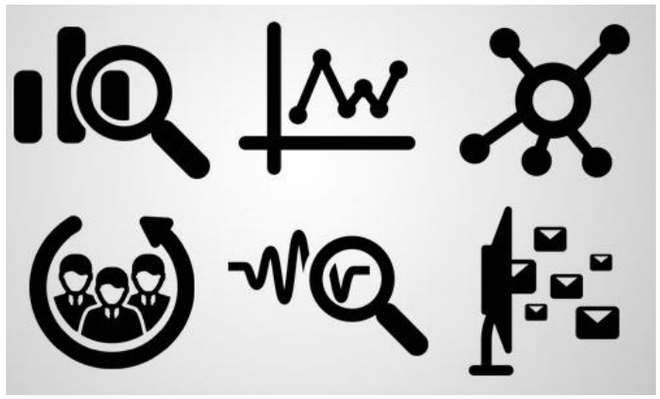
Bohr: Similarity Aware Geo-distributed Data Analytics

Hangyu Li, Hong Xu, Sarana Nutanong
City University of Hong Kong

Big Data Analytics



Analysis

Generate

Data are geo-distributed

Centralizing data is infeasible

- Moving data through wide area networks(WANS) can be extremely slow.
- Data sovereignty laws.

1. Vulimiri et al., NSDI'15
2. Pu et al., SIGCOMM'15
3. Viswanathan et al., OSDI'16
4. Hsieh et al., NSDI'17

Processing queries in-place

Bottleneck site exists.

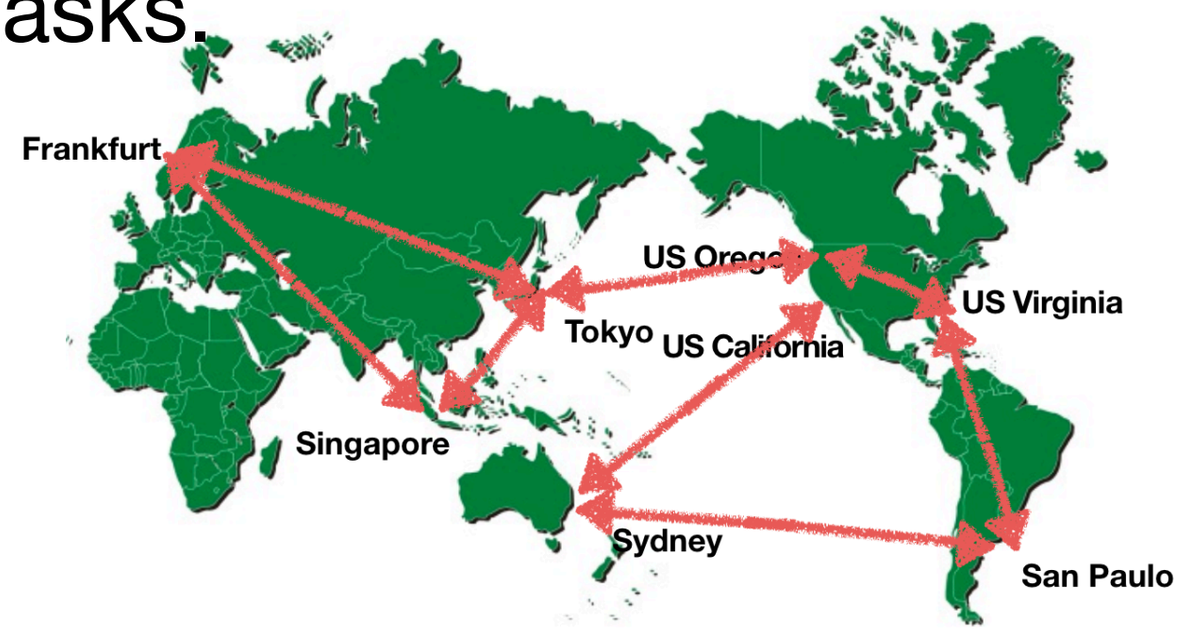
Processing queries in-place

System	Workload	Task / Data Placement	Optimize aspect
JetStream	streaming	no	WAN bandwidth usage
Geode	batch	data	WAN bandwidth usage
Iridium	batch	both	query response time

Processing queries in-place

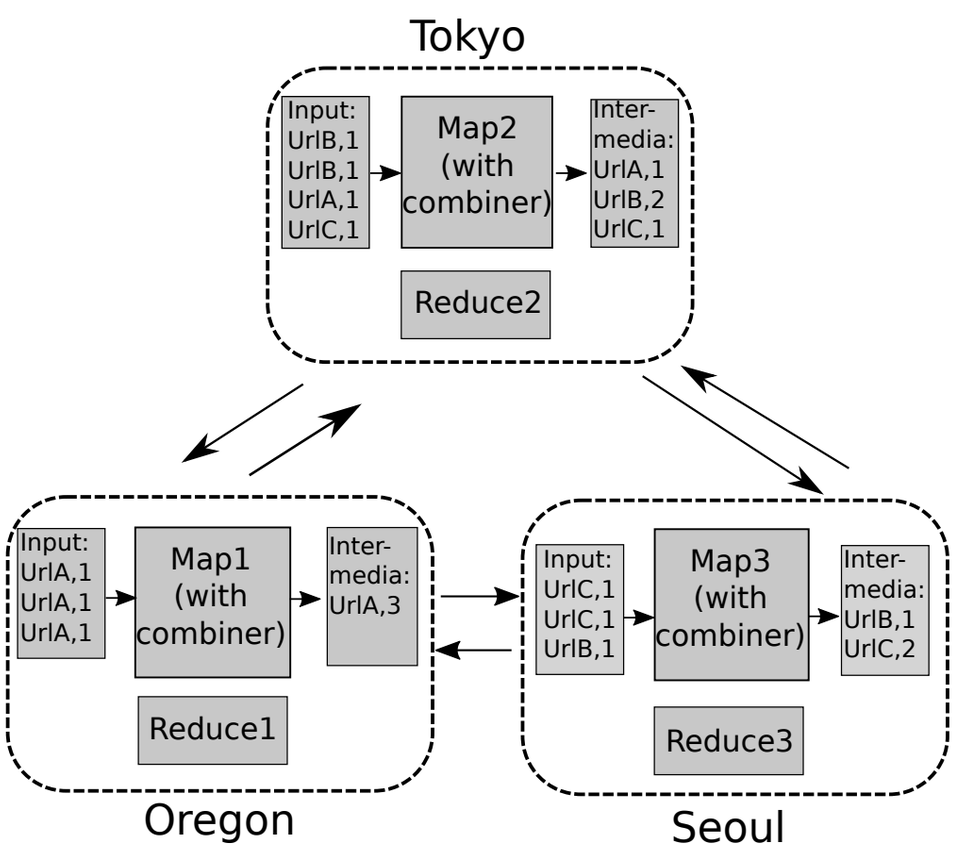
Iridium (Pu et al. SIGCOMM'15)

- Recurring queries, abundant computation resource.
- Transfer data out from the bottleneck site.
- Re-schedule reducer tasks.



Iridium can be improved

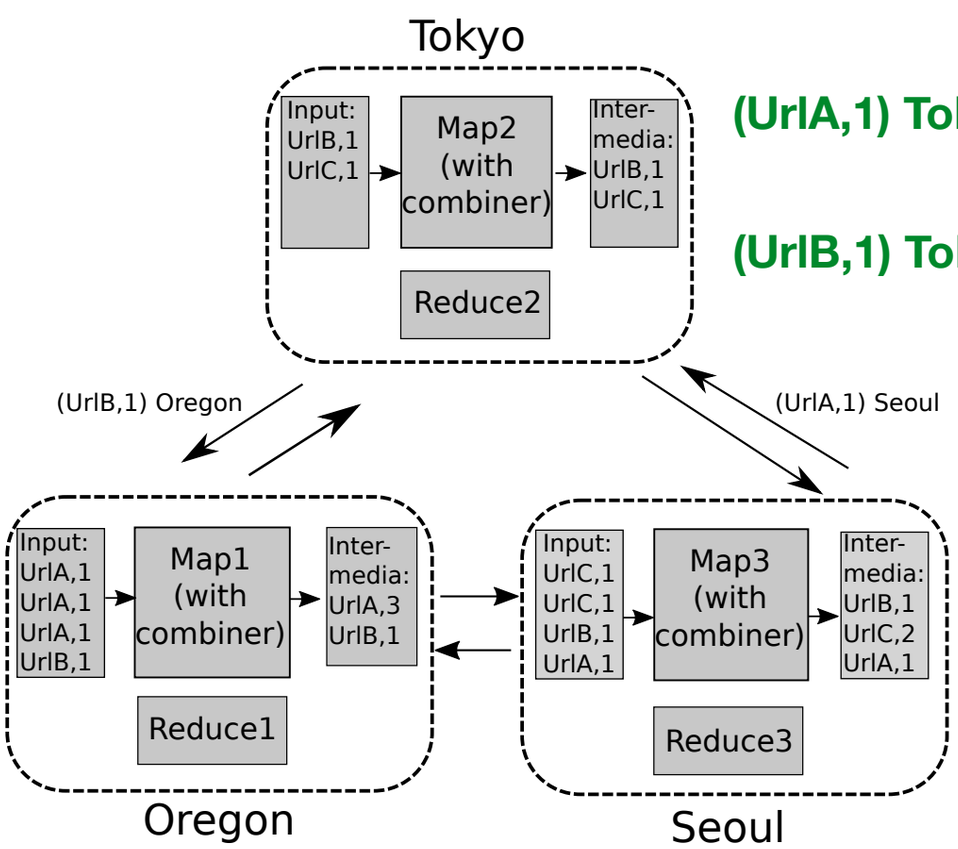
- **similarity oblivious**
- **data reduction ratio** is **constant**



Site	Intermed-iate record
Tokyo	3
Oregon	1
Seoul	2
Total	6

Iridium can be improved

- **similarity oblivious**
- **data reduction ratio** is **constant**



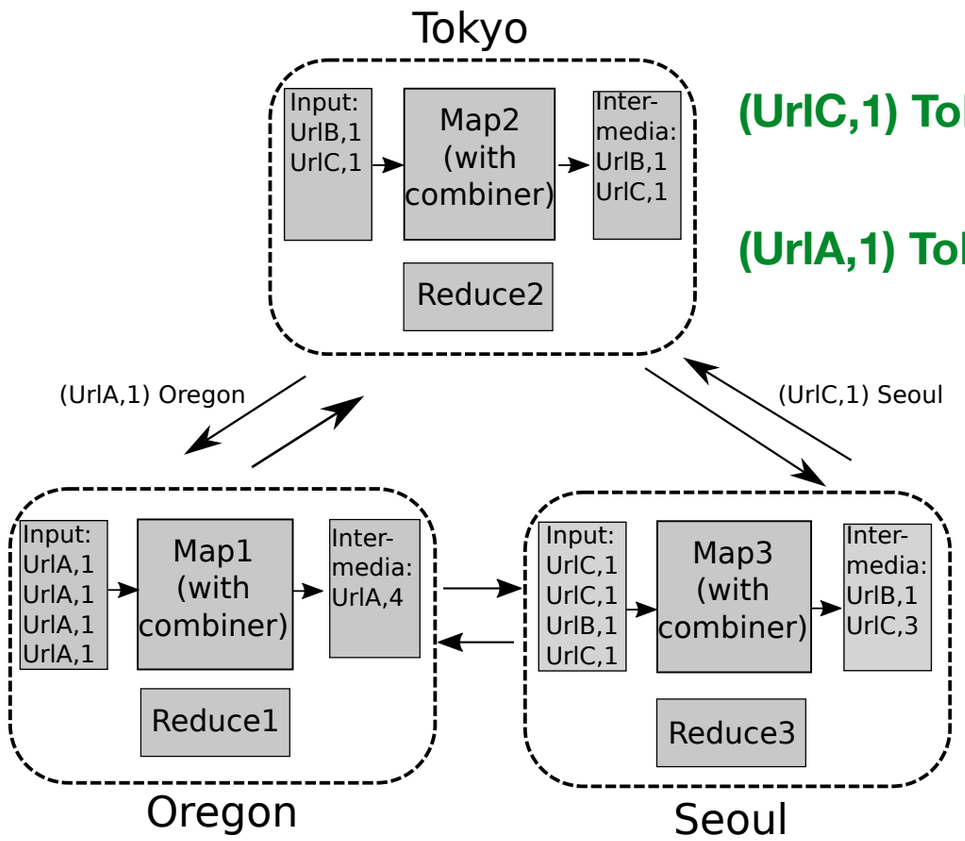
(UrlA,1) Tokyo to Seoul

(UrlB,1) Tokyo to Oregon

Site	Intermed-iate record
Tokyo	2
Oregon	2
Seoul	3
Total	7

Iridium can be improved

- **similarity oblivious**
- **data reduction ratio** is **constant**



(UrlC,1) Tokyo to Seoul

(UrlA,1) Tokyo to Oregon

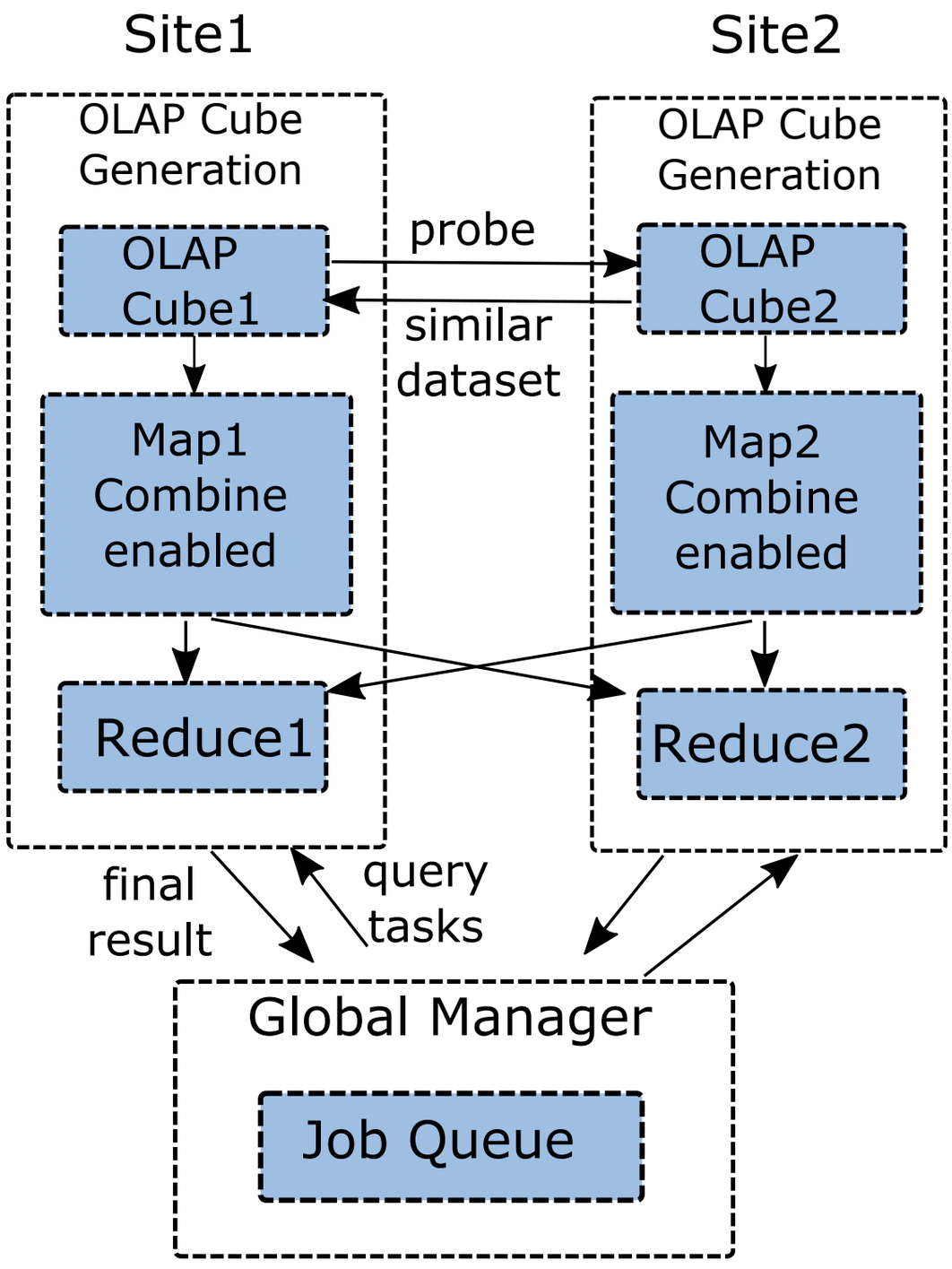
Site	Intermed-iate record
Tokyo	2
Oregon	1
Seoul	2
Total	**5**

Bohr Design

Key ideas:

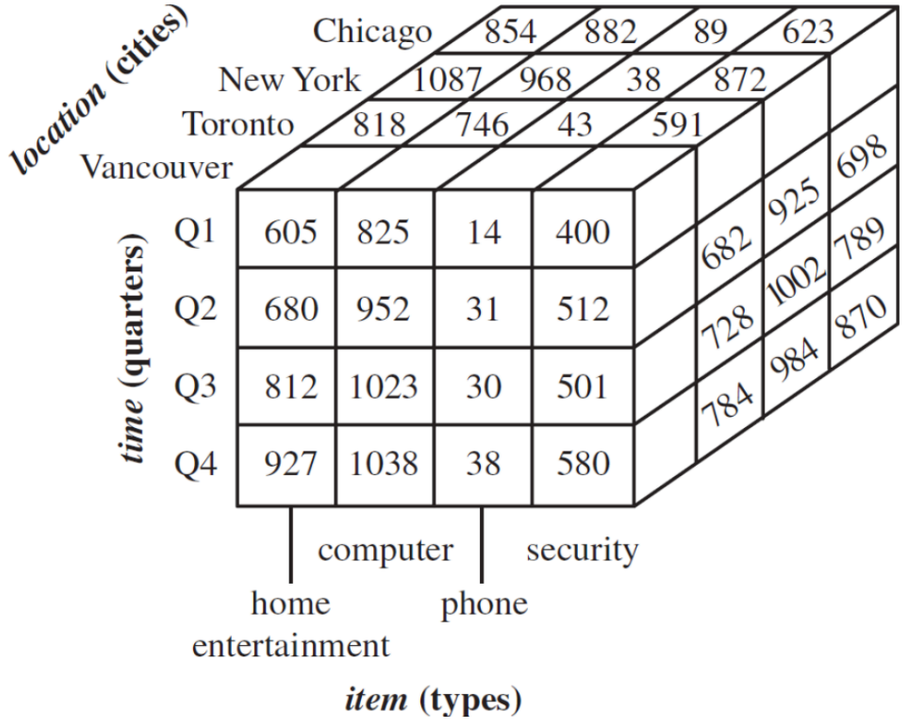
- **Recurring** queries, abundant computation resource.
- **Similarity aware** data transfer.
- **Measuring and predicting** the data reduction ratio.

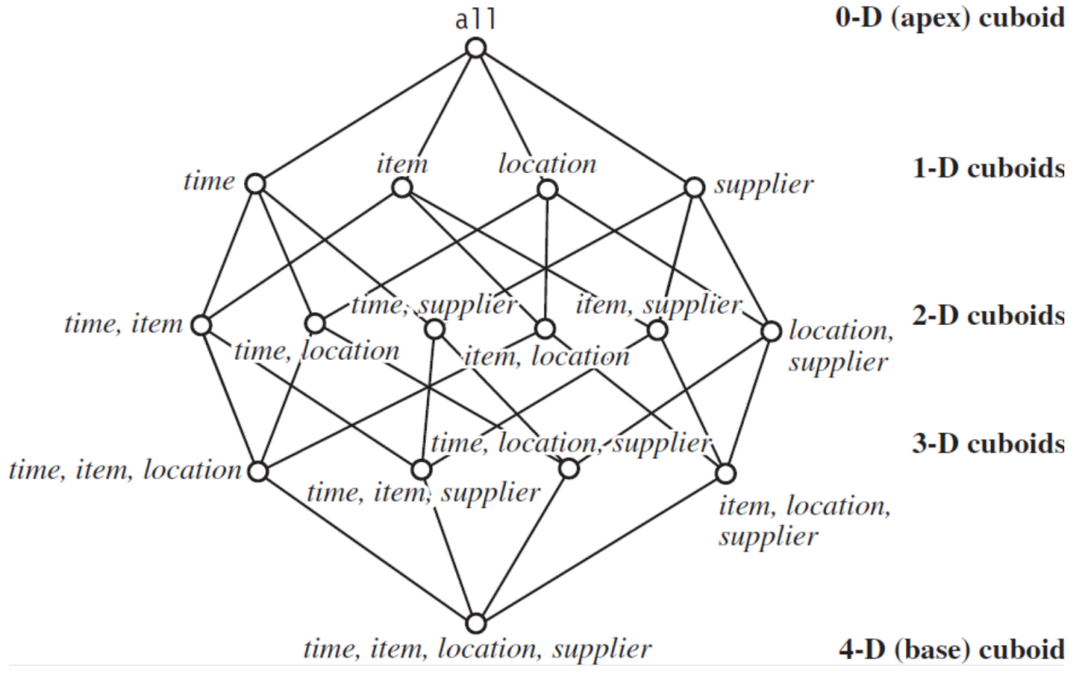
Bohr Design


Site1
Site2
OLAP Cube Generation
OLAP Cube Generation
probe
OLAP Cube1
OLAP Cube2
similar dataset
Map1 Combine enabled
Map2 Combine enabled
Reduce1
Reduce2
final result
query tasks
Global Manager
Job Queue

Why using OLAP Cube?

- Sorting dataset according to the attribute.
- Using record level similarity score.





Why using OLAP Cube?

Advantages:

- Format the unformatted raw data.
- Multiple dimension cube for one dataset.

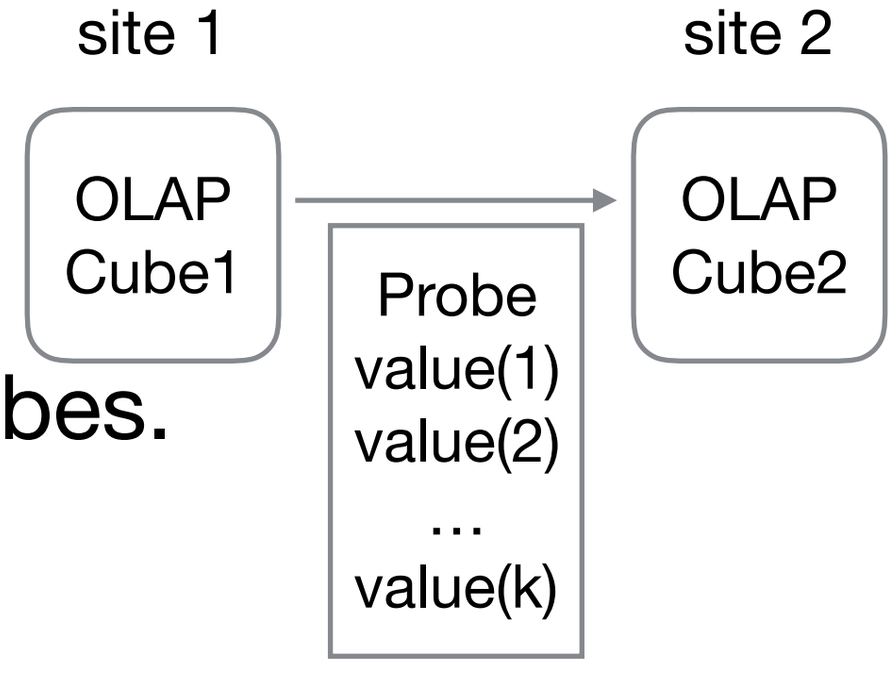
Similarity Search & Check

Search:

- OLAP instruction(eg. dice, slice, roll up) to retrieve the needed attributes.

Check:

- Cross site check: simple probes.
- The probe components.



Reduce tasks scheduling

t : data transfer time.
R: data reduction ratio.
D: downlink bandwidth.

I : input data size.
U: uplink bandwidth.
r : reduce tasks percentage.

$$\min \quad t \tag{1}$$

Goal : Minimize the data transfer time

$$\text{s.t.} \quad (1 - r_i) I_i R_i / U_i \leq t, \ \forall i, \tag{2}$$

Time to upload data from site i.

$$r_i (\sum_{j=1}^{N} I_j R_j - I_i R_i) / D_i \leq t, \ \forall i, \tag{3}$$

Time to download data from other sites.

$$\sum_i r_i = 1, r_i \geq 0, \forall i. \tag{4}$$

System Implementation

Based on Spark v2.1.0 and Iridium.

Utilize Apache Kylin OLAP data cube to deal with all the OLAP operation.

Use Gurobi Optimizer for optimize the LP function.

Evaluation Methodology

- **Workloads**:
  - TPC-DS.
  - Amplab big data benchmark.
- **Hardware platform:**

  A real EC2 deployment with 10 sites(Singapore,Tokyo…).

Evaluation Methodology

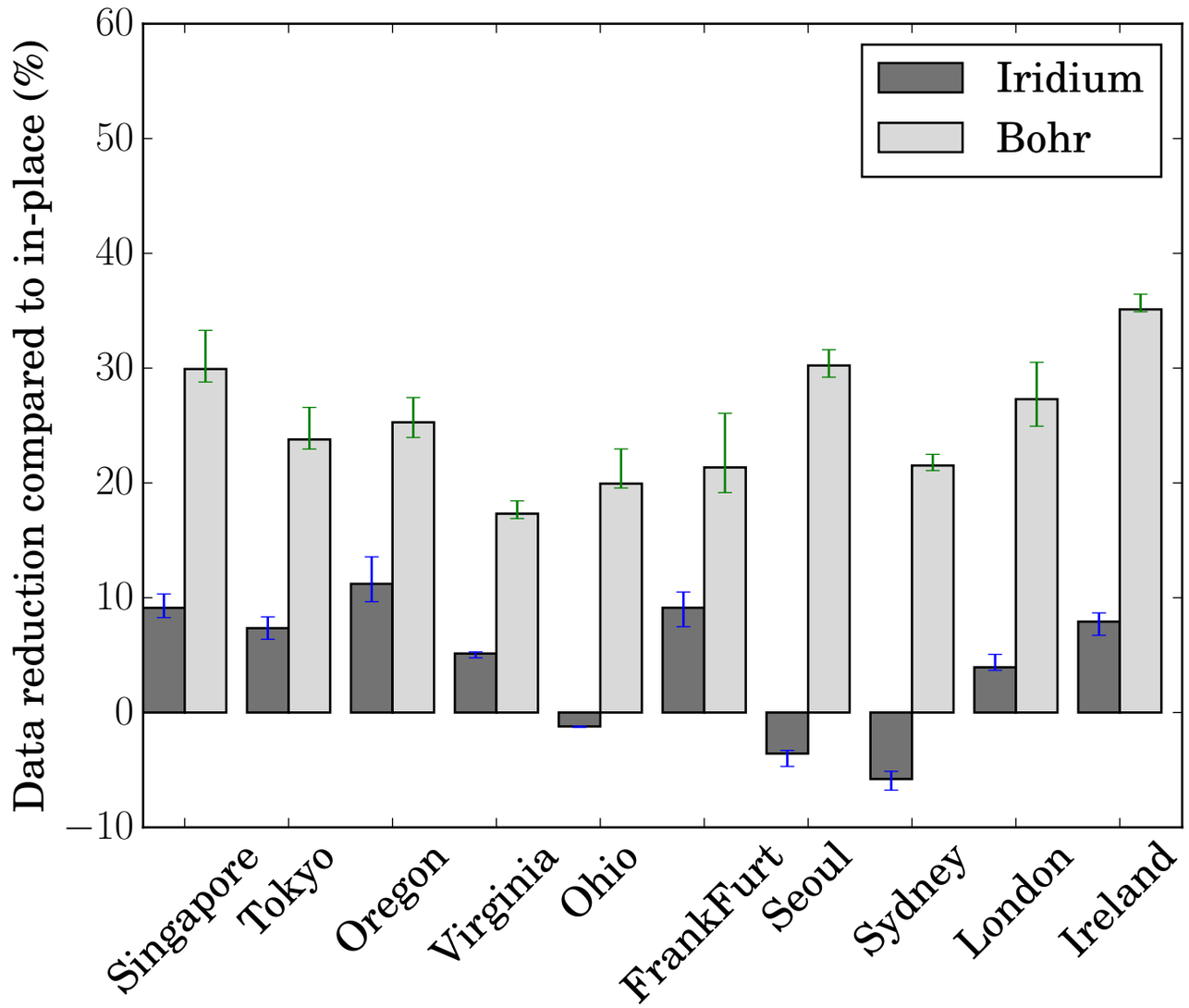
- **Baseline**:

  Iridium(Pu et al. Sigcomm’ 15).

- **Performance metrics**:

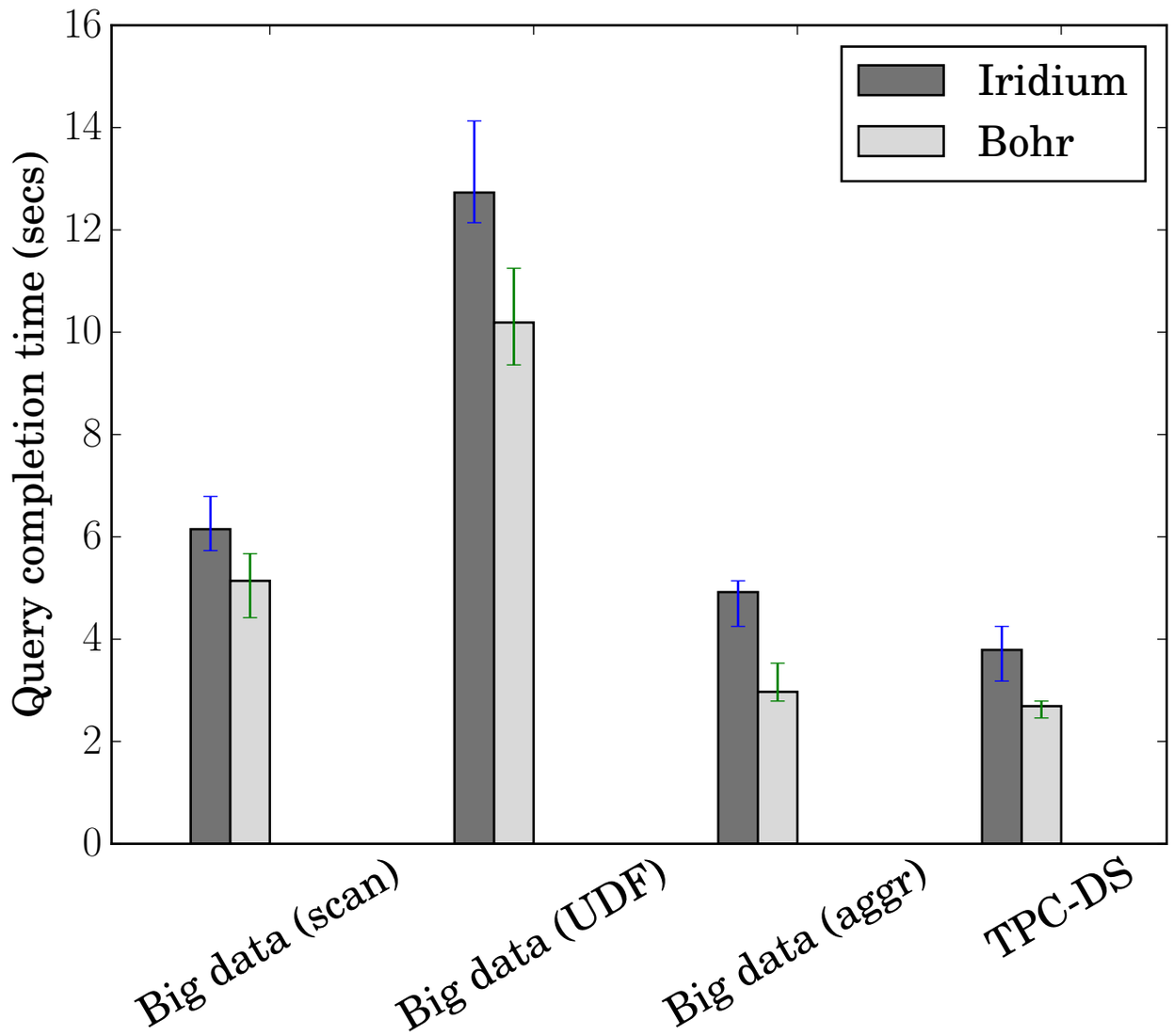
  - Query completion time.
  - Data reduction ratio.

Data Reduction Ratio



Bohr can increase the data reductio ratio in all sites.

Query Completion Time



Bohr saves 30% QCT compare to Iridium.

Conclusion

We propose Bohr:

- Similarity aware data transfer.
- Predicting and measuring data reduction ratio for each site.

Bohr is 30% faster than Iridium.

Thank you!

Future work

1. Generalize the basic idea to more workloads.
2. Dealing with the overhead bring by operate OLAP cubes.
3. Break recurring batch workload condition
4. Multiple types of queries accessing one dataset.