Harnessing Grid Resources with Data-Centric Task Farms

Ioan Raicu
Distributed Systems Laboratory
Computer Science Department
University of Chicago

Collaborators:
Ian Foster (UC/CI/ANL), Yong Zhao (MS), Mike Wilde (CI/ANL),
Zhao Zhang (CI), Rick Stevens (UC/CI/ANL), Alex Szalay (JHU),
Jerry Yan (NASA/AMES), Catalin Dumitrescu (FANL)

NASA, Ames Research Center
GSRP Fellowship Talk
May 13th, 2008
What will we do with 1+ Exaflops and 1M+ cores?
1) Tackle **Bigger and Bigger Problems**
2) Tackle **Increasingly Complex Problems**

Computational Scientist as Logistics Officer
“More Complex Problems”

- Use ensemble runs to quantify climate model uncertainty
- Identify potential drug targets by screening a database of ligand structures against target proteins
- Study economic model sensitivity to key parameters
- Analyze turbulence dataset from multiple perspectives
- Perform numerical optimization to determine optimal resource assignment in energy problems
Programming Model Issues

- Multicore processors
- Massive task parallelism
- Massive data parallelism
- Integrating black box applications
- Complex task dependencies (task graphs)
- Failure, and other execution management issues
- Data management: input, intermediate, output
- Dynamic task graphs
- Dynamic data access involving large amounts of data
- Documenting provenance of data products
Problem Types

<table>
<thead>
<tr>
<th>Input data size</th>
<th>Number of tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hi</strong></td>
<td>Data analysis, mining</td>
</tr>
<tr>
<td><strong>Med</strong></td>
<td>Big data and many tasks</td>
</tr>
<tr>
<td><strong>Lo</strong></td>
<td>Many loosely coupled tasks</td>
</tr>
</tbody>
</table>

- Data analysis, mining: Many loosely coupled tasks
- Big data and many tasks: Many loosely coupled tasks
- Many loosely coupled tasks: Many loosely coupled tasks
Motivating Example: 
AstroPortal Stacking Service

- **Purpose**
  - On-demand “stacks” of random locations within ~10TB dataset

- **Challenge**
  - Rapid access to 10-10K “random” files
  - Time-varying load

- **Solution**
  - Dynamic acquisition of compute, storage

---

5/12/2008 Harnessing Grid Resources with Data-Centric Task Farms
Challenge #1:
Long Queue Times

- Wait queue times are typically longer than the job duration times.
Challenge #2: Slow Job Dispatch Rates

- Production LRMs $\rightarrow$ $\sim$1 job/sec dispatch rates
- What job durations are needed for 90% efficiency:
  - Production LRMs: 900 sec
  - Development LRMs: 100 sec
  - Experimental LRMs: 50 sec
  - 1~10 sec should be possible

<table>
<thead>
<tr>
<th>System</th>
<th>Comments</th>
<th>Throughput (tasks/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condor (v6.7.2) - Production</td>
<td>Dual Xeon 2.4GHz, 4GB</td>
<td>0.49</td>
</tr>
<tr>
<td>PBS (v2.1.8) - Production</td>
<td>Dual Xeon 2.4GHz, 4GB</td>
<td>0.45</td>
</tr>
<tr>
<td>Condor (v6.7.2) - Production</td>
<td>Quad Xeon 3 GHz, 4GB</td>
<td>2</td>
</tr>
<tr>
<td>Condor (v6.8.2) - Production</td>
<td></td>
<td>0.42</td>
</tr>
<tr>
<td>Condor (v6.9.3) - Development</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Condor-J2 - Experimental</td>
<td>Quad Xeon 3 GHz, 4GB</td>
<td>22</td>
</tr>
</tbody>
</table>
Challenge #3: Poor Scalability of Shared File Systems

- **GPFS vs. LOCAL**
  - **Read Throughput**
    - 1 node: 0.48Gb/s vs. 1.03Gb/s \(\rightarrow 2.15x\)
    - 160 nodes: 3.4Gb/s vs. 165Gb/s \(\rightarrow 48x\)
  - **Read+Write Throughput**:
    - 1 node: 0.2Gb/s vs. 0.39Gb/s \(\rightarrow 1.95x\)
    - 160 nodes: 1.1Gb/s vs. 62Gb/s \(\rightarrow 55x\)
  - **Metadata (mkdir / rm -rf)**
    - 1 node: 151/sec vs. 199/sec \(\rightarrow 1.3x\)
    - 160 nodes: 21/sec vs. 31840/sec \(\rightarrow 1516x\)
Hypothesis

“Significant performance improvements can be obtained in the analysis of large dataset by leveraging information about data analysis workloads rather than individual data analysis tasks.”

• **Important concepts related to the hypothesis**
  – **Workload**: a complex query (or set of queries) decomposable into simpler tasks to answer broader analysis questions
  – **Data locality** is crucial to the efficient use of large scale distributed systems for scientific and data-intensive applications
  – Allocate computational and caching storage resources, **co-scheduled** to optimize workload performance
Proposed Solution: Part 1
Abstract Model and Validation

• AMDASK:
  – An Abstract Model for DAta-centric taSK farms
    • Task Farm: A common parallel pattern that drives independent computational tasks
    – Models the efficiency of data analysis workloads for the split/merge class of applications
    – Captures the following data diffusion properties
      • Resources are acquired in response to demand
      • Data and applications diffuse from archival storage to new resources
      • Resource “caching” allows faster responses to subsequent requests
      • Resources are released when demand drops
      • Considers both data and computations to optimize performance

• Model Validation
  – Implement the abstract model in a discrete event simulation
  – Validate model with statistical methods ($R^2$ Statistic, Residual Analysis)
Proposed Solution: Part 2
Practical Realization

• Falkon: a Fast and Light-weight task execution framework
  – Light-weight task dispatch mechanism
  – Dynamic resource provisioning to acquire and release resources
  – Data management capabilities including data-aware scheduling
  – Integration into Swift to leverage many Swift-based applications
    • Applications cover many domains: astronomy, astro-physics, medicine, chemistry, and economics
AMDASK: Performance Efficiency Model

• **B**: Average Task Execution Time:
  - K: Stream of tasks
  - \( \mu(k) \): Task k execution time
  \[
  B = \frac{1}{|K|} \sum_{k \in K} \mu(k)
  \]

• **Y**: Average Task Execution Time with Overheads:
  - \( o(k) \): Dispatch overhead
  - \( \zeta(\delta, \tau) \): Time to get data
  \[
  Y = \begin{cases}
    \frac{1}{|K|} \sum_{k \in K} [\mu(k) + o(k)], & \delta \in \phi(\tau), \delta \in \Omega \\
    \frac{1}{|K|} \sum_{k \in K} [\mu(k) + o(k) + \zeta(\delta, \tau)], & \delta \notin \phi(\tau), \delta \in \Omega
  \end{cases}
  \]

• **V**: Workload Execution Time:
  - A: Arrival rate of tasks
  - T: Transient Resources
  \[
  V = \max \left( \frac{B}{|T|}, \frac{1}{A} \right) * |K|
  \]

• **W**: Workload Execution Time with Overheads
  \[
  W = \max \left( \frac{Y}{|T|}, \frac{1}{A} \right) * |K|
  \]
AMDASK: Performance Efficiency Model

- **Efficiency**
  \[ E = \frac{V}{W} \]
  
  \[ E = \begin{cases} 
  1, & \frac{Y}{|T|} \leq \frac{1}{A} \\
  \max \left( \frac{B}{Y}, \frac{|T|}{A \cdot Y} \right), & \frac{Y}{|T|} > \frac{1}{A} 
  \end{cases} \]

- **Speedup**
  \[ S = E \cdot |T| \]

- **Optimizing Efficiency**
  - Easy to maximize either efficiency or speedup independently
  - Harder to maximize both at the same time
  
  - Find the smallest number of transient resources \(|T|\) while maximizing speedup*efficiency
Performance Efficiency Model
Example: 1K CPU Cluster

- Application: Angle - distributed data mining
- Testbed Characteristics:
  - Computational Resources: 1024
  - Transient Resource Bandwidth: 10MB/sec
  - Persistent Store Bandwidth: 426MB/sec
- Workload:
  - Number of Tasks: 128K
  - Arrival rate: 1000/sec
  - Average task execution time: 60 sec
  - Data Object Size: 40MB
Performance Efficiency Model
Example: 1K CPU Cluster

Falkon on ANL/UC TG Site:
Peak Dispatch Throughput: 500/sec
Scalability: 50~500 CPUs
Peak speedup: 623x

PBS on ANL/UC TG Site:
Peak Dispatch Throughput: 1/sec
Scalability: <50 CPUs
Peak speedup: 54x
Model Validation: Simulations

• Implement the abstract model in a discrete event simulation

• Simulation parameters
  – number of storage and computational resources
  – communication costs
  – management overhead
  – workloads (inter-arrival rates, query complexity, data set properties, and data locality)

• Model Validation
  – $R^2$ Statistic
  – Residual analysis
Falkon: a Fast and Light-weight task execution framework

- **Goal**: enable the *rapid and efficient* execution of many independent jobs on large compute clusters

- Combines three components:
  - a *streamlined task dispatcher* able to achieve order-of-magnitude higher task dispatch rates than conventional schedulers ➔ Challenge #1
  - *resource provisioning* through multi-level scheduling techniques ➔ Challenge #2
  - *data diffusion* and data-aware scheduling to leverage the co-located computational and storage resources ➔ Challenge #3
Falkon: The Streamlined Task Dispatcher

- Tier 1: Dispatcher
  - GT4 Web Service accepting task submissions from clients and sending them to available executors

- Tier 2: Executor
  - Run tasks on local resources

- Provisioner
  - Static and dynamic resource provisioning
Falkon: The Streamlined Task Dispatcher

- Falkon Message Exchanges
  - Description:
    - {1}: task(s) submit
    - {2}: task(s) submit confirmation
    - {3}: notification for work
    - {4}: request for task(s)
    - {5 or 7}: dispatch task(s)
    - {6}: deliver task(s) results to service
    - {8}: notification for task result(s)
    - {9}: request for task result(s)
    - {10}: deliver task(s) results to client
  - Worst case (process tasks individually, no optimizations):
    - 4 WS messages ({1,2}, {4,5}, {6,7}, {9,10}) and 2 notifications ({3}, {8}) per task
Falkon: The Streamlined Task Dispatcher

- Falkon Message Exchanges Enhancements
  - Bundling
    • Include multiple tasks per communication message
  - Piggy-Backing
    • Attach next task to acknowledgement of previous task
    • Include data management information in the task description and acknowledgement messages
  - Message reduction:
    • General Lower Bound: $10 \rightarrow 2 + c$, where $c$ is a small positive value
    • Application Specific Lower Bound: $10 \rightarrow 0 + c$, where $c$ is a small positive value
Falkon: Resource Provisioning

0. provisioner registration
1. task(s) submit
2. resource allocation to GRAM
3. resource allocation to LRM
4. executor registration
5. notification for work
6. pick up task(s)
7. deliver task(s) results
8. notification for task(s) result
9. pick up task(s) results
Falkon: Data Diffusion

- Resource acquired in response to demand
- Data and applications diffuse from archival storage to newly acquired resources
- Resource “caching” allows faster responses to subsequent requests
  - Cache Eviction Strategies: RANDOM, FIFO, LRU, LFU
- Resources are released when demand drops
Falkon: Data Diffusion

- Considers both data and computations to optimize performance
- Decrease dependency of a shared file system
  - Theoretical linear scalability with compute resources
  - Significantly increases meta-data creation and/or modification performance
- Completes the “data-centric task farm” realization
Related Work: Task Farms

• [Casanova99]: Adaptive Scheduling for Task Farming with Grid Middleware
• [Heymann00]: Adaptive Scheduling for Master-Worker Applications on the Computational Grid
• [Danelutto04]: Adaptive Task Farm Implementation Strategies
• [González-Vélez05]: An Adaptive Skeletal Task Farm for Grids
• [Petrou05]: Scheduling Speculative Tasks in a Compute Farm
• [Reid06]: Task farming on Blue Gene

Conclusion: none addressed the proposed “data-centric” part of task farms
Related Work: Task Dispatch

- [Zhou92]: LSF – Load Sharing Cluster Management
- [Bode00]: PBS – Portable Batch Scheduler and Maui Scheduler
- [Anderson04]: BOINC – Task Distribution for Volunteer Computing
- [Thain05]: Condor
- [Robinson07]: Condor-J2 – Turning Cluster Management into Data Management

**Conclusion:** related work is several orders of magnitude slower
Related Work: Resource Provisioning

- [Appleby01]: **Oceano** - SLA Based Management of a Computing Utility
- [Frey02, Mehta06]: **Condor glide-ins**
- [Walker06]: **MyCluster** (based on Condor glide-ins)
- [Ramakrishnan06]: Grid Hosting with Adaptive Resource Control
- [Bresnahan06]: Provisioning of bandwidth
- [Singh06]: Simulations

**Conclusion**: Allows dynamic resizing of resource pool (independent of application logic) based on system load and makes use of light-weight task dispatch
Related Work: Data Management

- [Beynon01]: DataCutter
- [Ranganathan03]: Simulations
- [Ghemawat03,Dean04,Chang06]: BigTable, GFS, MapReduce
- [Liu04]: GridDB
- [Chervenak04,Chervenak06]: RLS (Replica Location Service), DRS (Data Replication Service)
- [Tatebe04,Xiaohui05]: GFarm
- [Branco04,Adams06]: DIAL/ATLAS

**Conclusion:** Our work focuses on the co-location of storage and computations close to each other (i.e. on the same physical resource) while operating in a dynamic environment.
Results

• Abstract task farm model [Dissertation Proposal 2007]
• Practical Realization: Falkon
  – Task Dispatcher [Globus Incubator 2007, SC07, SC08]
  – Resource Provisioning [SC07, TG07]
  – Data Diffusion [NSF06, MSES07, DADC08]
  – Swift Integration [SWF07, NOVA08, SWF08, GW08]
• Applications [NASA06, TG06, SC06, NASA07, SWF07, NOVA08, SC08]
  – Astronomy, medical imaging, molecular dynamics (chemistry and pharmaceuticals), economic modeling
Dispatcher Throughput

- **Fast:**
  - Up to 3700 tasks/sec
- **Scalable:**
  - 54,000 processors
  - 1,500,000 tasks queued

- **Efficient:**
  - High efficiency with second long tasks on 1000s of processors
Dispatcher Performance Profiling

- **GT**: Java WS-Core 4.0.4
- **Java**: Sun JDK 1.6
- **Machine Hardware**: Dual Xeon 3GHz CPUs with HT
- **Machine OS**: Linux 2.6.13-15.16-smp
- **Executors Location**: ANL/UC TG Site, 100 dual CPU Xeon/Itanium nodes, ~2ms latency
- **Workload**: 10000 tasks, “/bin/sleep 0”

![Graph showing CPU Time per Task (ms) for JAVA and C]
- End-to-end execution time:
  - 1260 sec in ideal case
  - 4904 sec → 1276 sec
- Average task queue time:
  - 42.2 sec in ideal case
  - 611 sec → 43.5 sec
- Trade-off:
  - Resource Utilization for Execution Efficiency
Data Diffusion

- No Locality
  - Modest loss of read performance for small # of nodes (<8)
  - Comparable performance with large # of nodes
  - Modest gains in read+write performance

- Locality
  - Significant gains in performance beyond 8 nodes
  - Data-aware scheduler achieves near optimal performance and scalability

![Graphs showing read and read+write throughput for different models and policies.](image)
Falkon Integration with Swift

<table>
<thead>
<tr>
<th>Application</th>
<th>#Tasks/workflow</th>
<th>#Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLAS: High Energy Physics Event Simulation</td>
<td>500K</td>
<td>1</td>
</tr>
<tr>
<td>fMRI DBIC: AIRSN Image Processing</td>
<td>100s</td>
<td>12</td>
</tr>
<tr>
<td>FOAM: Ocean/Atmosphere Model</td>
<td>2000</td>
<td>3</td>
</tr>
<tr>
<td>GADU: Genomics</td>
<td>40K</td>
<td>4</td>
</tr>
<tr>
<td>HNL: fMRI Aphasia Study</td>
<td>500</td>
<td>4</td>
</tr>
<tr>
<td>NVO/NASA: Photorealistic Montage/Morphology</td>
<td>1000s</td>
<td>16</td>
</tr>
<tr>
<td>QuarkNet/l2U2: Physics Science Education</td>
<td>10s</td>
<td>3 – 6</td>
</tr>
<tr>
<td>RadCAD: Radiology Classifier Training</td>
<td>1000s</td>
<td>5</td>
</tr>
<tr>
<td>SIDGrid: EEG Wavelet Processing, Gaze Analysis</td>
<td>100s</td>
<td>20</td>
</tr>
<tr>
<td>SDSS: Coadd, Cluster Search</td>
<td>40K, 500K</td>
<td>2, 8</td>
</tr>
<tr>
<td>SDSS: Stacking, AstroPortal</td>
<td>10Ks ~ 100Ks</td>
<td>2 – 4</td>
</tr>
<tr>
<td>MolDyn: Molecular Dynamics</td>
<td>1Ks ~ 20Ks</td>
<td>8</td>
</tr>
</tbody>
</table>

Swift Architecture

- Specification
- Abstract computation
- SwiftScript Compiler
- Execution Engine (Karajan w/ Swift Runtime)
- Status reporting
- Virtual Node(s)

Provisioning
- Falkon Resource Provisioner
- Amazon EC2
Functional MRI (fMRI)

• Wide range of analyses
  – Testing, interactive analysis, production runs
  – Data mining
  – Parameter studies
fMRI Application

- GRAM vs. Falkon: 85%-90% lower run time
- GRAM/Clustering vs. Falkon: 40%-74% lower run time
Montage Application

- GRAM/Clustering vs. Falkon: 57% lower application run time
- MPI* vs. Falkon: 4% higher application run time
- * MPI should be lower bound

![Bar chart showing performance comparison between GRAM/Clustering, MPI, and Falkon for different components of the Montage Application. The chart indicates significant performance improvements for Falkon.](chart.png)
MolDyn Application

- 244 molecules → 20497 jobs
- 15091 seconds on 216 CPUs → 867.1 CPU hours
- Efficiency: 99.8%
- Speedup: \(\frac{206.9}{8.2}\) faster than GRAM/PBS
- 50 molecules w/ GRAM (4201 jobs) → 25.3 speedup
MARS Economic Modeling on IBM BG/P

- CPU Cores: 2048
- Tasks: 49152
- Micro-tasks: 7077888
- Elapsed time: 1601 secs
- CPU Hours: 894
- Speedup: 1993X (ideal 2048)
- Efficiency: 97.3%
Many Many Tasks: Identifying Potential Drug Targets

200+ Protein target(s) x 5M+ ligands

(Mike Kubal, Benoit Roux, and others)

5/12/2008 Harnessing Grid Resources with Data-Centric Task Farms
DOCK on SiCortex

- CPU cores: 5760
- Tasks: 92160
- Elapsed time: 12821 sec
- Compute time: 1.94 CPU years
- Average task time: 660.3 sec
- Speedup: 5650X (ideal 5760)
- Efficiency: 98.2%
AstroPortal Stacking Service

- **Purpose**
  - On-demand “stacks” of random locations within ~10TB dataset

- **Challenge**
  - Rapid access to 10-10K “random” files
  - Time-varying load

- **Sample Workloads**

<table>
<thead>
<tr>
<th>Locality</th>
<th>Number of Objects</th>
<th>Number of Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>111700</td>
<td>111700</td>
</tr>
<tr>
<td>1.38</td>
<td>154345</td>
<td>111699</td>
</tr>
<tr>
<td>2</td>
<td>97999</td>
<td>49000</td>
</tr>
<tr>
<td>3</td>
<td>88857</td>
<td>29620</td>
</tr>
<tr>
<td>4</td>
<td>76575</td>
<td>19145</td>
</tr>
<tr>
<td>5</td>
<td>60590</td>
<td>12120</td>
</tr>
<tr>
<td>10</td>
<td>46480</td>
<td>4650</td>
</tr>
<tr>
<td>20</td>
<td>40460</td>
<td>2025</td>
</tr>
<tr>
<td>30</td>
<td>23695</td>
<td>790</td>
</tr>
</tbody>
</table>

Sloan Data

Web page or Web Service
AstroPortal Stacking Service with Data Diffusion

Low data locality ➔

High data locality
→ Near perfect scalability
AstroPortal Stacking Service with Data Diffusion

- Big performance gains as locality increases

- 90%+ cache hit ratios
AstroPortal Stacking Service with Data Diffusion

• Aggregate throughput:
  - 39Gb/s
  - 10X higher than GPFS

• Reduced load on GPFS
  - 0.49Gb/s
  - 1/10 of the original load
Hadoop vs. Swift

- Classic benchmarks for MapReduce
  - Word Count
  - Sort
- Swift performs similar or better than Hadoop (on 32 processors)
Mythbusting

- **Embarrassingly** Happily parallel apps are trivial to run
  - Logistical problems can be tremendous
- Loosely coupled apps do not require “supercomputers”
  - Total computational requirements can be enormous
  - Individual tasks may be tightly coupled
  - Workloads frequently involve large amounts of I/O
- Loosely coupled apps do not require specialized system software
- Shared file systems are good all around solutions
  - They don’t scale proportionally with the compute resources
Conclusions & Contributions

• Defined an *abstract model for performance efficiency of data analysis workloads* using data-centric task farms
• Provide a reference implementation (Falkon)
  – Use a streamlined dispatcher to increase task throughput by several orders of magnitude over traditional LRMs
  – Use multi-level scheduling to reduce perceived wait queue time for tasks to execute on remote resources
  – Address data diffusion through co-scheduling of storage and computational resources to improve performance and scalability
  – Provide the benefits of dedicated hardware without the associated high cost
  – Show flexibility/effectiveness on real world applications
    • Astronomy, medical imaging, molecular dynamics (chemistry and pharmaceuticals), economic modeling
  – Runs on real systems with 1000s of processors:
    • TeraGrid, IBM BlueGene/P, SiCortex
More Information

• More information: http://people.cs.uchicago.edu/~iraicu/
• Related Projects:
  – Falkon: http://dev.globus.org/wiki/Incubator/Falkon
  – Swift: http://www.ci.uchicago.edu/swift/index.php
• Collaborators (relevant to this proposal):
  – Ian Foster, The University of Chicago & Argonne National Laboratory
  – Alex Szalay, The Johns Hopkins University
  – Rick Stevens, The University of Chicago & Argonne National Laboratory
  – Yong Zhao, Microsoft
  – Mike Wilde, Computation Institute, University of Chicago & Argonne National Laboratory
  – Catalin Dumitrescu, Fermi National Laboratory
  – Zhao Zhang, The University of Chicago
  – Jerry C. Yan, NASA, Ames Research Center
• Funding:
  – NASA: Ames Research Center, Graduate Student Research Program (GSRP)
  – NSF: TeraGrid

Ioan Raicu, Yong Zhao, Ian Foster, Alex Szalay. "Accelerating Large-scale Data Exploration through Data Diffusion", to appear at International Workshop on Data-Aware Distributed Computing 2008.

Ioan Raicu, Yong Zhao, Ian Foster, Mike Wilde, Zhao Zhang, Ben Clifford, Mihael Hategan, Sarah Kenny. “Managing and Executing Loosely Coupled Large Scale Applications on Clusters, Grids, and Supercomputers”, to appear at GlobusWorld08, part of Open Source Grid and Cluster Conference 2008.


