



Leibniz-Institut für
Astrophysik Potsdam

EREBOS: CLUES Research Environment

CosmoSim Database

Harry Enke
(Kristin Riebe, Jochen Klar, Adrian Partl)

CLUES Meeting 2015 , Copenhagen

Collaborative Research Environment (CRE)

Elements:

- huge data space
 - for raw data
 - for post processing product data
- post processing tools
 - scientific programs, compiler, libs, viewers...
- management procedures for a cross border collaboration
 - user accounts etc. managed by collaboration,
 - membership not bound by relation to institute or university
 - mailing list
- accessible, secure machines
- project web space for collaboration results
- selective file access on demand



EREBOS CRE

CLUES website: www.clues-project.org

User management: www.clues-project.org/management/

Membership: by CLUES manager

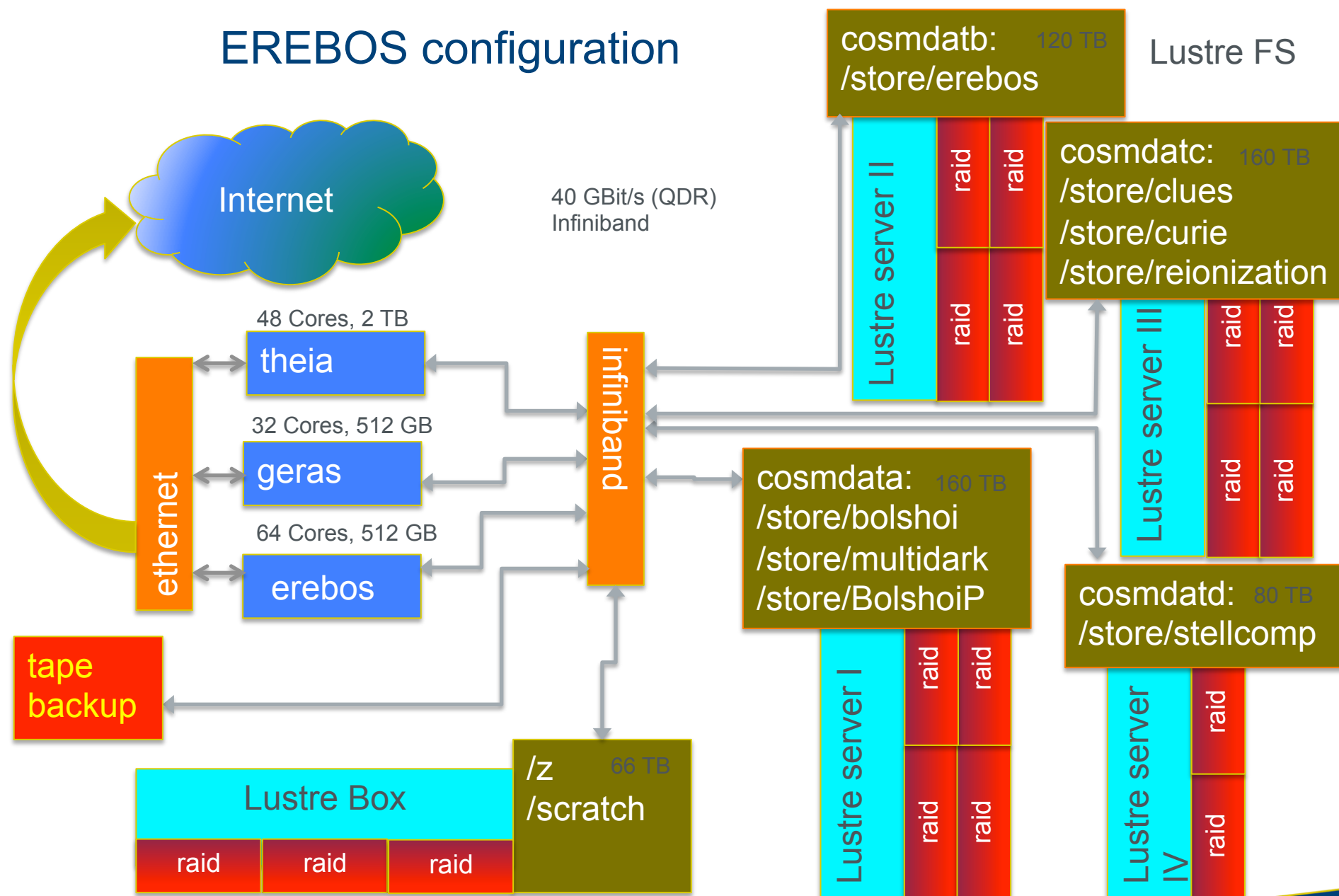
Mailing List: erebos-user@aip.de

Machines: erebos, geras, theia (.aip.de) {work space}
Lustre filesystems {data space }

Access: via ssh/scp on port 2222
from anywhere



EREBOS configuration



EREBOS CRE configuration

OS: CentOS 6.6

Env: modules
idl-8.1, gcc-4.4.7,
intel-14.02,
mpich3,
openmpi{intel,gcc}
python 2.6.6/2.7.5

Groups:
erebos, stellcomp

CPU usage:
32/32/48 cores max.
renice required

Filesys: Lustre 2.7.0 / ext4
~ 560 TB

Policies:
no quota

/store/*

archiving:

dirs not changed
for 6 month are set ro
and written to archive

/scratch/*

(soft) quota [~300GB]

/z {\${HOME}}

weekly backup

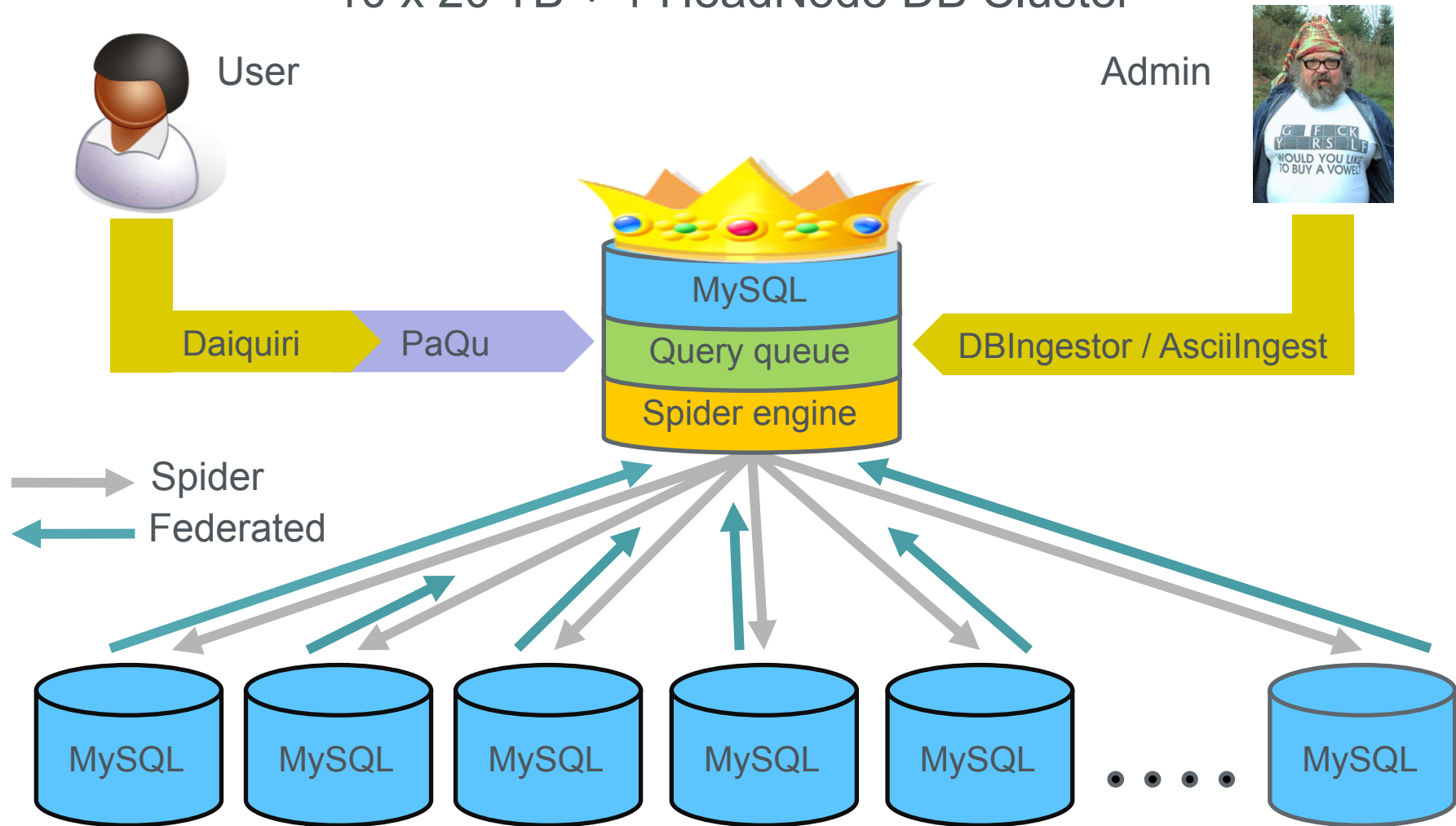
CosmoSim Database: Publishing simulation data



The screenshot shows the CosmoSim website in a Chromium browser window. The URL is www.cosmosim.org. The navigation bar includes links for CosmoSim, Blog, Simulations, Documentation, Query, Admin, Contact, My Account, and Logout. The main header features the "CosmoSim" logo. Below the header, a central text block states: "The CosmoSim database provides results from cosmological simulations performed within different projects: the [MultiDark project](#), the [BolshoiP project](#), and the [CLUES project](#)." To the right of this text is a "Register to CosmoSim" button. Below the central text are three project cards: "MULTIDARK" (Multimessenger Approach for Dark Matter Detection) with links to MDR1, MDPL, and Bolshoi; "BolshoiP" (Cosmological Simulations) with a link to BolshoiP; and "CLUES" (Constrained Local Universe Simulations) with links to Clues3_LGDM and Clues3_LGGas. To the right of these cards is a sidebar containing the AIP logo, text about hosting by the Leibniz-Institute for Astrophysics Potsdam (AIP), the GAVO logo (German Astrophysical Virtual Observatory), text about its contribution to the German Astrophysical Virtual Observatory, and text about the MultiDark and Bolshoi simulations being run on the NASA's Pleiades supercomputer at the NASA Ames Research Center. At the bottom of the sidebar is the PRACE logo. Below the project cards, a "Database access" section explains how to query the database using SQL statements, mentioning the "Query Form" and "Registration Form". It also notes that queries can be submitted as a guest, but results can be accessed and removed by any other guest as well. A final line of text states: "More information on the simulations and their projects is available in the [Simulations](#) section. Details about

CosmoSim : Machinery / Database

10 x 20 TB + 1 HeadNode DB Cluster



CosmoSim : Machinery / Webinterface

Daiquiri:

- Highly customizable framework for database publication
- Many applications – one code base, maintained by the E-Science group at AIP
- Modern interface
- Hardened SQL query interface
- State-of-the-art web technologies
- WordPress module for documentation
- Best of all: open source, on github



CosmoSim : Machinery / Data Management

Ingest + Curation

- from erebos environment
- libraries for insert of huge data
 - DBIngestor: all data with reader-plugin
 - AsciiIngest
- consistency and plausibility checks in collaboration with scientists
- extensive documentation of data, howtos
- Virtual Observatory Compliance with
 - Metadata
 - Data Formats + Data API
- Digital Object Identifier for data tables/sets



CosmoSim : Data Sets

The screenshot shows the CosmoSim website with a dark blue header containing navigation links: CosmoSim, Blog, Simulations, Documentation, and Query. The Simulations menu is open, displaying a list of simulation projects: Simulations overview, MultiDark Project (MDR1, MDPL, MDPL2, SMDPL, Bolshoi), BolshoiP Project (BolshoiP), CLUES Project (Clues3_LGDM, Clues3_LGGas), Galaxy Clusters, and Halo Finders. To the right of the menu, a list of data types is shown: halo catalogs (FoF, BDM AHF, Rockstar), merger trees, particles of snapshots, and densities, cosmic web. Below the menu, three project cards are visible: MULTIDARK (Multimessenger Approach for Dark Matter Detection), BolshoiP (Cosmological Simulations), and CLUES (Constrained Local Universe Simulations). Each card contains a brief description and a list of specific simulation names.

Simulations

- Simulations overview
- MultiDark Project
 - MDR1
 - MDPL
 - MDPL2
 - SMDPL
 - Bolshoi
- BolshoiP Project
 - BolshoiP
- CLUES Project
 - Clues3_LGDM
 - Clues3_LGGas
- Galaxy Clusters
- Halo Finders

data types:

- halo catalogs (FoF, BDM AHF, Rockstar)
- merger trees
- particles of snapshots
- densities, cosmic web

MULTIDARK
Multimessenger Approach for Dark Matter Detection

The Spanish MultiDark Consolider project supports efforts to identify and detect matter, including dark matter simulations of the universe.

[MDR1](#)
[MDPL](#)
[Bolshoi](#)

BolshoiP
Cosmological Simulations

The BolshoiP project contains a simulation like Bolshoi, with the same box size and resolution, but with Planck cosmology.

[BolshoiP](#)

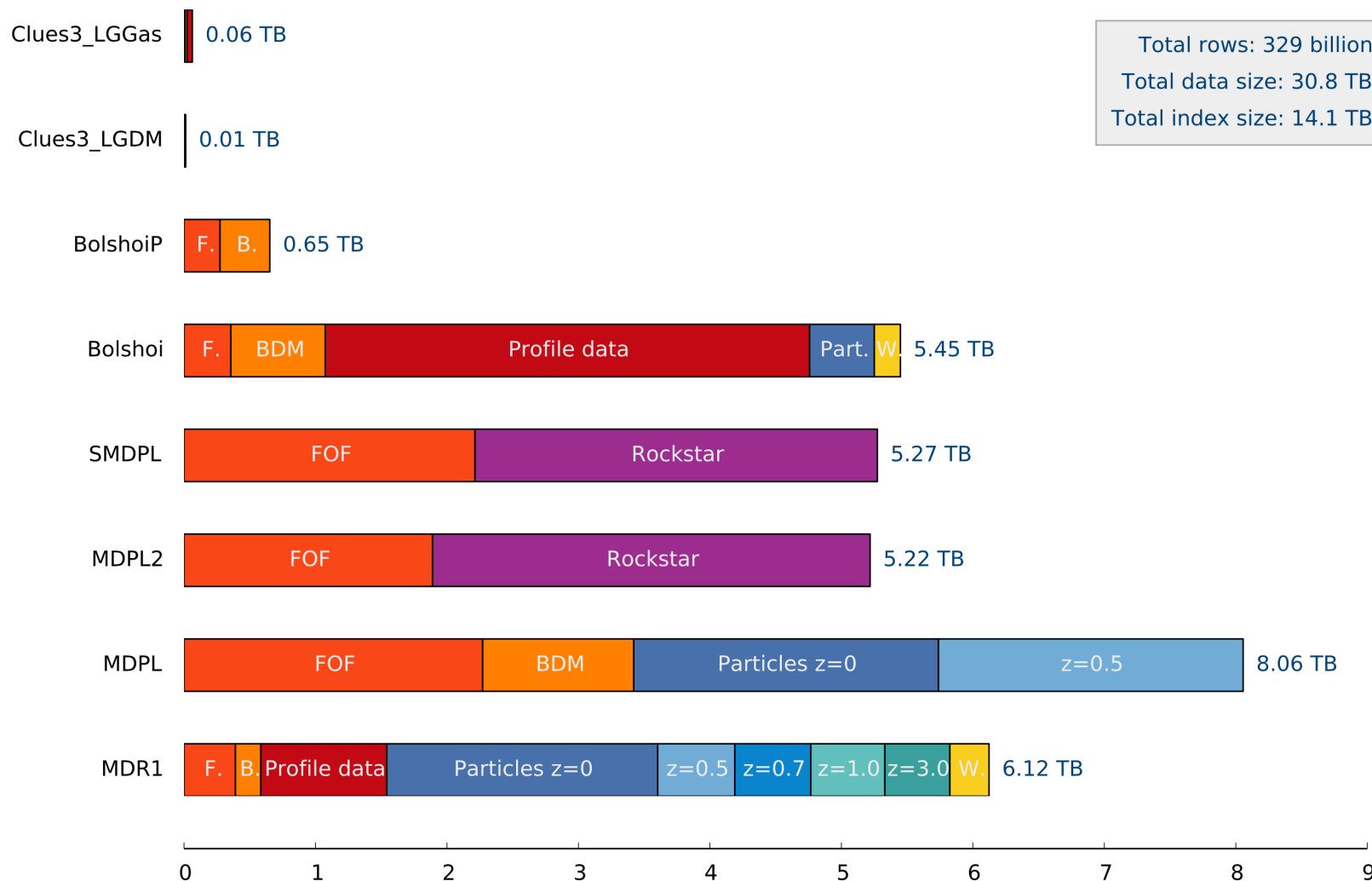
CLUES
Constrained Local Universe Simulations

The CLUES project deals with constrained simulations of the local universe, partially with gas and star formation.

[Clues3_LGDM](#)
[Clues3_LGGas](#)

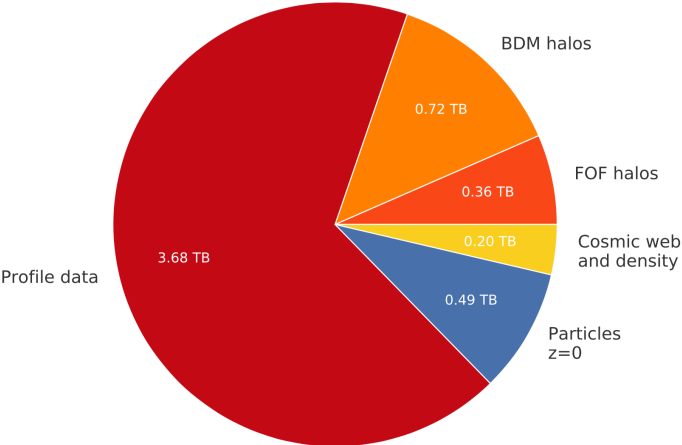
CosmoSim : Numbers, Stats

Data volumes for CosmoSim



CosmoSim : numbers, data sizes

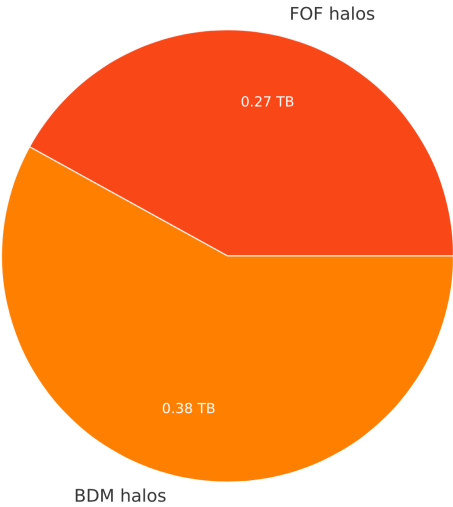
Data volume for Bolshoi



FOF halos	0.36 TB
BDM halos	0.72 TB
Profile data	3.68 TB
Particles z=0	0.49 TB
Cosmic web and density	0.20 TB

Total data size 5.45 TB
index size 1.76 TB
number of rows 45.77 billion

Data volume for BolshoiP

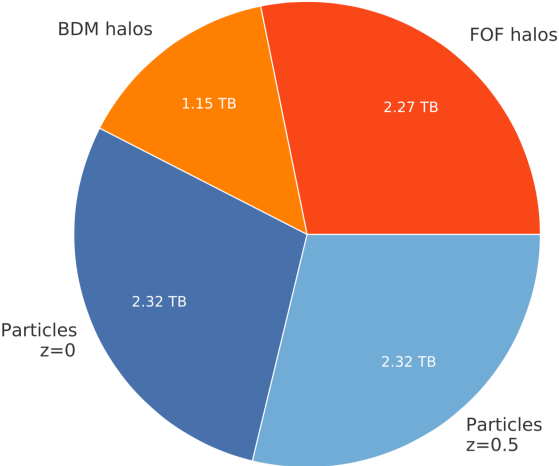


FOF halos	0.27 TB
BDM halos	0.38 TB

Total data size 0.65 TB
index size 0.32 TB
number of rows 3.96 billion

CosmoSim : numbers, data sizes

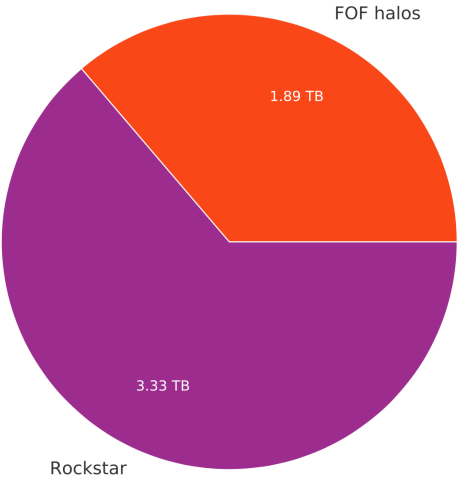
Data volume for MDPL



FOF halos	2.27 TB
BDM halos	1.15 TB
Particles z=0	2.32 TB
Particles z=0.5	2.32 TB

Total data size 8.06 TB
index size 6.27 TB
number of rows 135.89 billion

Data volume for MDPL2



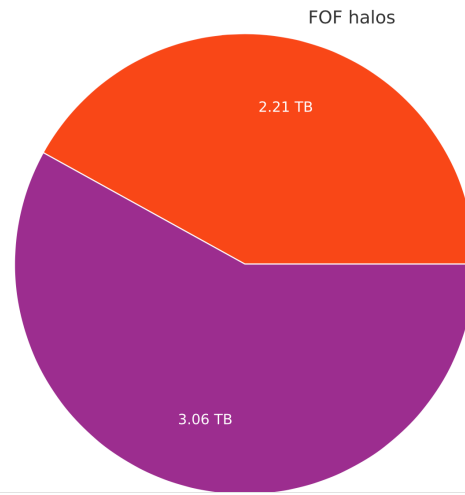
FOF halos	1.89 TB
Rockstar	3.33 TB

Total data size 5.22 TB
index size 1.01 TB
number of rows 24.09 billion



CosmoSim : numbers, data sizes

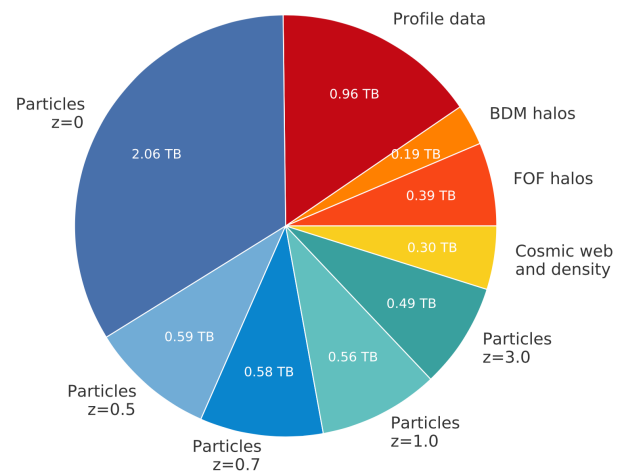
Data volume for SMDPL



FOF halos	2.21 TB
Rockstar	3.06 TB

Total data size 5.27 TB
index size 1.19 TB
number of rows 25.65 billion

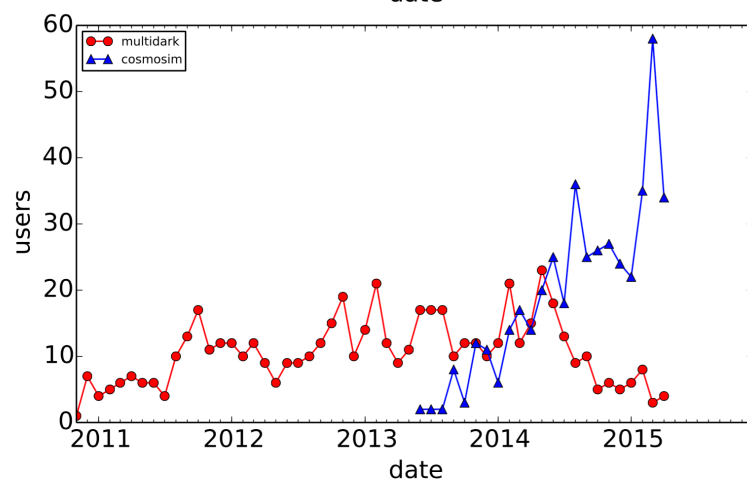
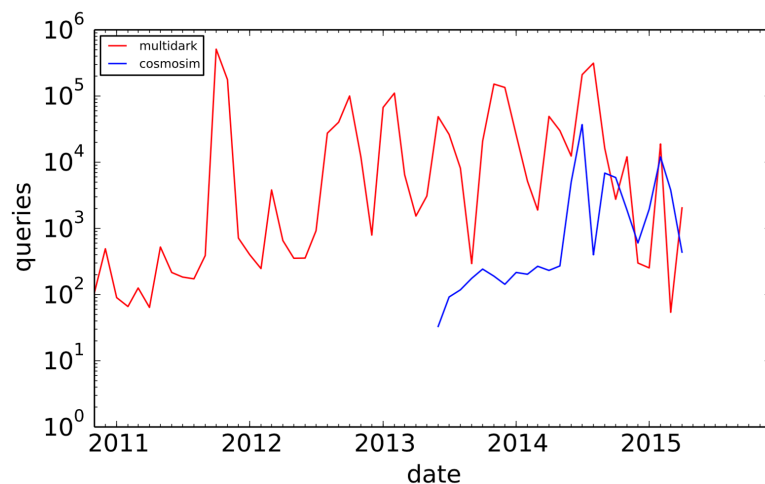
Data volume for MDR1



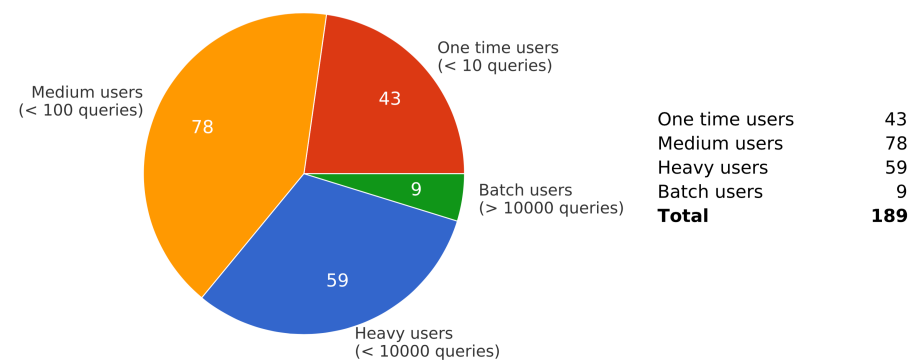
FOF halos	0.39 TB
BDM halos	0.19 TB
Profile data	0.96 TB
Particles z=0	2.06 TB
Particles z=0.5	0.59 TB
Particles z=0.7	0.58 TB
Particles z=1.0	0.56 TB
Particles z=3.0	0.49 TB
Cosmic web and density	0.30 TB

Total data size 6.12 TB
index size 3.54 TB
number of rows 93.44 billion

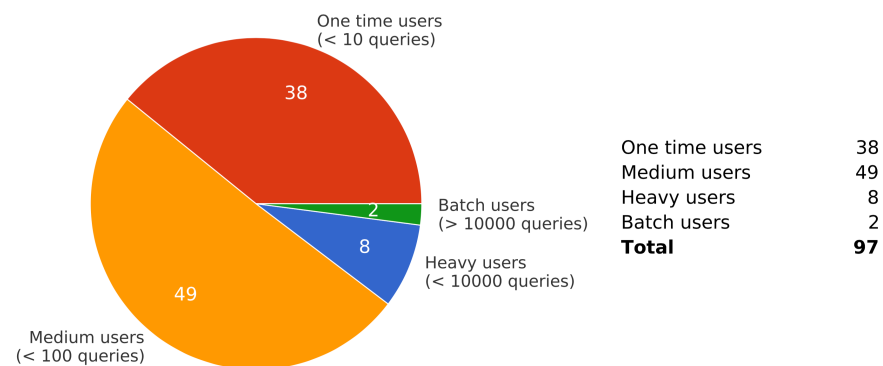
CosmoSim : numbers, usage stats



Queries per user, all



Queries per user, cosmosim



CosmoSim : CLUES data

Database browser

DATABASES	TABLES
BOISNOIR	AHF
Clues3_LGDM	AHFProf
Clues3_LGGas	Redshifts
MDPL	

- CLUES-Daten-Nutzung:
- mysql> select resultDBName, count(*)
as num from mysql.queue_history
where query like '%Clues%' group by resultDBName;
- resultDBName | num | +|
- cosmosim_user_4 | 4 | |
- cosmosim_user_1 | 1 | |
- cosmosim_user_j | 534 | |
- cosmosim_user_l | 140 | |
- cosmosim_user_s | 1 | |
- cosmosim_user_s | 7 | |
- cosmosim_user_f | 38 |
- +-----+-----+
- 7 rows in set (2.39 sec)

Questions:
other data?
RockStar?

New data?
usage?

CLUES
data locator ?

CosmoSim : CLUES data

Database browser

CLUES
Constrained Local Universe Simulations

AIP University of Liverpool NM STATE UNIVERSITY UAN

CLUES People Simulations Talks Articles Image Gallery Movies Observations

» Simulations » Box 64 - 186592 » Full Box

Box 64 - Full Box

For getting the large scale structure, the full box was simulated with 256^3 and 128^3 (dark matter) particles. The initial conditions were set up according to a WMAP 3 power spectrum and normalization. The constraints for simulating the Local Universe were generated as a 256^3 mesh of overdensities.

The full box was simulated with dark matter particles only. If you are interested in hydrodynamical simulations, please see the high resolution simulations at "Resimulation".

Click on the plus signs to explore the data tree below.

[Collapse All](#) | [Expand All](#)

- Dark Matter
 - 1024³
 - Snapshots
 - AIP (z=0)
erebos:/store/01/clues/B64_WM3_186592/BOX_DM/1024/SNAPS/SNAP_135/
 - LRZ (z=0)
archive:/ptmp2/h009z/h009zad/DATA/LG/YANIV4096/DARK_ONLY/64_1024/
 - AHF
 - AIP
erebos:/store/01/clues/B64_WM3_186592/BOX_DM/1024/AHF/
 - LRZ
archive:/ptmp1/h009z/h009zah/64_1024/analysis_final/
 - AHF Merger Tree
 - BDM
 - LRZ (z=0)
archive:/ptmp1/h009z/h009zaa/Box64CR/RUN1024/BDM/
 - FOF
 - FOF Merger Tree
 - FOF Substructure
 - GALICS Merger Tree

Questions:
other data?
RockStar?

New data?
usage?

CLUES
data locator ?

CosmoSim : CLUES data

Database browser	
DATABASES	TABLES
BolshoiP	AHF
Clues3_LGDM	AHFProf
Clues3_LGGas	Redshifts
MDPL	

Database browser	
DATABASES	TABLES
BolshoiP	AHF
Clues3_LGDM	AHFGas
Clues3_LGGas	AHFLum
MDPL	AHFMag
MDPL2	AHFMatch
MDR1	AHFProf
SMDPL	AHFStars
testBolshoiP	

Questions:
other data?
RockStar?

New data?
usage?

CLUES
data locator ?

CosmoSim : in preparation / plans / desiderata

- update-mainLeaflds for MDPL2, publication-announcement
 - ingest Rockstar for MDR1, Bolshoi, BolshoiP; update mainLeaflds
 - ingest Galacticus for MDPL2
 - density fields from Noam, for MDPL2 (1024^3)
 - question from LSST (Nelson Padilla)
 - Tweb / Vweb ??
 - new simulations: BigMD (Planck?, other?), HugeMD
 - with FOFs, Rockstar => on which timescale?
 - particles for MDPL2, SMDPL? (at least $z=0$)
 - do we want all snapshots for Galacticus
 - or just restricted subset (e.g: $z=0, 0.5, 0.7$)
- 