

# Scalable Processing of Aggregate Continuous Queries in a Distributed Environment

Anatoli U. Shein, Panos K. Chrysanthis, Alexandros Labrinidis

Advanced Data Management Technologies Laboratory  
Department of Computer Science, University of Pittsburgh

<http://db.cs.pitt.edu>



## 1) Online Analytics & Data Stream Management Systems

SELECT AVG(price) FROM Stock\_Stream  
Range = 100 msec AND Slide = 5 msec

SELECT MAX(heart\_rate) FROM Patient\_Stream  
Range = 30 min AND Slide = 10 min



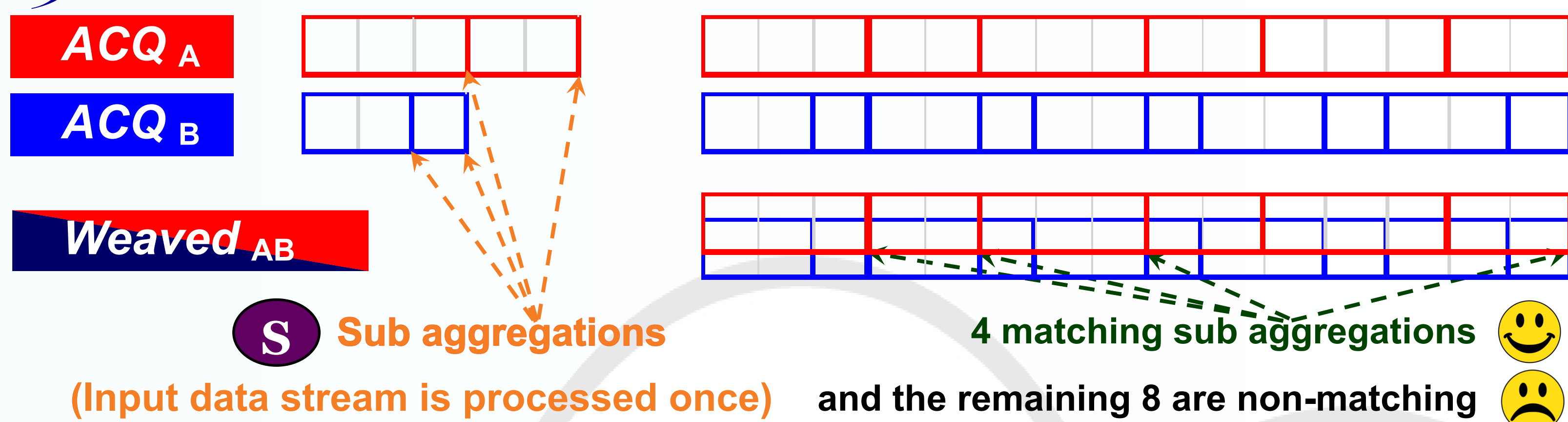
SELECT MIN(humidity) FROM Forest\_Stream  
Range = 1 day msec AND Slide = 2 hrs

SELECT MEDIAN(luminance) FROM Star\_Stream  
Range = 10 min msec AND Slide = 30 sec



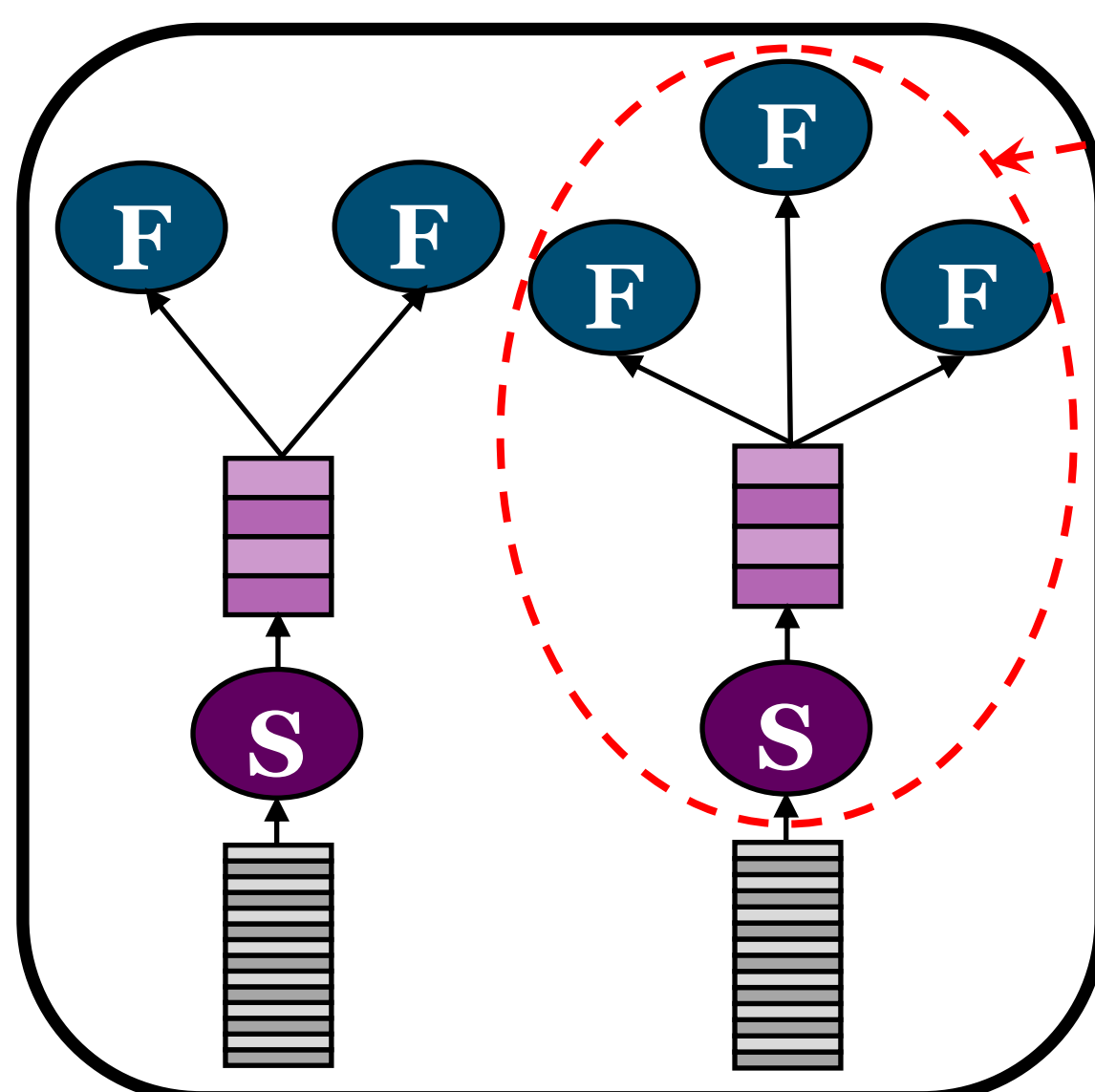
Aggregate Continuous Queries (ACQs)

## 2) Sharing Partial Aggregations (Weaving)



## 3) CLOUD

Node 1

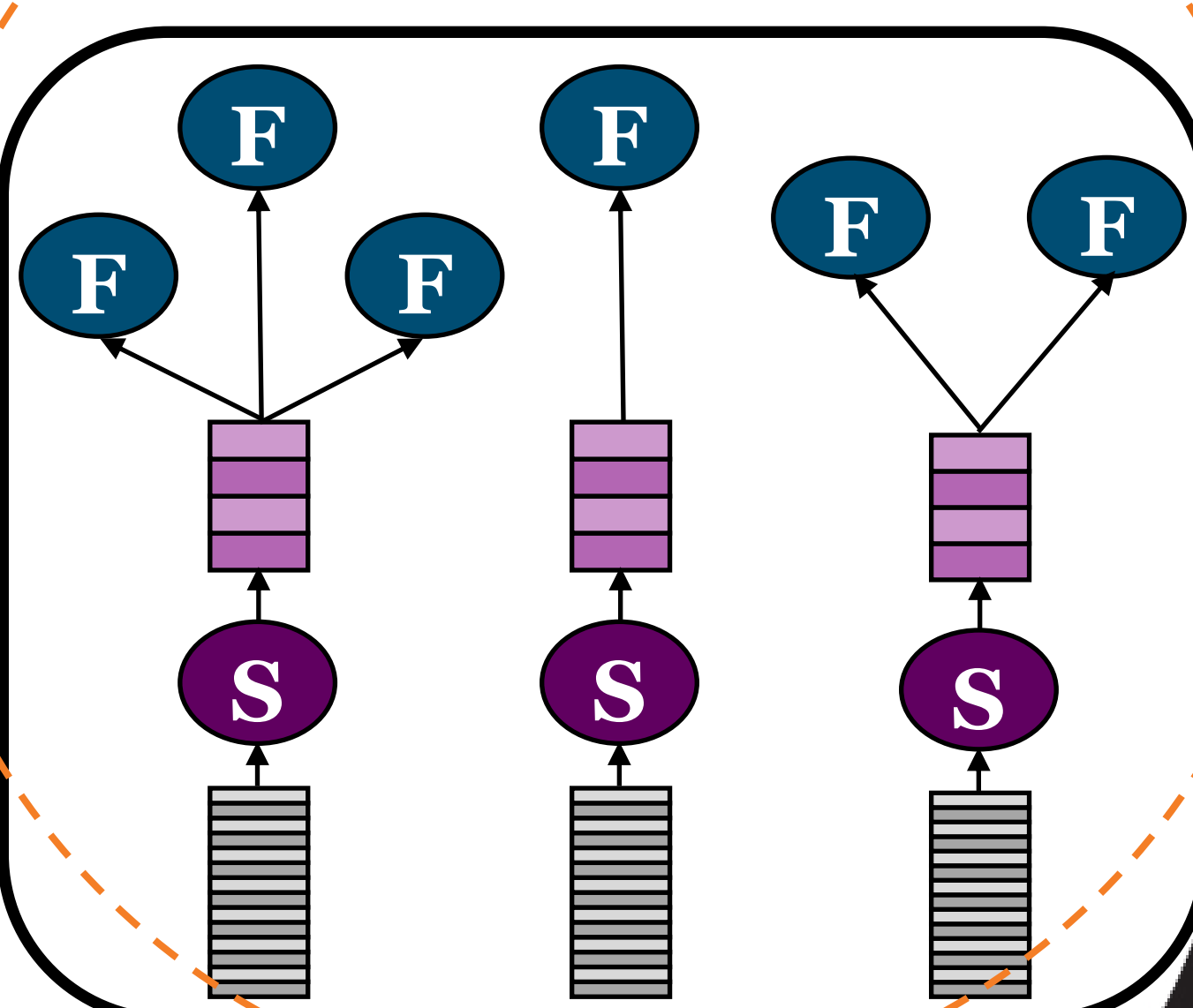


Weaving = combining multiple ACQs into execution trees by sharing:

S Sub aggregates and  
F Final aggregates

Grouping = collocating multiple execution trees within a computation node

Node 2



## 4) Taxonomy of Optimizers

Optimizers

Categories	Optimizers					
	Non Cost-based			Cost-based		
	Random	Round Robin	To Lowest	To Nodes	Inserted	
Group only	G <sub>RAND</sub>	G <sub>RR</sub>	G <sub>TL</sub>	-	-	
Weave only	W <sub>RAND</sub>	W <sub>RR</sub>	-	W <sub>TN</sub>	W <sub>I</sub>	
Weave + Group	WG <sub>RAND</sub>	WG <sub>RR</sub>	-	WG <sub>TN</sub>	WG <sub>I</sub>	

## 5) Goals of Optimization

Low Total Cost → More Client Requests

Clients → 😊

$$\text{Total Cost } (T) = \sum_{i=1}^n C_i$$

Balanced Workload → Energy Savings

Providers → 😊

$$\text{MaxCost} = \text{Max}_i^n C_i$$

## 6) Experimental Evaluation

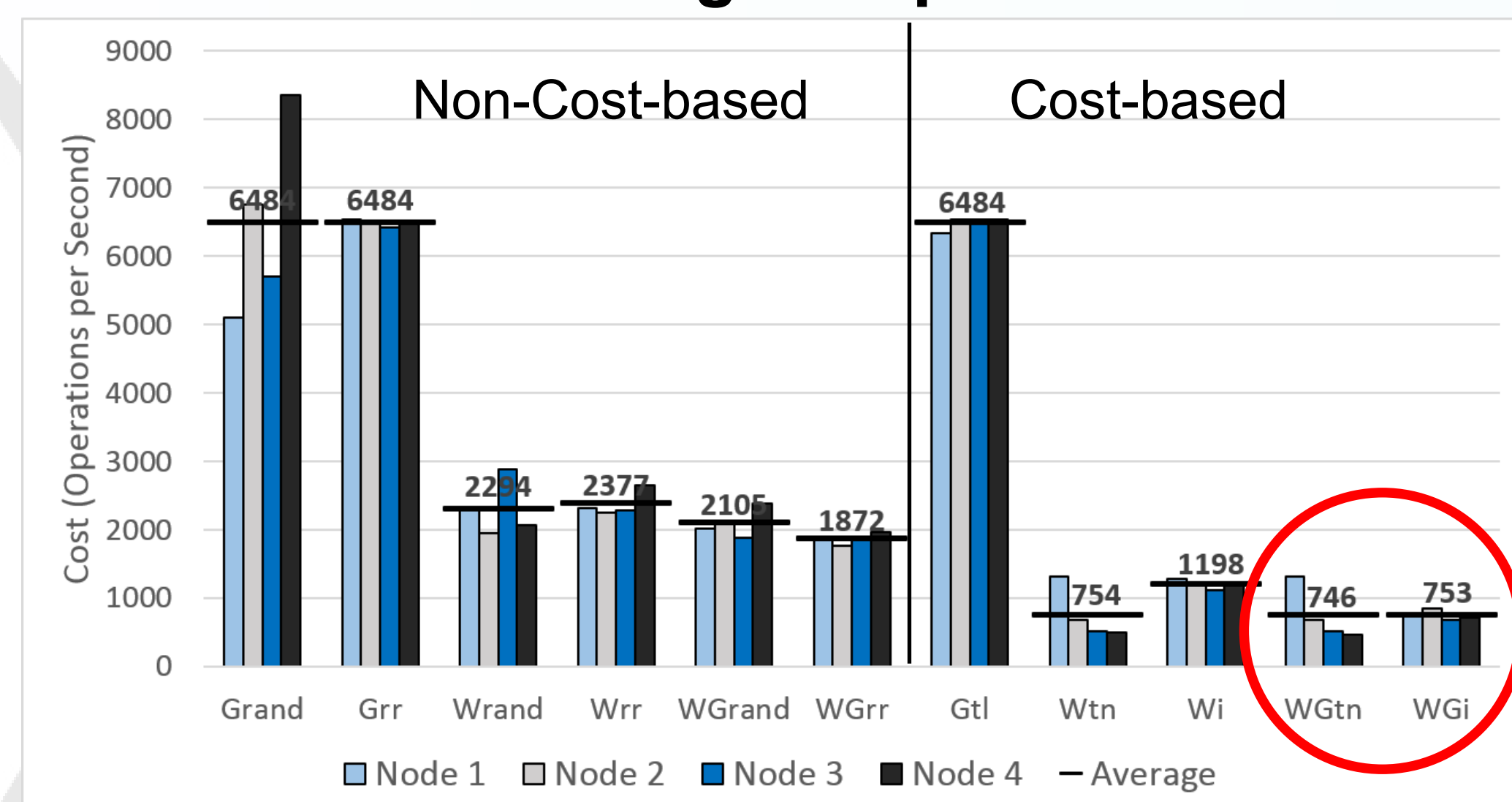
### Max Cost Comparison



### Total Cost Comparison



### Balancing Comparison



## Related Work

- S. Guirguis, M. A. Sharaf, P. K. Chrysanthis, and A. Labrinidis. Optimized processing of multiple aggregate continuous queries. In CIKM, 2011.
- A. U. Shein, P. K. Chrysanthis, and A. Labrinidis. Processing of aggregate continuous queries in a distributed environment. In BIRTE, 2015.
- A. U. Shein, P. K. Chrysanthis, and A. Labrinidis. F1: Accelerating the optimization of aggregate continuous queries in a distributed environment. In CIKM, 2015.

## Acknowledgements

NSF award CBET-1250171

Gift from EMC/Greenplum

