ENCE 667
Final Report

Project Performance Measurement Techniques in the Development and Implementation of a Data Warehouse

Presented By:
Nizar A. Al Rihani
Michael J. Casey
Overview

- Development of a Data Warehouse for major telecommunications company.
- A professional consultant was hired to develop the Warehouse.
- Two phases were considered.
  - Recurring task: Monthly downloading of data.
  - Development and Implementation
Project Performance Methods Used

- **Goal Programming.** Develop an optimal strategy for downloading data to DW on a monthly basis.
  - Software: LINDO
- **Monte Carlo Simulation.** Assess variability in schedule and cost due to uncertainty.
  - Software: MS Project, @Risk
Phase II – 6 mo. contract

- *5th – 25th of each month:* Receiving and downloading data files into the Warehouse.
- *25th of the current month – 5th of the next month:* Data processing and Warehouse development.
Goal Programming

- **Goal 1:** Downloading data files should be completed in 20 days. Otherwise, there is a penalty of $850/day.
- **Goal 2:** Data processing and evaluation should be completed in 10 days. Otherwise, there is a penalty of $1000/day.
- The Warehouse receives data files from two main offices, located in New York and Atlanta.
- A linear program is formulated to determine the optimal number of files to be received from each office.
Problem Formulation

- **Input:**
  
  \[
  \begin{align*}
  &\text{MIN } 850S_1P + 1000S_2P \\
  &\text{SUBJECT TO} \\
  &0.25X_1 + 0.1667X_2 + S_1N - S_1P = 20 \\
  &0.1667X_1 + 0.0833X_2 + S_2N - S_2P = 10 \\
  &455X_1 + 270X_2 = 22680 \\
  &\text{GIN } X_1 \\
  &\text{GIN } X_2
  \end{align*}
  \]

- **Result:** No penalties. 84 files received from Atlanta office, no files received from the New York office.

- **Input:**
  
  \[
  \begin{align*}
  &\text{MIN } 850S_1P + 1000S_2P \\
  &\text{SUBJECT TO} \\
  &0.25X_1 + 0.1667X_2 + S_1N - S_1P = 20 \\
  &0.1667X_1 + 0.0833X_2 + S_2N - S_2P = 10 \\
  &455X_1 + 270X_2 = 22680 \\
  &X_1 \geq 1 \\
  &X_2 \geq 1 \\
  &\text{END}
  \end{align*}
  \]

- **Result:** No penalties. 49 data files received from New York office, one file received from Atlanta office.
Phase I

- **Phase 1:**
  1. Determine the DW requirements.
  2. Design and develop the system code.
  3. Test the system performance.
  4. Redesign based on testing.
  5. Conduct user testing.
  6. Integrate the DW with existing systems.
  7. Document development & implementation process.
  8. Plan for future work.

- **AON Network:**
Monte Carlo Simulation

• **Uncertainties**: task durations in coding, DW development and resource utilization.

• **Apply Simulation to provide output effect on project duration and cost.**

• **Data from real project**: Tasks, monthly time sheet data.
Modeling Uncertainty

- **BestFit** used to fit distributions to employee utilization data.

- **@Risk Functions**
  - **Duration (Uniform)**

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Duration</th>
<th>Cost</th>
<th>@RISK Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data Warehouse</td>
<td>95.75 days</td>
<td>$983,498.00</td>
<td>Duration=RiskUniform(96.3,95)</td>
</tr>
<tr>
<td>2</td>
<td>Requirements Analysis</td>
<td>32 days</td>
<td>$63,595.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Development (R&amp;D and Loading)</td>
<td>18 days</td>
<td>$67,600.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Develop ETL code and processes for loading the</td>
<td>16 days</td>
<td>$17,560.00</td>
<td>Duration=RiskUniform(16,26)</td>
</tr>
<tr>
<td>5</td>
<td>Develop ETL code and processes for loading the</td>
<td>16 days</td>
<td>$17,560.00</td>
<td>Duration=RiskUniform(16,26)</td>
</tr>
<tr>
<td>6</td>
<td>Develop ETL code and processes for loading the</td>
<td>15 days</td>
<td>$9,220.00</td>
<td>Duration=RiskUniform(15,23)</td>
</tr>
<tr>
<td>7</td>
<td>Consolidate ETL code (processed)</td>
<td>6 days</td>
<td>$40,000.00</td>
<td>Duration=RiskUniform(5,10)</td>
</tr>
<tr>
<td>8</td>
<td>Add 4 new MicroStrategy reports (already existing)</td>
<td>6 days</td>
<td>$2,600.00</td>
<td>Duration=RiskUniform(4,10)</td>
</tr>
<tr>
<td>9</td>
<td>Develop 4 new JDRS modules (previously spent</td>
<td>5 days</td>
<td>$2,890.00</td>
<td>Duration=RiskUniform(5,10)</td>
</tr>
</tbody>
</table>

- **Utilization**

<table>
<thead>
<tr>
<th>ID</th>
<th>Resource Name</th>
<th>@RISK Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>Max. Units=RiskEXPON(10)</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>Max. Units=RiskUNIFORM(0,250)</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>Max. Units=RiskTRIANG(0,137.5,215.4)</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>Max. Units=RiskTRIANG(0,137.5,196)</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>Max. Units=RiskUNIFORM(0,176.3)</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>Max. Units=RiskTRIANG(0,0,159)</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>Max. Units=RiskUNIFORM(0,176.3)</td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>Max. Units=RiskTRIANG(0,0,193.8)</td>
</tr>
</tbody>
</table>
Results

- Cost Distribution data:
- Sensitivity Analysis (tornado diagram)
Conclusions

- Development of a Data Warehouse is a necessary step to optimize profits.
- Monte Carlo simulation is an effective method to measure changes in cost and schedule.
- "LINDO" and "@Risk" are simple, effective and user-friendly software.
- Optimal number and exact locations of data centers should be determined to minimize the cost of data transfer.