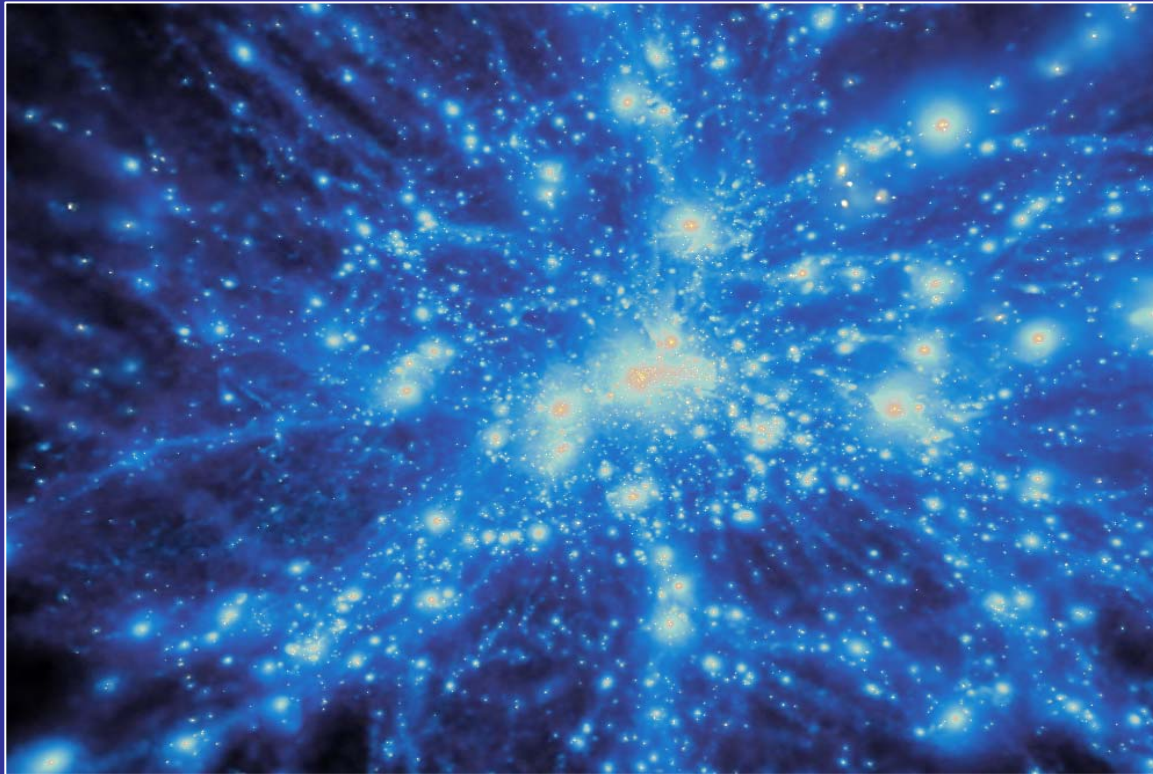


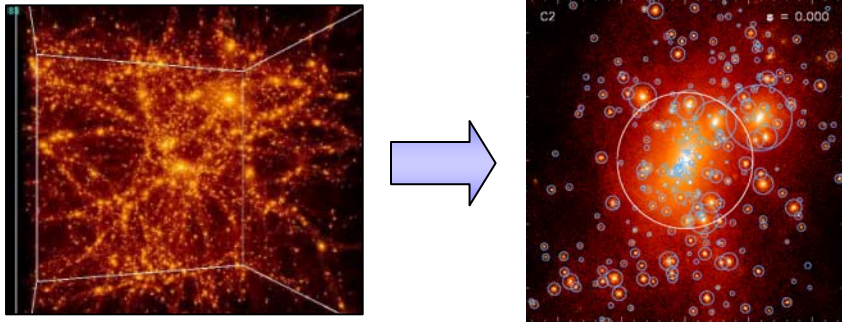
MultiDark Simulation Database



Kristin Riebe
CLUES Workshop, 18.06.2012
Lyon, France

Simulations

- produce Terabytes of data
- Post-processing results have various data formats
- Hard to handle & share!



Databases

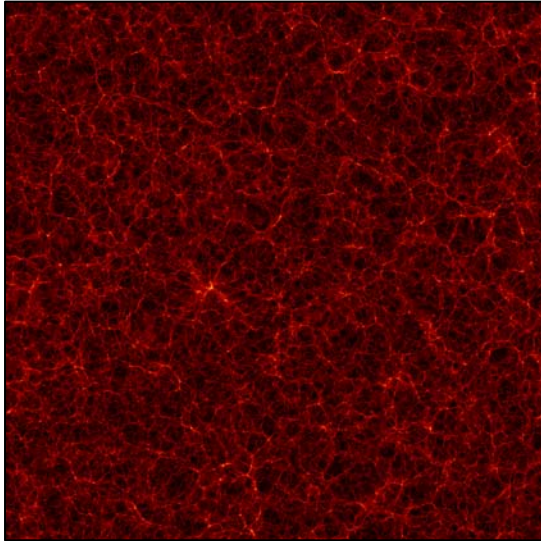
- store tables & relations
- Standard Query Language SQL to access data
- Uniform data format
- Download only subsets, easy sharing & publishing!

```
select top 20 * from MDR1..FOF
where snapnum=85
order by mass desc
```

extracts 20 most massive FOF
groups at $z=0$

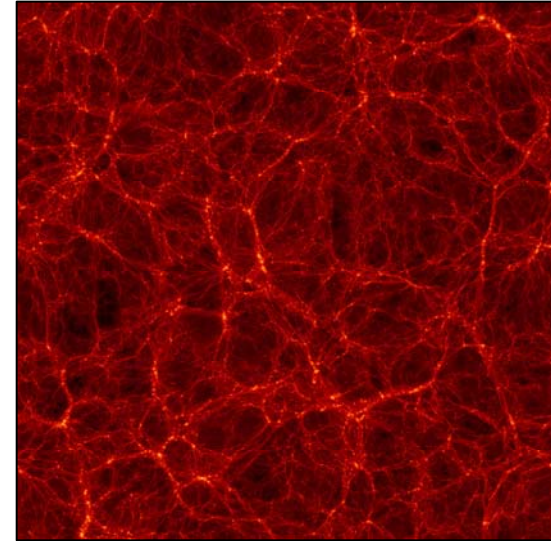
< 400 ms

MultiDark Run 1 (MDR1)



- WMAP5 cosmology
- Box: 1 Gpc/h
- 2048^3 particles (~ 8.6 billion)
- mass resolution $8.7 \cdot 10^9 M_{\text{sun}}/h$
- 43 snapshots between $10.34 < z < 0$

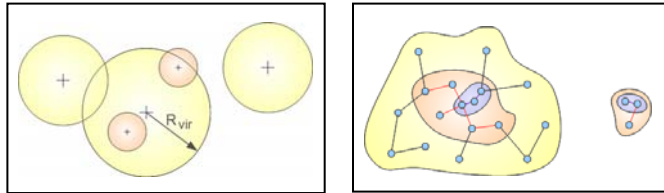
Bolshoi



- WMAP5 cosmology
- Box: 250 Mpc/h
- 2048^3 particles
- mass resolution $1.35 \cdot 10^8 M_{\text{sun}}/h$
- 73 snapshots between $15.57 < z < 0$

- Halo catalogues**

- BDM, FOF
- multiple density thresholds/ linking lengths



- BDM halo profiles**

- inner structure of halos

- Particle Snapshots**

MDR1: $z=0, 0.5, 1, 2.9$

Bolshoi: $z=0$

bdmId	snapnum	NInCat	hostFlag	x	y	z
8511186098	85	11186098	-1	548.8973	143.9528	895.0
8512166221	85	12166221	-1	384.92	468.0325	872.0
8512459068	85	12459068	-1	737.7123	482.5123	972.0
8505410295	85	5410295	-1	947.049	971.8178	267.0
8506742613	85	6742613	-1	276.4744	488.3719	548.0
8506941522	85	6941522	-1	665.9299	986.9252	815.0
8513458743	85	13458743	-1	665.9299	986.9252	815.0

1 halo/group per row

unique identifier for each halo/group

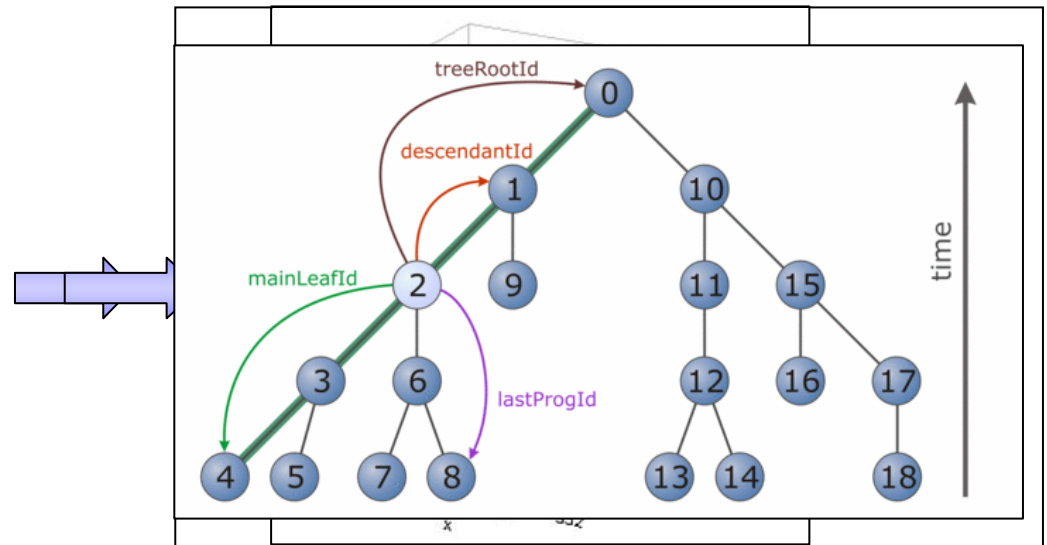
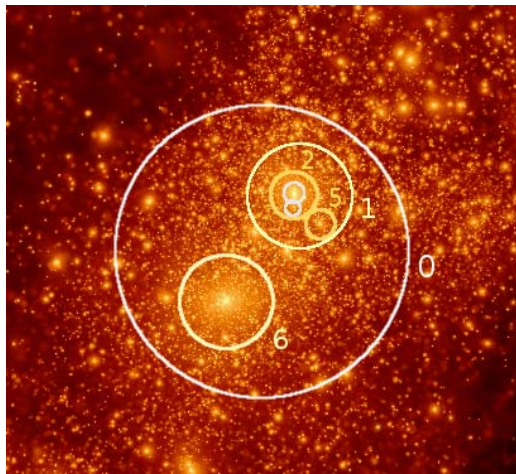
bdmId	snapnum	NinCat	R_Rvir	Rbin	np	mass
8506742613	85	6742613	1.0629464	2828.978	10582	1.880109
8506742613	85	6742613	1.1953862	3174.1658	12201	1.985920
8506742613	85	6742613	1.3382614	3561.4724	11217	2.083212
8506742613	85	6742613	1.4941747	3996.0378	7967	2.152309
8506742613	85	6742613	1.6838918	4483.628	6245	2.206469
8506742613	85	6742613	1.8875381	5030.7134	7078	2.267854
8506742613	85	6742613	0.9454057	2521.329	10907	1.788320

link to BDM halo

distance from center

Data sets: only MDR1

- FOF particles: link between FOF groups and particles
- FOF substructure trees, superclusters: hierarchy of structures
- FOF merger trees: evolution of FOF groups



MDR1

- FOF (6 link. lengths)
- BDMV, BDMW
- BDM profiles
- particle snapshots:
z=0, 0.5, 1, 2.9
- FOF Superclusters, 7 link.
length, substructure links;
z=0
- FOF Substructure (z=0)
- FOF Merger trees

Bolshoi

- FOF (6 link. lengths)
- BDMV, BDMW
- BDM profiles
- particle snapshots:
z=0

Bolshoi Merger Trees
(BDM)???

Galaxy Catalogues?

- Data side:
 - Data in variety of formats => must be properly documented!
 - Data curation and preparation needed

New Database Ingestor Library
=> easier & faster data ingest

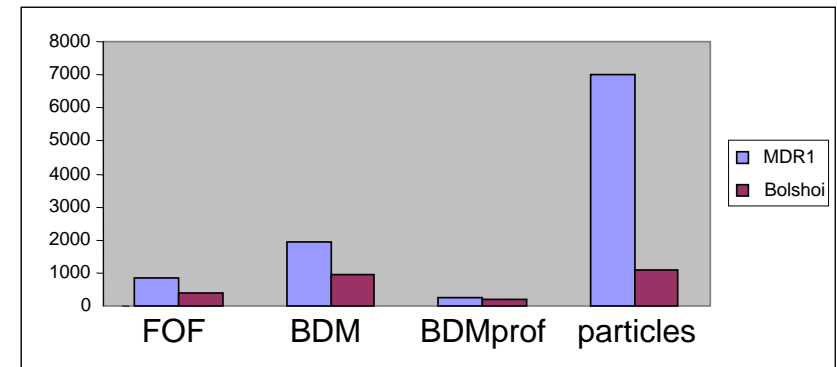
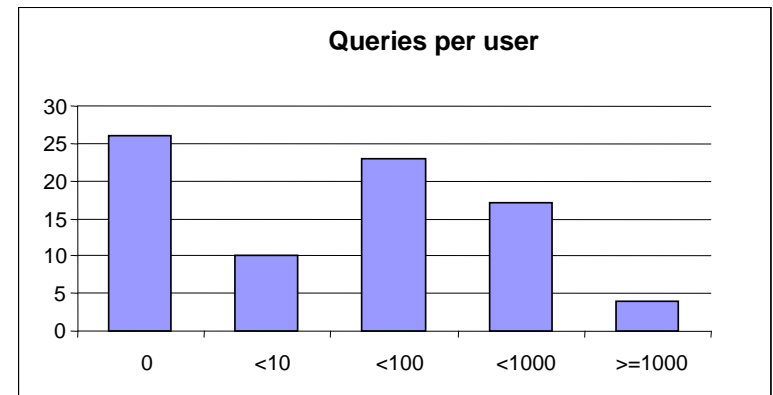
- Server side:
 - Long ingest times & index calculation
 - Performance limitations reached

ParaQuery for MySQL, Spider Engine
=> fast, open source solution, own plugins

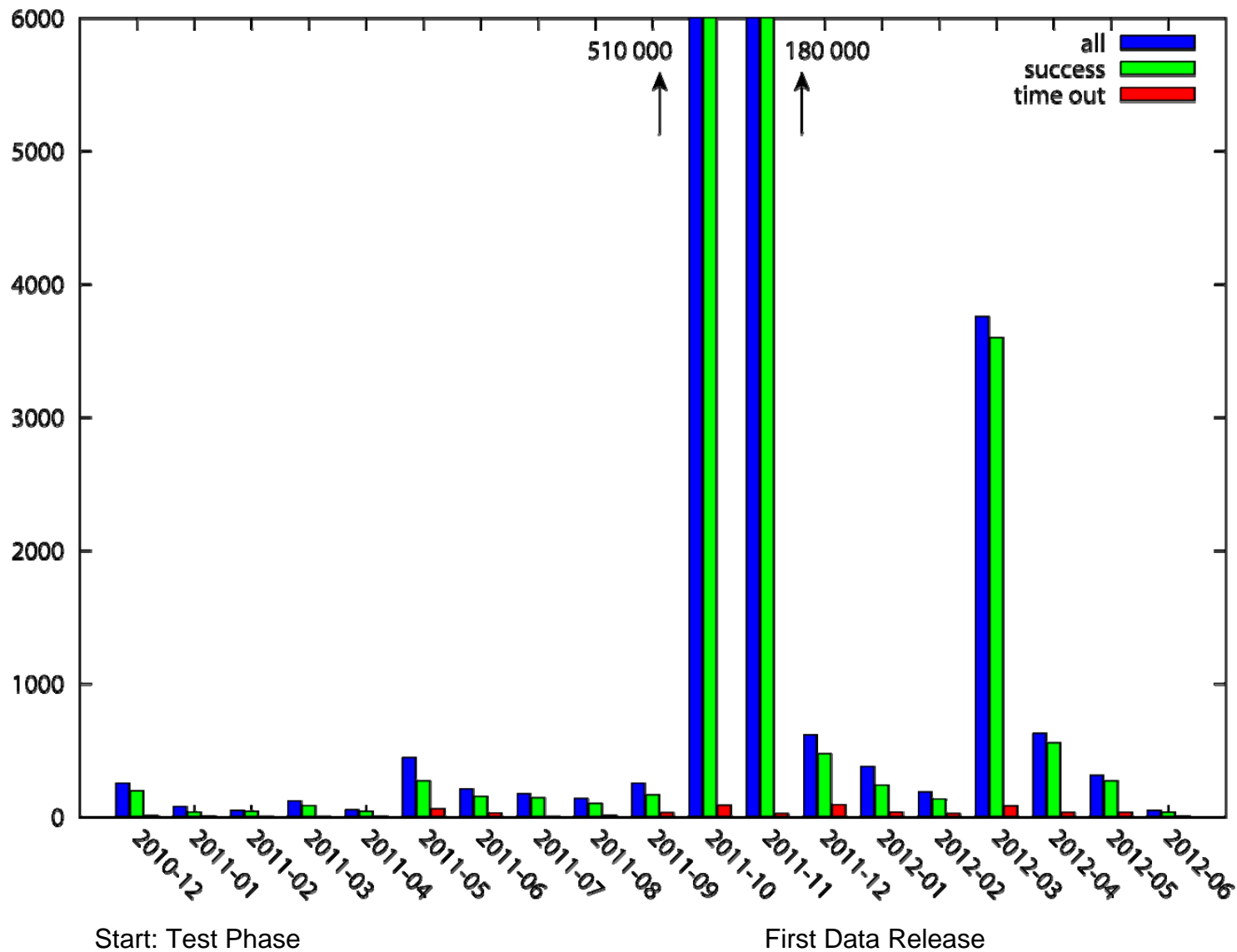
- User side:
 - Complicated queries with joins can take hours
 - Spatial queries only feasible in simple cases
 - Slow access to particle data
 - Time outs, no possibility to kill jobs

MySQL plugins for killing parallel jobs, MySQL job queue, New Web Application

- **80 registered users**
- **Total number of (succ.ful) queries:**
 - 692 000 queries,
 - 6.2 billion rows,
 - 205 GB
- 1 power user: 682 000 queries, 1.2 billion rows, 42 GB ~20%
- => remaining users:
 - 10 000 queries,
 - 5 billion rows,
 - 163 GB
- **Queries per Database (only 10k)**
 - 25% queries on Bolshoi,
 - 75% on MDR1
- **Most wanted tables:**
 - particles, BDM&FOF, BDMprof;
 - remaining tables < 20 queries



Queries per month



- How can we increase usage?
- Which data do YOU want?
- What science do you want to do, could it be done with a database?
- Which CLUES simulations shall be published?

- More data sets are welcome!
- Need proper documentation, clean link from your data to existing catalogues (if there is a connection)