

Improving IT Sustainability through Automated System Management: A Data-Based Approach

Calton Pu

with more than a dozen PhD, MS, and undergrad students (some graduated) at CERCS, Georgia Institute of Technology

C. Bash, X. Zhu, Z. Wang (*HP Labs*),
plus collaborators at *ATT Labs and Fujitsu Labs*



1

Large Application Trends

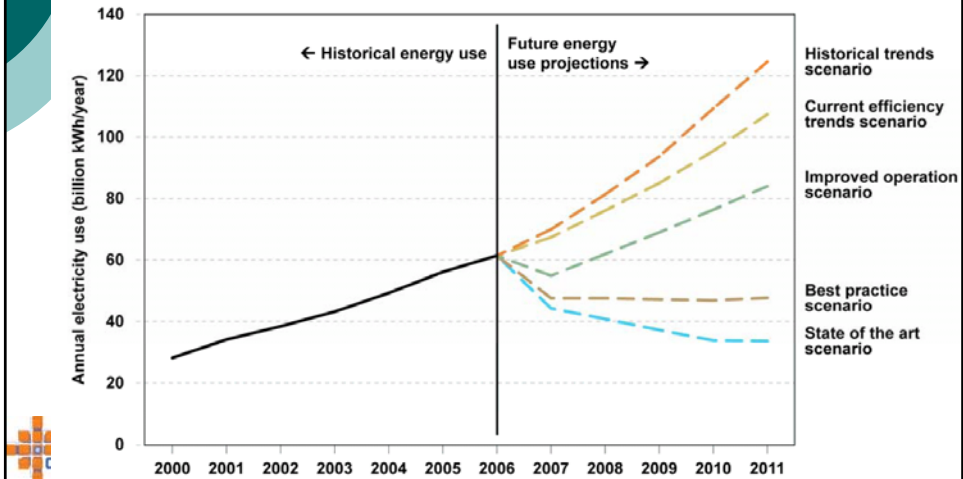
- Enterprise applications more complex
 - Enterprises grow more complex, larger
 - Data centers grow more complex, larger
 - Cloud computing, server consolidation
- EPA 2007 report on data centers
 - 61B kWh in 2006 (1.5% of electricity)



2

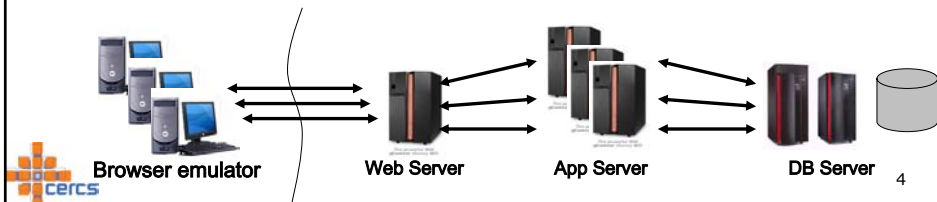
EPA Report Projections

Figure ES-1. Comparison of Projected Electricity Use, All Scenarios, 2007 to 2011

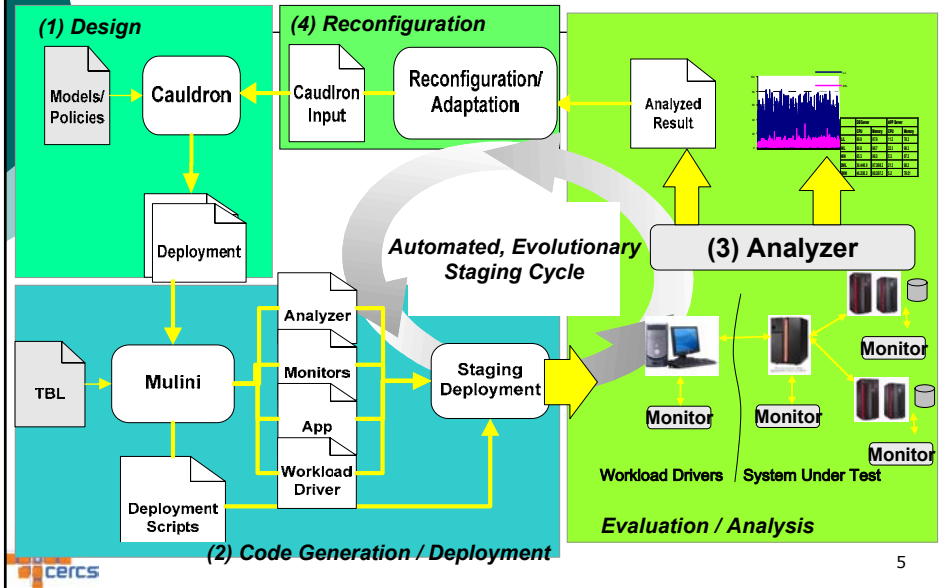


N-Tier Application Benchmarks

- Stressing database server
 - TPC-W: e-Commerce (bookstore)
 - RUBBoS: electronic bulletin board
- Stressing application server
 - RUBiS: e-Commerce (auction)
 - SpecjAppServer2004: manufacturing (auto)



Elba: Overall Picture



Complexity of Experiments

	Experiment Scale (W-A-D)				
	1-2-1	1-6-1	1-6-2	1-8-1	1-8-2
Script lines (per exper. set)	758	1298	1426	1564	1692
Configuration lines (per experiment set)	1168	1212	1212	1234	1234
Number of experiment sets	408	76	72	496	90
Total script lines in all sets	309K	98K	102K	775K	152K
Total configuration lines	476K	92K	87K	612K	111K
Total nodes used	3264	912	936	6944	1350
Total data points (million)	323M	100M	104M	785M	154M

6

RUBiS CPU Utilization (1-8-1)

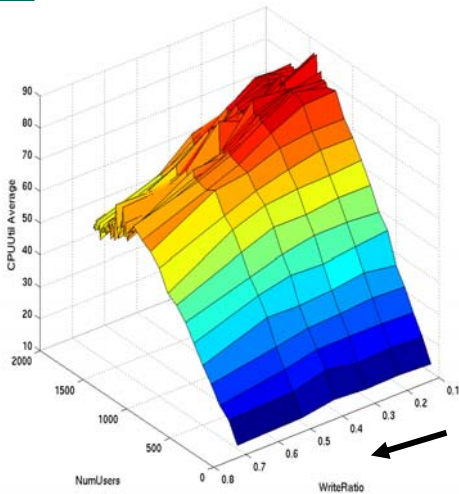


Figure (a) App Server CPU Utilization for 1-8-1

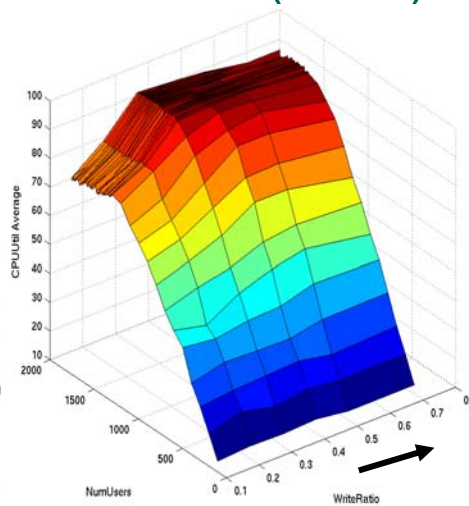


Figure (b) Database Server CPU Util. for 1-8-1

Comparison App & DB Servers

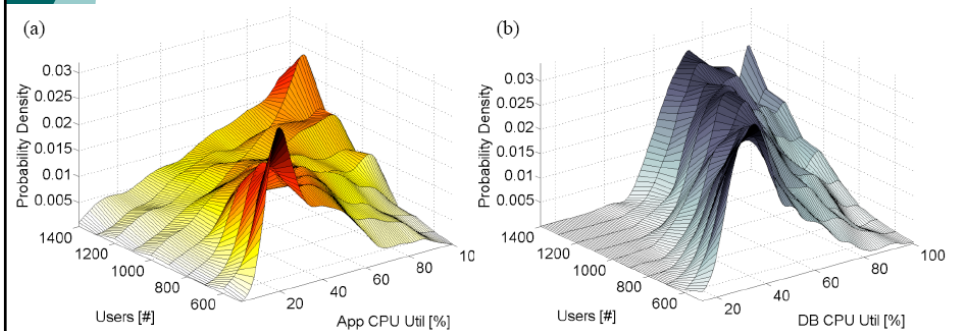
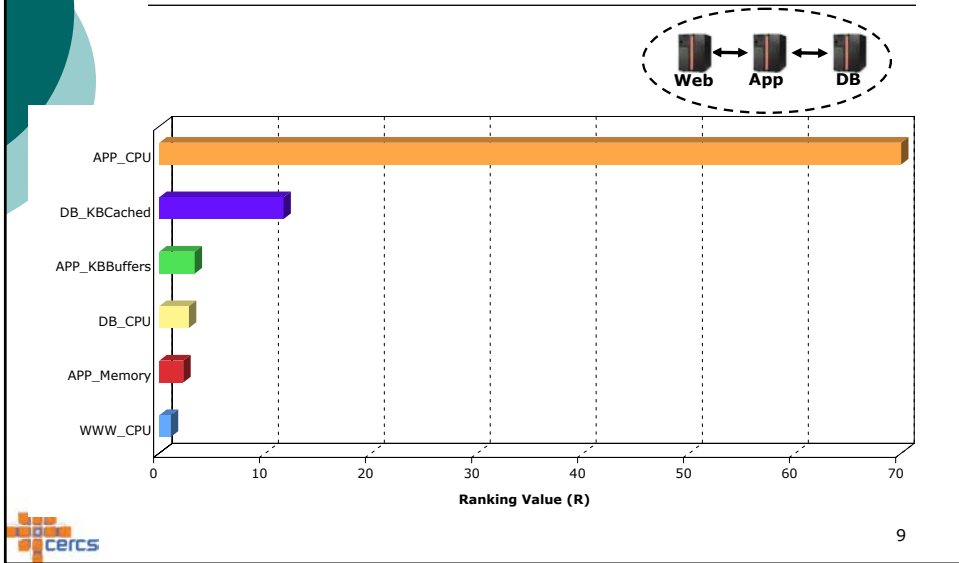
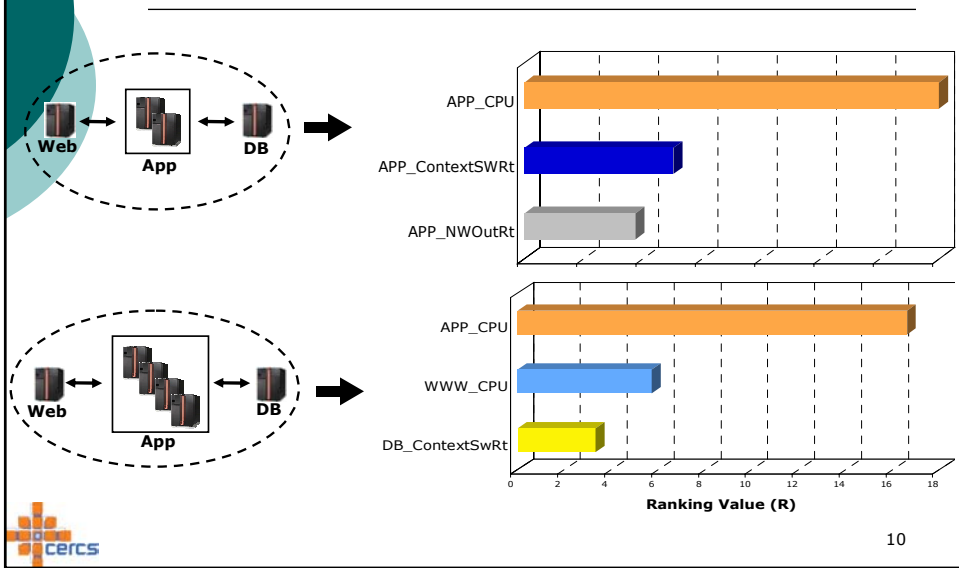


Figure 4: RUBiS on 1H/6H/1L configuration. Estimated probability density functions for CPU utilization in (a) application server for 70 % write ratio and (b) database server for 10 % write ratio.

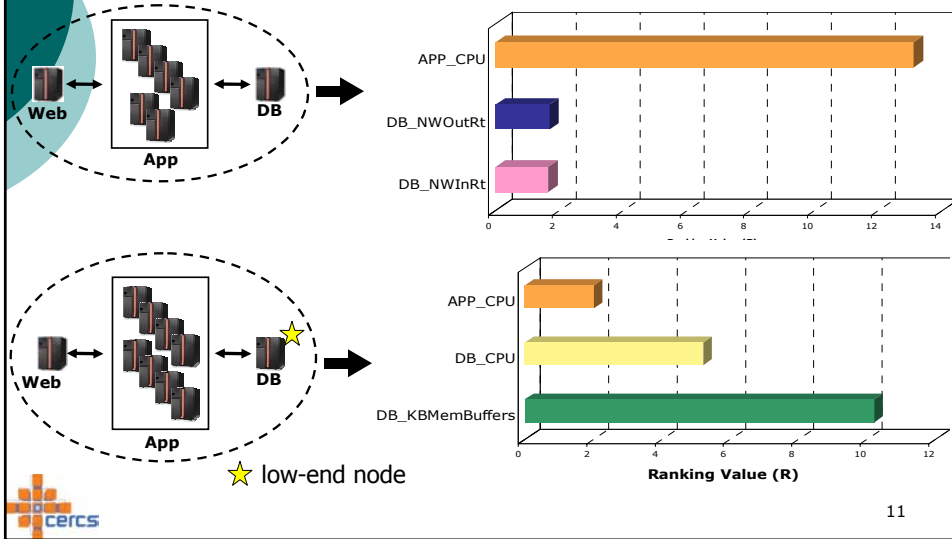
Results of Detection Process



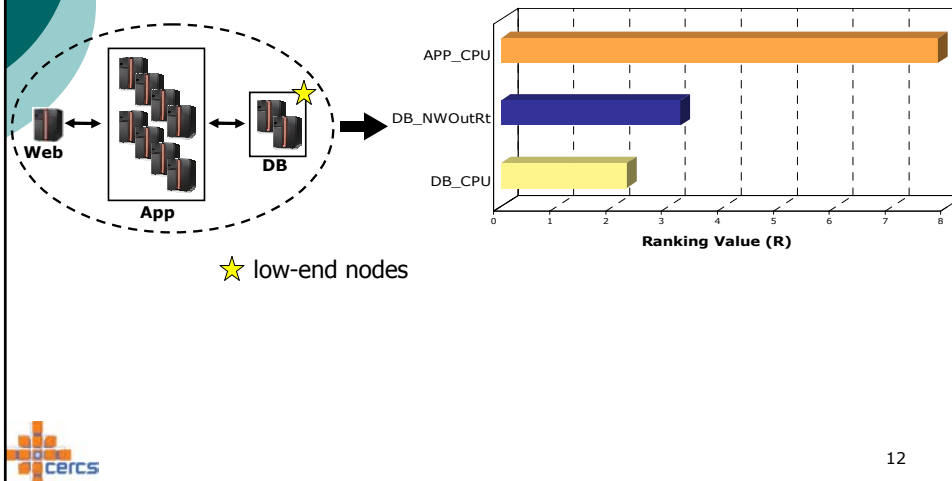
Scale-Out Analysis (a)



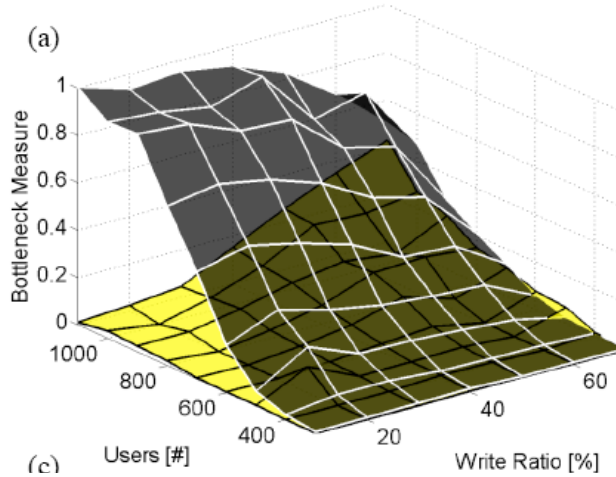
Scale-Out Analysis (b)



Scale-Out Analysis (c)

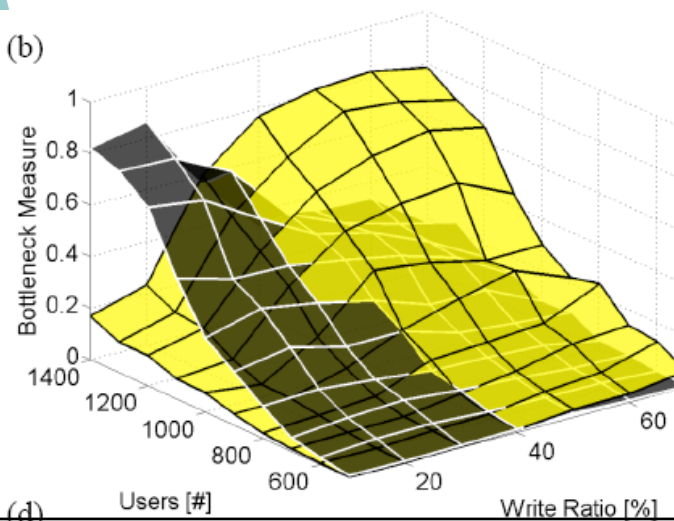


Bottleneck Indicator (1-4H-1L)



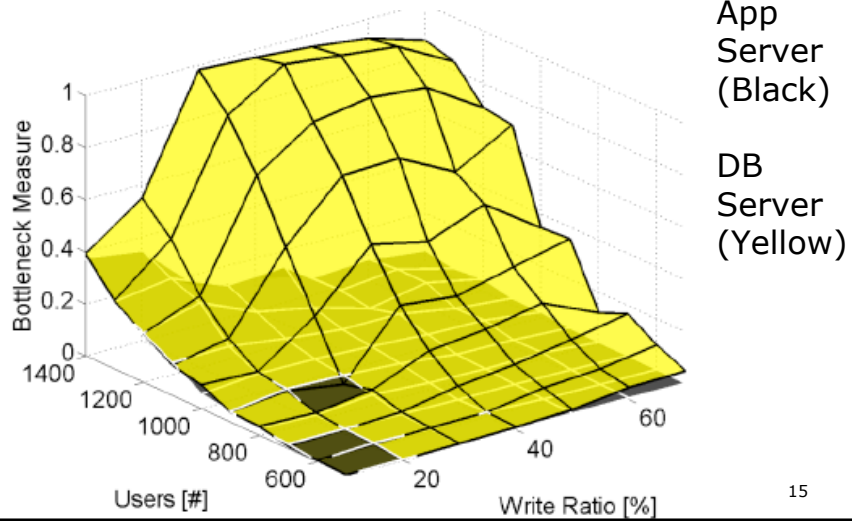
App Server (Black)
DB Server (Yellow)

Bottleneck Indicator (1-6H-1L)

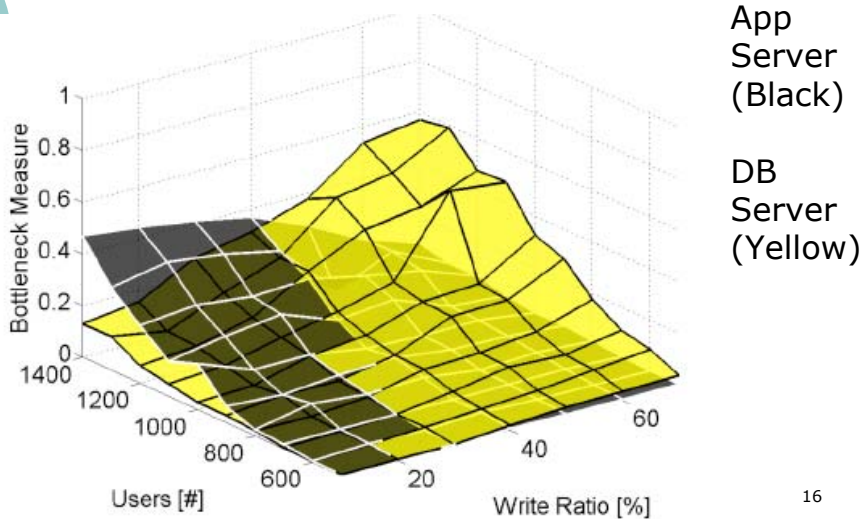


App Server (Black)
DB Server (Yellow)

Bottleneck Indicator (1-8H-1L)



Bottleneck Indicator (1-8H-2L)



Non-Stationary Workload

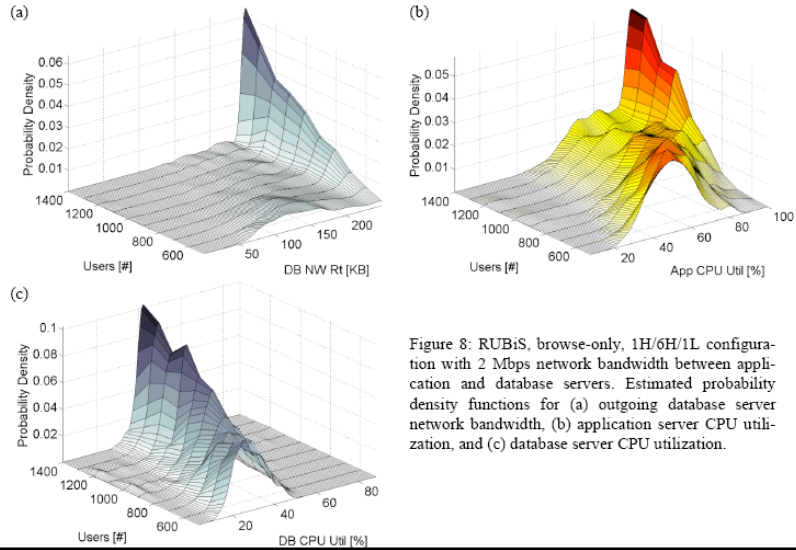


Figure 8: RUBiS, browse-only, 1H/6H/1L configuration with 2 Mbps network bandwidth between application and database servers. Estimated probability density functions for (a) outgoing database server network bandwidth, (b) application server CPU utilization, and (c) database server CPU utilization.

Multiple Bottleneck Detection

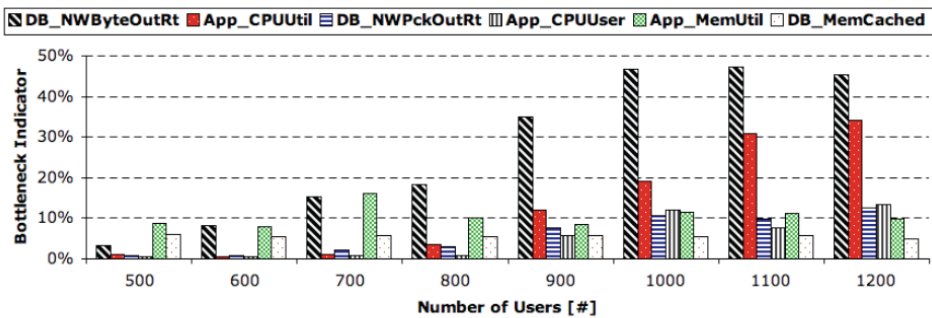


Figure 9: Top bottleneck indicators for RUBiS, browse-only, 1H/6H/1L configuration with limited network bandwidth to database.

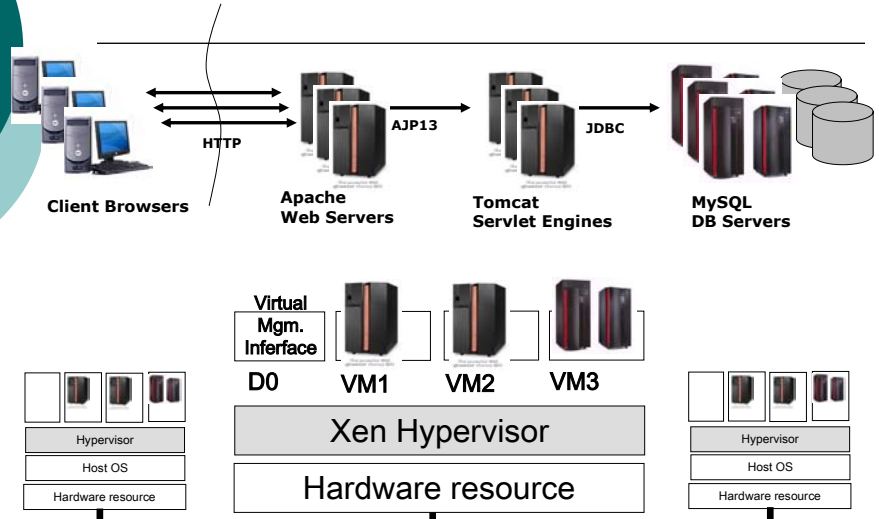
Data-Based System Performance Characterization

- Measured benchmark data
 - Experiments on a range of configurations and workloads
 - N-tier application “model”
- Data-based “prediction” of system behavior
 - Input: configuration and workload
 - Output: observed quality of service (response time, throughput, etc)
 - Output: resource utilization

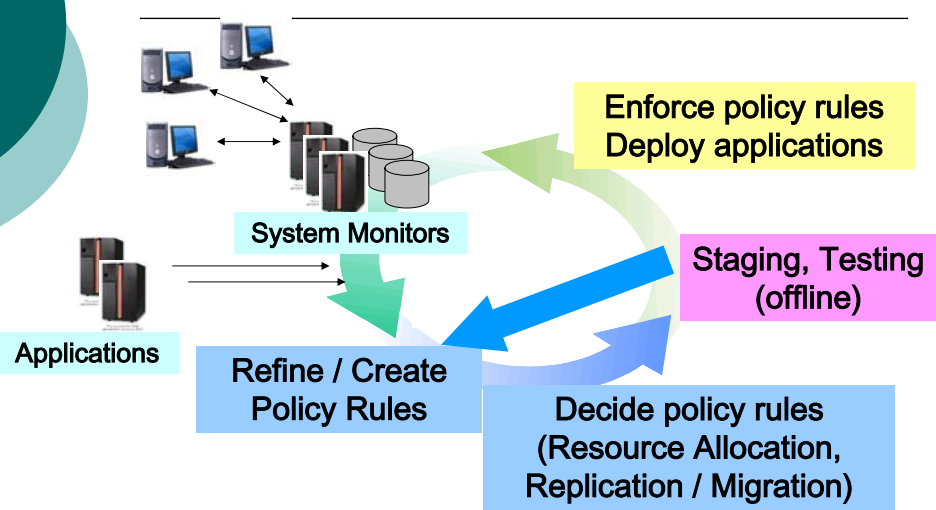
Data-Based Energy Management

- Current Elba experiments
 - Workload \Rightarrow resource utilization
- Measure energy consumption
 - Resource utilization \Rightarrow energy consumption (electricity and heat)
- Data-based prediction of energy consumption
 - Workload \Rightarrow Resource utilization \Rightarrow energy consumption

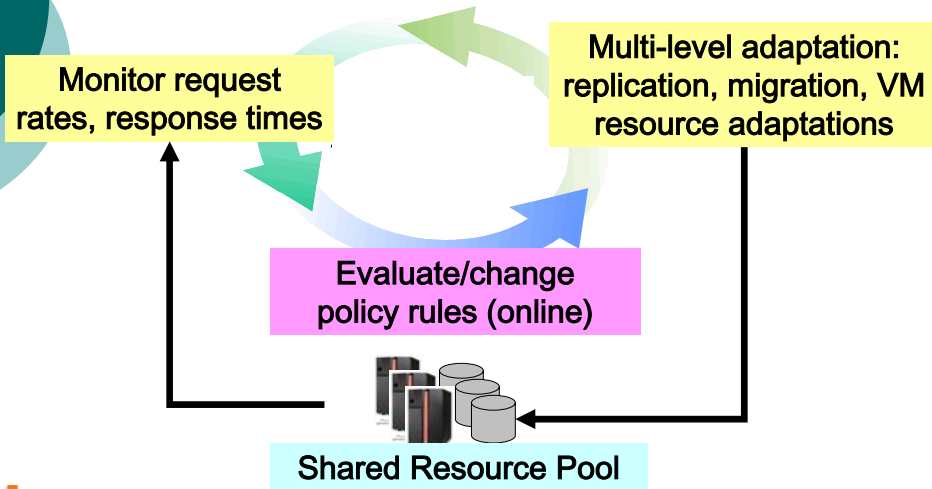
Consolidated (Virtualized) Servers



Adaptation Policies for System Deployment

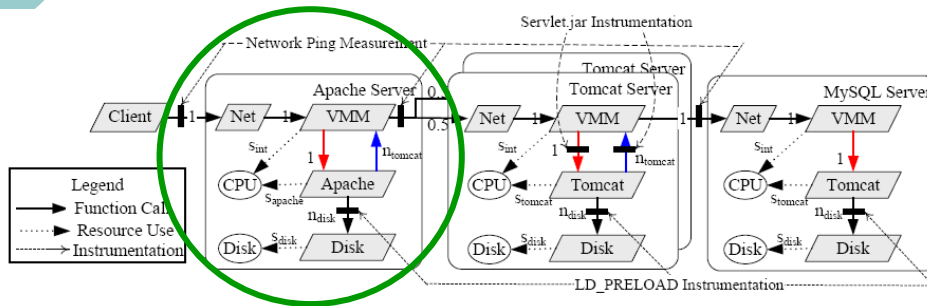


Adaptive Runtime Resource Management

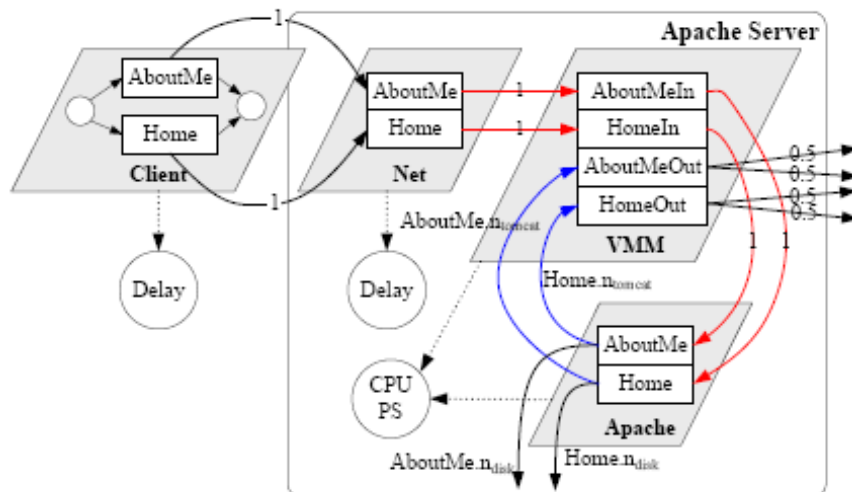


Hybrid Adaptation Policies

- o Layered Queuing Network Model of RUBiS (3-Tier)



Web Server Sub-Model



Quantitative Energy Management

- Data-based automated system management
 - Successful automation in experiment deployment, data collection, and performance analysis
- Measured energy management
 - Workload \Rightarrow resource utilization \Rightarrow energy consumption
 - Automated adaptation policies for performance and energy