A Fast Database for Large Observational or Simulation Datasets

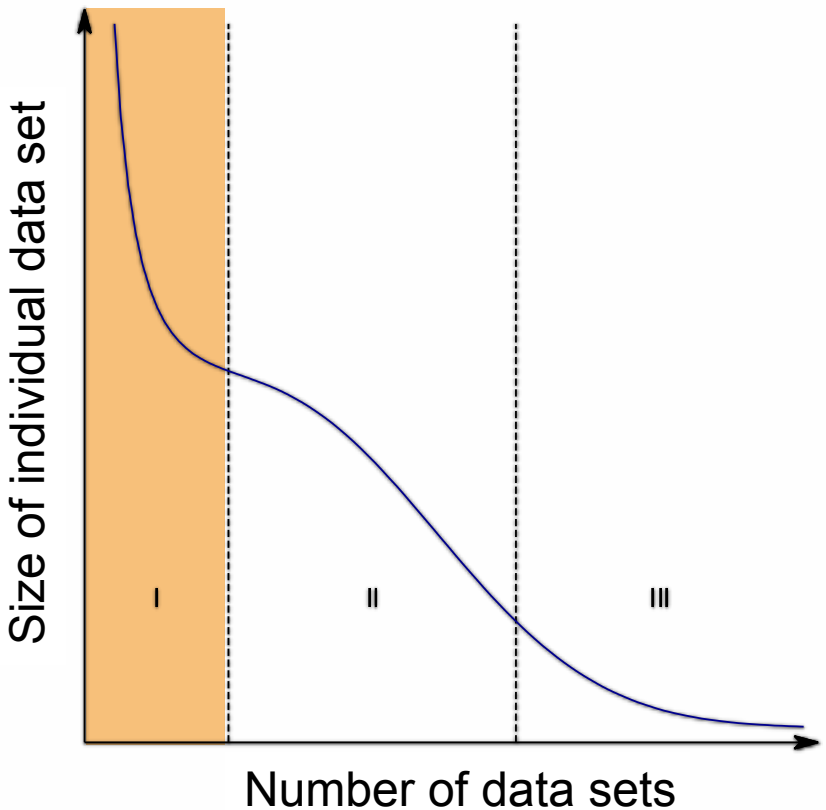
Adrian M. Partl

Leibniz-Institut für Astrophysik Potsdam (AIP)

Big Data in Astronomy and Astrophysics

- Raw data usually with clear data formats, huge in size

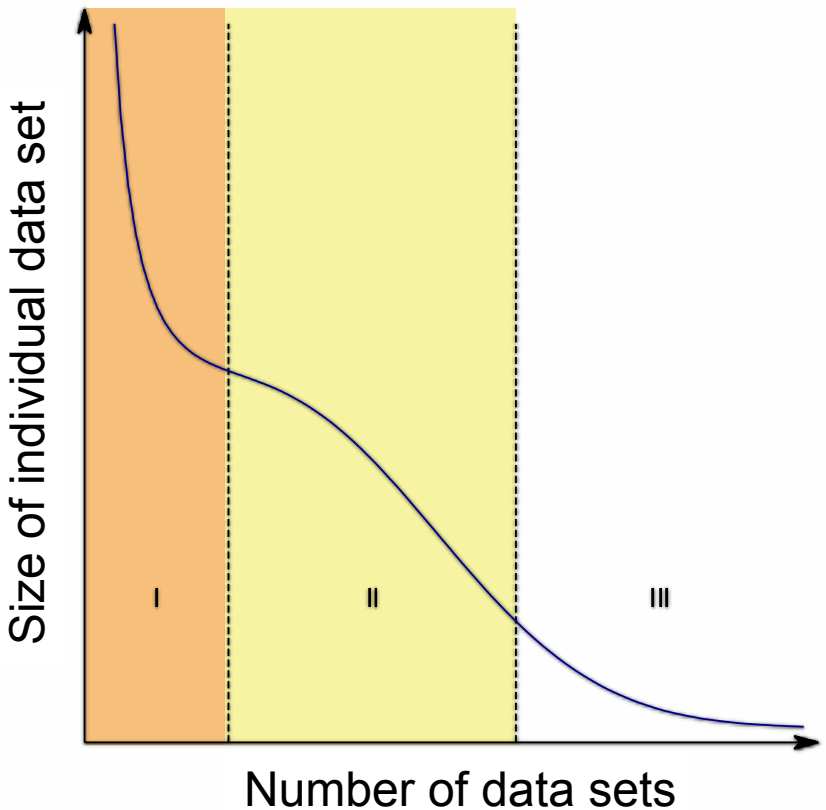
a tsunami of data...



Big Data in Astronomy and Astrophysics

- Raw data usually with clear data formats, huge in size
- Derived / processed data highly irregular in formats, large in size

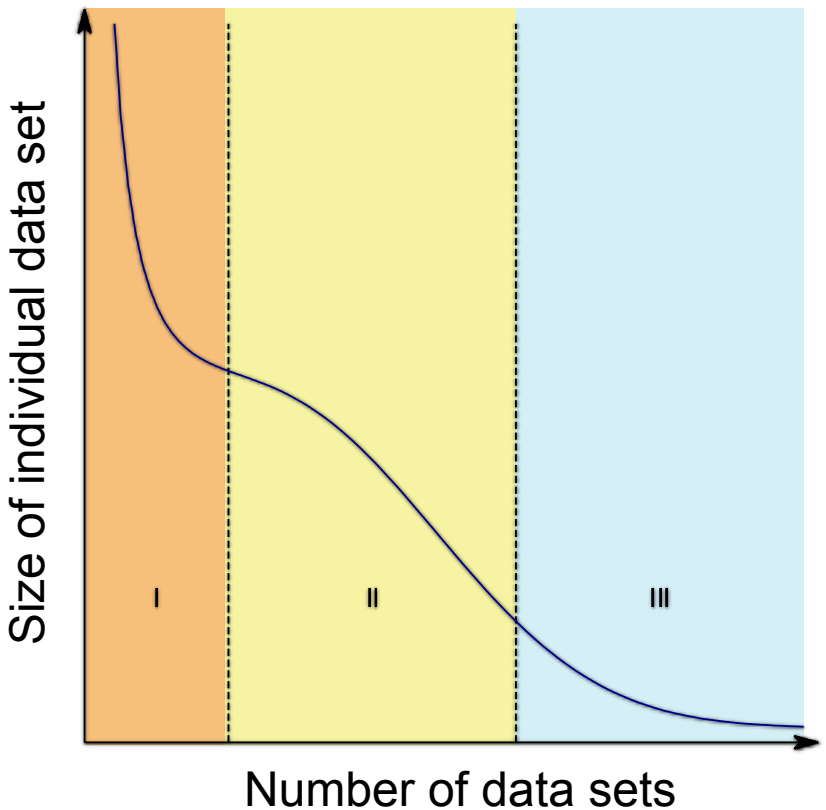
a tsunami of data...



Big Data in Astronomy and Astrophysics

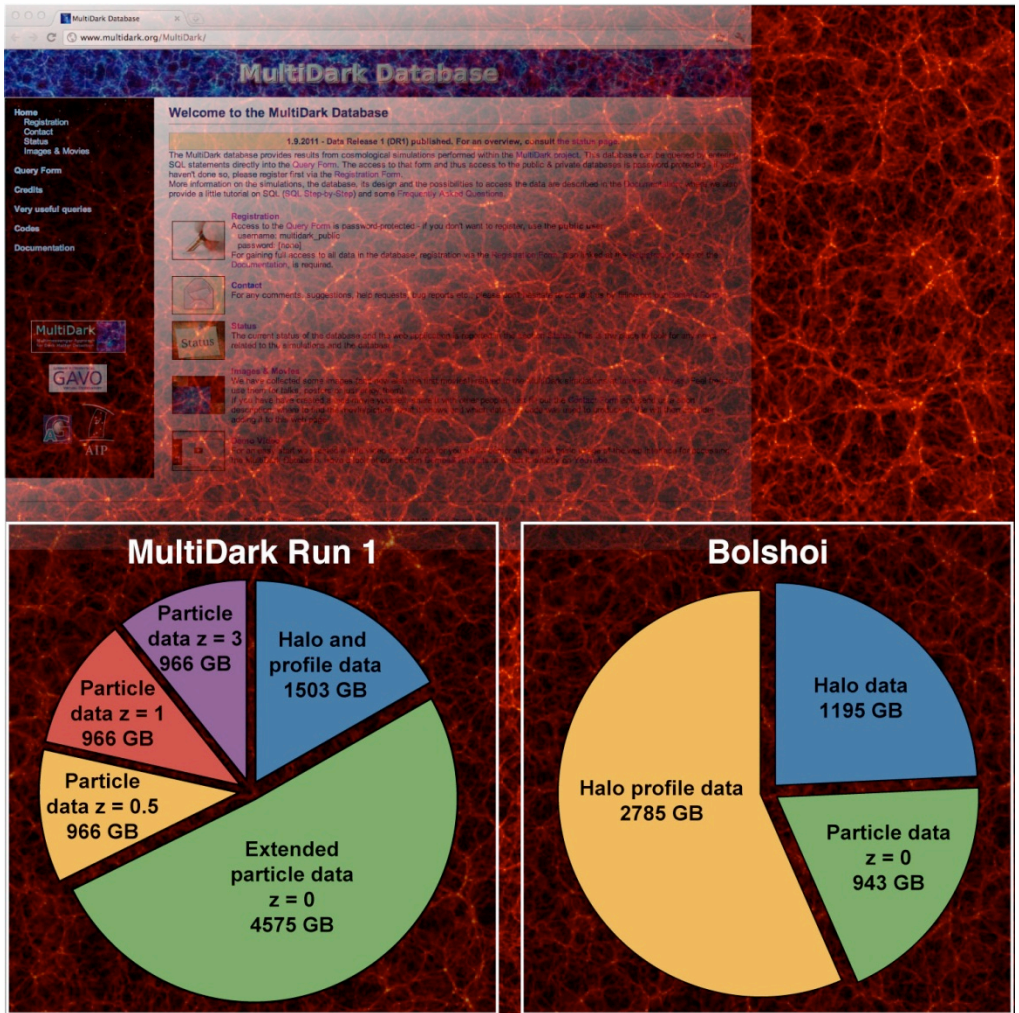
- Raw data usually with clear data formats, huge in size
- Derived / processed data highly irregular in formats, large in size
- Analysed results with complex formats, small in size

a tsunami of data...



www.multidark.org

- Total row count: $2.34\ 10^{11}$
- MS SQL Server
- Apache Tomcat
- similar setup to Millenium DB
- Most queries 100 sec
- significant amount of queries > 1000 sec



Problems while building the MultiDark DB

- Data ingest time:
  Need to convert binary to ASCII CSV format
  (highly inefficient)
- Data transformation:
  Computing values after ingest slow - best during ingest
- Data indexing:
  Index on particle data (~10^{10} particles) around one week
- Data retrieval times slow on full table scans:
  cannot build index for every query
- Spatial queries in 3D hard, impossible in 6D
  nearest neighbour search also inefficient

Why RDBMS?

- SQL - it took long time for the community to adopt SQL (we think this is the main problem with NoSQL)
- proven, widely available, large user base
- good for structured data

- Problems:
  - Built for different purposes (business, web, ...) result sets usually small - mostly in memory solutions
  - parallelisation of data / sharding
  - can be expensive

Our vision:

- Open source DB solution for scientific purposes:
  A one size fits all solution built by the community for the community

- Developments at AIP:
  - DB independent ingestion library and data transformation tool *(DBIngestor and AsciiIngest)*
  - MySQL sharding solution for scientific queries *(PaQu)*
  - MySQL plugins for data analysis, spatial queries and indexing, job queueing...
  - MySQL storage engine plugins for simulation raw data
  - Common web framework for data access *(Daiquiri)*

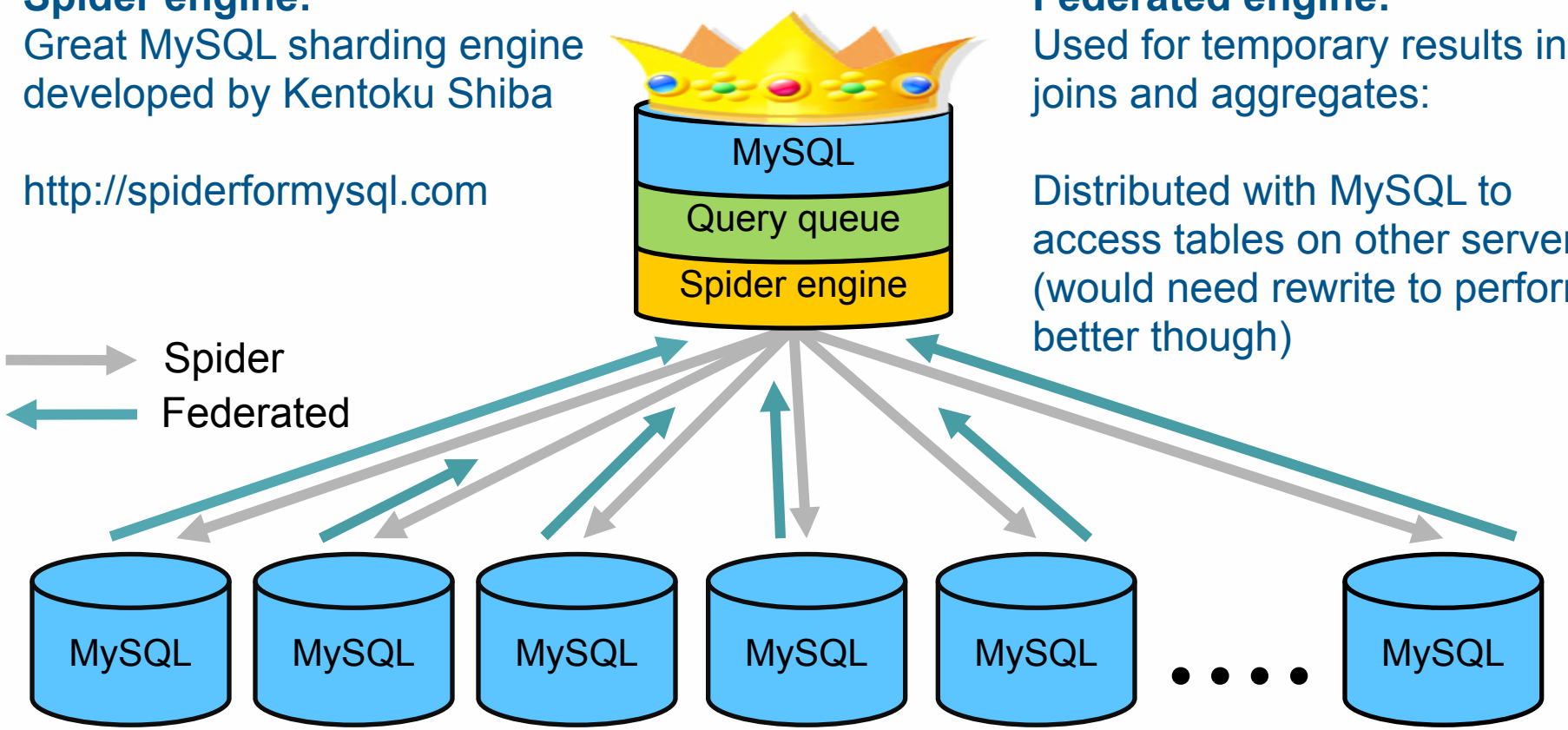
Our Developments with MySQL

Spider engine:
Great MySQL sharding engine developed by Kentoku Shiba

http://spiderformysql.com

Federated engine:
Used for temporary results in joins and aggregates:

Distributed with MySQL to access tables on other servers. (would need rewrite to perform better though)



Our Developments with MySQL

User

Admin

Daiquiri

PaQu

MySQL

Query queue

Spider engine

DBIngestor / AsciiIngest

Spider

Federated

MySQL

MySQL

MySQL

MySQL

MySQL

MySQL

PaQu: Parallel Query Reformulation

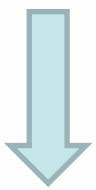
Implicit Joins:

```
SELECT    a.*, b.*, c.*
FROM      a, b, c
WHERE     b=2 AND
  b.id=c.b_id AND
  a.id=b.a_id;
```

```
SELECT    a.*, tmp.*
FROM      a,
    (SELECT  b.*, c.*
     FROM    c,
             (SELECT  b.*
              FROM    b
              WHERE b=2) as b
     WHERE   b.id=c.b_id)
WHERE     a.id=tmp.b.a_ids
```

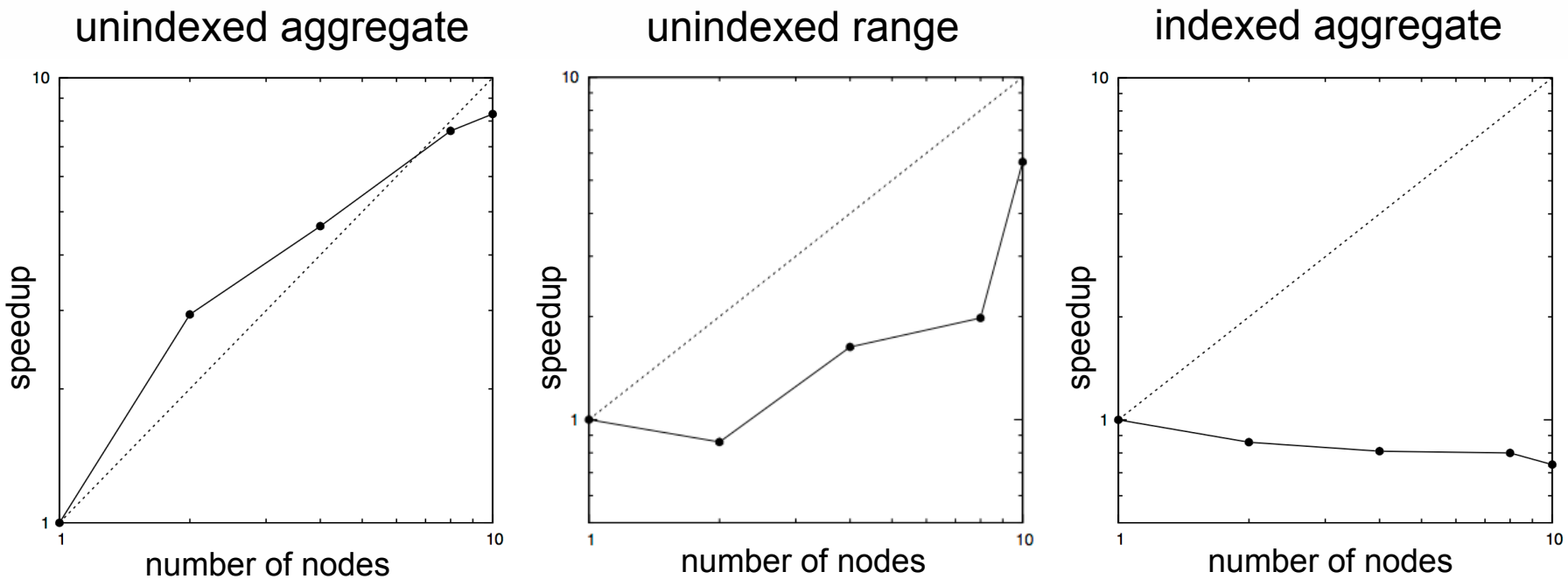
Aggregates:

```
SELECT    a.bar, AVG(a.foo)
FROM      a
GROUP BY  a.bar;
```



```
SELECT    a.bar,
          SUM(a.sum)/SUM(a.cnt)
FROM
   (SELECT   a.bar as bar,
             SUM(a.foo) as sum,
             COUNT(a.foo) as cnt
    FROM     a
    GROUP BY a.bar) as a
GROUP BY  a.bar;
```

Preliminary Performance Results



- Strong correlation with hardware setup:
  - Cache sizes, size of data files (smaller is better / partitioning?), network and I/O performance

Conclusions

- Scientific data increases rapidly in size
  ⇒ problems for off-the-shelf database systems
- Slow acceptance by scientific community of anything new poses problems for NoSQL solutions and favours SQL
- Open source solution for scientific use of open source databases needed
- Promising results for using MySQL + Spider + PaQu for scientific data mining and fast full table scans

AIP

Interested? Help us in the development!

DBIngestor/AsciiIngest: http://github.com/adrpar

Thank you!