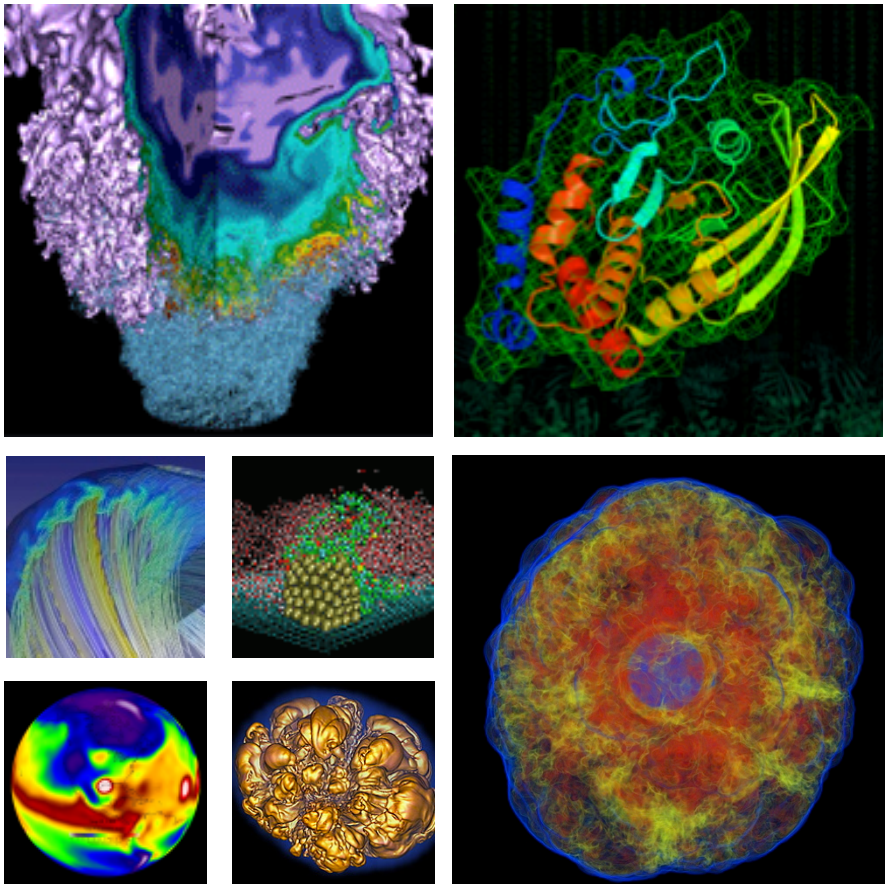
Accelerating Analysis with SciDB



Lisa Gerhardt NERSC, LBNL
Yushu Yao NERSC, LBNL
Carlos Faham LBNL

April 16, 2015

Experimental Data

- **Vast majority of experimental data is array-like**
  - Time series data from detectors (LHC, IceCube)
  - Image data from satellites and light sources
- **TBs to PBs of data, that is written once and read many times**
- **Reduced, aggregated, analyzed**
- **Strong need for a tool that can handle this kind of heavy workload**

SciDB

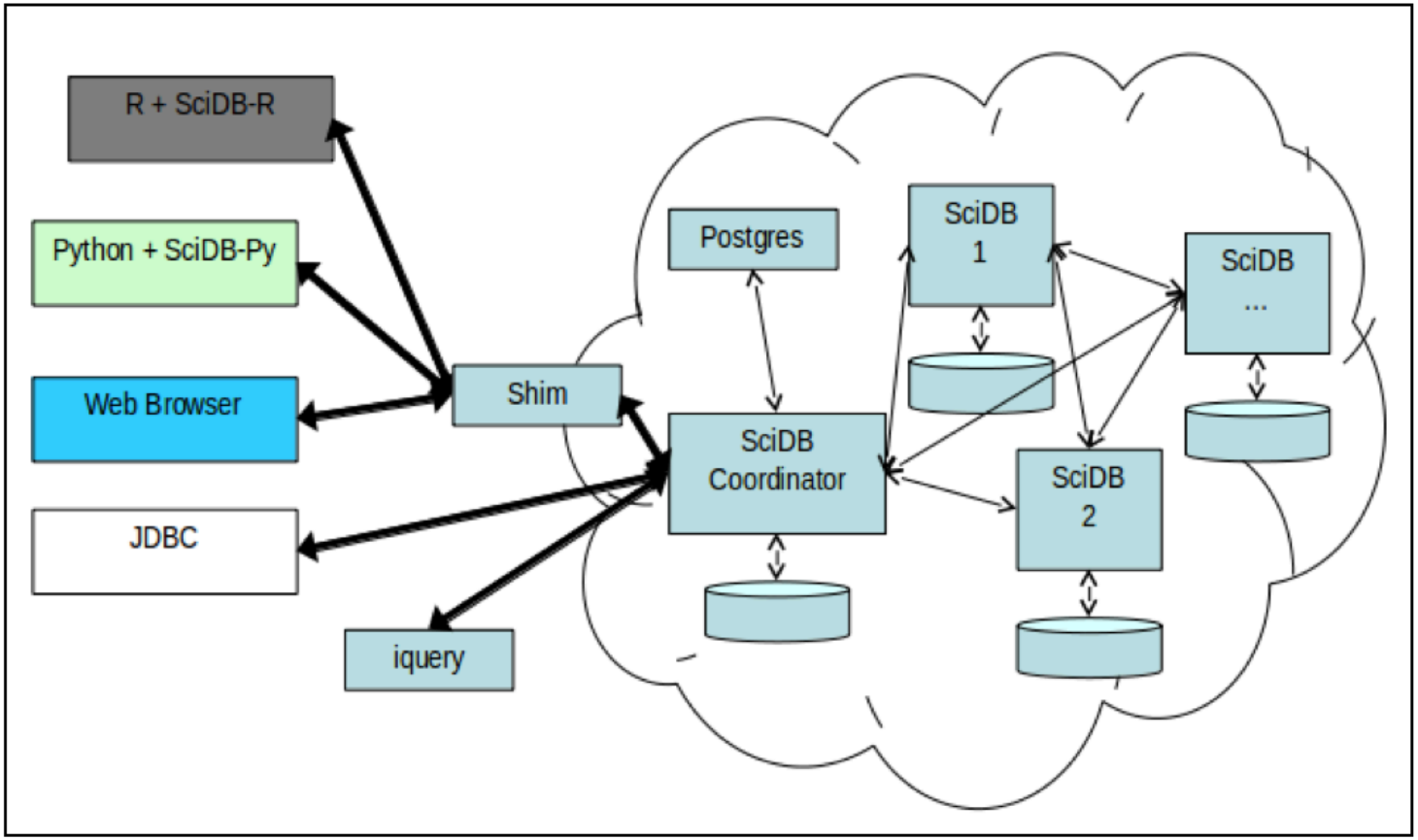
- **Array DB management system and analytic platform**
- **Shared nothing architecture**
- **Scalable out to 1,000s of processors and TBs - PBs of data**
- **Data is never overwritten, previous versions can be recovered**
- **Open source made by Paradigm4**
  - Enterprise version with a few more features

SciDB
paradigm4

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

SciDB Architecture



SciDB Structure

Arrays are chunked

Chunk 1

0.02	0.01	0.01	0.02
0.01	0.01	0.5	0.02
0.01	0.02	0.01	0.01
0.02	0.01	0.02	0.02

Chunk 2

0.02	0.01	0.01	0.02
0.01	0.01	0.5	0.02
0.01	0.02	0.01	0.01
0.02	0.01	0.02	0.02

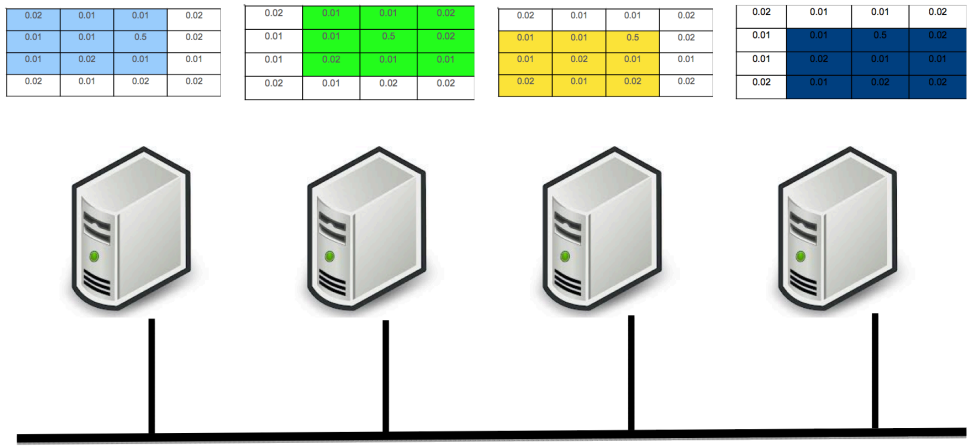
Chunk 3

0.02	0.01	0.01	0.02
0.01	0.01	0.5	0.02
0.01	0.02	0.01	0.01
0.02	0.01	0.02	0.02

Chunk 4

0.02	0.01	0.01	0.02
0.01	0.01	0.5	0.02
0.01	0.02	0.01	0.01
0.02	0.01	0.02	0.02

Each chunk is stored on a different instance



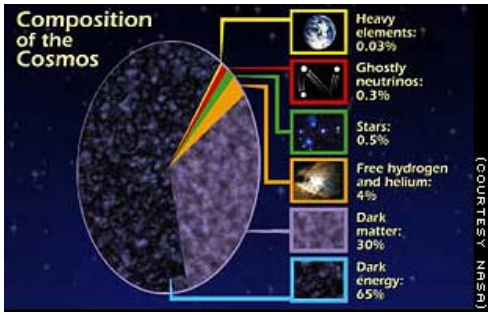
Data is scanned from array storage on all instances and streamed into and out of each operator one chunk at a time. “Hot” chunks are kept in memory.

Uploading Data

- **Load data as a flat array**
  - Supports binary, csv, tsv, SciDB text format
  - Re-dimension into desired array shape
- **Can do a parallel or single load**
  - Manually split data into pieces and upload to each instance
- **Choose commonly used data attributes to serve as array dimensions**

SciDB Case Study: LUX

Searching for dark matter with 370 kg of liquid Xenon



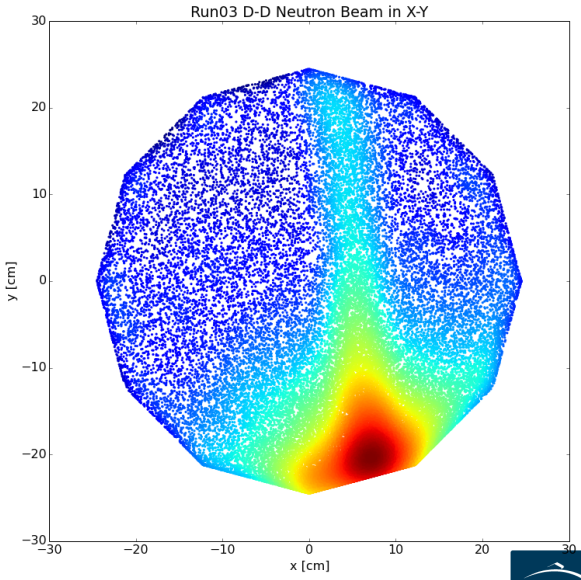
Precursor to LZ, chosen by DOE as one of three next generation dark matter detectors

U.S. DEPARTMENT OF ENERGY | Office Science

BERKELEY LAB

LUX's SciDB Setup at NERSC

- **Proof of concept with 32 instances spread over 16 nodes**
  - Work with LUX to develop representative solutions
- **Uploaded 10 TB of inaugural data**
- **83 Million events, 600 million pulses**
- **Dimension of array:**
  - Unique timestamp
  - Pulse type
  - Pulse number
- **Roughly 50 attributes**



Problem 1: Detector Stability

- Look at variation in measured electron pulse size over time
- Task: Histogram reported pulse sizes in each 5-hour window, fit it with a Gaussian, and return a plot of Gaussian mean versus time

How it's Done

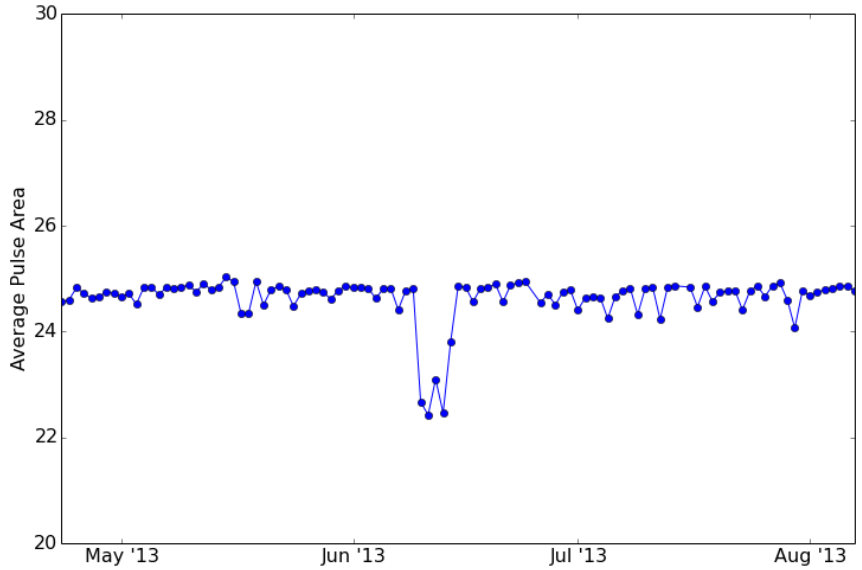
- **First, pull out all single electron pulse types and toss the rest of the attributes**

```
filter(project(pulse_array,pulse_size),
pulse_type==single_electron)
```

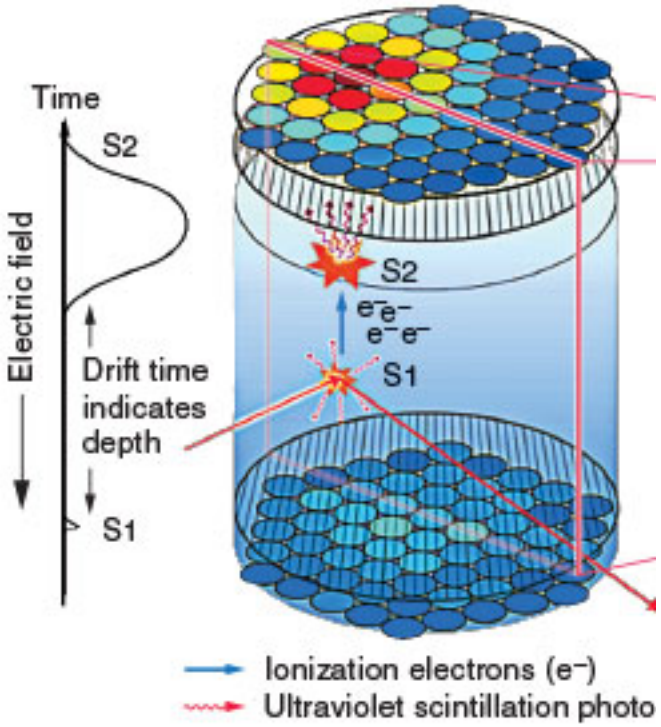
- **Gaussian fit needs a histogram, so we can leverage SciDB's regridding functionality**
- **Regrid to clump time into 5-hour window size and to get a count of pulse_size in each pulse_size bin**
  - Automatic count functionality is folded into regridding

Results

- **Result is ~2000 arrays that are time window histograms**
- **Spun over 200 M pulses (including scipy Gaussian fits for each slice) in 7 minutes**



Problem 2: Search for Signal



- **Task: Find every event with a small initial pulse followed by larger pulse within 1 ms**

The Beauty of Crossing

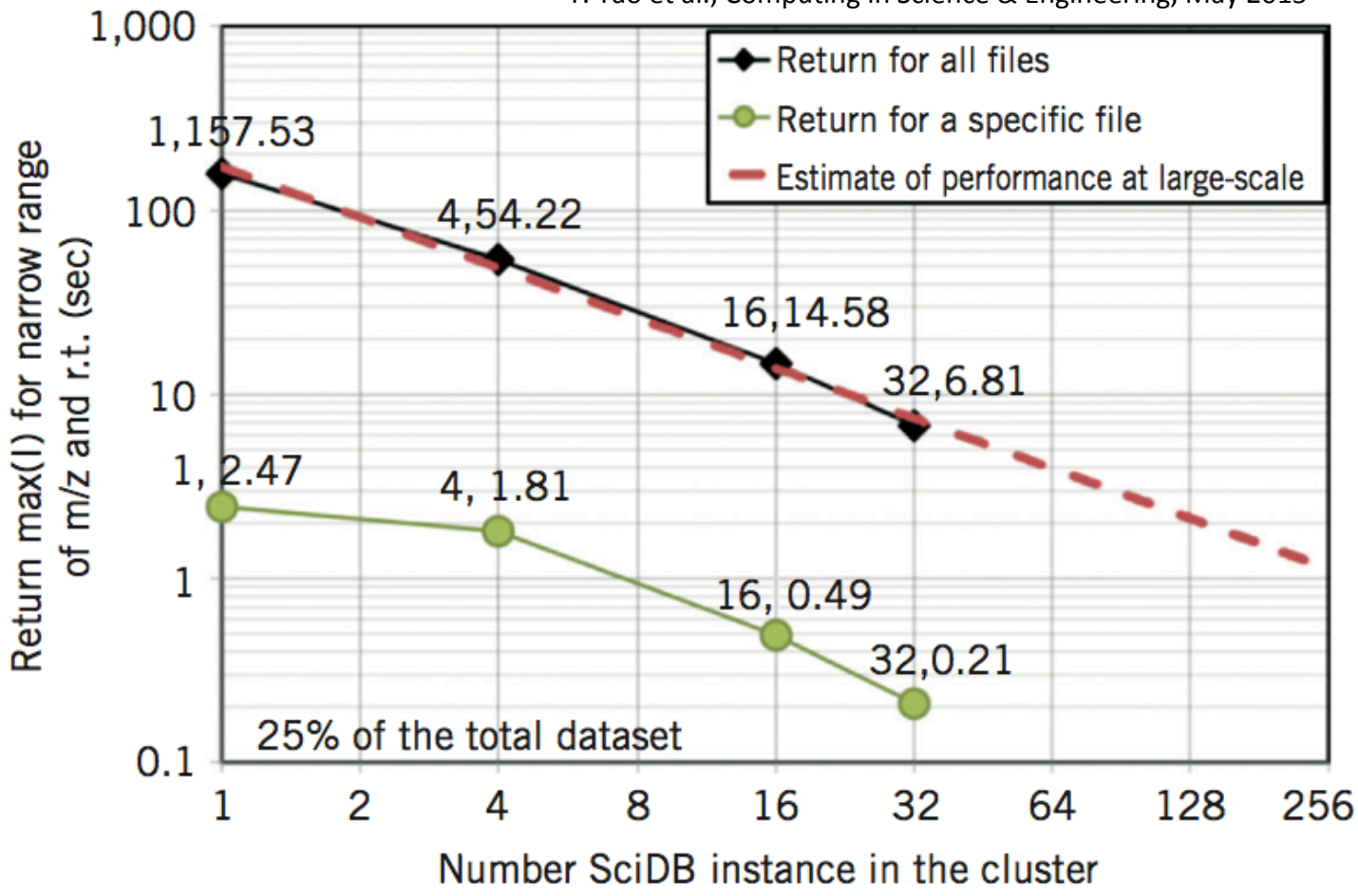
- **Produces an array with every possible combination of the two input arrays**

```
cross(small_pulse, big_pulse)
```

- **Split array into two sub arrays**
  - 300,000 entries with small peaks
  - 1,000,000 entries with large peaks
- **Use cross to join these two arrays**
- **Filter events that have a time difference of 1 ms or less**
- **Took only 4 hours to search 100 days of data for signal candidates**

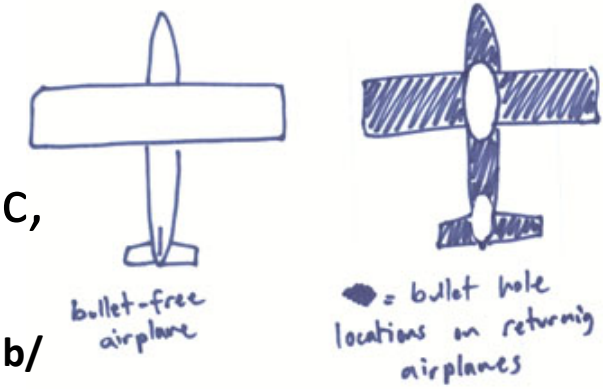
SciDB Scales Well

Y. Yao et al., Computing in Science & Engineering, May 2015



Production Level SciDB Cluster at NERSC

- **Wald cluster**
  - 12 worker nodes + 2 also serve as login nodes
  - 20-core, 64 GB RAM, IvyBridge processor
- **Roughly ~10 other groups, with space for more**
- **Also plans to run at scale on Edison**
- **Multiple Science Domains:**
  - Astronomy, Climate, Bio-imaging, Genomic, HEP



https://www.nersc.gov/users/computational-systems/testbeds/scidb/

Strengths / Weaknesses of SciDB

- **The Not-So-Good**
  - Can't read data from native formats
    - Loading data can take a while, though it can do it in parallel
  - Streaming analysis
    - Inefficient at outputting lots of data
  - Somewhat high threshold
    - Not everyone thinks in array joins
- **The Good**
  - Filtering and aggregating
  - SQL-like joins
  - Lots of optimized functions
    - Sampling
    - Matrix operations (ScaLAPACK)
    - Statistics
    - User designed functions
  - **Parallel architecture that is transparent to the user**

National Energy Research Scientific Computing Center

SciDB Test Cases at NERSC

Big Data Tasks in SciDB@NERSC Projects

	Bio-Informatics	Climate Simulation	Odetta	Variable Stars	OpenMSI	SDSS Spectrum Analysis
Querying	✔	✔	✔	✔	✔	✔
Density estimation						
Regression				✔		✔
Classification				✔		
Dimension reduction	✔		✔		✔	
Outlier detection						
Clustering	✔					
Time series analysis		✔		✔		
Feature selection & causality						
Fusion and matching					✔	✔

Y. Yao

Creating an Array

Create array ZOO <monkey:int64 zebra:int64 elk:int64> [cagenum=0:100,10,0, distemper=0:*,2,0]

Attributes: monkey, zebra, elk

Dimensions: cagenum, distemper (can be non-ints)

Dimension size: 0:100 (* unbounded)

Chunk size: 10

Chunk overlap: 0

Many Interfaces

- **Can talk directly to SciDB with special language**

  filter(between(ZOO,1,null,20,null),zebra=0))

- **Robust python and R interface**

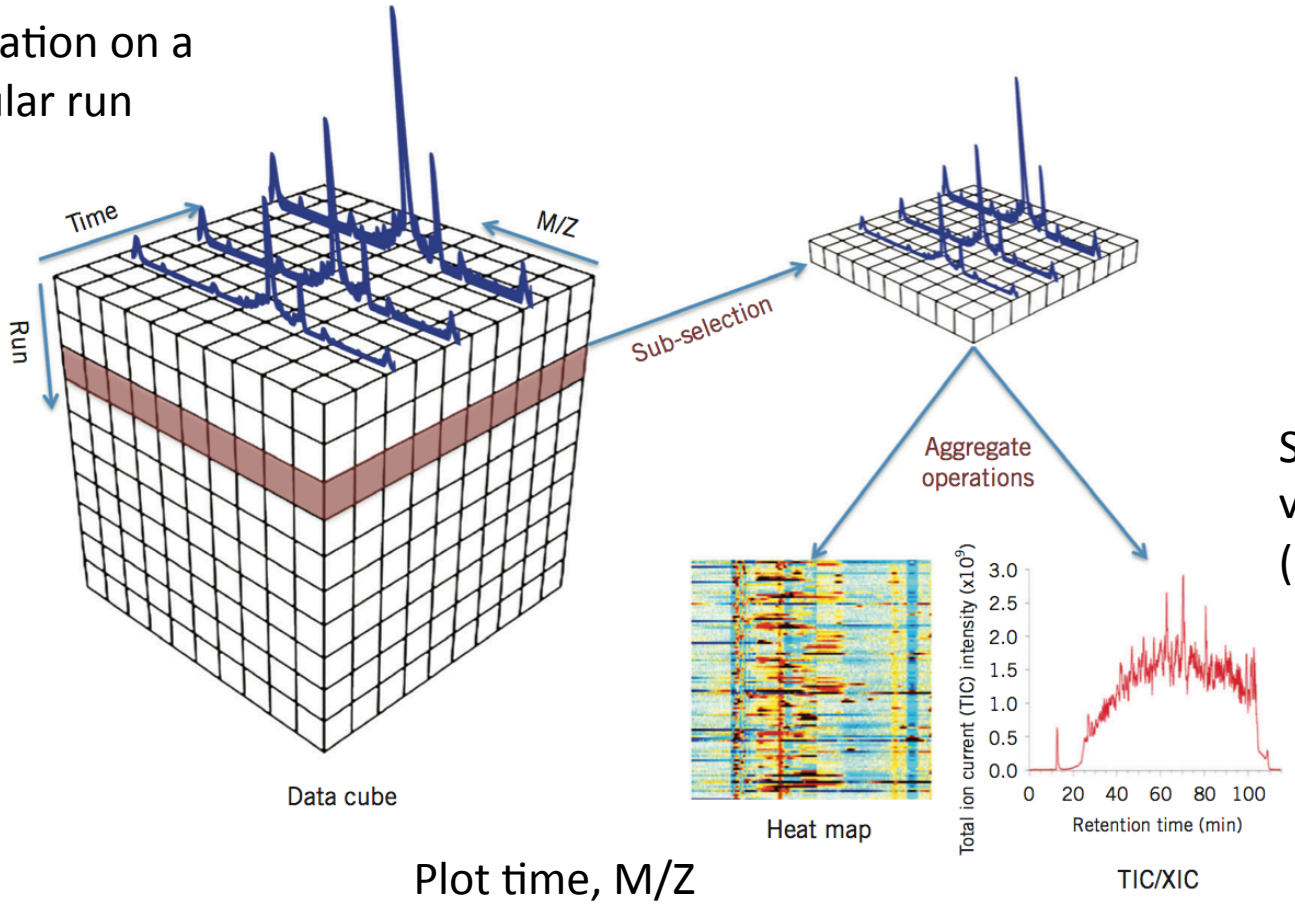
  R: subset(ZOO[1:20], "zebra == 0")

Metabolite Atlases

- **Profile molecular composition of metabolic by-products**
- **Many different metabolites and apparatus produce 10s of GBs of data a day**
  - At each timestep the intensity of mass spectrum reading for each M/Z is recorded
- **Want to use SciDB to compile an atlas of measurements that is searchable via web portal**

Common Uses

Pull all information on a particular run (filter)



Sum over all M/Z values (aggregate)

Plot time, M/Z heat map (regrid)

Picture from
Y. Yao et al., Computing in Science & Engineering, May 2015