A Cost-oriented methodology for the design of Web based IT architectures

Danilo Ardagna

Politecnico di Milano
Italy
Summary

- Research objectives and Motivation
- Presentation of proposed methodology
- ISIDE
- Empirical verifications
- Conclusions and Future work
Research objectives and Motivation

- **Objectives**
  - Support the Architectural Design of Web IT Systems minimizing TCO

- **Motivation**
  - The Architectural design of IT systems is traditionally led optimizing single subsystems separately
  - Fill literature gap
Design Methodology: architectural alternatives

- Thin vs. fat clients
- Number of tiers
- Total number of servers
- Allocation of applications
Design Phases

System Designer defines exploration domain

Technology Requirements Model

Infrastructural Model

Optimization

Physical Model
Technology Requirements Model

- Organization Sites $S_i$
- Buildings $B_i$
- Applications $A_i$
- User classes $C_i$
- Requests $R_i$
- Databases $D_i$
## Technology Requirements Model - An example

**Request R₁**
- \( f(R₁) = 3 \) req/min
- \( \Phi(R₁) = \begin{array}{c|cc}
  \text{Application} & \text{CPU-Time} & \text{disk-time} \\
  \hline
  A₂ & 1 & 0.2 \\
  A₃ & 1.5 & 0.1 \\
  A₄ & 2 & 1.3 \\
  A₅ & 2.4 & 1.1 \\
\end{array} \)

**Request R₂**
- \( f(R₂) = 1 \) req/min
- \( \Phi(R₂) = \begin{array}{c|cc}
  \text{Application} & \text{CPU-Time} & \text{disk-time} \\
  \hline
  A₆ & 0.5 & 0.2 \\
  A₇ & 1 & 0.3 \\
\end{array} \)

**System Specifications**
- Type(A₁) = monolithic
- \( \Delta₁(A₁) = \begin{array}{c|c|c|c}
  \text{MIPS} & \text{RAM} & \text{OS} \\
  \hline
  120 & 16 & W98 \\
  150 & 32 & W2000 \\
\end{array} \)
- Type(A₂) = server
  - Tuning-system(A₂) = (W2000, PIII 450, Raid-5 SCSI)
  - RAM(A₂) = 512 MB
  - \( d(A₂) = 150 \) MB

**Application Performance**
- **CPU-Time**
  - A₇
    - 100 B
    - 200 B
  - A₆
    - 20 KB
    - 100 KB
  - A₄
    - 50 KB
    - 1 KB
  - A₃
    - 20 KB
    - 1 KB
  - A₂
    - 150 KB
    - 150 B
  - A₁
    - 20 KB
    - 100 KB
  - A₅
    - 50 KB
    - 1 KB
  - A₈
    - d(A₈) = 50 MB
  - A₉
    - d(A₉) = 50 MB

---

D. Ardagna, C. Francalanci - SAC 2002
The Infrastructural Design Model

- Goal: Build a virtual IT architecture with components that can meet requirements with no approximation

- To a Technology Requirements Model are associated $N$ Infrastructural Models, vice versa to an Infrastructural Model a single Physical Model is assigned

- The solution domain is defined specifying:
  - Type of client computers
  - Number of computing levels of requests
## The Infrastructural design Process

<table>
<thead>
<tr>
<th>Virtual Computing Resource</th>
<th>Symbol</th>
<th>Variables</th>
<th>Analytical Formulation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Server</td>
<td>VS&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Frequency of requests</td>
<td>f(R&lt;sub&gt;i&lt;/sub&gt;)</td>
<td></td>
</tr>
<tr>
<td>Primary Memory</td>
<td></td>
<td>RAM = ∑&lt;sub&gt;VS&lt;sub&gt;i&lt;/sub&gt;&lt;/sub&gt;RAM(A&lt;sub&gt;i&lt;/sub&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual application server for WT/HFC</td>
<td>VAS&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Computing Power</td>
<td>MIPS = n(C&lt;sub&gt;i&lt;/sub&gt;)·p(C&lt;sub&gt;i&lt;/sub&gt;)·[\text{max}\ (\text{MIPS}(A&lt;sub&gt;i&lt;/sub&gt;, OS, RP, \text{think-time}(C&lt;sub&gt;i&lt;/sub&gt;)))]</td>
<td>MIPS(A&lt;sub&gt;i&lt;/sub&gt;, OS, RP, \text{think-time}(C&lt;sub&gt;i&lt;/sub&gt;)) returns MIPS required to support execution of application A&lt;sub&gt;i&lt;/sub&gt; on the target operating system OS under the specified remote protocol RP for a user of specified think-time; p(C&lt;sub&gt;i&lt;/sub&gt;) returns the percentage of concurrent users in class C&lt;sub&gt;i&lt;/sub&gt;</td>
</tr>
<tr>
<td>Primary Memory</td>
<td></td>
<td>RAM = n(C&lt;sub&gt;i&lt;/sub&gt;)·p(C&lt;sub&gt;i&lt;/sub&gt;)·[\sum\text{RAM} (A&lt;sub&gt;i&lt;/sub&gt;, OS, RP, \text{think-time}(C&lt;sub&gt;i&lt;/sub&gt;))]</td>
<td>RAM(A&lt;sub&gt;i&lt;/sub&gt;, OS, RP) returns RAM required to support execution of application A&lt;sub&gt;i&lt;/sub&gt; on the target operating system OS under the specified remote protocol RP for a user of specified think-time.</td>
<td></td>
</tr>
<tr>
<td>Virtual fat client</td>
<td>VFG&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Computing Power</td>
<td>MIPS = [\text{max}\ (\text{MIPS}(A&lt;sub&gt;i&lt;/sub&gt;, OS))]</td>
<td>MIPS(A&lt;sub&gt;i&lt;/sub&gt;, OS) returns MIPS required to support execution of application A&lt;sub&gt;i&lt;/sub&gt; on the target operating system OS</td>
</tr>
<tr>
<td>Primary Memory</td>
<td></td>
<td>RAM = ∑&lt;sub&gt;FG&lt;sub&gt;i&lt;/sub&gt;&lt;/sub&gt;RAM(A&lt;sub&gt;i&lt;/sub&gt;, OS)</td>
<td>RAM(A&lt;sub&gt;i&lt;/sub&gt;, OS) returns RAM required to support execution of application A&lt;sub&gt;i&lt;/sub&gt; on the target operating system OS</td>
<td></td>
</tr>
<tr>
<td>Virtual HFC</td>
<td>VHFC&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Computing Power</td>
<td>MIPS = [\text{max}\ (\text{MIPS}(A&lt;sub&gt;i&lt;/sub&gt;, OS))]</td>
<td>MIPS(A&lt;sub&gt;i&lt;/sub&gt;, OS) returns MIPS required to support execution of application A&lt;sub&gt;i&lt;/sub&gt; on the target operating system OS</td>
</tr>
<tr>
<td>Primary Memory</td>
<td></td>
<td>RAM = ∑&lt;sub&gt;FG&lt;sub&gt;i&lt;/sub&gt;&lt;/sub&gt;RAM(A&lt;sub&gt;i&lt;/sub&gt;, OS)</td>
<td>RAM(A&lt;sub&gt;i&lt;/sub&gt;, OS) returns RAM required to support execution of application A&lt;sub&gt;i&lt;/sub&gt; on the target operating system OS</td>
<td></td>
</tr>
</tbody>
</table>
The Physical Design Model

- Optimization iterates the following steps:
  - Selection of an Infrastructural Model
  - Association of commercial components with virtual computing resources

- Commercial components are selected:
  - Virtual Server: Resource Utilization lower than 60%
  - Virtual WT and HFC servers, FC, HFC: MIPS $\geq$ MIPS*
  - Virtual thin clients: design criteria that discriminate low and high performance devices are satisfied
Optimization algorithm: Tabu Search

- **Heuristic** algorithm
- **Initial solution**
- Defining possible **movements**
- **Neighborhood** exploration
- **Tabu list** check
Empirical Verifications

- Thin vs. fat clients

<table>
<thead>
<tr>
<th></th>
<th>20% Management Cost reduction</th>
<th>35% Management Cost reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEW</td>
<td>19%</td>
<td>35%</td>
</tr>
<tr>
<td>STW</td>
<td>23%</td>
<td>39%</td>
</tr>
<tr>
<td>KW</td>
<td>30%</td>
<td>44%</td>
</tr>
</tbody>
</table>

- Number of tiers

Minimum cost solution savings
Empirical Verifications

- Total number of servers:
  - Application server for raising frequencies
  - Server farm supporting WT users

![Configuration Cost](chart1)

![Minimum cost solution savings](chart2)

Average saving 79.92%
Conclusions and Future Work

- The methodology allows the identification of the architectural solution that minimizes costs.
- Preliminary results show that cost reductions can be significant.
- Cost oriented approach can be a complement to traditional performance evaluations.
- Future work: Network design alternatives and legacy systems.