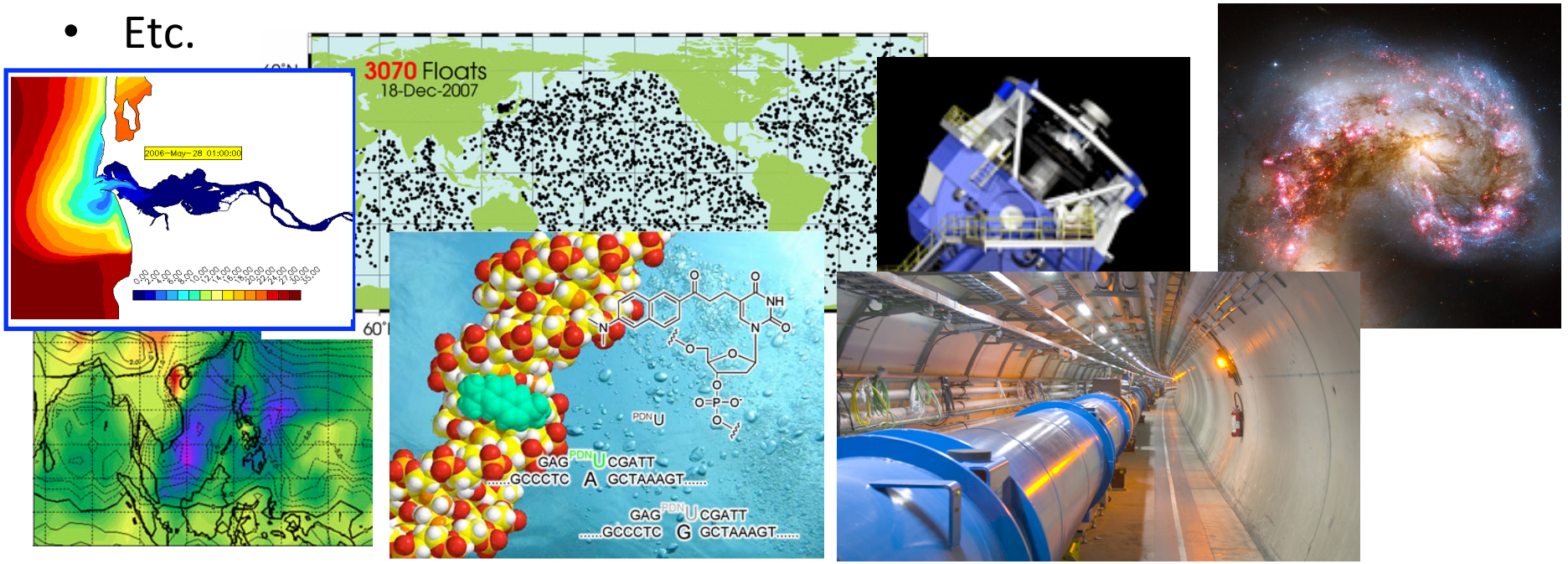
Big Data Management

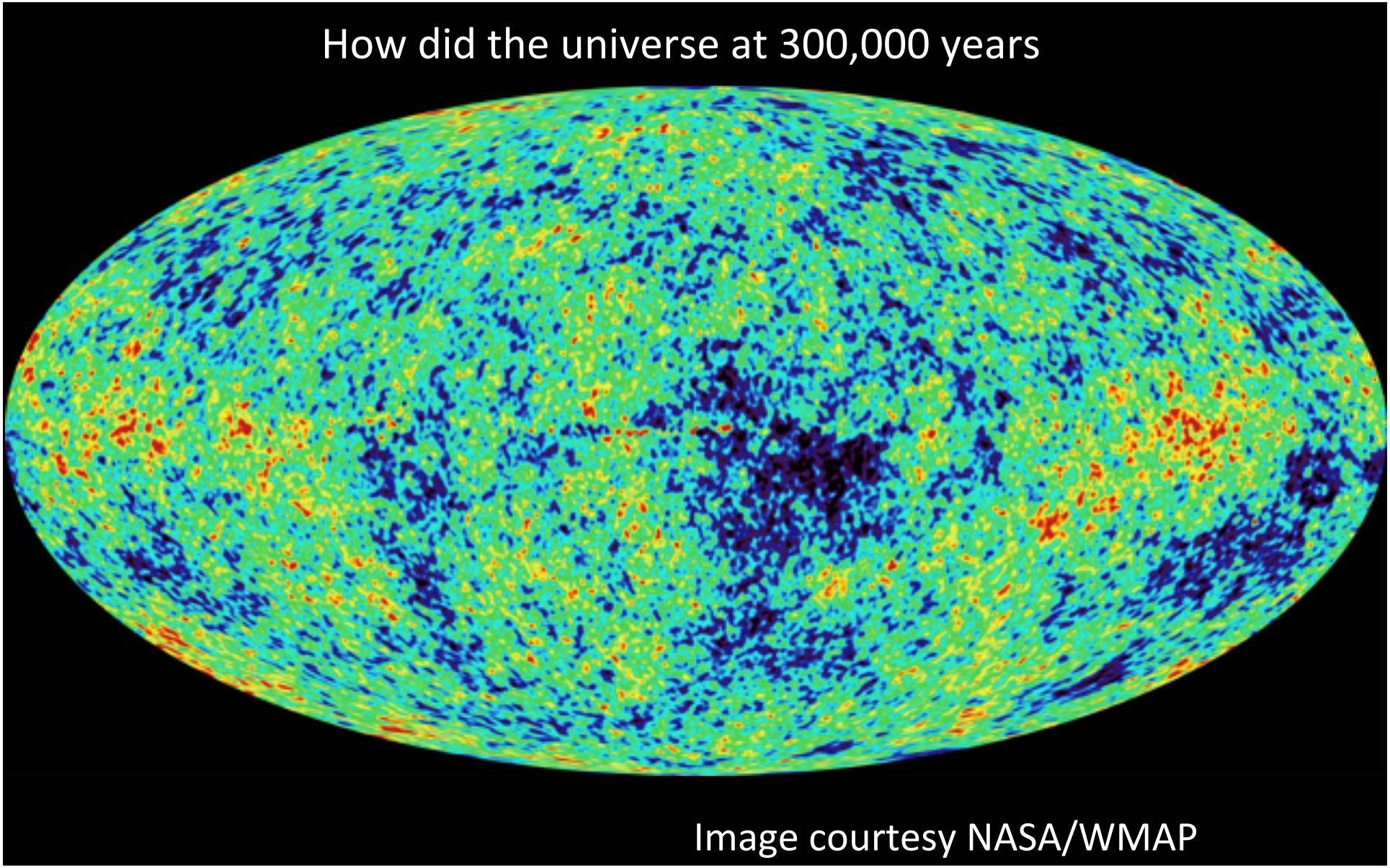
Magdalena Balazinska
Computer Science & Engineering
UNIVERSITY OF WASHINGTON
http://www.cs.washington.edu/people/faculty/magda

Science is Facing a Data Deluge!

- Astronomy: High-resolution, high-frequency sky surveys (SDSS, LSST)
- Medicine: ubiquitous digital records, MRI, ultrasound
- Biology: lab automation, high-throughput sequencing
- Oceanography: high-resolution models, cheap sensors, satellites
- Etc.

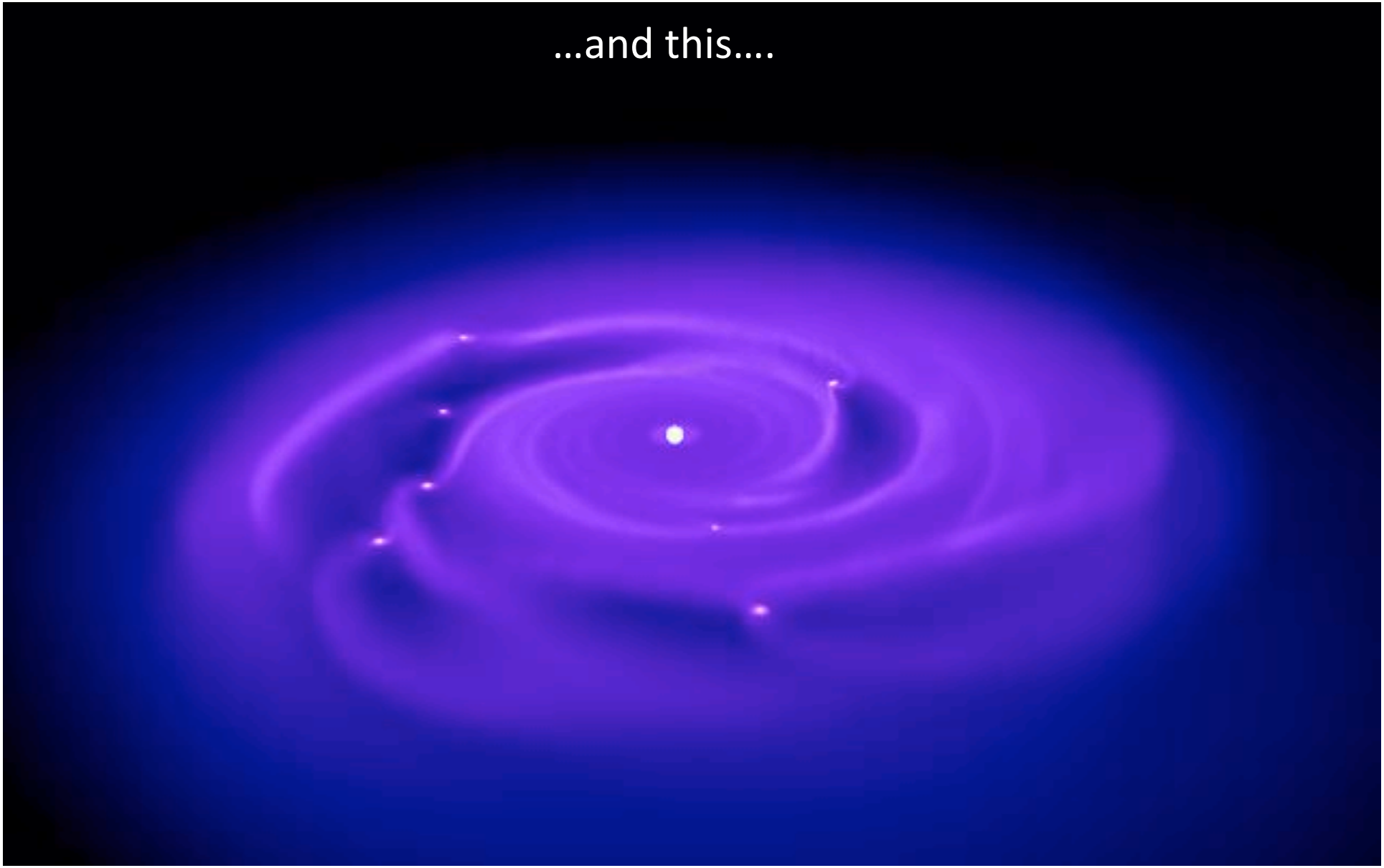


Example: Astronomy



Slide from Andrew Connolly (Astronomy Dept.)

Slide from Andrew Connolly (Astronomy Dept.)



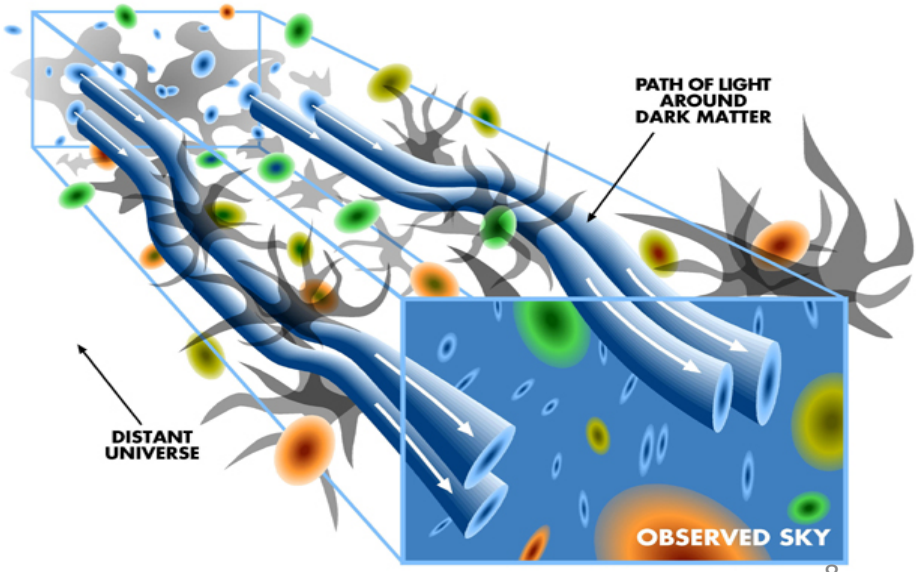
Slide from Andrew Connolly (Astronomy Dept.)

Slide from Andrew Connolly (Astronomy Dept.)

Dark Energy, Dark Matter and Baryons

Nature of the Universe

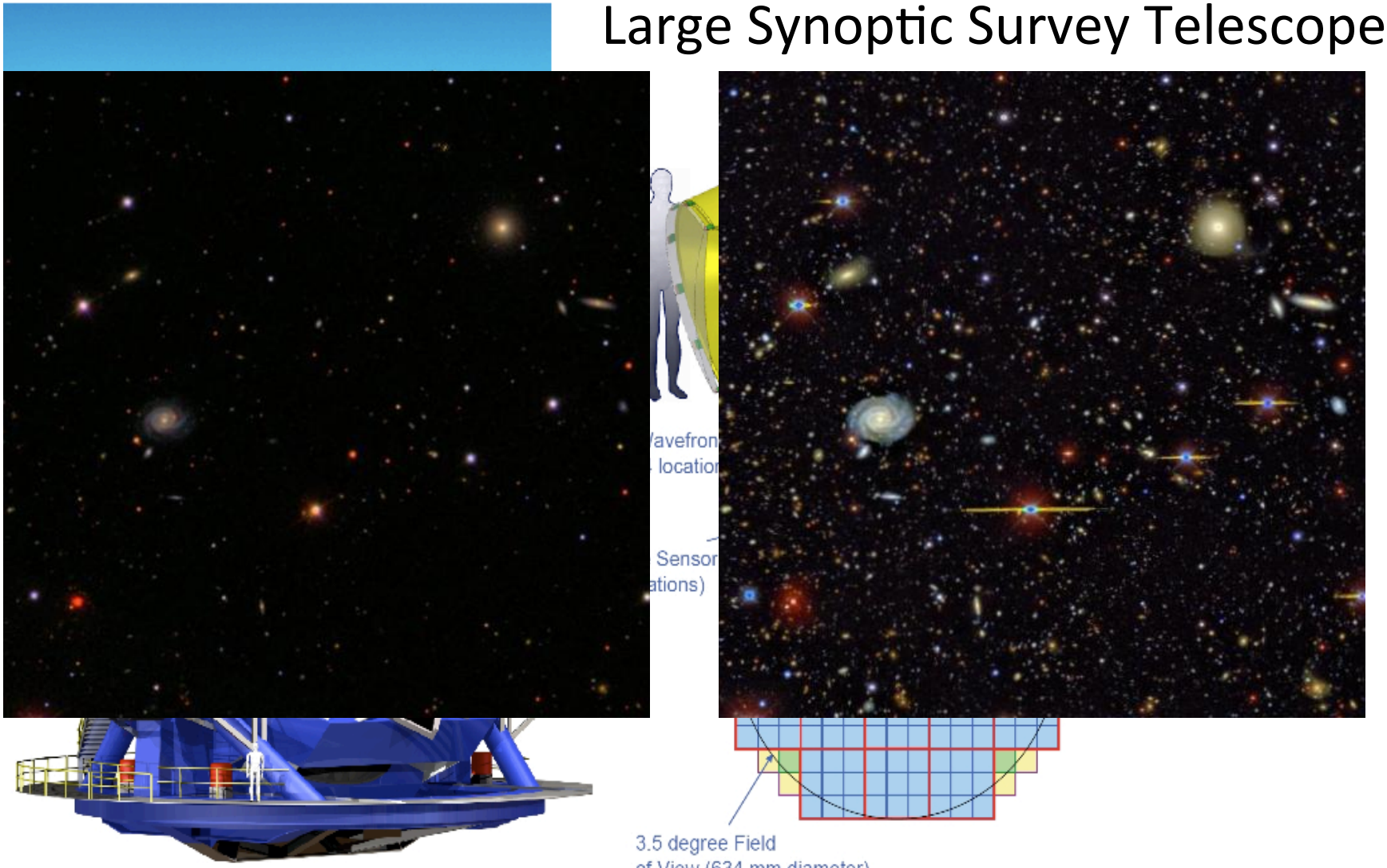
- Dark energy
  - 73% of energy density
  - Drives acceleration
  - **Physics unknown**
- Dark matter
  - 25% of energy density
  - Drives growth of structure
  - **Particle unknown**
- Small effects
  - Signals are small
  - Systematics can be large
  - Image distortions measured to 1 part in 10^5 (100x better than today)



Slide from Andrew Connolly (Astronomy Dept.)

Method 1: Observe the Sky

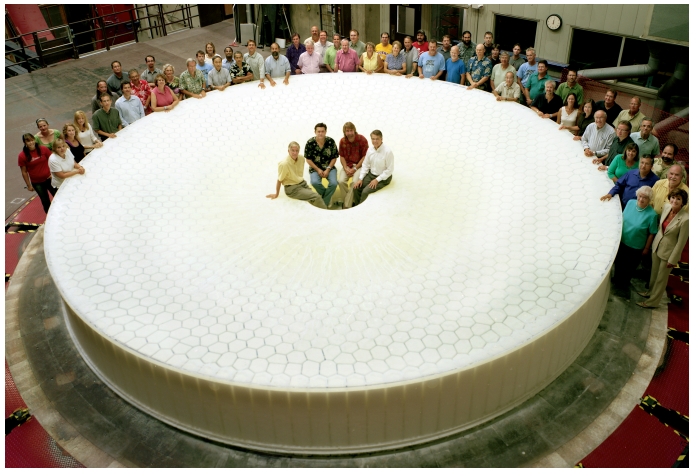
Large Synoptic Survey Telescope



Slide from Andrew Connolly (Astronomy Dept.)

Challenges from new astronomy

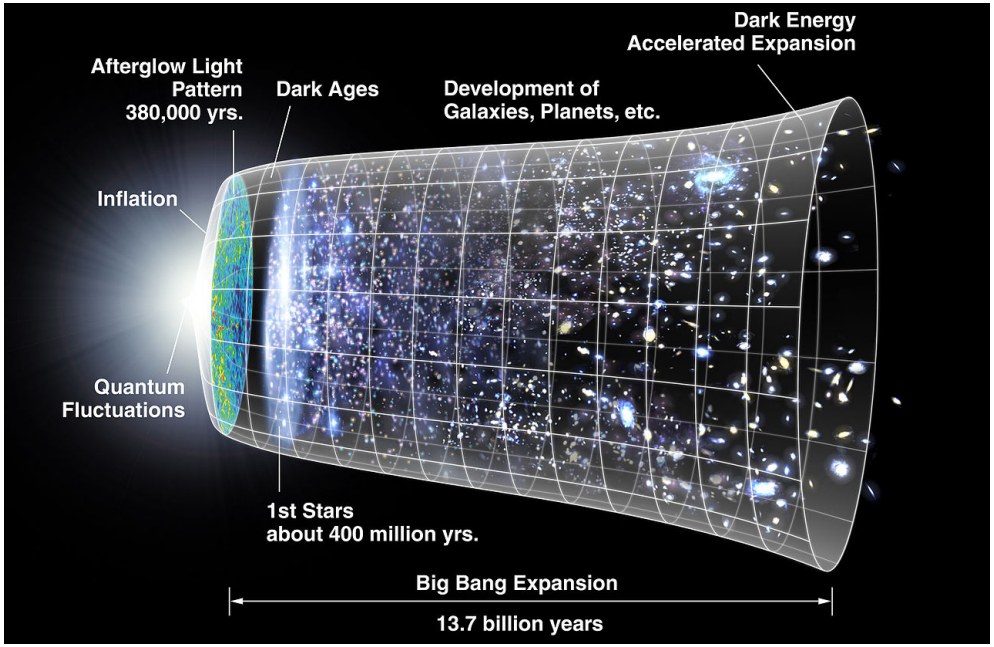
- Sloan Digital Sky Survey (SDSS)
  - 7 years of imaging
  - 8000 sq degrees of the sky (1/5th)
  - 200 million stars and galaxies
  - **80 TB raw images**
- LSST data flow
  - 20,000 sq degrees every 3 nights
  - **40 TB of imaging per night**
  - 10^8 sources a night (10^3 "events")
  - 1000 repeat observations over 10 years
  - 10 Petabytes of catalogs (10 years)
  - 100 PBs of images
  - 5 months to watch 1 year of data (HDTV)
  - Data public as soon as taken



Slide from Andrew Connolly (Astronomy Dept.)

Method 2: Simulate

- Evolution of large-scale structure in the universe
  - Universe is a set of particles (gas, dark matter, stars)
  - Output snapshot every few simulation timesteps



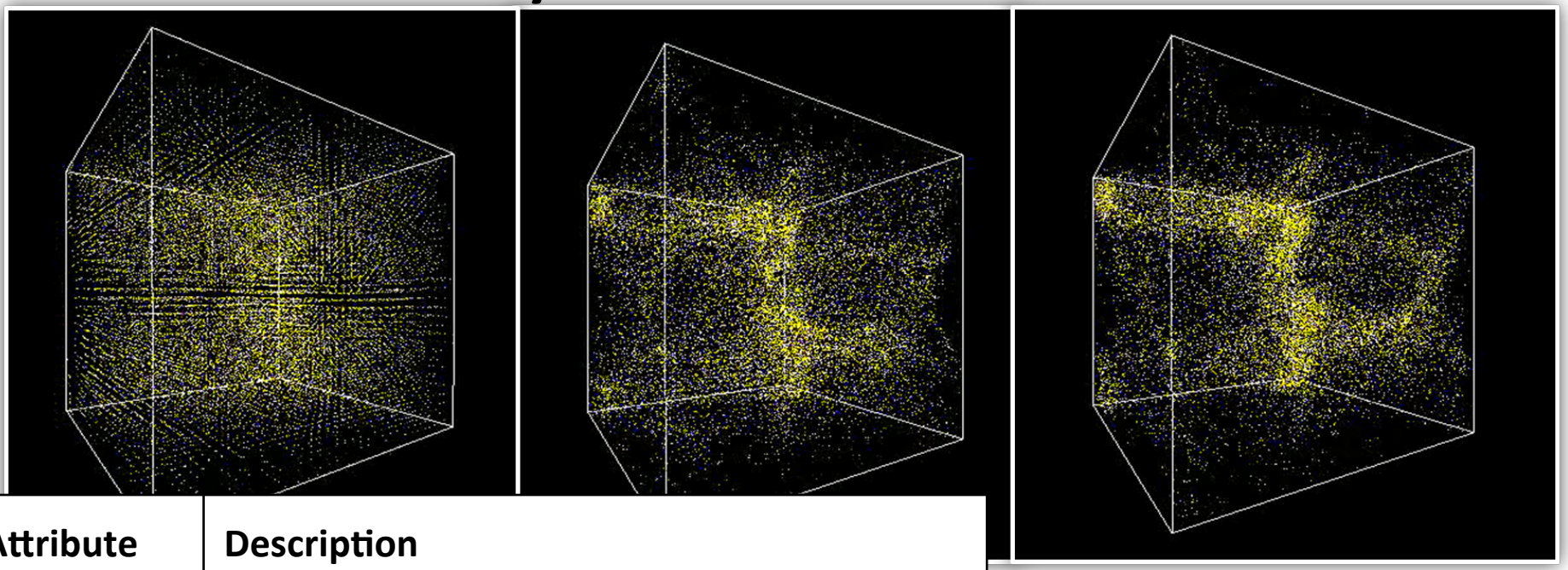
Typical simulations today

Spring 2014: 5TB

Fall 2014: 200TB

UW N-body group

N-Body Simulations Data

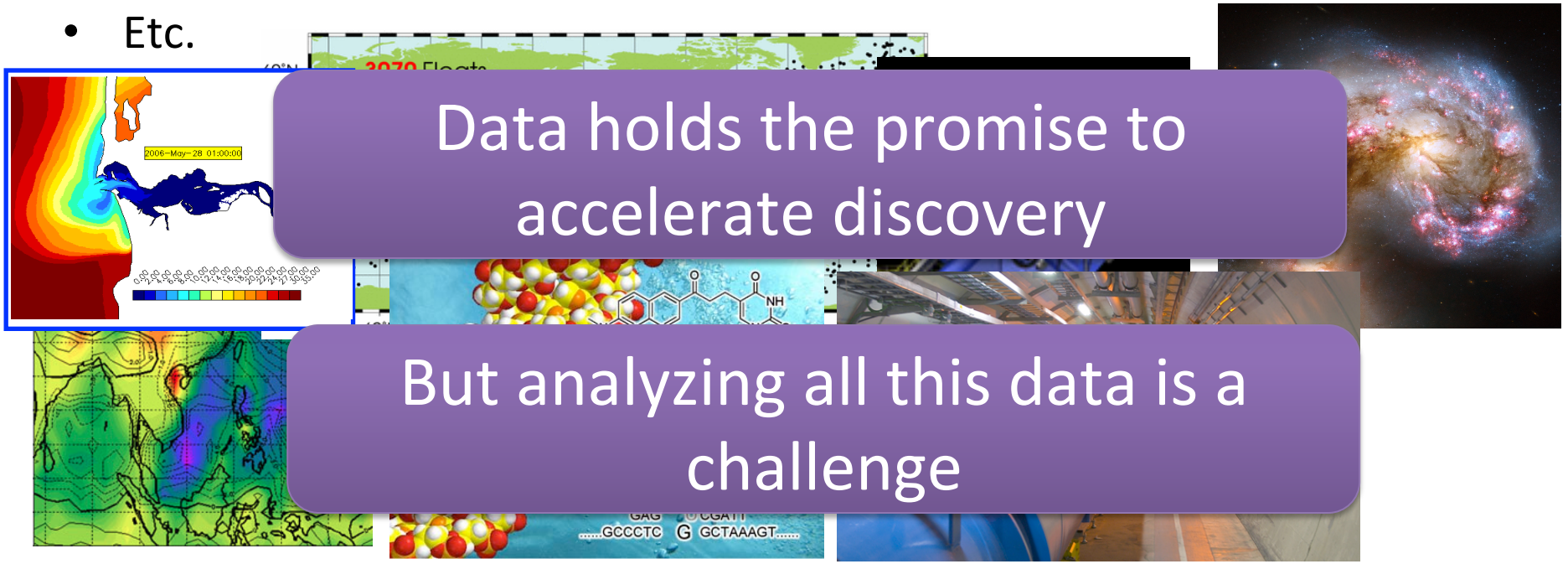


Attribute	Description
iOrder	unique identifier
X, Y, Z	position in Cartesian coordinates
Type	type of particle: either dark, star or gas
Grp	halo group particle belongs in
Time	Timestep the particle belongs in

Universe simulation
46 snapshots
5 TB

Big Data Need in All Sciences!

- Astronomy: High-resolution, high-frequency sky surveys (SDSS, LSST)
- Medicine: ubiquitous digital records, MRI, ultrasound
- Biology: lab automation, high-throughput sequencing
- Oceanography: high-resolution models, cheap sensors, satellites
- Etc.

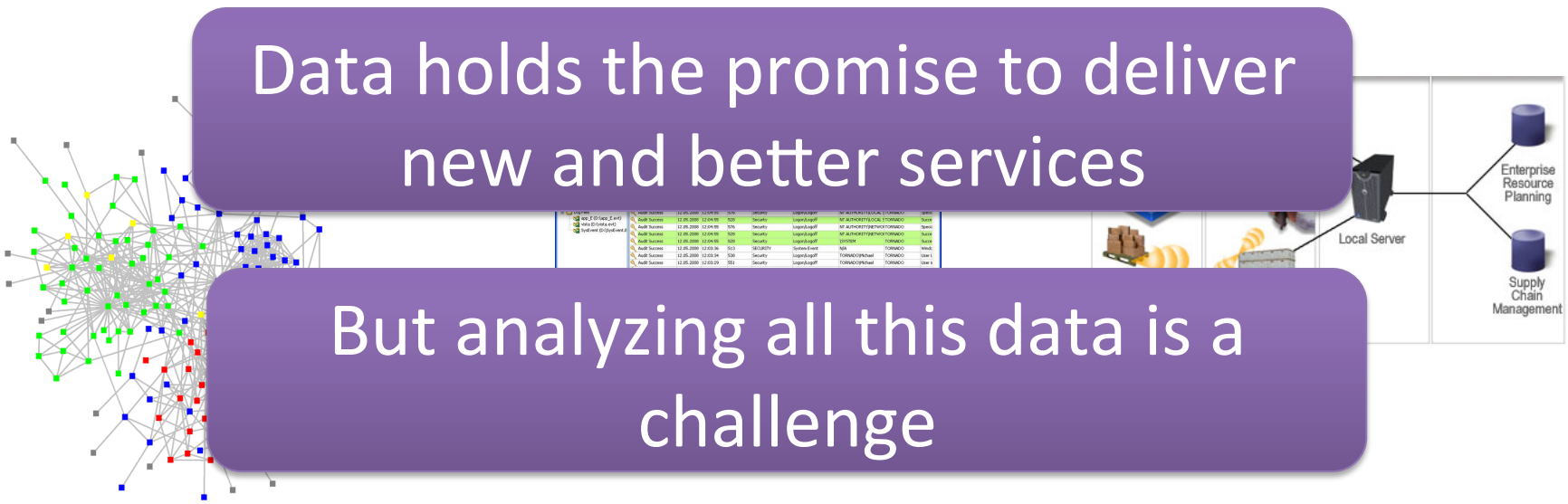


Data holds the promise to accelerate discovery

But analyzing all this data is a challenge

Industry is Facing a Data Deluge!

- Clickstreams, search logs, network logs, social networking data, RFID data, etc.
- Examples: Facebook, Twitter, Google, Microsoft, Amazon, Walmart, etc.



How Can We Analyze All this Data

- Excel? Limit of 65,536 rows by 256 columns
- Write my own Java/Python/other program?
  - Not all data fits in memory
  - May want to run in parallel in a cluster
  - Do I really want to write a program for each question?
- Use an existing database management system
  - Today's data **V**olume, **V**elocity, and **V**ariety are breaking these systems… new systems are appearing
- Build a new database management system!

Myria Big Data Management Service

Myria is a Cloud service: Just open browser and go!

Myria Editor Queries Datasets Report an issue vega.cs.washington.edu:1776 [72/72]

Write your code here, perhaps starting from one of the examples at the right.

```
1  OppData = scan(armbrustlab:seaflow:all_opp_v3);
2  VctData = scan(armbrustlab:seaflow:all_vct);
3
4  OppWithPop = select opp.*, vct.pop
5               from OppData as opp,
6                    VctData as vct
7               where opp.Cruise = vct.Cruise
8                 and opp.Day = vct.Day
9                 and opp.File_Id = vct.File_Id
10                and opp.Cell_Id = vct.Cell_Id;
11
12 PlanktonCount = select Cruise, COUNT(*) as Phytoplankton
13                 from OppWithPop
14                 where pop != "beads" and pop != "noise"
15                   and fsc_small > 10000;
16
17 store(PlanktonCount, JustX);
```

Parse | Execute the Query | Myria JSON

Query Language: MyriaL

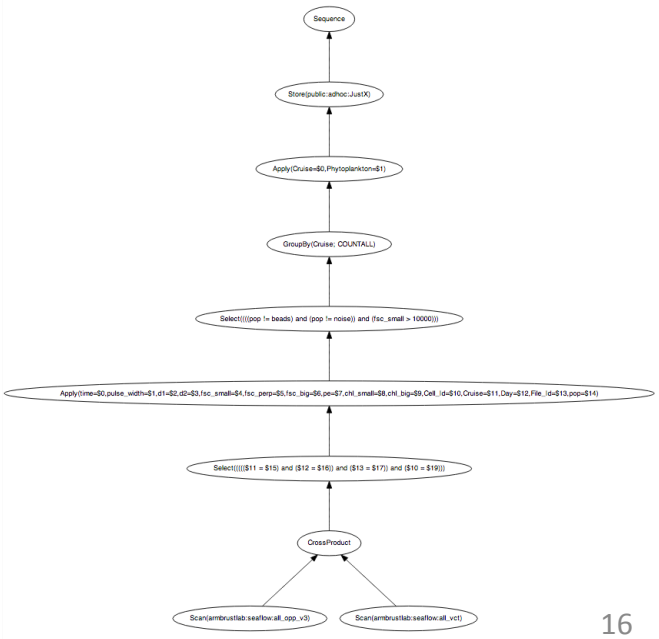
Profile query

Profiling will make the query run a little bit slower but allows you to examine exactly how the query was executed.

Examples | Datasets | Query Plan | Results

Relational Algebra | Physical Plan

Code parsed as Relational Algebra



Myria is a Cloud Service

MyMergerTree

Browser

Specialized Services

MyriaQ

Web UI

Language Parser

Logical Optimizer for RA+While

Myria Compiler | C Compiler | Grappa

Google App Engine

MyriaX

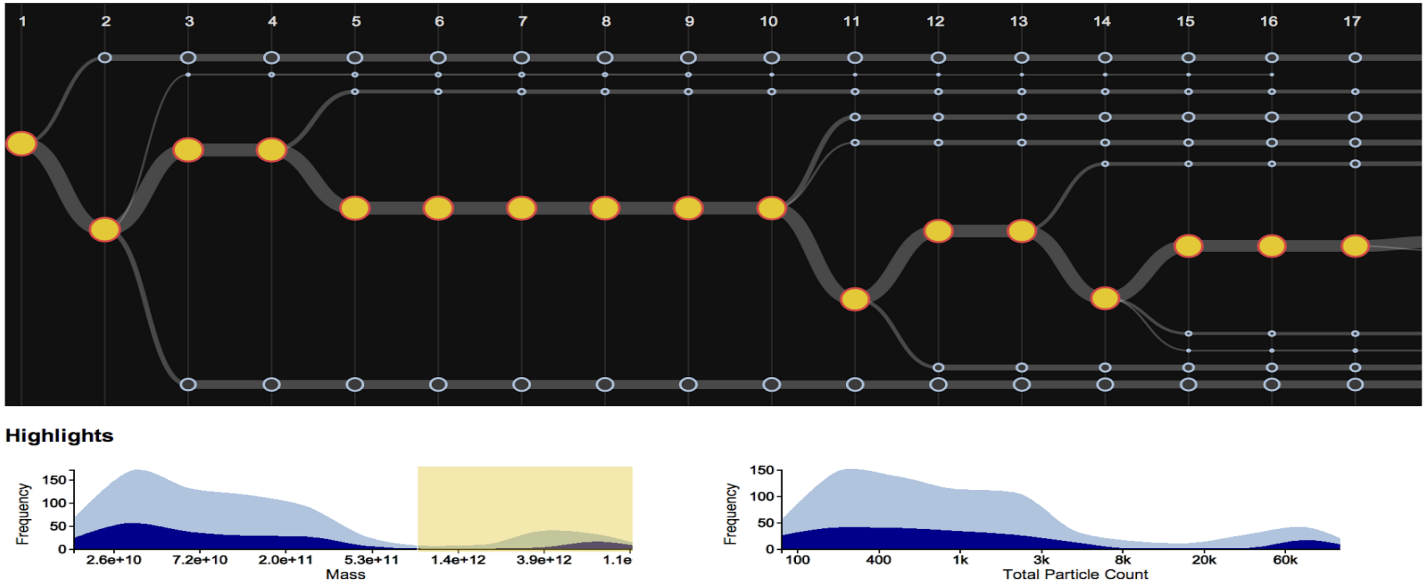
Query Execution

Example: MyMergerTree

Cloud Service for analyzing galactic merger trees

Built on top of Myria

Used to analyze a 5TB dataset



Myria is Expressive

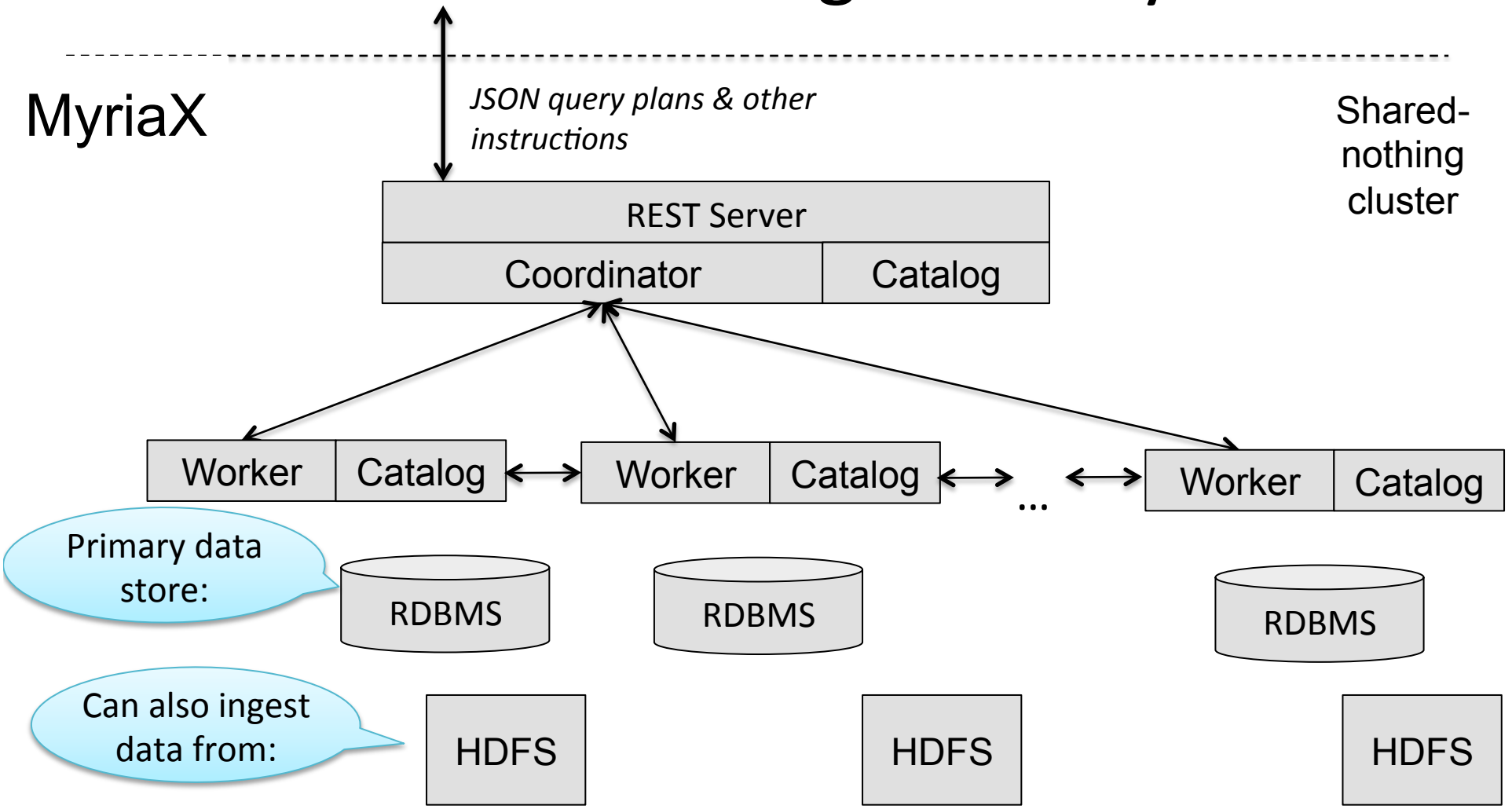
Myria is expressive

- SQL, Datalog, MyriaL

Quick example (connected components)

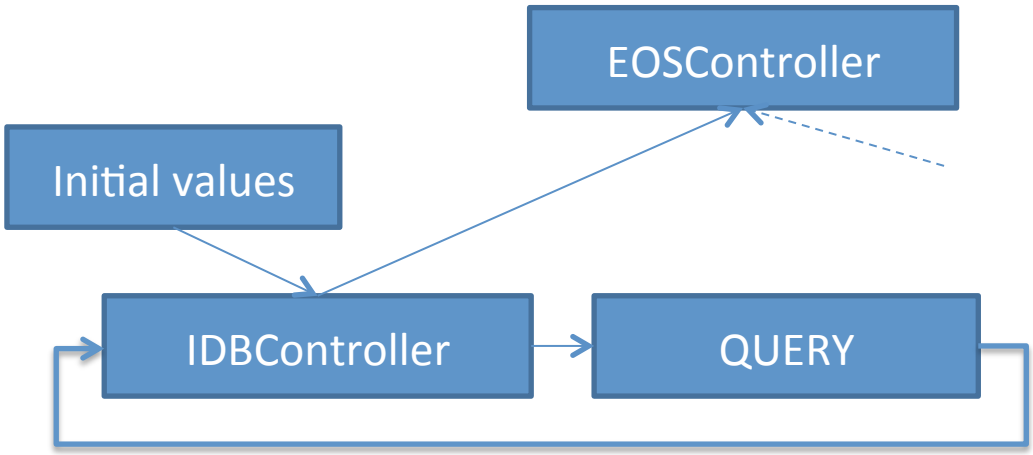
```
edgesRaw = scan(astro:cosmo50:cosmo50CompleteEdges);
edgesAll = [from edgesRaw as E where E.NowGroup = 52
            emit E.CurrentTime, E.CurrentHalo, E.NextHalo, E.SharedParticlesCount];
edgesNew = [from edgesAll as E where E.CurrentTime == 1
             emit E.CurrentTime, E.CurrentHalo, E.NextHalo, E.SharedParticlesCount];
edgesResult = edgesNew;
I = [1 as i];
do
   delta = [from edgesNew as E1, edgesAll as E2, I
   where E1.NextHalo == E2.CurrentHalo and E1.CurrentTime == I.i and E2.CurrentTime == I.i+1
   emit E2.CurrentTime as CurrentTime, E2.CurrentHalo as CurrentHalo, E2.NextHalo as NextHalo,
   E2.SharedParticlesCount as SharedParticlesCount];
   edgesResult = unionall(delta, edgesResult);
   edgesNew = delta;
   I = [from I emit i+1 as i];
while [from I emit min(i) <= 10];
store(edgesResult, finalAnswer);
```

Myria is a Parallel Data Management System



Example Specific Research Problems

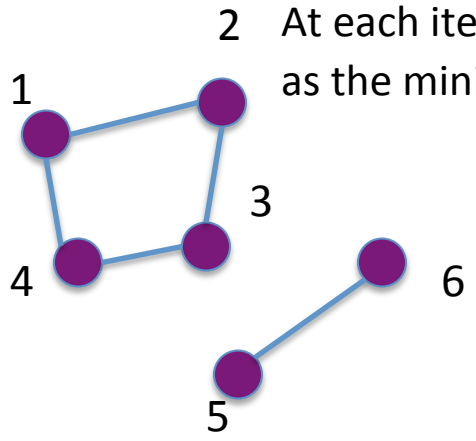
Efficient Iterative Processing with Failures



How to execute queries?

- Bulk synchronous
- Incremental synch.
- Prioritized incremental synch.
- Asynchronous

Connected components in a graph



At each iteration, update node ID as the minimum ID of all neighbors

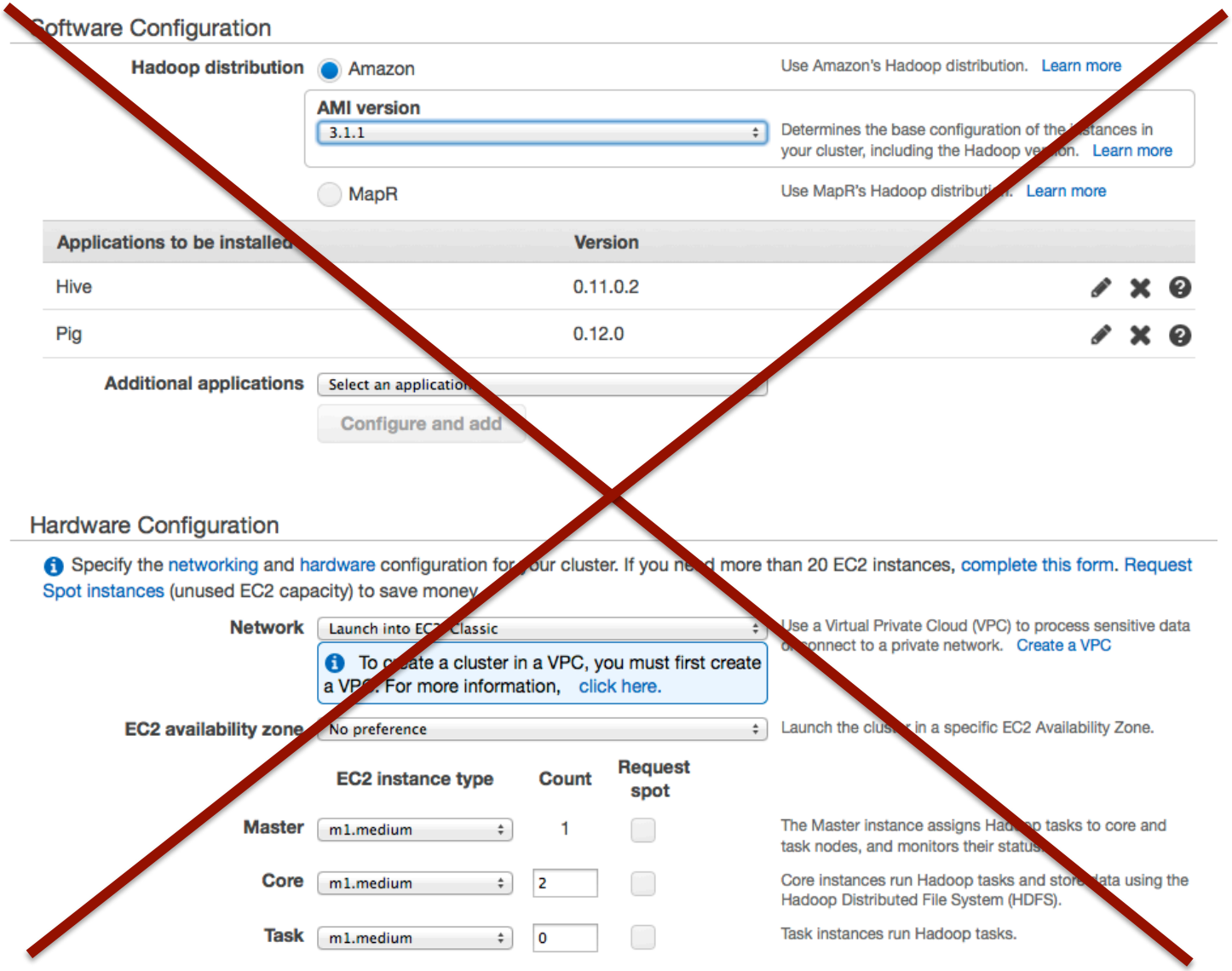
How to handle failures?

Extend shuffle operators

- Buffer and replay
- Buffer, combine, and replay
- Buffer, combine, and prioritize
- Buffer, combine, rebuild state first

Myria Changes the Face of Cloud Services

Today:



Myria Changes the Face of Cloud Services

With Myria: Personalized Service Level Agreements

Available SLAs

Tier #1

Query Template	Runtime (seconds)
SELECT (22 ATTR.) FROM (JOIN 4 Tables) WHERE '<10% CONDITION' SELECT (22 ATTR.) FROM (JOIN 5 Tables) WHERE '<1% CONDITION' SELECT (27 ATTR.) FROM (JOIN 5 Tables) WHERE '<0.1% CONDITION'	60
SELECT (9 ATTR.) FROM (JOIN 4 Tables) SELECT (27 ATTR.) FROM (JOIN 5 Tables) WHERE '<10% CONDITION'	120
SELECT (27 ATTR.) FROM (JOIN 5 Tables)	180

Purchase @ $0.35/hour

Tier #2, benefits in addition to Tier#1

Query Template	Runtime (seconds)
SELECT (27 ATTR.) FROM (JOIN 5 Tables) WHERE '<10% CONDITION'	60
SELECT (27 ATTR.) FROM (JOIN 5 Tables)	120

Purchase @ $0.89/hour

Conclusion

- We live in the “big data era”
- We need to build new data management and analytics systems to handle the data
- We work with domain scientists and companies to use these tools and enable new discoveries and new services

Acknowledgments

- The University of Washington (UW) Database Group
- UW eScience
- UW astronomy
- And many others

- Our sponsors: NSF, ISTC Big Data, HP Labs, Microsoft Research, Yahoo!, Google, Intel, Amazon, UW eScience