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

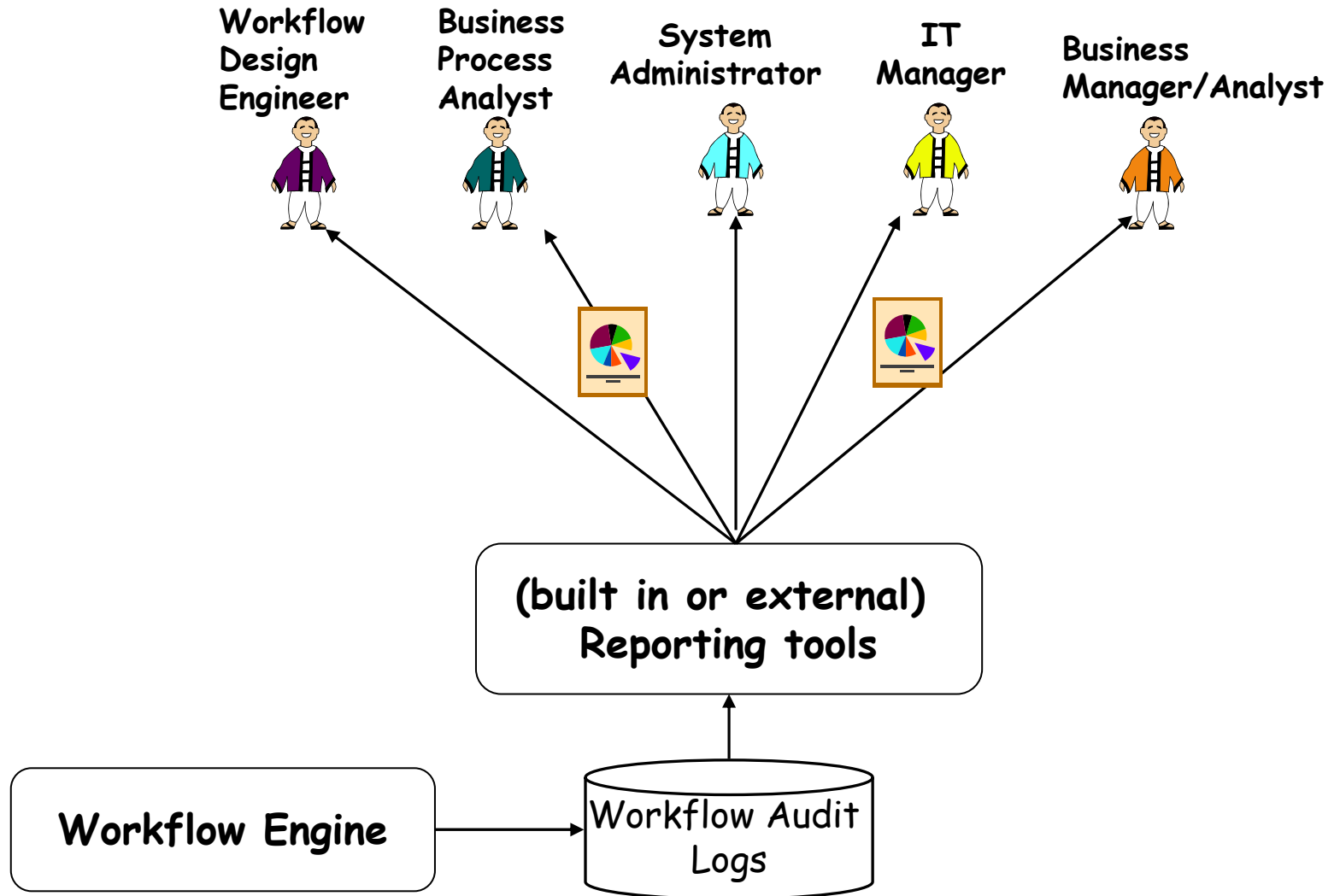- However, more comprehensive analysis is needed!
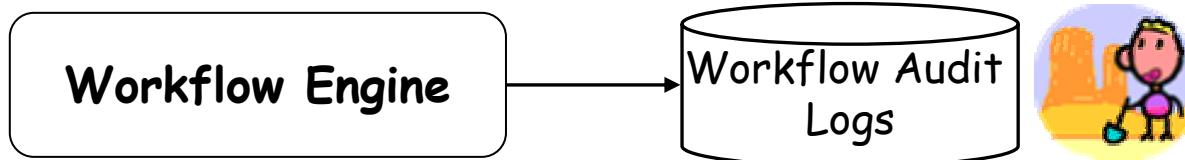
Samples of hidden knowledge

- 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 <u>lots</u> 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?

(built in or external) Reporting tools

Workflow Engine → Workflow Audit Logs
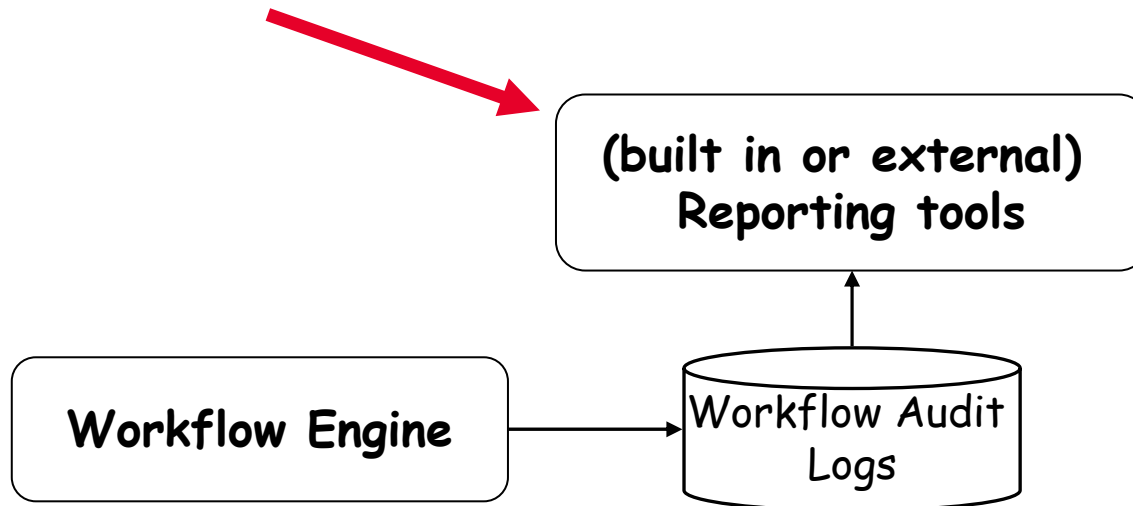
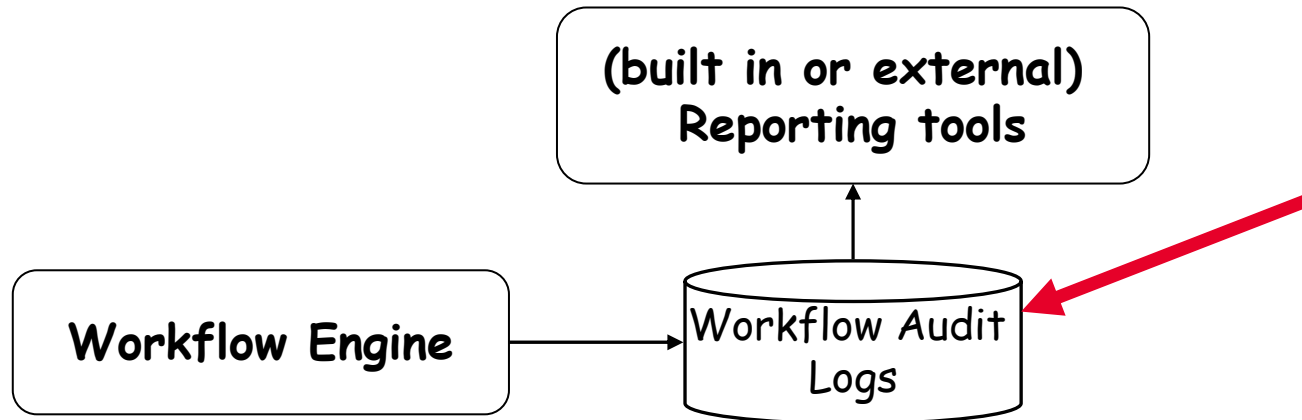# Inadequate support

- **Operates on dirty data**

  - Missing or erroneous values

  - Special codes (e.g., completion date "Jan 2038" represents a failure)

```
                    ┌─────────────────────────┐
                    │  (built in or external) │
                    │     Reporting tools     │
                    └─────────────────────────┘
                                 ▲
                                 │
┌──────────────────┐         ╔═══════════════╗         ⟵
│ Workflow Engine  │────────▶║ Workflow Audit ║
└──────────────────┘         ║     Logs      ║
                             ╚═══════════════╝
```
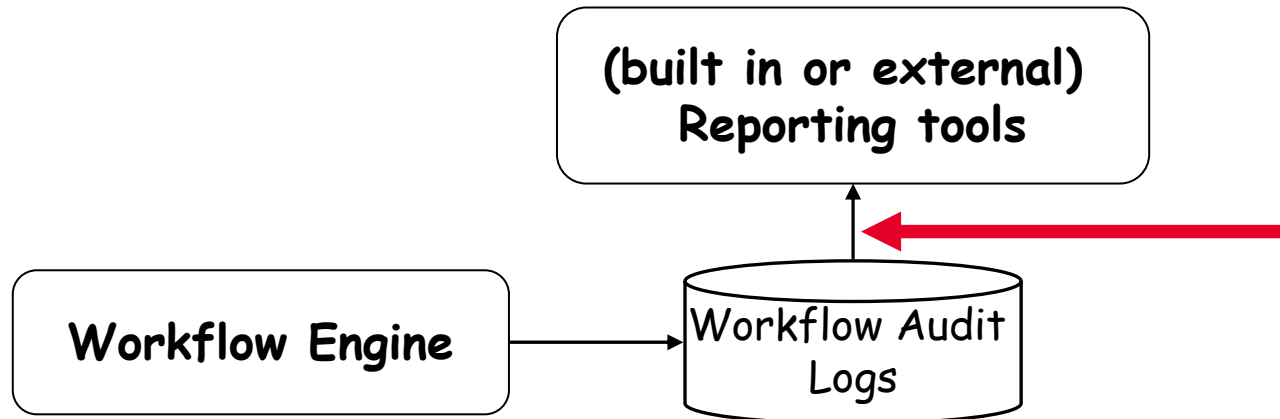
Inadequate support

- 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

??

(built in or external)
Reporting tools

Workflow Engine

Workflow Audit Logs
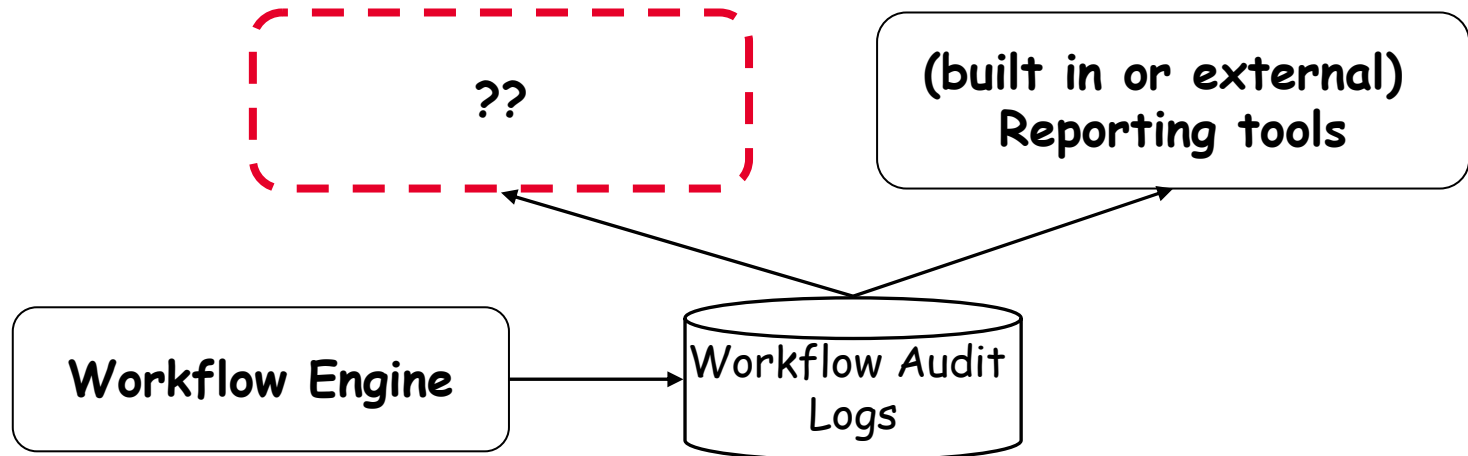
Our goal (in the large)

- 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*
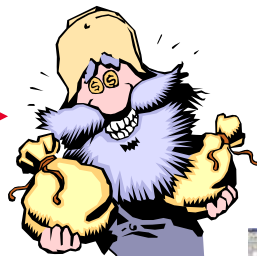
Approach

- 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.

Warehouse

Happy analyst! →

Can still use your favorite tool

BPI Console

Commercially available OLAP/reporting tool
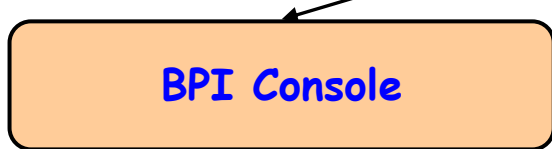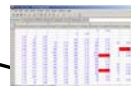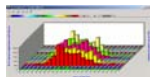
Integrates and cleans data from multiple sources

Workflow Engine A

Workflow Engine B

Workflow A Audit Logs

Workflow B Audit Logs

Other sources

ETL

Quality/cost data, data from other sources

Process definition and execution data
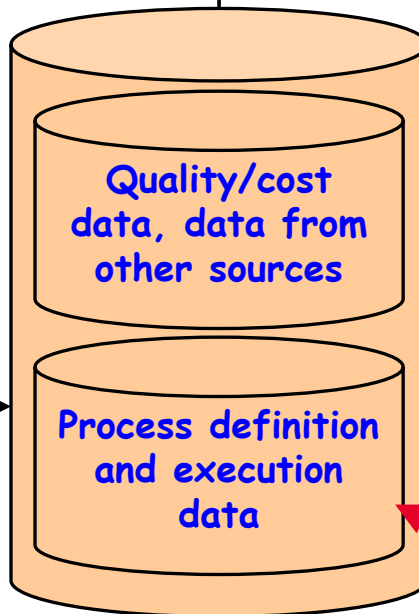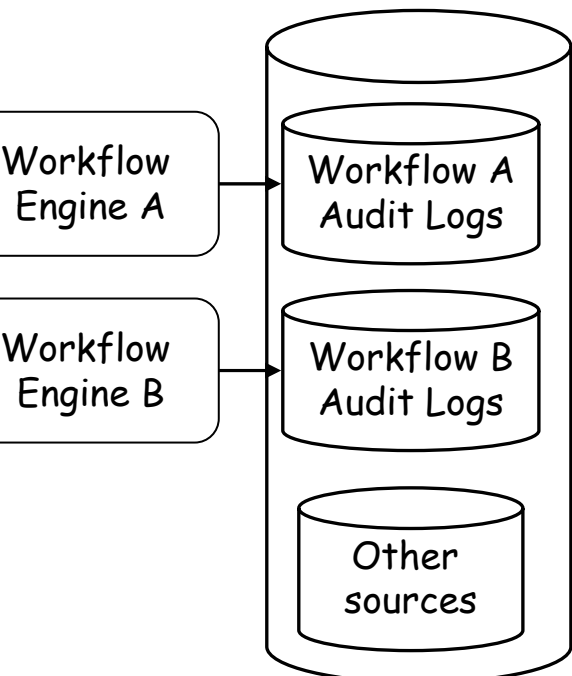
Warehouse

Aggregated data

Quality Criteria

Clean data

Design for performance

Allow OLAP analysis

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

- Rules (*Resource rules, Authorization rules, Domain Rules*)

CE Process model (by example)

File Edit View Format Tools Arrange Window Help

75%

1   2   3   4   5   6   7   8   9   10   11   12   13

AN Join

Approval done?   Send Noaction Email

Submit For Approval

Noaction Timer

Verify Approval Status

Approval/Final Join

Check Approval Status

Send all Approved Email

Update EOP Cancel   Send Cancel Email

Update EOP Failed   Send Failed Email

Update EOP Reject

Send Reject Email

Notify Approver of Work   Approval Join

Get Approval Join

Done

Does Approver Exist?   Add Approver to System

Update EOP Approve   Notify Requester of Approval

Get next Approver

Check for Approver Existence

Set Account Name

Notify Requester of Initiation

Initiate   Validate Form Data

Notify Support of Bad Form

Layer 1

List

How CE allocates resources…..
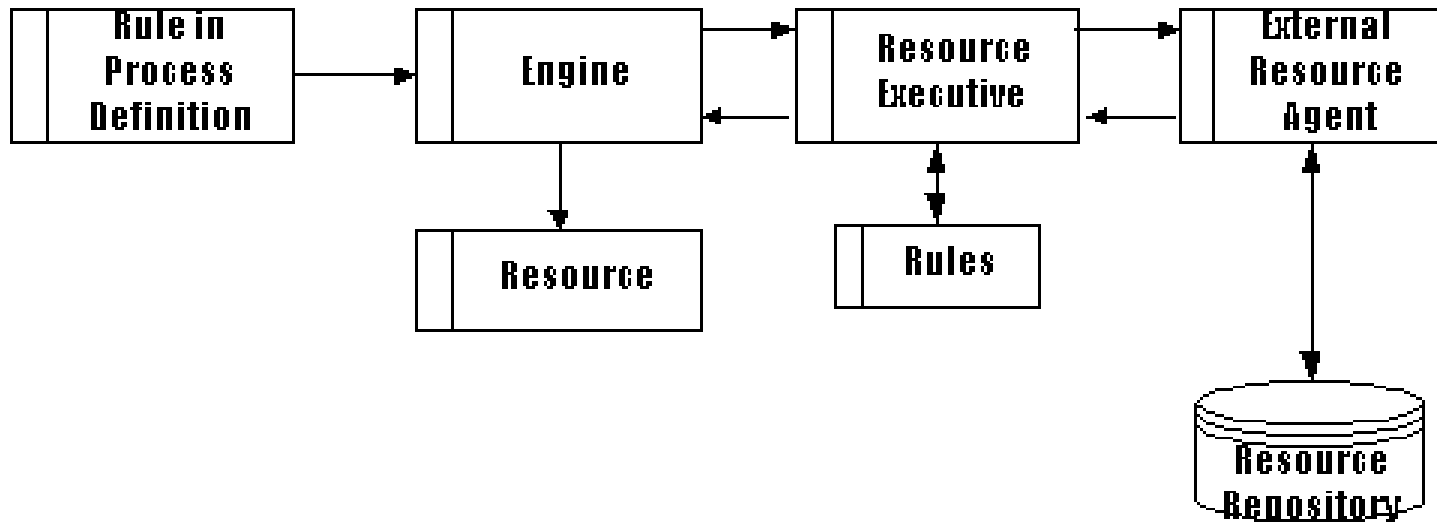
CE Audit Logger

CE Logs

Activities

ActInstID
ProcInstID
ActInstID
ActDefID
Sent
Finished
Resource

Case packet data

ActInstId
ProcInstID
TimeStamp
NameID
Value
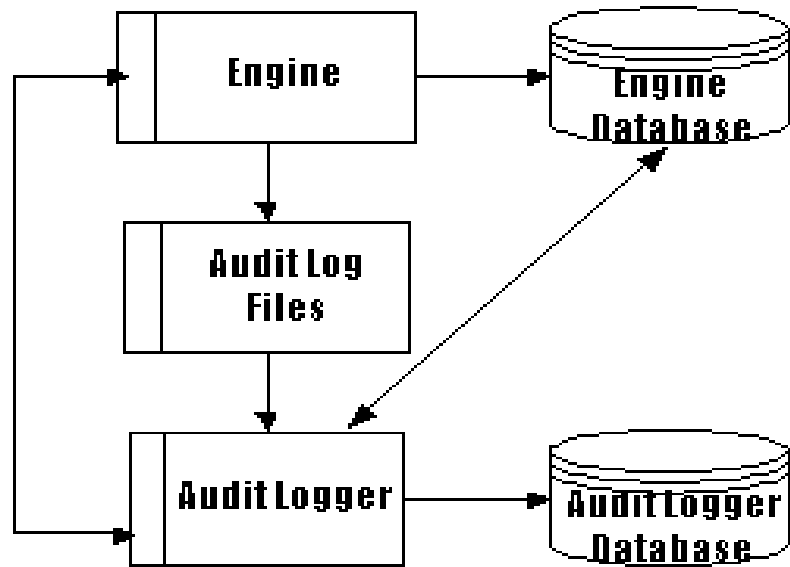
Process instances

ProcInstID
ProcDefID
StartTime
Completed
CreatorAddr
ProcessComplete

Process Definition

ProcDefID
SourceNode
DestNode

Reporting Screenshot

Reporting Screenshot (graph)
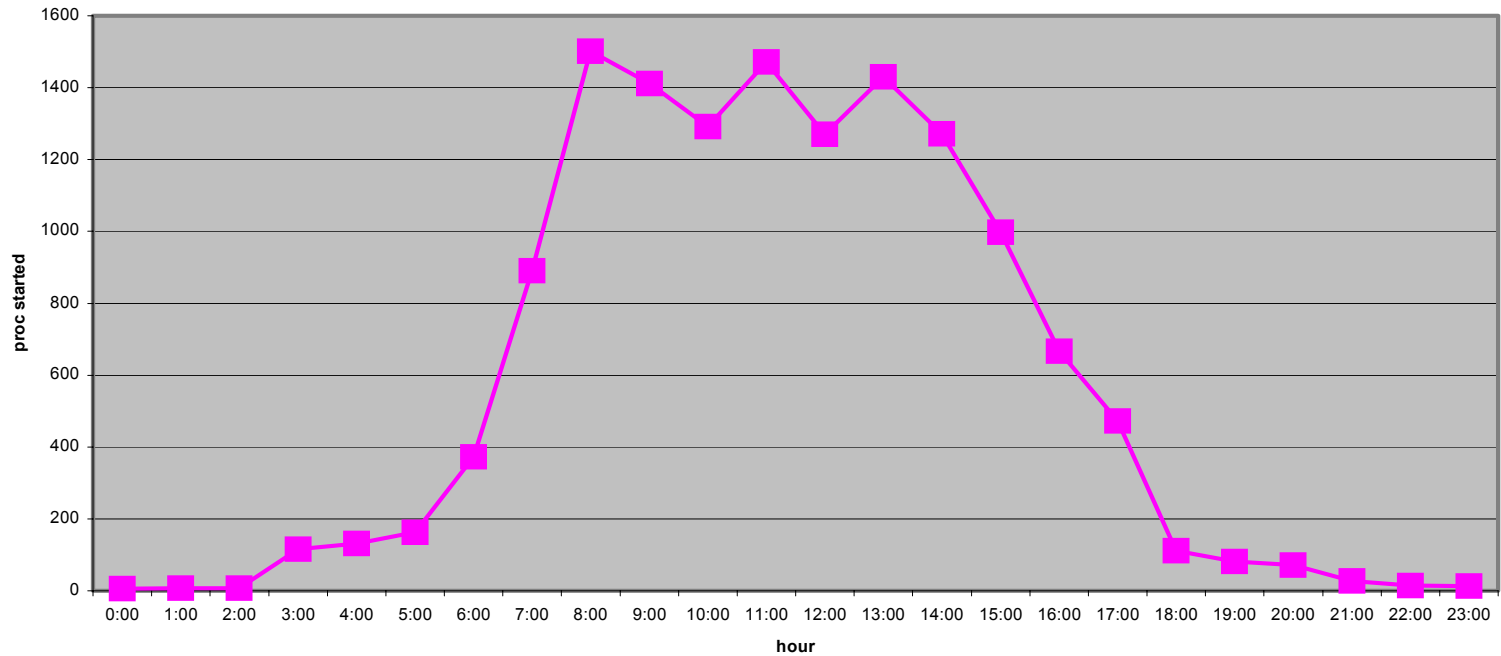
Reporting Screenshot (table)

Oracle Discoverer - [TimeAnalysis.DIS]

File   Edit   View   Sheet   Format   Tools   Graph   Window   Help

Page Items: Processname: <All> ▼   Nodename: <All> ▼   Servicename: <All> ▼

	3		4		5		6		7	
	# Activated	% completed	# Activated	% completed	# Activated	% completed	# Activated	% completed	# Activated	% completed
0	9	0.89			19	0.84	17	0.88		
1	12	0.92			10	0.90	16	0.94		
2	8	0.88			9	0.89	8	0.88		
3	27	0.89	18	0.83	74	0.85	14	0.86		
4	109	0.84	7	0.71	190	0.89	140	0.90		
5	365	0.78	54	0.72	319	0.76	259	0.78	14	0.86
6	391	0.89	95	0.77	788	0.80	502	0.80		
7	715	0.83	1167	0.85	714	0.85	852	0.85	3	0.67
8	1138	0.84	1462	0.83	1372	0.76	1783	0.75	15	0.87
9	1149	0.82	1066	0.79	1458	0.74	1421	0.75	22	0.82
10	1077	0.82	1265	0.78	1001	0.79	876	0.85	12	0.92
11	1266	0.77	1108	0.82	1234	0.76	972	0.79		
12	974	0.83	1323	0.79	1213	0.82	574	0.88	35	0.77
13	911	0.84	1449	0.80	1026	0.76	735	0.82	29	0.76
14	809	0.77	1058	0.77	664	0.84	1260	0.72	26	0.85
15	1003	0.76	883	0.81	640	0.78	524	0.83	15	0.87

Service activations and their completion rate   Sheet

Reporting Screenshot (drill down)

Oracle Discoverer - [TimeAnalysis.DIS]

File Edit View Sheet Format Tools Graph Window Help

Page Items: Processname: STREAM ColSprings ▼ Nodename: <All> ▼ Servicename: <All> ▼

|		3		4		5		6		7	
		% completed	# Activated	% completed	# Activated	% completed	# Activated	% completed	# Activated	% completed	# Activated	% completed
0								3	1.00		
2						3	1.00				
3		1.00	3	1.00							
4		1.00	81	0.84			145	0.90	76	0.87		
5		0.93	288	0.74	26	0.77	248	0.73	149	0.71	12	0.83
6		0.82	193	0.90	63	0.73	538	0.77	336	0.74		
7		0.83	418	0.82	535	0.81	392	0.79	471	0.78	3	0.67
8		0.75	668	0.80	1072	0.80	802	0.74	1244	0.71	6	0.83
9		0.80	729	0.79	500	0.76	835	0.71	1048	0.69	6	0.83
10		0.71	418	0.88	738	0.75	622	0.73	352	0.78		
11		0.78	727	0.76	645	0.79	824	0.71	585	0.77		
12		0.81	478	0.80	614	0.77	516	0.78	336	0.85	35	0.77
13		0.72	473	0.81	656	0.86	432	0.76	303	0.72	28	0.75
14		0.75	474	0.76	478	0.78	247	0.77	792	0.68	13	0.85
15		0.83	574	0.76	353	0.74	190	0.83	218	0.77	5	0.80
16		0.76	303	0.81	173	0.77	62	0.82	83	0.83		
17		0.75	74	0.76	20	0.85	75	0.83	13	0.85	12	1.00
18		0.74	35	0.86	70	0.80	27	0.85				

Service activations and their comple

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

$$PR_{Resource,Activity,BusyPeriod} \propto (Avg\#1, Avg\#2, StdD\#1, StdD\#2, SDInt, SDComp)$$

Busy periods

Busy period #1
(resource X, activity Y)

Busy period #2
(resource X, activity Y)
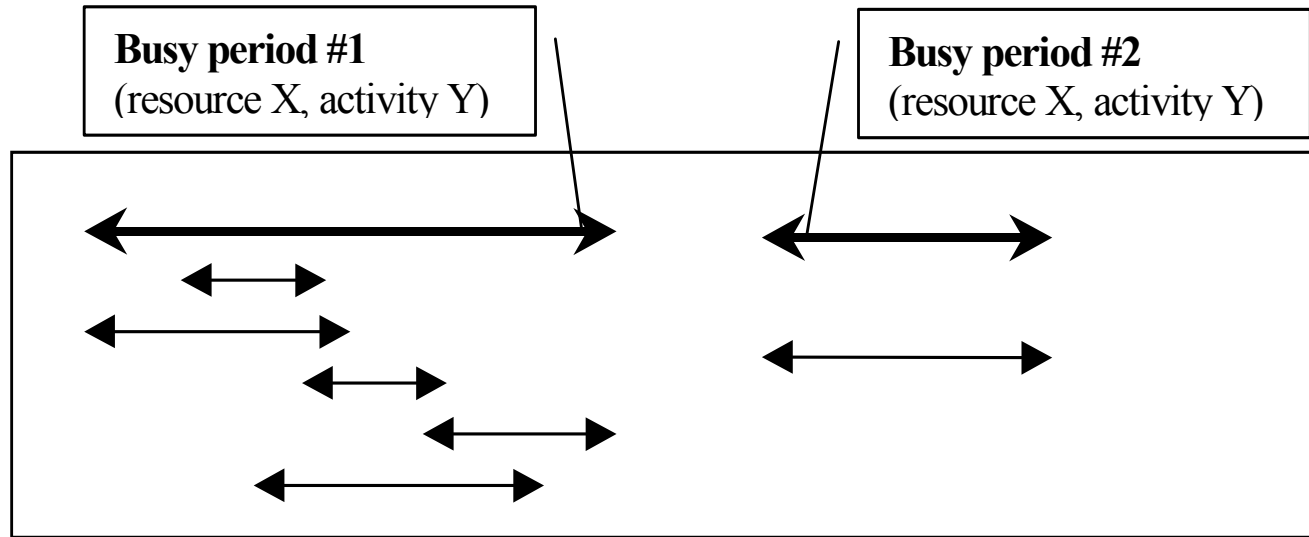
For each busy period $BP_{(s,e)}$

    Calculate $Average\#1 = Avg(\sum_{n=s}^{e}(Actend - Actstart))$

    Calculate $Average\#2 = (e\text{-}s)/n$

    Calculate $StdDev$ *on* $Avg\#1 = StdDev(Average\#1)$

   *Calculate* $StdDev$ *on* $Avg\#2 = StdDev_e(Average\#2)$

    Calculate $StdDevInt = StdDev(Avg(\sum_{n=s}^{e}(Actend - Actstart^{*})))$

    Calculate $StdDevCompleted = StdDev(nCompleted/nActivated)$

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
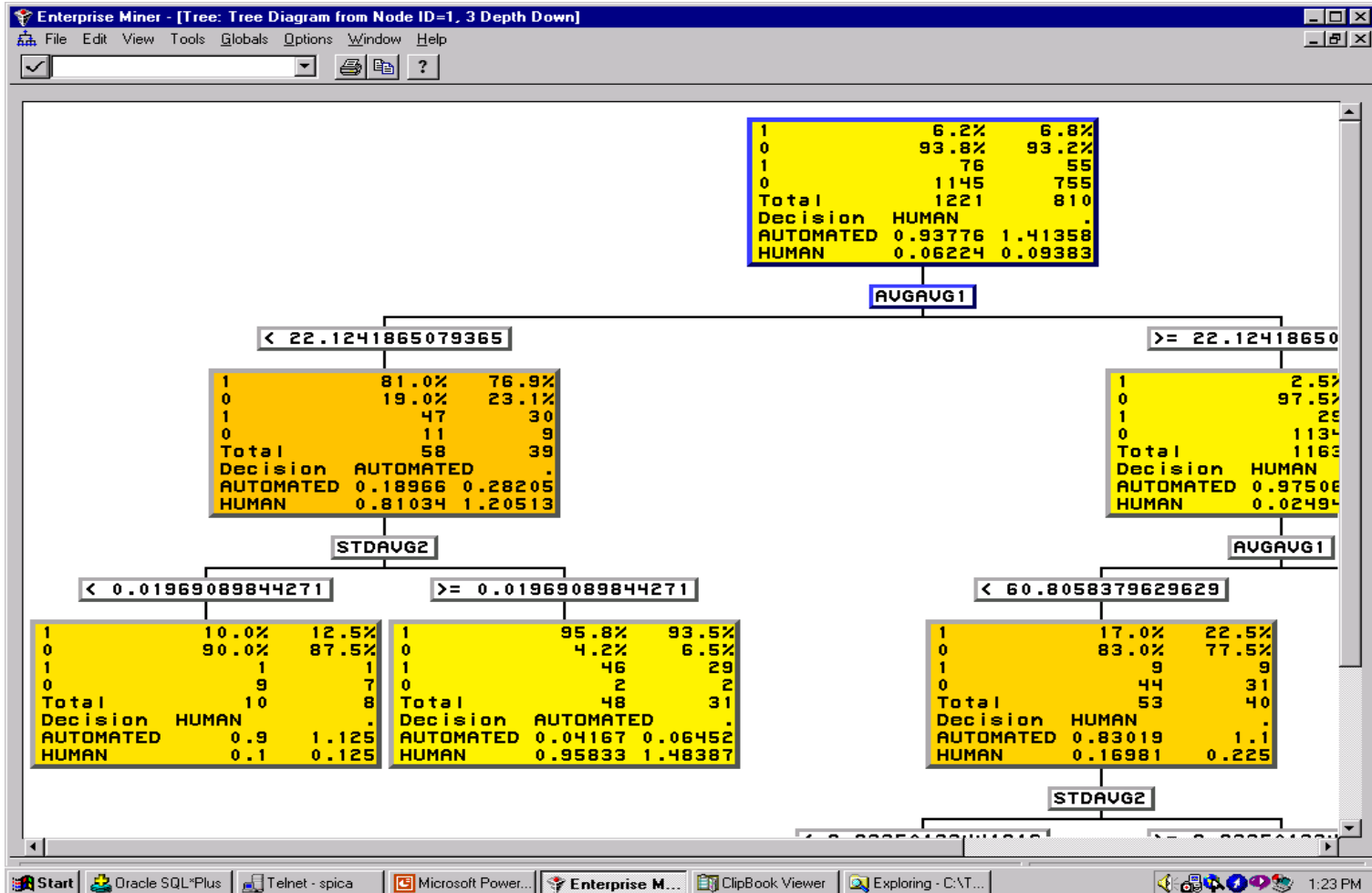
Conclusions and future developments

- 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

- A. Bonifati, F. Casati, U. Dayal, M. Chan "Warehousing Workflow Data: Challenges and Opportunities", VLDB 2001, Rome, September 2001.

- A. Bonifati "Resource Analysis and Classification in Business Processes" TR #2001-48, DEI, Politecnico di Milano.