Collaborative Query Coordination in Community-Driven Data Grids

Tobias Scholl, Angelika Reiser, and Alfons Kemper

Department of Computer Science, Technische Universität München
Germany
Community-Driven Data Grids (HiSbase)
The AstroGrid-D Project

- German Astronomy Community Grid
  http://www.gac-grid.org/
- Funded by the German Ministry of Education and Research
- Part of D-Grid
Up-Coming Data-Intensive Applications

• Alex Szalay, Jim Gray (Nature, 2006): “Science in an exponential world”
• Data rates
  – Terabytes a day/night
  – Petabytes a year
• LHC
• LSST
• LOFAR
• Pan-STARRS
The Multiwavelength Milky Way

http://adc.gsfc.nasa.gov/mw/
Research Challenges

• Directly deal with Terabyte/Petabyte-scale data sets
• Integrate with existing community infrastructures
• High throughput for growing user communities
Current Sharing in Data Grids

- Data autonomy
- Policies allow partners to access data
- Each institution ensures
  - Availability (replication)
  - Scalability
- Various organizational structures [Venugopal et al. 2006]:
  - Centralized
  - Hierarchical
  - Federated
  - Hybrid
Community-Driven Data Grids (HiSbase)
“Distribute by Region – not by Archive!”
“Distribute by Region – not by Archive!”

Histogram regions
“Distribute by Region – not by Archive!”
"Distribute by Region – not by Archive!"

Quadtree
Mapping Data to Nodes
Submission Characteristics

- **Portal-based submission**
- Browser in every researcher’s "tool box"
- Scalability depends on portal

- **Institution-based submission**
- All data nodes accept queries
- Submission via local data node
Coordinator Selection Strategies

• The node submitting the query
  – SelfStrategy (SS)
• A node containing relevant data (region-based strategies)
  – FirstRegionStrategy (FRS)
  – SelfOrFirstRegionStrategy (SOFRS)
  – CenterOfGravityStrategy (COGS)
  – RandomRegionStrategy (RRS)
SelfStrategy (SS)

\[ \text{P}_S \]

FullQueryMsg -> \[ \text{D}_1 \]

FullQueryMsg -> \[ \text{D}_i \]

FullQueryMsg -> \[ \text{D}_n \]

FullAnswerMsg

PartialQueryMsg

PartialAnswerMsg
FirstRegionStrategy (FRS)
SelfOrFirstRegionStrategy (SOFRS)

- Combination from SelfStrategy and FirstRegionStrategy
- Submit node is coordinator if it covers data
- Avoids unnecessary data transport
- With many partitions and many nodes basically the same as FirstRegionStrategy (as probability of Self-case decreases)
CenterOfGravityStrategy (COGS)

- Further reduce amount of data shipping
- "Perfect spot" for minimizing data transfer
RandomRegionStrategy (RRS)

- Select random relevant region
- Tradeoff between balancing coordination load and reducing data shipping

- Probability(a) = 2/9
- Probability(b) = 5/9
- Probability(c) = 2/9
Evaluation

• Coordination Strategies: SS, FRS, SOFRS, COGS, RRS
• Submission Strategies: portal-based, institution-based
• Observational data sets
• Two workloads
  – SDSS query log ($Q_{obs}$)
  – Synthetic ($Q_{scaled}$)
• Network size
• Network traffic measurements
  – Number of routed messages
  – Coordination load balancing
• Throughput Measurements
Query Workloads

The diagram shows the distribution of queries based on the number of relevant partitions. The x-axis represents the number of relevant partitions, ranging from 1 to more than 100, out of a total of 262144 partitions. The y-axis represents the percentage of queries, ranging from 0% to 100%.

There are two sets of bars: Q_{obs} and Q_{scaled}. Q_{obs} is represented by green bars and Q_{scaled} is represented by orange bars.

- For Q_{obs}, approximately 90% of queries have a single relevant partition, with a smaller percentage having 1-10 partitions and an even smaller percentage having more than 100 partitions.
- For Q_{scaled}, the distribution is similar but with a slightly different percentage for each range of partitions.

This distribution highlights the importance of efficient query processing strategies to handle queries with varying partition sizes.
Routed Messages per Query ($Q_{obs}$)
Routed Messages per Query ($Q_{scaled}$)
Portal-based Coordination Load

![Graph showing accumulated coordination load vs. proportion of nodes for different methods: SS, FRS, SOFRS, COGS, RRS. The line labeled "uniform load balancing" indicates the ideal scenario.]
Institution-based Coordination Load

The graph illustrates the accumulated coordination load as a function of the proportion of nodes. The lines represent different methods and strategies for load balancing:

- **SS**: Simple Sharing
- **FRS, SOFRS**: Federated Resource Sharing
- **COGS**: Coordinated Global Sharing
- **RRS**: Round-Robin Sharing

A note indicates "uniform load balancing."
Throughput

- Throughput dependent on query complexity
- No clear winner in terms of throughput
Workload-Aware Data Partitioning

• Query skew (hot spots) triggered by increased interest in particular subsets of the data
• Two well-known query load balancing techniques:
  – Data partitioning
  – Data replication
• Finding trade-offs between both (see EDBT ’09 paper)
Load Balancing During Runtime

• Complement workload-aware partitioning with runtime load-balancing
• Short-term peaks
  – Master-slave approach
  – Load monitoring
• Long-term trends
  – Based on load monitoring
  – Histogram evolution
Related Work

• On-line load balancing
• Hundreds of thousands to millions of nodes
• Reacting fast
• Treating objects individually
Who Is the Query Coordinator?

• Many challenges and opportunities in e-science for distributed computing and database research
  – High-throughput data management
  – Correlation of distributed data sources
• Collaborative Query Coordination
  – Region-based strategies reduce number of messages
  – Load balancing independent of submission characteristic
Special Thanks To …

- Ella Qiu, University of British Columbia
  - DAAD Rise Internship
  - Support during implementation
  - Initial measurements
Get in Touch

• Database systems group, TU München
  – Web site: http://www-db.in.tum.de
  – E-mail: scholl@in.tum.de
• The HiSbase project
  – http://www-db.in.tum.de/research/projects/hisbase/

Thank You for Your Attention