



Grid aprēķinu vide

teorija • metodes • aprēķini

DatuGrid izveides iespējas Latvijā

D.Ludviga, K.Krampis

LU 66. konference, 2008.gada 13.februāris, LU MII



Kas ir *DatuGrid*?

- A Data Grid is a grid computing system that deals with data — the **controlled sharing and management of large amounts of distributed data**. These are often, but not always, combined with computational grid computing systems. *[Wikipedia]*
- A Data Grid provides **seamless access to the local or remote data required to complete compute intensive calculations**. *[www.gridconsortium.com]*
- Data Grids primarily deal with providing services and **infrastructure for distributed data-intensive applications that need to access, transfer, and modify massive datasets stored in distributed storage resources**. *[Chervenak et al 2000; Hosceh et al. 2000]*
- Data Grid allows institutions to **collaborate and share resources across geographic and administrative boundaries**. *[S.Venugopal, R.Buyya, and K.Ramamohanarao, "A Taxonomy of Data Grids for Distributed Data Sharing, Management, and Processing. "]*



DatuGrid iespējamie modeļi

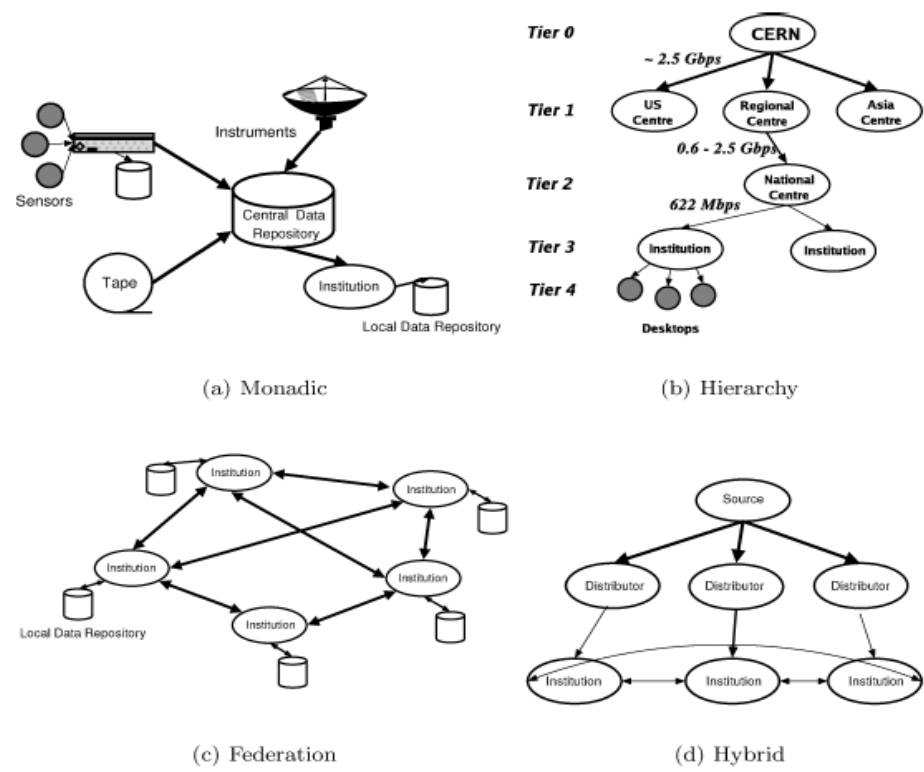


Fig. 5. Possible models for organization of Data Grids.



Projekti un piemēri.

Nosaukums	Nozare	<i>DatuGrid</i> modelis
LCG [2005]	High Energy Physics	Hierarchical
EGEE [2005]	High Energy Physics, Biomedical Sciences	Hierarchical
BIRN [2005]	Bio-Informatics	Federated
NEES grid [Pearlman et al. 2004]	Earthquake Engineering	Monadic
GriPhyn [Avery and Foster 2001]	High Energy Physics	Hierarchical



Citi risinājumi



■ Wuala

- bezmaksas Java bāzētā lietojumprogramma
- dati tiek glabāti centralizētā P2P tīklā
- lokālās atmiņas maiņa pret tiešsaistes
- datu šifrēšana
- datu pieejamība izmantojot *erasure code*
- neierobežots faila izmērs



Bez problēmām nekur!

- Drošība
- Piekļuves tiesības
- Vienota vārdtelpa (*namespace*)
- Apjomīgu datu uzglabāšana, koplietošana



Jaunākās problēmas

■ Terenas *task force*:

- Enerģijas patēriņš
- Datu aprakstīšana
- Autorizācijas Autentifikācijas Infrastruktūra (AAI)
- Atkopšanās
- Datu koplietošana

■ TCP protokola aizvietošana



Status quo

■ Esošo risinājumu izpēte

- Dalītās klasteru failu sistēmas (PVFS, Lustre, GPFS, TerraFS)
- Krātuves resursu vadības sistēmas:
 - ▶ dCache/SRM
 - ▶ SRB
- Grid krātuves resursu vadības protokols:
 - ▶ GsiFTP (GridFTP)



Status quo (2)

■ Esošo risinājumu izpēte

□ Dalīto failu sistēmu pieejas protokoli:

- ▶ SMB
- ▶ NFS
- ▶ SSHFS
- ▶ WebDAV
- ▶ FTP
- ▶ GridFTP

■ LUMII dCache testa datu klastera izveide

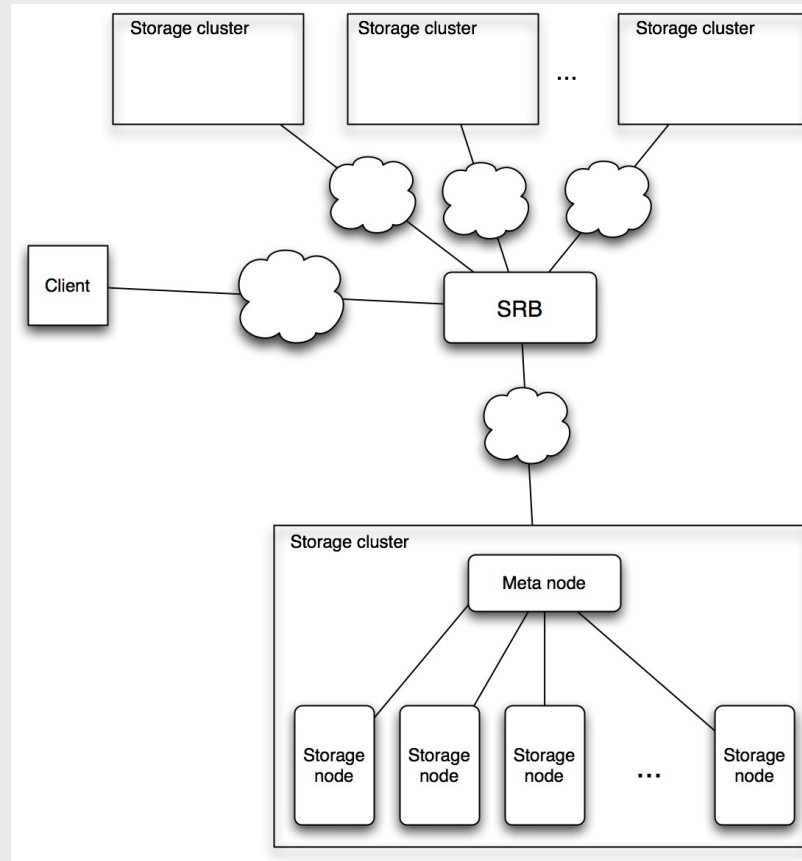


Nākotnes plāni/vīzija

- Izveidot lokālu LUMII dalīto datu krātuvi, kas savietojama ar krātuves vadības sistēmu (piem: SRB), apvienojot vairāku institūciju vai valstu jau esošās krātuves.
- Nodrošināt zinātniekiem plašas iespējas dalīties ar savām zināšanām un sadarboties, uzglabājot savu pētījumu datus kopīgā, drošā krātuvē.



Nākotnes plāni/vīzija (2)





Mājas darbs :)

- “The design of a world-wide, fully transparent distributed file system for simultaneous use by millions of mobile and frequently disconnected users is left as an exercise for the reader. ”

[A. Tanenbaum, *Distributed Operating Systems*, 1995]



Paldies par uzmanību!