Sampling Estimators for Parallel Online Aggregation

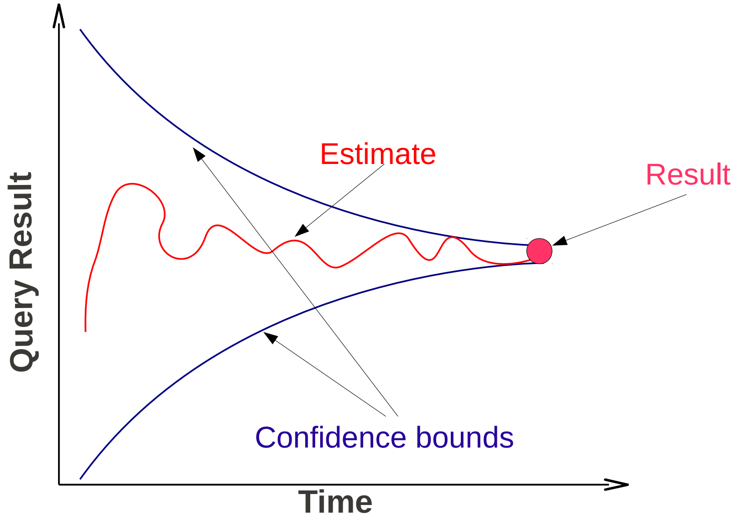
Chengjie Qin and **Florin Rusu**
University of California, Merced

July 10, 2013

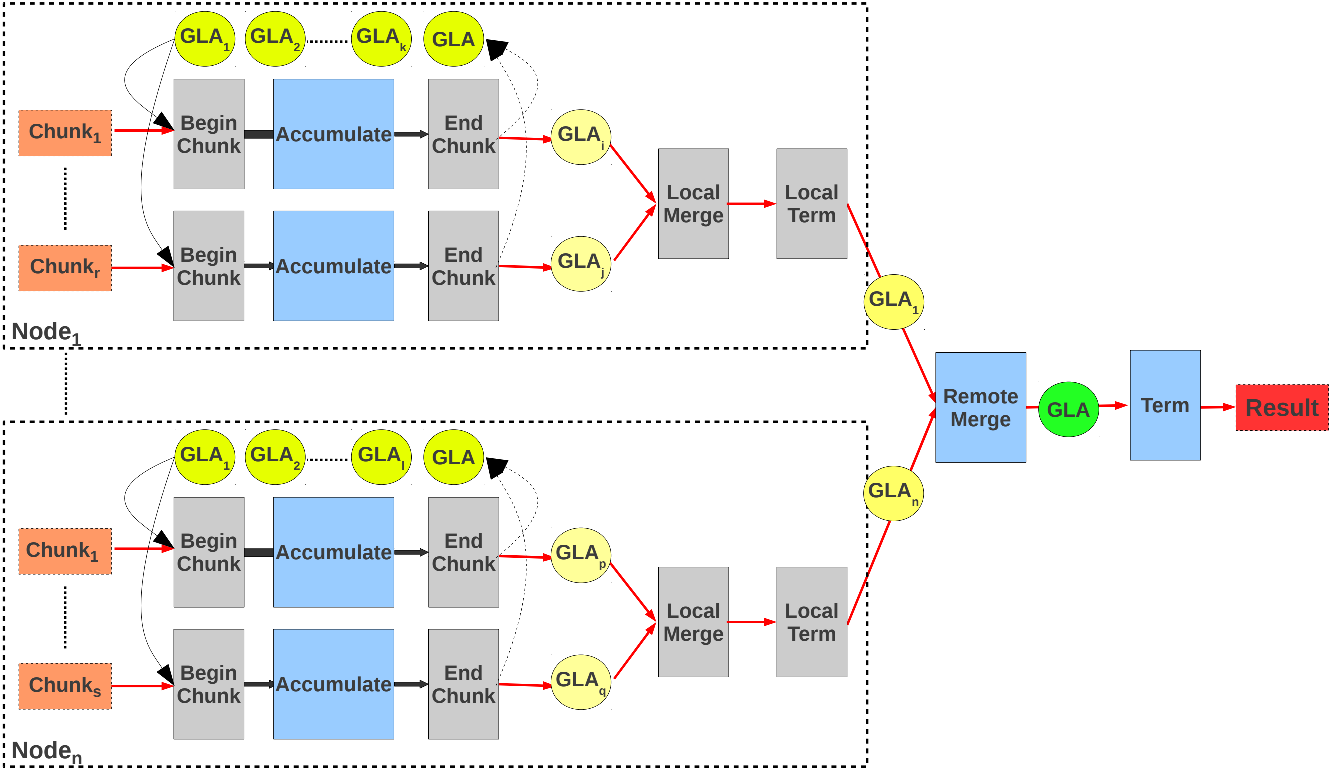
Online Aggregation

$$\begin{aligned} \text{AGG} \Longleftarrow & \text{SELECT SUM}(f(T_1 \odot T_2)) \\ & \text{FROM TABLE}_1 \text{ AS } T_1, \text{ TABLE}_2 \text{ AS } T_2 \\ & \text{WHERE } P(T_1 \odot T_2) \end{aligned}$$

- $\odot$ is concatenation operator
- f is arithmetic expression over concatenated tuple
- P is boolean predicate with selections and join conditions



GLADE



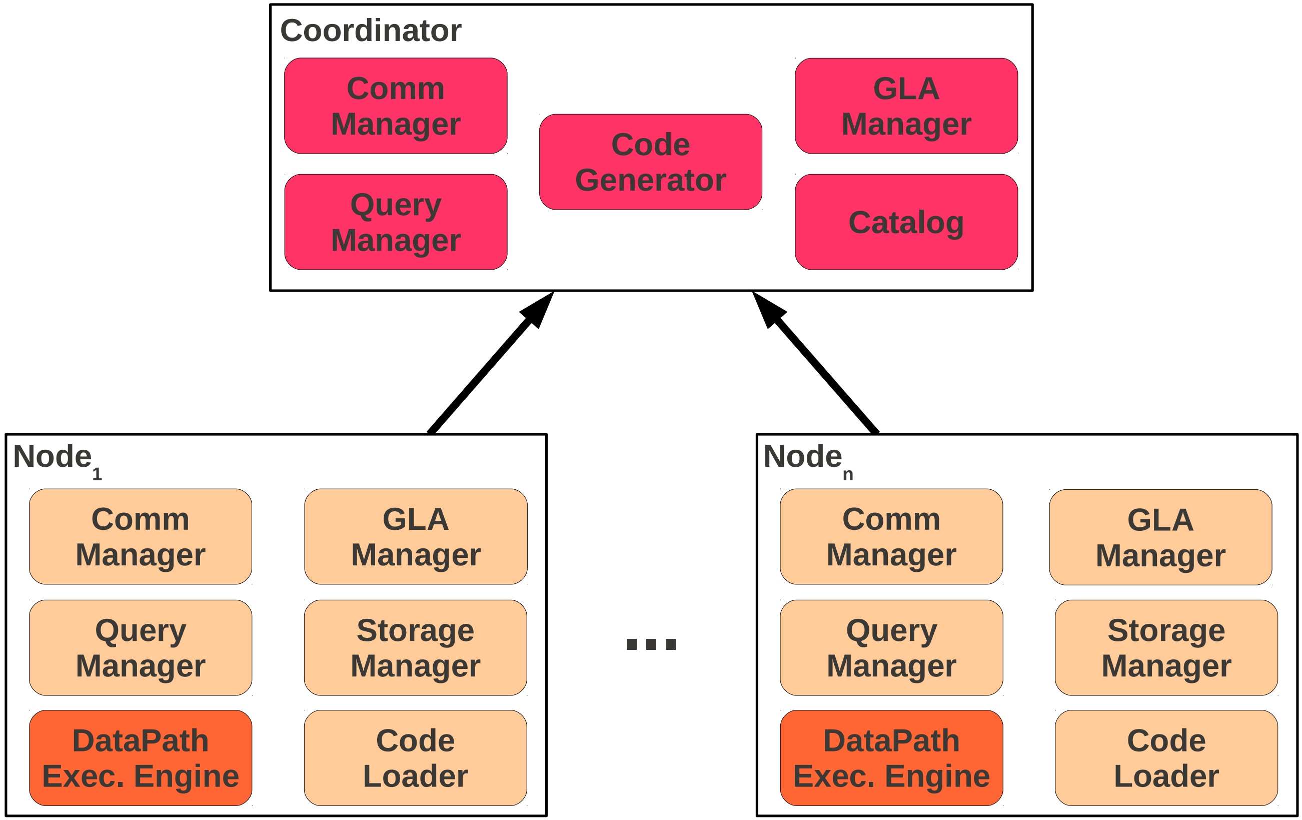
GLA_1
GLA_2
GLA_k
GLA
$Chunk_1$
$Chunk_r$
Begin Chunk
Accumulate
End Chunk
GLA_i
GLA_j
Local Merge
Local Term
$Node_1$
GLA_l
$Chunk_s$
GLA_p
GLA_q
$Node_n$
GLA_n
Remote Merge
Term
Result

Research Question

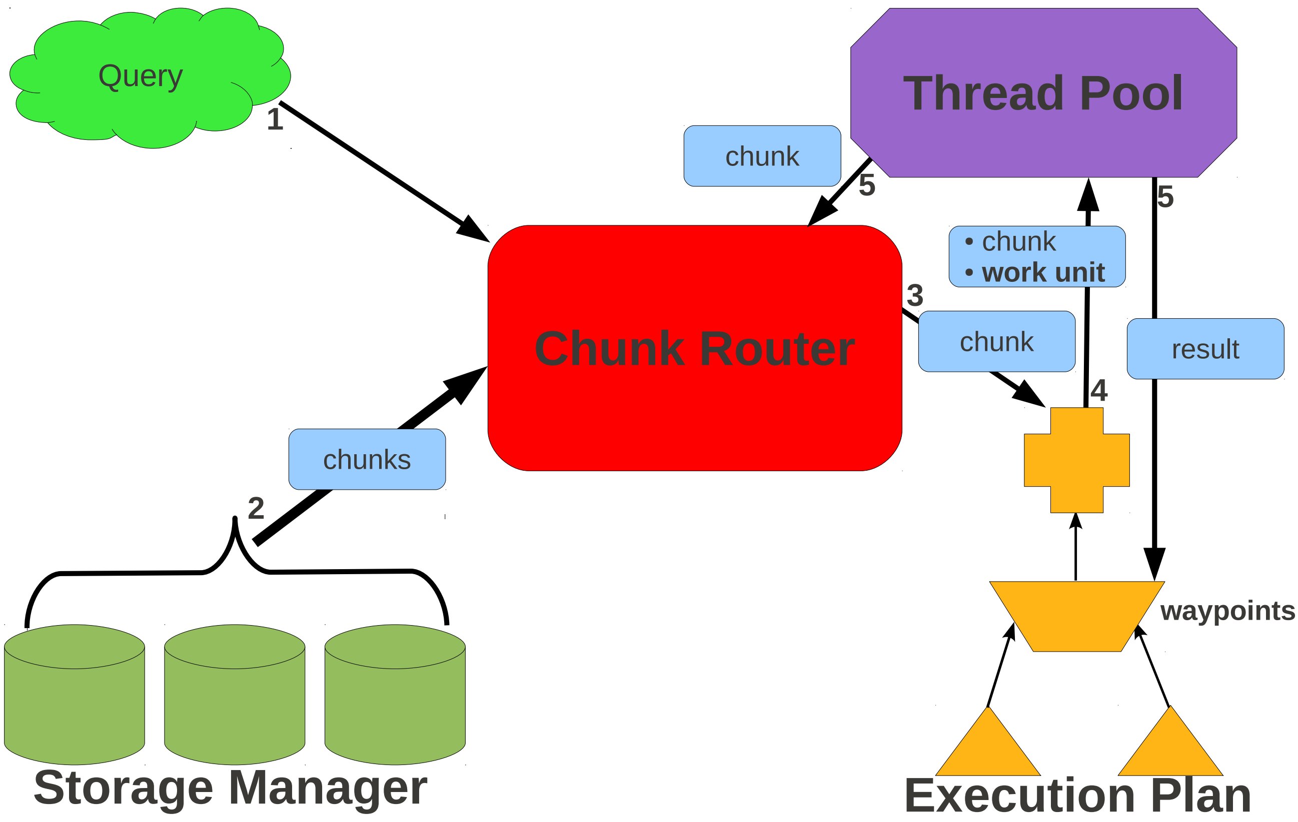
- **How to provide online aggregation in GLADE?**
  - Partial aggregation
  - Parallel sampling
  - Estimators and confidence bounds

- **Contributions**
  - Analyze existent estimators
  - Introduce new scalable and reliable estimator
  - Provide I/O-bound implementation where estimation is virtually free

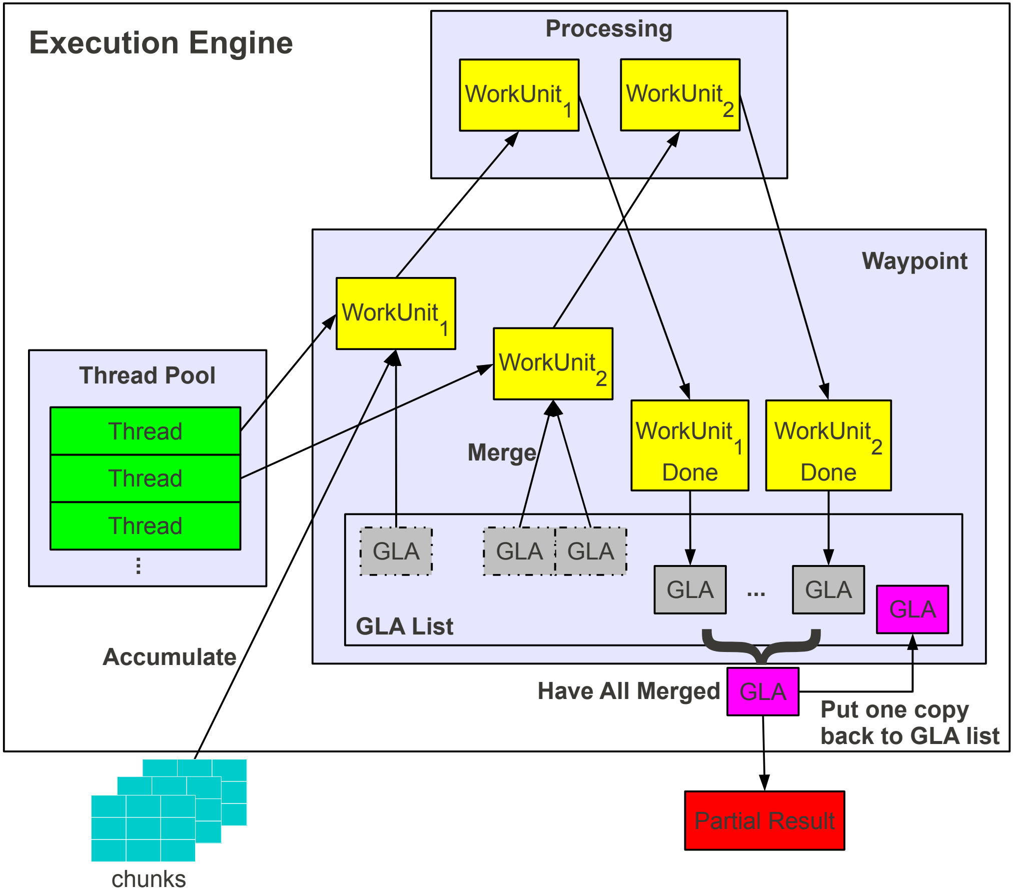
GLADE Architecture



DataPath Execution Engine



Partial Aggregation



Parallel Sampling

Centralized random shuffling

- Permute data randomly at loading
- Scan produces larger samples
- Sample size is not important

Stratified sampling

- Permute data randomly in each partition
- Direct extension of random shuffling to partitioned data

Global data randomization at loading

- Split data randomly at each node
- Permute all received data randomly
- Standard hash-based data partitioning

Generic Sampling Estimator

$$\begin{aligned} \text{AGG} \Longleftarrow & \text{SELECT SUM}(f(d)) \\ & \text{FROM } D \\ & \text{WHERE } P(d) \end{aligned}$$

- S is simple random sample without replacement from D
- Estimator $X = \frac{|D|}{|S|} \sum_{s \in S, P(s)} f(s)$

$$E[X] = \text{AGG}$$

$$\text{Var}[X] = \frac{|D| - |S|}{(|D| - 1)|S|} \left[|D| \sum_{d \in D, P(d)} f^2(d) - \left(\sum_{d \in D, P(d)} f(d) \right)^2 \right]$$

$$\text{Est}_{\text{Var}[X]} = \frac{|D|(|D| - |S|)}{|S|^2(|S| - 1)} \left[|S| \sum_{s \in S, P(s)} f^2(s) - \left(\sum_{s \in S, P(s)} f(s) \right)^2 \right]$$

Parallel Sampling Estimators

- Data are partitioned across N nodes: $D = D_1 \cup D_2 \cup \cdots \cup D_N$
- Take samples S_i, $1 \le i \le N$ independently at each node

Single Estimator

- Guarantee $S = S_1 \cup S_2 \cup \cdots \cup S_N$ is a sample from D
- Apply generic estimator directly

Synchronized estimator

- $\frac{S_i}{D_i} = k$ (const), $1 \le i \le N$

Our estimator

- Global data randomization

Multiple Estimators

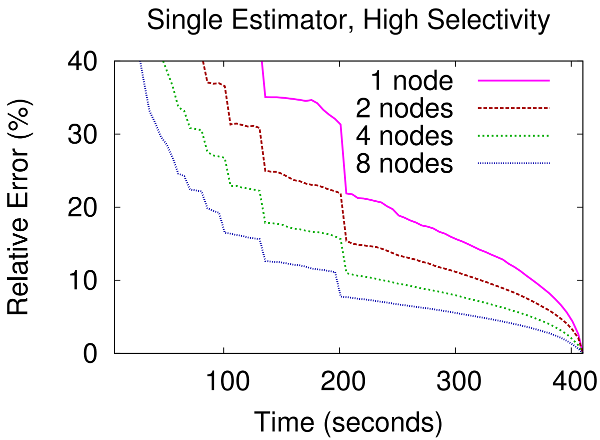
- Stratified sampling
- Build an estimator X_i for each partition D_i, $1 \le i \le N$: $X_i = \frac{|D_i|}{|S_i|} \sum_{s \in S_i, P(s)} f(s)$
- $X = \sum_{i=1}^{N} X_i$ is unbiased
- $\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}\left[X_i\right]$

Estimator Comparison

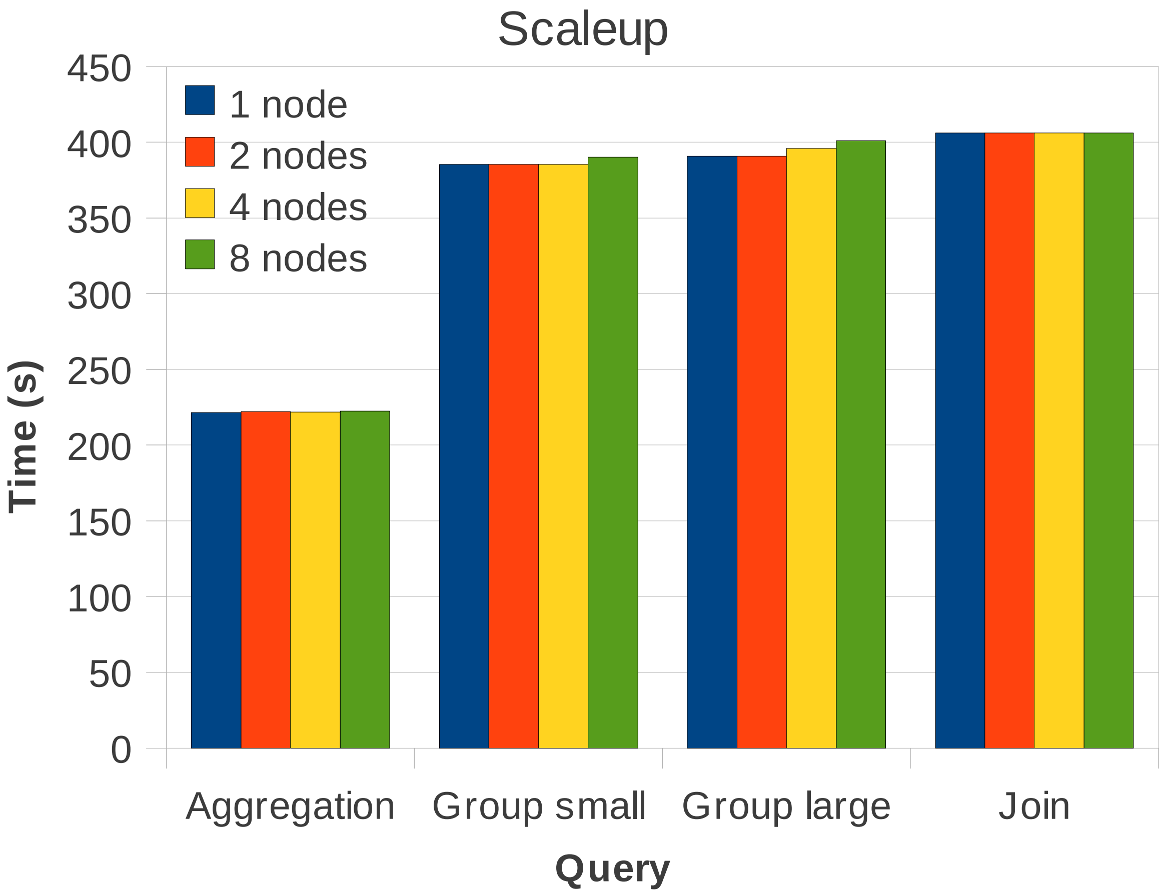
	Single estimator	**Multiple estimators**
Randomization	Global	Local
Dataset information	Dataset cardinality	Local partition cardinality
Accuracy	Generic estimator	Optimal for same sampling ratio across partitions
Convergence rate	Generic estimator	Sensitive to differences across partitions
Fault tolerance	No convergence to exact result	No estimation is possible in the case of node failure without complicated data replication

Empirical Evaluation

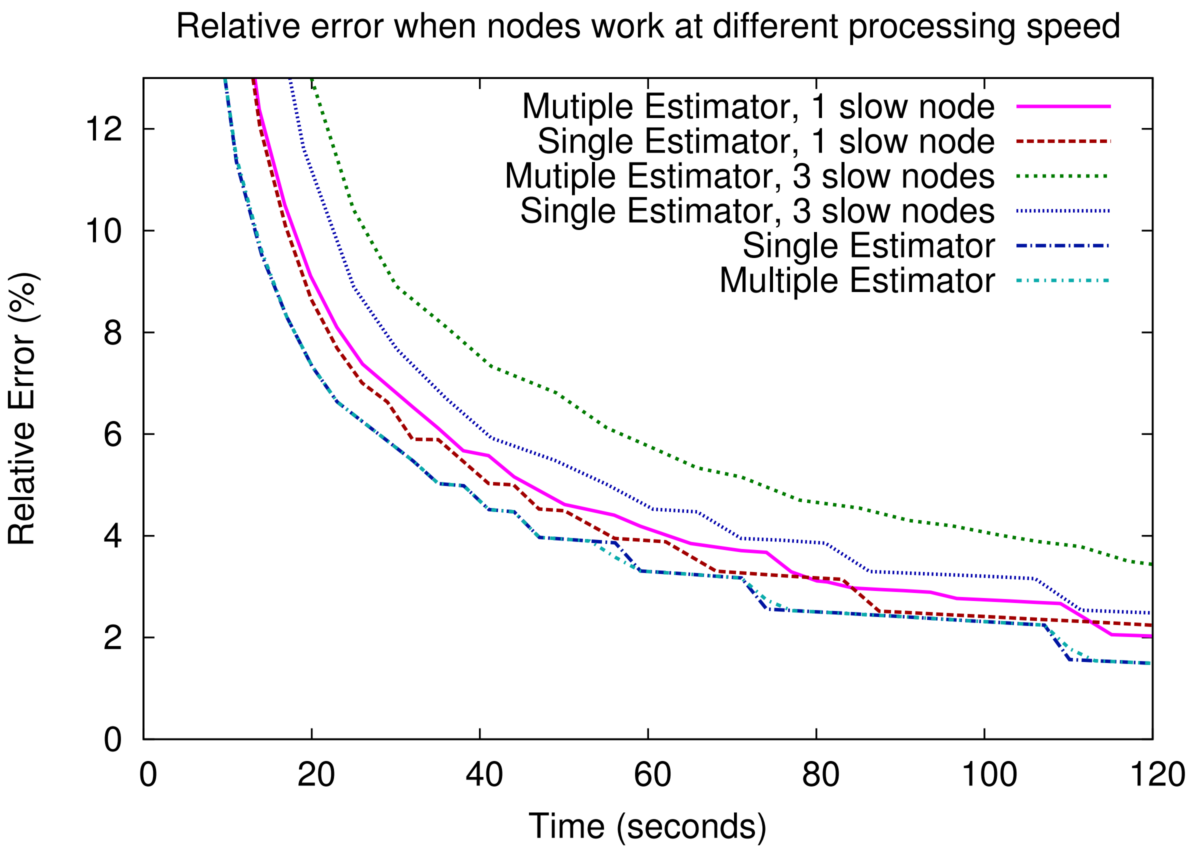
```
SELECT n_name, SUM(l_extendprice*(1-l_discount)*(1+l_tax))
FROM lineitem, supplier, nation
WHERE l_shipdate = 1993-02-26 AND l_quantity = 1 AND
 l_discount between [0.02,0.03] AND
 l_suppkey = s_suppkey AND s_nationkey = n_nationkey
GROUP BY n_name
```



- TPC-H scale **8,000 (8TB)**
- Single node: 16 cores @ 2GHz; 16GB RAM; 4 disks @ 110MB/s throughput/disk
- Cluster: 8 X worker + coordinator (9 nodes); Gigabit Ethernet; same rack


Scaleup
450
400
350
300
250
200
150
100
50
0
Time (s)
1 node
2 nodes
4 nodes
8 nodes
Aggregation
Group small
Group large
Join
Query

Estimator Robustness



Estimation Overhead

Query	Execution Time (seconds)		
	No estimation	Single estimator	Multiple estimators
Aggregate	222	222	222
$\texttt{Group}_{\texttt{small}}$	344	345	344
$\texttt{Group}_{\texttt{large}}$	404	407	407
Join	409	411	411

Questions