STLCMG

Best Paper Award CMG December 2009

Leveraging the Cloud for Green IT:

Predicting the Energy, Cost & Performance of Cloud Computing

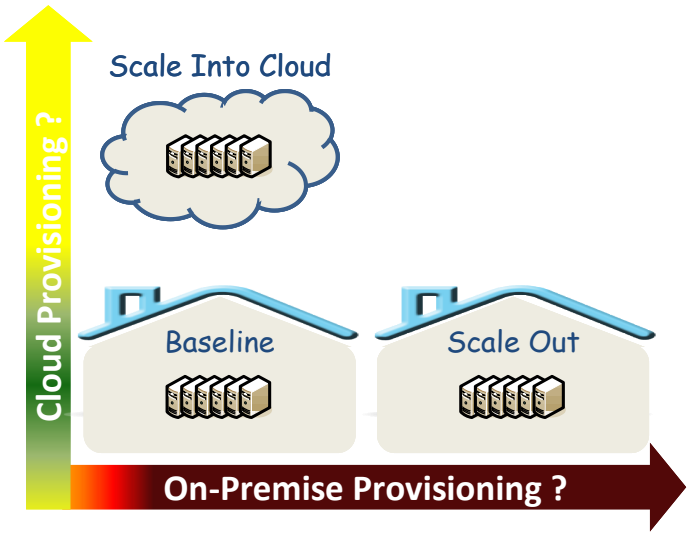
Amy Spellmann, Optimal Innovations
Richard Gimarc, Hyperformix, Inc.
Mark Preston, RS Performance

May 18, 2010

Leveraging the Cloud for Green IT:
What are we going to talk about?

Quantitative Methodology vs. Qualitative 'Chatter'

We will describe & demonstrate a methodology for predicting energy, cost and performance while expanding on-premise IT into the Cloud.



Cut through the hype:

- View Cloud as a new platform in your infrastructure
- Quantify the effect of utilizing Cloud
  - ✓ Energy
  - ✓ Cost
  - ✓ Performance
- Utilize a repeatable, scientific methodology

Cloud Computing Promises

- Flexible, on-demand IT services
- Low entry cost (OpEx vs. CapEx)
- Time-to-Market
- Scalable
- Elastic
- Extend capacity of a “full” data center
- Higher Efficiency
- Greener
- Energy saving incentives, rebates, carbon penalties

Why Consider Cloud Computing?

- IT infrastructure at capacity, need to accommodate growth
- Available power to the existing data center at capacity
- New data center must be cost effective
- Need to avoid peak demands and penalties
- Facilities below efficiency targets
- Pay-as-you go cash outlay vs. large investment up front
- Government & utility energy saving incentives: rebates, carbon penalties

Energy Consumption: Are Clouds Greener?

- Energy consumption in the data center is often not well understood [Koomey]
- Economies of scale enable Cloud providers to be more efficient
  - Scalability & efficiency must be balanced to meet profit margins
  - Energy usage can be reduced through free cooling and specialty hardware (e.g. Amazon, IBM, Google, Microsoft)
  - 'Follow-the-Moon' can be a viable strategy

Leveraging On-Demand Cloud with a Hybrid Architecture

- Reduces energy consumption in our data center (greener)
- Decreases overall energy consumption due Cloud's economies of scale

Cloud Considerations (1 of 2)

Attribute	Traditional On-Premise	Cloud Computing
Infrastructure Scalability	▪ Server upgrade (CapEx) ▪ Potential data center expansion (CapEx)	▪ Cloud resources (OpEx) ▪ Limits data center expansion (CapEx)
Deployment Timeline	▪ Weeks to months	▪ On-demand in minutes/hours
Infrastructure Management	▪ Physical and virtualized components	▪ (Virtual) Cloud resources
Business Continuity	▪ Responsibility of on-premise IT staff	▪ Partner with Cloud provider
Physical Resource Utilization	▪ Size to handle peak loads ▪ Low utilization during non-peak time	▪ Scale dynamically with on-demand resources ▪ Eliminates the need to overbuild

Cloud Considerations (2 of 2)

Attribute	Traditional On-Premise	Cloud Computing
Network Infrastructure	▪ Data center IT staff responsible for networks and network expansion (CapEx)	▪ Utilize Cloud network infrastructure (OpEx) ▪ Additional on-premise network may be required (CapEx)
Performance	▪ Response time depends on workload volumes & supporting infrastructure	▪ Dependent on responsiveness of the virtual resources ▪ Potential new network delay between on-premise and Cloud
Energy	▪ Driven by IT infrastructure & supporting facility ▪ Fixed limit on power draw ▪ Demand charges (kW) sized for Peak	▪ Energy consumption managed by the provider ▪ Limits on-premise energy growth ▪ Limits on-premise demand charges
IT Budget Categories	▪ CapEx for infrastructure ▪ OpEx for facility & energy	▪ Primarily OpEx for infrastructure

Cloud Consideration – Set of Attributes

Addressed in Case Study

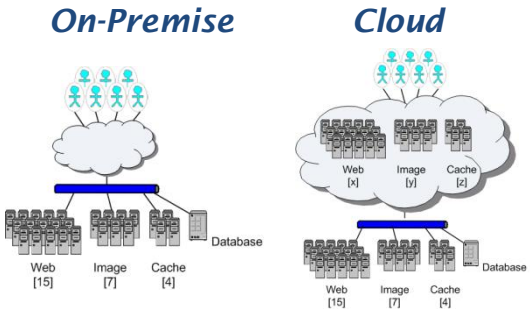
- ✓ Infrastructure Scalability
- ✓ Physical Resource Utilization
- ✓ Network Infrastructure
- ✓ Performance
- ✓ Energy
- ✓ IT Budget Categories

Not Addressed in Case Study

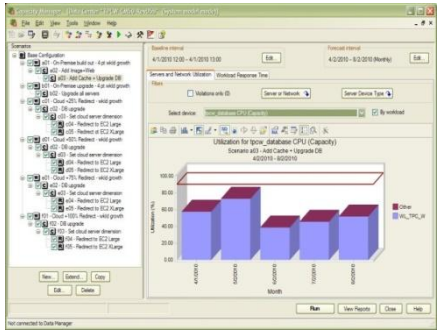
- Deployment Timeline
- Infrastructure Management
- Business Continuity
- Security
- Software Licensing
- Operations Staffing
- Facilities Cost
- Demand vs. Peak utility charges
- Maintenance

Decision Support for Leveraging the Cloud
Predictive Performance & Capacity Management

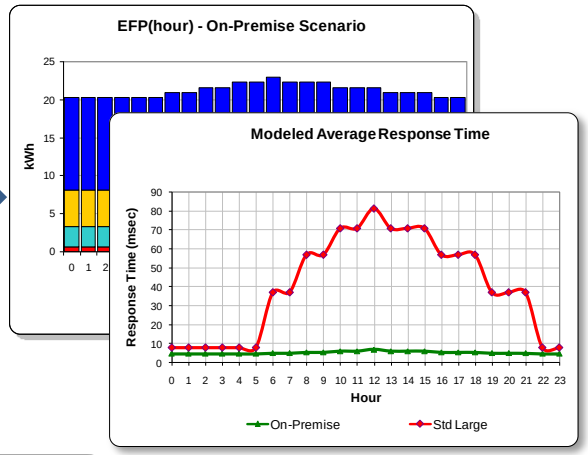
Business Service Analysis & Planning



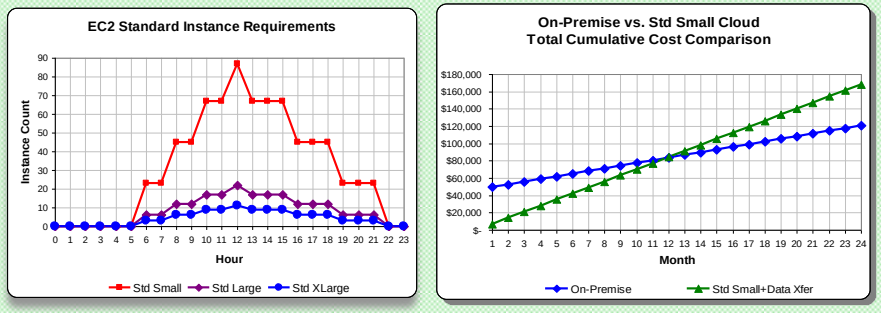
Predictive Capacity Planning



Energy Footprint & Performance Projection

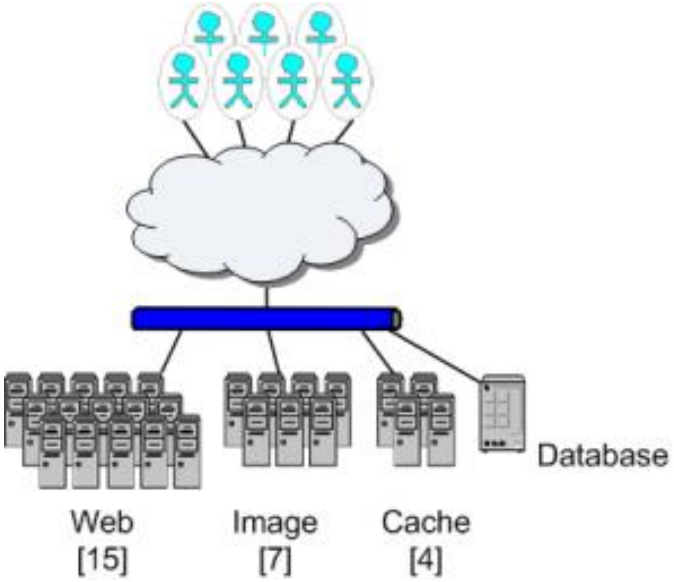


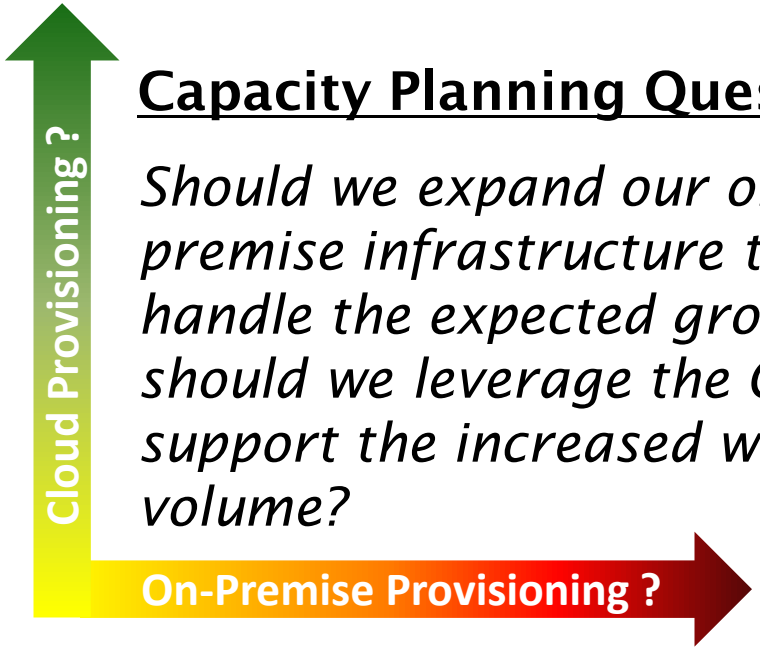
Cost, Energy, Performance Comparison



Case Study
Data Center & Application Growth Challenges

- Web-based eCommerce application
- Current server infrastructure at capacity limit
- Expecting 2x workload growth over next year
- Data center reaching power limits





Capacity Planning Question:

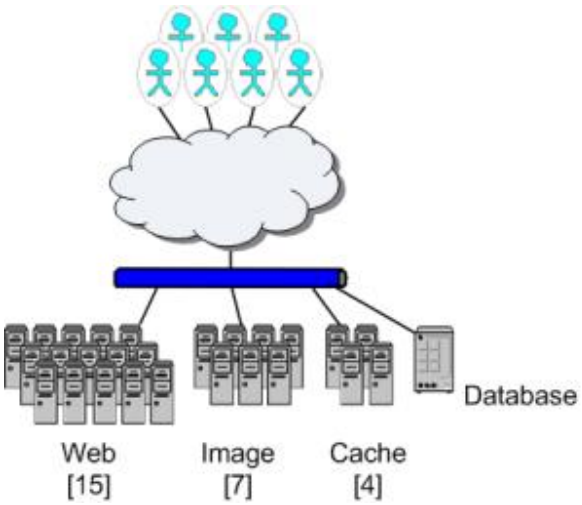
Should we expand our on-premise infrastructure to handle the expected growth, or should we leverage the Cloud to support the increased workload volume?

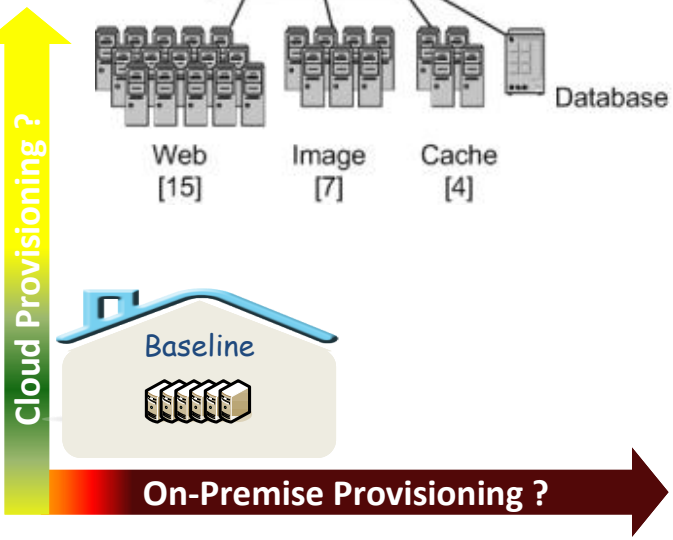
Case Study
Methodology

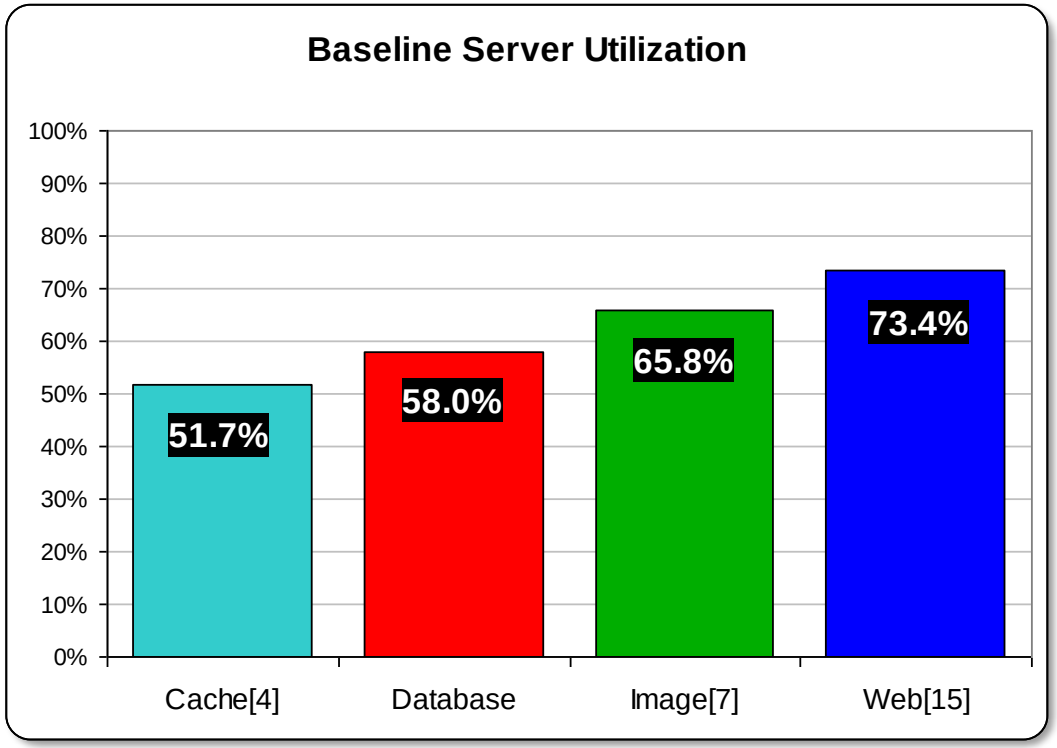
Workload growth	What on-premise infrastructure do we need to support workload growth? *Model the workload growth & on-premise infrastructure*
Cloud resources	What Cloud resources are required to support increased demand? *Model the on-demand growth & the Cloud resources*
Energy, cost & performance	How do energy, cost & performance compare; on-premise versus Cloud? *Use a model to predict & compare*

Case Study
Today's Current Baseline (On-Premise Only)

- 27 servers
- Maintain 70% utilization threshold





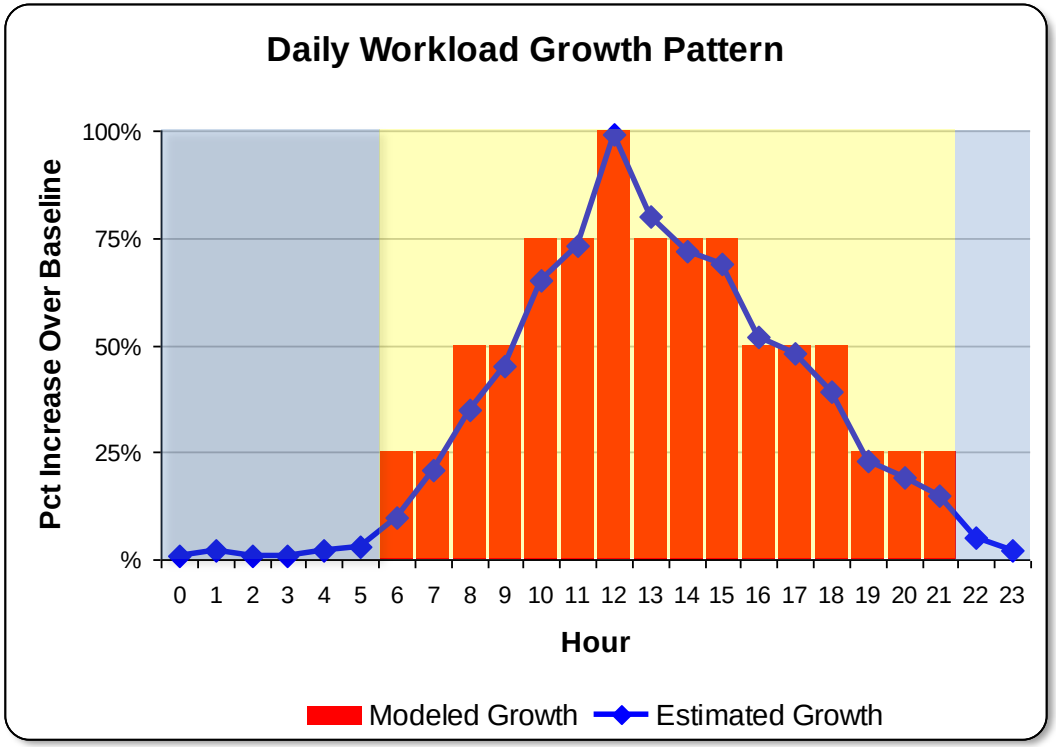


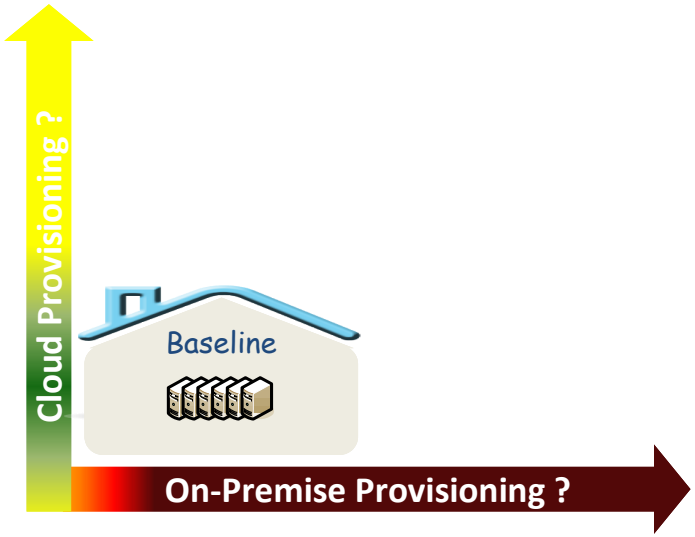
Quantity	Tier	Make/Model	Relative Speed
4	Cache	Dell PowerEdge 1850 2.8GHz	20
1	Database	Dell PowerEdge 2950 1.8GHz	81
7	Image	Dell PowerEdge 1850 2.8GHz	20
15	Web	Dell PowerEdge 1850 2.8GHz	20

Case Study
Workload Growth Pattern

➢ **Workload Growth**
- Cloud Resources
- Energy, Cost, Performance

- Historical data shows workload volume by hour
- Business expects 100% growth at peak hour of the day





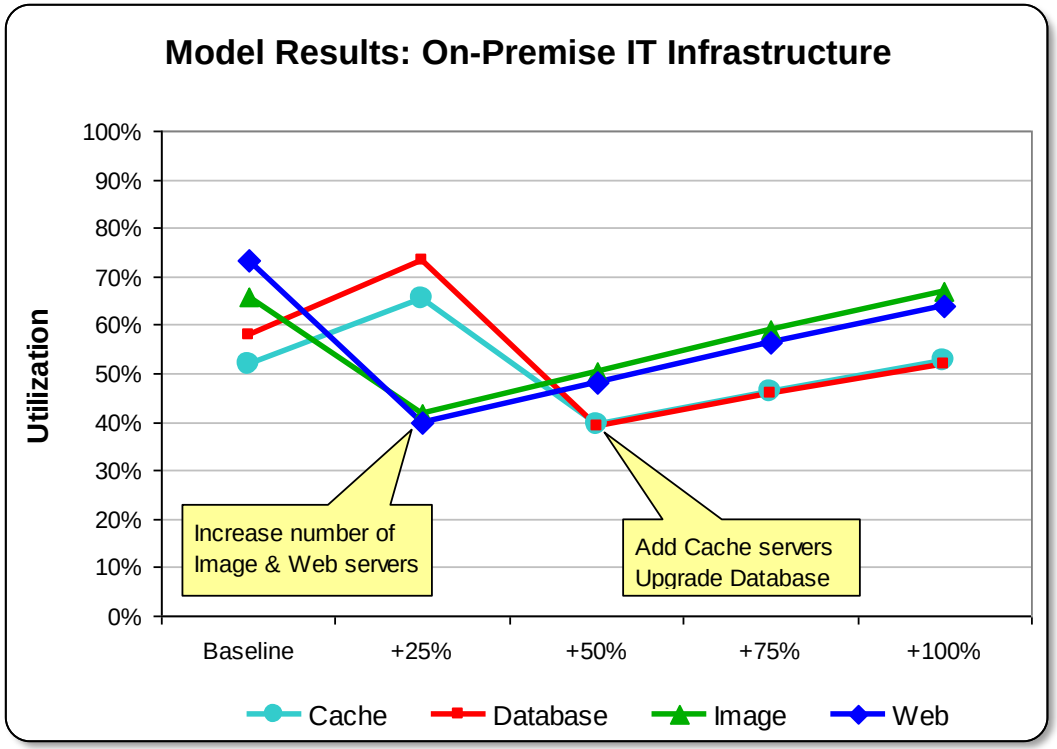
- Increased workload from hours 6-21
- On-premise infrastructure must be scaled to support peak load

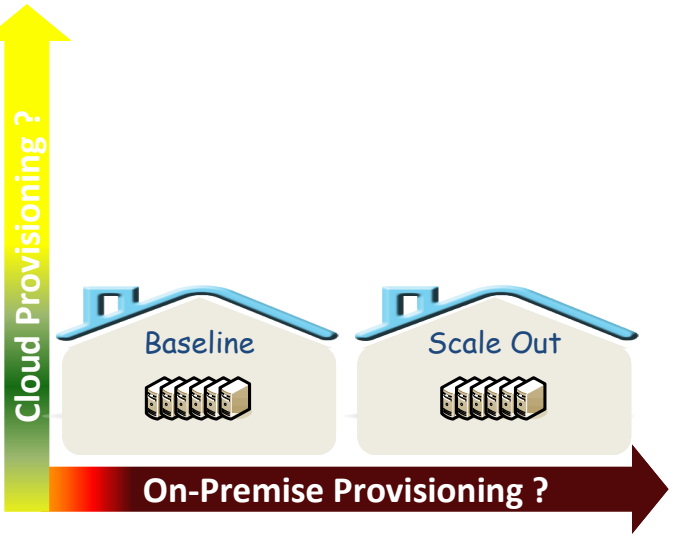
Do we add on-premise capacity or route to the Cloud?

Case Study
Size On-Premise to Handle Growth

- **Workload Growth**
  - Cloud Resources
  - Energy, Cost, Performance

- Increase workload by 2x
- 25% growth steps
- Upgrade DB server
- Add +31 other servers
- 58 total on-premise servers
- Doubling processing power on-premise



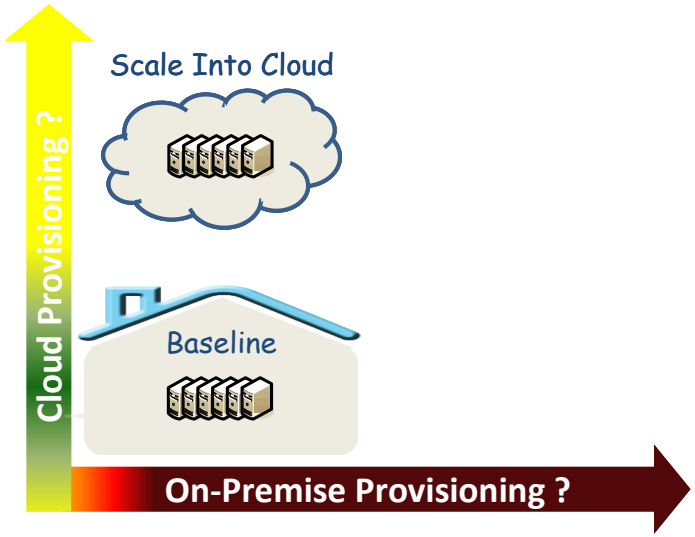


Quantity	Change	Tier	Make/Model	Relative Speed
8	+4	Cache	Dell PowerEdge 1850 2.8GHz	20
1	-	Database	Dell PowerEdge R900 2.4GHz	182
14	+7	Image	Dell PowerEdge 1850 2.8GHz	20
35	+20	Web	Dell PowerEdge 1850 2.8GHz	20

Case Study
Evaluate Cloud Services

✓Workload Growth
➢ **Cloud Resources**
· Energy, Cost, Performance

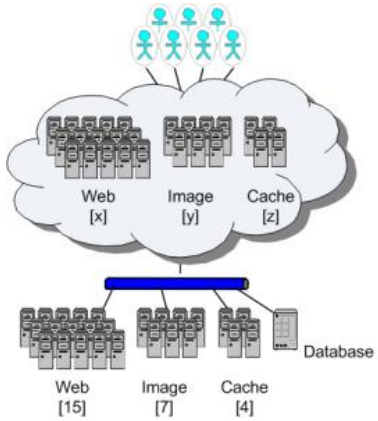
Standard Instance	EC2 Compute Units	Number of Virtual Cores	Description
Small	1	1	1.7 GB of memory, 1 Amazon EC2 Compute Unit (1 virtual core with 1 EC2 Compute Unit), 160 GB of instance storage, 32-bit platform
Large	4	2	7.5 GB of memory, 4 Amazon EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each), 850 GB of instance storage, 64-bit platform
Extra Large	8	4	15 GB of memory, 8 Amazon EC2 Compute Units (4 virtual cores with 2 EC2 Compute Units each), 1690 GB of instance storage, 64-bit platform



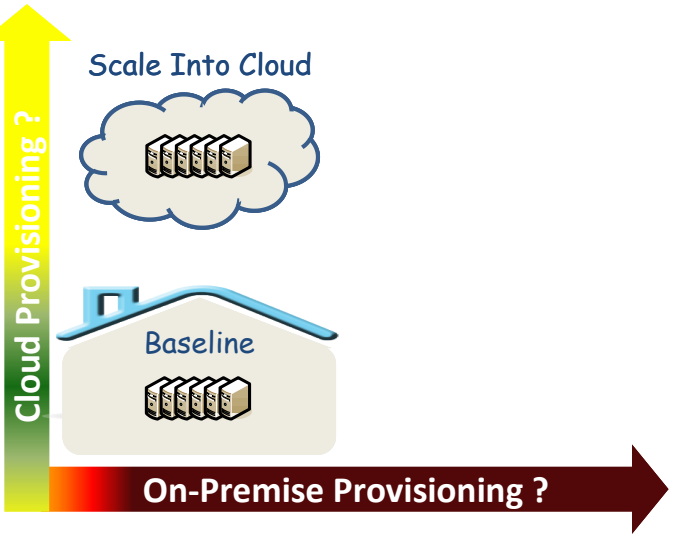
- Amazon Web Services (AWS)
- Elastic Compute Cloud (EC2) - PaaS
  - ✓ Standard Instances
  - High-CPU Instances
- EC2 Compute Unit has the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor

Case Study
Prepare to Model Cloud Resources

✓Workload Growth
➢ **Cloud Resources**
· Energy, Cost, Performance



Standard Instance	EC2 Compute Units	Number of Virtual Cores	Description
Small	1	1	1.7 GB of memory, 1 Amazon EC2 Compute Unit (1 virtual core with 1 EC2 Compute Unit), 160 GB of instance storage, 32-bit platform
Large	4	2	7.5 GB of memory, 4 Amazon EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each), 850 GB of instance storage, 64-bit platform
Extra Large	8	4	15 GB of memory, 8 Amazon EC2 Compute Units (4 virtual cores with 2 EC2 Compute Units each), 1690 GB of instance storage, 64-bit platform



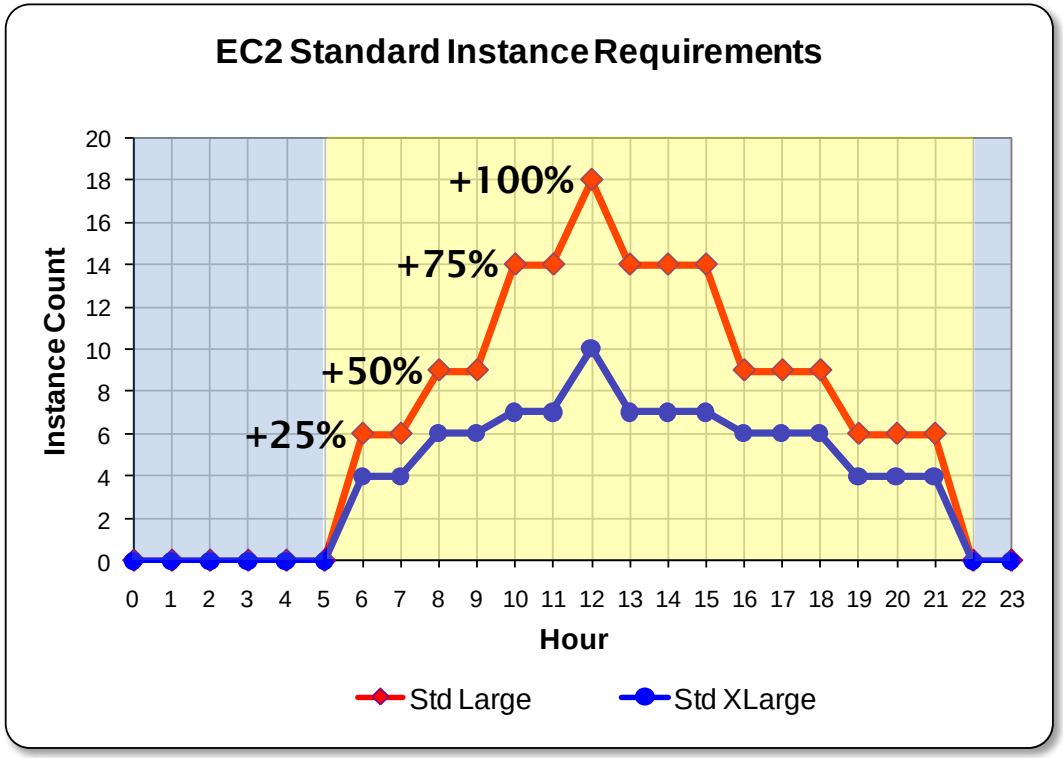
- Develop model representations of EC2 Standard Instances

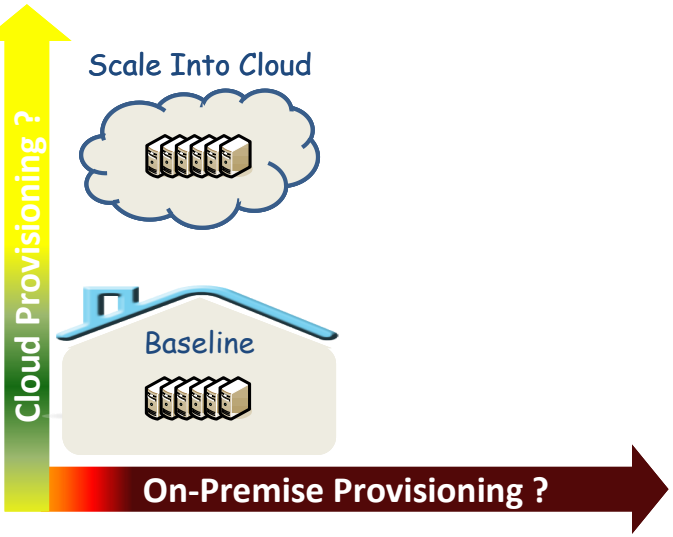
	Standard Instance	EC2 Compute Units	Number of Processors	Relative Speed
	Small	1	1	8
✓	Large	4	2	31
✓	Extra Large	8	4	61

Case Study
Quantify Required Cloud Resources

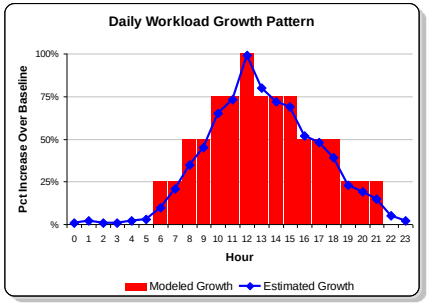
✓Workload Growth
➢ **Cloud Resources**
· Energy, Cost, Performance

- Number of instances required to satisfy on-demand
- Large & XLarge instances
- Maintain 70% utilization threshold
- Leave DB on-premise





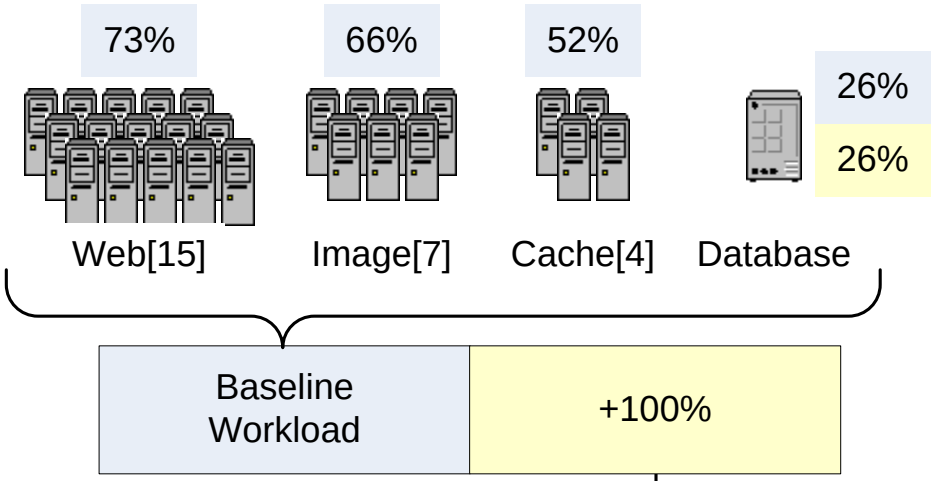
On-Demand Provisioning

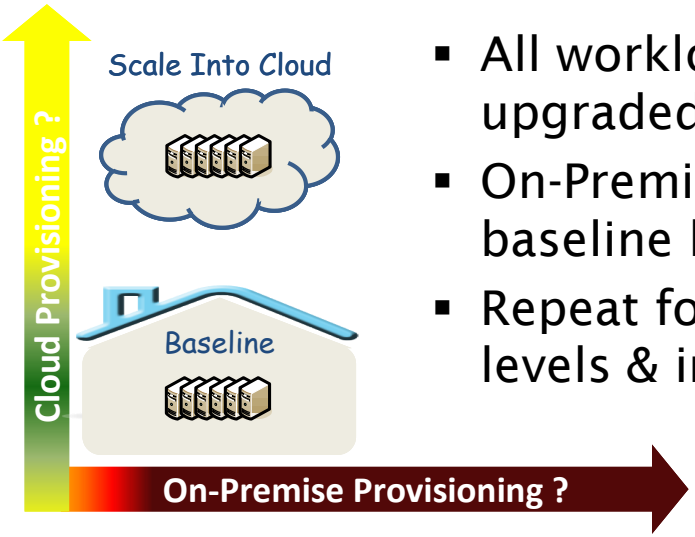


Case Study
'Large' Cloud Resources for +100%

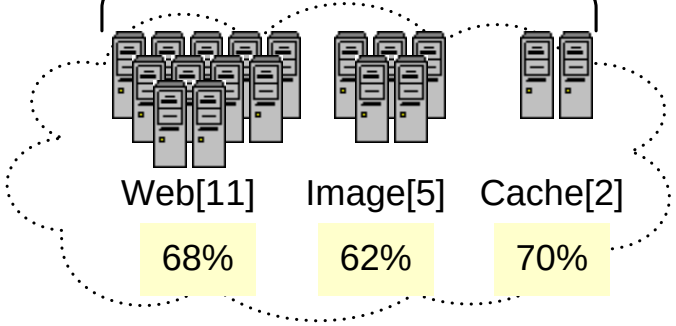
✓Workload Growth
➢ **Cloud Resources**
· Energy, Cost, Performance

- Consider 100% growth
- Route growth to the Cloud
- Estimate number of required Large Cloud instances
- Maintain 70% utilization threshold





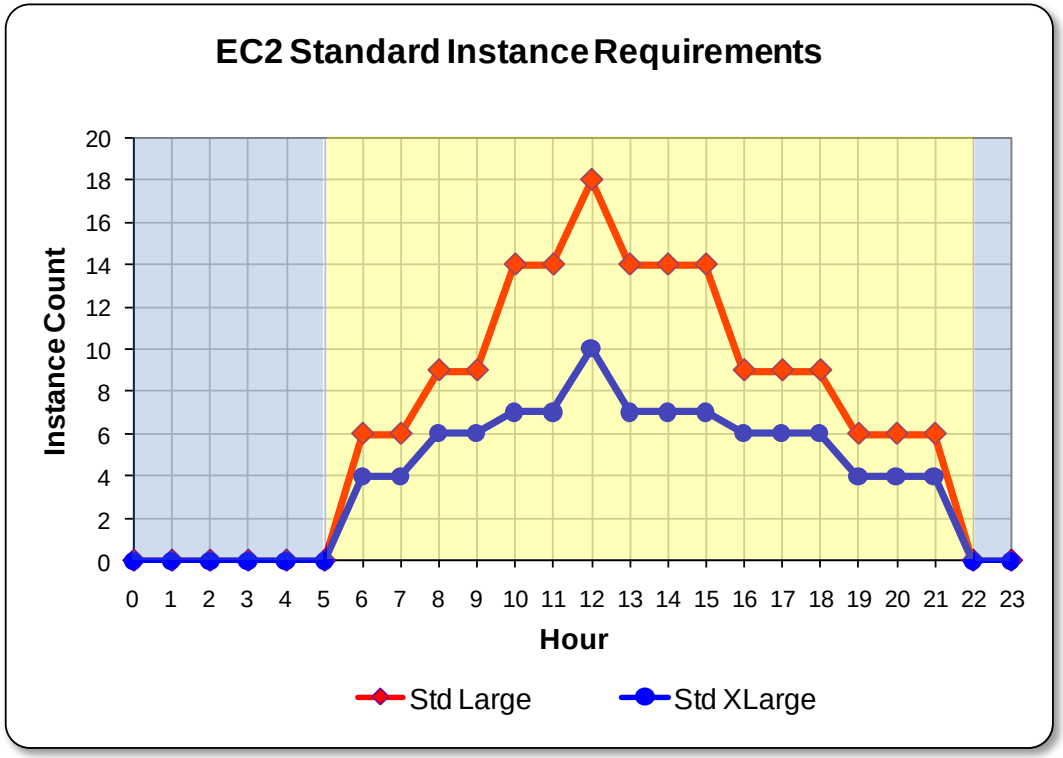
- All workload uses the upgraded database
- On-Premise servers are at baseline levels
- Repeat for all workload levels & instance sizes

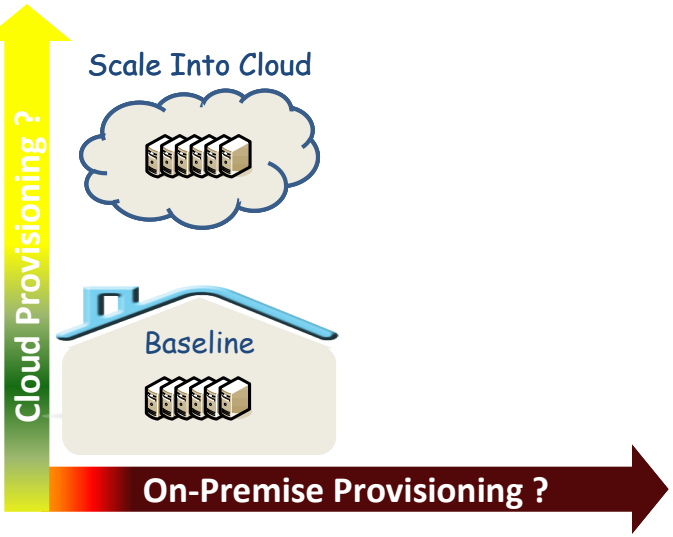


Case Study
Cloud Workload Routing

✓Workload Growth
➢ **Cloud Resources**
· Energy, Cost, Performance

- On-premise servers will handle workload from hours 0-5 and 22-23
- Additional workload from hours 6-21 will be routed to the Cloud

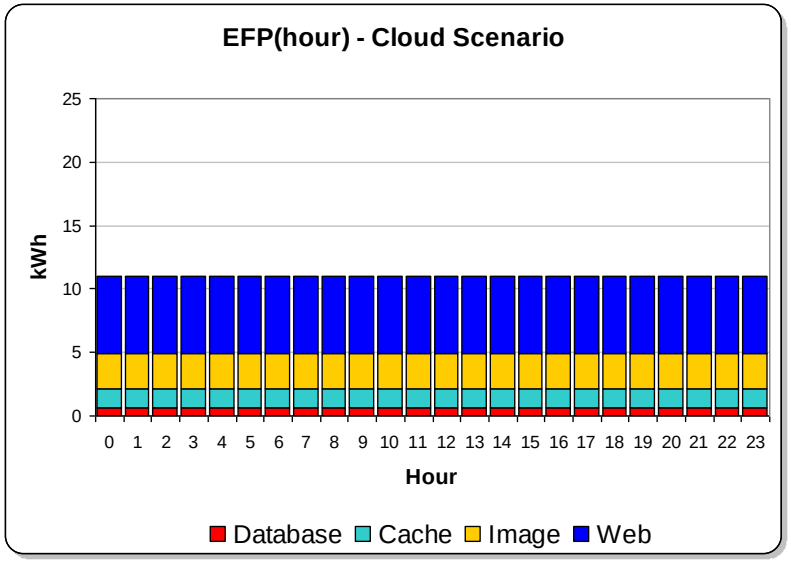




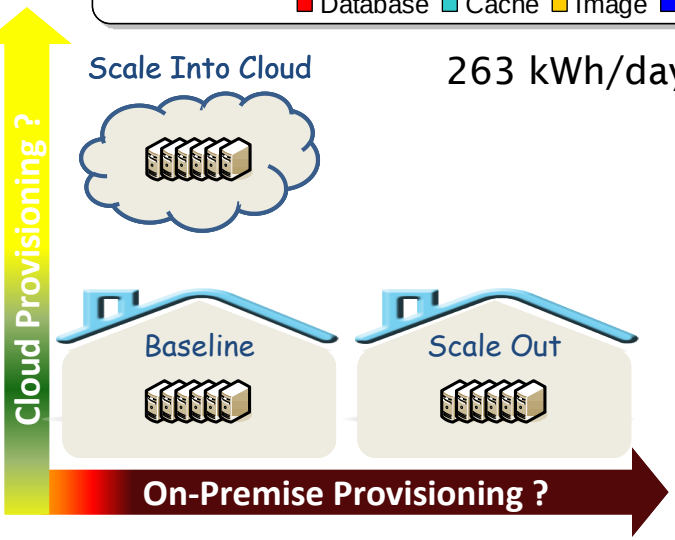
Standard Instance	Total Instance Hours per Day
Large	163
Extra Large	95

Case Study
On-Premise Energy Footprint (EFP)

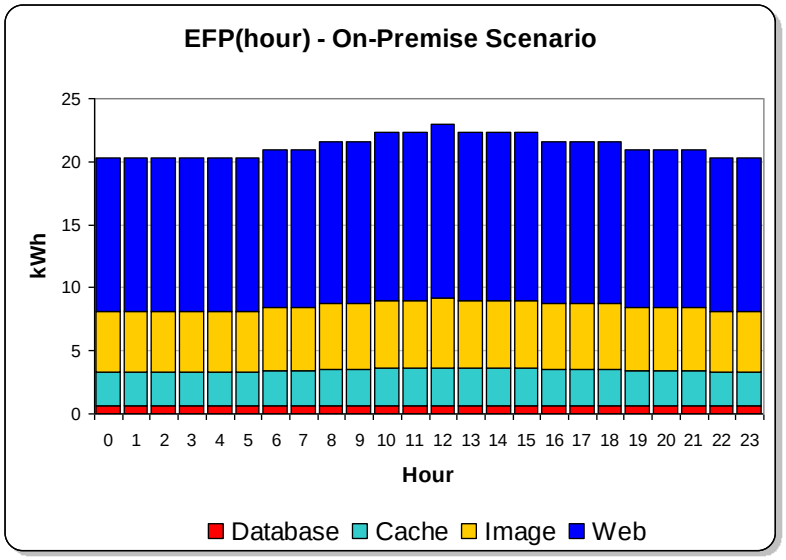
✓Workload Growth
✓ Cloud Resources
➢ **Energy**, Cost, Performance



263 kWh/day



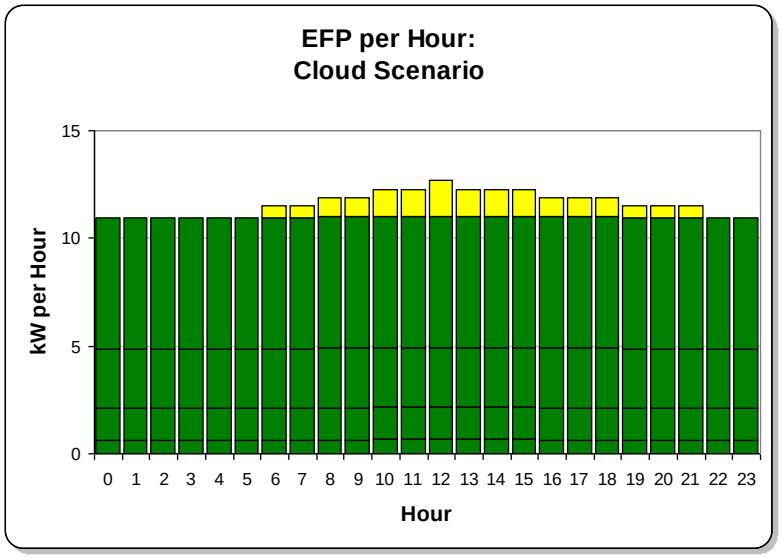
- On-Premise EFP using Cloud stays constant across the day – additional work routed to the cloud
- On-Premise scenario EFP follows workload volume
- +31 additional servers increase On-Premise energy use

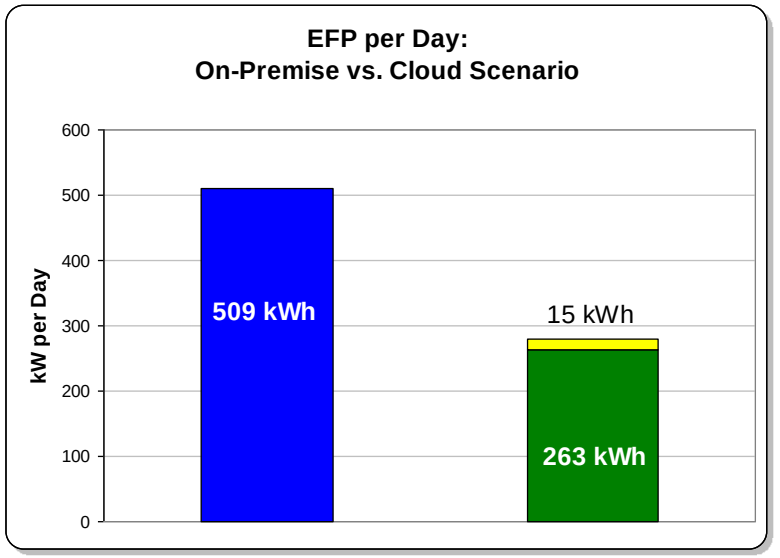


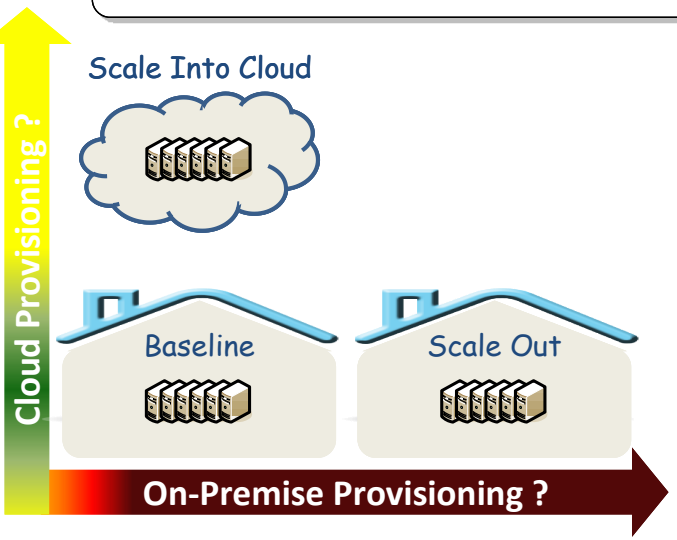
509 kWh/day

Case Study
Estimated EFP in the Cloud

✓Workload Growth
✓ Cloud Resources
➢ **Energy**, Cost, Performance







- ~~On-demand characteristics of Cloud drastically reduces overall EFP~~
- Servicing peak workload with Cloud drastically reduces on-premise energy demand
- Average 7 Large instances per hour
- Assumes 4 Large instances per physical server
- Estimated daily Cloud EFP is 15 kWh

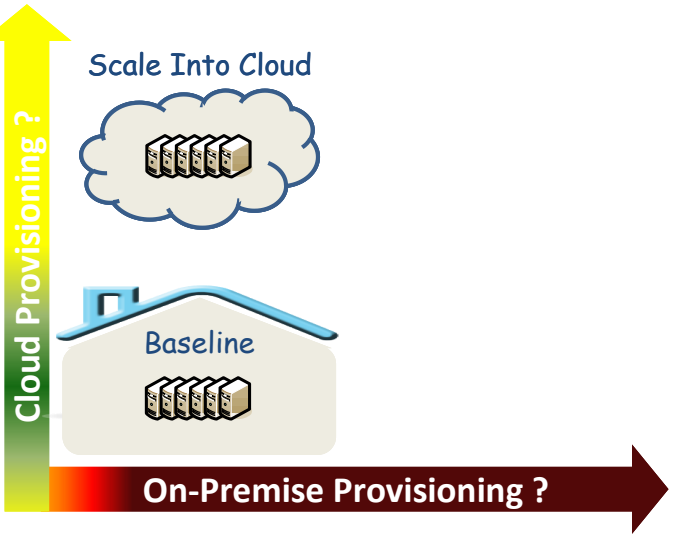
Case Study
Compute Cloud Cost per Month

✓Workload Growth
✓ Cloud Resources
➢ Energy, **Cost**, Performance

- Instance pricing per hour of use
- Data transfer cost based on GB quantity of in/out per month

Standard Instance	Total Number of Hourly Instances per Day	Windows Usage per Hour	Total Cost per Month
Large	163	$0.48	$2,380
Extra Large	95	$0.96	$2,774

Fee Category	Price per GB	GB per Month	Total Cost per Month
Transfer IN	$0.10	885	$89
Transfer OUT			
First 10 TB/Month	$0.17	10,000	$1,700
Next 40 TB/Month	$0.13	6,816	$818
Next 100 TB/Month	$0.11	-	-
Over 150 TB/Month	$0.10	-	-
TOTAL		17,701	$2,607



Standard Instance	Instance Cost per Month	Data Transfer Cost per Month	Total Cost per Month
Large	$2,380	$2,607	$4,987
Extra Large	$2,774	$2,607	$5,381

Comparisons
On-Premise vs. Cloud Scenarios

- Energy

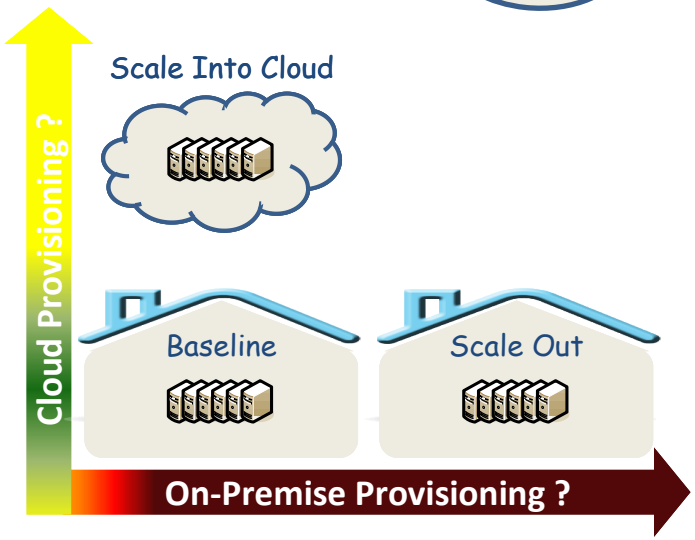
Cloud provides a way to cap our on-premise energy consumption

- Cost

Cloud requires a lower initial investment, but costs accumulate over time

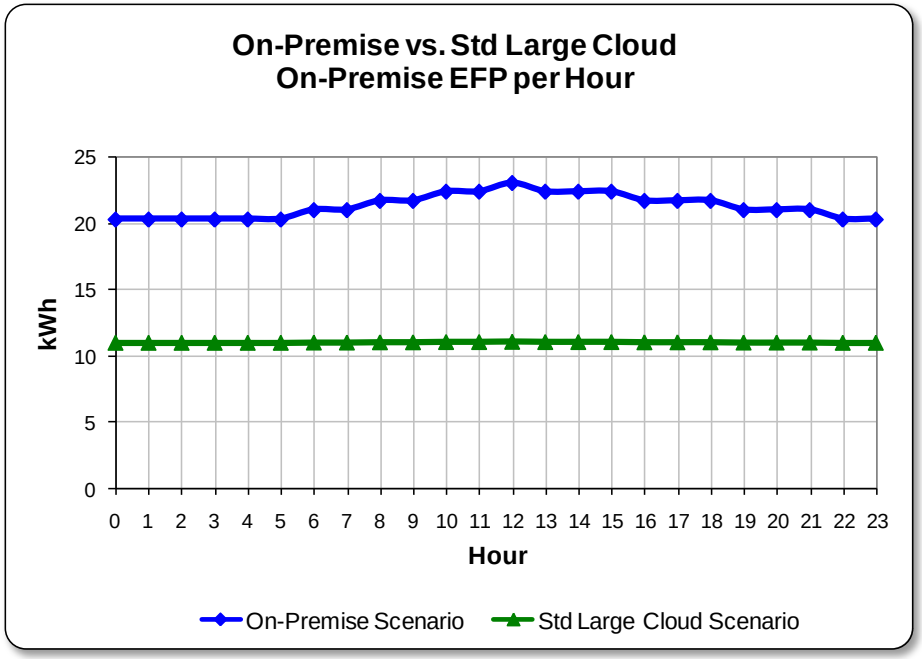
- Performance

Cloud may increase our business service response times



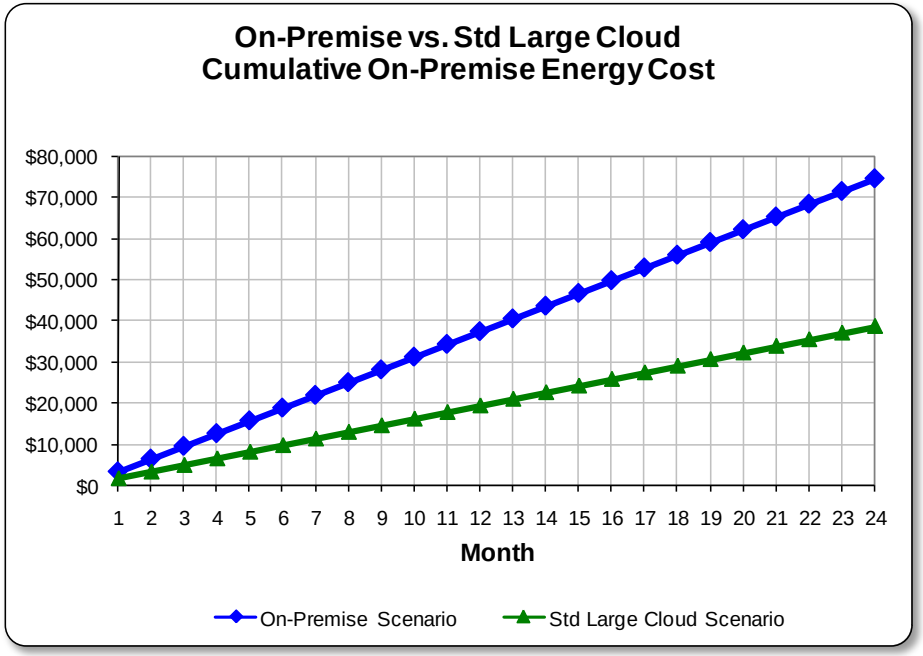
Comparisons
On-Premise Energy Only

✓Workload Growth
✓ Cloud Resources
➢ **Energy**, Cost, Performance



Hour-by-hour energy consumption

The peak on-premise EFP is 50% lower using the Cloud



On-Premise Scenario energy cost is twice that for the on-premise portion of the Cloud Scenario:

58 vs. 27 on-premise servers

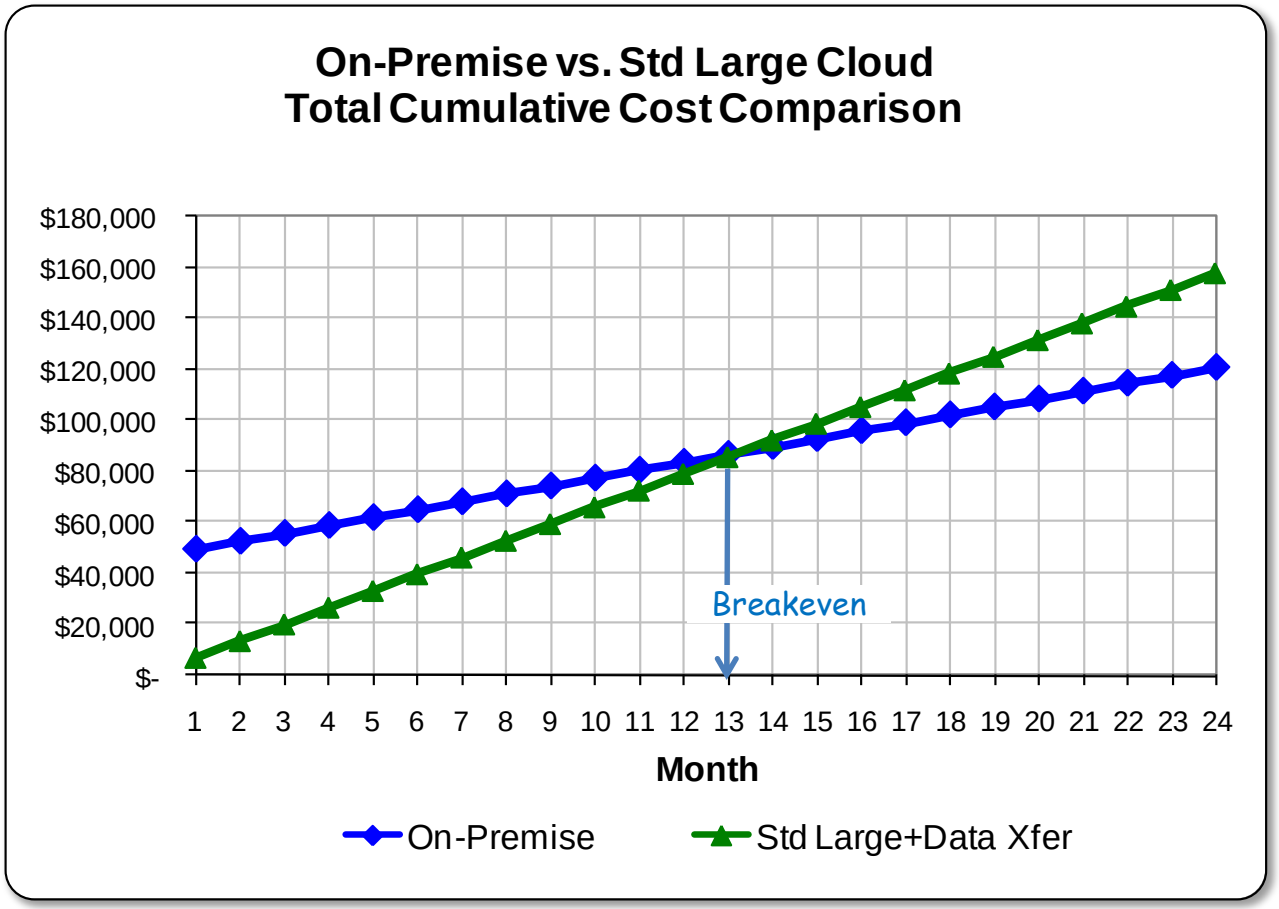
$3,098 vs. $1,602 per month

Comparisons
Total Cost On-Premise vs. Cloud

✓Workload Growth
✓ Cloud Resources
➢ Energy, **Cost**, Performance

On-Premise Scenario
- CapEx for 31 new servers
- Energy cost for built-out infrastructure (58 servers)

Cloud Scenario
- Energy cost for on-premise servers (27)
- On-demand monthly cost of Cloud instances & data transfer (includes Cloud vendor energy cost)

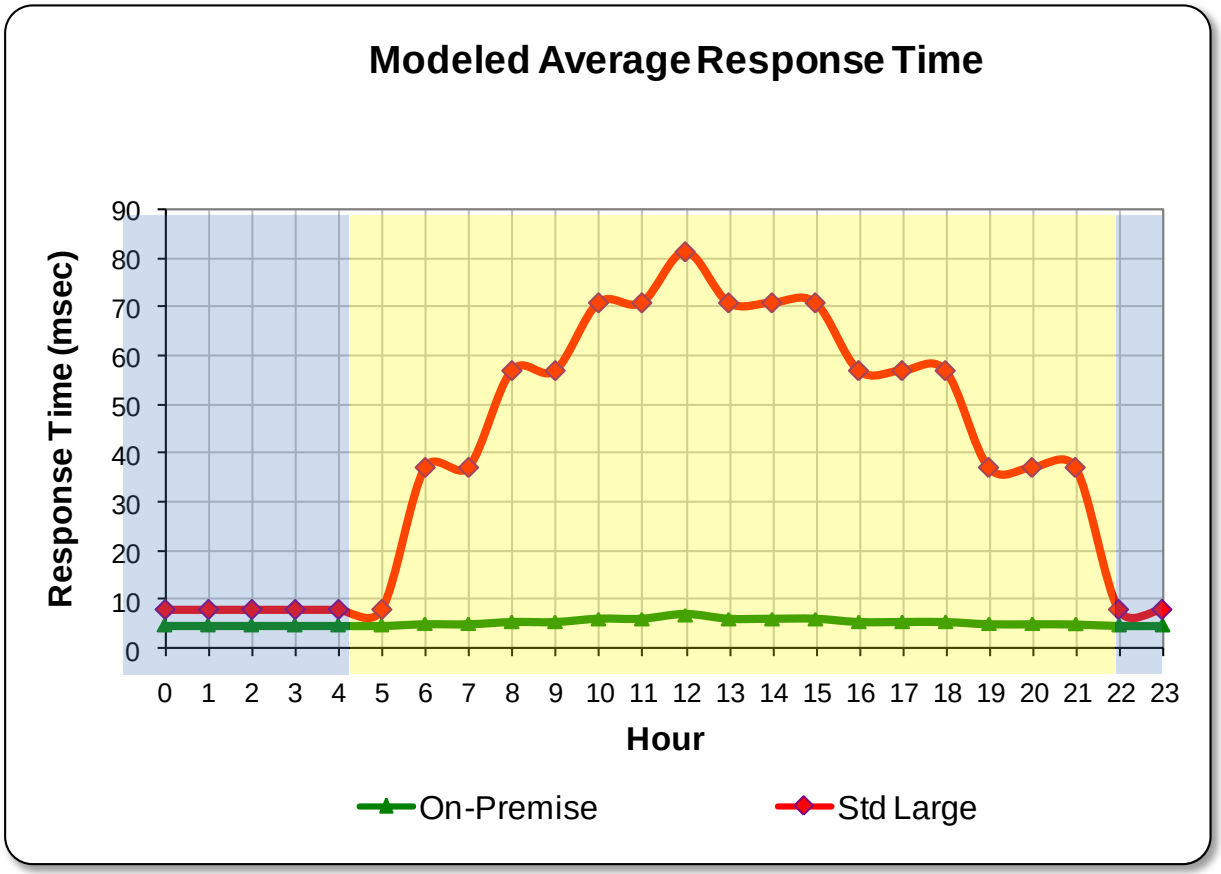


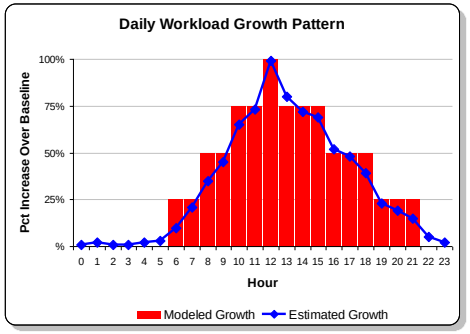
- Slope of Cloud curve is steeper, implying higher incremental cost (OpEx)
- On-Premise OpEx is $3,098 - Cloud OpEx is $6,589

Comparisons
Average Response Time

✓Workload Growth
✓ Cloud Resources
➢ Energy, Cost, **Performance**

- Keeping DB on-premise incurs additional network overhead
- Hybrid architectures
  - ➢ May increase response time beyond acceptable limits
  - ➢ May not be good candidates for chatty applications



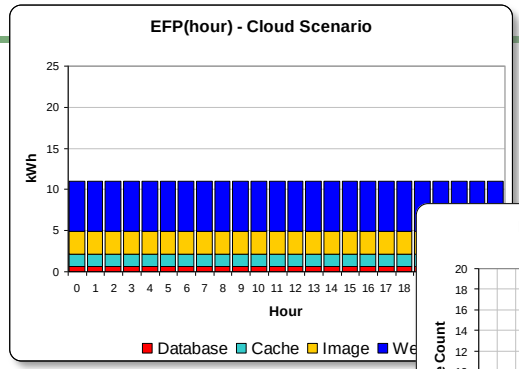


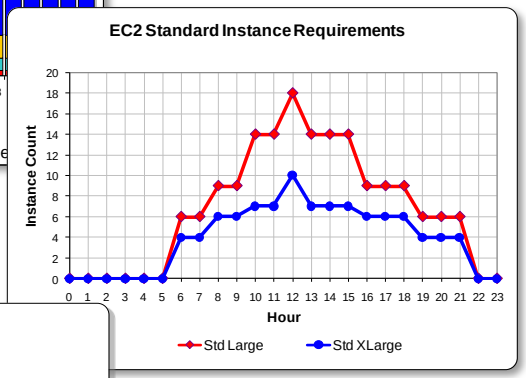
- On average 1.5 DB accesses per transaction (max of 5)
- Assume 100 msec round trip latency
- Max response time increase > 0.5 seconds

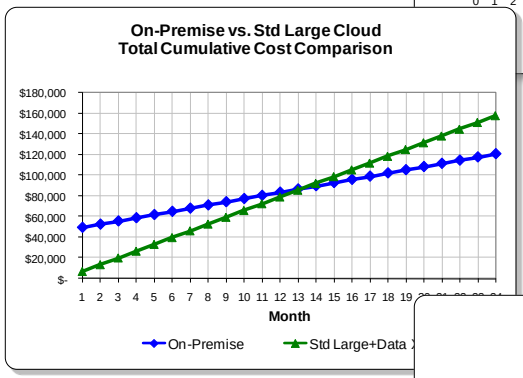
Summary

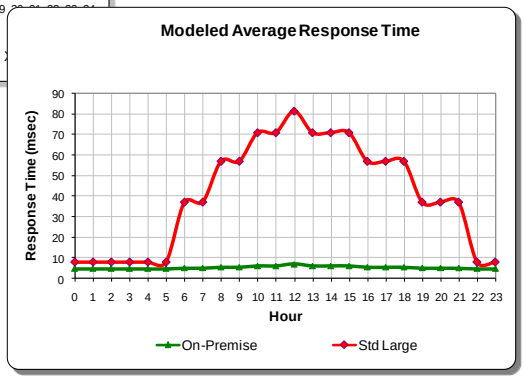
Did we meet our goals?

- ✓ Cloud enabled us to keep our on-premise energy requirements flat
- ✓ Dynamic provisioning enabled us to more closely match infrastructure & demand for energy, reduce the number of servers (KPI)
- ✓ Growing into the Cloud requires less up-front (CapEx) investment
- ? Cloud recurring cost can end up exceeding traditional data center cost
- ? Hybrid models can suffer an increase in response time due to the Cloud instance interactions with the on-premise servers









Summary
Capacity Planning Question

Should we expand our on-premise infrastructure to handle the expected growth, or should we leverage the Cloud to support the increased workload volume?

✓Grow into the Cloud

- ✓ Utilize on-demand Cloud to handle workload growth
- ✓ On-Premise energy usage will be effectively capped or reduced
- ✓ Defer or eliminate need for data center expansion
- ✓ Eliminates CapEx for required On-Premise infrastructure

➢But

- Cloud monthly OpEx are 2x larger than On-Premise
- Cumulative cost of Cloud overtakes on-premise solution
- Cloud increases average transaction response time

Recommendations
A Formal Methodology is Required

- Focus your analysis on the Cloud provider's pricing model(s)
  - Evaluate data transfer cost & any other miscellaneous costs carefully
- Manage your energy footprint by leveraging the Cloud
  - Dynamic allocation in the Cloud reduces full time resource requirements
  - Utilizing Cloud resources effectively caps or reduces your on-premise energy cost & consumption
  - Key Performance Indicator (KPI) for energy = Number of servers
- Don't ignore performance
  - Cloud response times must meet business requirements
- Balance CapEx & OpEx based on your business requirements
  - Cloud allows you to shift large upfront infrastructure investments into predictable recurring monthly costs

Amy Spellmann
amy@optimalinnovations.com

Richard Gimarc
rgimarc@hyperformix.com

Mark Preston
mark.preston@rsperform.com

Based on the CMG 2009 Best Paper

Leveraging the Cloud for Green IT:

Predicting the Energy, Cost & Performance of Cloud Computing

Applying Capacity Planning Expertise to Cloud & Virtualization

Amy Spellmann
amy@optimalinnovations.com

OPTIMAL INNOVATIONS

Founded in 2008, Optimal Innovations is a solutions company dedicated to saving Information Technology dollars by reducing IT and Data Center power consumption. 20+ years in IT Capacity Planning provides the basis for our Green Capacity Planning services.

Hyperformix offers products and solutions for capacity planning and management at progressive levels of sophistication and automation. Hyperformix provides software and services for application profiling, IT decision support and detailed performance operations.

Richard Gimarc
rgimarc@hyperformix.com

Hyperformix

Mark Preston
mark.preston@rsperform.com

RSPerformance

Founded in 2009, RS Performance provides companies the opportunity to Right-Size solutions early in the life-cycle with tailored services. We offer unique Product & Technology Management, Capacity, Performance and Reliability Services to ensure goals are achieved.