Analytical treatment of a system

• Find ways to predict the behavior of a system through a model
  – Know what will happen without having to run an experiment
  – Use it to predict behavior
  – Gain more understanding about the system
  – Extrapolate behavior and results beyond what is feasible to test empirically

• Use experimental data to refine and validate the model
Tools

• Statistics
• Queuing theory and models
• Operational laws
• Intuition and insight on how the system works
For milestone 2

• A template will be made available
• Many possibilities for building models
  – Depends on the architecture of the system
  – Depends on what has been measured
  – Depends on what works
  – Depends on what aspects of the system are relevant
• Choosing what to model and how is part of the exercise
Example 1

• For each load level (number of clients), five experiments are run and the table shows the throughput for each experiment in operations per second, the average, and the standard deviation. Clients have a think time of 1 second and the experiments are run as a closed system.
### Experimental data

<table>
<thead>
<tr>
<th>Load</th>
<th>Repetition</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>AVG</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4.60</td>
<td>4.90</td>
<td>5.00</td>
<td>4.70</td>
<td>4.80</td>
<td>4.80</td>
<td>0.16</td>
</tr>
<tr>
<td>10</td>
<td>9.40</td>
<td>9.60</td>
<td>9.20</td>
<td>9.92</td>
<td>9.80</td>
<td>9.58</td>
<td>0.29</td>
</tr>
<tr>
<td>20</td>
<td>16.10</td>
<td>15.30</td>
<td>17.40</td>
<td>14.80</td>
<td>15.20</td>
<td>15.76</td>
<td>1.03</td>
</tr>
<tr>
<td>25</td>
<td>17.20</td>
<td>14.90</td>
<td>17.01</td>
<td>15.90</td>
<td>19.30</td>
<td>16.86</td>
<td>1.65</td>
</tr>
<tr>
<td>30</td>
<td>16.50</td>
<td>17.00</td>
<td>12.80</td>
<td>16.80</td>
<td>13.50</td>
<td>15.32</td>
<td>2.00</td>
</tr>
<tr>
<td>35</td>
<td>8.30</td>
<td>14.10</td>
<td>10.25</td>
<td>11.20</td>
<td>17.30</td>
<td>12.23</td>
<td>3.52</td>
</tr>
<tr>
<td>40</td>
<td>12.50</td>
<td>4.30</td>
<td>5.20</td>
<td>4.90</td>
<td>9.80</td>
<td>7.34</td>
<td>3.62</td>
</tr>
</tbody>
</table>
Throughput (just plot it)
Response time

Response time graph showing an increasing trend with time.
Pay attention to what the numbers say

<table>
<thead>
<tr>
<th>LOAD</th>
<th>X</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4,80</td>
<td>0,041667</td>
</tr>
<tr>
<td>10</td>
<td>9,58</td>
<td>0,043406</td>
</tr>
<tr>
<td>15</td>
<td>14,26</td>
<td>0,051893</td>
</tr>
<tr>
<td>20</td>
<td>15,76</td>
<td>0,269036</td>
</tr>
</tbody>
</table>

At 20 clients the response time is 5 times higher (system is 5 times slower)
Service rate

• Based on those numbers:
  – What would be the service rate of the system (modeled as an M/M/1)
  – How would you calculate the number of jobs in the system?
    • 5  0,2
    • 10  0,416
    • 15  0,74
    • 20  4,24
    • 25  8,138
And now what?

• You have characterized the system
  – Maximum performance
  – Inflexion points
  – Service rates at different loads
  – Queue length
  – …

• Can you explain why these things are the way they are? => map to the design
Further analysis

• Start with system as one block, then divide in smaller parts and see what you can say about each one of them
  – Where is the bottleneck (the database or the middleware?)
  – In the middleware, where is the bottleneck?
• Use 2k experiments approach to evaluate more aspects of the system once you know where the bottlenecks are