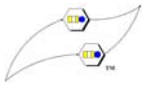


A Simalytic® Approach to Modeling Virtualized Environments

CMG2008 Paper 8199, December 12, 2008

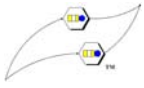
Dr. Tim R. Norton

Simalytic Solutions, LLC

719-635-5825

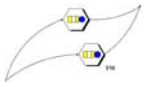
email: tim.norton@simalytic.com

http://www.simalytic.com



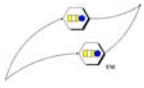
Agenda

- **Introduction**
  - Capacity Planning and Virtualization
- **The Problem**
  - What Virtualization does to Planning
- **A Proposed Solution**
  - A Simalytic Approach
- **Conclusion**



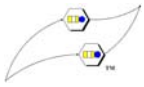
Introduction

- **Capacity Planning**
  - Capacity is Measured by Business Performance Objectives
    - Making decisions about resource requirements
  - What do we have to buy and when do we have to buy it to make sure that the business applications perform at the level required to insure the business succeeds?



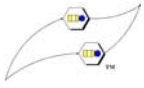
Introduction

- **Capacity Planning**
  - Two key aspects:
    - Demand for available resources
      - What do we have to buy
    - Effective completion of business work
      - When to buy it?
  - Requires some predictive technique
    - Usually some form of model
      - Simple: trend or other statistical techniques
      - Advanced: simulation or queuing network



Introduction

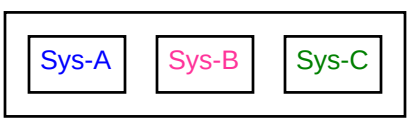
- **Virtualization**
  - Increasing parallelization within the host system
    - Increase productive business-related work
    - Increase the usage of resources
  - Virtualized environment control program
    - Hypervisor
      - Usually implies a hardware implementation
    - VMM (Virtual Machine Monitor)
      - VMM usually implies a software implementation
        - Will use VMM for all virtualization control programs

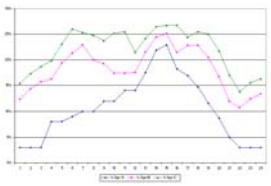


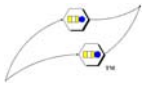
Introduction

- **Virtualization**
  - From:

  - To:

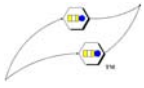






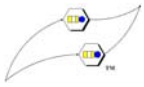
The Problem

- **What Virtualization does to Planning**
  - Generalizes the measurements
    - Virtualization abstracts the underlying resources
  - Increases the scope for sharing
    - Dedicated resources become shared
  - Complicates Workload Characterization
    - Virtualization allocates resources as needed
      - Not in the same ratio from one time interval to the next
  - Affects measurements needed for models



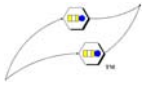
Capacity Models

- **Models are Key Tools**
  - Require detailed measurements
    - How work is grouped – Workload Characterization
    - How resources are used – Service time
    - How work changes – Arrival rate
  - Allow Capacity Planner to predict the future
    - Show future resource usage
    - Show future performance



Capacity Models

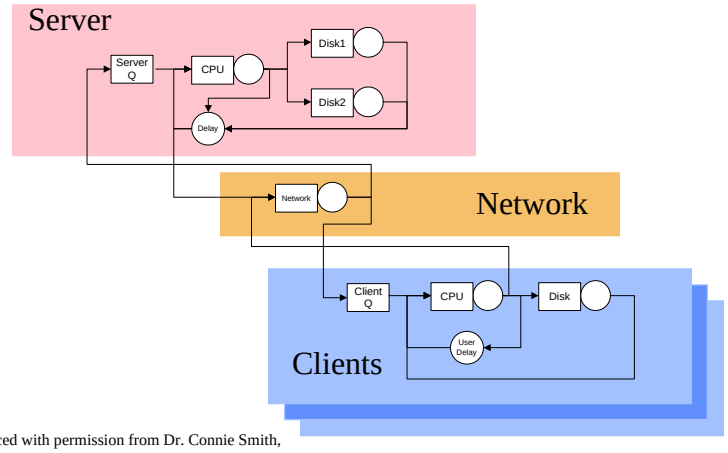
- **What Virtualization does to Models**
  - Reduces accuracy – Measurement issues:
    - System clock
    - Accounting for dispatch time of other VMs
    - Interrupts – delays and re-driven to VM
    - VM vs. process priority
    - Delays – where are they accounted for
    - Virtualization overhead
      - VMM (Virtual Machine Monitor) and VM context switch
    - Interference from other VMs
  - Complicates Workload Characterization
    - Shared resources



Capacity Models

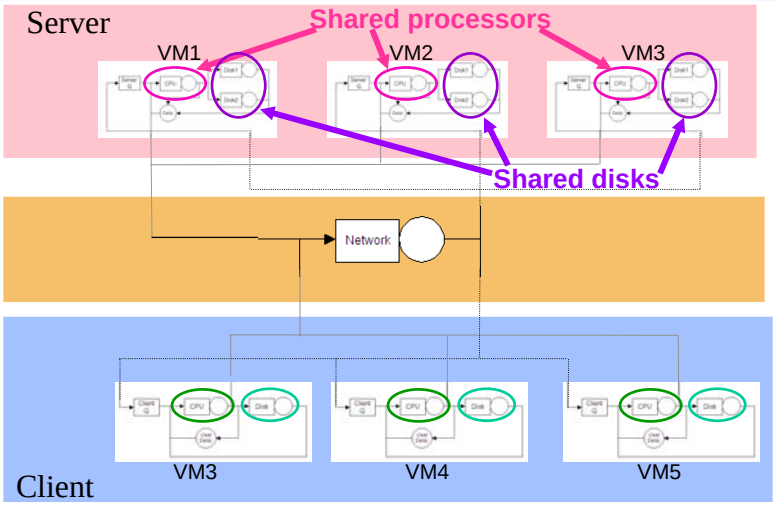
- **What Virtualization does to Models**
  - Complicates model design
    - More resources
    - More relationships
    - Sharing rules
    - Increases model scenarios

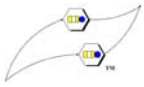
Application Model



Reproduced with permission from Dr. Connie Smith, Performance Engineering Services.

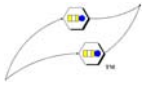
Virtualization Model





A Proposed Solution

- **A Simalytic Approach**
  - Use known "good" measurement
    - VM utilization
      - from the VMM (Virtual Machine Monitor)
    - Transaction arrivals
      - from application measurements
    - Transaction response times
      - from application measurements
  - Use a Simalytic Model
    - Builds relationships
    - Iterative design
    - Leverages existing tools



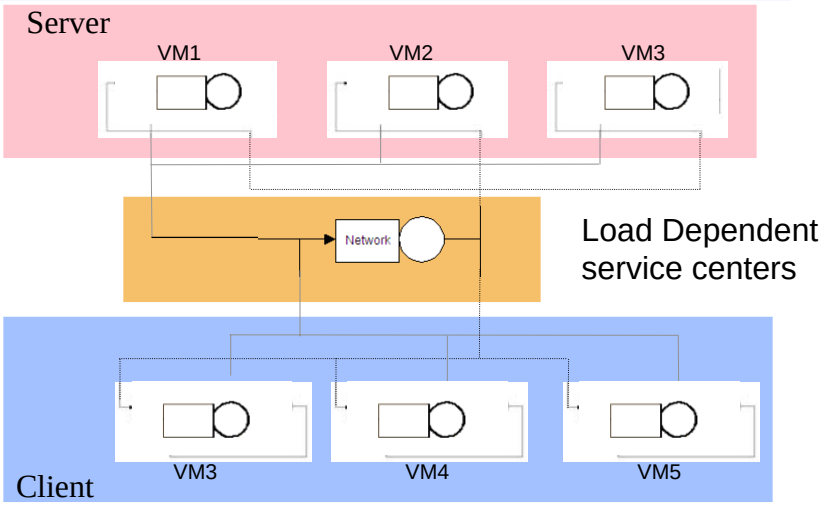
What is Simalytic Modeling?

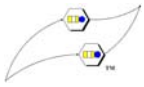
- **"Simalytic" (Sim**ulation/An**alytic)**
  - Hybrid modeling technique
    - General purpose simulation framework
    - Platform-centric analytic node models
  - Existing tools
  - Predict capacity requirements
  - Heterogeneous computer systems
  - Enterprise level application model
- **Goal**
  - Model computer application requirements at the enterprise level

Hybrid Modeling

- **Combination of Techniques**
  - Simulation model as framework
  - Analytic queuing theory node models
- **Transform Function Bridges Techniques**
  - Simalytic Function maps arrival rates to response times
    - Load dependent Service Centers change service times with changes in arrival rates

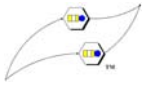
Simalytic Model





Useful Modeling Tools

- **Platform-Centric Tools**
  - Narrow focus
    - Detailed Information about single platform
    - Easier to build but only environments built into tool
  - Tend to be Analytic based
- **General Purpose Tools**
  - Broad focus
    - Features to model anything
    - No “Built-in” platform understanding
    - Level of granularity = Level of effort
  - Tend to be Simulation based
- **Both Available as Commercial Tools**

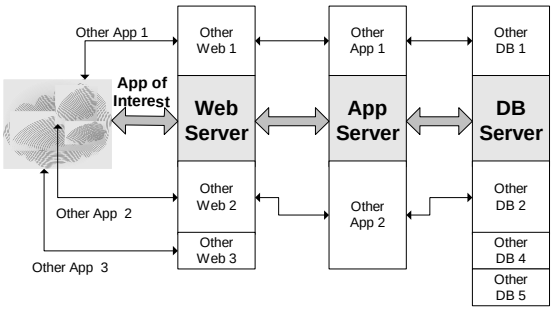


Virtualization Implementation

- **Hybrid Modeling**
  - Combines data and results from existing tools
    - Traditional operating system measurements
    - VMM measurements (mix of vendors)
  - Adjusts model based on measured results
    - Reduces requirements for precise measurements
  - Application focus across tiers
    - Mixed virtualization and non-virtualization systems

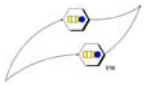
Virtualization Environment

- **Multiple virtual systems on few physical hosts**
  - Many challenges for selecting systems to combine
  - Resource usage has direct impact on response times



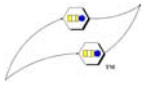
Modeling Issues

- **Different possible modeling scenarios**
  - Single Application of interest
    - Model only the transaction workload of interest
      - Treat all other workloads as static interference
    - Effective when the other workloads are relatively static
  - Total Environment
    - Model all transaction work-loads that share the same virtual environment
      - Creating and running complex models is seldom practical
      - Many sceneries to find balance to support reasonable arrival rate for all workloads simultaneously



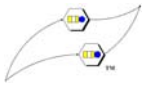
Modeling Issues

- **Different possible VM scenarios**
  - Single Workload Guest
    - VM is only processing the application's transactions
      - Virtualization can be included in "overhead"
  - Multiple Workload Guest
    - VM processes other significant workloads in addition to the transactions of the application of interest
      - Service time of each workload must be adjusted in some way that minimizes the influence from the other workloads



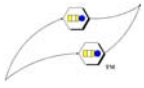
Modeling Examples

- **Single Application of Interest in the Single Workload Guest**
  - Account for the interference
    - From the other VMs and the VMM overhead
  - Adjust service times for the Application of Interest
    - Model response times match measured response times as the utilization of the other VMs increases
    - Other VMs reduces the capacity of the AoI
      - Issues:
        - How VMM prioritizes VM dispatch
        - How and where interrupts are processed
        - Use of shared resources by multiple VMs



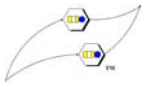
Modeling Examples

- **Single Application of Interest in the Multiple Workload Guest**
  - Account for the interference
    - From the other VMs and the VMM overhead
    - From the other workloads
  - Adjust service times for the Application of Interest
    - Model response times match measured as
      - the utilizations of the other VMs increase
      - the arrival rates of the other workloads increase
    - Other VMs and workloads reduce capacity of the AoI
      - Issues, same as last example plus:
        - Accuracy of measurements within VM
        - Prioritization of other workloads



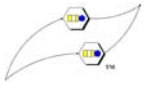
Modeling Examples

- **Total Environment with Multiple Workload Guests**
  - Account for the interference on each VM
    - From all other VMs and VMM overhead
    - From all the other workloads
  - Adjust service times for all Applications
    - Model response times match measured as
      - the utilizations of the other VMs increase
      - the arrival rates of the other workloads increase
    - Other VMs and workloads reduce capacity of the AoI
      - Issues, same as last example plus:
        - Work distribution over time
        - Appropriate arrival rates to use for each application



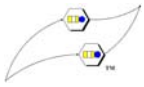
Model Validation

- **Use "known good" measurements**
  - Application response time
  - VM utilizations from the VMM
- **Variety of intervals**
  - Select actual production or load-test measurements and arrival rates
    - If model results do not match actual response times
      - Adjust load dependent service center profiles until they do
    - Once all of test cases have been validated
      - Use model to explore other scenarios



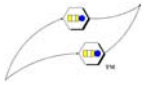
Conclusion

- **Virtualization**
  - Makes planning harder and less precise
  - Back to the basics of capacity planning
    - Understand the needs of the business
    - Think in terms of achieving the business objectives
      - Efficient use of resources isn't the issue
        - it's almost impossible to understand
    - Holistic view that balances Costs against Benefits
      - Technical, financial, human and political
      - There will always be localized inefficiencies



Conclusion

- **Simalytic Modeling for Virtualization**
  - Technique that allows
    - Using available tools in combination
      - Incorporate measurements from VMM with existing tools
    - Reduces complexity of overall environment
      - Fewer model components and scenarios
    - Explores both performance and capacity constraints
      - Focused on business objectives
    - Leverages existing tools
      - Modeling and data collection
      - Nothing to buy!
    - Addresses mixed environments
      - Virtualized and non-virtualized systems



Questions

?

Presentation will be available at:
http://www.simalytic.com/Papers.html