Harnessing Organizationally Distributed Data with VeilFS

L. Dutka, R. Slota, M. Wrzeszcz, D. Krol, L. Opiola, R. Slota, J. Kitowski

ACK Cyfronet AGH

Department of Computer Science, AGH UST

KUKDM 2014
Zakopane, Poland, March 12-14, 2014

Agenda

- Addressed problem

- Overview of VeilFS

- VeilFS – user point of view

- Exemplary use case

- Integration with other services

- Preliminary tests results

- Conclusions

Addressed problem

The user is typically interested in results - he expects that data access will be simple while data management in distributed environment is often too complicated for many users.

> The access to the data in large scale computational environment should be the same as on personal computer.

Administrator is interested in effective utilization of resources, he expects that data management will be simple.

> Data management in large distributed environments where many problems appear (several different technologies, accounting, users' groups) should be automated and supported by unified tools.
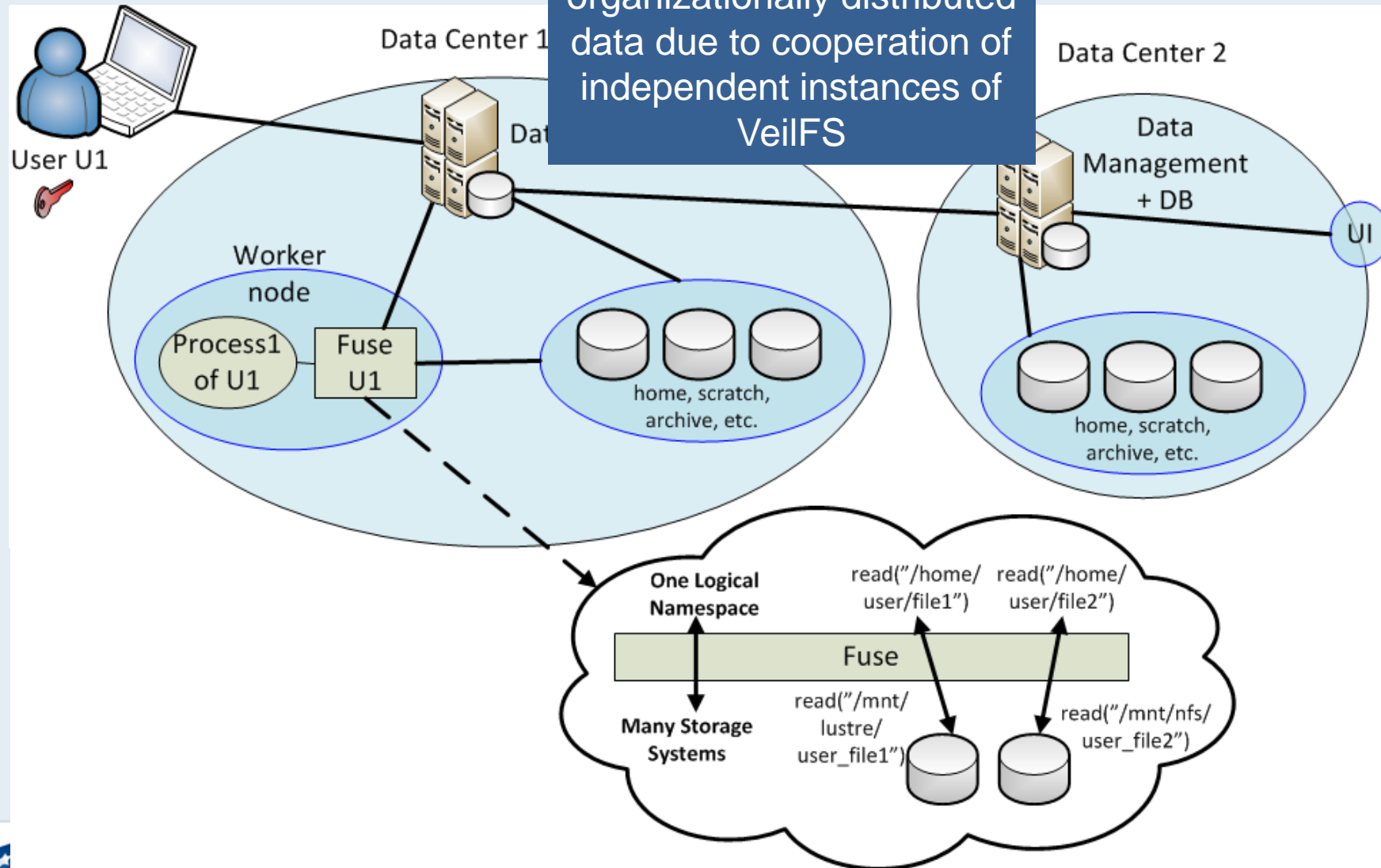
How to do that?

Overview of VeilFS

Management of organizationally distributed data due to cooperation of independent instances of VeilFS
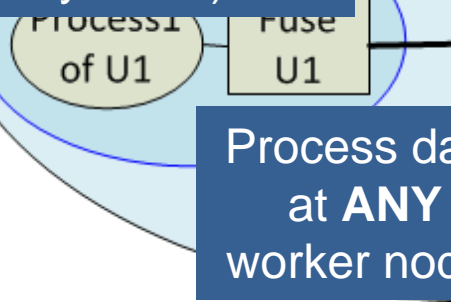
VeilFS – user point of view

PL GRID+

Ease of use and unified access from anywhere

Mount VeilFS at [...]
PC and use your [...]
as if they were [...]
(currently only Li[...])

Or simply use Internet browser (all operating systems).

User U1

Process1 of U1

Fuse U1

Process da[...]
at **ANY** [...]
worker nod[...]

File manager

https://veilfs.plgrid.pl/file_manager

studia doktoranckie · agh · rozrywka · TV · praca · Programowanie · banki · tłumacz · różne · budowa · konkurenci VeilFS

File manager   Shared files   Michał Wrzeszcz

~   Search

Sort

Name	Size	Access	Modification
groups		2013-10-21 22:58:06	2013-10-21 22:58:06
1org manipulacja drzewem.png	35.60 KB	2013-10-24 09:33:23	2013-10-24 09:33:23
1org plik lokalny.png	21.90 KB	2013-10-24 09:33:24	2013-10-24 09:33:24
1org plik zdalny.png	42.12 KB	2013-10-24 09:33:25	2013-10-24 09:33:25
instal client.txt	430 B	2013-10-24 09:33:26	2013-10-24 09:33:26
obecna koncepcja.png	134.34 KB	2013-10-24 09:33:27	2013-10-24 09:33:27
rozproszenie.txt	310 B	2013-10-24 09:33:28	2013-10-24 09:33:28

VeilFS – user point of view

- Ease of use and unified access from anywhere
- Data publication

Simple sharing in groups (throw data to group directory)

files and stop sharing if needed

File manager   Shared files   Michał Wrzeszcz

https://veilfs.com/share/04e8ed45063fddeecae55f39eaa80...

Google

~ / groups

Search

Sort

Name	Size	Access
..		1970-01-01 01:00:00
plggqstorman		2013-10-17 14:27:04
plggveilfs		2013-10-17 14:27:17

File manager   Shared files   Michał Wrzeszcz

Manage account

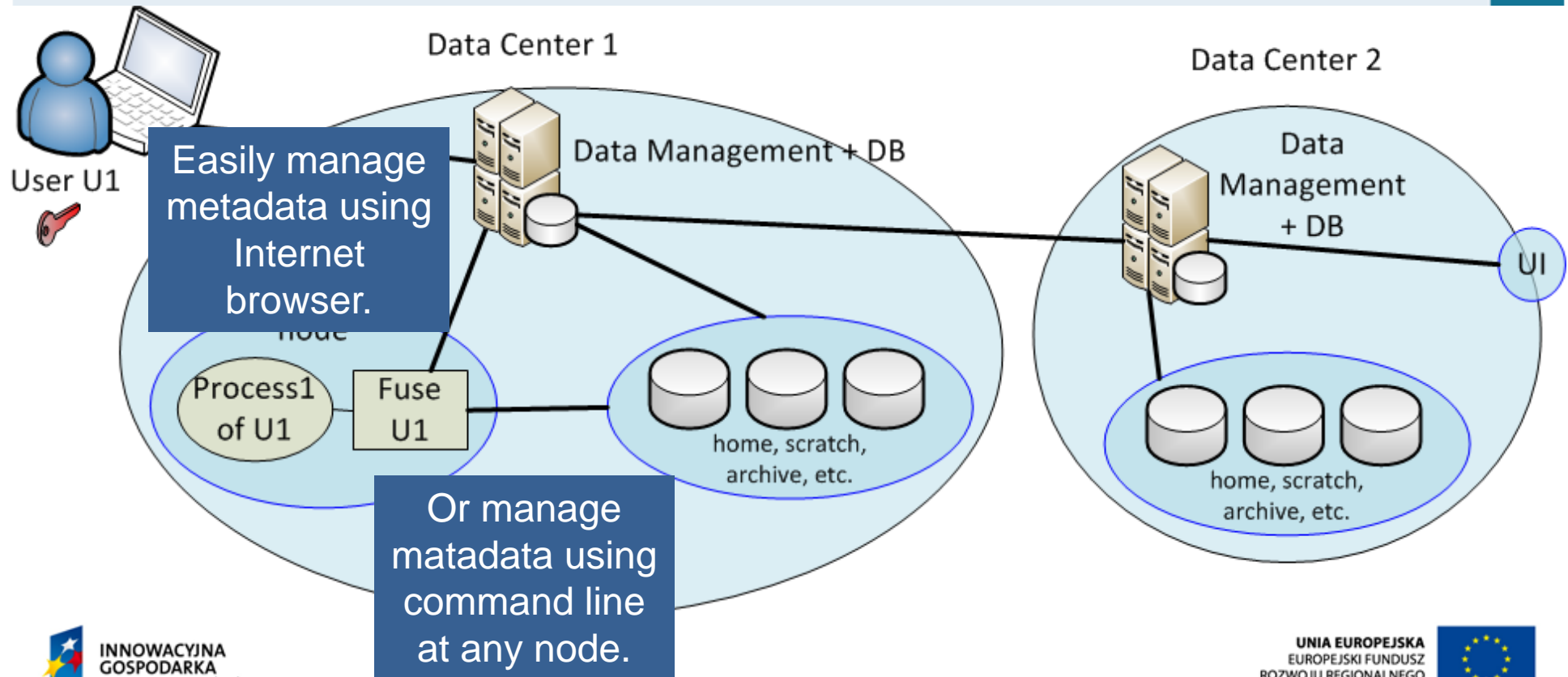| Login	plgwrzeszcz
Name	Michał Wrzeszcz
Teams	1. plggqstorman (QStorMan)
	2. plggveilfs (VeilFS)
E-mails	1. michalwrzeszcz@gmail.com ✕
	2. +
Certificates' DNs	1. CN=plgwrzeszcz,CN=Michal Wrzeszcz,O=CYFRONET,O=Uzytkownik,O=PL-Grid,C=PL ✕
	2. +

VeilFS – user point of view

■Ease of use and unified access from anywhere
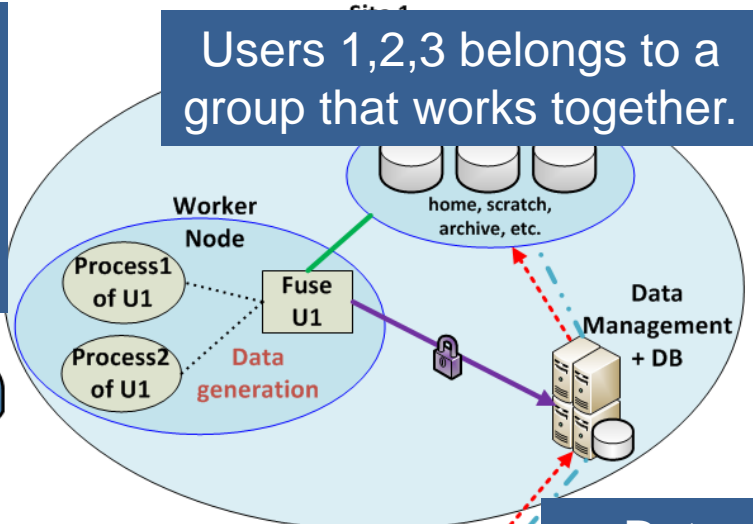■Data publication and sharing within groups
■Metadata support

Exemplary use case with 2 sites

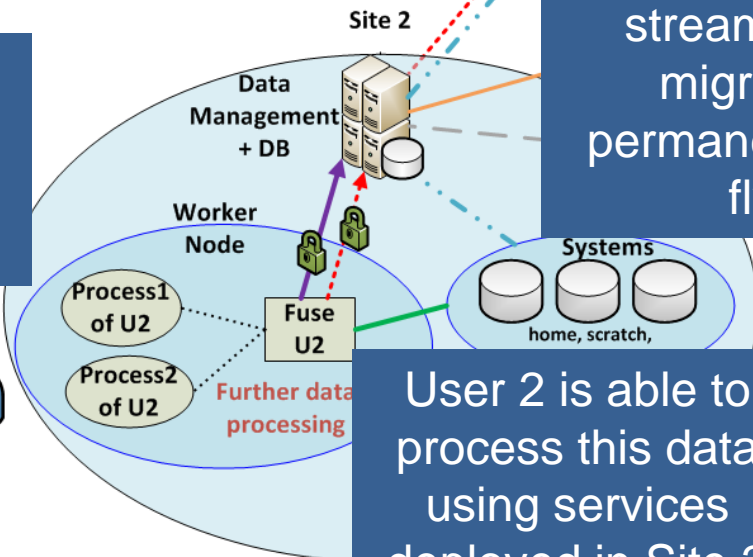User 1 processes data using services deployed in Site 1.

Users 1,2,3 belongs to a group that works together.

User 3 can anayse data produced by coworkers and publish chosen results using Web-based GUI.

User 2 sees data produced by User 1 in any site.

Data can be streamed or migrated permanently on fly

Anyone is able to download chosen results.

User 2 is able to process this data using services deployed in Site 2

Legend:
- Getting physical location of file on the basis of logical name
- Direct access to data
- Access to files stored in another center
- Data migration between various storage systems
- Results analysis and publication via GUI
- Download

Site 1 — Worker Node — Process1 of U1, Process2 of U1 — Fuse U1 — Data generation — home, scratch, archive, etc. — Data Management + DB — User U1

Site 2 — Data Management + DB — Worker Node — Process1 of U2, Process2 of U2 — Fuse U2 — Further data processing — Systems — home, scratch, — User U2

# Integration with other services

- VeilFS provides REST API
  - API allows direct interfacing from third party applications
  - third party applications can interact with VeilFS on behalf of the users with standard X509 proxy certificates

- Domain Grids – main client of REST API
  - VeilFS can provide data to computing elements where services are deployed
  - VeilFS can manage data storing and migration when services do not have access to shared storage

- REST API evolution – VeilFS integration with Grid middleware
  - VeilFS can provide information about data that helps job schedulers in choice of best site for the job

Preliminary tests results

Transfer rates measured by SysBench [Mb/s]
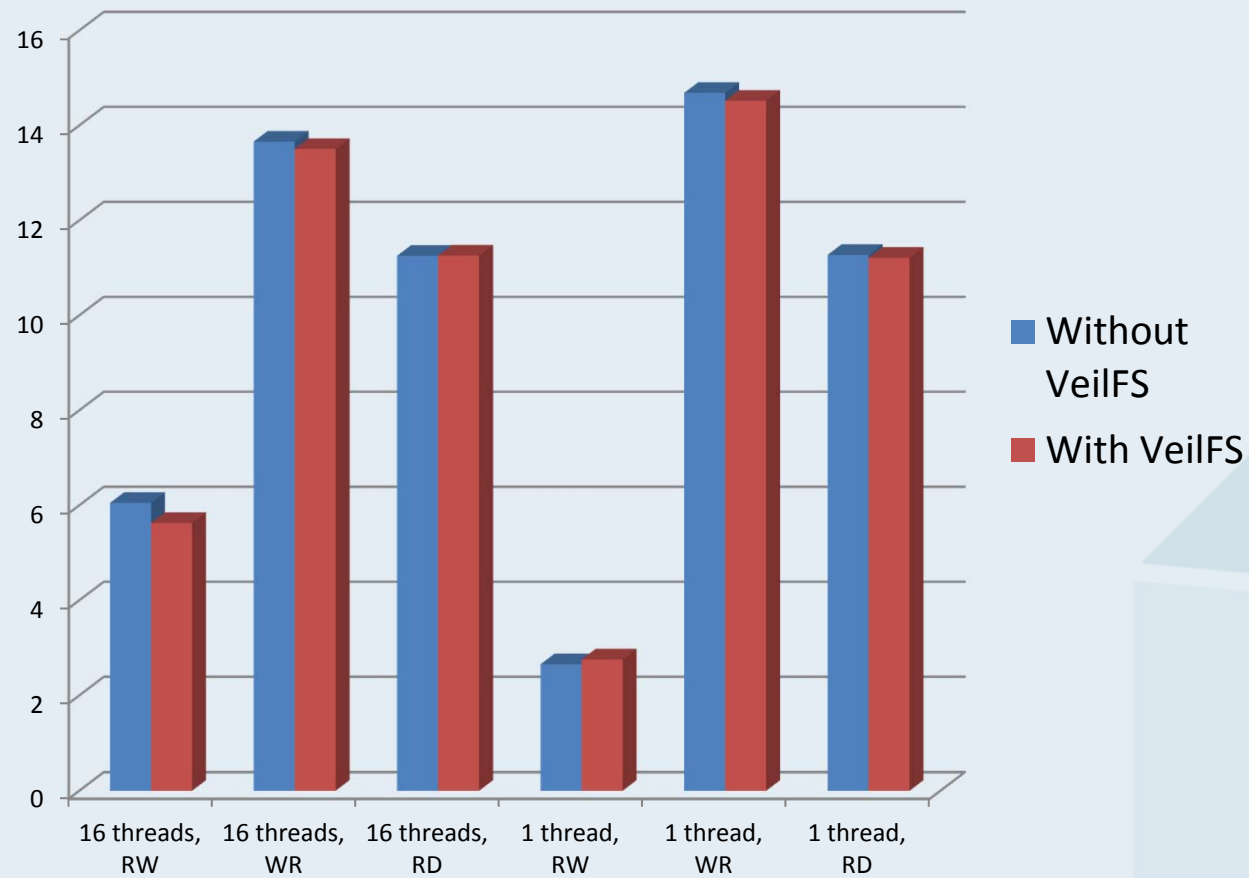
Conclusions

- From user point of view the system:
  - is easy in use,
  - provides unified access to data,
  - supports publication and sharing of data,
  - **overcomes all existing barriers.**

- VeilFS has also second, not shown today face – it helps administrators during the data and infrastructure management**.**

- First ongoing deployment in PL-Grid base at one site resources:
  - multi-site deployment is planned after further tests.