Best Execution Venue

Where is the best place to run a workload?

**External Public Cloud:**
Servers and storage on demand (AWS, Rackspace)

**Software-as-a-Service:**
Applications on demand (Salesforce, Google Docs)

**Private Cloud:**
We own the servers, network and storage - can be traditional or “cloud” (CloudStack, OpenStack)

**Hybrid:**
Private and public resources managed together as needed (RightScale, Dell)

**External Public Cloud:**
Servers and storage on demand (AWS, Rackspace)

**SaaS:**
Applications on demand (Salesforce, Google Docs)
Capacity Planning in Best Execution Venues

The Challenge

**Challenge**
- How do you apply Capacity Planning to different BEVs with the Stack?
- How does the Business formulate and communicate their plans to the Application, Infrastructure and Facilities planners?
- How do the levels communicate?

**Capacity Planning Stack**
- Describes & supports a structured approach
- Includes all elements of today's Digital Infrastructure
- Promotes transparency & communication
- Applies to all execution environments
Capacity Planning in Best Execution Venues

*Our Approach*

- Understand workload demand from the Business perspective
- Translate & communicate between levels of the stack (Demand & Feedback)
- Stack levels may be on-premise or remote as required to support BEV (e.g., IaaS)
- Evaluate efficiency & delivery at each candidate BEV
- Select the BEV that cost-effectively meets the Business demand
The Capacity Planning Stack

*Demand & Feedback*

**Demand (↓)**
- Volume & priorities
- Logical resource requirements
- Performance requirements & SLAs
- Budget

**Demand (↓)**
- Physical resource footprint & instances
- Performance requirements & SLAs
- Budget

**Feedback (↑)**
- Cost
- Total time to satisfy
- Expected performance
The Capacity Planning Stack Taxonomy

- Total business transactions (includes IT and non-IT based transactions)
- Total business revenue
- Number of business customers

- Business transactions per minute
- Total IT resources required to support the business
  - Application transaction volume for the Business
  - Response time for the part of the Business supported by Application C

- CPU utilization per server
- Total CPU capacity
- SAN I/Os per second
- Memory used per server

- Total power consumed by the data center
- Data center capacity
- PUE (Power Usage Effectiveness)

- Total power used to directly support the IT infrastructure

- Application transactions per kWh
- Application power footprint
The Stack & Stack Taxonomy

- Provides a structured approach to evaluating today’s Digital Infrastructure
- Simplifies & guides the selection of metrics required to support decision making
- Equips the capacity planner with a structured way to think about, organize, communicate and collect the metrics that apply to the decision support process
- Paves the way for federated capacity planning across today’s Digital Infrastructure
Best Execution Venue
Where is the best place to run a workload?

- **External Public Cloud**: Servers and storage on demand (AWS, Rackspace)
- **Software-as-a-Service**
- **Private Cloud**: We own the servers, network and storage - can be traditional or “cloud” (CloudStack, OpenStack)
- **Hybrid**: Private and public resources managed together as needed (RightScale, Dell)
- **External Public Cloud**: Servers and storage on demand (AWS, Rackspace)
- **SaaS**: Applications on demand (Salesforce, Google Docs)
• A portion of the infrastructure always remains in-house
• External communication of demand to IaaS provider
  - In-house Infrastructure to Cloud Provider
  - Physical hardware requirements
  - Feedback is cost based on instance usage & type, storage and network actuals
SaaS

Applying the Stack to BEV

- In-house Application and Infrastructure levels are not involved
- Business communicates directly to the SaaS Provider
  - Number of users, size of data stores, network traffic requirements, any specific SaaS customization or add-on features
  - Feedback is cost based on actual users, special features, storage and network
ITaaS – Best Execution Venues

*Vision of the Future*

- Every application has a best execution venue
  - Some are mature, others are evolving
  - All are headed toward “the cloud”

- **Cloud computing is mainstream**
  - This means choice, access and diversity for IT
  - Benefits are material - business drives choice of venue

- **There are ways to systematically make the best choice of venue for each workload**
  - Enterprises are making these choices today
  - Providers are/should be targeting these workloads and creating the venues
  - There is a snowball effect across the IT landscape as cloud begets automation and automation begets growth
ITaaS will choose the best execution venue

IT Brokerage model
Case Study
Applying the Capacity Planning Stack

- The Business projects a 6% monthly growth in their workload
- Management Directive: No new in-house hardware – grow into the cloud (IaaS)

- The capacity planner’s job is to determine the resource and infrastructure requirements for the next 24 months
- 3-tier Web application
- Compare IaaS vs all in-house
Applying the Capacity Planning Stack

Application: Translating Transaction Volumes

- Start with Business demand
- Translate to Application transaction rate
- Estimate volume over the 24-month planning horizon
- Next step is to estimate current and future resource requirements
**Applying the Capacity Planning Stack**

**Application: Baseline Resource Requirements**

**Application** - Profile application resource requirements

<table>
<thead>
<tr>
<th>Application Baseline</th>
<th>Web Tier</th>
<th>App Tier</th>
<th>DB Tier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS Instance Count</td>
<td>10</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total CPU Footprint</td>
<td>325</td>
<td>57</td>
<td>84</td>
</tr>
</tbody>
</table>

**CPU Footprint:**
Normalized measure of resource requirements
Applying the Capacity Planning Stack

Application: Estimate Future Resource Demand

- 24-month projection from baseline – average demand per month
- Instance count & CPU footprint requirements
- Next step – Application demand passed to the Infrastructure layer
Applying the Capacity Planning Stack

*Infrastructure: Estimate On-Premise Requirements*

- Infrastructure team uses demand from Application & Shared Services
- Database upgrade required immediately (>75% utilized)
- Estimate infrastructure requirements for 24-month planning horizon
- Next step - Infrastructure demand is passed to the Facilities layer
Applying the Capacity Planning Stack

*Facilities: On-premise power estimate*

- Data center upgrade is required to support the power demand
- Feedback up the stack to communicate power constraint
- Validates executive decision to move to the cloud
Next Steps
• Utilize the cloud for additional Web and App tier servers
• Maintain current data center infrastructure
• Size cloud instances based on CPU footprint
• Plan for the max workload volume – potentially ignore savings with elasticity
BEV Capacity Planning with the Stack

Application: Estimate Cloud Resource Requirements

- 24-month projection from baseline
- New Web & App instances grow into the cloud
**BEV Capacity Planning with the Stack**

*Infrastructure: CPU Footprint On Premise vs. Cloud*

- Infrastructure CPU requirements are identical (blue line)
- Cloud headroom closely follows CPU demand (add new cloud instances)
- Headroom is an efficiency metric for the Infrastructure layer
On Premise does not include the cost for a data center upgrade (gray area)
Note the recurring monthly cost for cloud instances
Total cumulative cost:
On Premise: $108k
Including data center upgrade: $237k
Cloud: $123k
• CPU capacity available versus what is used
• Headroom is the area between the available & used lines
• Goal is to have a predictable and steady amount of headroom
Number of Business Transactions per DI dollar (want large numbers)
On Premise does not include the cost of a data center upgrade
Cloud efficiency decreases due to the recurring monthly cost of cloud instances
Capacity Planning in Best Execution Venues

Summary

- Reviewed the Capacity Planning Stack
  - Simplify
  - Structure
  - Focus
- Mapped the Stack to BEV
- Demonstrated how to apply the Stack to compare BEV in the case study
  - On-premise
  - IaaS BEV

How can you use the Stack?

- Straightforward description of capacity planning methodology
- Includes all elements of today's Digital Infrastructure
- Evaluates & compares execution environments
- Promotes transparency & communication
Contact Info

- Amy Spellmann   amy@spe-ed.com
- Richard Gimarc  rgimarc@featherfall.com