



Enabling Grids for E-scienceE

Design of an Expert System for Enhancing Grid Fault Detection based on Grid Monitoring Data

Gerhild Maier

March 2nd 2008

www.eu-egee.org



Information Society
and Media



- problem description

- approach
 - association rule mining
 - design of an expert system

- current status, example

- outlook and summary

- Dashboard database: a lot of information about jobs
- Dashboard monitoring tools: find faulty Grid components
- exit codes

- detect error source underneath the exit codes
- fast, to solve problems quickly

- automatization

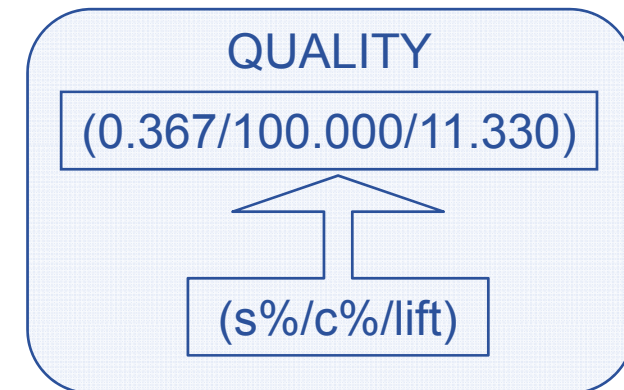
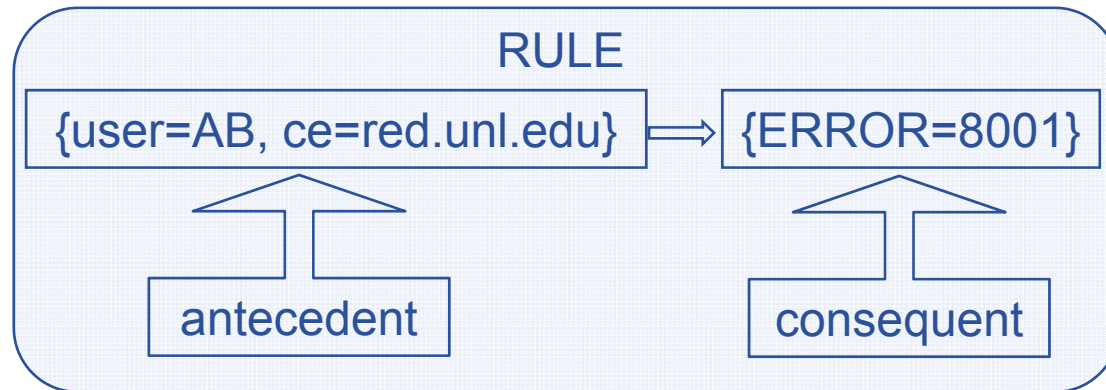
- combine machine created knowledge with human knowledge to an expert system

AR

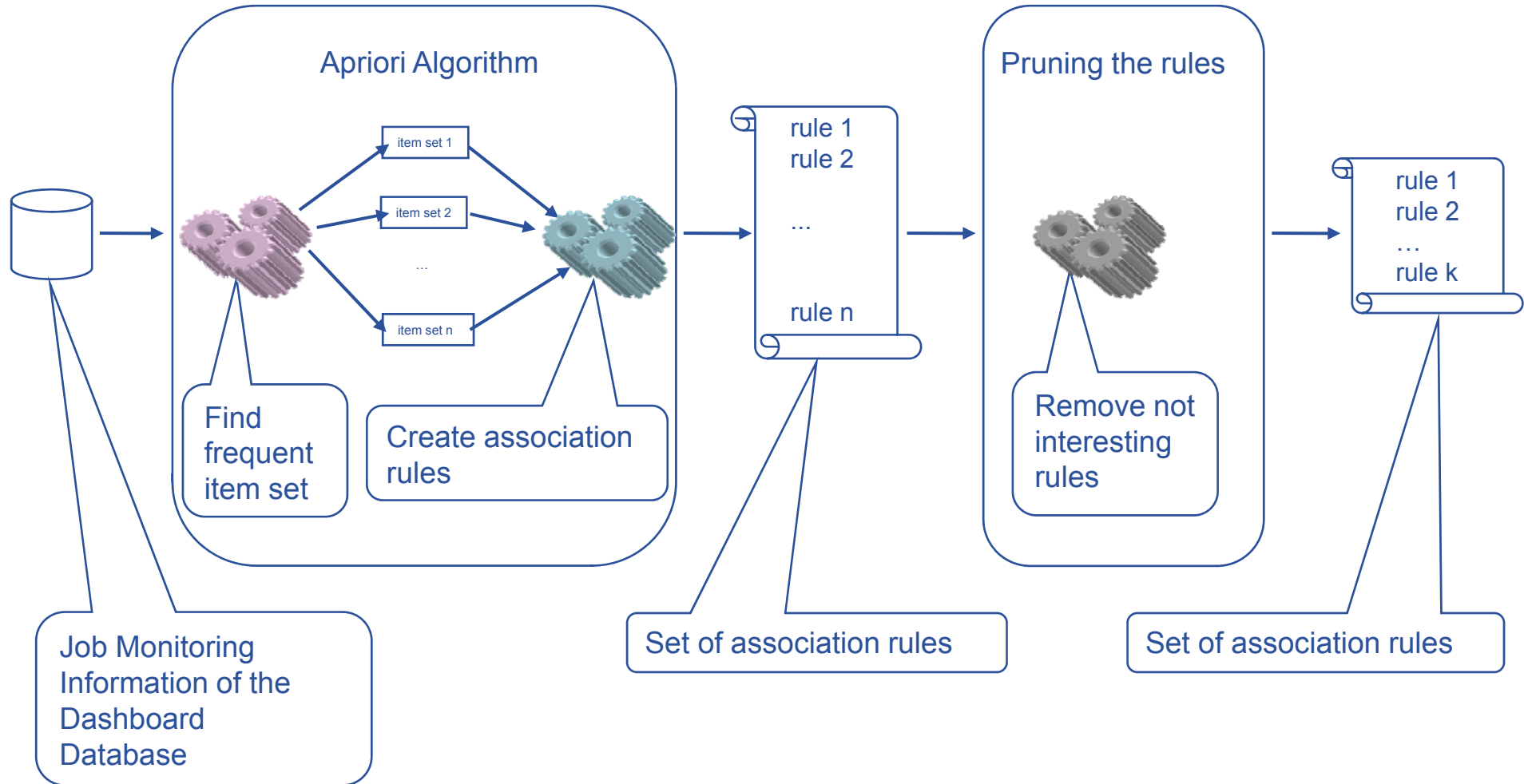
HK

ES

- **Association Rule Mining on Monitoring Data**
(machine created knowledge)
- **Human Knowledge**
(the rule interpretation)
- **Expert System**

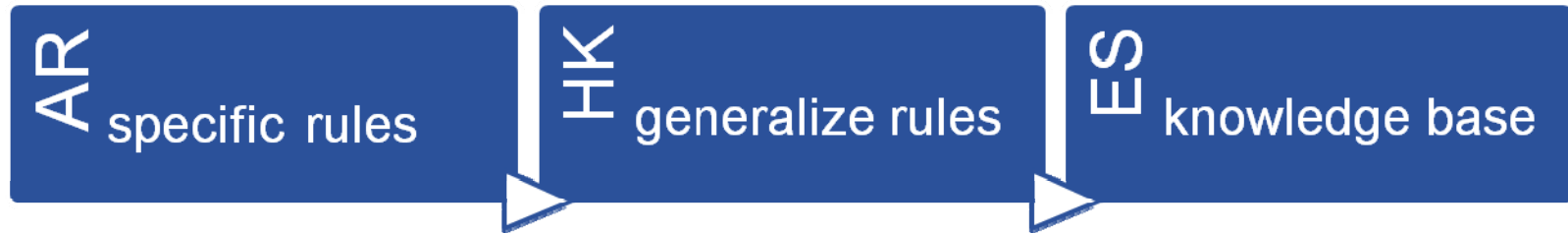


- rule: set of items
- item: attribute-value pair
- support: s% of the data includes all items
- confidence: c% of the data including the antecedent also include the consequent
- lift: measurement of interestingness
 $lift = support(AB) / (support(A) * support(B))$



- a program solving problems like an expert
- example: decision support system to detect a problem
 - ...
 - Did you plug in the printer? → yes
 - Did you install a driver? → no
 - ...
- 2 components:
 1. knowledge base: collection of human expert knowledge in a problem domain
 2. inference engine: defines how to use the knowledge

building the ES



using the ES



maintaining the ES



- QAOES = Quick Analysis Of Error Sources



QAOES

[PLEASE SEND ME YOUR COMMENTS!](#)

The page will be refreshed every hour.

Timerange:	<input type="text" value="last 12 hours"/>	Description: jobs in the selected time range are the data set input for the data mining process.
Min number of jobs:	<input type="text" value="100"/>	Description: minimum number of jobs, to list a created rule. Low number results in long calculation time!
Confidence:	<input type="text" value="80"/>	Description: confidence% of jobs in the input data set that contain the attributes in the antecedent also contain the consequent attribute.

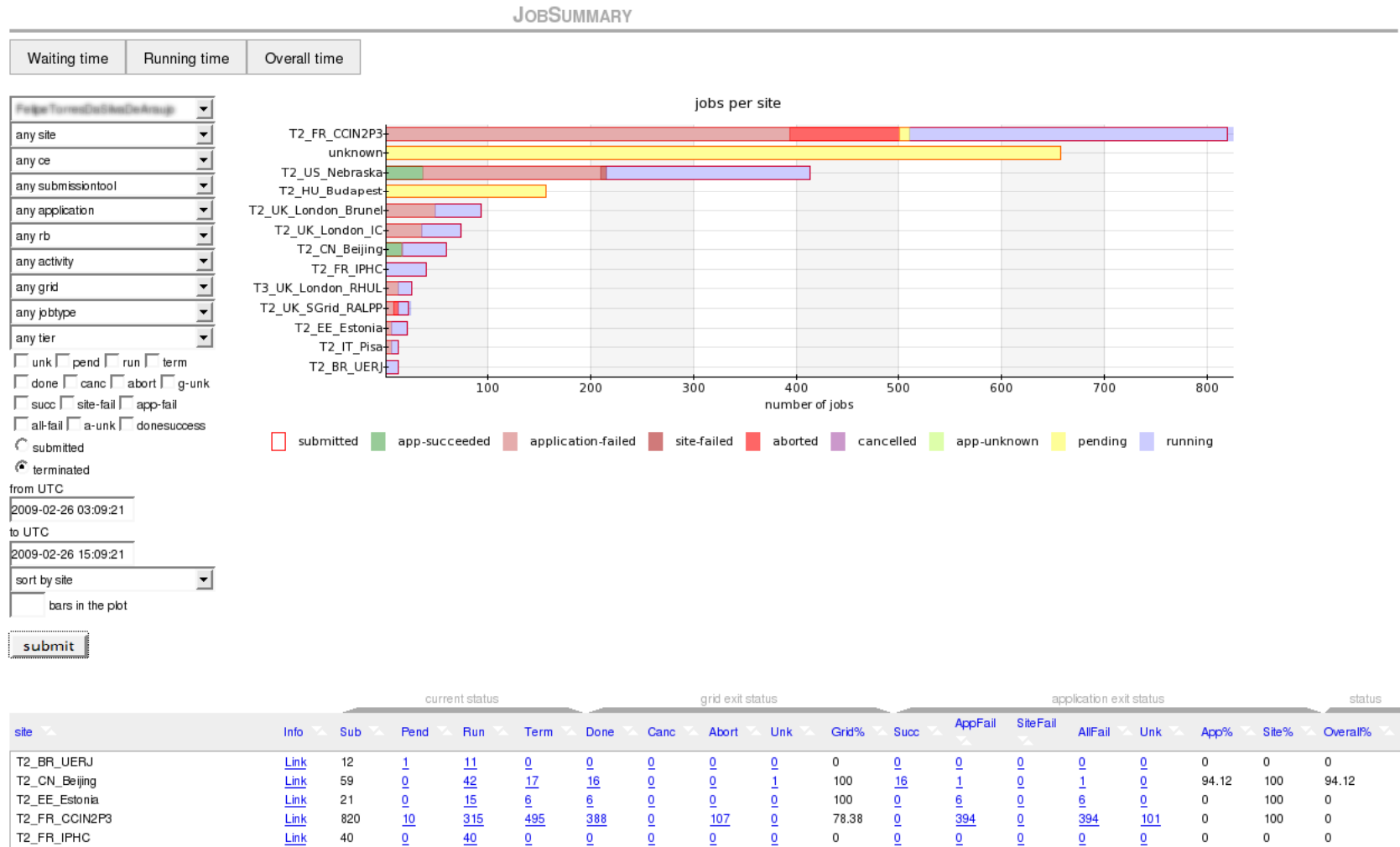
Show Rules

- background: 8 job attributes
 - site, ce, queue, worker node
 - dataset
 - user, application
 - exit code

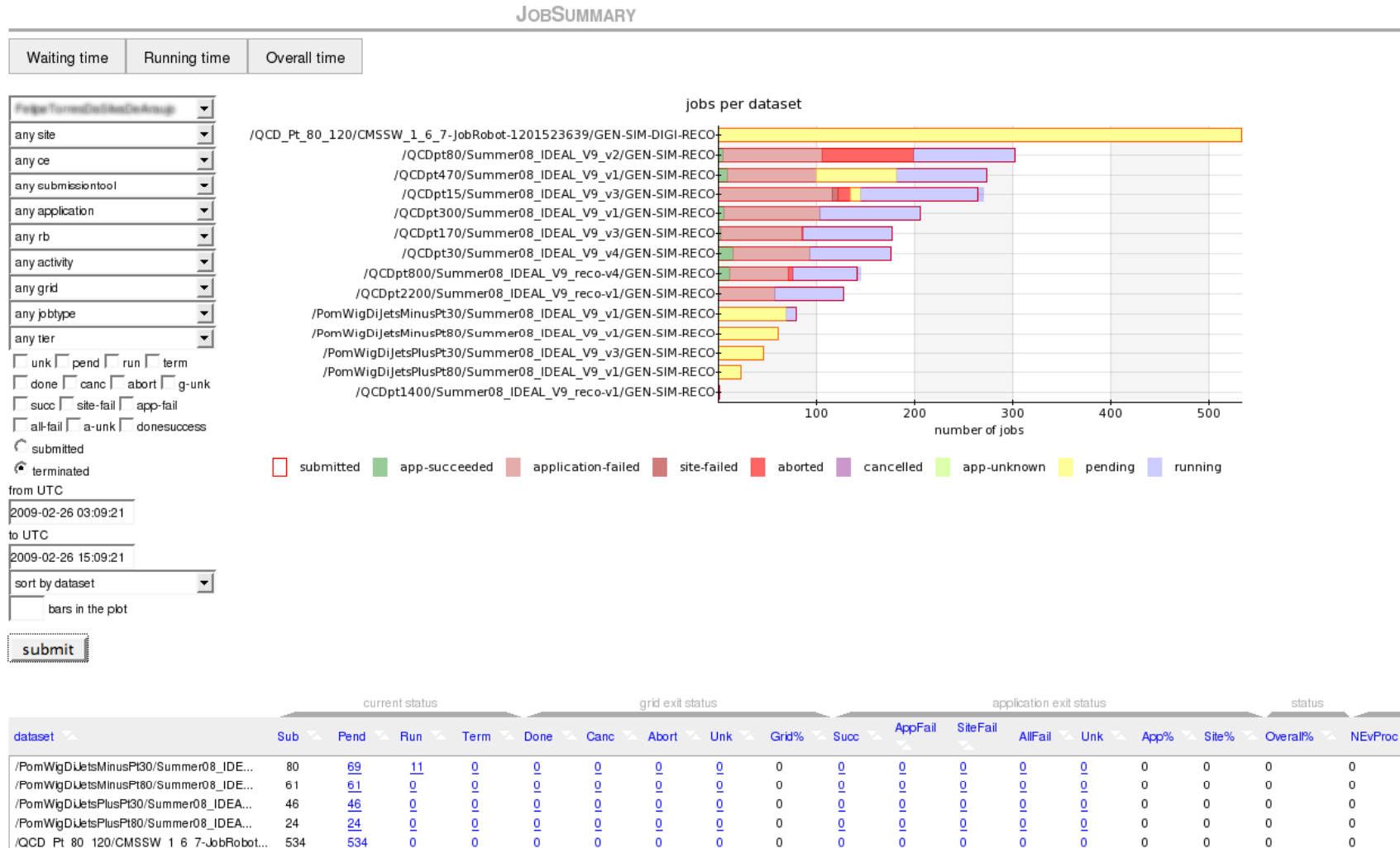
- list of rules, with quality measures: support, confidence, lift
- association rules:
 - interesting dependencies of job attributes
 - unusual patterns in the dataset
- link to dashboard job summary page
- example:
 - CMS analysis jobs from 12 hours: 30761
 - min 100 jobs => support = 0.26 %
 - confidence: 90%
 - runtime: 5 min
 - number of rules: 7

Link to jobSummary	Antecedent	Consequent	Support in % /number of jobs	Confidence in %	Lift
click!	user= RobertSchoeffbeck dataset=/QCDpt170/...	ERROR=8009	0.410/124	91.912	160.068
click!	user= FelipeTorresDaSilvaDeAraujo ce=cclcgceli05.in2p3.fr queue=cclcgceli0... site=T2_FR_CCIN2P3	ERROR=60307	1.273/388	98.477	27.790
 click!	user= FelipeTorresDaSilvaDeAraujo	ERROR=60307	2.189/666	89.892	25.367
click!	dataset=/QCDpt80/S...	ERROR=60307	0.331/100	88.596	25.001
click!	dataset=/QCDpt15/S...	ERROR=60307	0.335/101	83.607	23.593
click!	user= Aarabta site=T2_RU_SINP dataset=/QCD_pt_0_...	ERROR=8001	0.538/163	82.000	14.564

one user has problems on different sites



one user has problems with different datasets



- user has problems on different site, with different datasets
→ problem in his code?
- exit code 60xxx → stage out problem
→ problem with the storage element?
- ...
- ...
- collection of rule interpretations
- rule generalization
- input for the knowledge base

- continuous adaptation of the association rule mining parameters
- building the knowledge base
- development of the inference engine

- building the **Expert System**



- **A**ssociation **R**ule mining completed
- collecting **H**uman **K**nowledge
- web interface currently deployed for analysing CMS analysis jobs
- QAOES easy to adapt to different VOs job data

- **QAOES:**
<http://dashb-cms-mining-devel.cern.ch/dashboard/request.py/rules>
- **Twiki:**
<https://twiki.cern.ch/twiki//bin/view/ArdaGrid/AutomaticFaultDetection>
- **Association Rule Mining:**
article: Mining Association Rules between Sets of Items in Large Databases, Agrawal R, Imielinski T, Swami AN.
- **Pruning Association Rules:**
article: Efficient Statistical Pruning of Association Rules, Alan Ableson, Janice Glasgow
- **Expert Systems:**
book: Introduction to Expert Systems, Peter Jackson