BioQuant's Large Scale Data Facility

Usage Patterns, Problems, Solutions

Christian Thiemann Marc Hemberger

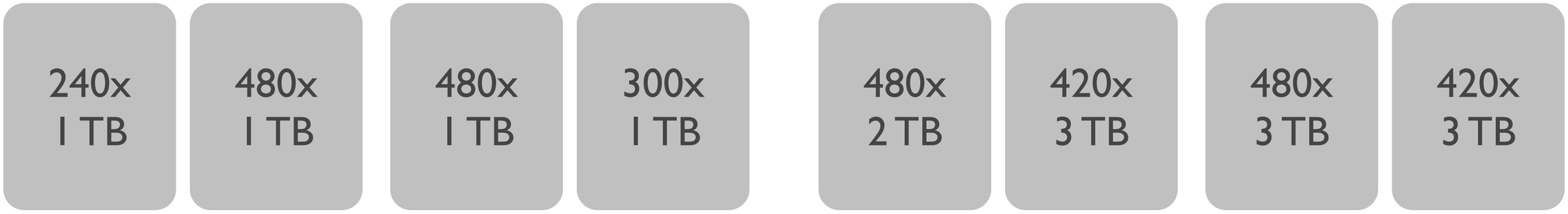
BioQuant IT, Ruprecht-Karls-Universität Heidelberg

LSDMA 2013 Spring Meeting

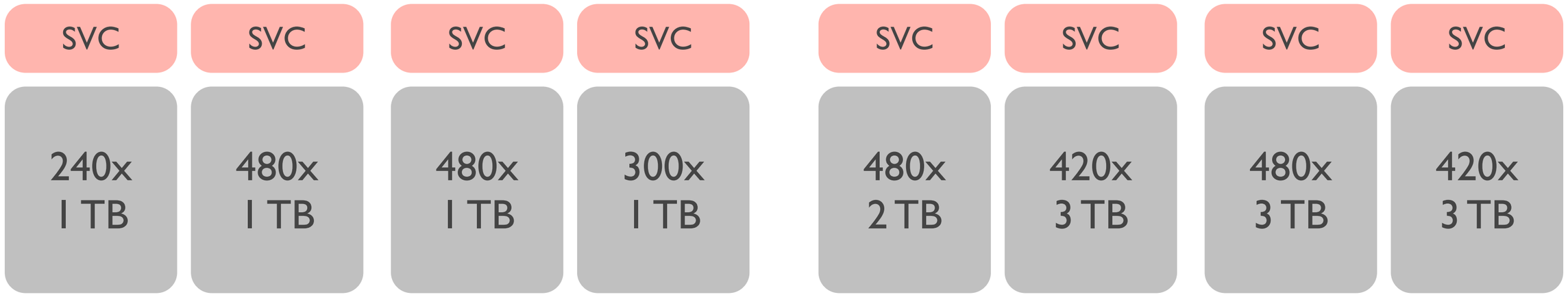
BioQuant LSDF
IBM SONAS

BioQuant LSDF

IBM SONAS



BioQuant LSDF
IBM SONAS



BioQuant LSDF
IBM SONAS

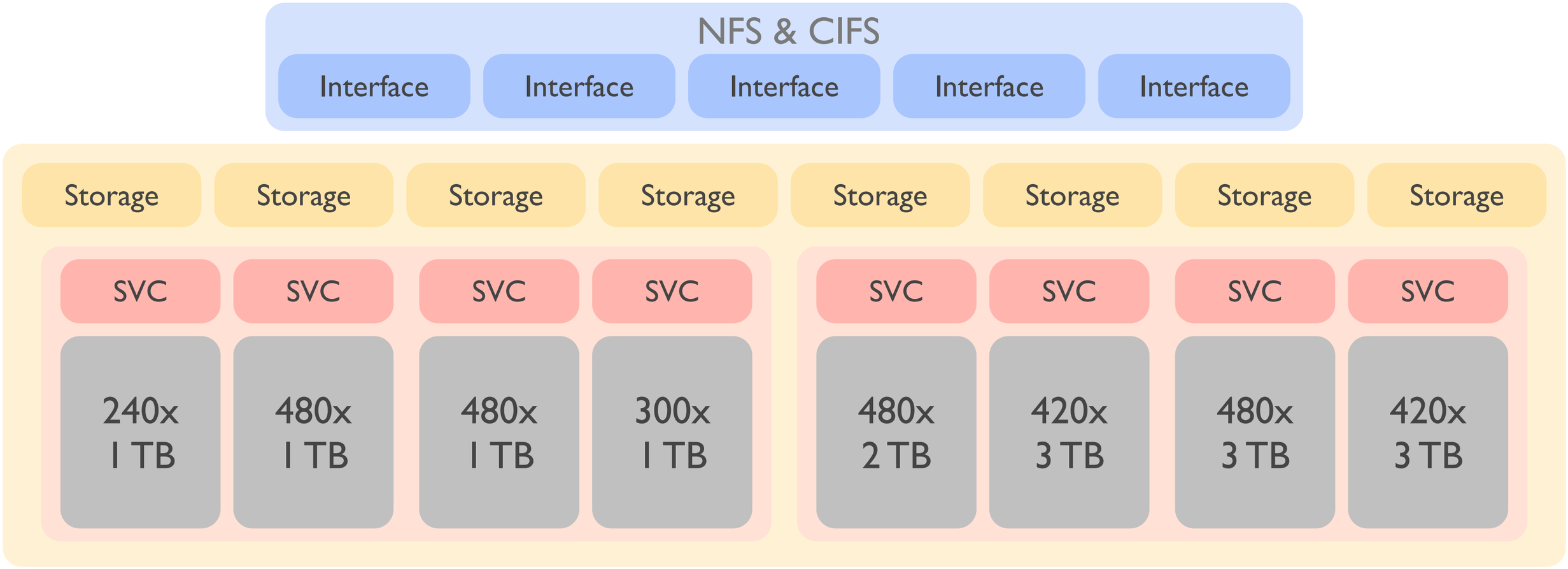
SVC	SVC	SVC	SVC
240x 1 TB	480x 1 TB	480x 1 TB	300x 1 TB

SVC	SVC	SVC	SVC
480x 2 TB	420x 3 TB	480x 3 TB	420x 3 TB

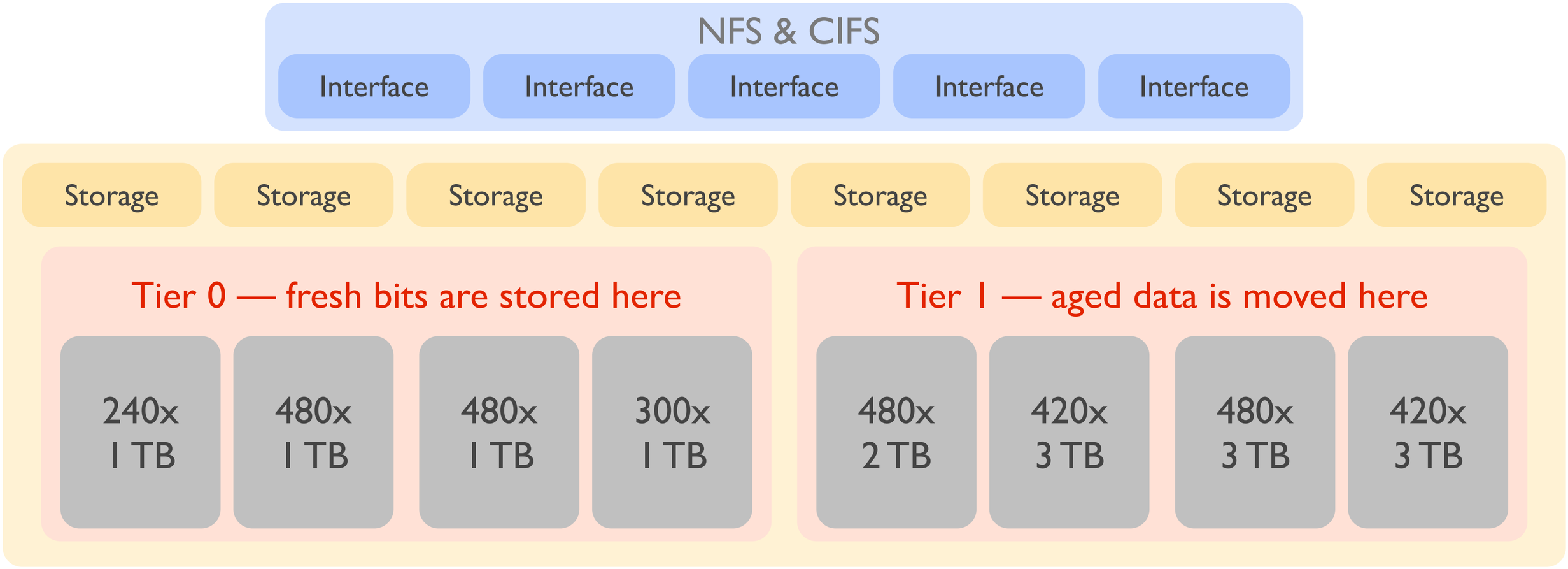
BioQuant LSDF
IBM SONAS

Storage	Storage	Storage	Storage	Storage	Storage	Storage	Storage
SVC	SVC	SVC	SVC	SVC	SVC	SVC	SVC
240x 1 TB	480x 1 TB	480x 1 TB	300x 1 TB	480x 2 TB	420x 3 TB	480x 3 TB	420x 3 TB

BioQuant LSDF
IBM SONAS



BioQuant LSDF
IBM SONAS



BioQuant LSDF
IBM SONAS

NFS & CIFS

2x 10 GBit/s — 1.1 GB/s (read) — 700 MB/s (write)

3,300 HDDs in RAID6 (8+2) — 6.1 PB (physical) — 4.3 PB (logical)

Tier 0 — fresh bits are stored here

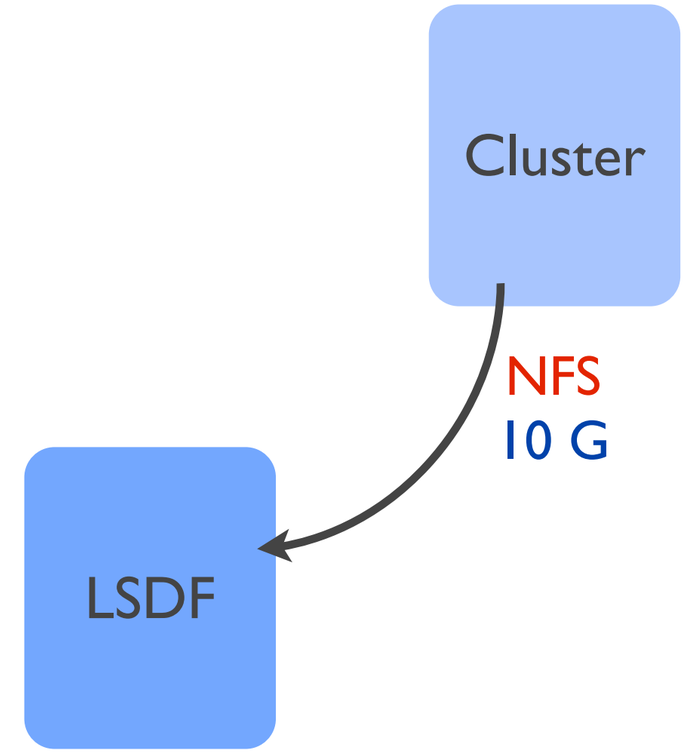
240x 1 TB	480x 1 TB	480x 1 TB	300x 1 TB

Tier 1 — aged data is moved here

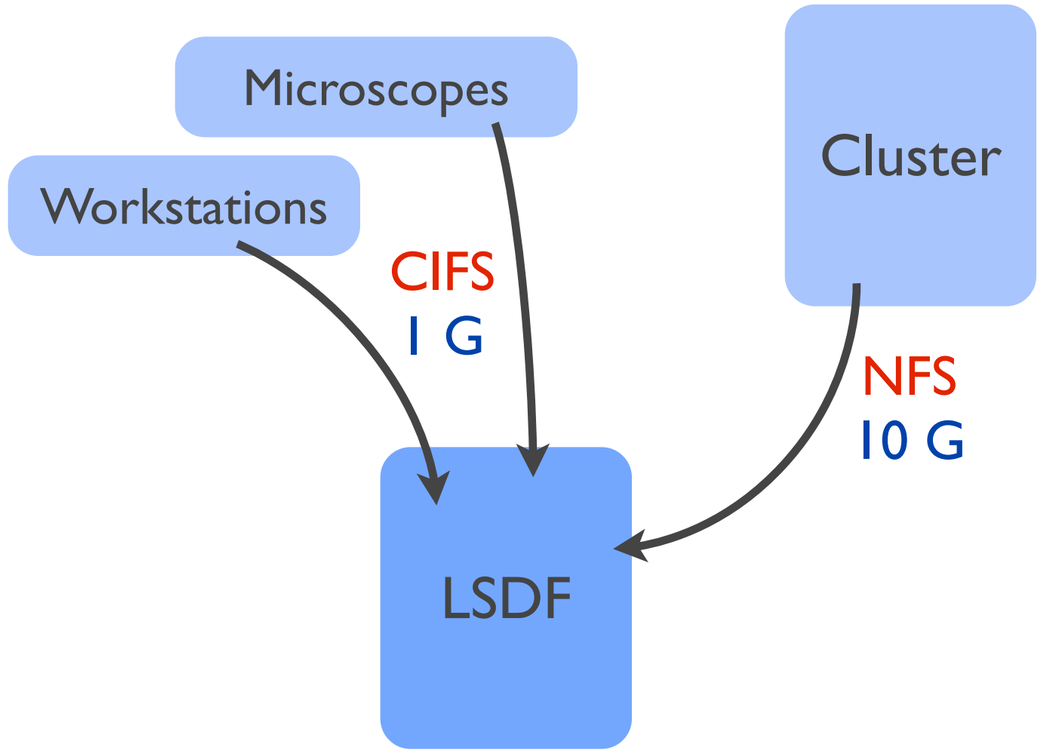
480x 2 TB	420x 3 TB	480x 3 TB	420x 3 TB

BioQuant LSDF

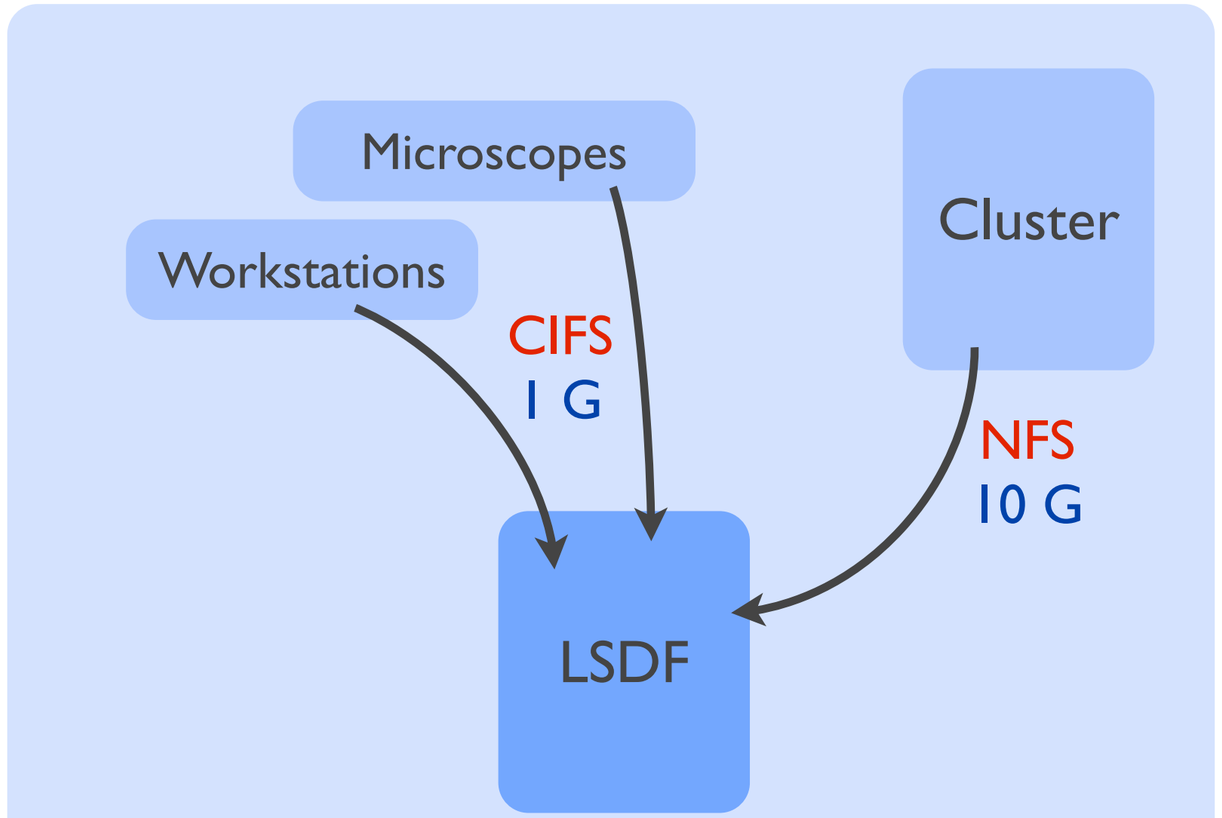
BioQuant LSDF



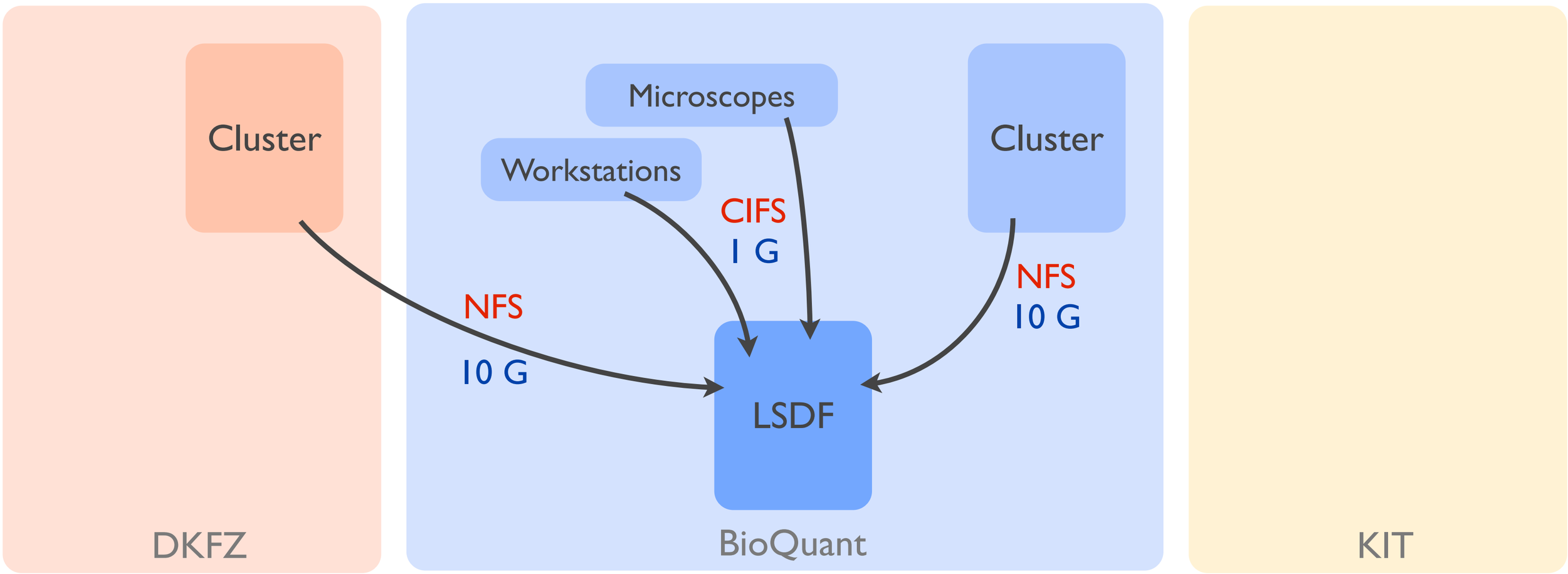
BioQuant LSDF



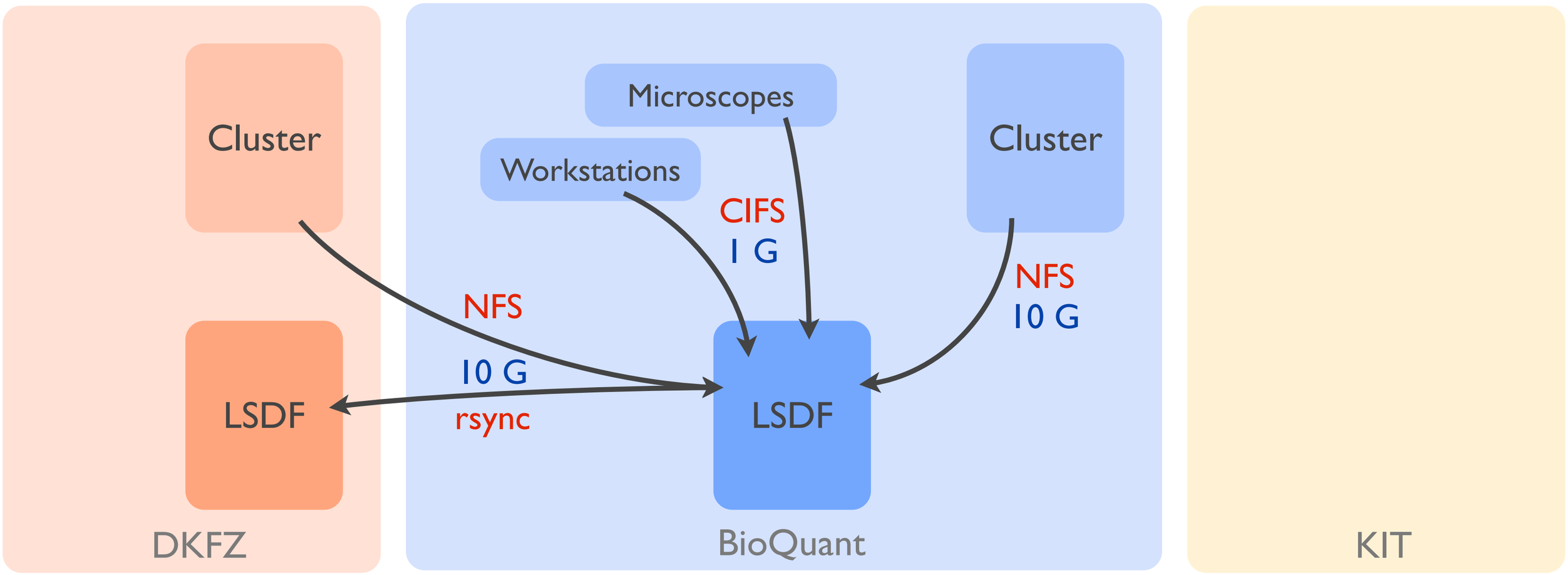
BioQuant LSDF



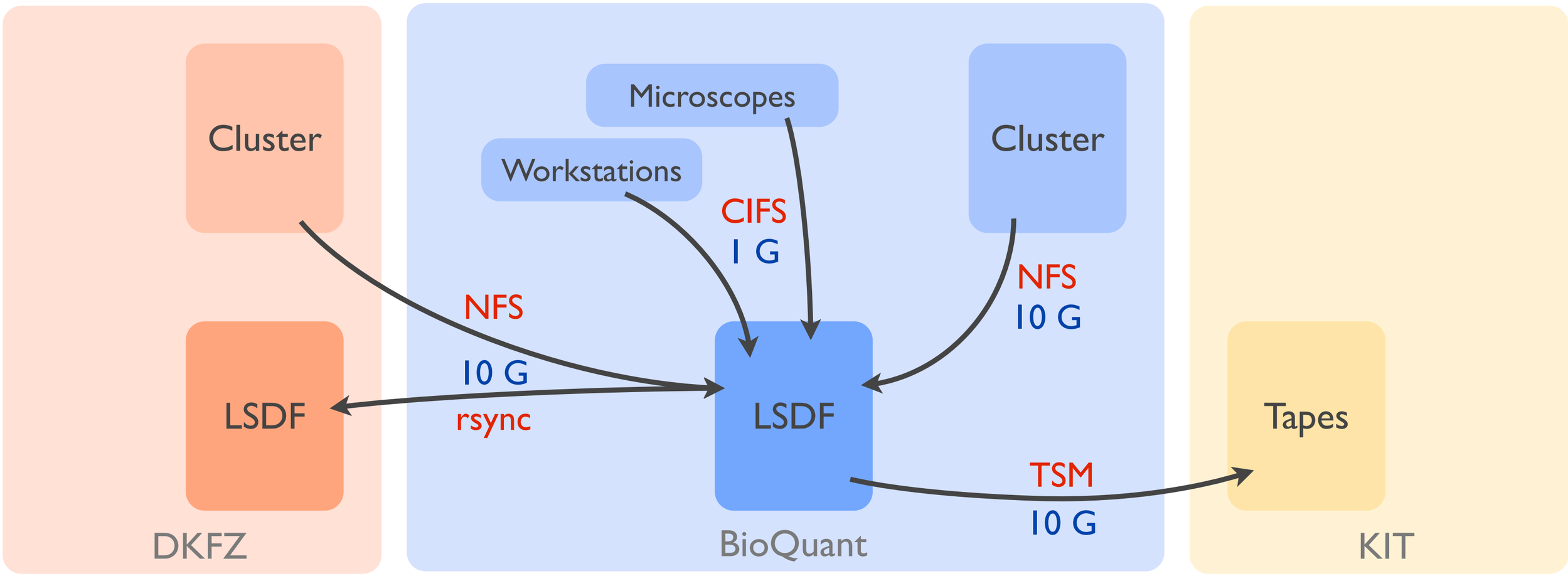
BioQuant LSDF



BioQuant LSDF



BioQuant LSDF



BioQuant LSDF

Group Shares

Microscopy

Sequencing

BioQuant LSDF

Group Shares	~ 80 TB
Microscopy	~ 70 TB
Sequencing	1.3 PB

BioQuant LSDF

Group Shares	~ 80 TB
Microscopy	~ 70 TB
Sequencing	1.3 PB
total	1.6 PB

BioQuant LSDF

Group Shares

Microscopy

Sequencing

BioQuant LSDF

Group Shares

RNA interference

Microscopy

Sequencing

BioQuant LSDF

Group Shares

RNA interference

Microscopy

image size ~ 2 MB

Sequencing

BioQuant LSDF

Group Shares

RNA interference

Microscopy

image size ~ 2 MB

Sequencing

~ 2,000 images per sample / run

future runs up to 70,000 images

BioQuant LSDF

Group Shares

RNA interference

Microscopy

image size ~ 2 MB

Sequencing

~ 2,000 images per sample / run

future runs up to 70,000 images

many samples per study / experiment

"typical" throughput: 20,000 images per day

peak throughput: 3,000,000 images (3 TB) in three weeks

BioQuant LSDF

Group Shares

RNA interference

Microscopy

image size ~ 2 MB

Sequencing

~ 2,000 images per sample / run

future runs up to 70,000 images

many samples per study / experiment

"typical" throughput: 20,000 images per day

peak throughput: 3,000,000 images (3 TB) in three weeks

many small files

BioQuant LSDF

Group Shares

Microscopy

Sequencing

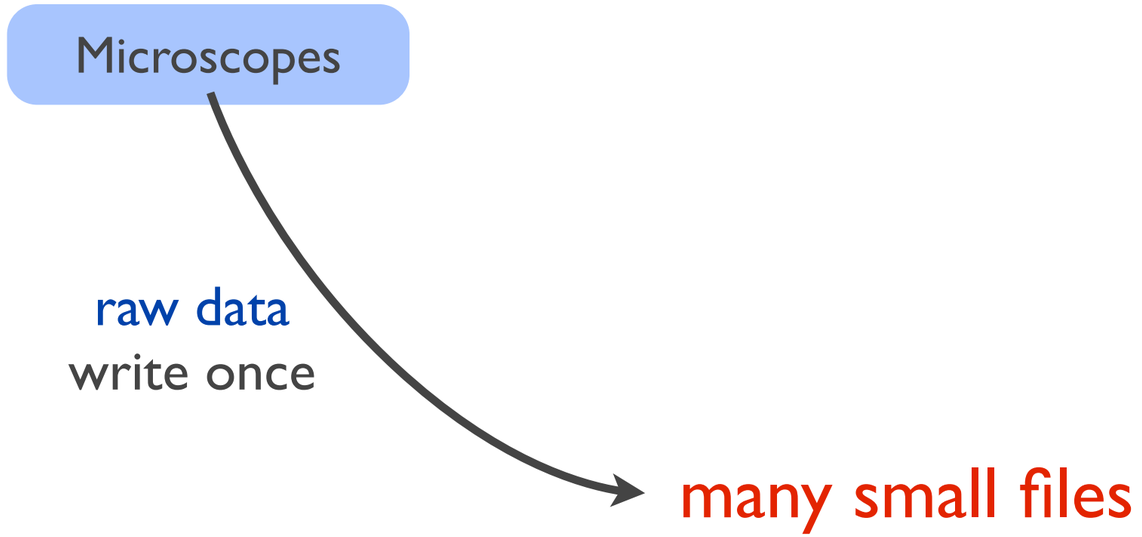
many small files

BioQuant LSDF

Group Shares

Microscopy

Sequencing

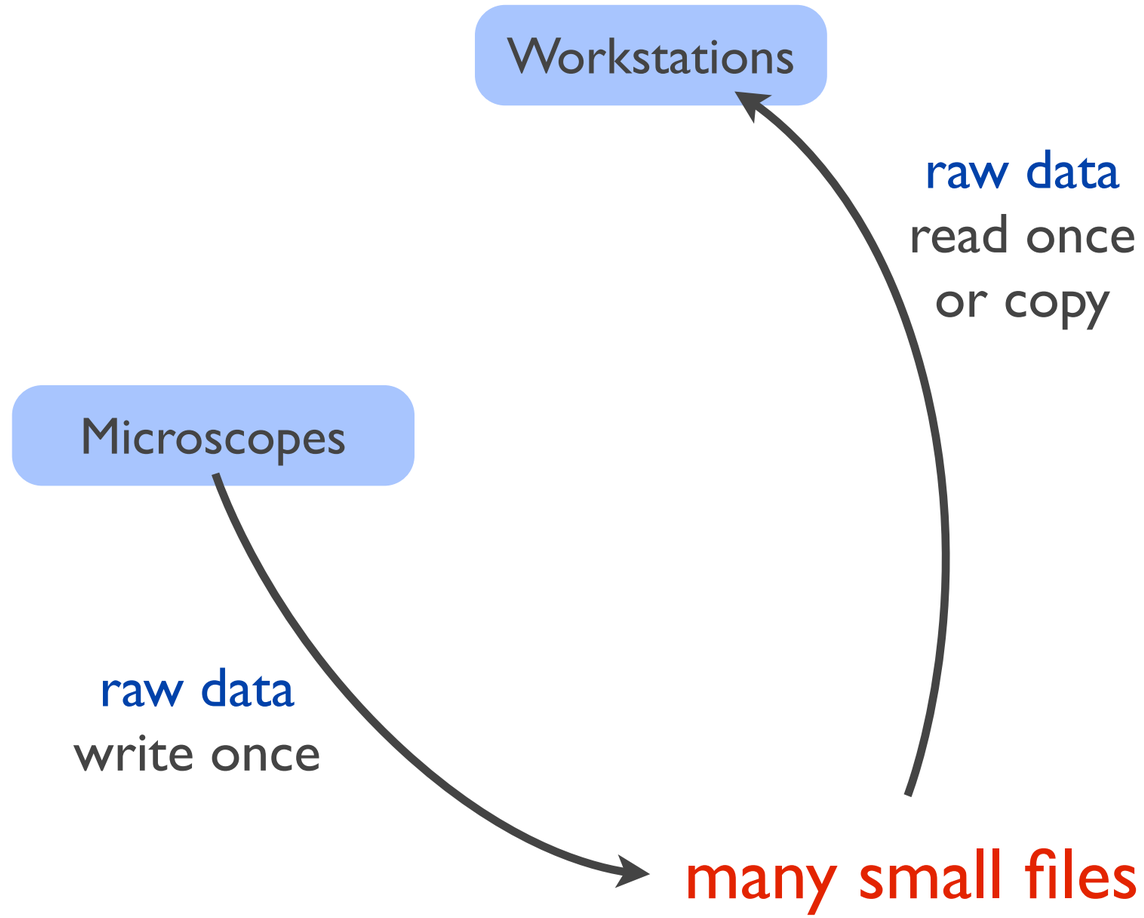


BioQuant LSDF

Group Shares

Microscopy

Sequencing

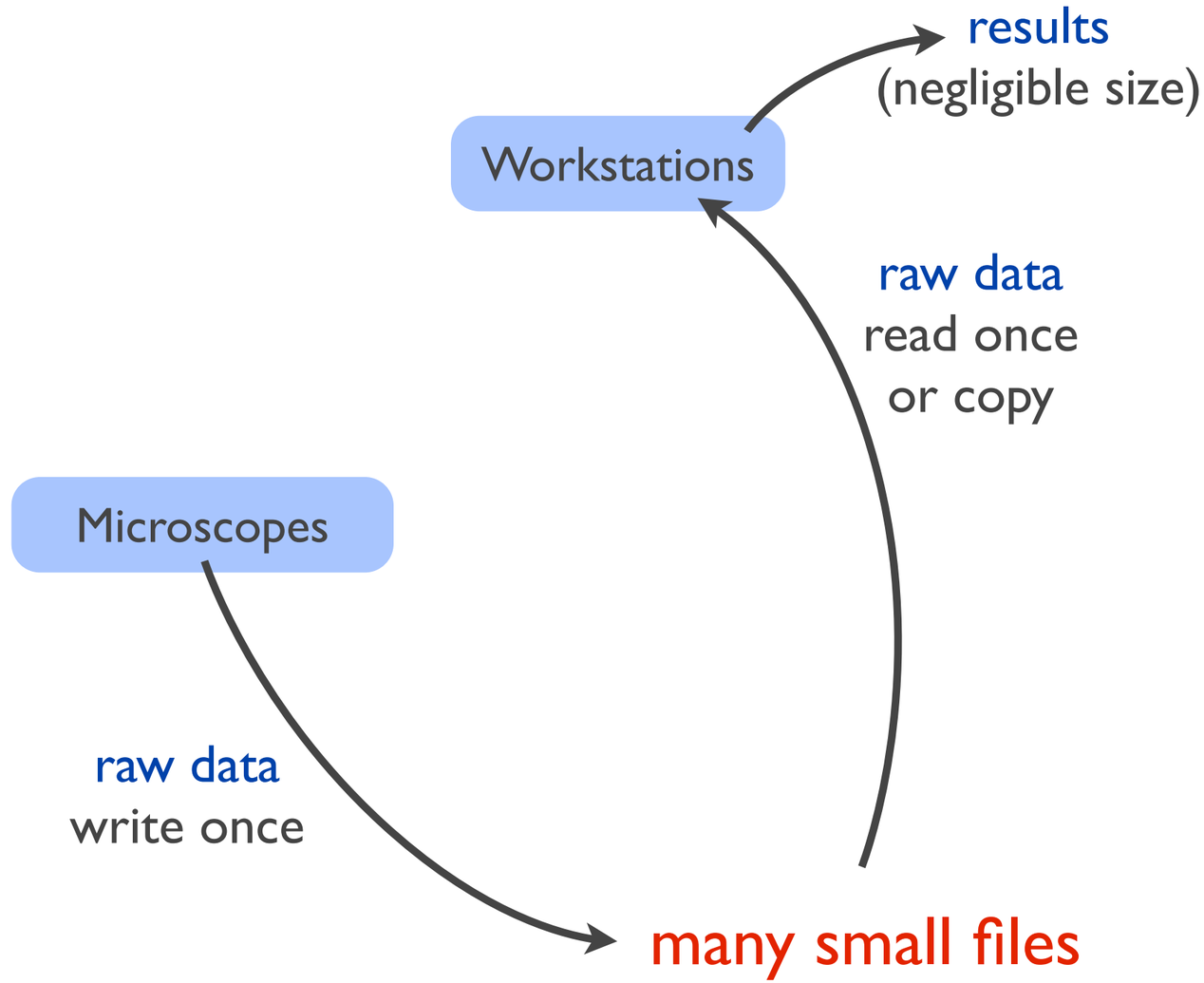


BioQuant LSDF

Group Shares

Microscopy

Sequencing

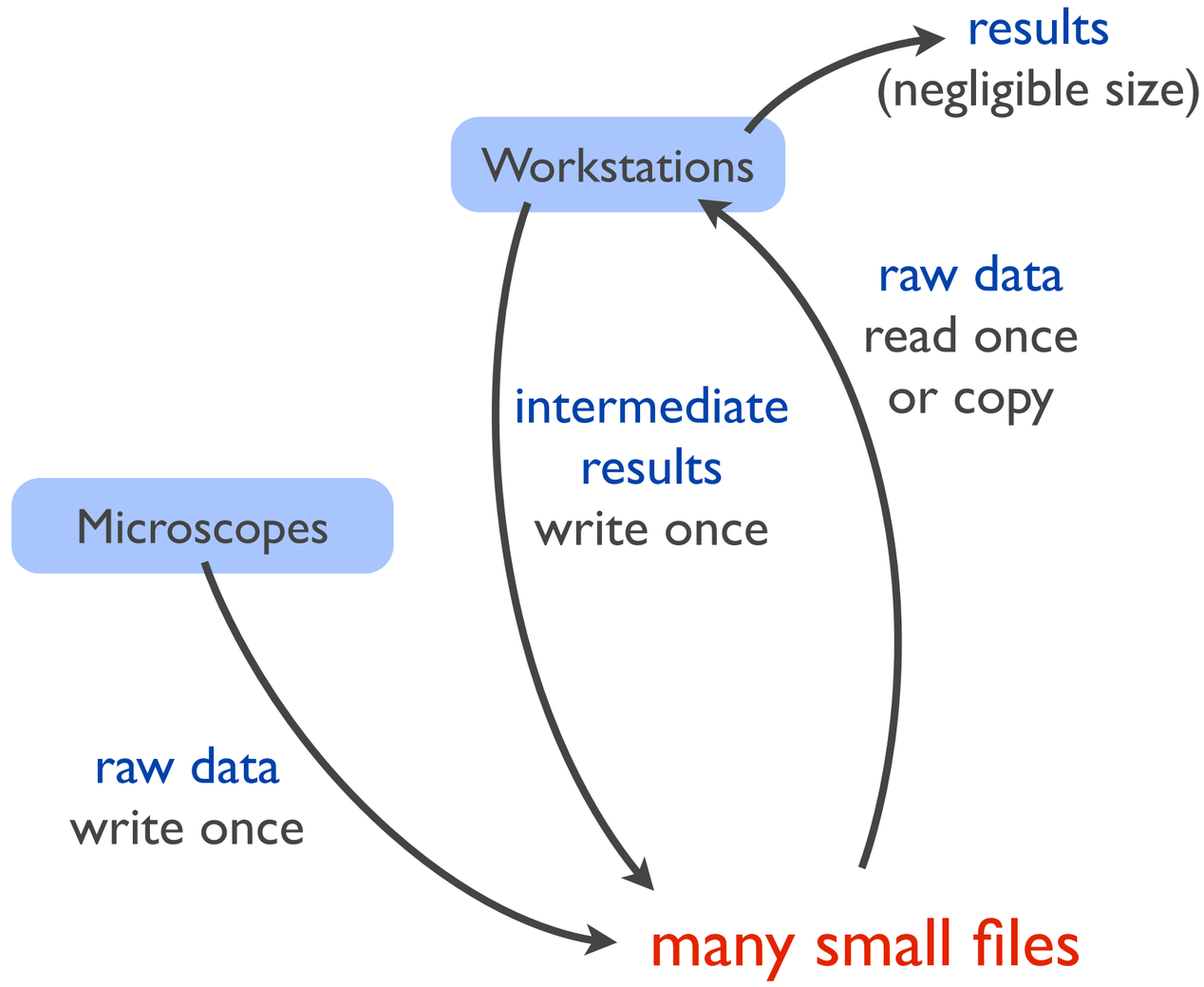


BioQuant LSDF

Group Shares

Microscopy

Sequencing



BioQuant LSDF

Group Shares

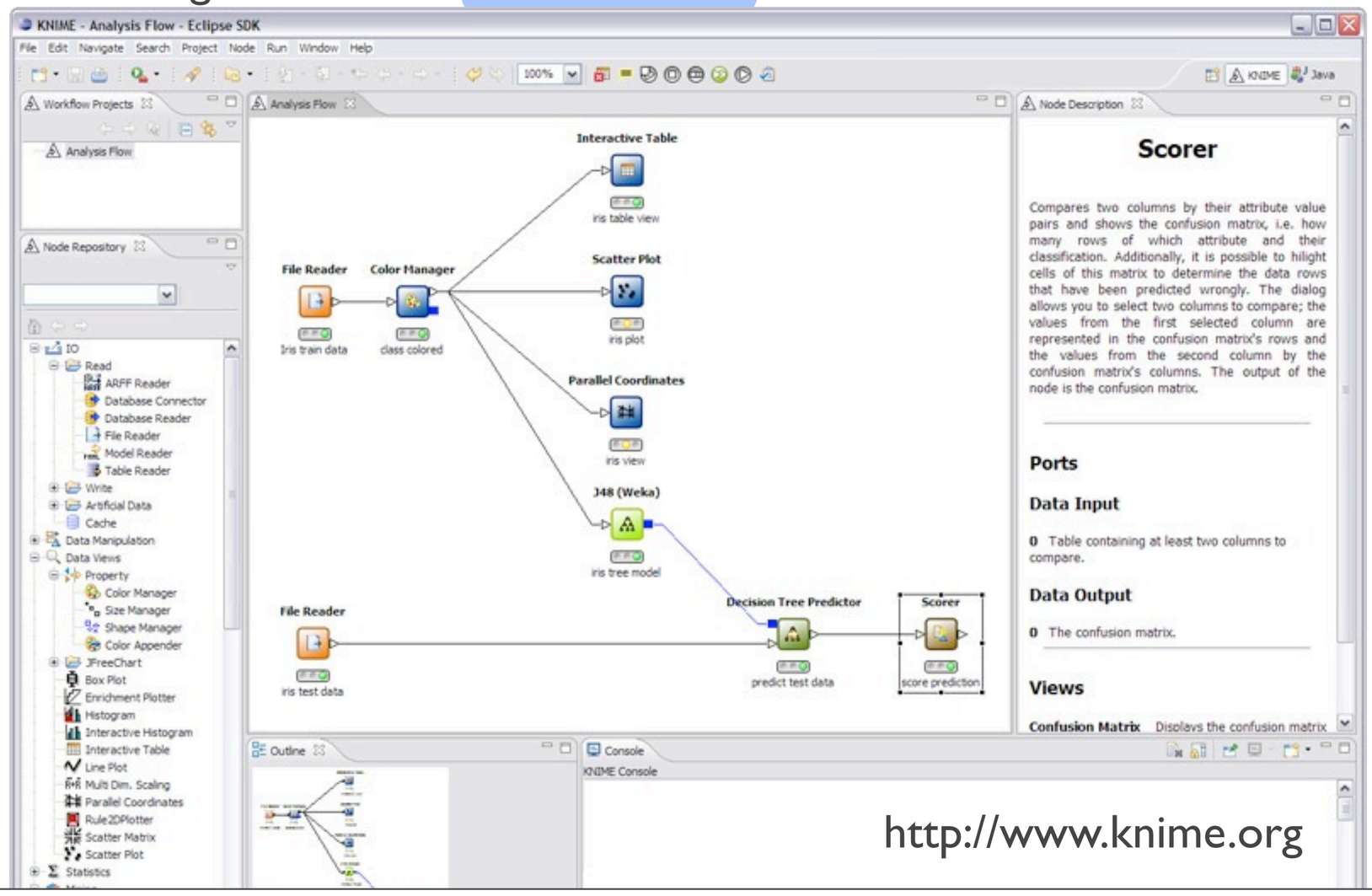
Microscopy

Sequencing

e.g. **KNIME**

Workstations

results
(negligible size)



http://www.knime.org

BioQuant LSDF

Group Shares

Microscopy

Sequencing

BioQuant LSDF

Group Shares

Microscopy

Sequencing

International Cancer Genome Consortium

BioQuant LSDF

Group Shares

Microscopy

Sequencing

International Cancer Genome Consortium

~ 200 GB raw data per genome

BioQuant LSDF

Group Shares

Microscopy

Sequencing

International Cancer Genome Consortium

~ 200 GB raw data per genome

~ 2 TB after processing

intermediate results are retained

BioQuant LSDF

Group Shares

Microscopy

Sequencing

International Cancer Genome Consortium

~ 200 GB raw data per genome

~ 2 TB after processing

intermediate results are retained

throughput: 10–20 genomes per week

BioQuant LSDF

Group Shares

Microscopy

Sequencing

International Cancer Genome Consortium

~ 200 GB raw data per genome

~ 2 TB after processing

intermediate results are retained

throughput: 10–20 genomes per week

total ~ 2 PB of data

~ 1 PB at BioQuant LSDF

BioQuant LSDF

Group Shares

Microscopy

Sequencing

International Cancer Genome Consortium

~ 200 GB raw data per genome

~ 2 TB after processing

intermediate results are retained

throughput: 10–20 genomes per week

total ~ 2 PB of data

~ 1 PB at BioQuant LSDF

few large files

BioQuant LSDF

Group Shares

Microscopy

Sequencing

International Cancer Genome Consortium

~ 200 GB raw data per genome

~ 2 TB after processing

intermediate results are retained

throughput: 10–20 genomes per week

total ~ 2 PB of data

~ 1 PB at BioQuant LSDF

few large files

only one ICGC project stored at BQ-LSDF

BioQuant LSDF

Group Shares

Microscopy

Sequencing

International Cancer Genome Consortium

~ 200 GB raw data per genome

~ 2 TB after processing

intermediate results are retained

throughput: 10–20 genomes per week

total ~ 2 PB of data

~ 1 PB at BioQuant LSDF

few large files

different "views" of same data via symbolic links

BioQuant LSDF

Group Shares

Microscopy

Sequencing

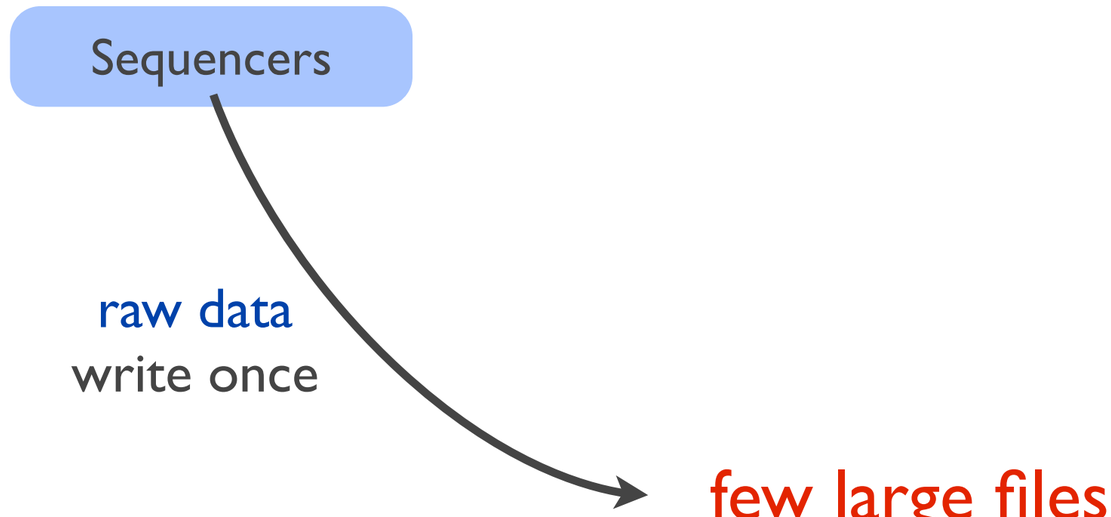
few large files

BioQuant LSDF

Group Shares

Microscopy

Sequencing

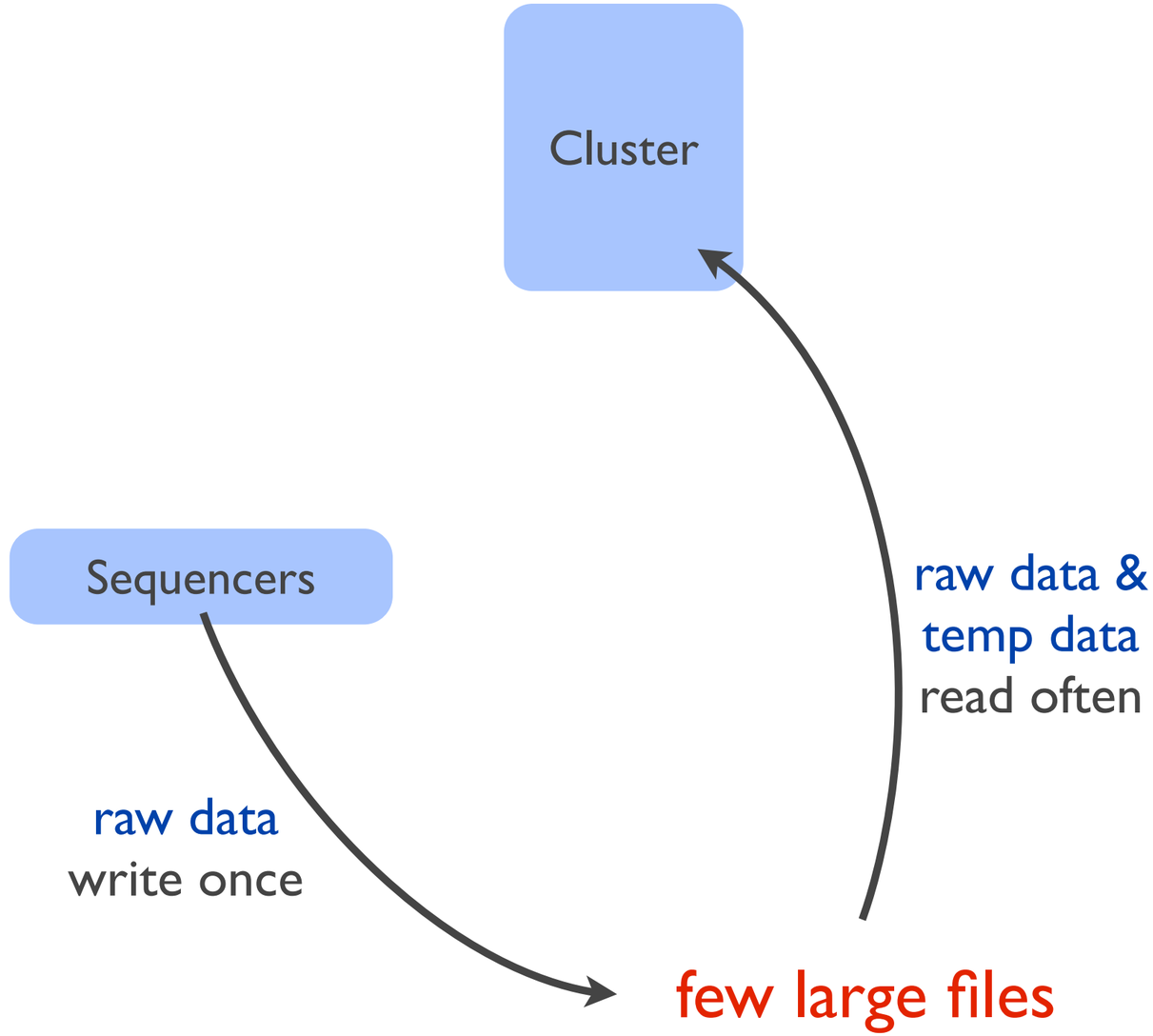


BioQuant LSDF

Group Shares

Microscopy

Sequencing

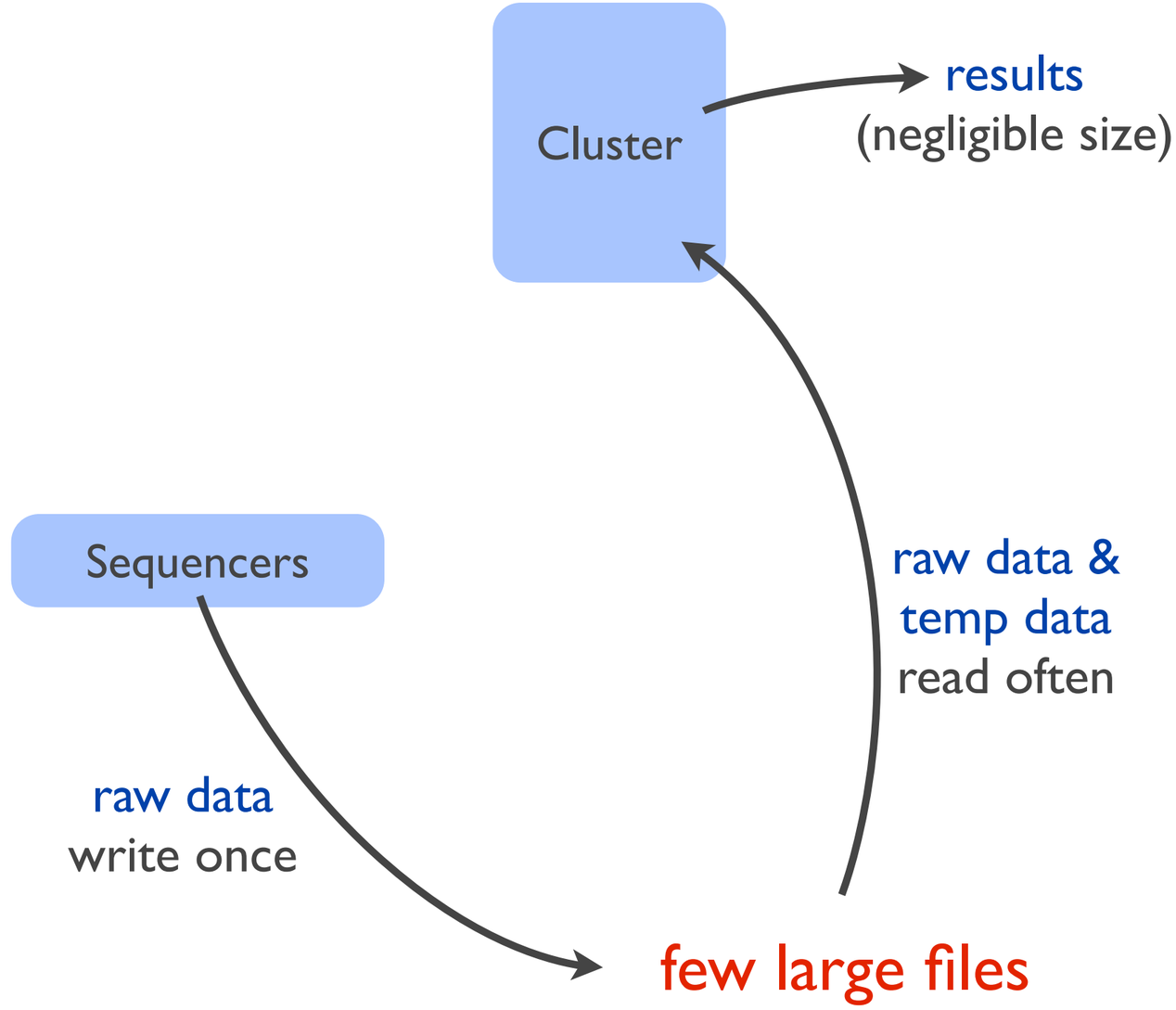


BioQuant LSDF

Group Shares

Microscopy

Sequencing

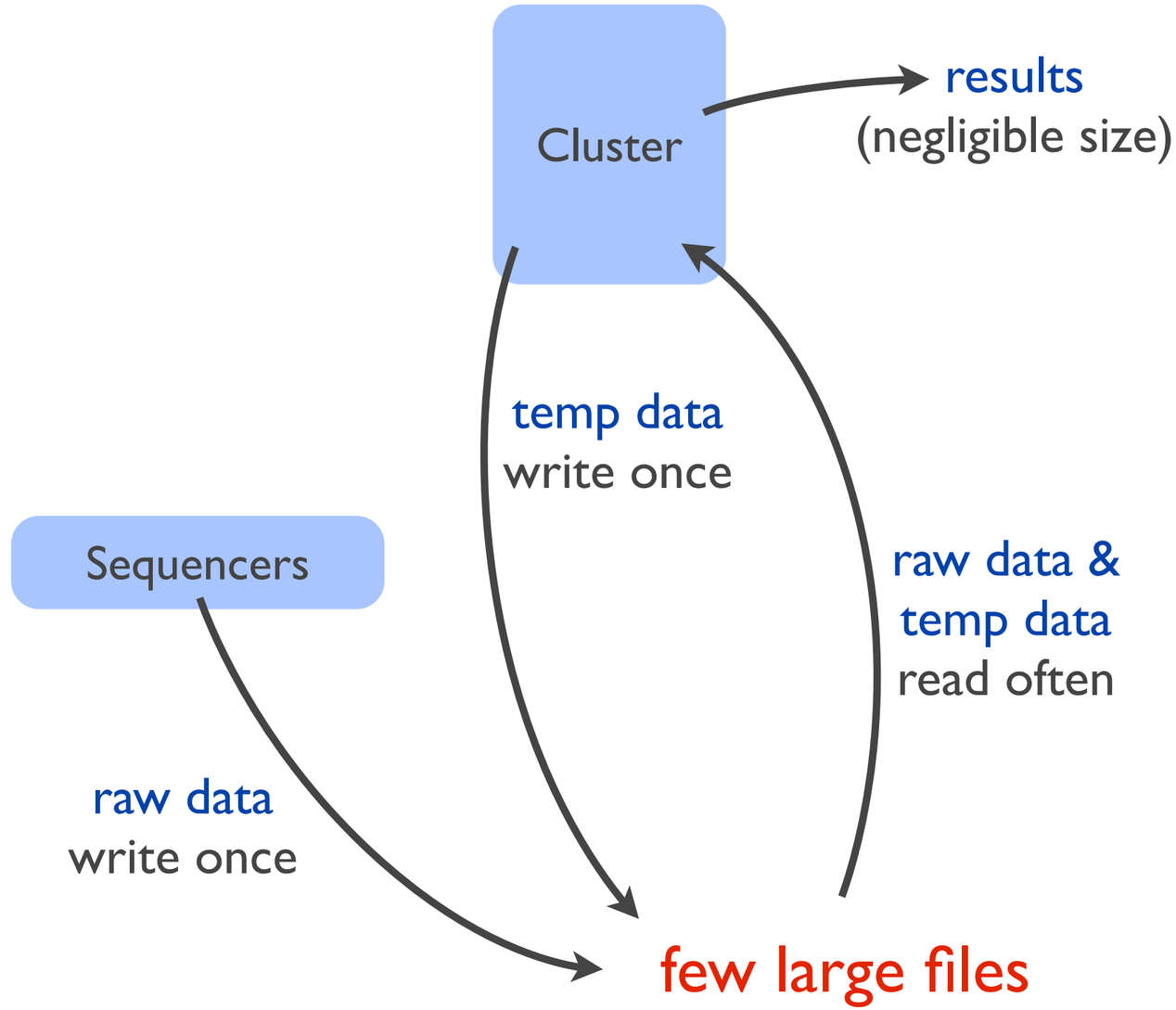


BioQuant LSDF

Group Shares

Microscopy

Sequencing



BioQuant LSDF

Group Shares

Microscopy

Sequencing

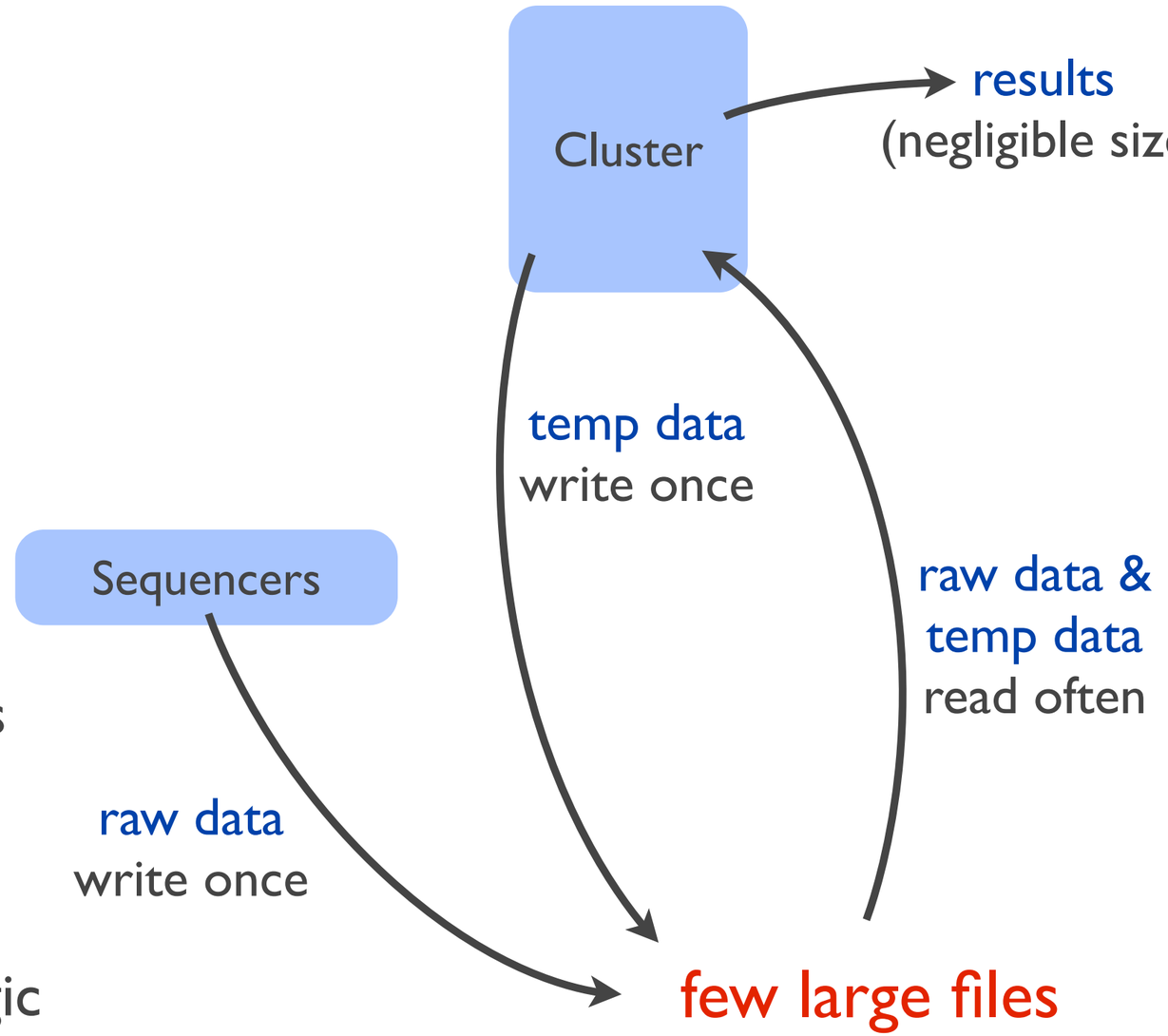
store as provided by sequencers

organize semantically (symlinks)

check for data corruption

perform genome alignment magic

compute standard measures

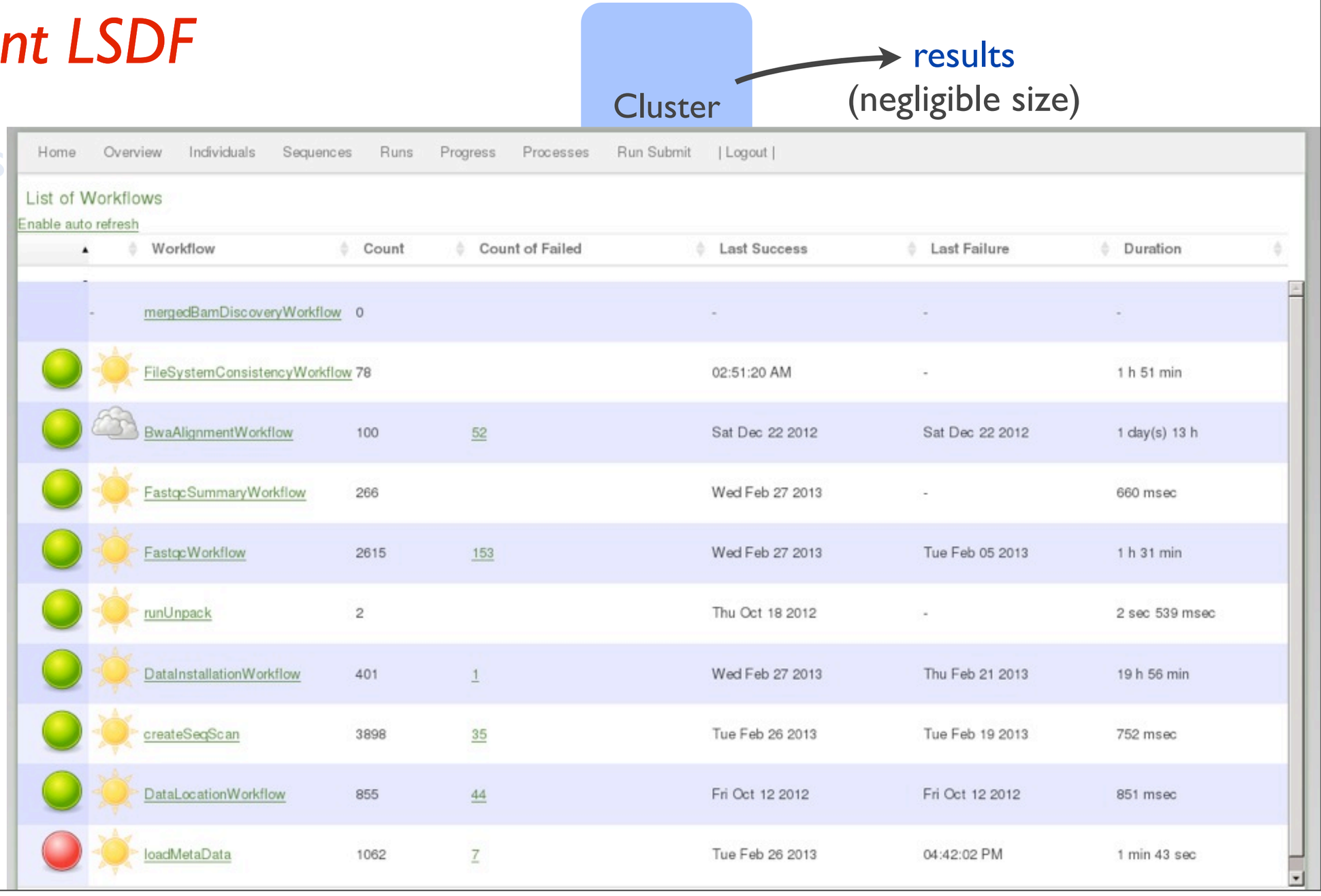


BioQuant LSDF

Group Shares

Microscopy

Sequencing
One Touch Pipeline



Sylwester Radomski, DKFZ

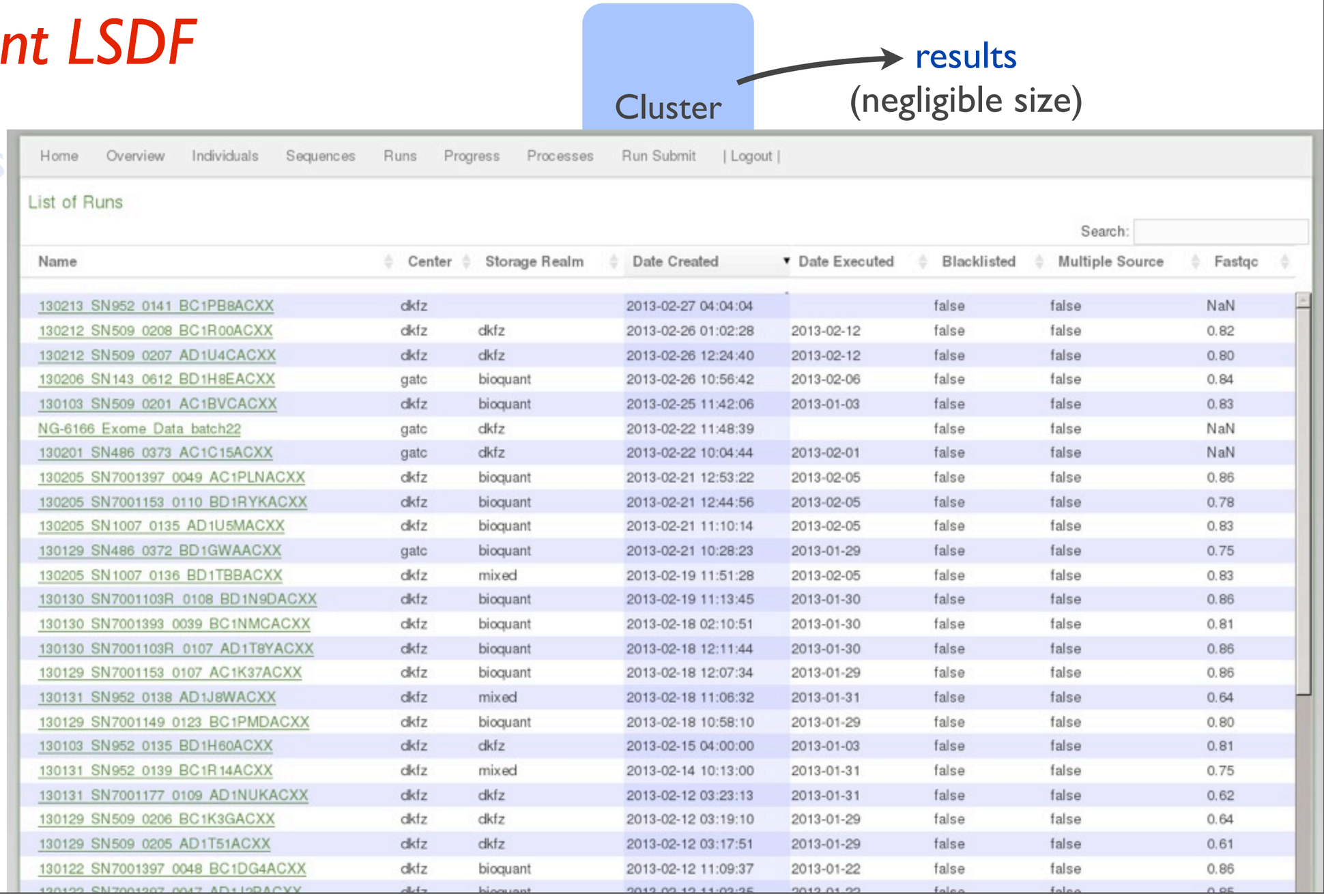
BioQuant LSDF

Group Shares

Microscopy

Sequencing
One Touch Pipeline

Sylwester Radomski, DKFZ



Name	Center	Storage Realm	Date Created	Date Executed	Blacklisted	Multiple Source	Fastqc
130213_SN952_0141_BC1PB8ACXX	dkfz		2013-02-27 04:04:04		false	false	NaN
130212_SN509_0208_BC1R00ACXX	dkfz	dkfz	2013-02-26 01:02:28	2013-02-12	false	false	0.82
130212_SN509_0207_AD1U4CACXX	dkfz	dkfz	2013-02-26 12:24:40	2013-02-12	false	false	0.80
130206_SN143_0612_BD1H8EACXX	gatc	bioquant	2013-02-26 10:56:42	2013-02-06	false	false	0.84
130103_SN509_0201_AC1BVCACXX	dkfz	bioquant	2013-02-25 11:42:06	2013-01-03	false	false	0.83
NG-6166_Exome_Data_batch22	gatc	dkfz	2013-02-22 11:48:39		false	false	NaN
130201_SN486_0373_AC1C15ACXX	gatc	dkfz	2013-02-22 10:04:44	2013-02-01	false	false	NaN
130205_SN7001397_0049_AC1PLNACXX	dkfz	bioquant	2013-02-21 12:53:22	2013-02-05	false	false	0.86
130205_SN7001153_0110_BD1RYKACXX	dkfz	bioquant	2013-02-21 12:44:56	2013-02-05	false	false	0.78
130205_SN1007_0135_AD1U5MACXX	dkfz	bioquant	2013-02-21 11:10:14	2013-02-05	false	false	0.83
130129_SN486_0372_BD1GWAACXX	gatc	bioquant	2013-02-21 10:28:23	2013-01-29	false	false	0.75
130205_SN1007_0136_BD1TBBACXX	dkfz	mixed	2013-02-19 11:51:28	2013-02-05	false	false	0.83
130130_SN7001103R_0108_BD1N9DACXX	dkfz	bioquant	2013-02-19 11:13:45	2013-01-30	false	false	0.86
130130_SN7001393_0039_BC1NMCACXX	dkfz	bioquant	2013-02-18 02:10:51	2013-01-30	false	false	0.81
130130_SN7001103R_0107_AD1T8YACXX	dkfz	bioquant	2013-02-18 12:11:44	2013-01-30	false	false	0.86
130129_SN7001153_0107_AC1K37ACXX	dkfz	bioquant	2013-02-18 12:07:34	2013-01-29	false	false	0.86
130131_SN952_0138_AD1J8WACXX	dkfz	mixed	2013-02-18 11:06:32	2013-01-31	false	false	0.64
130129_SN7001149_0123_BC1PMDACXX	dkfz	bioquant	2013-02-18 10:58:10	2013-01-29	false	false	0.80
130103_SN952_0135_BD1H60ACXX	dkfz	dkfz	2013-02-15 04:00:00	2013-01-03	false	false	0.81
130131_SN952_0139_BC1R14ACXX	dkfz	mixed	2013-02-14 10:13:00	2013-01-31	false	false	0.75
130131_SN7001177_0109_AD1NUKACXX	dkfz	dkfz	2013-02-12 03:23:13	2013-01-31	false	false	0.62
130129_SN509_0206_BC1K3GACXX	dkfz	dkfz	2013-02-12 03:19:10	2013-01-29	false	false	0.64
130129_SN509_0205_AD1T51ACXX	dkfz	dkfz	2013-02-12 03:17:51	2013-01-29	false	false	0.61
130122_SN7001397_0048_BC1DG4ACXX	dkfz	bioquant	2013-02-12 11:09:37	2013-01-22	false	false	0.86

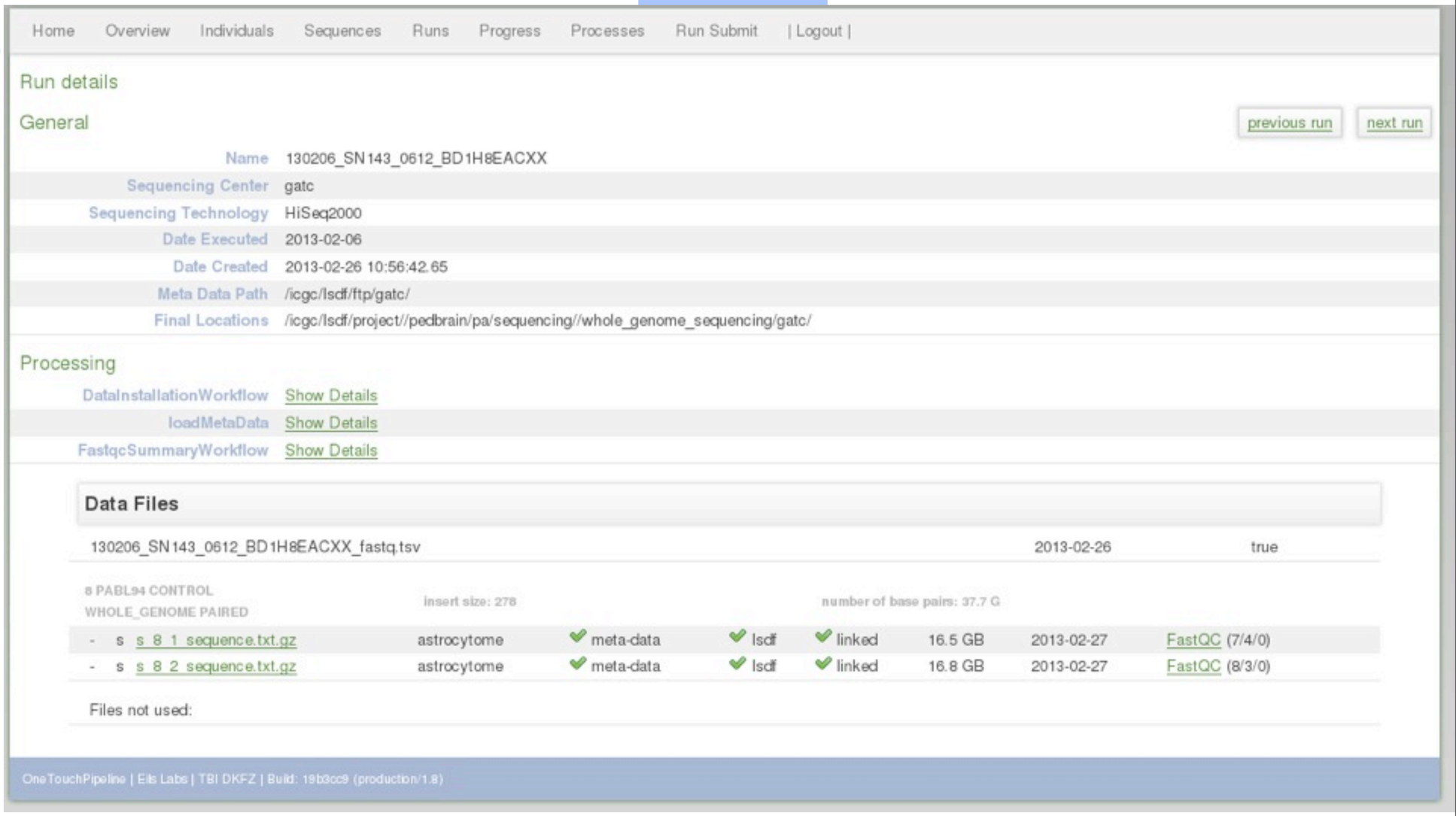
BioQuant LSDF

Cluster → results
(negligible size)

Group Shares

Microscopy

Sequencing
One Touch Pipeline



Sylwester Radomski, DKFZ

Showing details of jobs

BioQuan

Group Shares

Microscopy

Sequencing
Cluster Monitoring

Rolf Kabbe, DKFZ

Tbi-cluster

Logged in : Rolf Kabbe | Log out

Active jobs | Completed jobs | Nodes | Job categories | Projects | Preferences | Admin | About

Showing details for job : 929752

Use as MinJobId | Use as MaxJobId

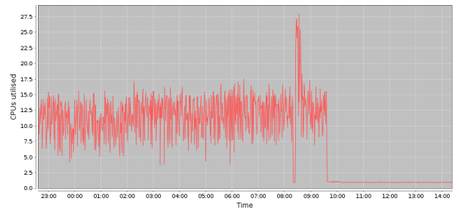
Modify this job

Delete Job | Purge Job | Extend Walltime

Requested and used resources

Attribute	Value
Owner	wangq
Name	LINA463K_LINA463K_merge_removedup
Submit arguments	-o /icgc/lsdf/mb/analysis/LiNA/analysis/DKFZLSDF/results_per_pid/LINA463K_logs -j oe -M q.wang@dkfz.de -N LINA463K_LINA463K_merge_removedup -v CONFIG_FILE=/home/wangq/SVNrepository/ngs2/branches/pipelines/QCPipeline_BSseq/DKFZLSDF_LINA_CONFIG/pipelineParams_TEST.txt,pid=LINA463K,INPUT_BAMS=:/icgc/lsdf/mb/analysis/LiNA/analysis/DKFZLSDF/results_per_pid/LINA463K_temp/alignment/LINA463K_run121121_SN7001153_0093_BD1H88ACXX_2.rmdup.bam:/icgc/lsdf/mb/analysis/LiNA/analysis/DKFZLSDF/results_per_pid/LINA463K_temp/alignment/LINA463K_run121127_SN952_0127_AD1HCEACXX_6.rmdup.bam:/icgc/lsdf/mb/analysis/LiNA/analysis/DKFZLSDF/results_per_pid/LINA463K_temp/alignment/LINA463K_run130117_SN7001149_0121_BD1J20ACXX_2.rmdup.bam:/icgc/lsdf/mb/analysis/LiNA/analysis/DKFZLSDF/results_per_pid/LINA463K_temp/alignment/LINA463K_run130117_SN7001149_0121_BD1J20ACXX_3.rmdup.bam:/icgc/lsdf/mb/analysis/LiNA/analysis/DKFZLSDF/results_per_pid/LINA463K_temp/alignment/LINA463K_run130117_SN7001149_0121_BD1J20ACXX_4.rmdup.bam,OUTPUT_BAM=/icgc/lsdf/mb/analysis/LiNA/analysis/DKFZLSDF/results_per_pid/LINA463K_temp/alignment/LINA463K_5lanes_merged.rmdup.bam,METRICS_FILE=/icgc/lsdf/mb/analysis/LiNA/analysis/DKFZLSDF/results_per_pid/LINA463K_temp/alignment/LINA463K_5lanes_merged.rmdup.bam.dupmark_metrics.txt /home/wangq/SVNrepository/ngs2/branches/pipelines/QCPipeline_BSseq/analysis_tools/MergeBAM_MarkDup.sh
Queue	verylong
Job Type	BATCH
Status	RUNNING
Queued	Tue Feb 5 22:37:09 2013
Started	Tue Feb 5 22:37:09 2013
Ended	--:--:--
Exit Code	
Start count	1
Exec Host	tbi-dsx08
Requested walltime	50:00:00
Elapsed walltime	15:45:47
CPU time	135:19:48
CPU time / cores	135:19:48
CPUs avg. utilised	8.59
Cores	1
Memory Requested	5.0 GB
Maximum Memory Used	4.4 GB
Memory Efficiency	0.89
File Size	250.0 GB
Job Category	wangq:LINA463K-LINA463K-merge-removedup:cores=1:mem=.5.0 GB:walltime=50:00:00

CPU profile



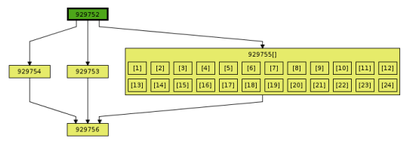
Download profile

Dependencies In	Dependencies Out
	ACTIVE BEFOREOK 929753
	ACTIVE BEFOREOK 929755[]
	ACTIVE BEFOREOK 929754

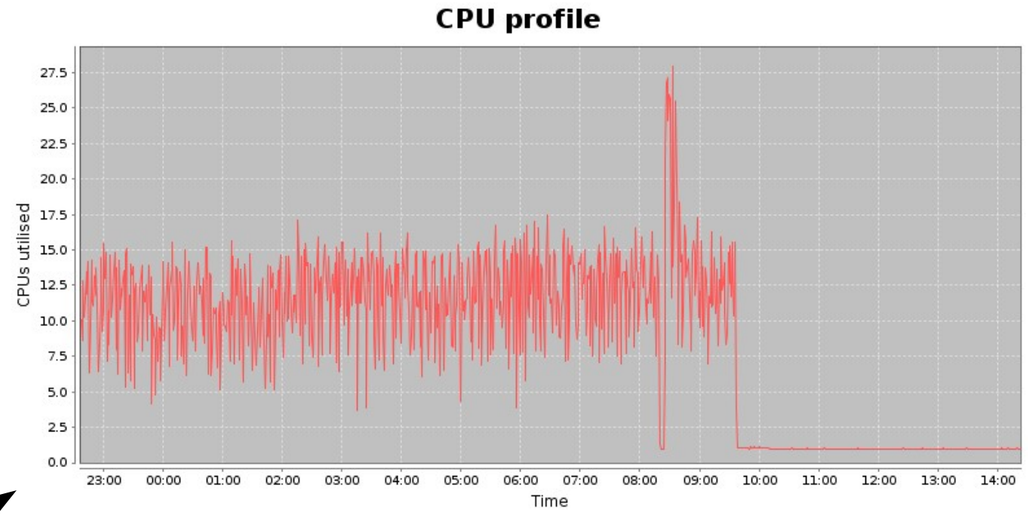
Warnings

CPU	SEVERE	More CPU-time used (135:19:48) than expected (15:45:47), ratio : 8.59
SPECIFICATION	INFO	PBS will prevent that this job can create files larger than 250.0 GB

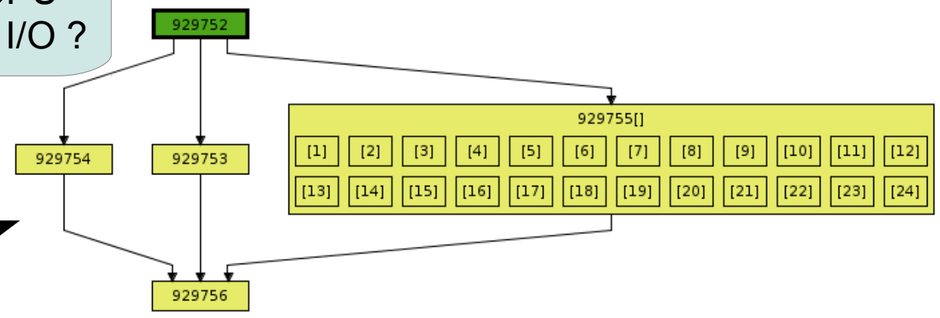
Dependency graph



CPU profile



Drop of CPU
Caused by I/O ?



Dependency-graph
Shows progress
of workflows

BioQua

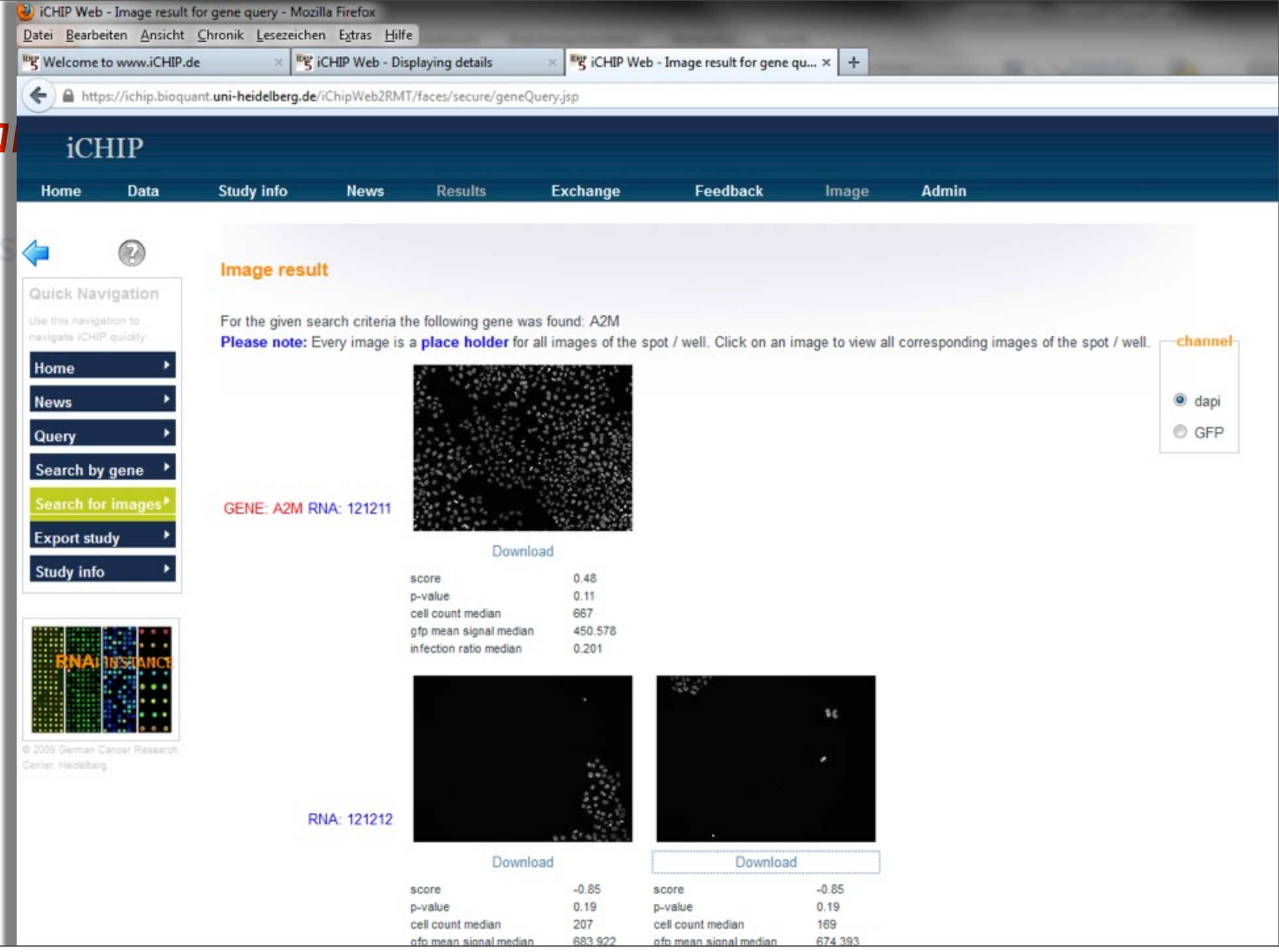
Group Shares

Microscopy

Sequencing

Results Database

Jürgen Eils, DKFZ



BioQua

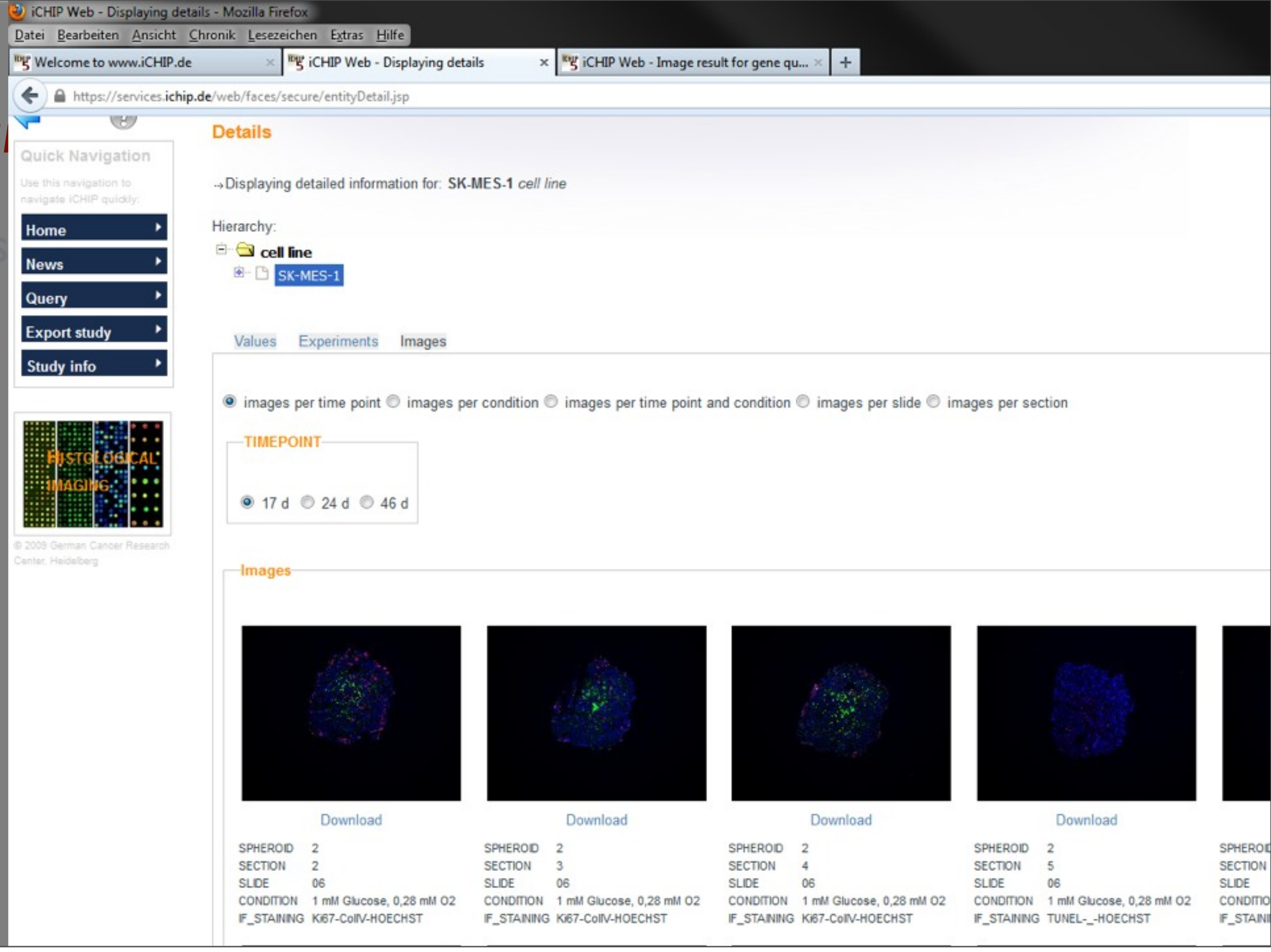
Group Shares

Microscopy

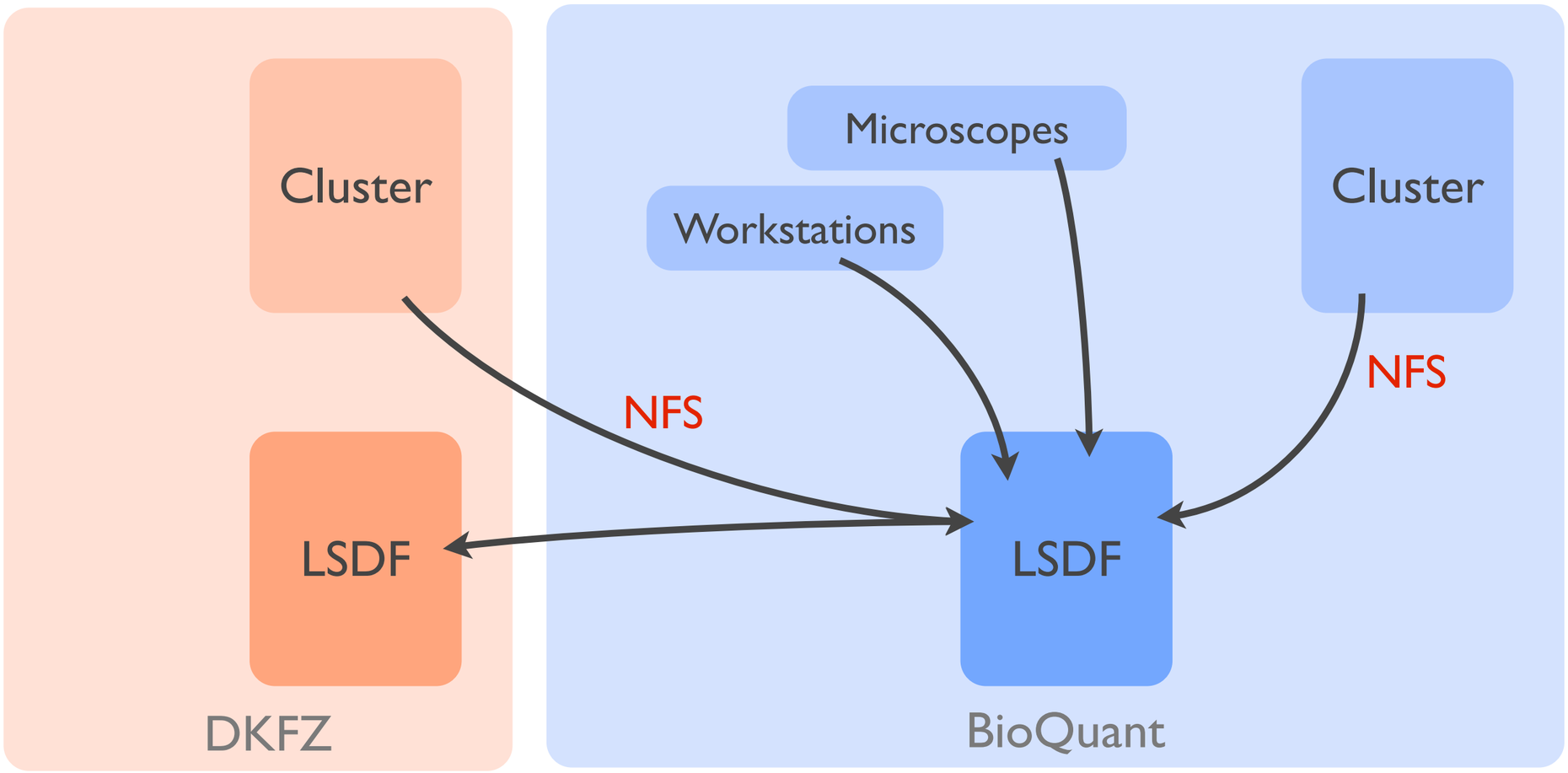
Sequencing

Results Database

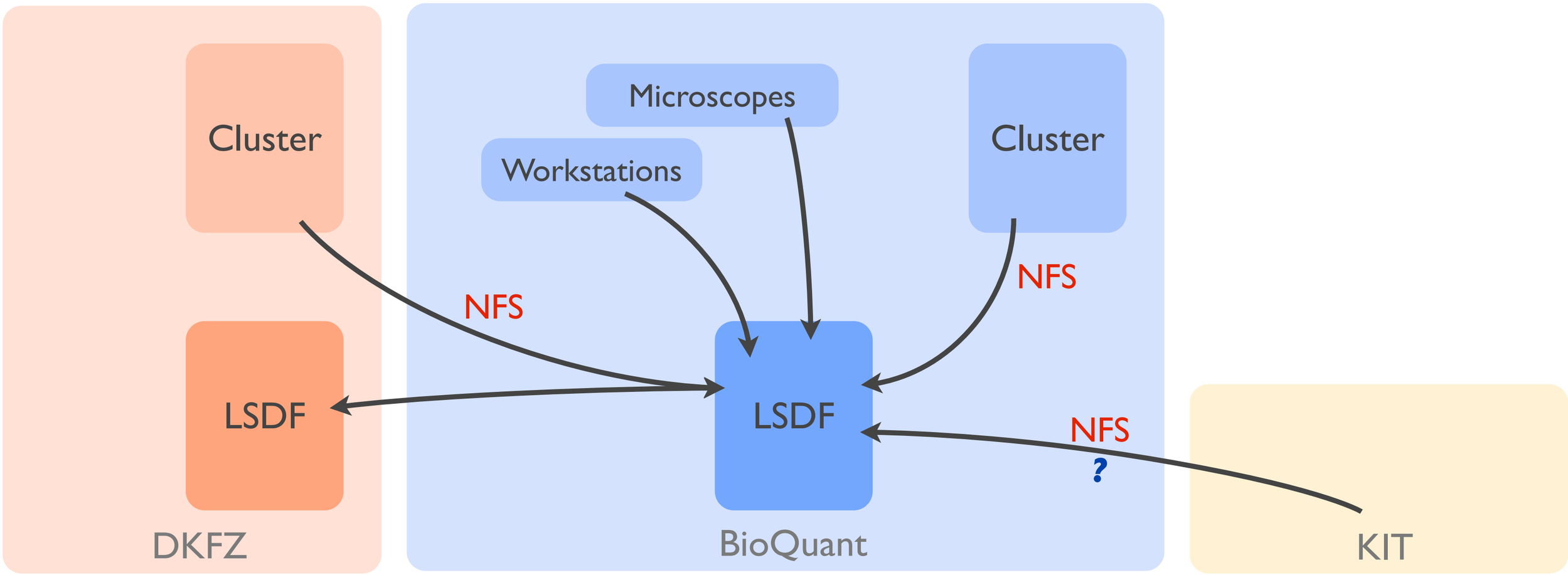
Jürgen Eils, DKFZ



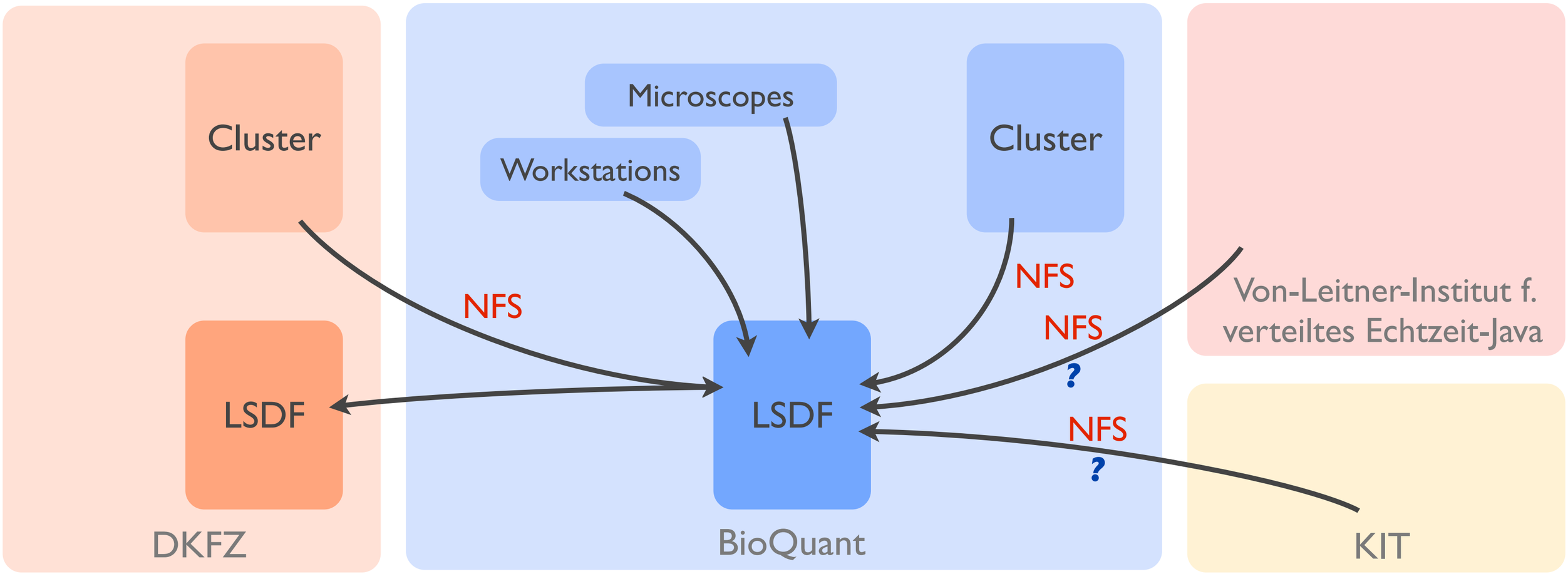
BioQuant LSDF



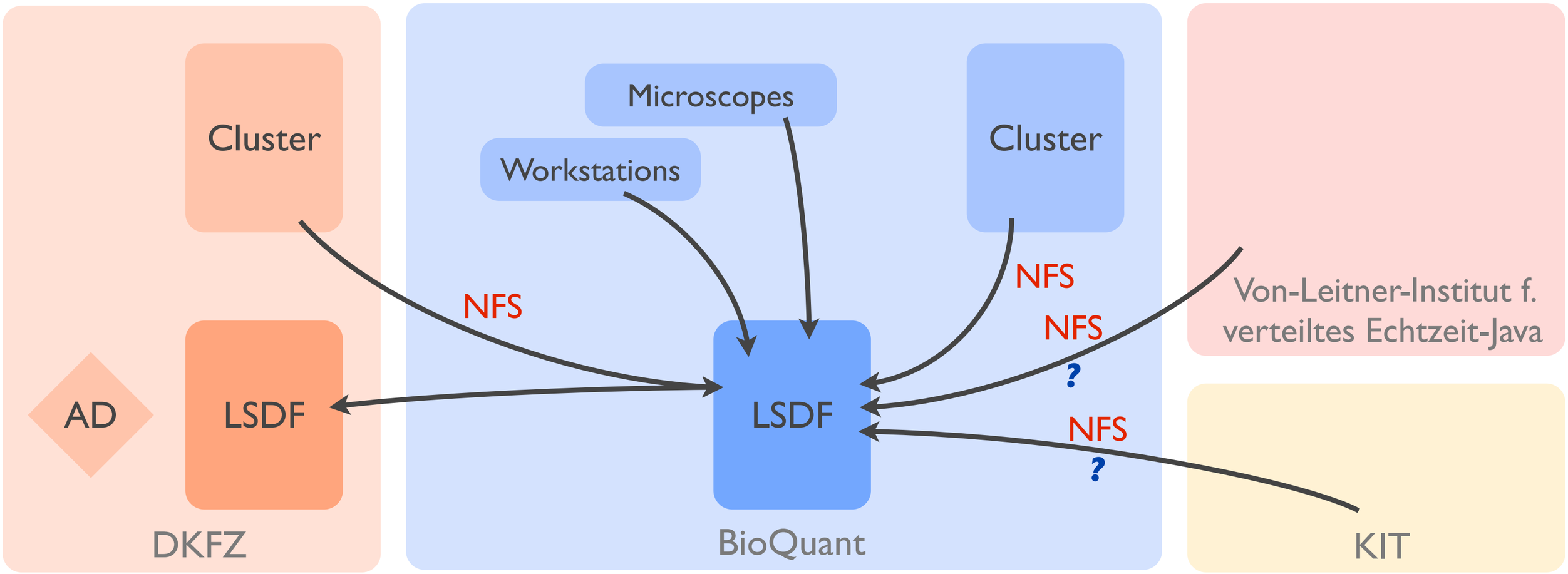
BioQuant LSDF



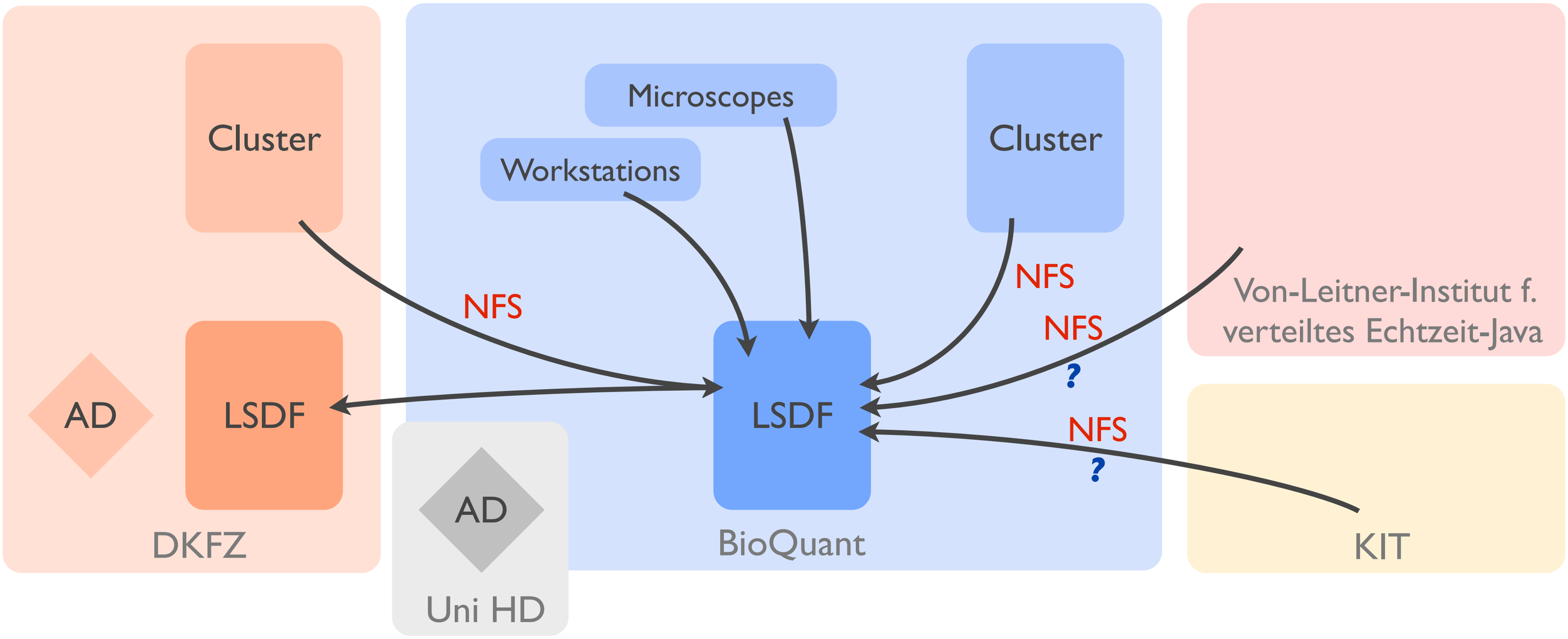
BioQuant LSDF



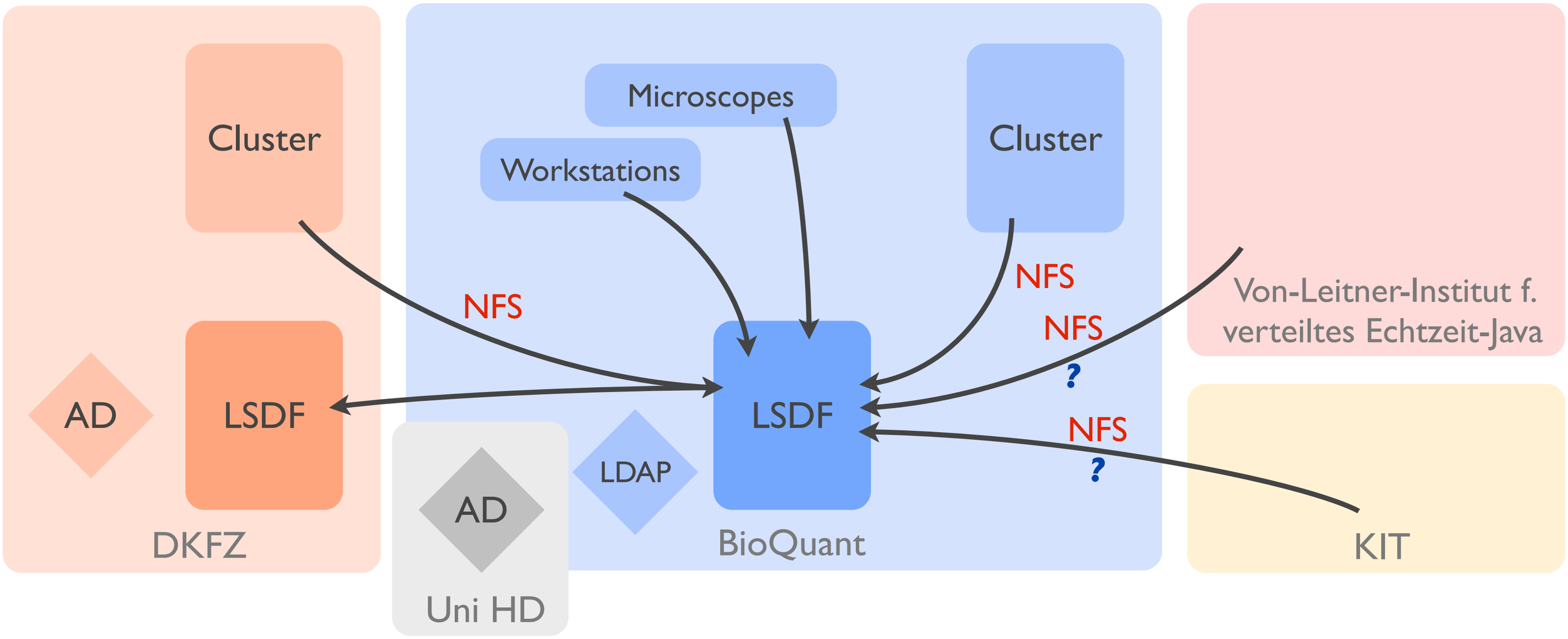
BioQuant LSDF



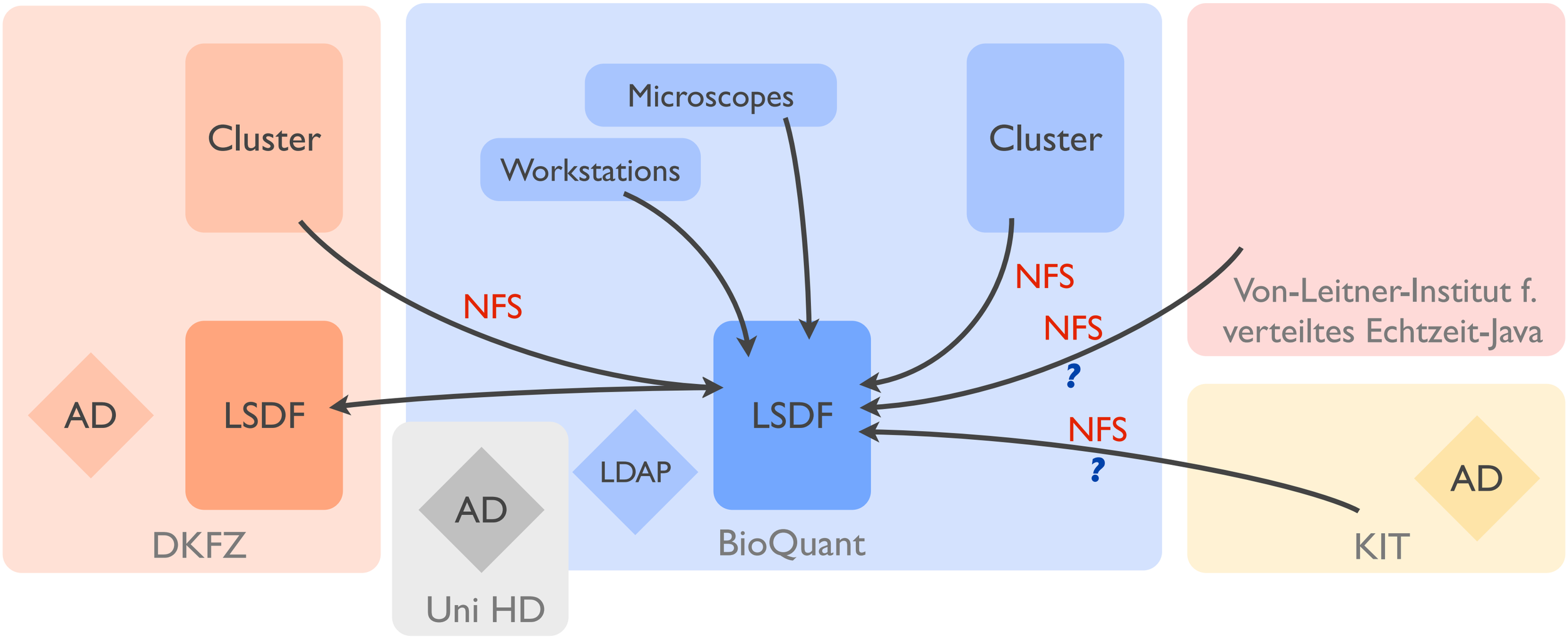
BioQuant LSDF



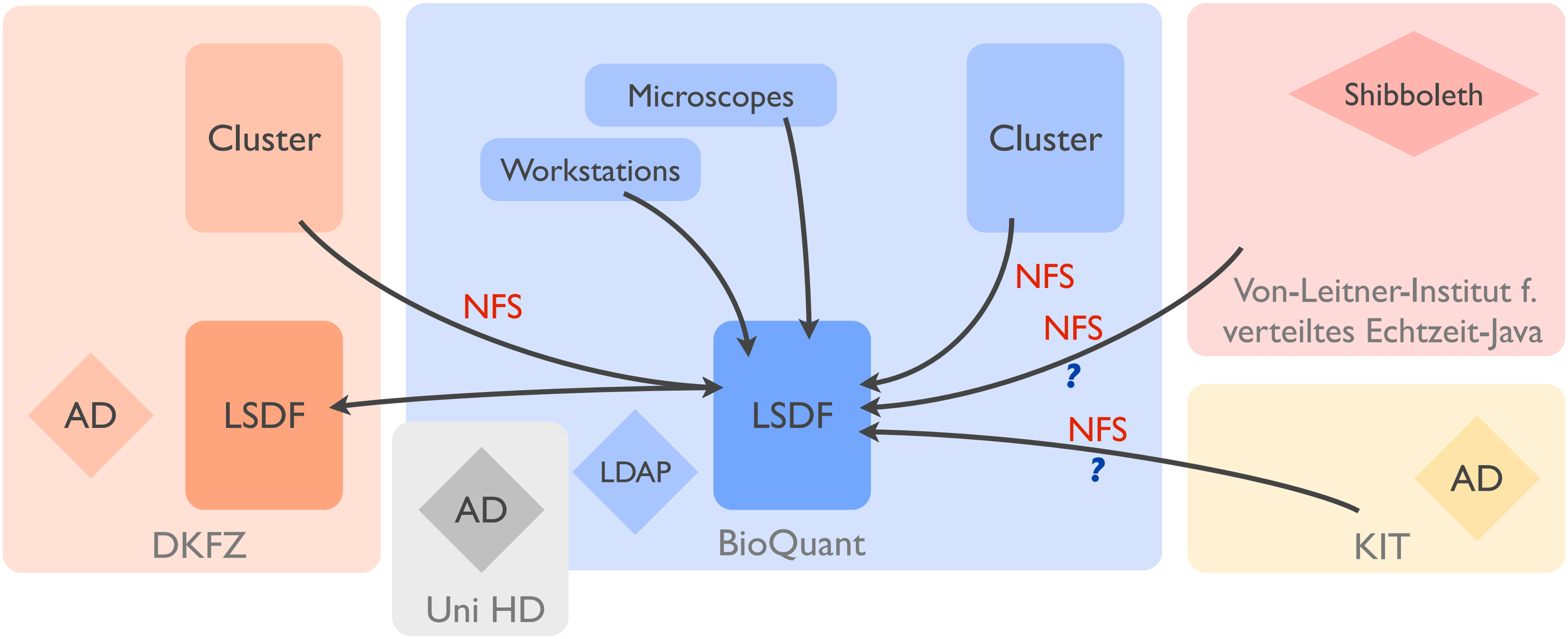
BioQuant LSDF



BioQuant LSDF



BioQuant LSDF



The Problem

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Problem
Authentication
NFS via RPC AUTH_UNIX

The Problem

Authentication

NFS via RPC AUTH_UNIX

The Solution

Authentication

NFS via RPCSEC_GSS_KRB5

The Solution

Authentication

NFS via RPCSEC_GSS_KRB5

The Solution

Authentication

NFS via RPCSEC_GSS_KRB5

BioQuant:

Christian Thiemann

is 3224

The Problem

Differentiation

NFS via RPCSEC_GSS_KRB5 and multiple authorities

The Problem

Differentiation

NFS via RPCSEC_GSS_KRB5 and multiple authorities

COMMUNITY BANK
SAFE DEPOSIT BOXES

Ok, my records show that you are– Aw crap.

DKFZ:

Alice Smith
is 3224

The Solution

Differentiation

NFS via RPCSEC_GSS_KRB5 and multiple authorities and mapped ID spaces

BioQuant:
Christian Thiemann
is 3224

DKFZ:
Alice Smith
is 3224

The Solution
Differentiation

NFS via RPCSEC_GSS_KRB5 and multiple authorities and mapped ID spaces

My records show
that you are 103224.
Here is your box.

My records show
that you are 203224.
Here is your box.

BioQuant: Christian Thiemann is 3224

DKFZ: Alice Smith is 3224

Bank Policy: BioQuant = 10xxxx DKFZ = 20xxxx

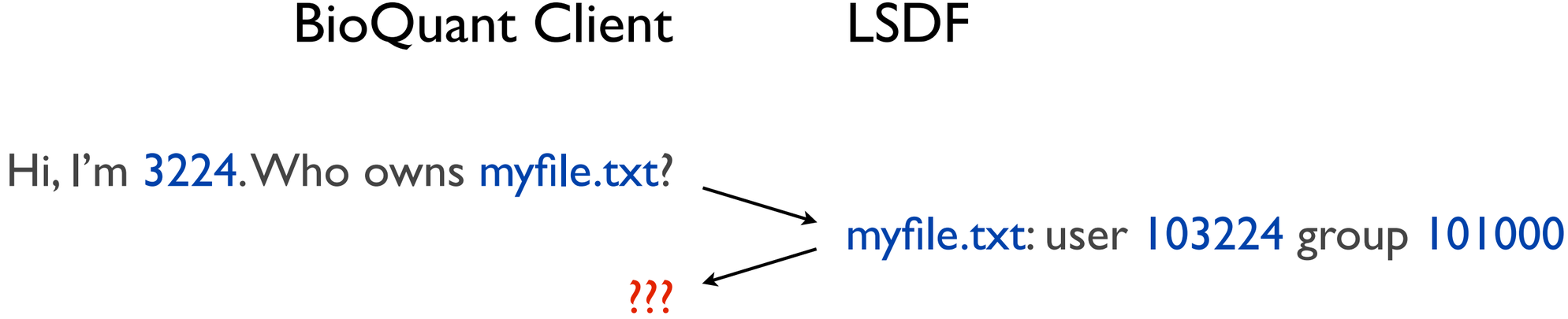
The Problem

NFSv3 uses numeric UID/GIDs

BioQuant Client LSDF

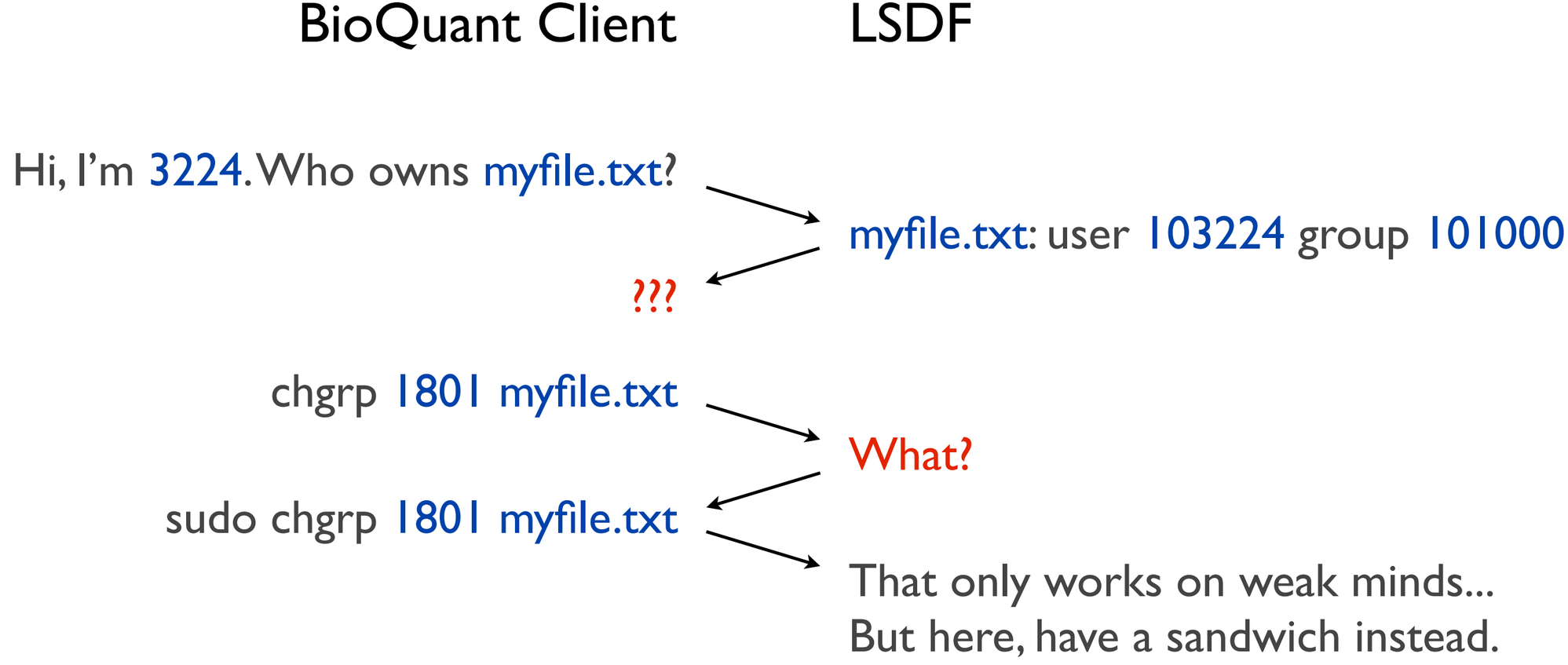
The Problem

NFSv3 uses numeric UID/GIDs



The Problem

NFSv3 uses numeric UID/GIDs



The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

Hi, I'm 103224. Who owns myfile.txt?

myfile.txt: user 3224 group 1000

Ok

myfile.txt: user 103224 group 101000

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client | NFS Proxy | LSDF

Hi, I'm 3224. Who owns myfile.txt?

Hi, I'm 103224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

myfile.txt: user 3224 group 1000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

Ok

sudo chgrp 1801 myfile.txt

sudo chgrp 101801 myfile.txt

Still ok

The Solution

NFSv4 usernames

BioQuant Client

LSDF

The Solution

NFSv4 usernames

BioQuant Client

LSDF

Hi, I'm 3224. Who owns myfile.txt?

Hi, I'm bq_cthiemann@bioquant. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

myfile.txt: user bq_cthiemann@BQ group bq_admins@BQ

Ok

The Solution

NFSv4 usernames

BioQuant Client

LSDF

Hi, I'm 3224. Who owns myfile.txt?

Hi, I'm bq_cthiemann@bioquant. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

myfile.txt: user bq_cthiemann@BQ group bq_admins@BQ

Ok

chgrp 1801 myfile.txt

chgrp rattorturers@BQ myfile.txt

Ok

The Solution

NFSv4 usernames

NFSv4 development is slow:

“no economical incentive” (IBM / EMC)

The Solution

NFSv4 usernames

invisible hand

NFSv4 development is slow:

“no economical incentive” (IBM / EMC)

The Solution

NFSv4 usernames

invisible hand

NFSv4 development is slow:

“no economical incentive” (IBM / EMC)

IBM SONAS will be NFSv4-capable in 2014

hopefully...

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

Ok

sudo chgrp 1801 myfile.txt

sudo chgrp 101801 myfile.txt

Still ok

NFSv3 Ticket Agent:
Kerberized Inter-Realm NFS Service

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

Ok

sudo chgrp 1801 myfile.txt

sudo chgrp 101801 myfile.txt

Still ok

NFSv3 Ticket Agent:
Kerberized Inter-Realm NFS Service

acts as a trusted client to the LSDF (using AUTH_SYS)

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

sudo chgrp 1801 myfile.txt

sudo chgrp 101801 myfile.txt

Still ok

NFSv3 Ticket Agent:
Kerberized Inter-Realm NFS Service

acts as a trusted client to the LSDF (using AUTH_SYS)

authenticates external clients using RPCSEC_GSS_KRB5

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

sudo chgrp 1801 myfile.txt

Still ok

NFSv3 Ticket Agent:
Kerberized Inter-Realm NFS Service

acts as a trusted client to the LSDF (using AUTH_SYS)

authenticates external clients using RPCSEC_GSS_KRB5

transparently translates UIDs & GIDs between LSDF and external ID spaces

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

Ok

sudo chgrp 1801 myfile.txt

sudo chgrp 101801 myfile.txt

Still ok

NFSv3 Ticket Agent:
Kerberized Inter-Realm NFS Service

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

Ok

sudo chgrp 1801 myfile.txt

sudo chgrp 101801 myfile.txt

Still ok

NFSv3 Ticket Agent:
Kerberized Inter-Realm NFS Service

user-space daemon intercepting TCP traffic

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

Ok

sudo chgrp 1801 myfile.txt

sudo chgrp 101801 myfile.txt

Still ok

NFSv3 Ticket Agent:
Kerberized Inter-Realm NFS Service

user-space daemon intercepting TCP traffic

slows down filesystem operations (directory listings etc.) factor 2–3

The Workaround

NFSv3 on-route UID/GID translation

BioQuant Client NFS Proxy LSDF

Hi, I'm 3224. Who owns myfile.txt?

myfile.txt: user 103224 group 101000

Ok

chgrp 1801 myfile.txt

chgrp 101801 myfile.txt

Ok

sudo chgrp 1801 myfile.txt

Still ok

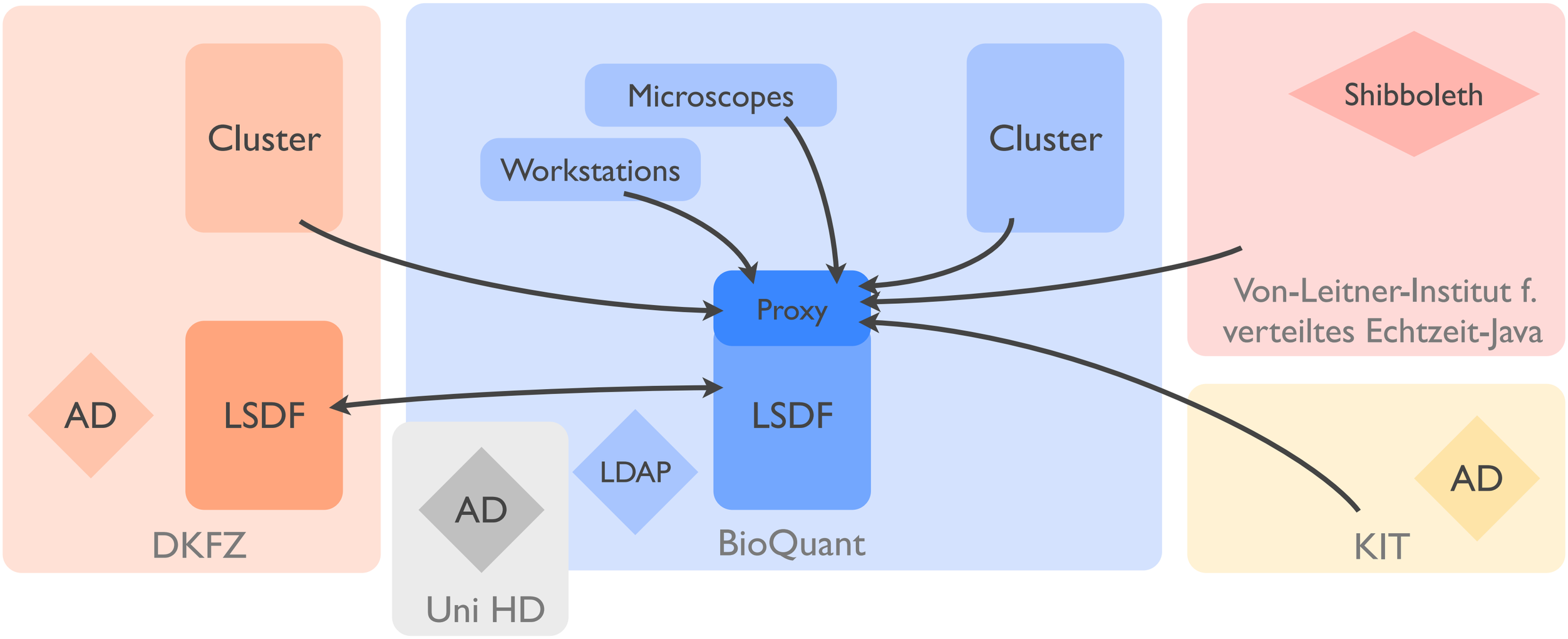
NFSv3 Ticket Agent:
Kerberized Inter-Realm NFS Service

user-space daemon intercepting TCP traffic

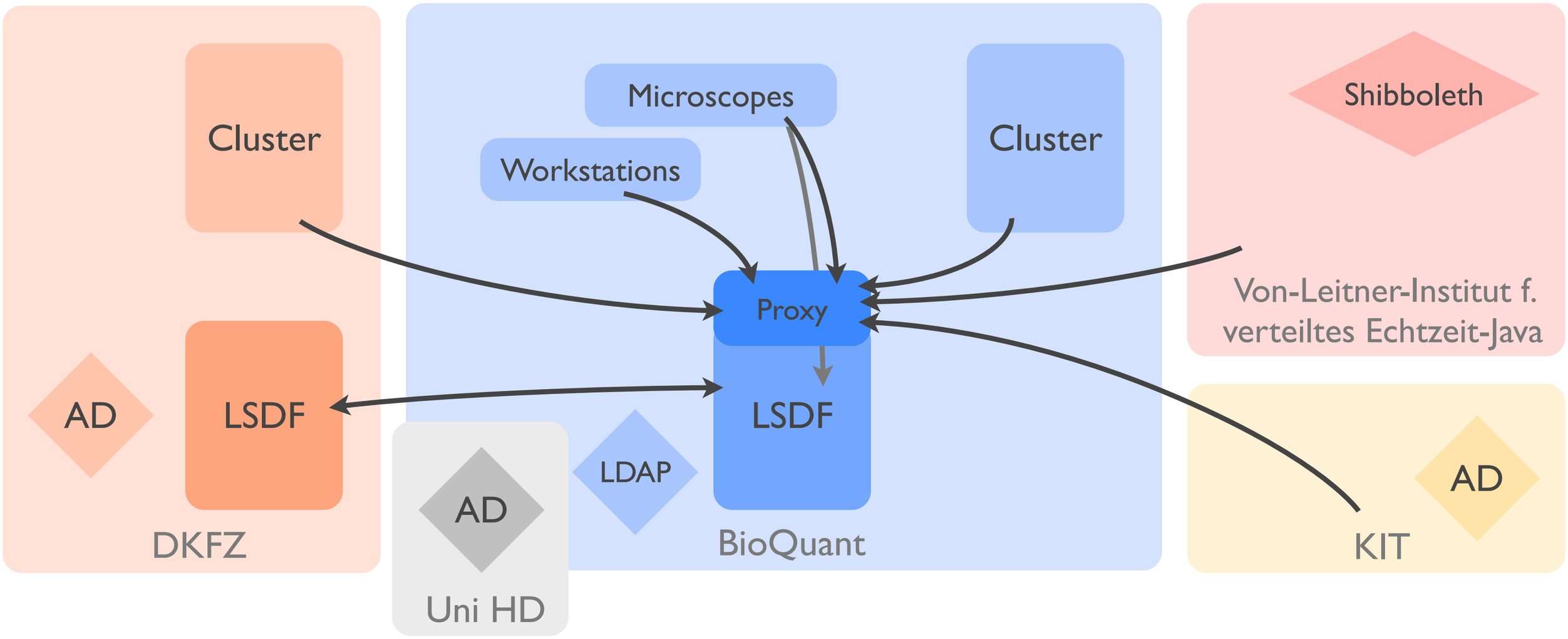
slows down filesystem operations (directory listings etc.) factor 2–3

no throughput penalty on read/write operations 10–20 GBit/s

BioQuant LSDF



BioQuant LSDF



BioQuant LSDF