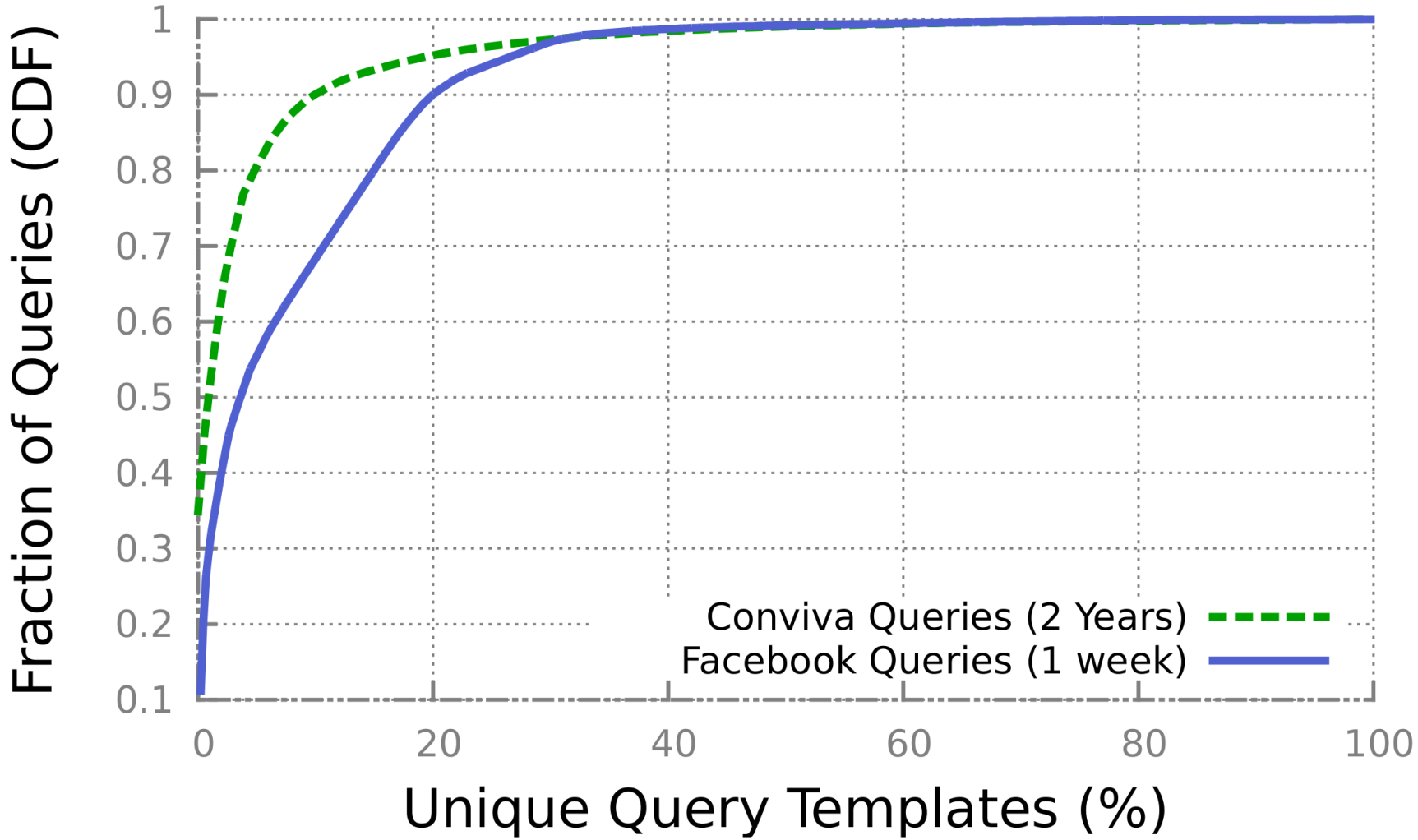
Query Column Set

The set of columns used by queries in WHERE, GROUP BY, HAVING.

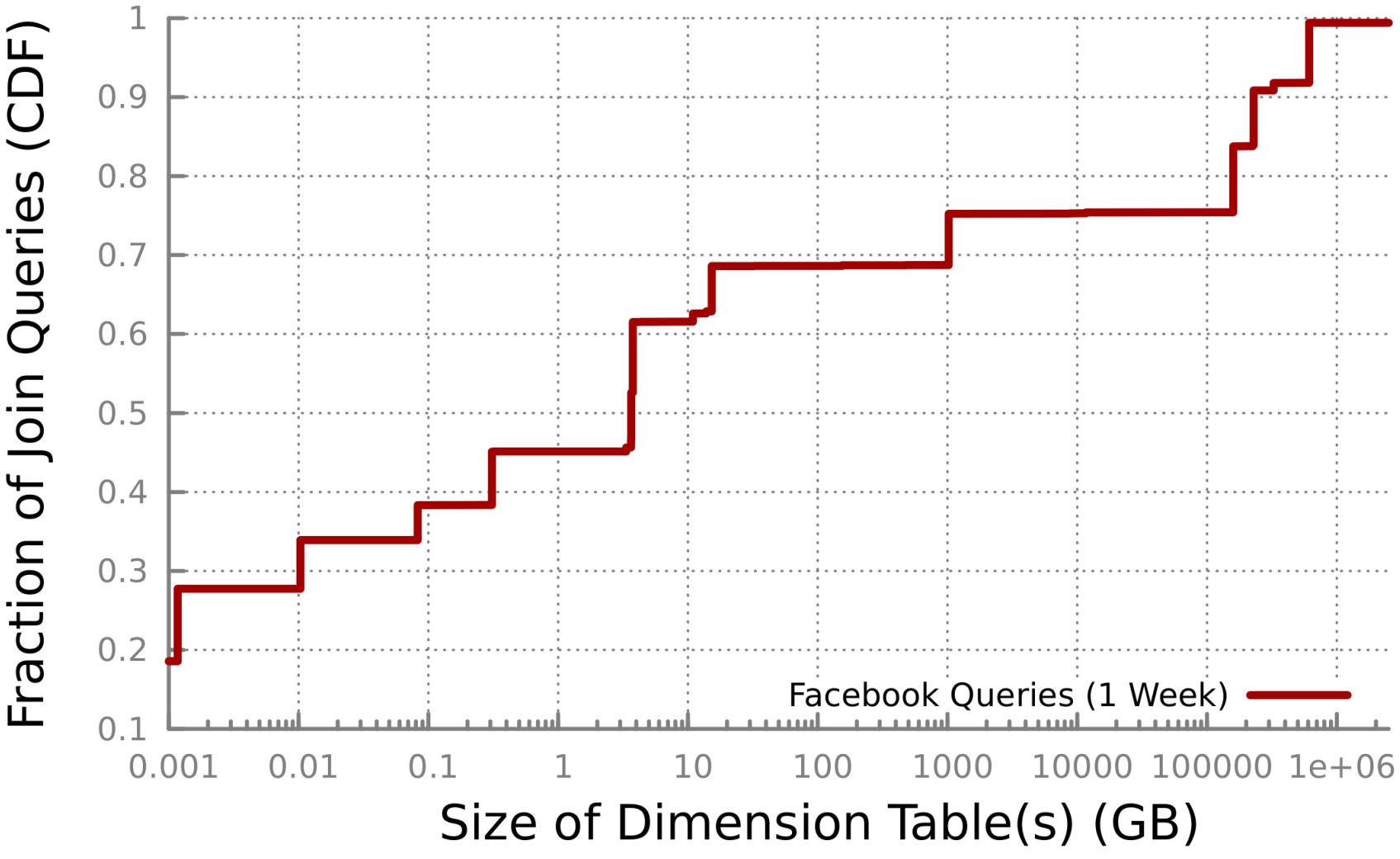
Approaches

- Uniform Sampling
- Online Aggregation
- Sketches

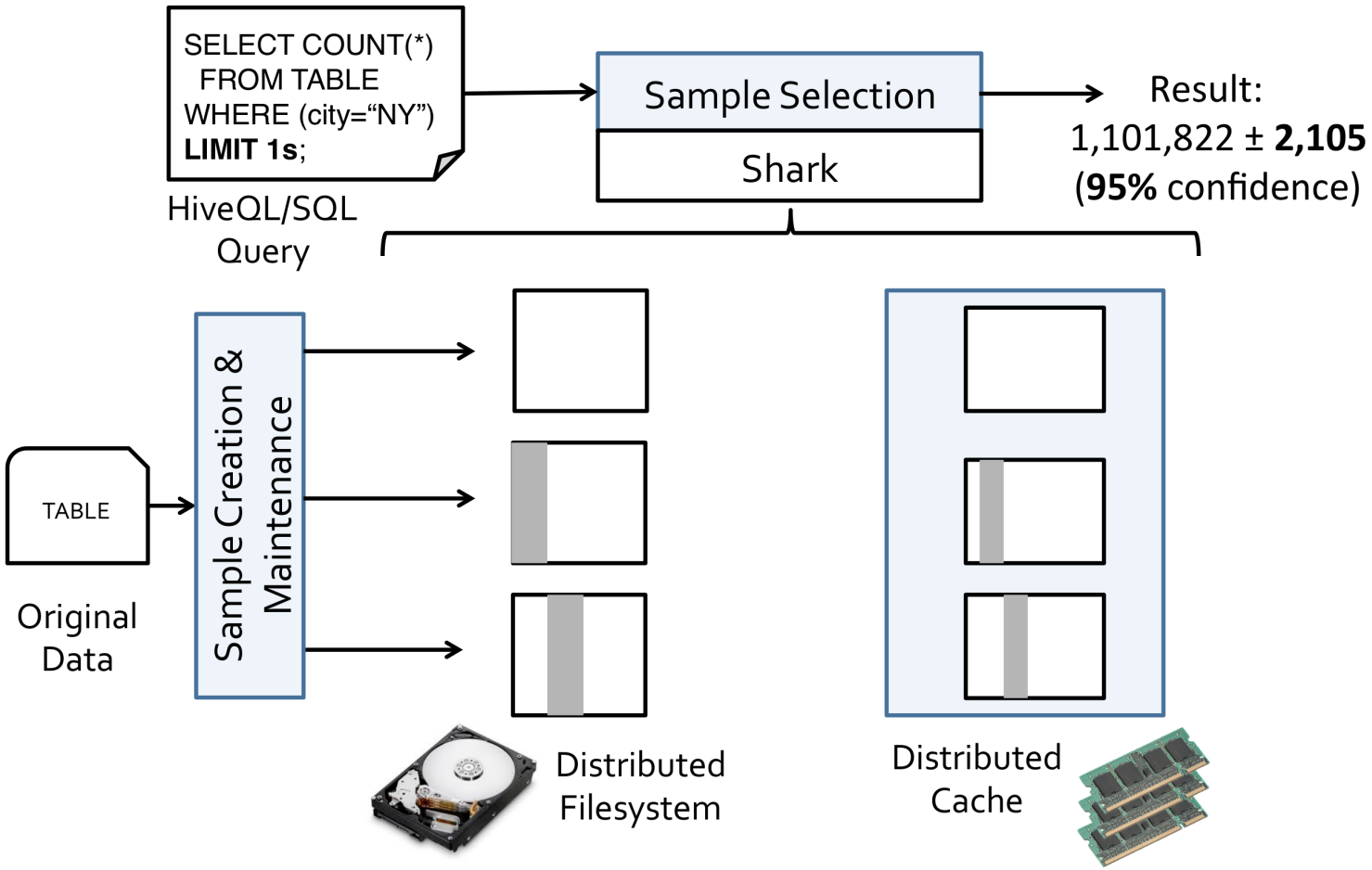
unique - all


Fraction of Queries (CDF)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
20
40
60
80
100
Unique Query Templates (%)
Conviva Queries (2 Years)
Facebook Queries (1 week)

Big joins


Fraction of Join Queries (CDF)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.001
0.01
0.1
1
10
100
1000
10000
100000
1e+06
Size of Dimension Table(s) (GB)
Facebook Queries (1 Week)

BlinkDB architecture



Example

```
SELECT COUNT(*)
FROM Sessions
WHERE city='London'
GROUP BY OS
ERROR WITIHIN 10% AT CONFIDENCE 95%
/WITHIN 5 SECONDS
```

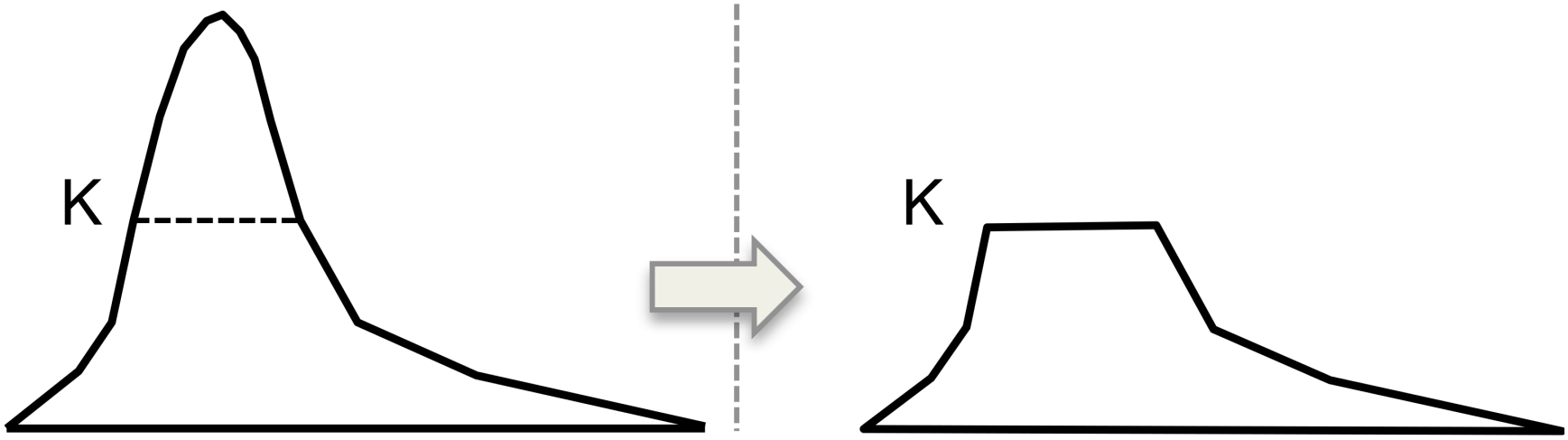
Samples

- Uniform Samples
- Single-Column Stratified Samples
- Multi-Column Stratified Samples

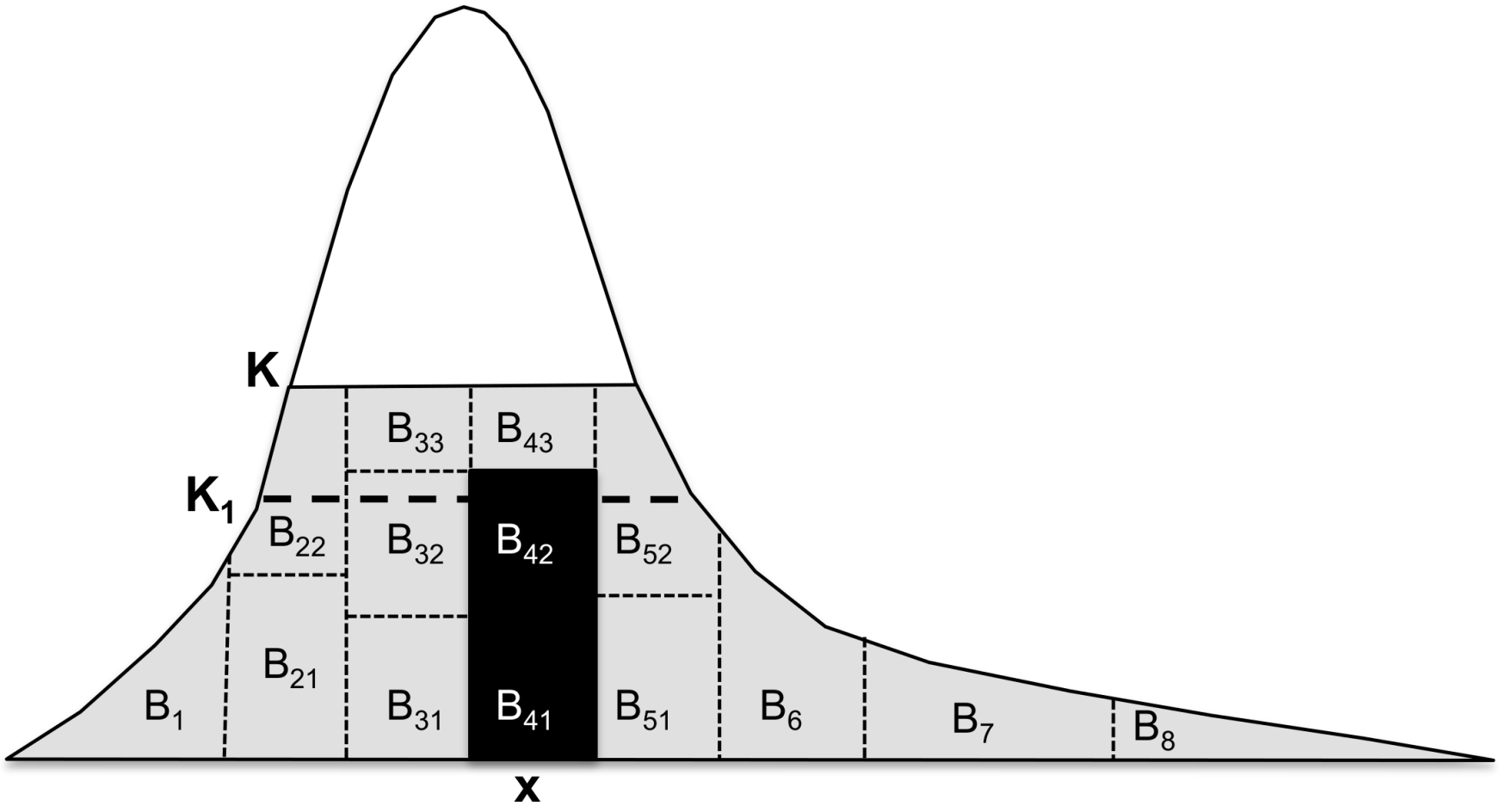
Samples

- Uniform Samples GROUP
- Single-Column Stratified Samples efficiency
- Multi-Column Stratified Samples

Stratified Samples - cutting



Machine memory spacing



Choose QCS

Choose QCS

- Use less memory than we have in whole system
- Coverage as many query QCS as possible
- Try to not sample big groups

Selecting Sample

- satisfiability vs all in memory
- disk vs memory

Selecting Right Sample

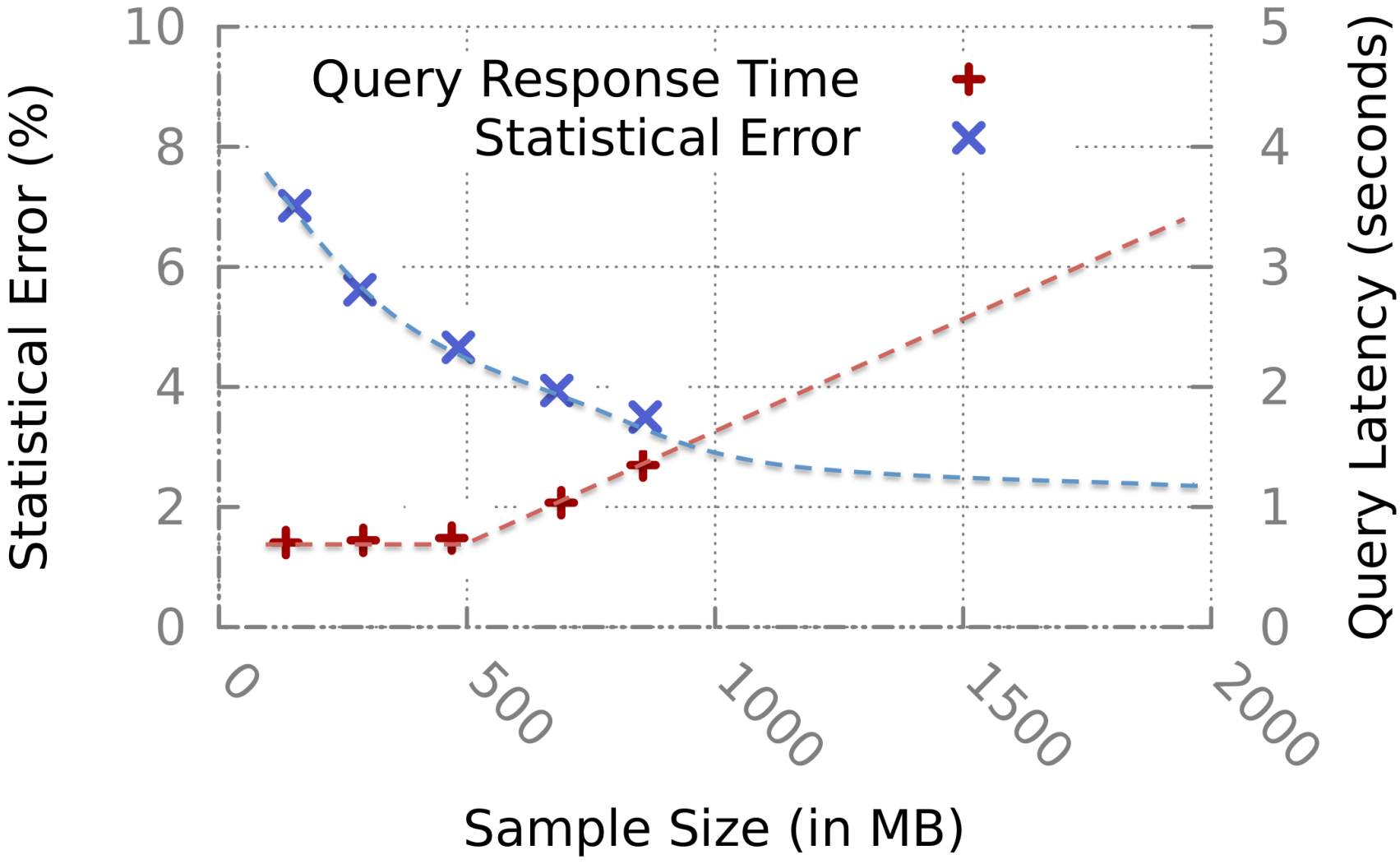
- Error Profile - variance $\sim \frac{1}{numberofrows}$
- Latency Profile - I/O scales lineary + initial cost

Example

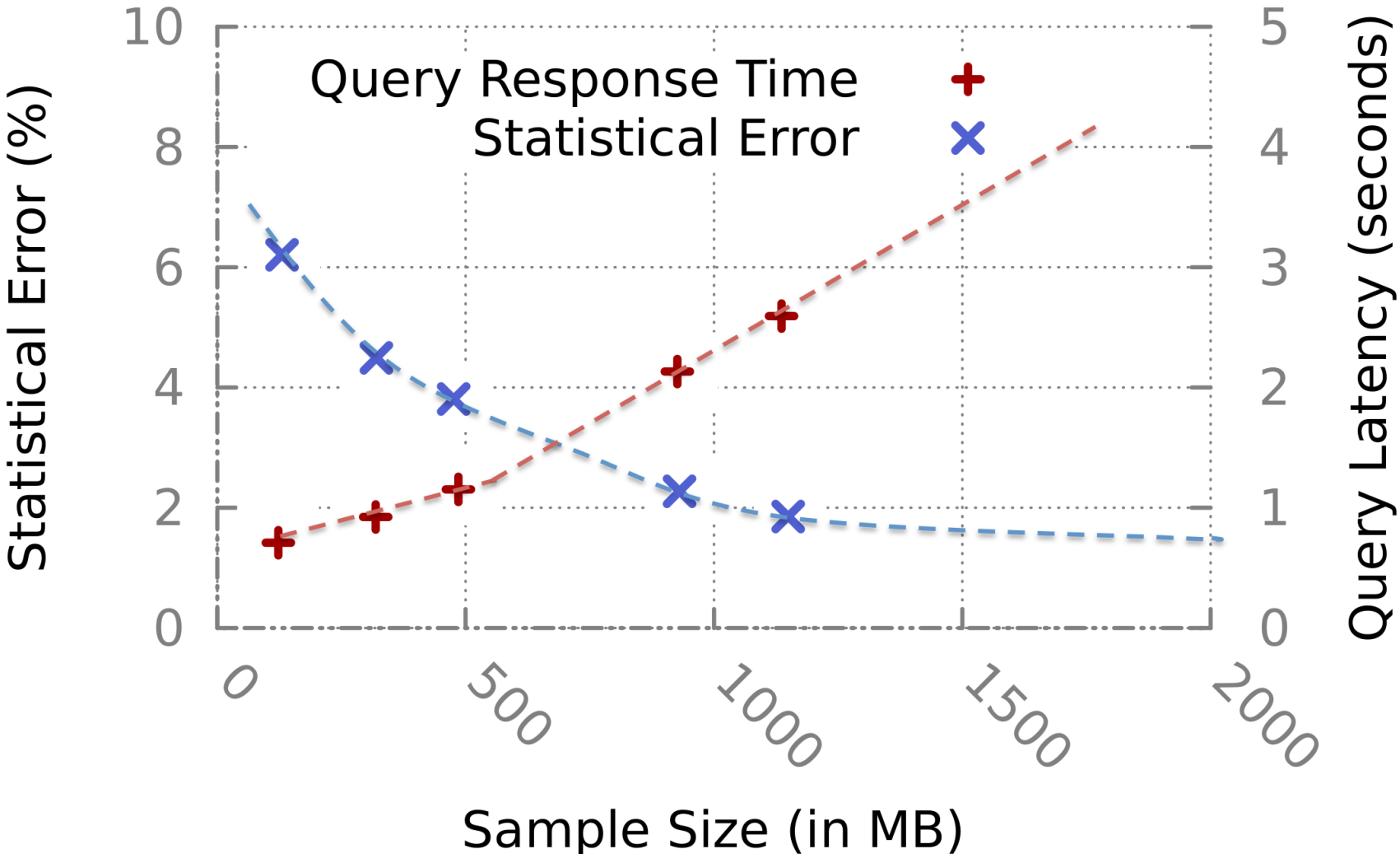
Task: get average todays connection time in Pontefract
QCS:

- date/country
- date/designated media
- date/ended flag

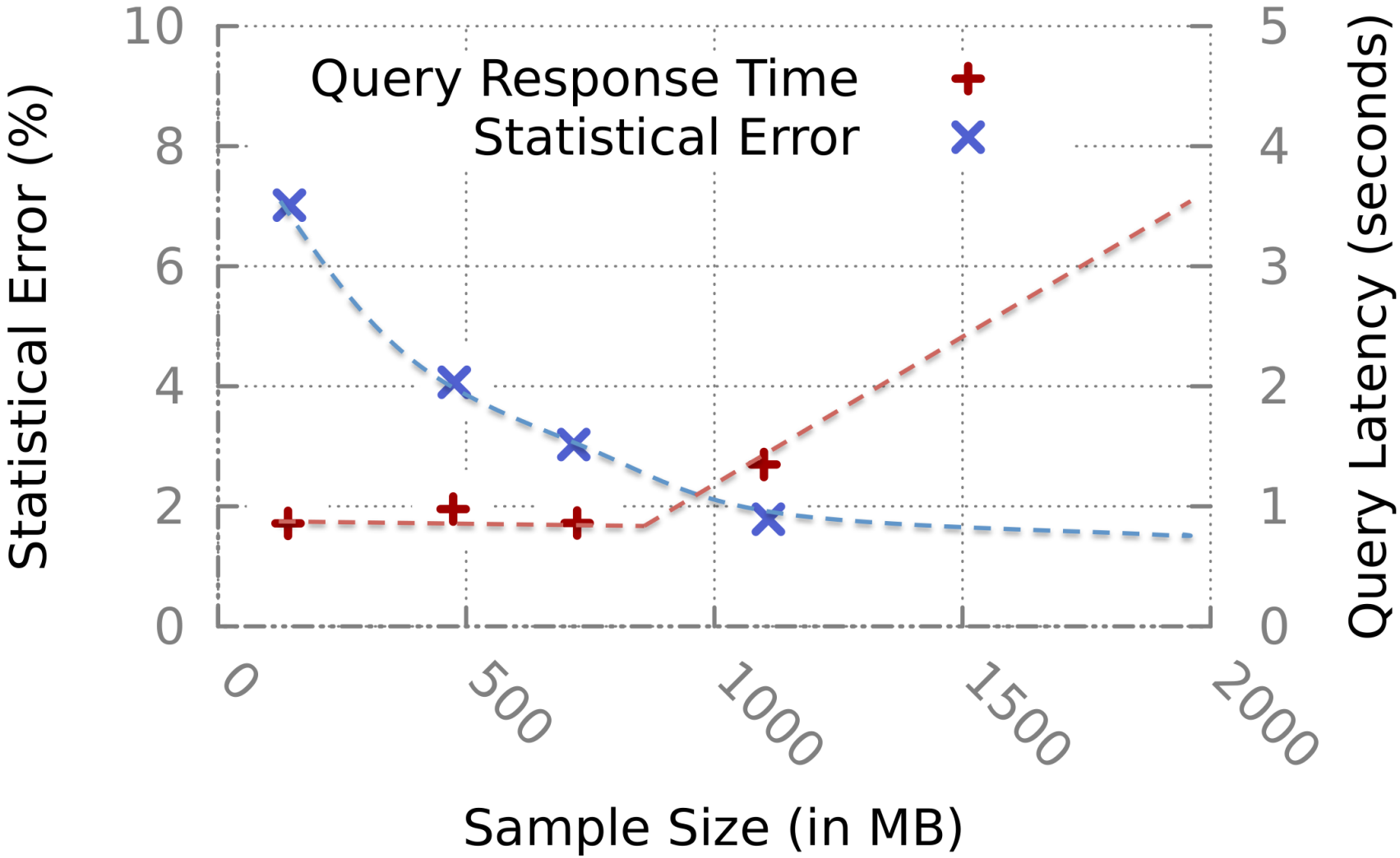
date/country



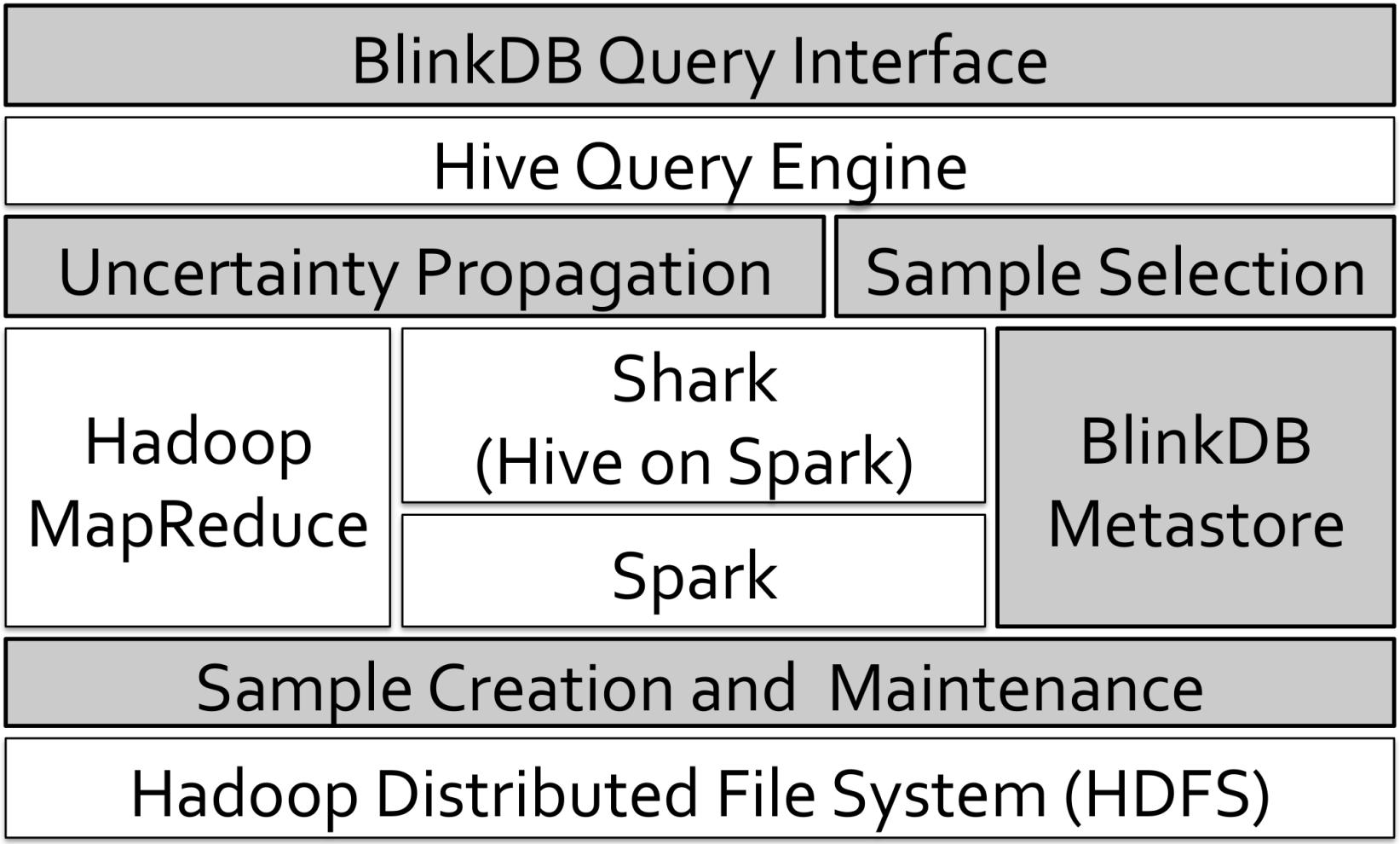
date/designated media



date/ended flag



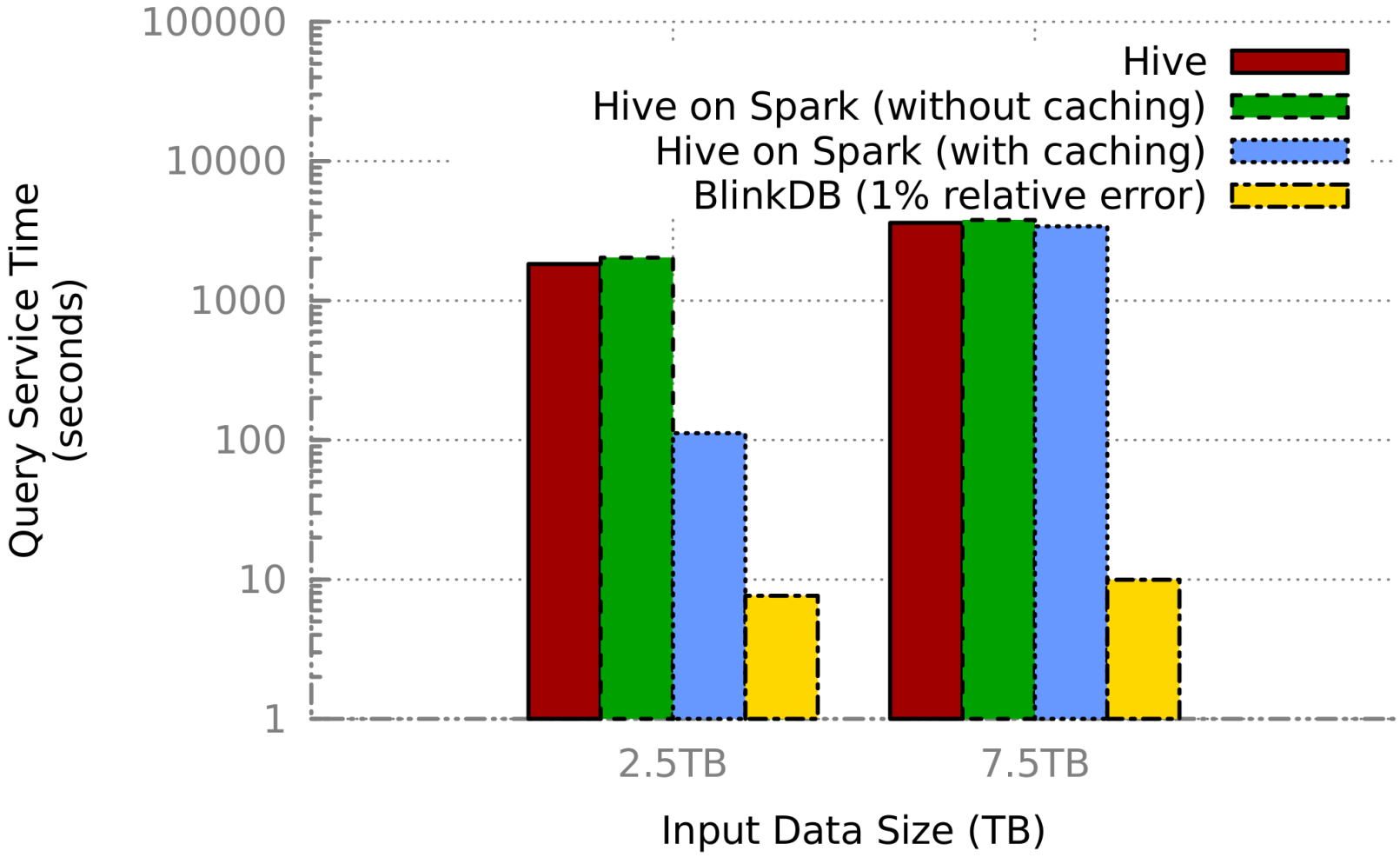
Implementation stack



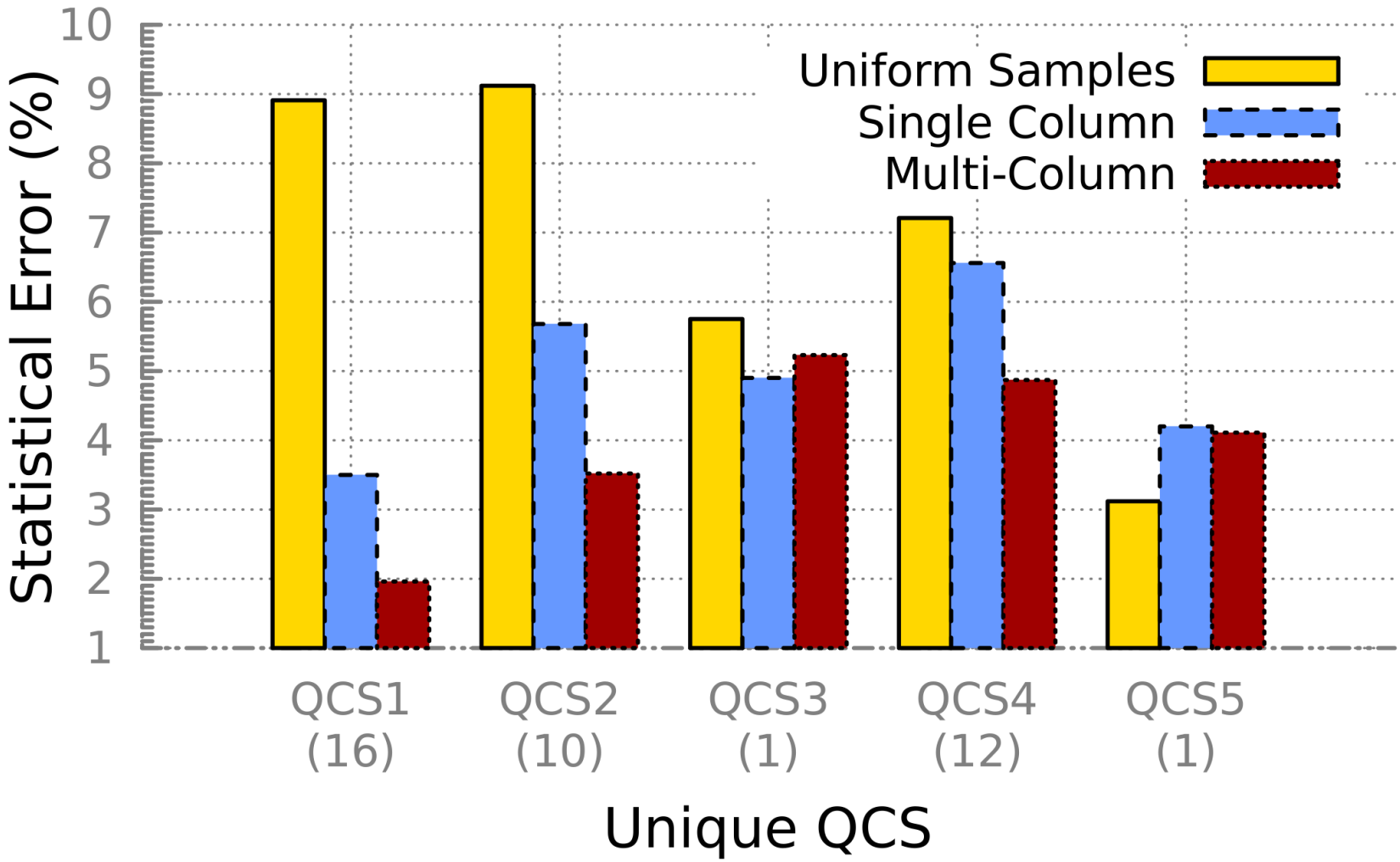
Resources - tests

- EC2 100 nodes 8 CPU 2,66 GHZ 70 GB RAM and 800 GB disc, what gives 75 TB dics storage and 6TB distributed RAM cache.
- Tests - Conviva 17TB, TPC-H 1TB
- Samples +50% memory

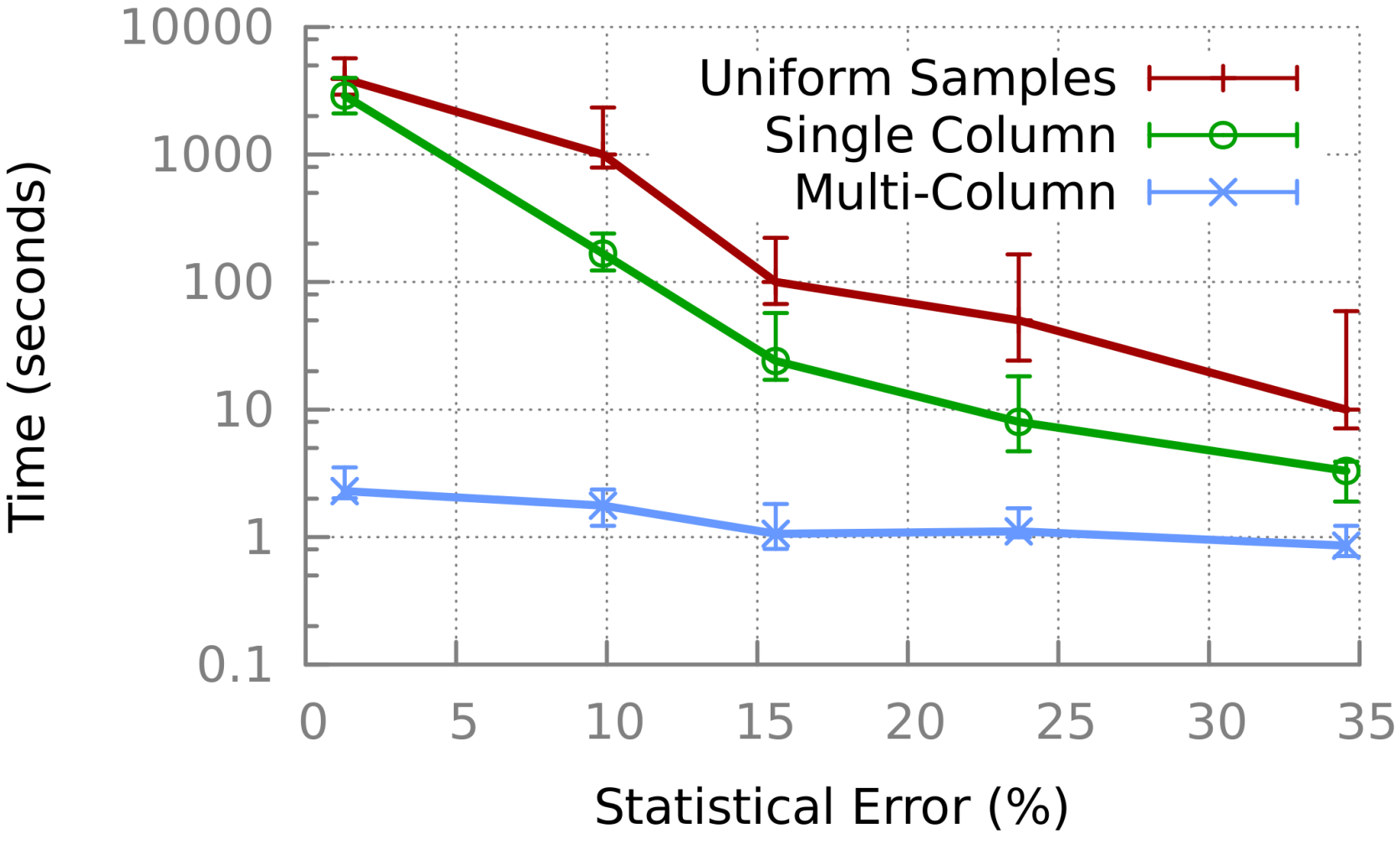
BlinkDB vs No sampling



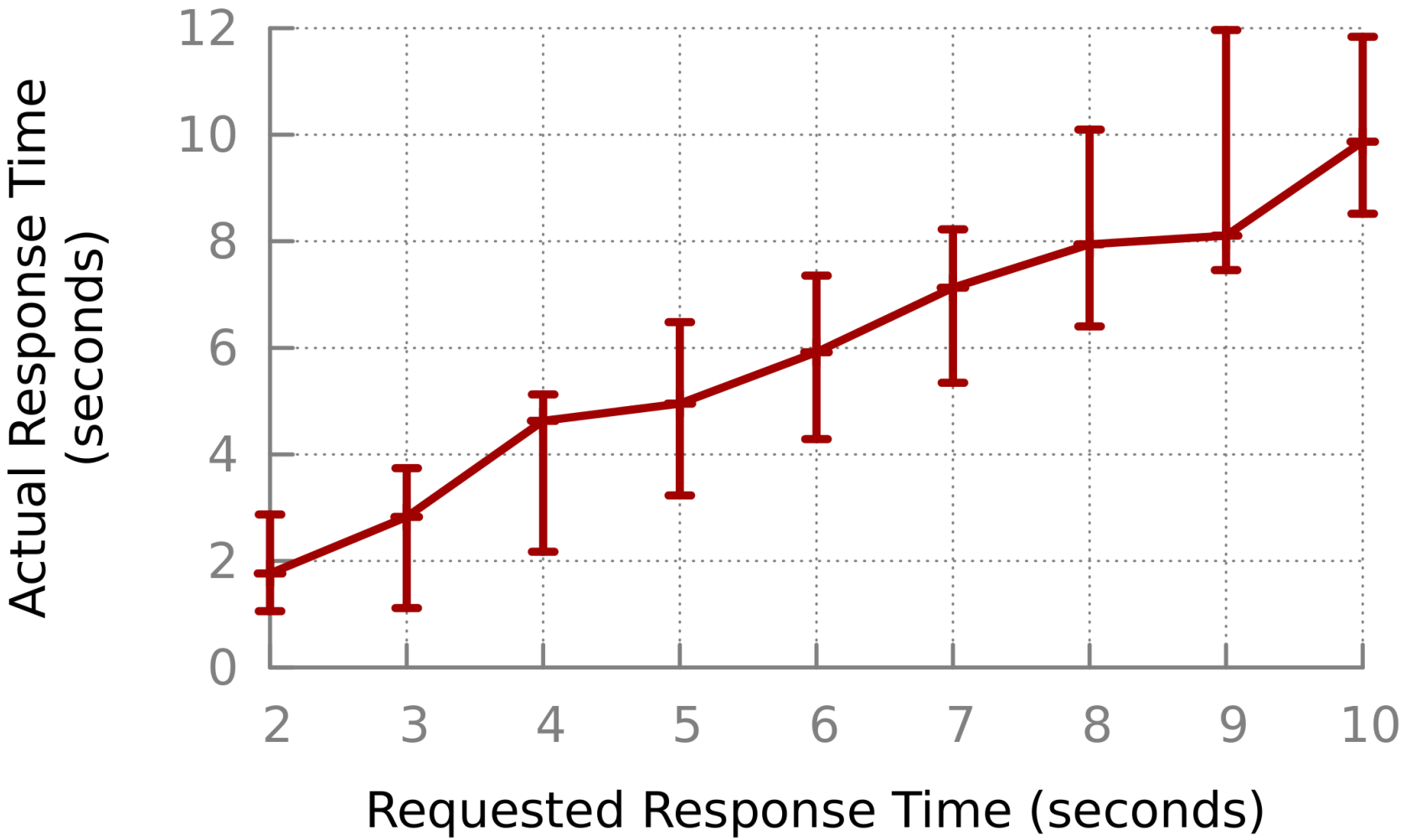
Error comparition (Conviva)



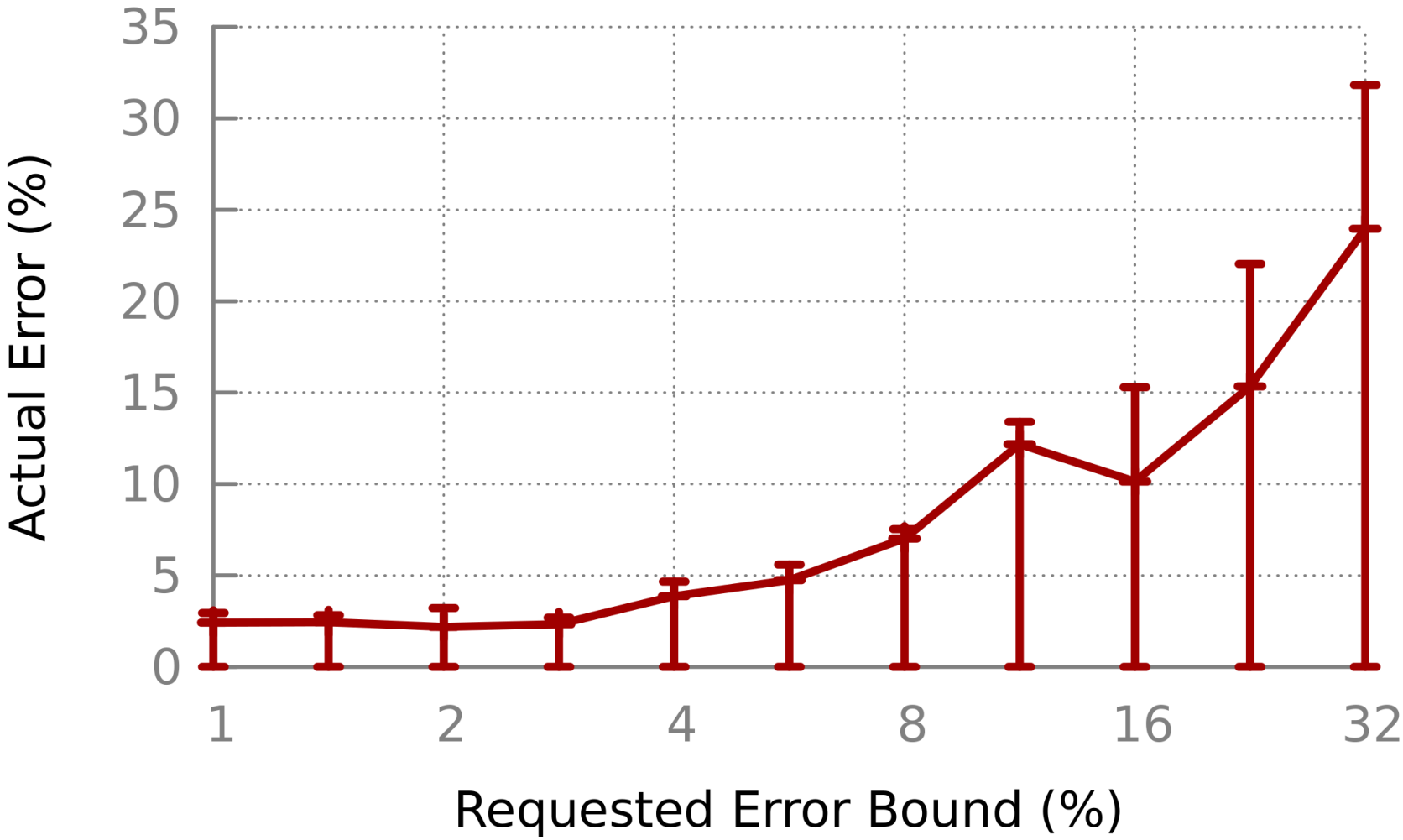
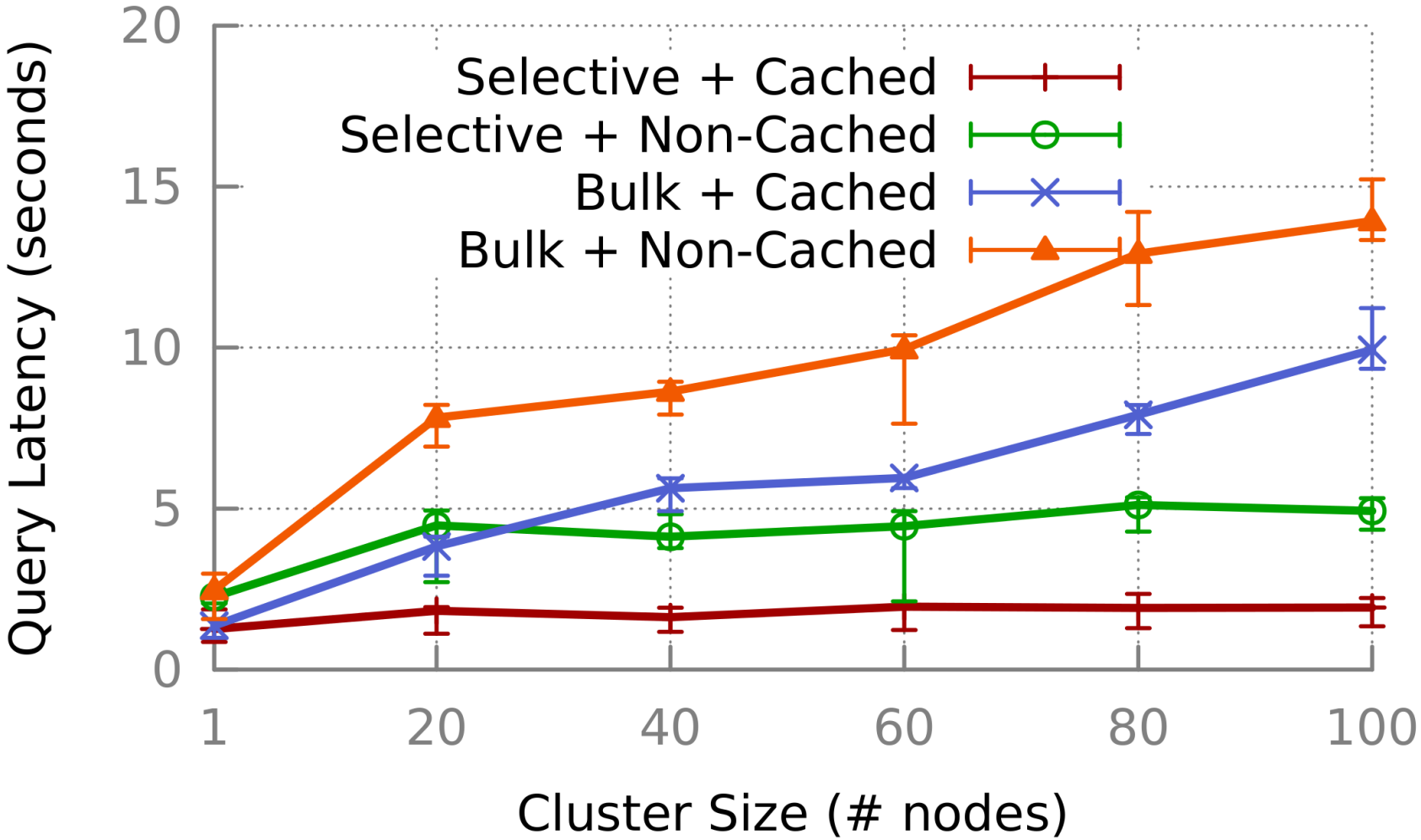
Error Convergence



Benchmark - scaling up



Benchmark - scaling up



Benchmark - scaling up