



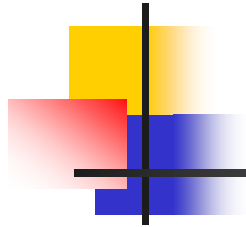
Výpočetní gridy, 2008

Z různých materiálů sestavil
L. Pešička



Obsah

- Motivace gridu
- Projekty distrib. počítání
- Architektura gridových služeb
- OGSA, OGSI, Globus Toolkit
- EGEE Grid
- Metacentrum



Motivace – proč gridy?



Reálné využití výpočetních zdrojů - cluster

- **výkonný cluster**

- periody intenzivního využití x idle
- nárazová potřeba výkonu CPU

- cena pořízení
- náklady provozní infrastruktury (energie, klimatizace)
- náklady na správu



Reálné využití výpočetních zdrojů - pracovní stanice

- **pracovní stanice**

- pořizovaná konfigurace (delší období)
- výkon potřebný pro běžnou agendu
- idle time
 - mimopracovní doba
 - oběd
 - čekání na interakci uživatele (např. word)
 - "screen savers"



Pronájem potřebného výkonu

- **zapůjčení** strojového času clusteru
- **od koho** – spřátelené organizace (**omezený** počet)
- často potřebují výkon ve stejné době
 - organizace stejného charakteru, podobné termíny..
- **administrativa**
 - přístupová konta (security policy)
 - metody přístupu, vytížení sítě, přenos velkých souborů
 - monitorování průběhu výpočtu



Využití CPU pracovních stanic

- Služba na pracovních stanicích, v době nečinnosti CPU provádí výpočet dané úlohy
- Příjem dávky
- Zpracování
- Odeslání výsledků

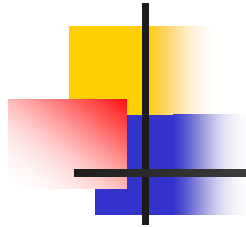
- Seti@home, World Community Grid
- vnitrofiremní použití – např. i Intel
 - data, výpočty, výsledky neopustí firmu



Sdružení výpočetních prostředků – virtuální organizace (VO)

- začlenění (výpočetních) prostředků do zdrojů sdílených v rámci **VO**
- definování politiky přístupu k prostředkům

- úloha do fronty
- **resource broker** rozhodne o vhodném **CE** pro danou úlohu
- řeší otázku bezpečnosti, přístupových práv
- větší množství organizací



Oblíbené projekty distribuovaného počítání



Distribuované počítání

- **SETI@HOME**
 - <http://setiweb.ssl.berkeley.edu/>
 - *původně SETI@home/Classic*
 - *od 15.12.05 SETI@home/BOINC*
- **BOINC** – Berkeley Open Infrastructure for Network Computing
 - <http://www.boinc.cz/>
- **World Community Grid**
 - <http://www.worldcommunitygrid.org/>
 - United Devices client OR Boinc



World Community Grid

- *wcg_boinc_5.10.30_windows_intelx86.exe*
(9 408 KB)
použit k instalaci (3.12.2007)
- Linux (x86)
- Apple Mac (PowerPC, x86)
- United Devices for Windows (Vista, XP, 2000,..)



Setup Type

Choose the setup type that best suits your needs.



Please select a setup type.

Single-User Installation

BOINC runs only when you're logged on, and only you can manage BOINC (recommended).

Shared Installation

BOINC runs whenever anyone is logged on, and anyone can manage BOINC.

Service Installation

BOINC runs even when no one is logged on. Only you can manage BOINC.
The screensaver does NOT work in this configuration.
The show graphics feature does NOT work in this configuration.

InstallShield

< Back

Next >

Cancel



World Community Grid - BOINC Agent - InstallShield Wizard



Single-User Installation Configuration

Allows you to configure the settings for the single-user installation configuration



Please select which options you would like to enable.

- Make BOINC your default screensaver.
- Launch BOINC when logging on.


InstallShield

< Back


Next >

Cancel

World Community Grid - BOINC Client




world community grid. Powered by IBM.
technology solving problems




Výpočty pozastaveny: Probíhají testy procesoru.

My Projects: [Add Project](#) ?




[Messages](#) | [Snooze](#) | [Preferences](#) | [Advanced View](#)

World Community Grid - BOINC Client



world community grid. Powered by IBM.
technology solving problems



Stahuji práci ze serveru.

My Projects: [Add Project](#) ?



[Messages](#) | [Snooze](#) | [Preferences](#) | [Advanced View](#)

World Community Grid - BOINC Client

 world community grid. Powered by IBM.
technology solving problems

● dddt

World Community Grid

Application: Discovering Dengue Drugs - Together



Running Graphics Available

Elapsed Time: 0 h 0 m 17 s
Time Remaining: 1 h 29 m 18 s

0,0 %

My Projects: [Add Project](#) ?



[Messages](#) | [Snooze](#) | [Preferences](#) | [Advanced View](#)

World Community Grid - BOINC Client

 world community grid. Powered by IBM.
technology solving problems

● dddt

World Community Grid

Application: Discovering Dengue Drugs - Together



Running Graphics Available

Elapsed Time: 0 h 0 m 22 s
Time Remaining: 1 h 29 m 18 s

0,0 %

My Projects: [Add Project](#) ?



[Messages](#) | [Snooze](#) | [Preferences](#) | [Advanced View](#)

World Community Grid - BOINC Client

world community grid. Powered by IBM. technology solving problems

● dddt

World Community Grid

Application: Discovering Dengue Drugs - Together



Dengue outbreaks, 2006 (WHO)

Running Graphics Available

Elapsed Time: 0 h 0 m 26 s
Time Remaining: 1 h 29 m 18 s

0,0 %

My Projects: [Add Project](#) ?



Messages | Snooze | Preferences | Advanced View

World Community Grid - BOINC Client

world community grid. Powered by IBM. technology solving problems

● dddt ● dddt

World Community Grid

Application: Discovering Dengue Drugs - Together



Rational drug targets in the dengue virus polyprotein

Running Graphics Available

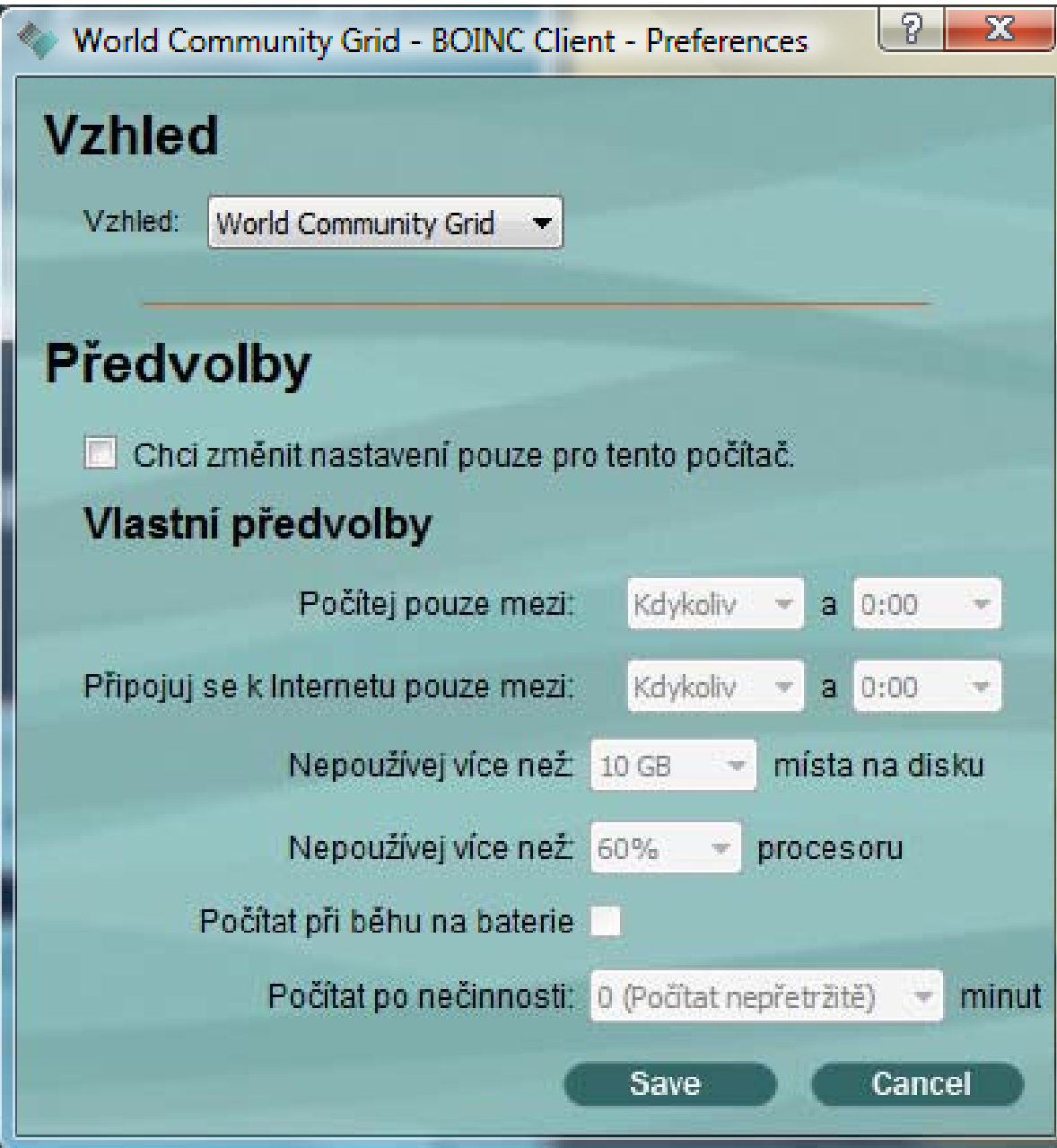
Elapsed Time: 0 h 0 m 26 s
Time Remaining: 1 h 29 m 18 s

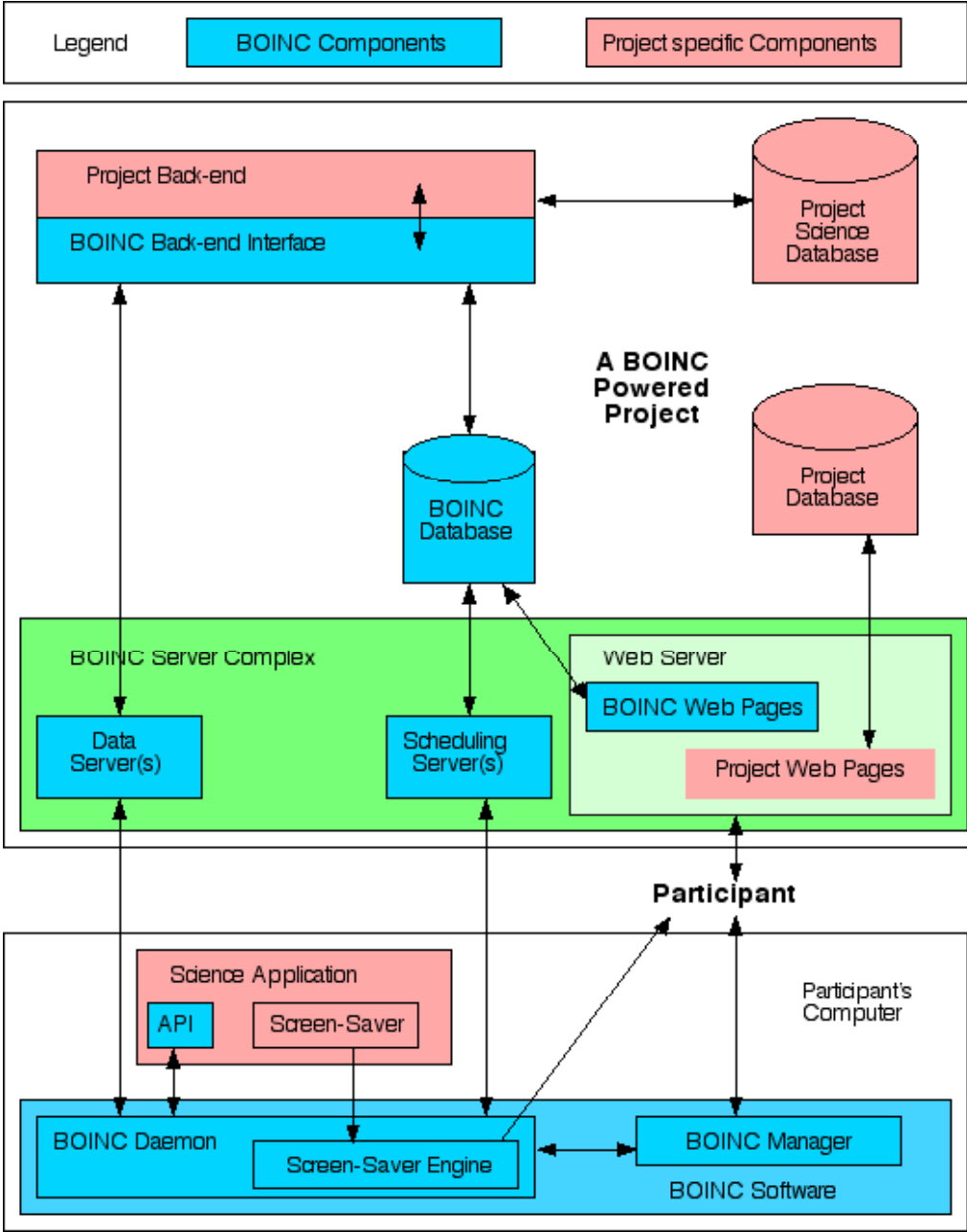
0,0 %

My Projects: [Add Project](#) ?



Messages | Snooze | Preferences | Advanced View

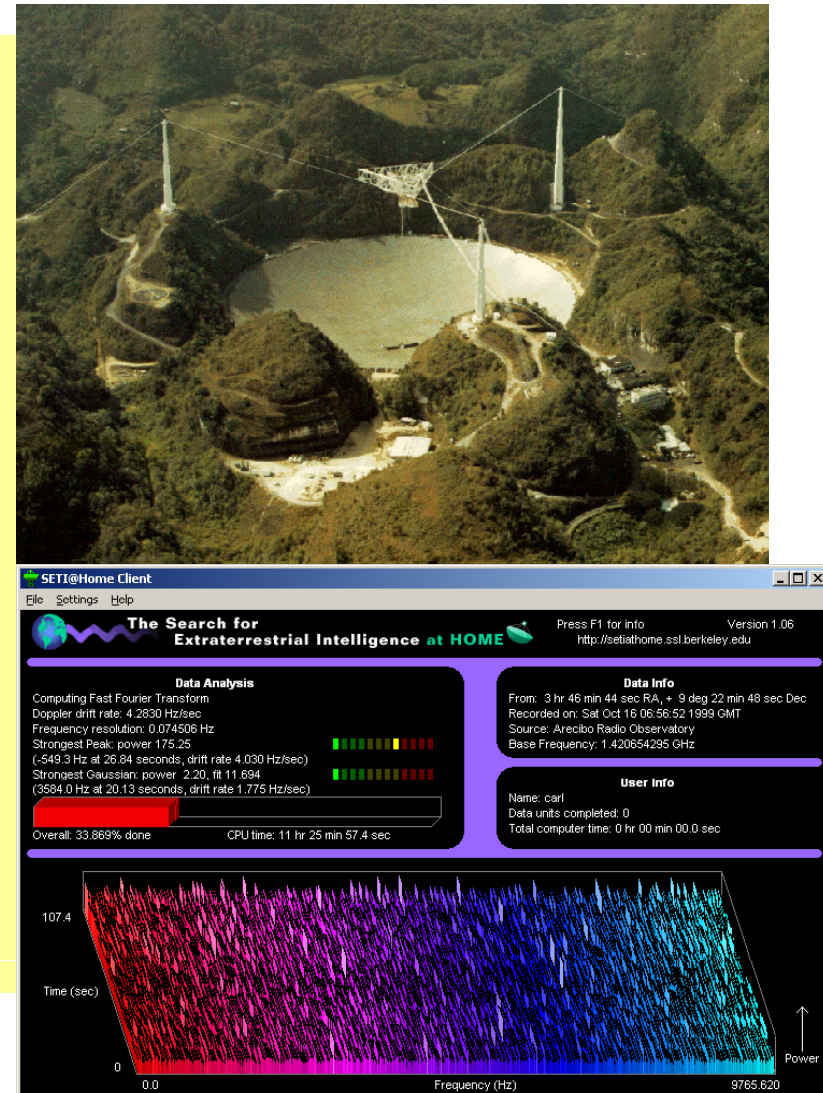


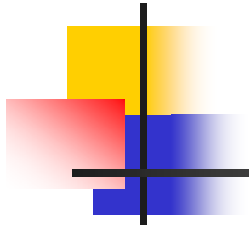


SETI: a global desktop grid

■ SETI@home

- 3.8M users in 226 countries
- 1200 CPU years/day
- 38 TF sustained (Japanese Earth Simulator is 32 TF sustained)
- Highly heterogeneous:
>77 different processor types

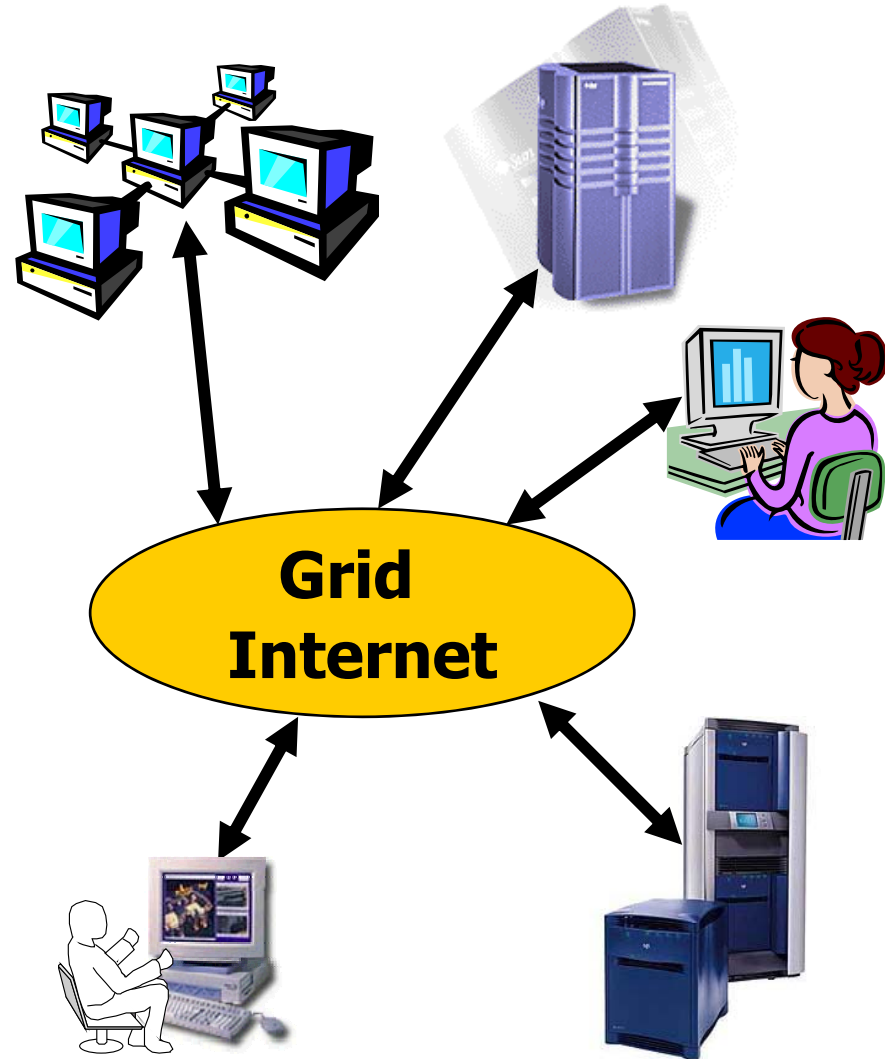




Gridy

What is Grid?

- A Grid is a collection of computers, storages, special devices, services that can **dynamically join and leave** the Grid
- They are **heterogeneous** in every aspect
- They are geographically **distributed** and connected by a **wide-area network**
- They can be accessed **on-demand** by a set of users





Definice gridu

I. Foster, C. Kesselman:

Výpočetní grid je hardwarová a softwarová infrastruktura, která poskytuje spolehlivý, standardizovaný, všudypřítomný a levný přístup ke špičkovým výpočetním službám.



Analogie – rozvod elektřiny

- původně každá budova vlastní generátor elektřiny
 - cca 1910, drahé, neefektivní
- zavedení elektráren a rozvodné sítě
- podobný vývoj ve využití výpočetních prostředků



Vlastnosti Gridu

- koordinuje zdroje **nepodléhající centralizované správě**
- používá standardní, otevřené, obecné protokoly a rozhraní
- poskytuje netriviální kvalitu i kvantitu služeb (více než jednotlivé části zvláště)
- Geografická vzdálenost nehraje roli



Vlastnosti Gridu

- **různé druhy zdrojů**
 - CPU, disk. prostor, přenosová kapacita sítí
 - speciální hw (senzory, mikroskopy..)
- různý hw participujících zařízení
- různé druhy interakcí
- různé uživatelské skupiny a aplikace
- **dynamičnost**
 - zdroje a uživatelé přibývají-ubývají-mění se



Typy gridů

- **Výpočetní**
 - spouštění aplikací na distribuovaných zdrojích
- **Datové**
 - sdílení velkého množství dat, replikované datové katalogy
- **Informační (znalostní)**
- Celosvětové
- Interní v rámci organizace
 - V pracovní době – po pracovní době - víkendy



Gridy - historie

- Cca od 90tých let 20.století ..
- Distribuovaná výpočetní infrastruktura pro vědecké a inženýrské výpočty
- Vytvářeny **virtuální organizace** (VO)
 - Správa a monitorování distribuovaných zdrojů
 - Bezpečnost
 - Důvěra
 - Ochrana soukromí



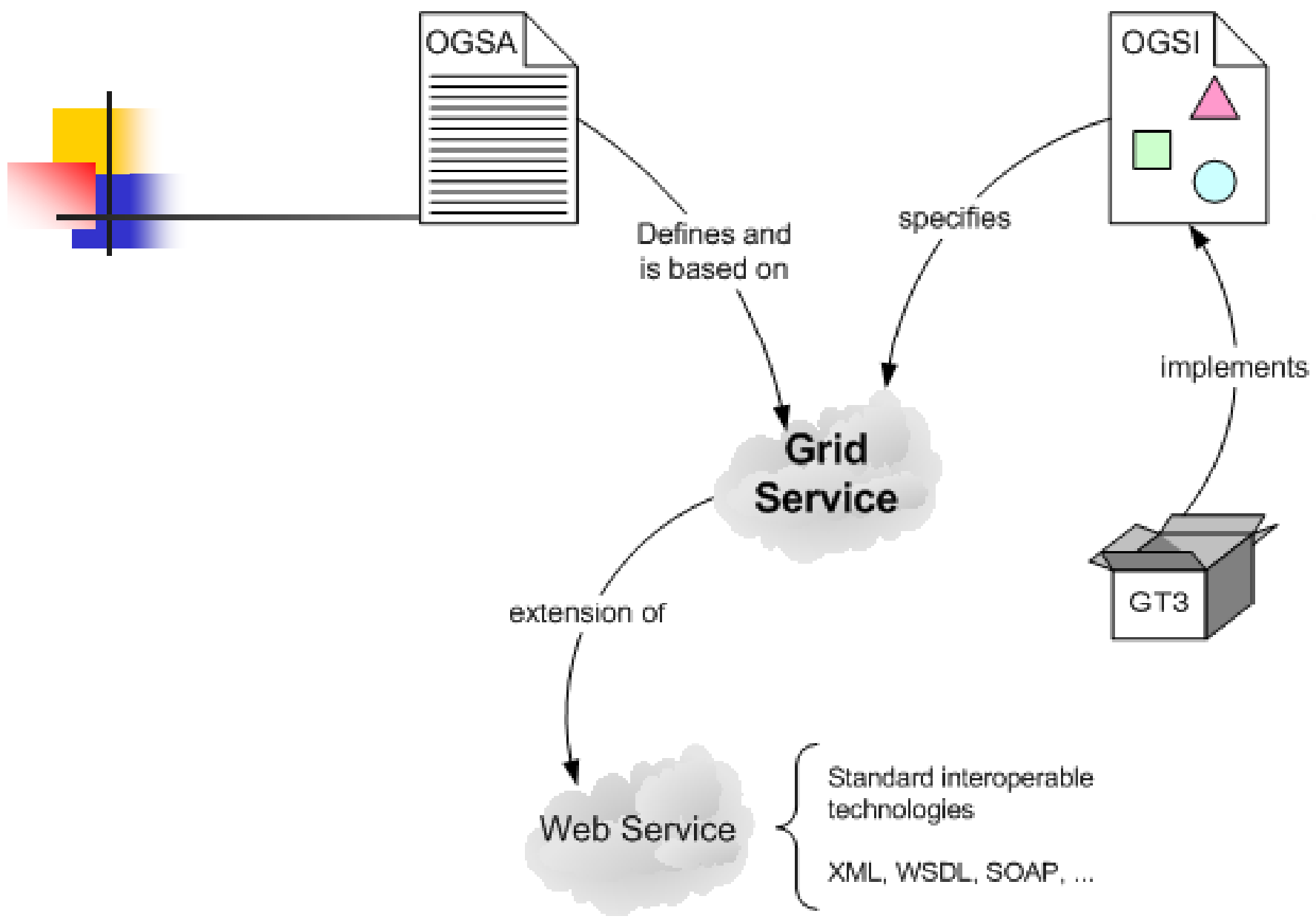
Virtuální organizace

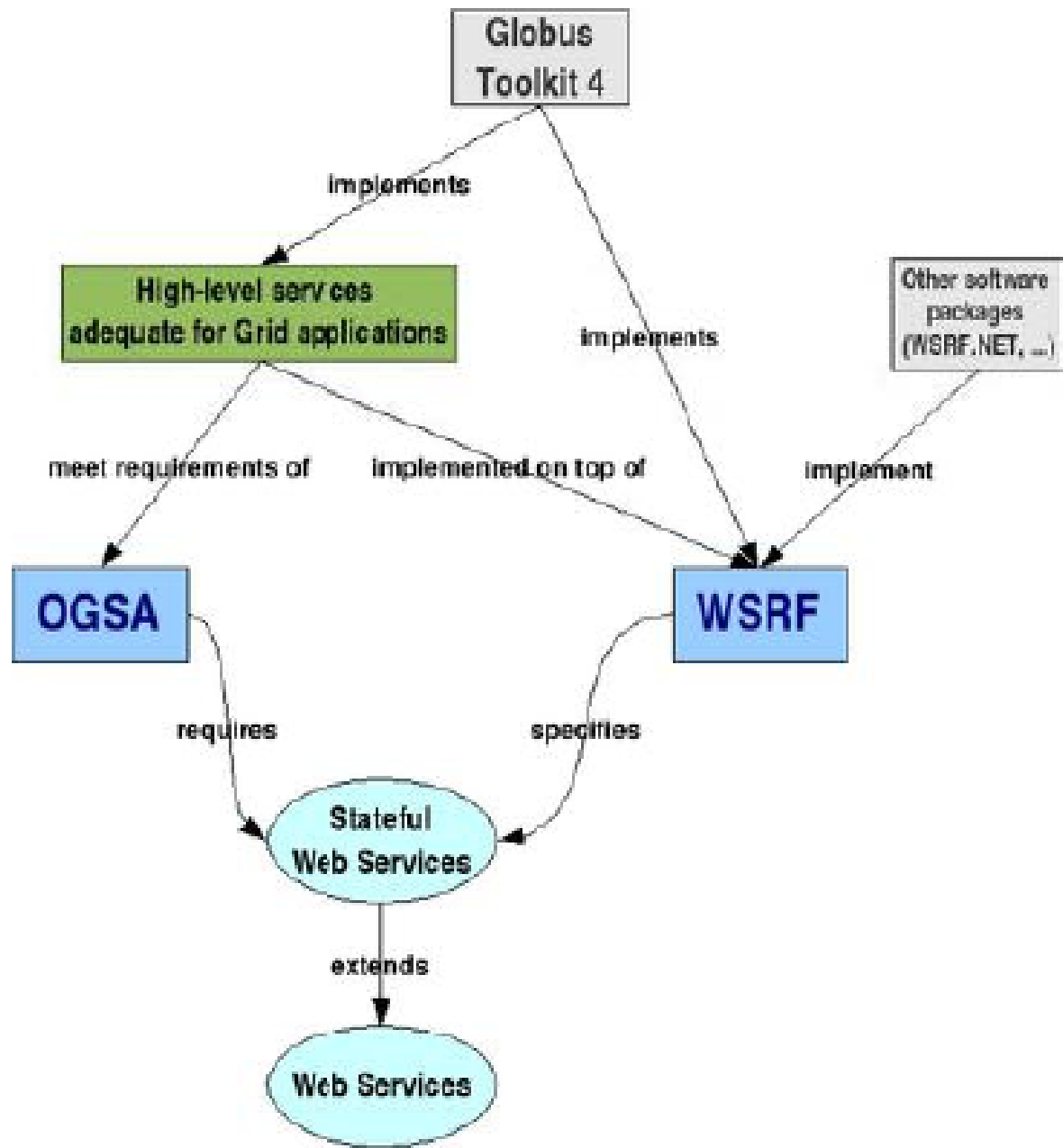
- Skutečné organizace
 - Mohou participovat v jedné nebo více virtuálních org.
- Sdílení zdrojů – podmíněné
 - Podmínky udané vlastníkem zdroje
 - Kdy, kde a co je možné se zdrojem dělat
- Nový účastník
 - K jakým zdrojům je možné přistupovat
 - Např. i dle publikací s odkazem na VO
 - Charakteristika zdrojů
 - Definice politiky, která řídí přístup ke zdrojům



Standardy

- OGSA (Open Grid Service Architecture)
 - Definuje gridové služby, žádné technické detailní specifikace
- OGSI (Open Grid Service Infrastructure)
 - Formální a technická specifikace
 - Nahrazena WSRF a WS-Management
- Globus Toolkit
 - Referenční implementace OGSI





The logo graphic consists of a vertical black line intersecting a horizontal black line. To the left of the intersection, there are three overlapping squares: a yellow one at the top, a red one in the middle, and a blue one at the bottom. The text 'OGSI' is positioned to the right of the vertical line, in a blue, sans-serif font.

OGSI

- OGSI definuje mechanismy pro vytváření, správu a výměnu informací mezi entitami – **gridové služby**
- Gridová služba – webová služba, která splňuje množinu specifikací (rozhraní a chování), které definují, jak klient komunikuje s touto službou
- Založena na WSDL



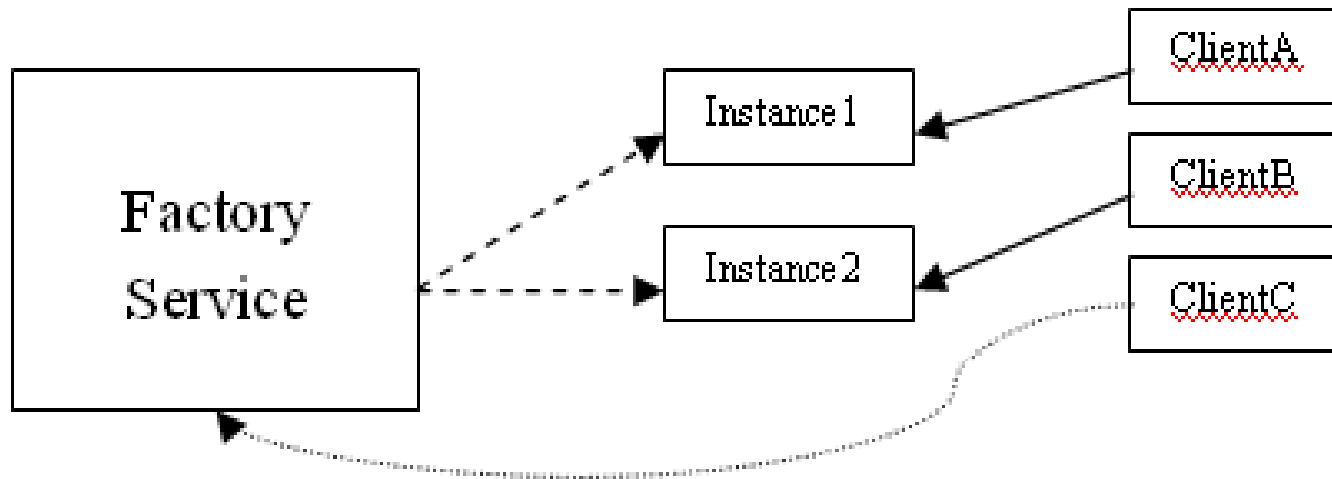
Webová služba

- Umístění – entita aplikačního serveru
- Rozhraní webové služby – popsané WSDL
 - Množina vykonatelných operací
- **Stateless**
 - Nepamatuje si stav mezi jednotlivými voláními
- **Non transient**
 - Klienti se připojují ke stejné instanci



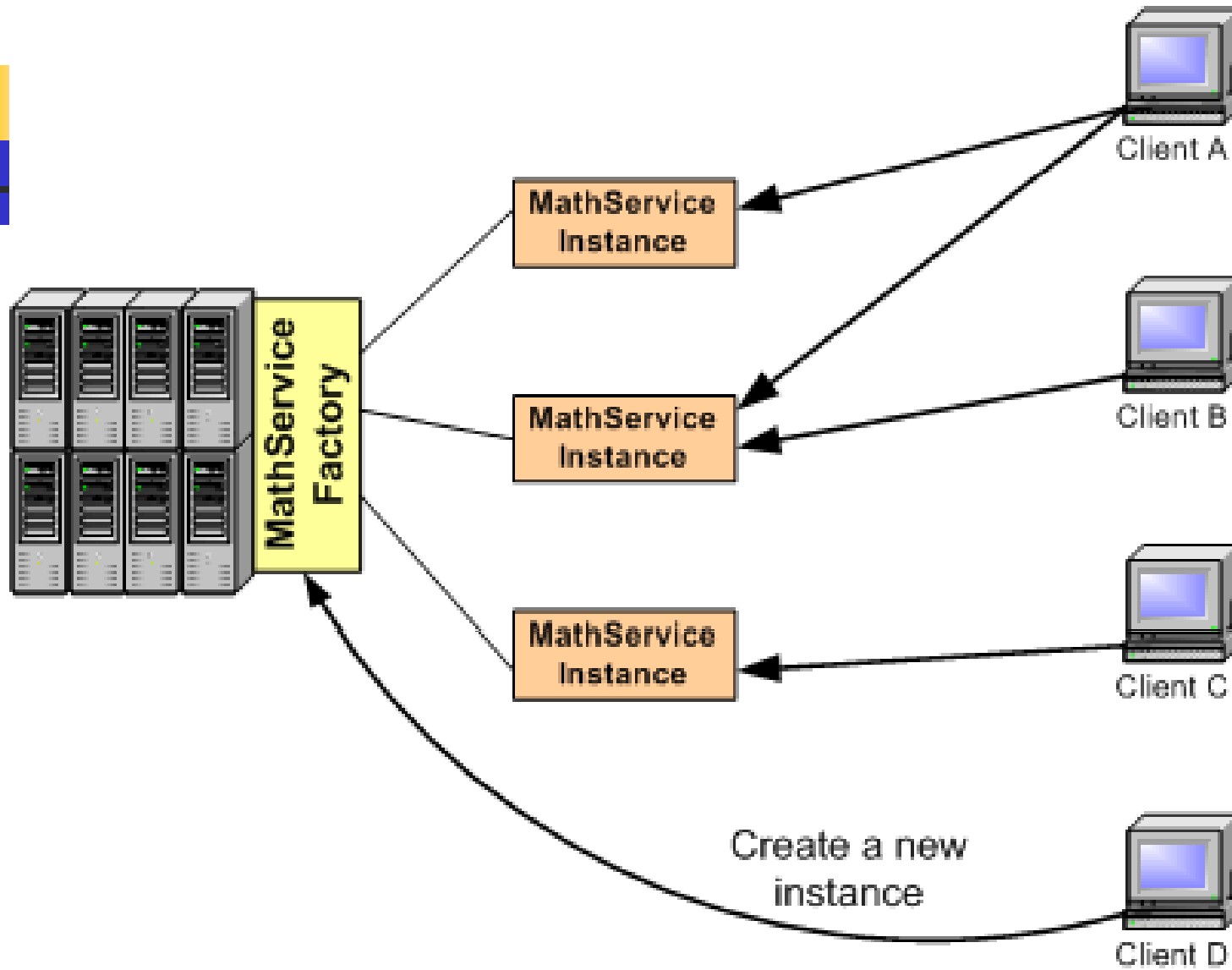
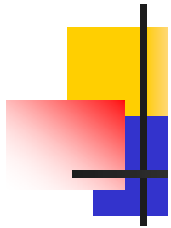
Gridová služba

- **Grid Service Factory**
 - Zodpovědná za vytváření a rušení objektů
- Klient vykonává operace na stejné instanci GS
- Jedna instance – většinou využívána jedním klientem
- **Service data elements** - popisují stav služby
 - State information – aktuální stav služby, výsledky operací
 - Service metadata – info o službě, např. náklady



Factory:

Vytvořit instanci, Zrušit instanci





GS – jednoznačná jména

- Odlišení různých instancí
- GSH (Grid Service Handler) – pojmenování, URI

<http://localhost:8080/ogsa/services/samples/counter/basic/CounterFactoryService/hash-31889293-1079702176271>

- GSR (Grid Service Reference)
 - Popisuje vše potřebné pro vyvolání služby
 - Ve formě WSDL dokumentu

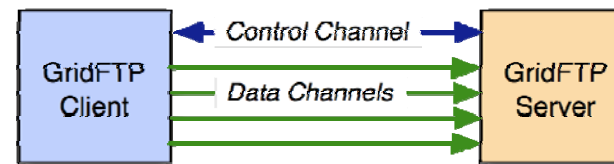


Globus Toolkit 3->4

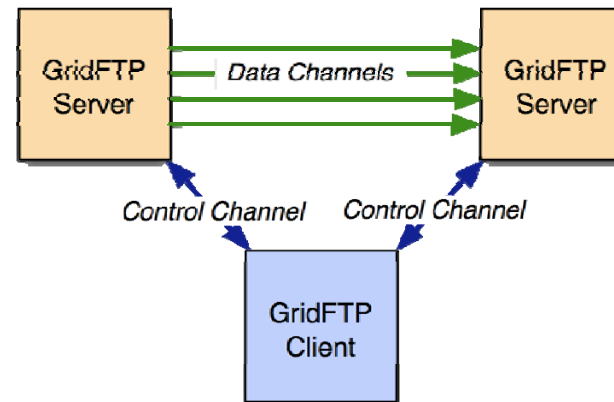
- Grid Service Middleware
- Implementace základních služeb založených na OGSI
 - Java
- Další služby – jazyk C (pouze na Unixech)
 - **GRAM** (Globus Resource Allocation Manager)
 - **GridFTP** (File Transfer Protocol, similar to FTP)
 - **MDS3** (Monitoring and Discovery Service)
 - **GSI** (Grid Security Infrastructure)

GridFTP ([z http://www-unix.mcs.anl.gov/~liming/primer/](http://www-unix.mcs.anl.gov/~liming/primer/))

- A high-performance, secure data transfer service optimized for high-bandwidth wide-area networks
 - FTP with extensions
 - Uses basic Grid security (control and data channels)
 - Multiple data channels for parallel transfers
 - Partial file transfers
 - Third-party (direct server-to-server) transfers
- OGF recommendation GFD.20

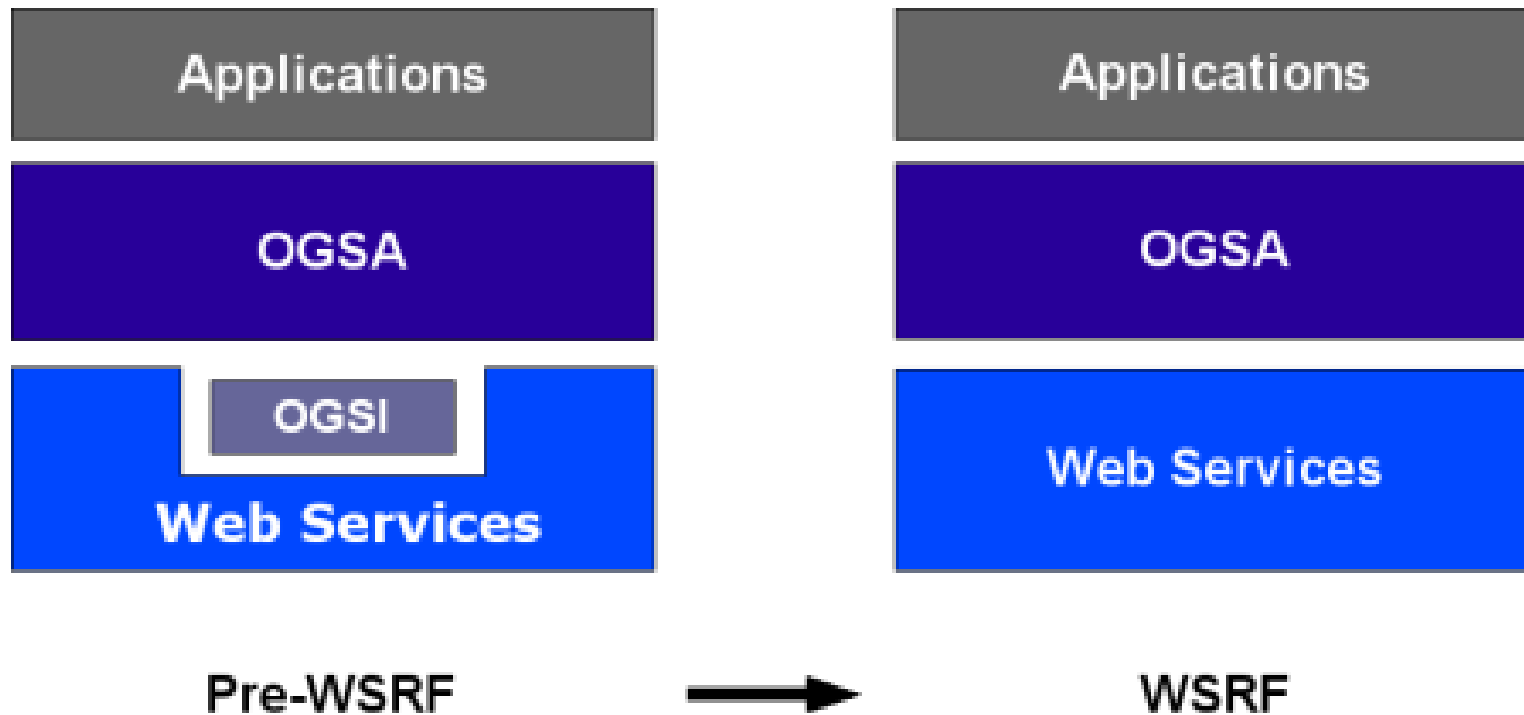


Basic Transfer
One control channel, several parallel data channels



Third-party Transfer
Control channels to each server, several parallel data channels between servers

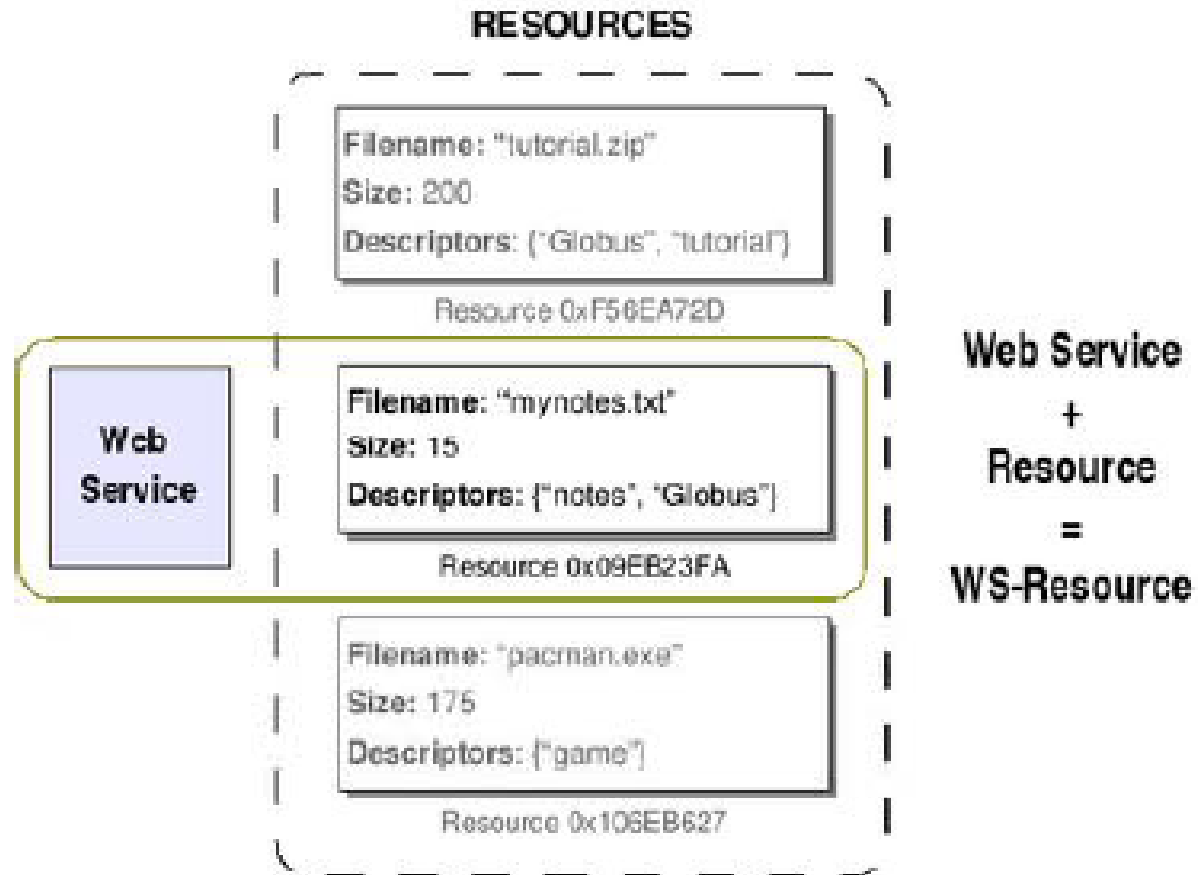
WSRF (Web Service Resource Framework)





WSRF

- společný standard pro webové a gridové služby
- využívá middleware Globus Toolkit 4
- jak vytvořit z bezstavových webových služeb služby stavové
 - OGSII – řešeno přidáním SDE ke každé gridové službě
 - WSRF – specifikuje resource, které jsou od služby odděleny a obsahují info dříve uložené v SDE
každý resource – identifikován klíčem

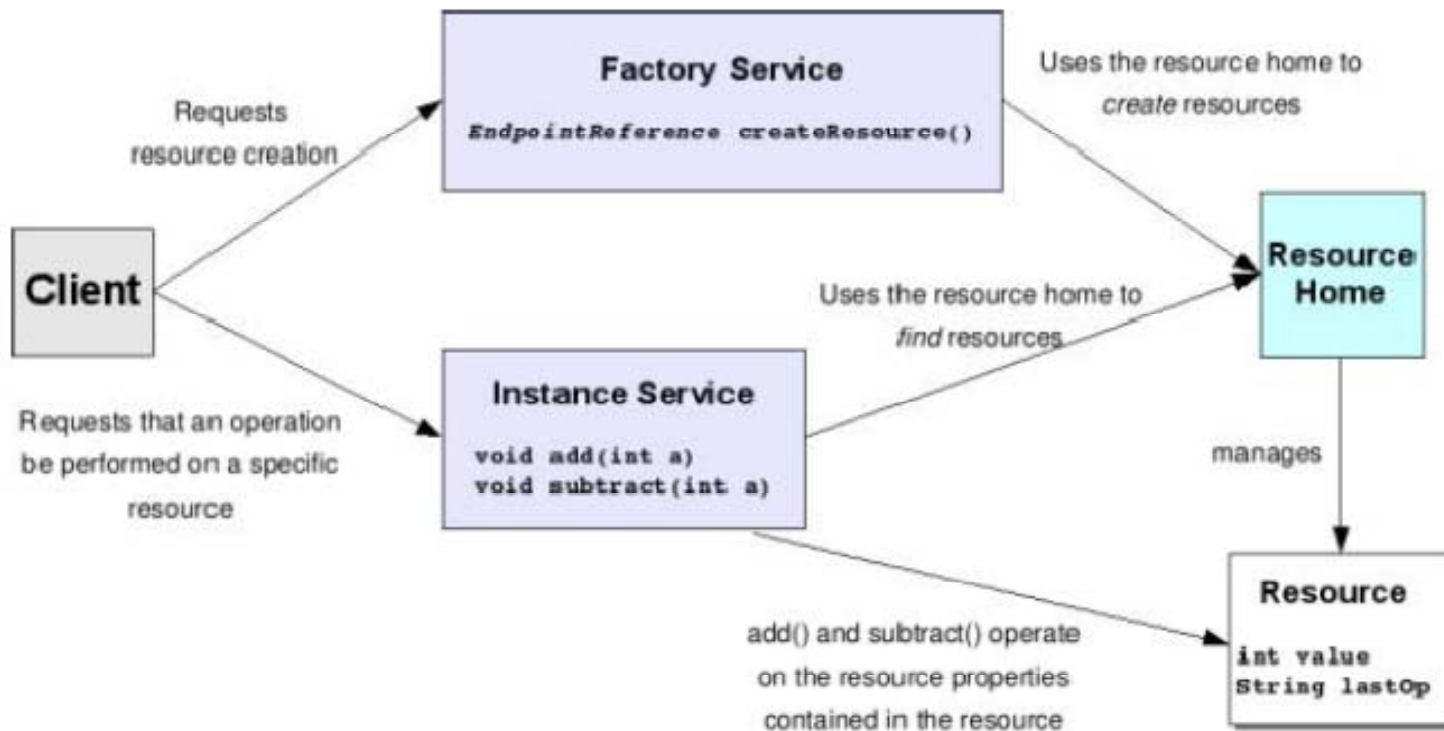




WSRF

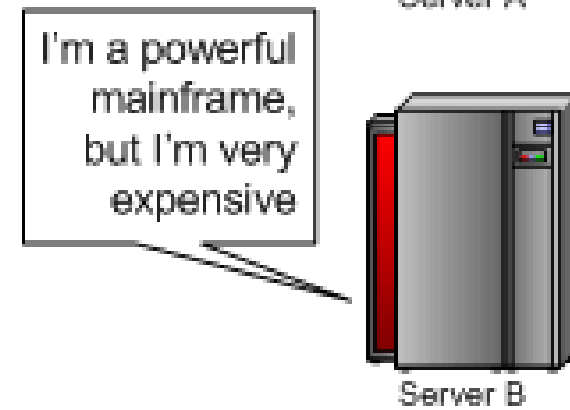
- klient se nepřipojuje ke specifické službě, ale k obecné službě s předem daným klíčem zdroje
- v GT4 se pomocí **factory** nevytvářejí instance služeb, ale pouze **instance zdrojů** (resource)
- instance služby – vytvořena při startu kontejneru a dále již pracuje se zdroji
- každý zdroj – unikátní klíč EPR (endpointReference)
 - dvojice služba - klíč

Implementace služby v GT4





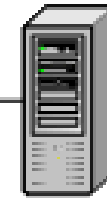
SDE



Více SDE

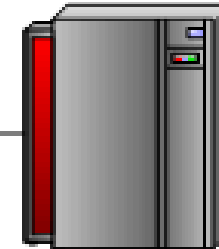


SDE:MathData
Speed: 5
Cost: \$30
Statistic: false
Service Data Type:
MathDataType



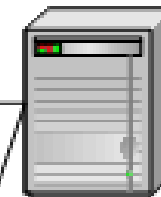
Server A

SDE:MathData
Speed: 9
Cost: \$125
Statistic: false
Service Data Type:
MathDataType



Server B

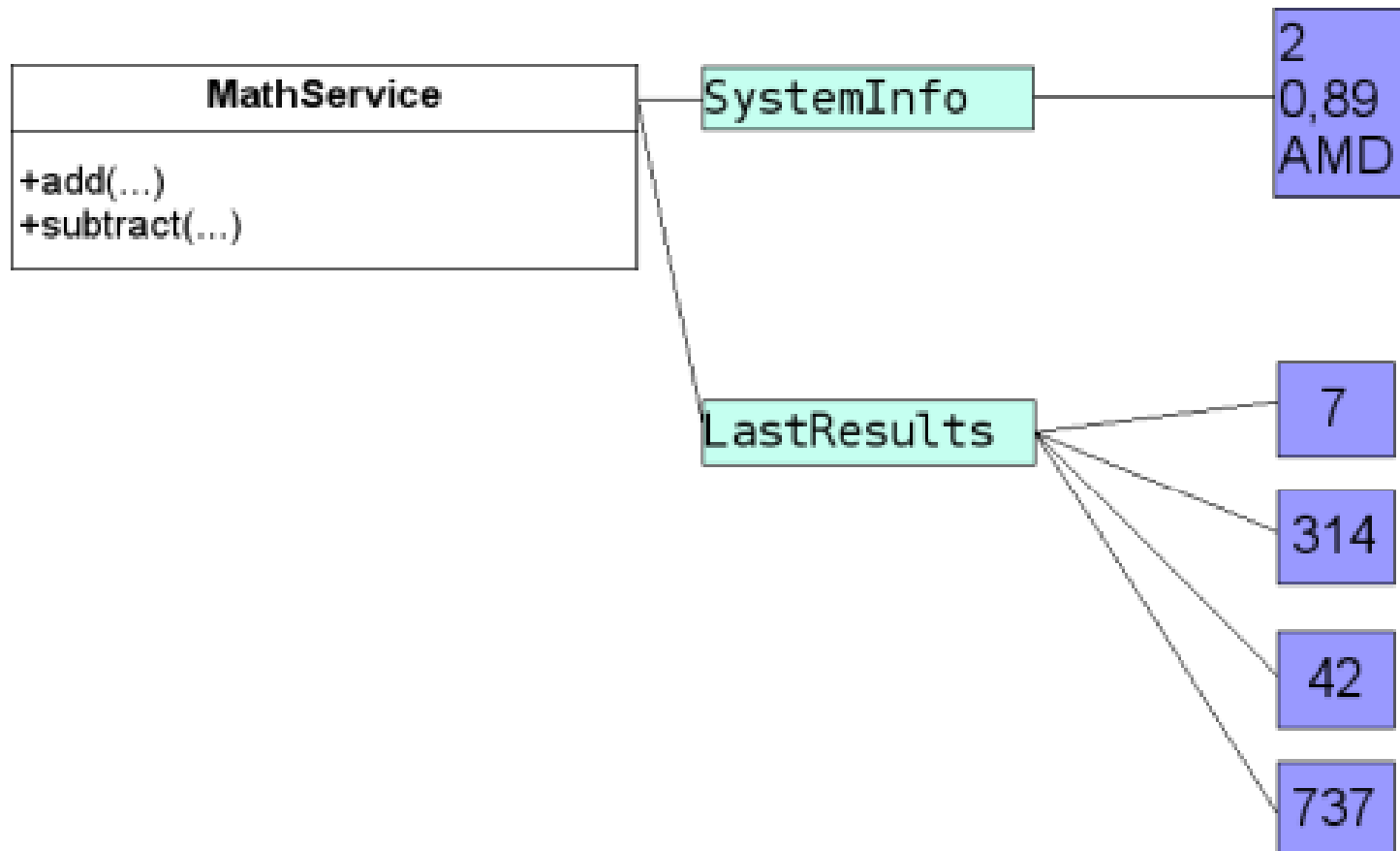
SDE:MathData
Speed: 7
Cost: \$90
Statistic: true
Service Data Type:
MathDataType



Server C

SDE:StatisticsData
Software Version: 3
Supported clients:
{A,B,C,F,T,Y}
Service Data Type:
StatisticsDataType

Více SDE





Volání služby - vytvoření

- ogssi-create-service
`http://127.0.0.1:8080/ogsa/services/progtutorial/core/first/MathFactoryService`
- *Service successfully created:*
Handle:
`http://127.0.0.1:8080/ogsa/services/progtutorial/core/first/MathFactoryService/hash-24981262-1078167170769`
Termination Time: infinity



Volání služby - klient

- `java -classpath ./build/classes/:$CLASSPATH \`
`org.globus.progtutorial.clients.MathService.Client \`
`http://127.0.0.1:8080/ogsa/services/progtutorial/core/first/Math`
`FactoryService/hash-24981262-1078167170769 5`
- *Added 5*
Current value: 5



Zrušení instance

- ogsi-destroy-service

<http://127.0.0.1:8080/ogsa/services/progtutorial/core/first/MathFactoryService/hash-24981262-1078167170769>



EGEE grid

- Funkce prvků je podobná u všech gridů
- Pojmenování specifické pro EGEE
- Cílem vývoj a integrace gridového prostředí
- > 70 institucí z Evropy, Rusko, USA

EGEE Grid

- The first EGEE infrastructure - Largest functioning Grid of the world:
 - more than 100 sites, over 10,000 CPUs, 4 PB
 - 5,000 jobs simultaneously

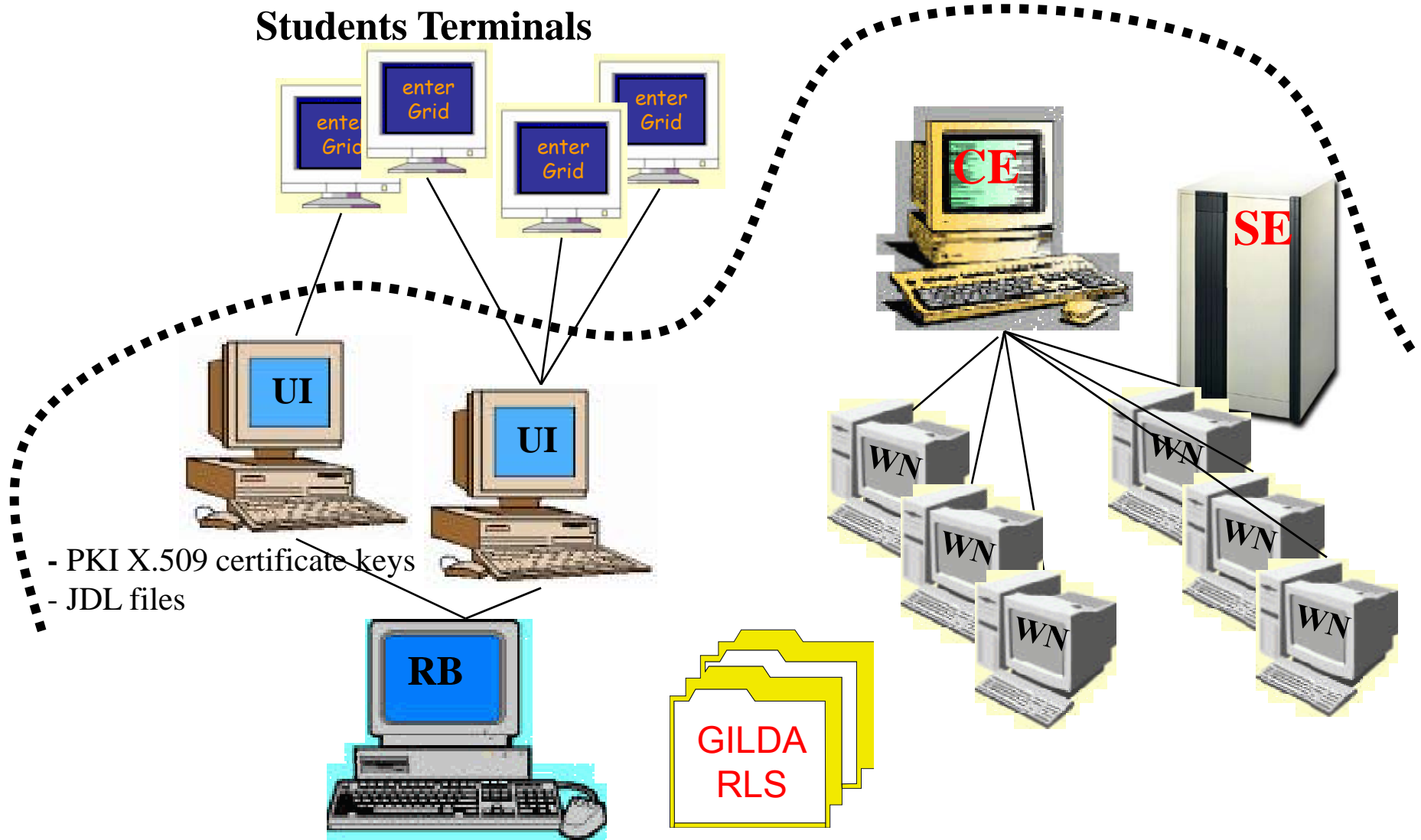




EGEE komponenty

- Resource Broker (RB)
- Compute Element (CE)
- Working Node (WN)
- Storage Element (SE)
- User Interface (UI)
- Replica Catalog (RC)
- Replica Location Server (RLS)

Students Terminals





UI – User Interface

- připojení klienta ke gridovému systému
- vytvořit novou úlohu (jdl)
- monitorování stavu úlohy
- přístup k uživatelským datům



CE – computing element

- přijme úlohu pro danou množinu homogenních uzlů
 - 1 PC, cluster, ...
- detailní informace o výkonu a instalovaném sw
- lokální dávkový systém
 - PBS, LSF, NQE, Condor



SE – Storage Element

- datové úložiště
- vzdálený přístup k datům
- repliky, přístup k nejbližší replice v gridu

- každý soubor
 - registrovaný
 - vlastní identifikace v gridu
 - identifikace nezávislá na jménu a lokaci



RC, RLS

- Informace o replikách souborů
- RC (Replica Catalog)
- RLS (Replica Location Server)



WN – Worker Node

- provádí vlastní výpočet
- přístup k aplikačnímu sw
 - lokálně instalovaný
 - dostupný přes sdílení
- není k nim přímý přístup
- množina WNs je reprezentovaná CE



RB – Resource Broker

- hlavní komponenta
- plánovač
- řídí distribuci zdrojů mezi výpočetní úlohy
 - jaké CE bude pro danou úlohu použito
 - pošle zvolenému CE tzv. InputSandBox (JDL,..)
- rozhoduje dle informací z IS (Information Service)



Životní cyklus úlohy v gridu

- **Submitted**

- úloha je vytvořena uživatelem, popsána .jdl souborem

- **Wait**

- RB najde vhodný CE
- může také najít nejbližší repliku požadovaných dat

- **Ready**

- RB připraví úlohu k běhu
- Přidá potřebné administrativní informace
- Vše pošle CE



Životní cyklus úlohy v gridu

- **Scheduled**

- CE přijme úlohu a předá ji lokálnímu dávkovému systému

- **Running**

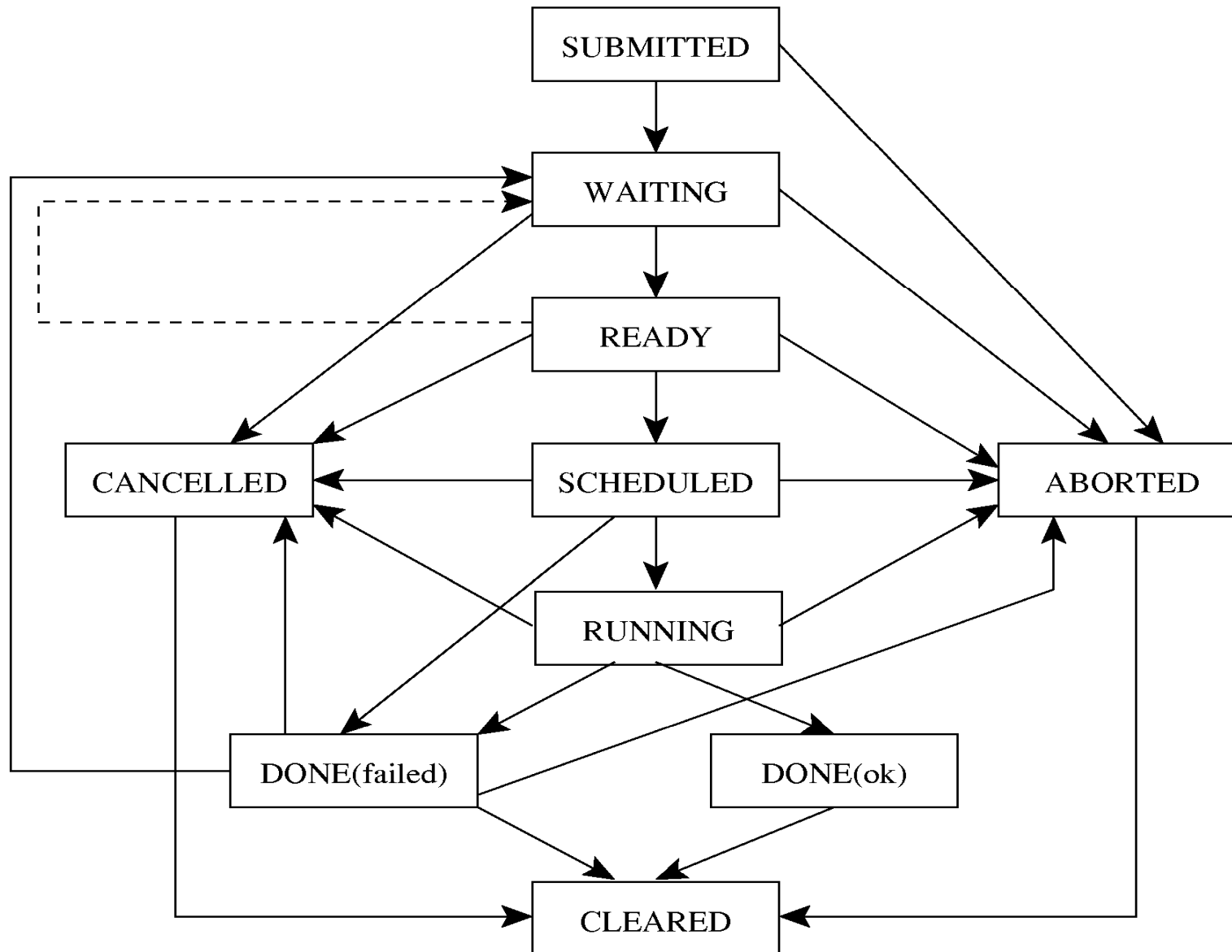
- úloha běží na dostupném WN
- uživatelská data zkopírována RB -> WN
- může využívat data ze SE
- nově vytvářená data – na SE a registrována v RLS



Životní cyklus úlohy v gridu

- **Done**
 - úloha je hotova
 - výstup – OutputSandBox (stdout,stderr) kopírován zpět na resource broker
- **Aborted**
 - úloha je zrušena uživatelem

Possible job states





JDL – Job Description Language

- popis požadavků a závislostí úlohy
- textový soubor .jdl
- vyžadované a volitelné parametry



JDL - parametry

- **Type** – “Job”
- **JobType**
 - Normal, Interactive, MPICH
- **Executable**
 - co se bude vykonávat
- **VirtualOrganization**
 - aktuální VO
- **NodeNumber**
 - počet vyžadovaných uzlů (MPICH)
- **Requirements** – další požadavky



JDL - parametry

- **Arguments**
- **StdInput, StdOutput, StdError**
 - definice I/O streamů, jména souborů
- **Environment**
- **InputSandBox**
 - které soubory pro běh potřebujeme, přeneseny do CE(WN) spolu s programem
- **OutputSandBox**
 - co bude přeneseno z CE do UI po skončení úlohy



JDL – příklad – myjob.sh

```
Executable           = "/bin/bash";
StdOutput            = "myjob.out";
StdError             = "myjob.err";
InputSandbox         = {"myjob.sh"};
OutputSandbox        = {"myjob.err", "myjob.out"};
RetryCount           = 1;
Arguments            = "myjob.sh 1 2 3";
```



JDL příklad

- úloha – spuštění myjob.sh
- výstup a chybový výstup – myjob.out, myjob.err
- skript bude poslán na cílový uzel jako součást InputSandBoxu

MPI Job

```
[JobType = "MPICH";  
  Executable = "cpi";  
  NodeNumber = 2;  
  StdOutput = "test.out";  
  StdError = "test.err";  
  InputSandbox = {"cpi"};  
  OutputSandbox = {"test.out", "test.err"};  
]
```

- The **NodeNumber** entry is the number of threads of MPI job
- The more processors you require the longer your job will stay in the queue waiting for free resources



Genius Portal, GILDA

- **Genius Portal**

- standardní grafický UI pro přístup k EGEE gridu

- **Gilda**

- virtuální laboratoře pro předvedení vlastností gridu

- Gilda Testbed

- RB, II, RLS, CE, SE
- INFN Grid Middleware (kompatibilní s LCG)



Gilda Grid Demonstrator

- <https://gilda.ct.infn.it/grid-demo.html>
- editace / prohlížení souboru
- VO Services – Job Services – Job Submission
- 09 Simple Hello World
helloworld.jdl



- GILDA Grid Services
 - Job Services
 - Job Submission
 - Job Queue
 - Job Data
 - Clean Job Queue
 - HadronTherapy Services
 - Raster-3D
 - WATERMARKING
 - CODESA3D
 - EMPIRE
 - GA4ts
 - GATE
 - gMOD
 - MAGIC
 - NOAH
 - PATSEARCH
 - Sonification
 - SW-Alignment
 - Back home

Job Submission

```
Selected Virtual Organisation name (from proxy certificate extension): gilda
Connecting to host glite-rb2.ct.infn.it, port 7772
Logging to host glite-rb2.ct.infn.it, port 9002

===== glite-job-submit Success =====
The job has been successfully submitted to the Network Server.
Use glite-job-status command to check job current status. Your job identifier is:

- https://glite-rb2.ct.infn.it:9000/-ZlREWtazysiY7Wnh-aziQ

The job identifier has been saved in the following file:
/home/demo43/.genius/.tmp_submittedjob_demo43
```

[Click here to see the status of this job](#)



- LDA Grid Services
- Job Services
 - Job Submission
 - Job Queue
 - Job Data
 - Clean Job Queue
- HadronTherapy Services
- Raster-3D
- WATERMARKING
- CODESA3D
- EMPIRE
- GA4ts
- GATE
- gMOD
- MAGIC
- NOAH
- PATSEARCH
- Sonification
- SW-Alignment
- Back home

Job Queue

#	Globus JobID	Last Update Time	Status	Destination	Exit Code	Name
1	-ZIREWtayzsiY7Wnh-aziQ	16:25:39 Mon Dec 3:	Scheduled	grid010.ct.infn.it:2119/jobmanager-lcgpbs-long		helloworld.jd

Details

Submission:

- Job manager: genius
- Time: 16:25:39 Mon Dec 3:
- Host:
- Directory:

Execution:

- Time:
- Host: grid010.ct.infn.it:2119/jobmanager-lcgpbs-long
- Directory:



Middleware

- Přístup aplikací ke kapacitám gridu
- Management prostředků gridu

- zdroj gridu
 - výpadek
 - saturace



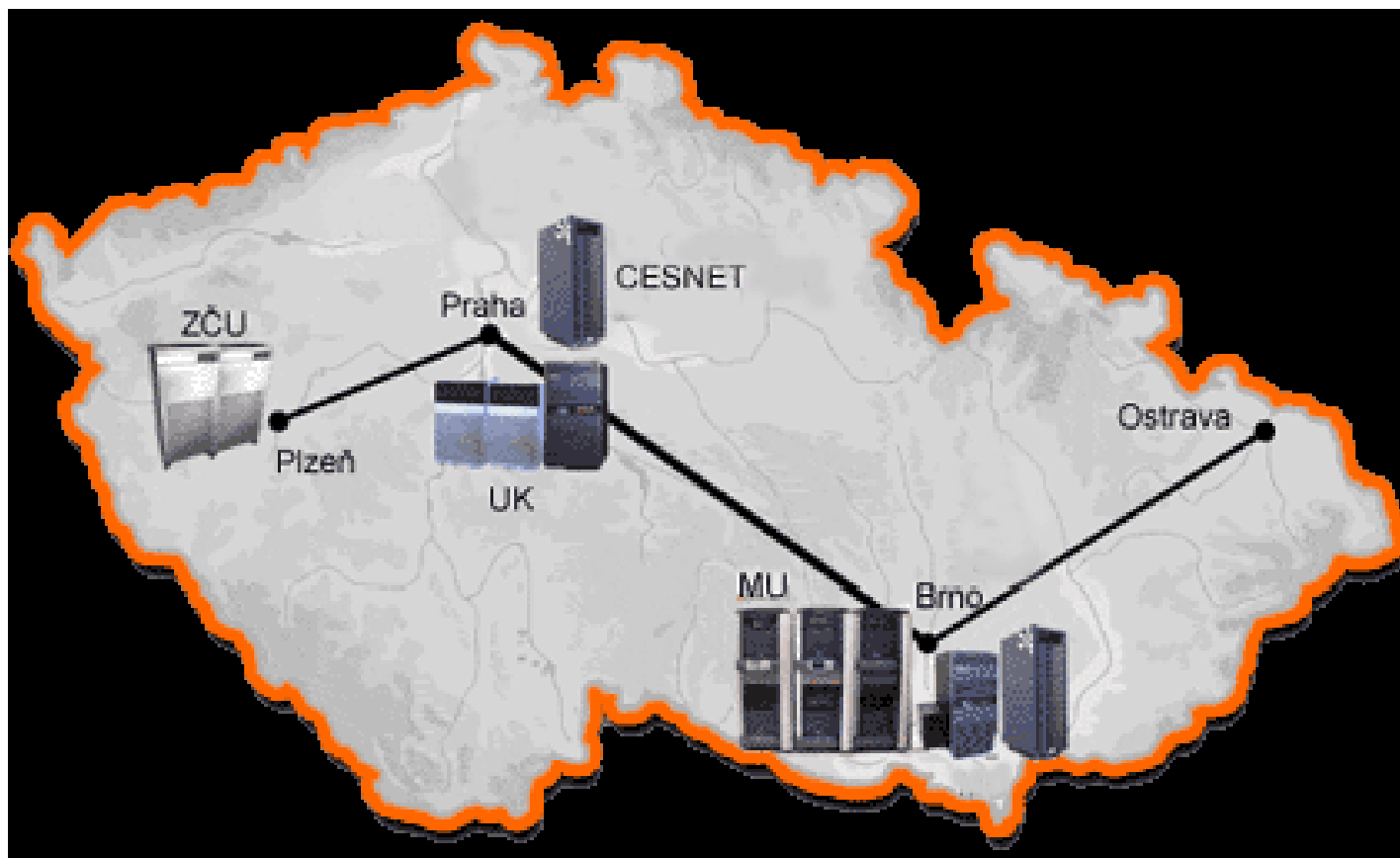
Použitá literatura, odkazy

- <http://www.ics.muni.cz/zpravodaj/articles/343.html>
- <http://www.egee.hu/grid05/>



MetaCentrum

- <http://meta.cesnet.cz>
- EGEE
- MediGrid
 - CESNET, FN Motol, MN Ústí nad Labem
- CoreGrid



Struktura
Ukázky informací z meta.cesnet.cz