Inspecting Workflow Data: Challenges and Opportunities

Angela Bonifati,
DEI-Politecnico di Milano
bonifati@elet.polimi.it
Roadmap

• Problem’s overlook

• The goals:
  • Broadening the effort: the BPI Cockpit
  • Narrowing the effort: Resource Rating

• Envisioned solutions

• Results and future challenges
Problem statement.....

- WfMS are software platforms tailored to define, monitor and manage business processes
- WfMS have a great potential for data analysis, but they do not fully exploit it
  - they solely provide internal basic report functionality
  - external analysis tools can be plugged in, but need to be heavily "customized"
  - resources data play a crucial role and may reveal useful trends....
Toward a more general problem...

- Companies are always engaged in the effort of improving the **quality** of their business processes and services

  Less costs →  
  - **Internal** quality, as perceived by the service provider (e.g. reduced operating costs)

  More revenues →  
  - **External** quality, as perceived by the user (e.g., better service)
Business Processes, Services, and Workflows

• Business processes and services are supported by Workflow tools.
• These tools monitor many aspects of process execution and service delivery
  • Who does what, when, how long do they take

• Most WfMS offer basic log analysis functionality, such as the # of processes completed in a given period and their average execution time
• However, more comprehensive analysis is needed!
Samples of hidden knowledge

- **Identify** low-quality executions, understand their “cost” and understand the causes
  - Maybe resources are overloaded, especially on Mondays? Assigned to the wrong job?
- **Optimize** process definitions, executions, resource assignment
- **Predict** risks of missing a SLA or inability to handle loads
- At a more basic level: *Quickly build accurate* reports to view processes and resources across several dimensions (time, cost,..)

Extracting this kind of knowledge today is impossible or, in the best case, requires lots of knowledge and implementation effort and is error-prone
Current situation: reporting tools

- Workflow Design Engineer
- Business Process Analyst
- System Administrator
- IT Manager
- Business Manager/Analyst

(built in or external)
Reporting tools

Workflow Engine
Workflow Audit Logs
Inadequate support

Writing the “right” queries is very difficult and time-consuming

• What is the performance and outcome of activities executed on Fridays? Or in the early morning? Is it better than Mondays?

• Which resources perform best for a given activity?

• How the relative performance of a resource changes with the execution time of day, day of week, or with the activity at hand?
Inadequate support

- Operates on dirty data
  - Missing or erroneous values
  - Special codes (e.g., completion date “Jan 2038” represents a failure)
Inadequate support

• Query performances are poor
  • Database not designed to support analysis over millions of tuples
  • Queries could take days even on a dedicated, high-performance server
Inadequate support

- No support for multidimensional analysis
- No support for understanding the causes of problems
- No support for predicting problems
- No support for static and dynamic optimizations
- No support for resource analysis

Workflow Engine

Workflow Audit Logs

(built in or external) Reporting tools

??
Our goal (in the large)

• **Develop a set of techniques and tools that supports business and IT analysts in understanding and improving the quality of their processes**
  
  • Term “quality” used with a very broad meaning
    
    - No a priori, fixed definition
    
    - Service providers are able to define what they consider as “good” or “bad” service (process) execution.

• **Solution: BPI (Business Process Intelligence)**
  
  Cockpit to enable business and IT analysts to extract knowledge hidden in WfMS logs
Our goal (in the small)

• Develop a set of techniques and tools that provides business and IT analysts with the right valuable information about resources
  - Conceive suitable statistics, summary tables, reports
  - "Good" versus "bad", "human" versus "automated", "fair" versus "unfair", "idle" versus "overwhelmed" resources

• Solution: BPI (Business Process Intelligence) Cockpit to enable business and IT analysts to extract knowledge (on resources) hidden in WfMS logs
Approach

• Apply data warehousing and business intelligence techniques to process execution data to:
  • Enable effective, multidimensional analysis of process executions.
  • Analyze causes of “good” and “bad” executions
  • Predict process execution quality degradations, to prevent their occurrence.
  • Suggest and apply static and dynamic optimizations
    – On process definitions, resource assignment, path selection

• It is a very large space. Luckily, we can proceed step by step
  • Provide initial solutions that add value, and progressively extend functionalities.
Other sources

Workflow A
Workflow B
Other sources

Workflow Engine A
Workflow Engine B

Audit Logs
Audit Logs

Quality/cost data, data from other sources
Process definition and execution data

Aggregated data
Quality Criteria
Clean data
Design for performance
Allow OLAP analysis

Happy analyst!
Can still use your favorite tool

BPI Console
Commercially available OLAP/reporting tool

Integrates and cleans data from multiple sources

Warehouse

ETL
Lingering on Resource Rating

- Resource data is seldom clean and “summarized” for analysis purposes:
  - **Analyze** the roles, quality and behavior of resources
  - **Dynamically optimize** resource assignment, to **prevent** their low performance.
  - Derive **relevant resource rate indexes**:
    - On performance, workload and quality/fairness

- Finally, integrate “summarized” tables into the business process intelligence tool and make them available to the other kinds of analysis
Focusing on CE resources....

• *HP Changengine* is an environment for the rapid creation of automated processes within an enterprise.

• It includes a set of components, addressing the definition and execution of processes and the management of resources.

• Resources in HP CE are assigned by the *Resource Executive* (can be persons, adapters, sub-processes and are assigned by resource rules).
CE Process Model

- Process Nodes (*Start, Work, Complete, Route, Abort Nodes*)
- Connectors
- Services
- Resources
CE Process model (by example)
How CE allocates resources.....
CE Audit Logger
CE Logs

Process instances

- ProcInstID
- ProcDefID
- StartTime
- Completed
- CreatorAddr
- ProcessComplete

Activities

- ActInstID
- ProcInstID
- ActInstID
- ActDefID
- Sent
- Finished
- Resource

Case packet data

- ActInstId
- ProcInstID
- TimeStamp
- NameID
- Value

Process Definition

- ProcDefID
- SourceNode
- DestNode
Reporting Screenshot

 PROC started
Reporting Screenshot (graph)
### Reporting Screenshot (table)

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Activated</td>
<td>% completed</td>
<td>Activated</td>
<td>% completed</td>
<td>Activated</td>
<td>% completed</td>
<td>Activated</td>
</tr>
<tr>
<td>0</td>
<td>9</td>
<td>0.89</td>
<td>19</td>
<td>0.84</td>
<td>17</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>0.92</td>
<td>10</td>
<td>0.90</td>
<td>16</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.88</td>
<td>9</td>
<td>0.89</td>
<td>8</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>0.89</td>
<td>18</td>
<td>0.83</td>
<td>74</td>
<td>0.85</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>109</td>
<td>0.84</td>
<td>7</td>
<td>0.71</td>
<td>190</td>
<td>0.89</td>
<td>140</td>
</tr>
<tr>
<td>5</td>
<td>365</td>
<td>0.73</td>
<td>54</td>
<td>0.72</td>
<td>319</td>
<td>0.76</td>
<td>259</td>
</tr>
<tr>
<td>6</td>
<td>391</td>
<td>0.89</td>
<td>95</td>
<td>0.77</td>
<td>788</td>
<td>0.80</td>
<td>502</td>
</tr>
<tr>
<td>7</td>
<td>715</td>
<td>0.83</td>
<td>1167</td>
<td>0.85</td>
<td>714</td>
<td>0.85</td>
<td>382</td>
</tr>
<tr>
<td>8</td>
<td>1138</td>
<td>0.84</td>
<td>1462</td>
<td>0.83</td>
<td>1372</td>
<td>0.76</td>
<td>1783</td>
</tr>
<tr>
<td>9</td>
<td>1149</td>
<td>0.82</td>
<td>1066</td>
<td>0.79</td>
<td>1453</td>
<td>0.74</td>
<td>1421</td>
</tr>
<tr>
<td>10</td>
<td>1077</td>
<td>0.82</td>
<td>1265</td>
<td>0.78</td>
<td>1001</td>
<td>0.79</td>
<td>376</td>
</tr>
<tr>
<td>11</td>
<td>1266</td>
<td>0.77</td>
<td>1108</td>
<td>0.82</td>
<td>1234</td>
<td>0.76</td>
<td>972</td>
</tr>
<tr>
<td>12</td>
<td>974</td>
<td>0.83</td>
<td>1323</td>
<td>0.79</td>
<td>1213</td>
<td>0.82</td>
<td>574</td>
</tr>
<tr>
<td>13</td>
<td>911</td>
<td>0.84</td>
<td>1449</td>
<td>0.80</td>
<td>1026</td>
<td>0.76</td>
<td>735</td>
</tr>
<tr>
<td>14</td>
<td>809</td>
<td>0.77</td>
<td>1058</td>
<td>0.77</td>
<td>664</td>
<td>0.84</td>
<td>1260</td>
</tr>
<tr>
<td>15</td>
<td>1003</td>
<td>0.76</td>
<td>883</td>
<td>0.81</td>
<td>640</td>
<td>0.78</td>
<td>524</td>
</tr>
</tbody>
</table>
### Reporting Screenshot (drill down)

![Oracle Discoverer - TimeAnalysis.DIS](image)

<table>
<thead>
<tr>
<th>Page</th>
<th>Items</th>
<th>Processname: STREAM Co/Springs</th>
<th>NodeName: &lt;All&gt;</th>
<th>ServiceName: &lt;All&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>% completed</td>
<td>#Activated</td>
<td>% completed</td>
</tr>
</tbody>
</table>

### Service activations and their completion

- **% completed**: Percentage of completion.
- **#Activated**: Number of activations.
- **% completed**: Percentage completion for each page item.
Performance evaluation of resources

• History of execution as from the CE Audit Logger Database
• Performance is tied up to the activities and to the “busy periods” of resources
• Performance rate depends on a set of statistics

\[ PR_{\text{Resource,Activity,BusyPeriod}} \propto (\text{Avg#1, Avg#2, StdD#1, StdD#2, SDInt, SDComp}) \]
For each busy period $BP_{(s,e)}$

- Calculate $\text{Average#1} = \text{Avg} \left( \sum_{n=s}^{e} (\text{Actend} - \text{Actstart}) \right)$
- Calculate $\text{Average#2} = (e-s)/n$
- Calculate StdDev on Avg#1 = $\text{StdDev} (\text{Average#1})$
- Calculate StdDev on Avg#2 = $\text{StdDev} (\text{Average#2})$
- Calculate StdDevInt = $\text{StdDev} (\text{Avg}(\sum_{n=s}^{e} (\text{Actend} - \text{Actstart*})))$
- Calculate StdDevCompleted = $\text{StdDev}(\text{nCompleted}/\text{nActivated})$
Human versus Automated Resources

- A *human* resource is generally devoted to the execution of one kind of task, makes a lot of breaks:-) and makes more difference in execution times among holidays and workdays.
- It may also execute tasks in parallel.

- An *automated* resource executes different kinds of tasks, takes the same time to execute instances of the same activity and makes no difference in execution times among holidays and workdays.
- It generally executes tasks in sequence.
A decision tree for H/A classification
Conclusions and future developments

• Examine the associations between resources and flow paths to discover more on resource behavior (quality/fairness, idleness, goodness)
• Envision summary tables of resource behaviors that are to be imported in the BPI Data Warehouse
• Use these information in prediction of business processes future evolution
For more information
