SCC:
Cluster Storage Provisioning Informed by Application Characteristics and SLAs

Harsha V. Madhyastha*, John C. McCullough, George Porter, Rishi Kapoor, Stefan Savage, Alex C. Snoeren, and Amin Vahdat
UC Riverside* and UC San Diego
Provisioning Hardware for Cluster Applications
Provisioning Hardware for Cluster Applications

“Need 100 Queries/s”
Provisioning Hardware for Cluster Applications

“Need 100 Queries/s”
Provisioning Hardware for Cluster Applications

“Need 100 Queries/s”

“Need 1000 Views/s”
Provisioning Hardware for Cluster Applications

“Need 100 Queries/s”

“Need 1000 Views/s”
Goals for Provisioning at Low-cost

- High performance
- Redundancy
- Supporting multi-tenancy
- High availability
- ...

Our focus: reach performance goal and minimize cost for a single application, emphasizing storage
## Challenge: Large Configuration Space

- Diverse server enclosures/architectures
- Diverse storage options

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th>MB/s (r/w)</th>
<th>IOPS</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2k-rpm</td>
<td>500GB</td>
<td>90/90</td>
<td>125/125</td>
<td>$213</td>
</tr>
<tr>
<td>15k-rpm</td>
<td>146GB</td>
<td>150/150</td>
<td>285/285</td>
<td>$296</td>
</tr>
<tr>
<td>SSD</td>
<td>32GB</td>
<td>250/80</td>
<td>2500/1000</td>
<td>$496</td>
</tr>
<tr>
<td>DRAM</td>
<td>1GB</td>
<td>13k/13k</td>
<td>1.6B/1.6B</td>
<td>$36</td>
</tr>
</tbody>
</table>
Challenge: Large Configuration Space

- Diverse server enclosures/architectures
- Diverse storage options

<table>
<thead>
<tr>
<th>Size</th>
<th>MB/s (r/w)</th>
<th>IOPS</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2k-rpm</td>
<td>500GB</td>
<td>90/90</td>
<td>125/125</td>
</tr>
<tr>
<td>15k-rpm</td>
<td>146GB</td>
<td>150/150</td>
<td>285/285</td>
</tr>
<tr>
<td>SSD</td>
<td>32GB</td>
<td>250/80</td>
<td>2500/1000</td>
</tr>
<tr>
<td>DRAM</td>
<td>1GB</td>
<td>13k/13k</td>
<td>1.6B/1.6B</td>
</tr>
</tbody>
</table>

- Current state-of-the-art:
Challenge: Large Configuration Space

- Diverse server enclosures/architectures
- Diverse storage options

<table>
<thead>
<tr>
<th>Size</th>
<th>MB/s (r/w)</th>
<th>IOPS</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2k-rpm</td>
<td>500GB</td>
<td>90/90</td>
<td>125/125</td>
</tr>
<tr>
<td>15k-rpm</td>
<td>146GB</td>
<td>150/150</td>
<td>285/285</td>
</tr>
<tr>
<td>SSD</td>
<td>32GB</td>
<td>250/80</td>
<td>2500/1000</td>
</tr>
<tr>
<td>DRAM</td>
<td>1GB</td>
<td>13k/13k</td>
<td>1.6B/1.6B</td>
</tr>
</tbody>
</table>

- Current state-of-the-art:
  - Apply rules-of-thumb from experience
  - Trial-and-error with various configurations
  - Configuration duplicated to scale-out
Scale-out Shortcomings for Photo-sharing Application

![Graph showing cost versus SLA (uploads/s)]

- **scale-out**
- **scc**

Cost (k$) vs. SLA (uploads/s) with a 4x increase in cost at higher SLA values.
Scale-out Shortcomings for Photo-sharing Application

**Insight:** Match storage to workload, leverage heterogeneity
Goal: Understand Configuration Space

• What is a low-cost configuration now?
• What will low-cost configurations look like in the future?

Measure “in-the-small” → Model Application Performance → Predict “in-the-large”
scc: Storage Configuration Compiler

Cluster Building-Blocks

scc
scc: Storage Configuration Compiler

Cluster Building-Blocks

Application Model

scc
scc: Storage Configuration Compiler

Cluster Building-Blocks → scc
Application Model → scc
SLA Specification → scc
scc: Storage Configuration Compiler

Cluster Building-Blocks

Application Model

SLA Specification

Cost

SLA
scc: Storage Configuration Compiler

Cluster Building-Blocks

Application Model

SLA Specification

Cost

SLA
Outline

- Modeling Applications and Hardware
- Computing low-cost configurations
- Example
- Validation
- Applications of scc
Cluster Building Blocks

- Many types of servers
Cluster Building Blocks

- Many types of servers
- Cores
Cluster Building Blocks

- Many types of servers
  - Cores
  - RAM

CPU
RAM
Cluster Building Blocks

• Many types of servers
  • Cores
  • RAM
  • Storage
Cluster Building Blocks

- Many types of servers
  - Cores
  - RAM
  - Storage
  - I/O & Network
Application Model

• Breakdown application into:
  • Tasks (Computation)
  • Datasets (Storage)
  • Edges between Tasks and Datasets (I/O)
  • Edges among Tasks (dependencies)
Example Model: Photo-Sharing

Tasks  Datasets
Example Model: Photo-Sharing

Tasks

Datasets

- Photos
- Thumbnails
- (Tag, Photo)
Example Model: Photo-Sharing

Tasks

- Photo Upload
- Thumbnail
- Write Tags

Datasets

- Photos
- Thumbnails
- (Tag, Photo)
Example Model: Photo-Sharing

Tasks

- Photo Upload
- Thumbnail
- Write Tags

Datasets

- Photos
- Thumbnails
- (Tag, Photo)
Example Model: Photo-Sharing

Tasks

Photo Upload
Thumbnail
Write Tags

Datasets

Photos
Thumbnails
(Tag, Photo)
Example Model: Photo-Sharing

Tasks

- Photo Upload
- Thumbnail
- Write Tags
- View Photo

Datasets

- Photos
- Thumbnails
- (Tag, Photo)
Example Model: Photo-Sharing

Tasks

Photo Upload
Thumbnail
Write Tags
View Photo

Datasets

Photos
Thumbnails
(Tag, Photo)
Example Model: Photo-Sharing

Tasks

Photo Upload
Thumbnail
Write Tags
View Photo
View Tag

Datasets

Photos
Thumbnails
(Tag, Photo)
Example Model: Photo-Sharing

Tasks

Photo Upload
Thumbnail
Write Tags
View Photo
View Tag

Datasets

Photos
Thumbnails
(Tag, Photo)
Example Model: Photo-Sharing

Tasks

Photo Upload
Thumbnail
Write Tags
View Photo
View Tag

Datasets

Photos
Thumbnails
(Tag, Photo)
Example Model: Photo-Sharing

Tasks

Photo Upload
Thumbnail
Write Tags
View Photo
View Tag

Datasets

Photos
Thumbnails
(Tag, Photo)
Example Model: Photo-Sharing

Tasks

Photo Upload
Thumbnail
Write Tags
View Photo
View Tag

Datasets

Photos
 Thumbnails
(Tag, Photo)

11x200kB
1x4kB
10x1kB
10x1kB
10x1kB
1x200kB
10x4kB
10x1kB
10x1kB

5ms
100ms
10ms
1ms
10ms

Friday, February 17, 12
Example Model: Photo-Sharing

Photo Upload

Thumbnail

Write Tags

View Photo

View Tag

Tasks

Datasets

Photos

1TB

 Thumbnails

20GB

(Tag, Photo)

2GB
Photo-Sharing SLA

Tasks
- Photo Upload
- Thumbnail
- Write Tags
- View Photo
- View Tag

Datasets
- Photos
- Thumbnails
- (Tag, Photo)
Photo-Sharing SLA

Tasks
- Photo Upload
- Thumbnail
- Write Tags
- View Photo
- View Tag

Datasets
- Photos
- Thumbnails
- (Tag, Photo)
Outline

- Modeling Applications and Hardware
- Computing low-cost configurations
- Example
- Validation
- Applications of scc
Navigating the Configuration Space
Navigating the Configuration Space
Navigating the Configuration Space
Navigating the Configuration Space

- Cost
- SLA

Friday, February 17, 12
Guiding Principle to Meet SLA

- Complex interaction across storage-type and dataset assignments
- Need to consider costs of meeting SLA for each permutation
- Our configuration space is:
  - $D$ datasets, $S$ storage-types $\rightarrow D^S$ configs
### Meeting the SLA at Low-cost

<table>
<thead>
<tr>
<th>Photos</th>
<th>Thumbs</th>
<th>Tags</th>
</tr>
</thead>
</table>
Meeting the SLA at Low-cost

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Photos</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Thumbs</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Tags</td>
<td>7.2krpm</td>
</tr>
</tbody>
</table>
Meeting the SLA at Low-cost

<table>
<thead>
<tr>
<th></th>
<th>7.2krpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photos</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Thumbs</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Tags</td>
<td>7.2krpm</td>
</tr>
</tbody>
</table>

# storage units per-dataset
Meeting the SLA at Low-cost

<table>
<thead>
<tr>
<th>Photos</th>
<th>7.2krpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thumbs</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Tags</td>
<td>7.2krpm</td>
</tr>
</tbody>
</table>

# storage units per-dataset

Assign CPUs for computation
Meeting the SLA at Low-cost

<table>
<thead>
<tr>
<th>Photos</th>
<th>7.2krpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thumbs</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Tags</td>
<td>7.2krpm</td>
</tr>
</tbody>
</table>

# storage units per-dataset

Assign CPUs for computation

Pack into servers
Meeting the SLA at Low-cost

<table>
<thead>
<tr>
<th>Photos</th>
<th>7.2krpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thumbs</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Tags</td>
<td>7.2krpm</td>
</tr>
</tbody>
</table>

# storage units per-dataset

Assign CPUs for computation

Pack into servers

Compute cost
# Meeting the SLA at Low-cost

<table>
<thead>
<tr>
<th>Photos</th>
<th>7.2krpm</th>
<th>7.2krpm</th>
<th>7.2krpm</th>
<th>7.2krpm</th>
<th>7.2krpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thumbs</td>
<td>7.2krpm</td>
<td>7.2krpm</td>
<td>7.2krpm</td>
<td>7.2krpm</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Tags</td>
<td>7.2krpm</td>
<td>15krpm</td>
<td>SSD</td>
<td>7.2k+DRAM</td>
<td>15k+DRAM</td>
</tr>
</tbody>
</table>

- Pack into servers
- Compute cost

**Assign CPUs for computation**

**# storage units per-dataset**
Meeting the SLA at Low-cost

<table>
<thead>
<tr>
<th>Photos</th>
<th>7.2krpm</th>
<th>7.2krpm</th>
<th>7.2krpm</th>
<th>7.2krpm</th>
<th>7.2krpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thumbs</td>
<td>7.2krpm</td>
<td>7.2krpm</td>
<td>7.2krpm</td>
<td>7.2krpm</td>
<td>7.2krpm</td>
</tr>
<tr>
<td>Tags</td>
<td>7.2krpm</td>
<td>15krpm</td>
<td>SSD</td>
<td>7.2k+DRAM</td>
<td>15k+DRAM</td>
</tr>
</tbody>
</table>

Find minimum cost

# storage units per-dataset

Assign CPUs for computation

Pack into servers

Compute cost
Detail: How Many Storage Units?

- # of storage units per-dataset
- Assign CPUs for computation
- Pack into servers
- Compute cost
Detail: How Many Storage Units?

- SLA+Model=Requirements
- Need to match requirements to storage units

# of storage units per-dataset
Assign CPUs for computation
Pack into servers
Compute cost
Detail: How Many Storage Units?

- SLA+Model=Requirements
- Need to match requirements to storage units
  - Capacity
  - IOPS
    - Write-heavy: Short-stroke disks
    - Read-heavy: Consider RAM for caching
Detail: How Many CPUs?

- Enough CPUs to satisfy computation requirements
- Linearly extrapolate computation time
- Minimum one core/server
- (Details in paper)
Detail: Fits in how many servers?

- # of storage units per-dataset
- Assign CPUs for computation
- Pack into servers
- Compute cost
Detail: Fits in how many servers?

- Integer Linear Programming
- Multiple server types, each with different constraints:
  - I/O Bus Bandwidth
  - I/O Slots
  - CPU Cores
  - Network Bandwidth
Detail: Fits in how many servers?

- Integer Linear Programming
- Multiple server types, each with different constraints:
  - I/O Bus Bandwidth
  - I/O Slots
  - CPU Cores
  - Network Bandwidth
- Sequential-workload → Bus/Network constrained
- IOPS-workload → Slot capacity constrained
Detail: How much does it cost?

- ILP minimizes cost:
  - Capital expenses
    - Each component
  - Operation expenses
    - Power & cooling
Outline

- Modeling Applications and Hardware
- Computing low-cost configurations
- Example
- Validation
- Applications of scc
Example Application: Photo-Sharing

Tasks

- Photo Upload
- Thumbnail
- Write Tags
- View Photo
- View Tag

Datasets

- Photos
- Thumbnails
- (Tag, Photo)
Example: Building Blocks

- Building blocks
  - 3GHz Intel Xeon (max 4 cores)
  - 146GB 15k-rpm Disks
  - 500GB 7.2k-rpm Disks (max 4 per machine)
  - 32GB SSD
  - 1 GB RAM (max 15 per machine)
  - Gigabit Ethernet
Example: SLA

- SLA Requirement
  - 100 image uploads/s
  - 300 photo views/s
  - 100 tags views/s
Example: Final Stages for Photo-sharing

Photos
- 15k-rpm
- 15k-rpm
- 15k-rpm
- 15k-rpm

Thumbnails
- SSD

Tags
- 7200rpm w/DRAM

WWW
- CPU
- CPU
- RAM
- CPU
- CPU
- RAM
Example: Final Stages for Photo-sharing

Photos

15k-rpm

15k-rpm

15k-rpm

15k-rpm

 Thumbnails

SSD

Tags

7200rpm w/DRAM

WWW

CPU

CPU

CPU

CPU

RAM

RAM
Example: Final Stages for Photo-sharing

Photos
- 15k-rpm
- 15k-rpm
- 15k-rpm
- 15k-rpm

Thumbnails
- SSD

Tags
- 7200rpm w/DRAM

WWW
- CPU
- CPU
- RAM
- CPU
- CPU
- RAM
Example: Final Stages for Photo-sharing

Photos
- 15k-rpm
- 15k-rpm
- 15k-rpm
- CPU
- RAM

Thumbnails
- SSD

Tags
- 7200rpm w/DRAM
- CPU
- RAM

WWW
- CPU
- CPU
- CPU
- CPU
- RAM
- RAM
- RAM
- RAM
Example: Final Stages for Photo-sharing

Photos

15k-rpm
15k-rpm
15k-rpm

CPU
RAM

Thumbnails

SSD

Tags

7200rpm w/DRAM

CPU
RAM

WWW

CPU
CPU
RAM
CPU
CPU
RAM
Outline

- Modeling Applications and Hardware
- Computing low-cost configurations
- Example
- Validation
- Applications of scc
Experimental Methodology

- Built three applications
  - Photo-sharing
  - Product Search
  - Terasort
- Micro-benchmarked each to create model
- Deployed scc output to cluster of machines
Validation: scc meets SLA at lower cost for Photo-sharing

Cost relative to scc

Fraction of SLA satisfied

0.96 1.00 1.04 1.32
Validation: scc meets SLA at lower cost for Photo-sharing
Validation: scc meets SLA at lower cost for Photo-sharing

Fraction of SLA satisfied

Cost relative to scc
Validation: scc meets SLA at lower cost for Photo-sharing

Cost relative to scc

Fraction of SLA satisfied

0.96 1.00 1.04 1.32

0.5 0.6 0.7 0.8 0.9 1.0 1.1

SCC
Validation: scc meets SLA at lower cost for Photo-sharing
Validation: scc meets SLA at lower cost for Photo-sharing

![Graph showing the comparison of different alternatives against the SLA satisfaction and cost relative to scc.](image.png)
Validation: SCC meets SLA at lower cost for Product Search

![Graph showing fraction of SLA satisfied vs cost relative to SCC]
Validation: scc meets SLA at lower cost for Terasort
Validation: SCC meets SLA at lower cost for Terasort

![Graph showing cost vs. SLA satisfaction for different options. The graph indicates that SCC meets SLA at a lower cost compared to other alternatives. The Best Practice is highlighted.]
Validation: scc meets SLA at lower cost for Terasort

Fraction of SLA satisfied vs Cost relative to scc

Best Practice

- scc
- Alt1
- Alt2
- Alt3
- Alt4
- Alt5
- 2M 2R
- 2M 3R
- 1M 3R
- 2M 2R+
Outline

- Modeling Applications and Hardware
- Computing low-cost configurations
- Experimental methodology
- Validation
- Applications of scc
### Storage Type Regimes

<table>
<thead>
<tr>
<th>Uploads/s</th>
<th>Photos</th>
<th>Thumbnails</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5</td>
<td>Disk</td>
<td>Disk</td>
<td>Disk</td>
</tr>
<tr>
<td>5-25</td>
<td>Disk</td>
<td>Disk</td>
<td>Disk w/ DRAM</td>
</tr>
<tr>
<td>25-330</td>
<td>Disk</td>
<td>SSD</td>
<td>Disk w/ DRAM</td>
</tr>
<tr>
<td>330-930</td>
<td>SSD</td>
<td>Disk w/ DRAM</td>
<td>Disk w/ DRAM</td>
</tr>
<tr>
<td>930-10k</td>
<td>Disk w/ DRAM</td>
<td>Disk w/ DRAM</td>
<td>Disk w/ DRAM</td>
</tr>
</tbody>
</table>
Output Sensitivity to Model Parameters

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Range with same architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photo size</td>
<td>50kB $\leftrightarrow$ 200kB $\rightarrow$ 850kB</td>
</tr>
<tr>
<td>Thumbnail size</td>
<td>1kB $\leftrightarrow$ 4kB $\rightarrow$ 30kB</td>
</tr>
<tr>
<td>SSD unit price</td>
<td>$200$ $\leftrightarrow$ $450$ $\rightarrow$ $900$</td>
</tr>
</tbody>
</table>
Output Sensitivity to Model Parameters

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Range with same architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photo size</td>
<td>50kB ← 200kB → 850kB</td>
</tr>
<tr>
<td>Thumbnail size</td>
<td>1kB ← 4kB → 30kB</td>
</tr>
<tr>
<td>SSD unit price</td>
<td>$200 ← $450 → $900</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sensitive to what hardware cost?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photos</td>
<td>20% drop in 7.2k-rpm drive price</td>
</tr>
<tr>
<td>Thumbnails</td>
<td>92% drop in DRAM price</td>
</tr>
<tr>
<td>Tags</td>
<td>31% drop in 15k-rpm drive price</td>
</tr>
</tbody>
</table>
Conclusion
Conclusion

• Our scc tool finds low-cost cluster configurations
  • Saves 2-4.5x over simple scale-out for Photo-share and Product Search
  • 16% lower cost for Terasort vs. Map Reduce best-practice configuration
Conclusion

- Our scc tool finds low-cost cluster configurations
  - Saves 2-4.5x over simple scale-out for Photo-share and Product Search
  - 16% lower cost for Terasort vs. Map Reduce best-practice configuration
- Better matches hardware to application requirements
Conclusion

• Our scc tool finds low-cost cluster configurations
  • Saves 2-4.5x over simple scale-out for Photo-share and Product Search
  • 16% lower cost for Terasort vs. Map Reduce best-practice configuration
• Better matches hardware to application requirements
• Useful for predicting output sensitivity and future needs
Conclusion

- Our scc tool finds low-cost cluster configurations
  - Saves 2-4.5x over simple scale-out for Photo-share and Product Search
  - 16% lower cost for Terasort vs. Map Reduce best-practice configuration
- Better matches hardware to application requirements
- Useful for predicting output sensitivity and future needs
- Future work
  - More precise network models
  - Cloud deployment