GridMiner: Design and underlying grid technology

Ivan Janciak

University of Vienna
Institute for Software Science

email: janciak@par.univie.ac.at

Outline

- Architecture
  - Motivation/Use Case
- Grid Services
  - Service description
  - Implementation
- Data sources
  - Mediator
- Knowledge Discovery Workflow
- GridMiner Application Demo

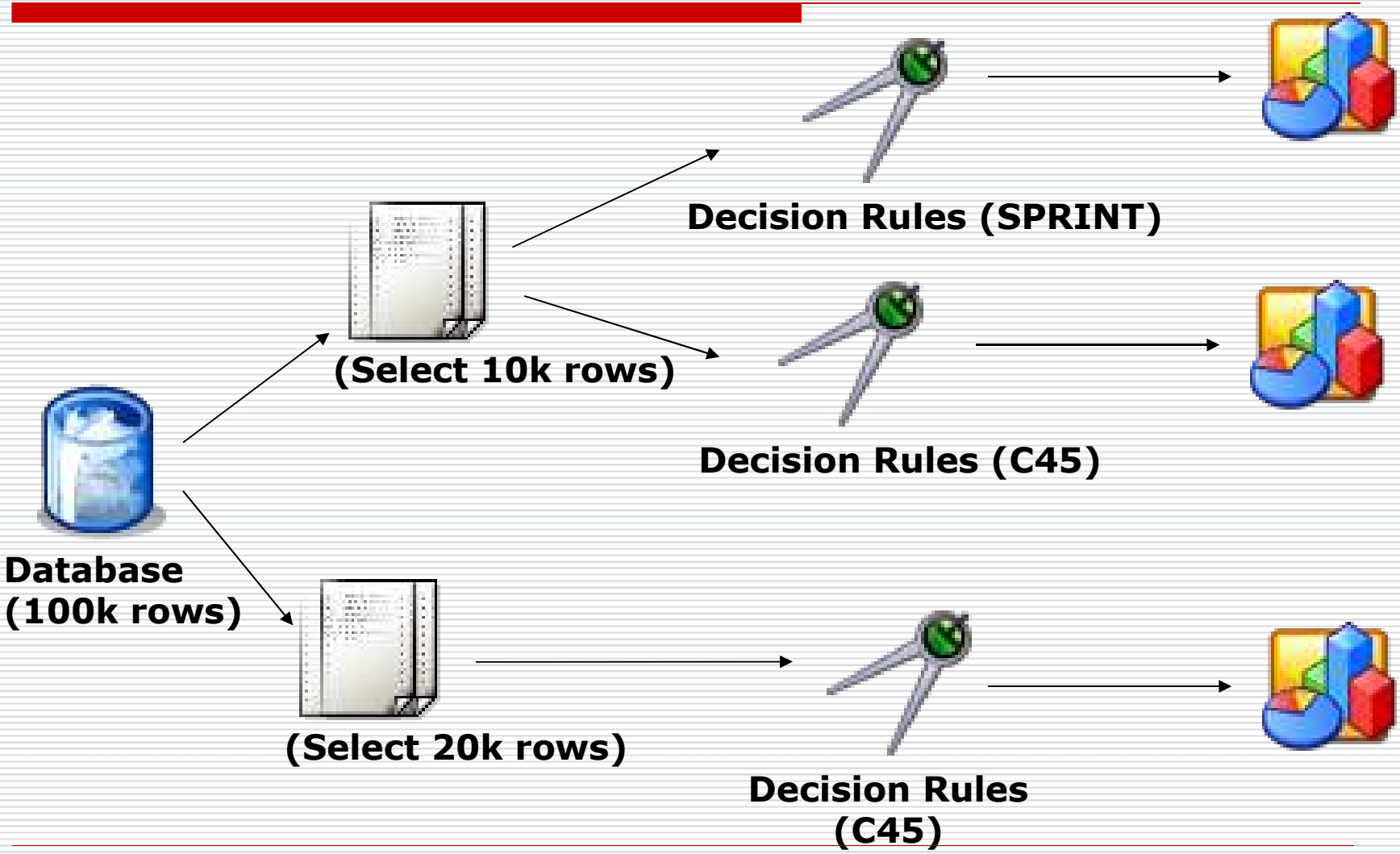
Motivation

- Grid Based data mining tool
- Open/Expandable
- Easy to use

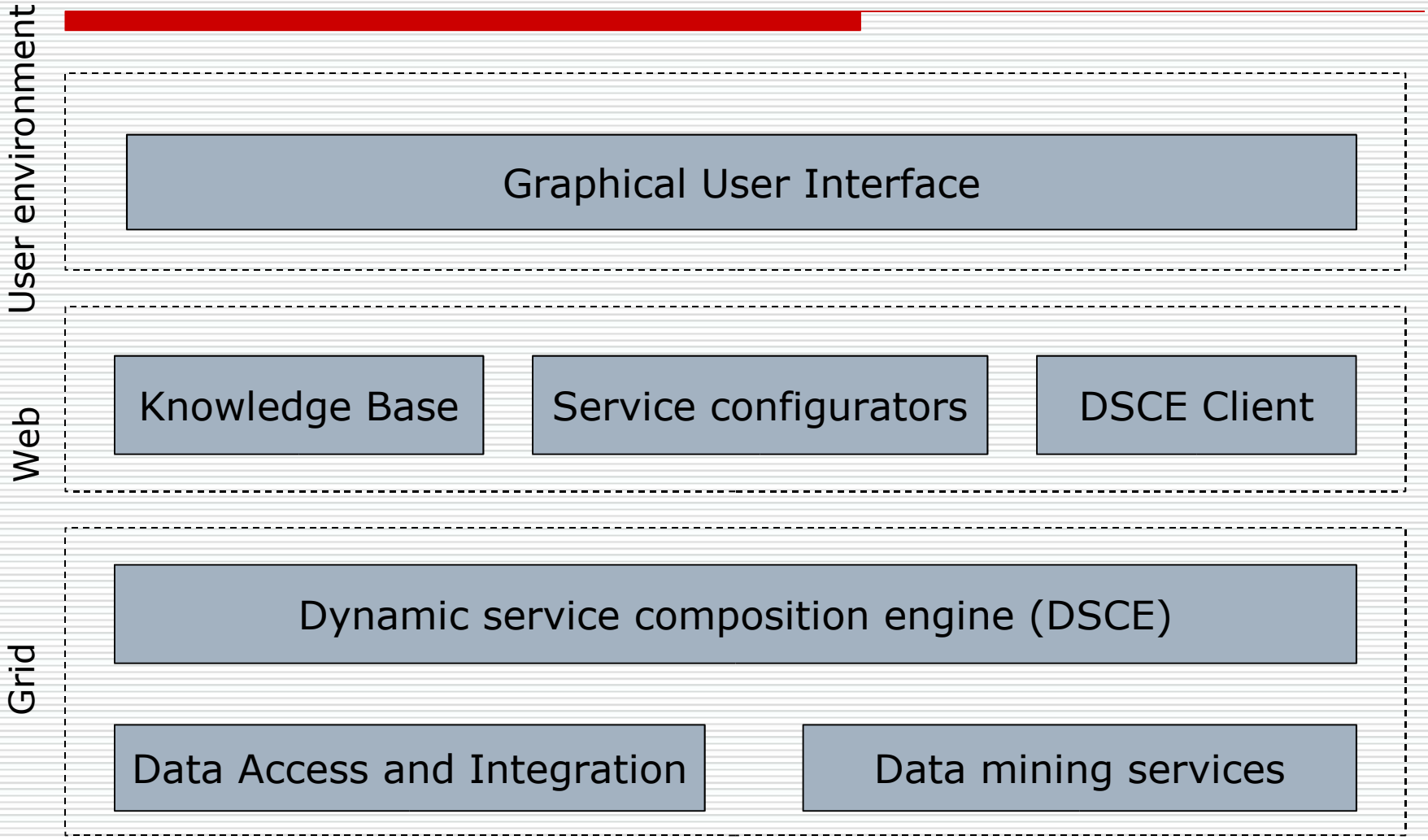
Requirements

- Ability to access and analyze a huge amount of information – typically heterogeneous and geographically distributed
- Intelligent behavior ability to maintain, discover, extend, present and communicate knowledge
- High performance (real-time or soft real-time) query processing
- High security guarantee

Simple Use Case



GridMiner Architecture



Implementation/Technology

- Globus 3.2
- OGSA/DAI ver 3
- GUI – Workflow constructions/Results visualization (JGraph, Java web Start, Java server pages)
- Service Configurators (Java server pages)
- Workflow management – DSCE Client (OGSA)
- Knowledge base – Configurations (XML,OWL)
- Data mediation service (OGSA-DAIS activity)

Implemented Services

- Data mining services
  - Sequences (SPADE)
  - Clustering (SimpleKMeans)
  - Decision rules (SPRINT)
  - OLAP (sequential/parallel version)
  - Association rules on OLAP
- Dynamic Workflow Composition Service
- Mediator (Activity in DAIS ver5)

Service Description

- GWSDL
- Semantic Description (OWL-S)
  - Activity ontology
  - Data mining ontology

- Service Configurator
  - Dynamic service composition language
    - variables/composition

Datasources

- WebRowSet (XML file) produced by OGSA-DAI
- Data Streams

Metadata

- WebRowSet
- Predictive Model Markup Language (PMML)
- Mediating schema
- Datasource ontology

Knowledge Discovery process

1. Data preparation/integration

2. Setting data mining task

3. Data mining task execution

4. Results visualization

Grid service execution

1.

webRowSet, number of cluster

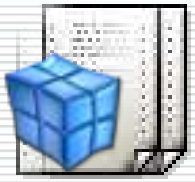
PMML

Clustering
- Create service
- Execute method (process clustering)
- Query SDE
- Destroy service

2.

PMML, webRowSet

GSH



OLAP
- Create service
- Execute method (build cube)

OLAP Query
- Execute query

Components interaction