MODEL ENSEMBLE TOOLS FOR SELF-MANAGEMENT IN DATA CENTERS

Jin Chen

*Gokul Soundararajan**

Saeed Ghanbari

Cristiana Amza

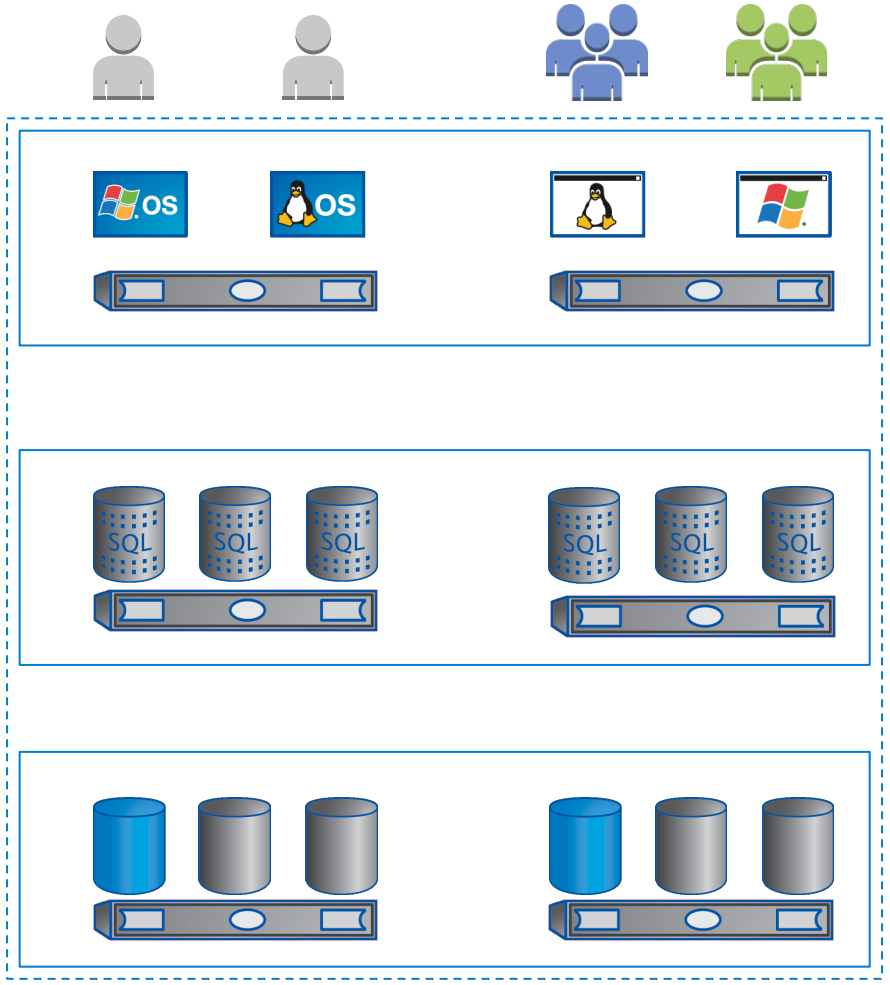
University of Toronto

**currently at NetApp Advanced Technology Group*

Shared datacenter platforms

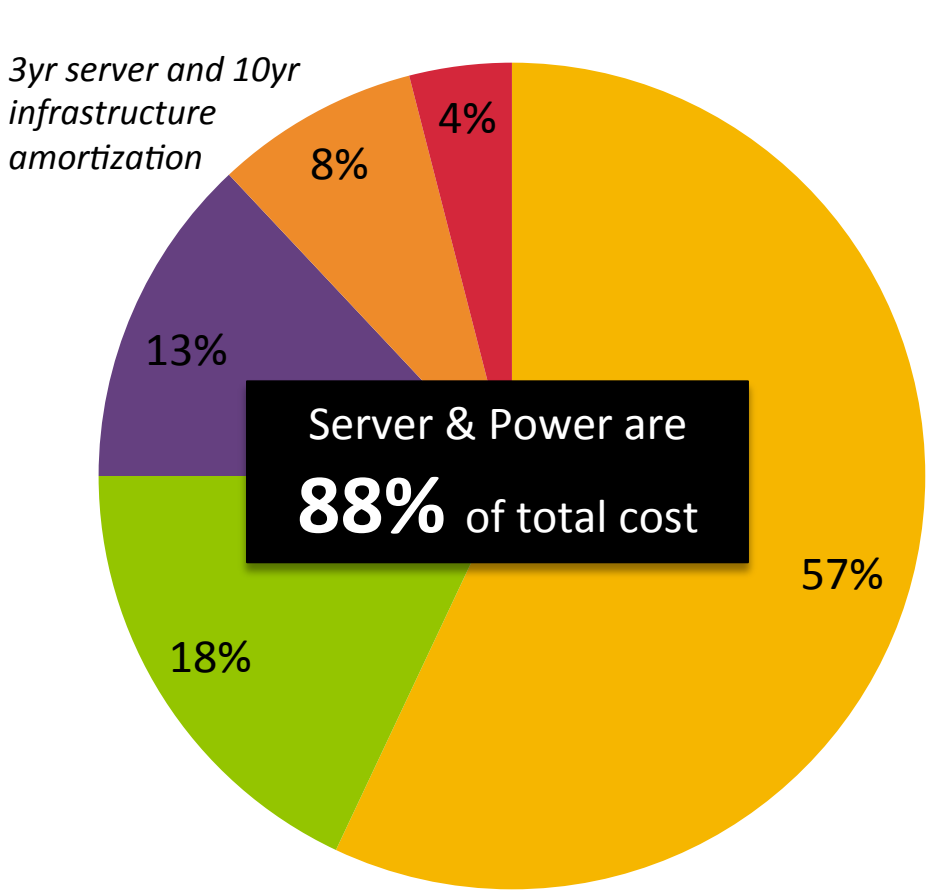
Amazon Web Services
(Relational Data Service)

Google Compute Engine
(Cloud SQL)



Datacenter

Costs in a datacenter



Estimated costs of datacenter:

46,000 servers

$3,500,000 per month to run

Data courtesy of James Hamilton [SIGMOD'11 Keynote]

Ways to reduce costs

Service Provider Administrator

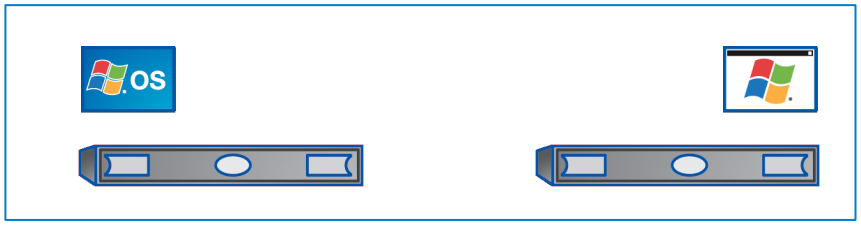
- **Use less resources:** Customer wants 1000 TPS. What is the most efficient (e.g., CPU/Memory) to deliver it?
- **Share resources:** Can I place customer A's DB along side customer B's DB? Will their service-levels be met?

Customer DBA

Provider: Support multiple customers

More customers arrive!

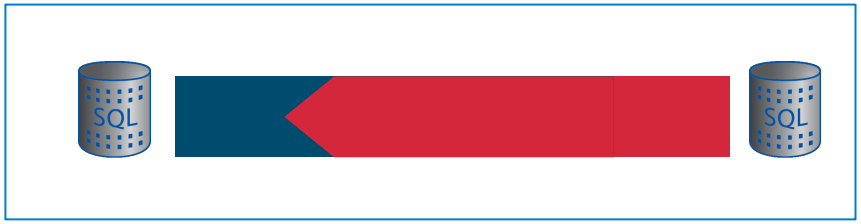
Blue is running
High-priority app
Strict SLOs



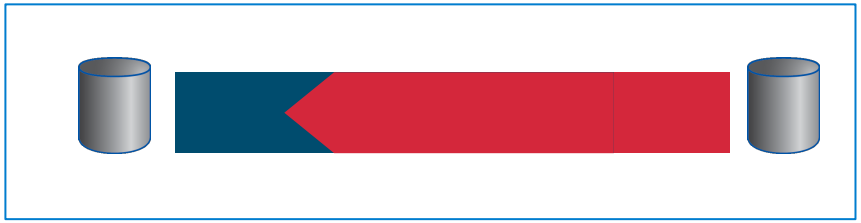
Red is running
Low-priority app
Best-effort SLO

Servers

Expects
< 5ms query response time



Databases



Storage

How to reduce costs

Service Provider Administrator

- **Use less resources:** Customer wants 1000 TPS. What is the most efficient (e.g., CPU/Memory) to deliver it?
- **Share resources:** Can I place customer A's DB along side customer B's DB? Will their service-levels be met?

Customer DBA

- **Use the right amount of resources:** What will be the performance (e.g., query latency) if I use 8GB of RAM instead of 16GB?
- **Solve performance problems:** I'm only getting 500 TPS. What's wrong? Is the cloud to blame?

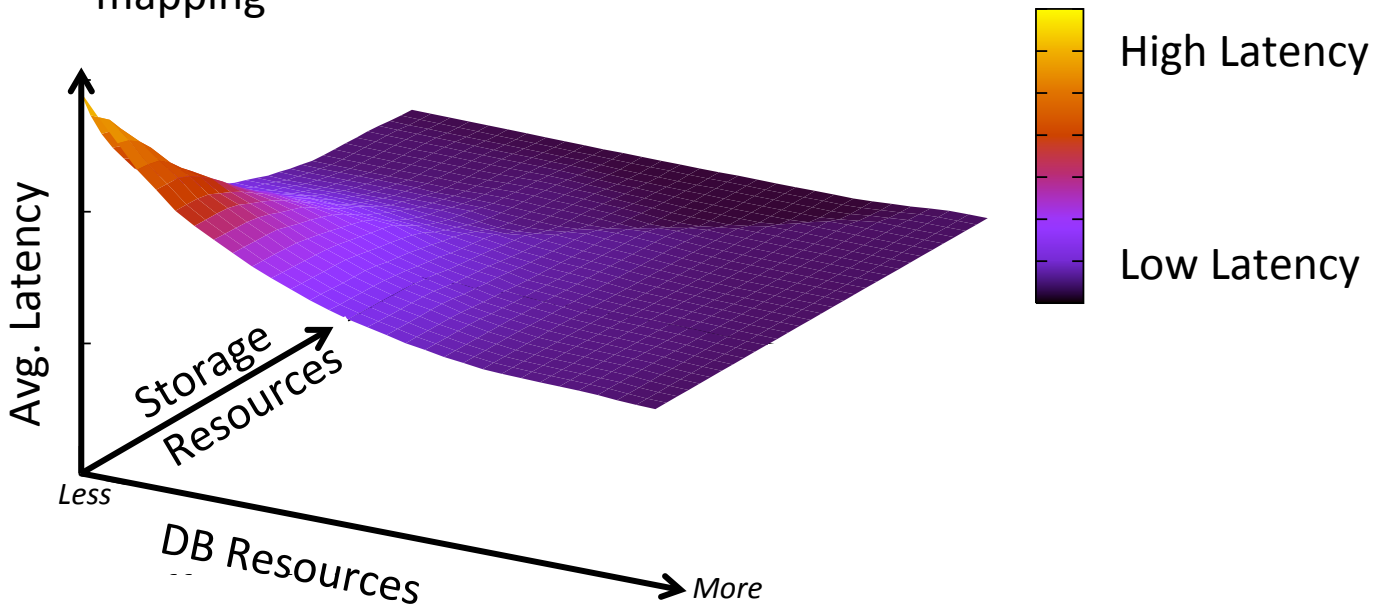
Need to build **performance models** to understand

What are performance models?

Provide

resource-to-performance

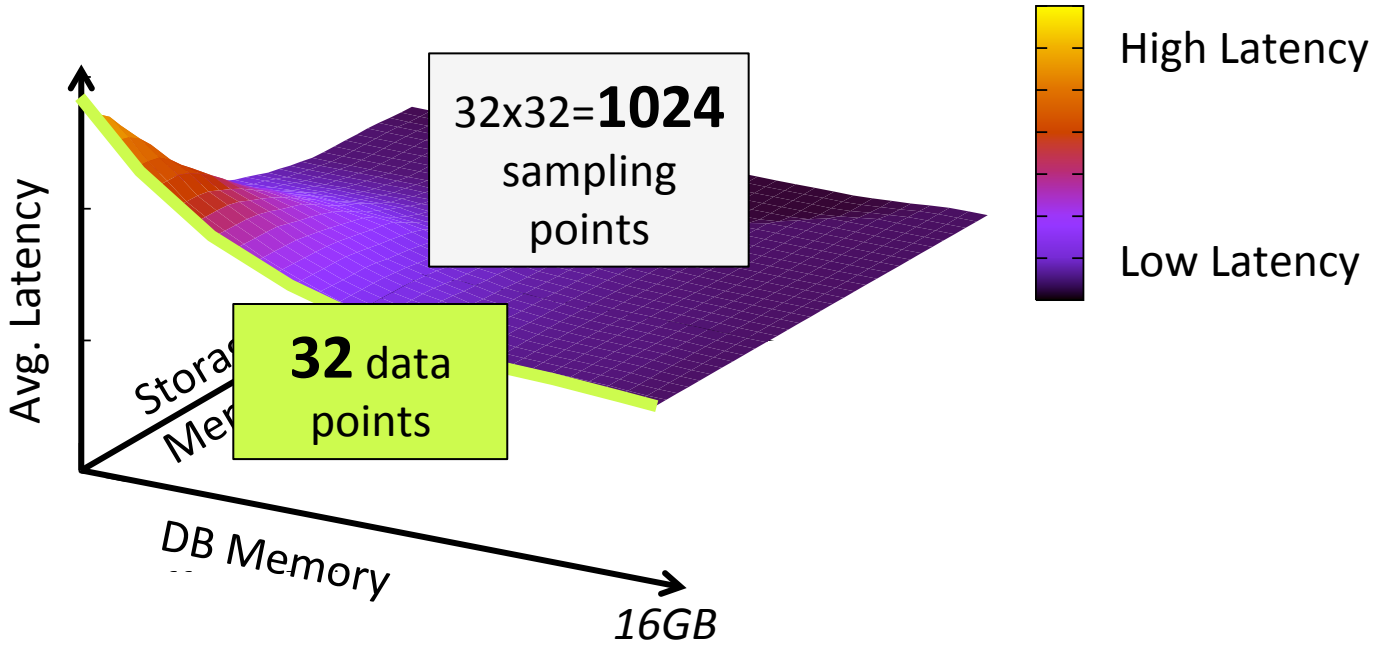
mapping



- ✓ Dynamic Resource Allocation [FAST'09]
- ✓ Capacity Planning [SIGMOD'13]
- ✓ What-if Queries [SIGMETRICS'10]
- ✓ Performance Anomaly Detection [SIGMETRICS'10]

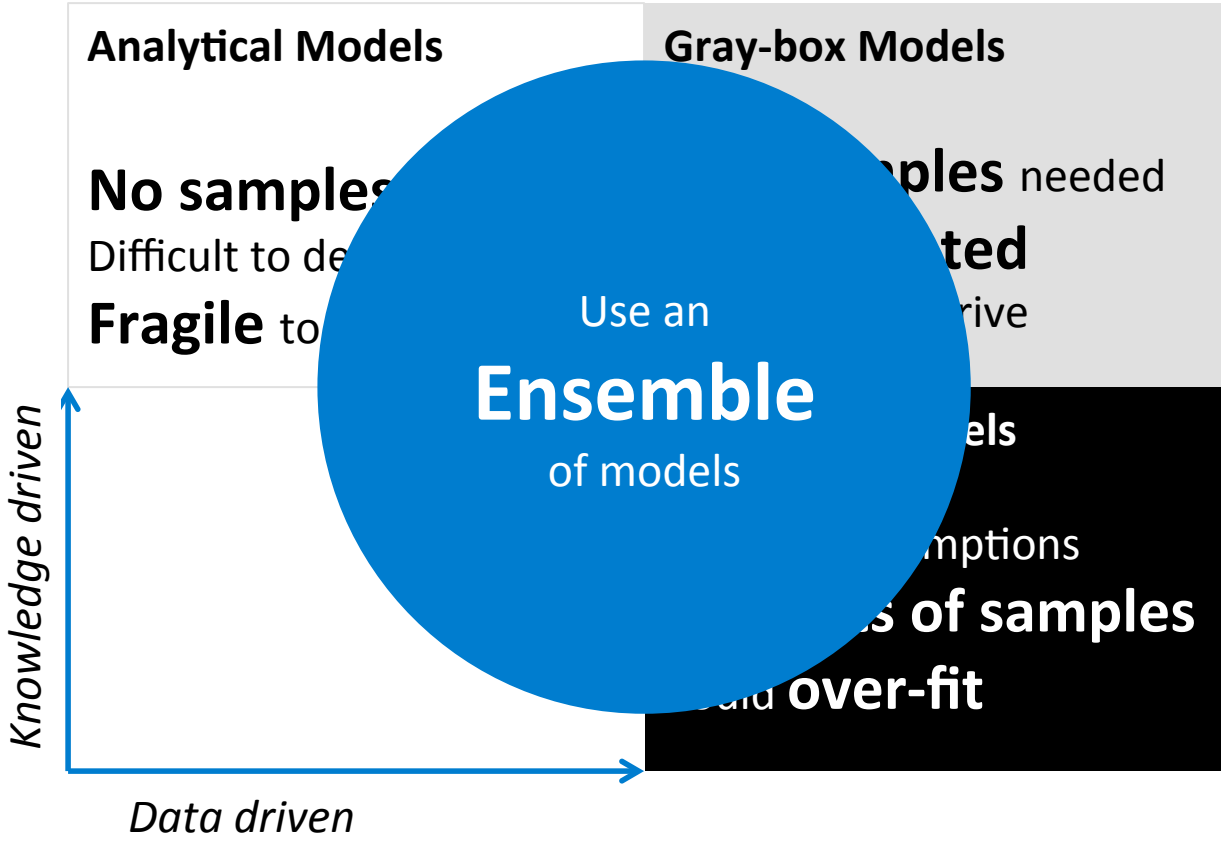
Building performance models takes time

*Exhaustive sampling takes **11 days!***



Actuate a live system and take experimental samples.
Sample in 512MB chunks; 15 minutes for each point

Types of performance models

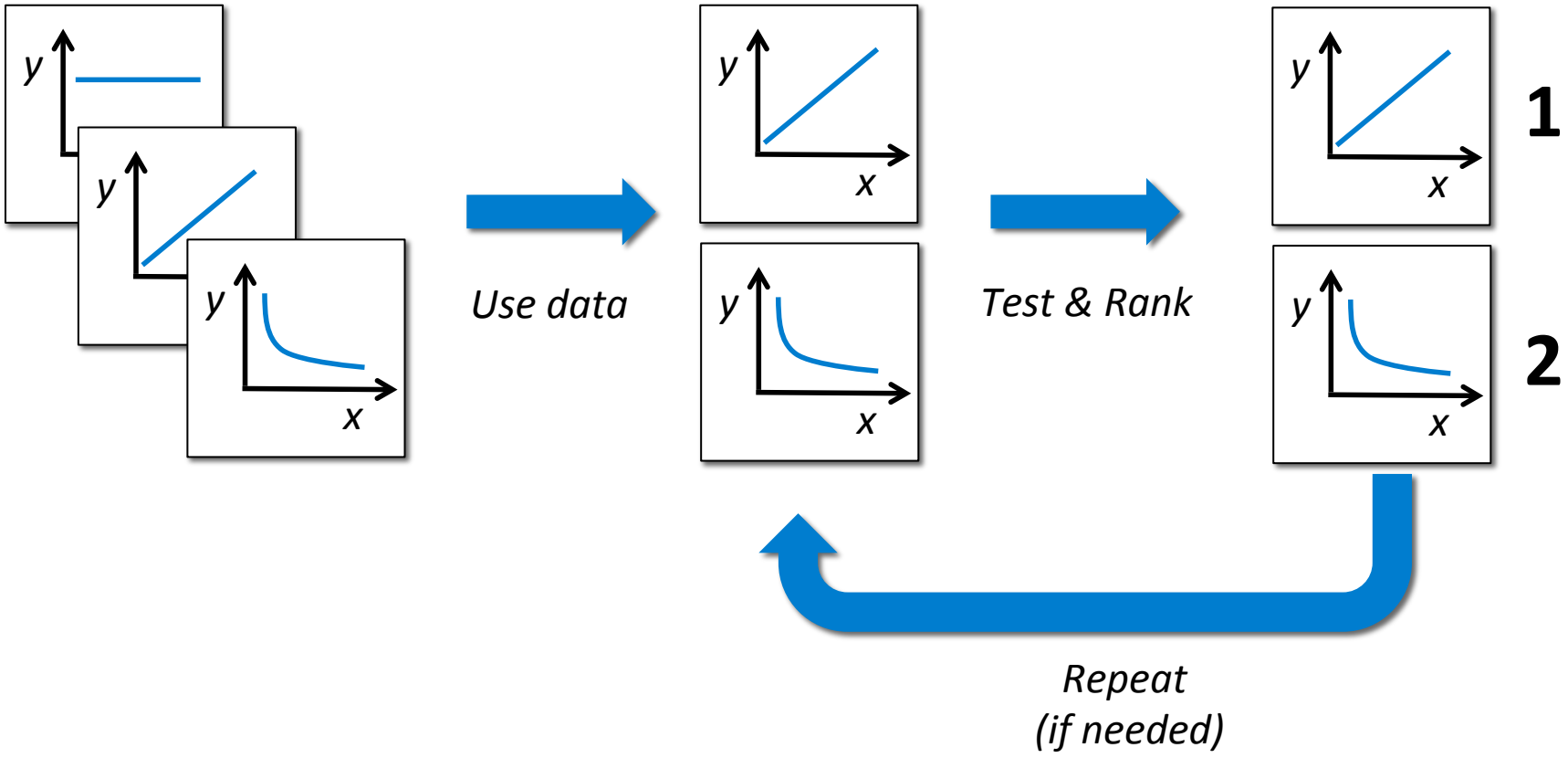


Ensemble approach

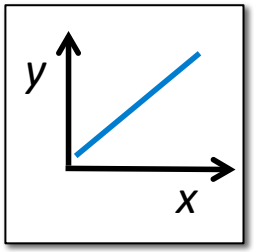
1. Specify **trends and patterns**

2. Automatically **tune** the models

3. **Rank** the models use a **blend**



Specifying trends and patterns



$$\hat{y} = \alpha x_i + \beta$$

SelfTalk

Language to **describe relationships**

Provide a **catalog** of **common functions**

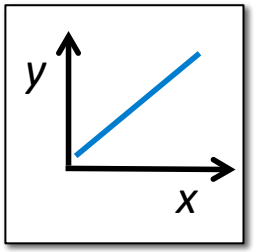
```
DEFINE RELATION linear {
  PARAMETER a,b : number,
  INPUT x:number-array, y:number-array,
  ...
  FUNCTION confidence
  {
    OUTPUT confidence:number
    LANGUAGE 'matlab'
    SCRIPT
    $                          Goodness-of-fit
      y_hat = a.*x .+ b;
      confidence = R2(y,y_hat);
      //calculate residuals
      ...
    $
  }
  ...
}
```

Specifies model **inputs** and **parameters**

Curve-fitting and validation **algorithms**

Details in SIGMETRICS'10 paper

Refine models using data



$$\hat{y} = \alpha x_i + \beta$$

Use **hints** to **link** relations to metrics

```
HINT myHint
RELATION LINEAR(x,y)
METRIC (x,y) {
  x.name=MySQL.CPU
  y.name=MySQL.QPS
}
CONTEXT (a) {
   a.name=MySQL.BufPoolAlloc
   a.value >= 512MB
}
```

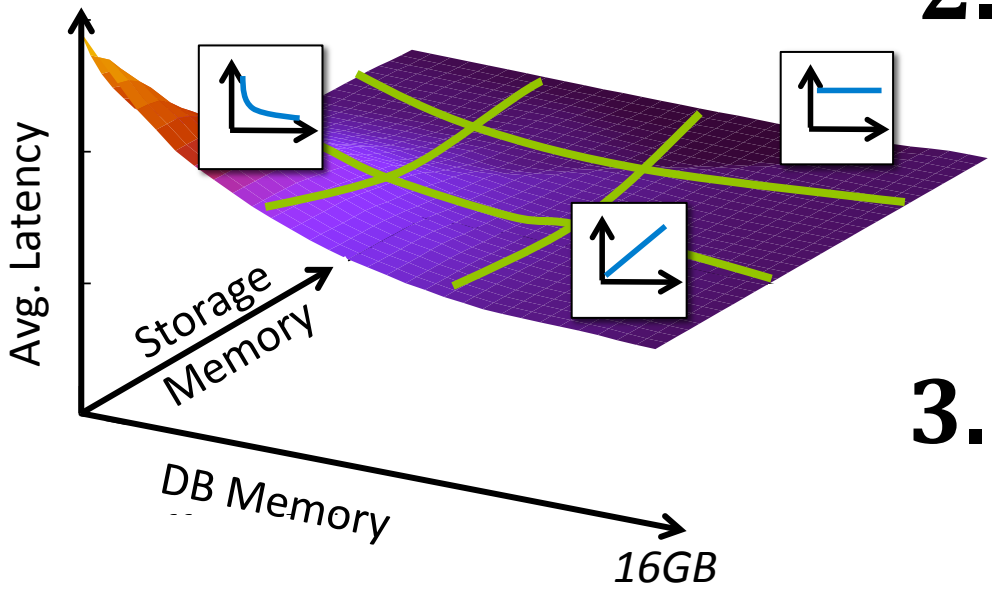
Hints that CPU **linearly correlated** to QPS

But **working-set** should be in **RAM**

Learns parameters using data
(or requests more data)

Rank models and blend

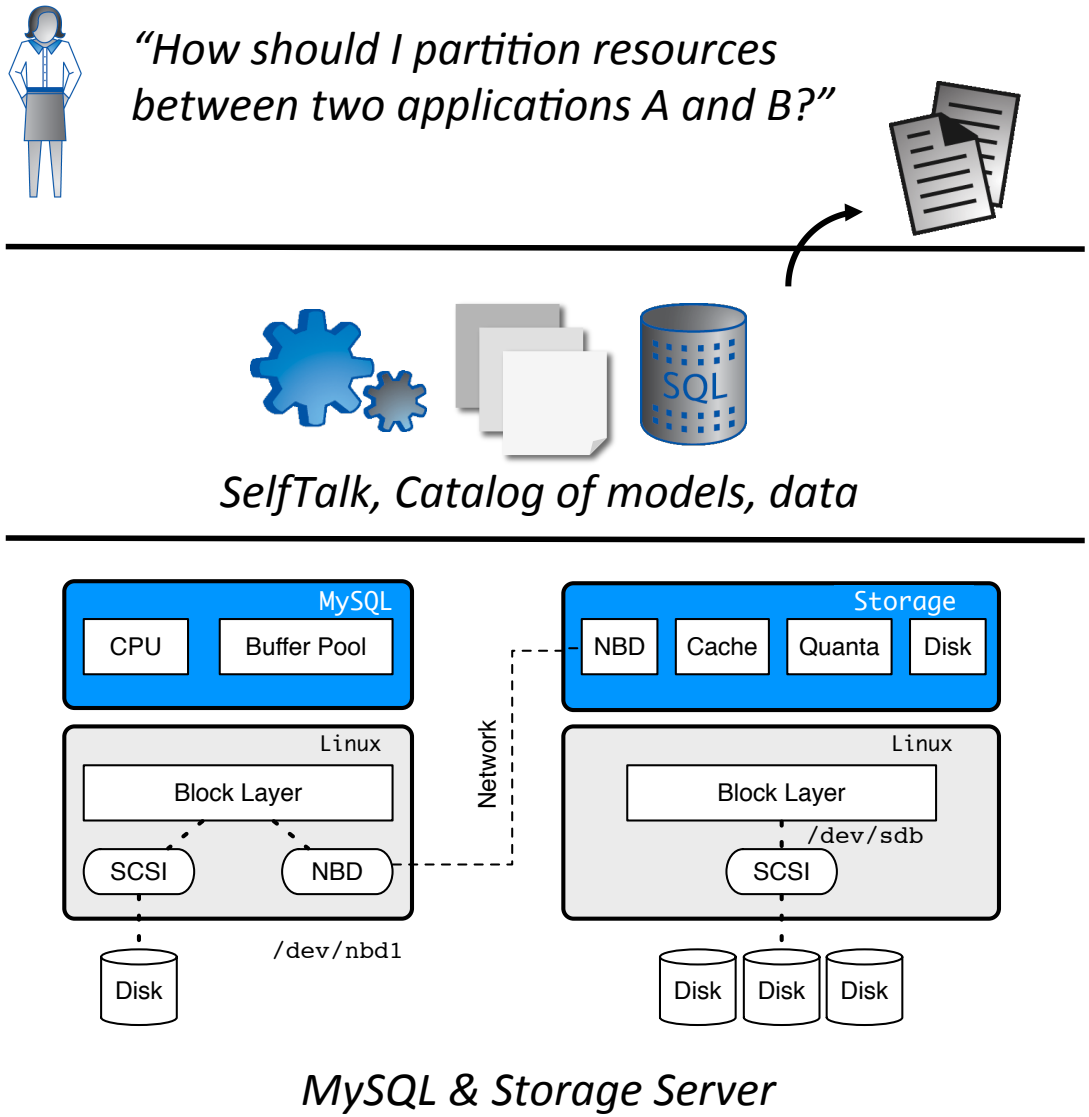
1. Divide search space into **regions**



2. n-fold **cross-validation** to **rank**

3. Associate **best-model** to **region**

Prototype


"How should I partition resources between two applications A and B?"
SQL
SelfTalk, Catalog of models, data
MySQL
CPU
Buffer Pool
Linux
Block Layer
SCSI
NBD
/dev/nbd1
Disk
Network
Storage
NBD
Cache
Quanta
Disk
Linux
Block Layer
/dev/sdb
SCSI
Disk
Disk
Disk
MySQL & Storage Server

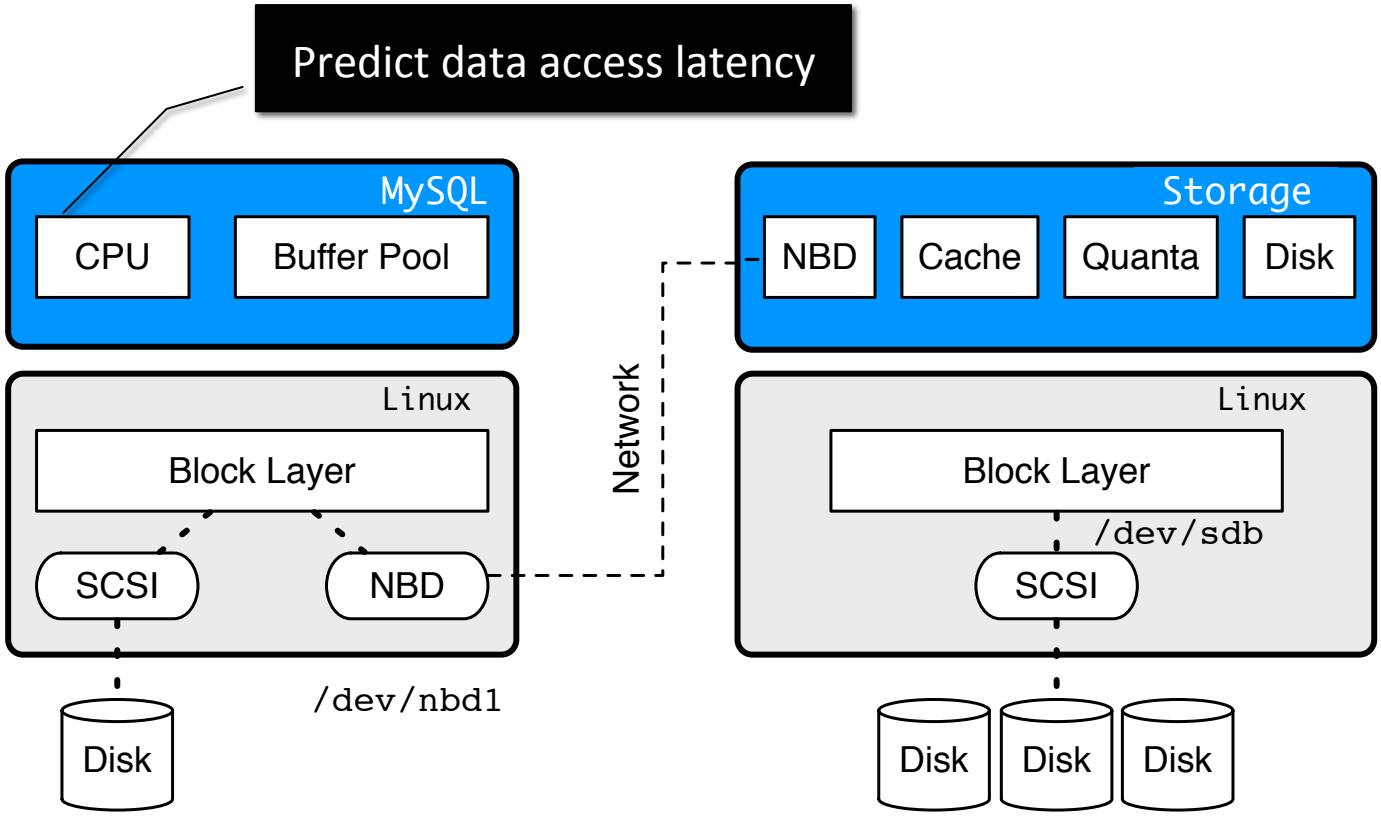
Set of model templates

- Analytical models
  - A-STOR: analytical model for predicting storage hierarchy latency [FAST'09]
- Black-box models
  - B-SVM: support vector machine regression
  - B-CNST: average values as prediction results
- Gray-box models
  - G-RGN: region-based curve fitting model
  - G-INV: inversely-proportional curve fitting model

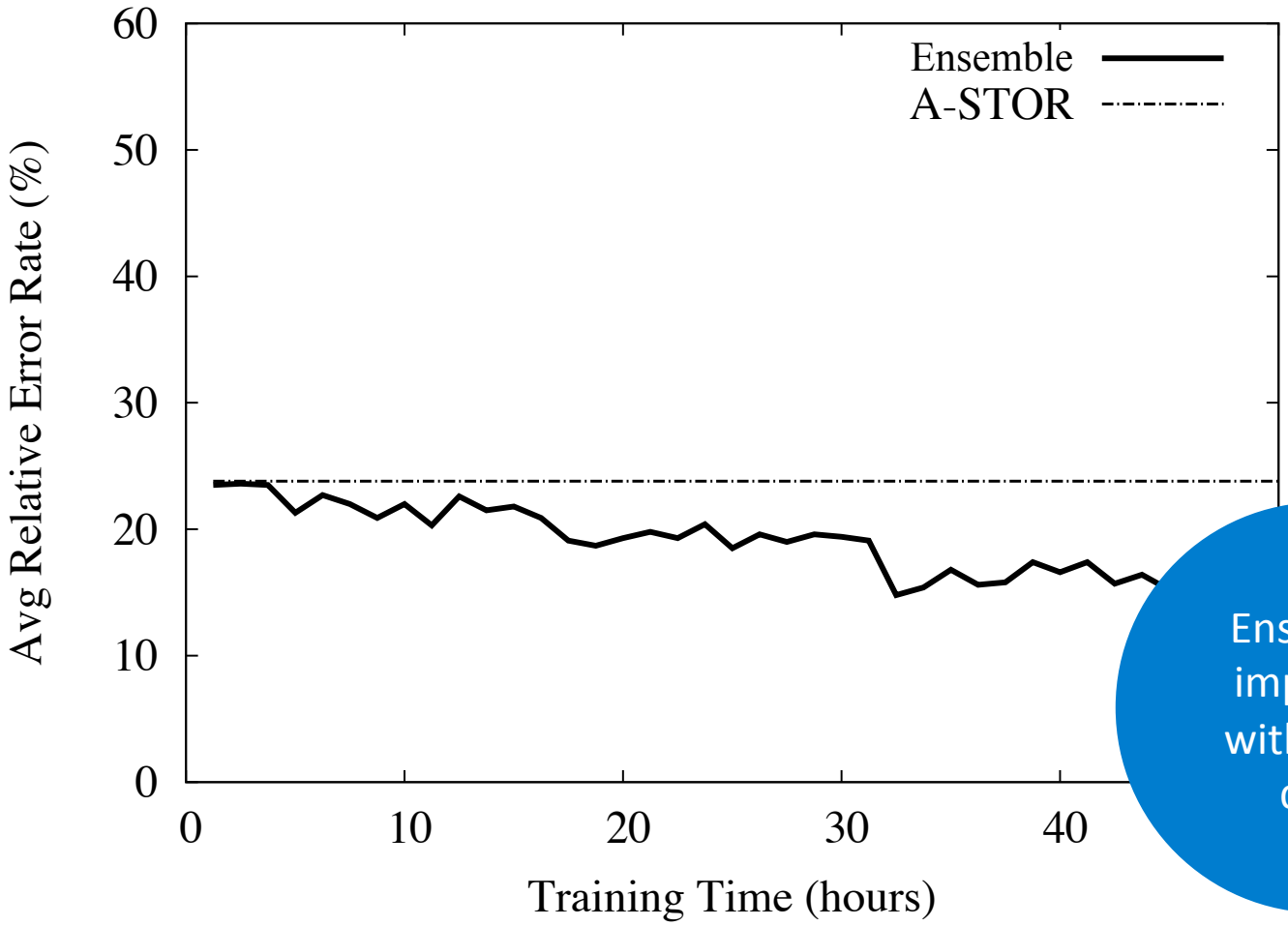
Workloads and datasets

- Benchmarks
  - **TPC-W:** E-commerce (online bookstore) benchmark
  - **TPC-C:** Wholesale parts supplier with warehouses and sales districts
- Experiments
  - Conducted over a period of **6 months**
  - Data from OS, MySQL, and storage server
  - At the end, the structured data was **10GB**

Modeling MySQL IO subsystem



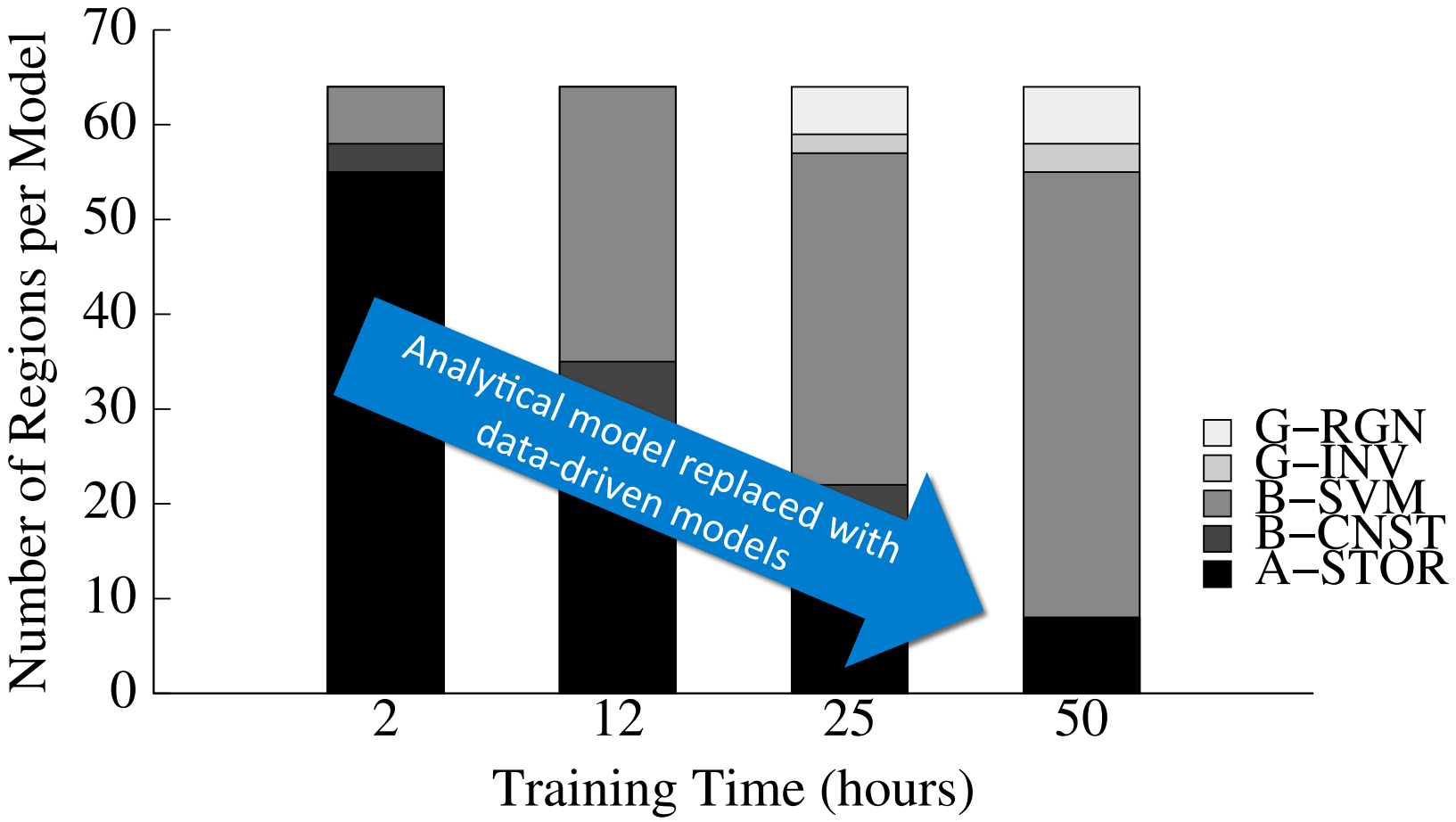
Predicting TPC-W buffer pool latencies



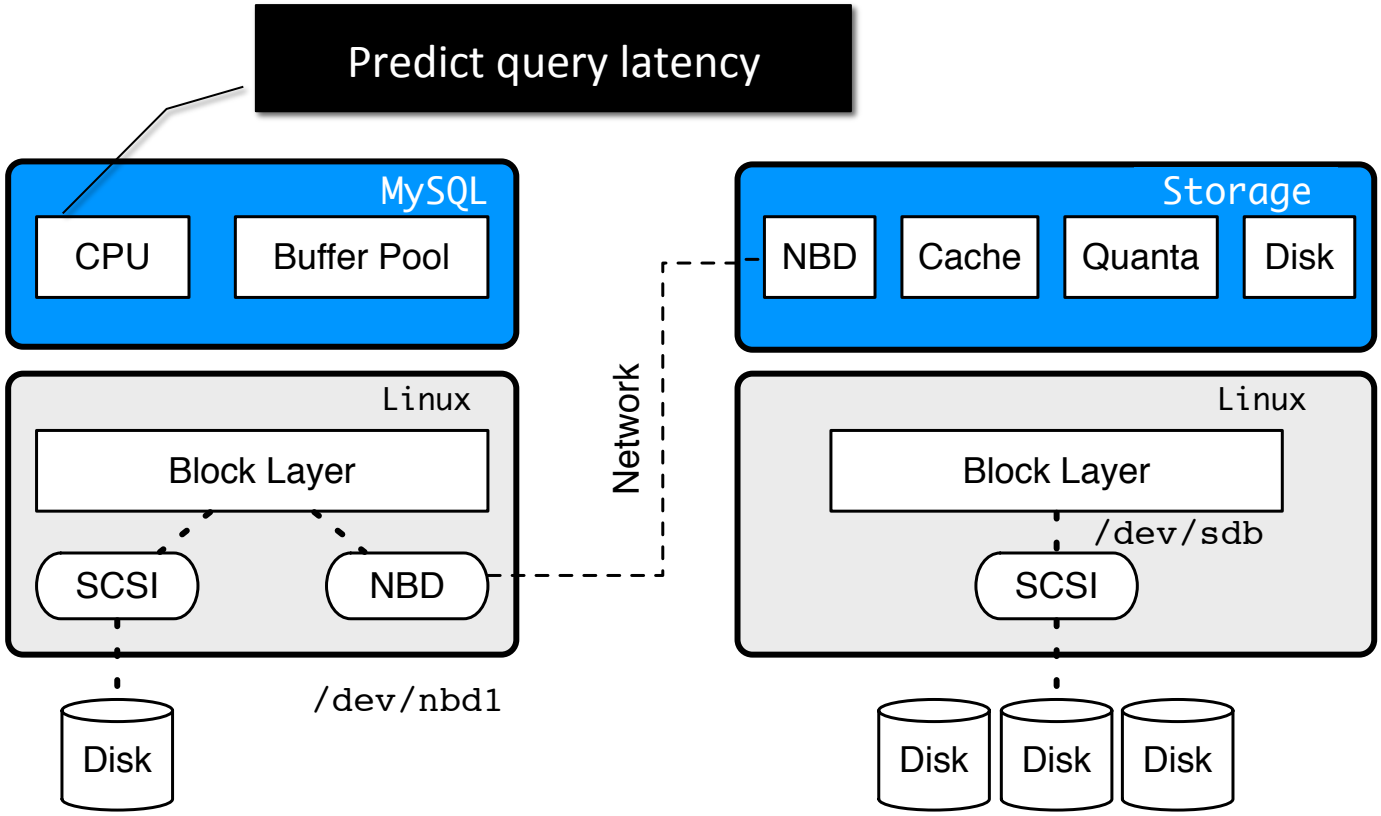
A-STOR Model in FAST'09 paper

Ensemble improves with more data

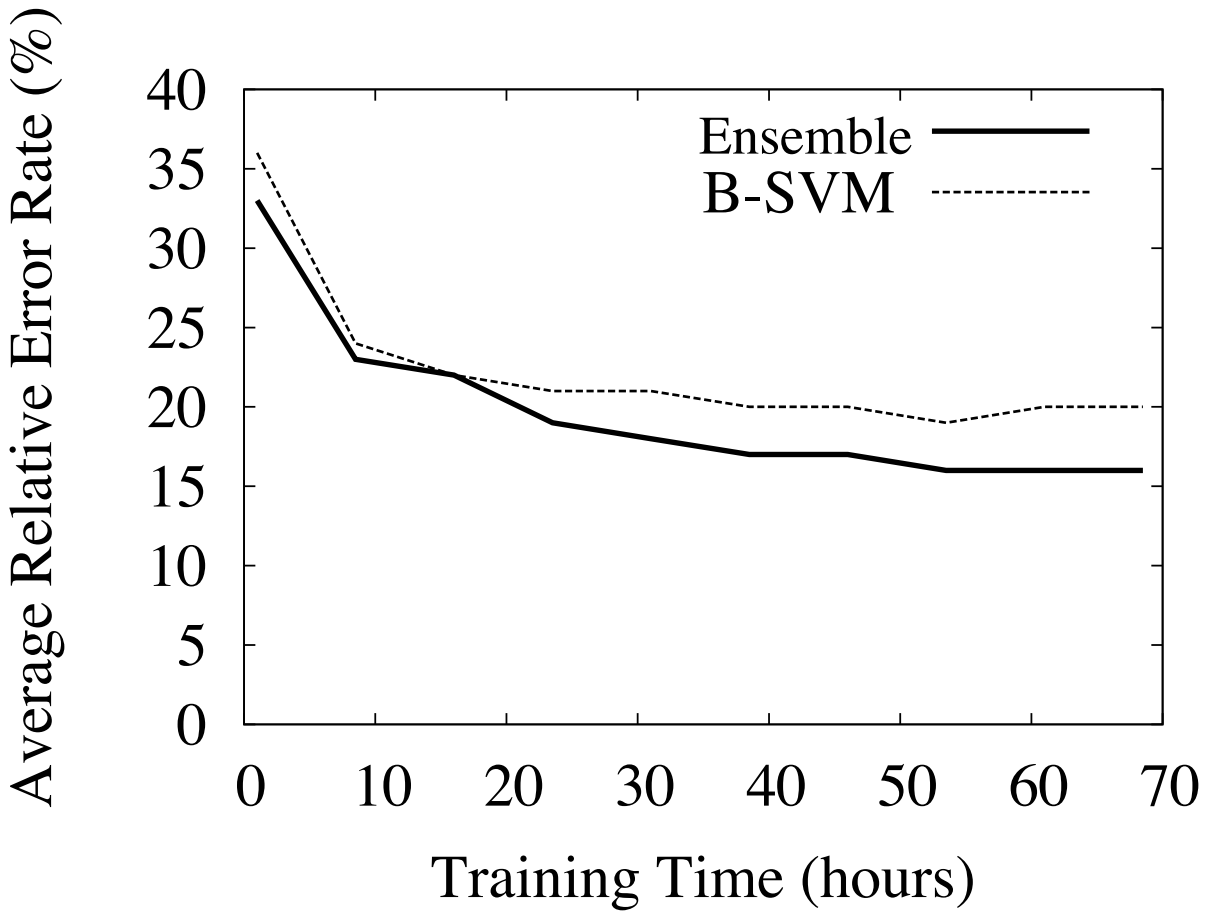
Predicting TPC-W buffer pool latencies



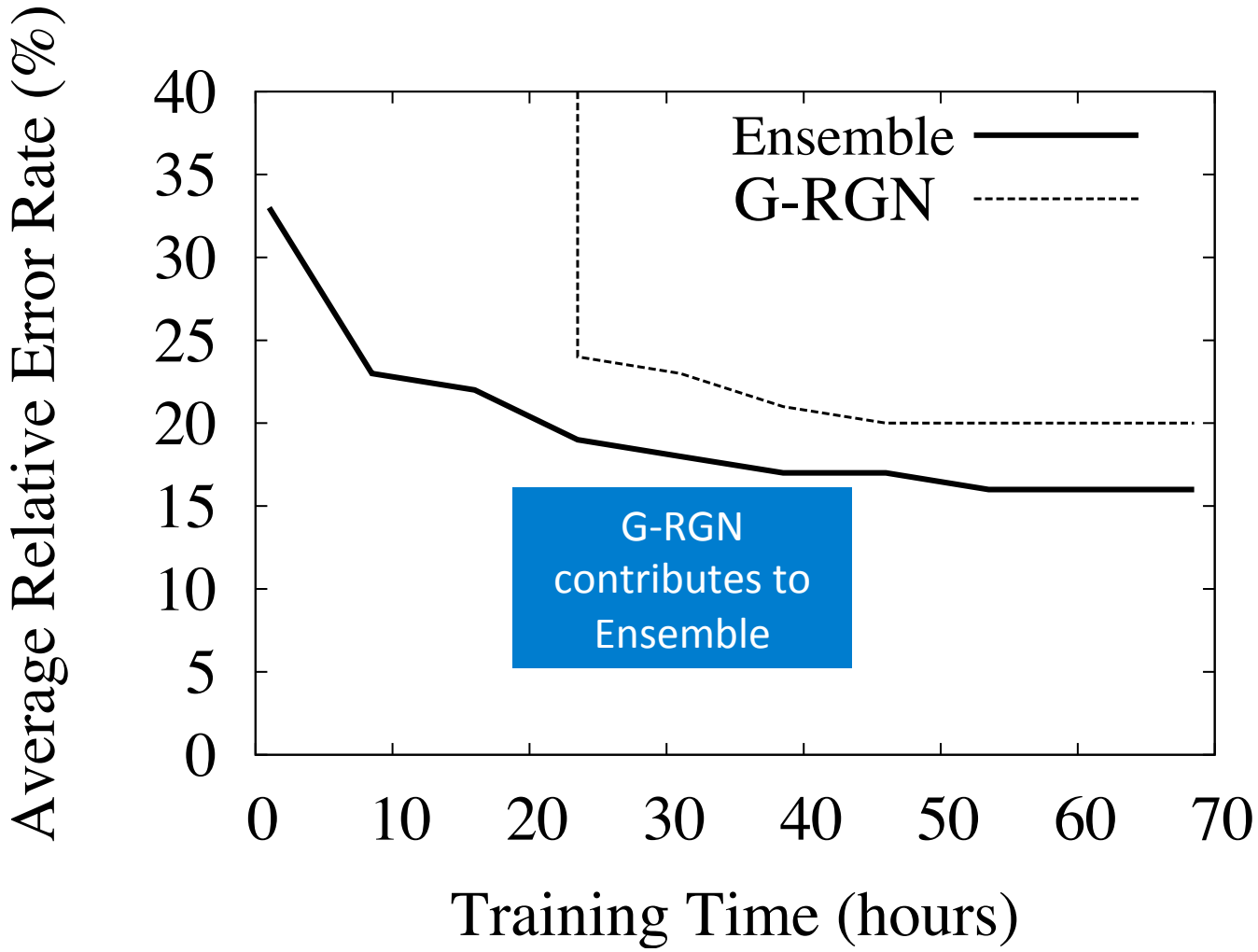
Modeling MySQL end-to-end



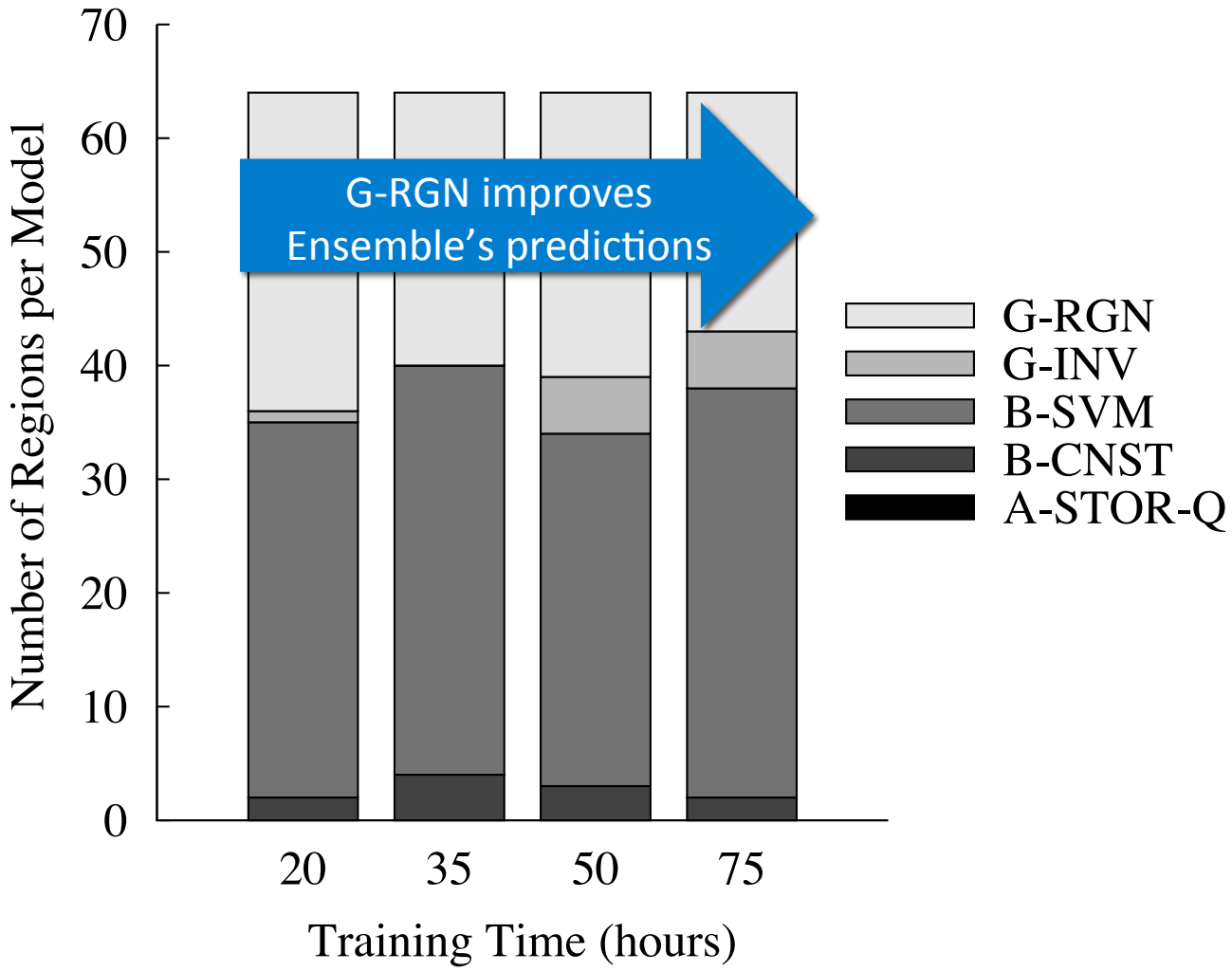
Predicting TPC-C query latencies



Predicting TPC-C query latencies



Predicting TPC-C query latencies



Conclusions

- We built Ensemble
  - An **interactive modeling** framework
  - Uses **semantic knowledge** as model guides
  - Runtime engine iteratively validates, selectively reuses and refines models
- Applied Ensemble to build performance models
  - Allows one to **explore tradeoffs** of analytical, black-box, and gray-box models
  - Showed that we can create models with better accuracy in shorter periods of time

References

- "Ensembles of Models for Automated Diagnosis of System Performance Problems" DSN'05
- "IRONModel: Robust Performance Models in the Wild" SIGMETRICS'08
- *"Dynamic Partitioning of the Cache Hierarchy in Shared Data Centers" VLDB'08*
- *"Dynamic Resource Allocation for Database Servers Running on Virtual Storage" FAST'09*
- *"A Query Language and Runtime Tool for Evaluating Behavior of Multi-tier Servers" SIGMETRICS'10*
- "Performance and Resource Modeling in Highly-Concurrent OLTP Workloads" SIGMOD'13

Thank you!

Mixing CPU and Disk models