Scaling from **Big Data** to *Fast Data*

Emerging Challenges from eScience and eEngineering

Yogesh Simmhan

simmhan@serc.iisc.in

COMAD, 2013, Ahmedabad

How do you react when the *next big thing* is here?

- Bah, humbug!
- Me too, Me too
- Hmmm, lets examine this…

mcswhispers.wordpress.com

bit.ly/Jbwv9O

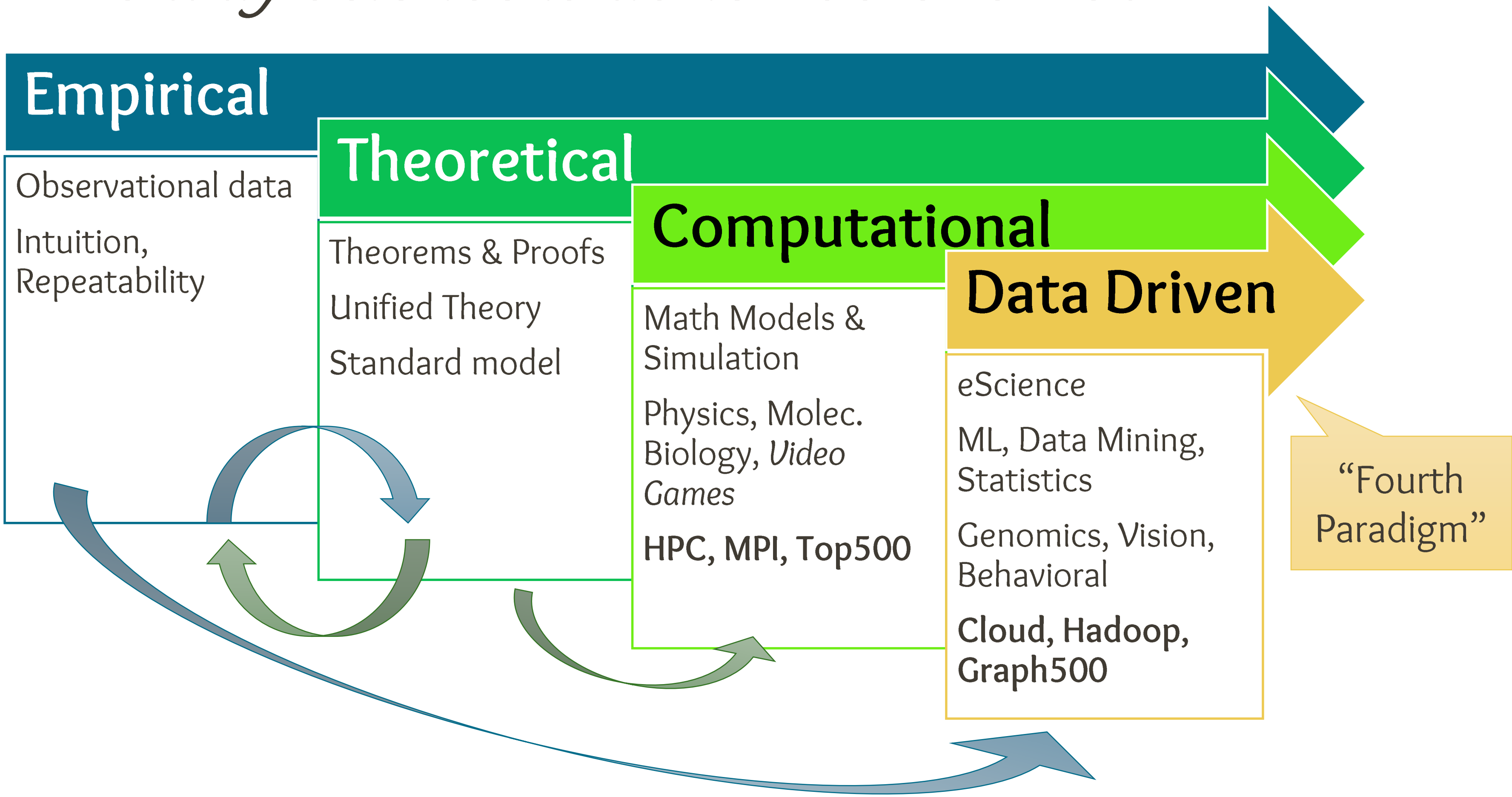
/bit.ly/1etUwra

Bah, humbug!

- There are enough of these around…too many to list

"Big Compute, Big Data, Big Science"†

- The *way science is done* has evolved



† Title from Robert Harrison (Stony Brook/Brookhaven Lab)'s *HiPC 2013 Keynote*

Easing into eScience

IISc
SERC
M.Tech.
Comput'nal
Science

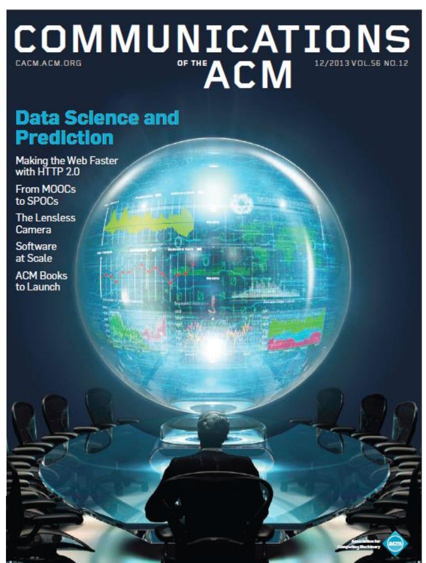
The Data Grid

JNCA, 2000

IU Informatics PhD, 2006

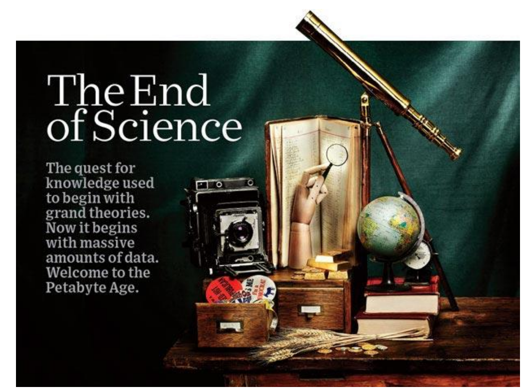
Towards
Data Intensive
Scientific
Discoveries!

4PI.in, 2013



Dec, 2013

4 Sept, 2008



Wired, July 2008

The Obligatory 3D's

- **Volume**
  - Sheer size of data. Storage, mgmt., bandwidth
- *Velocity*
  - Realtime processing, ephemeral, latency
- Variety
  - Complexity, linked data analysis, compute+I/O
- Not exclusive dimensions, but useful
- Helps shape some of the interesting eScience and eEngineering activity

Volume | Pan-STARRS Sky Survey, *2008*

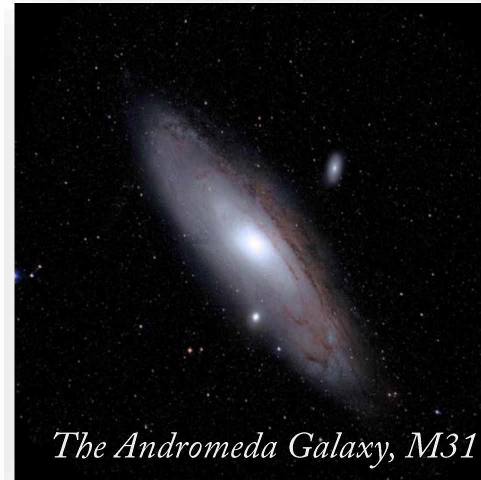
"Me Too"

bit.ly/Jbwv9O

Pan-STARRS Sky Survey

www.ps1sc.org

- *Discover & characterize Earth-approaching objects that might pose a danger to our planet.*
- One of the largest telescopes
  - 1.4 Gigapix camera world's largest!
- Scan $^{2}/_{3}{}^{rds}$ of sky, 3 times/month
  - **1 PB** of images, 30 TB of processed data/year
  - **150 M** detections / night
  - **5.5 Billion** objects, 350 Billion detections



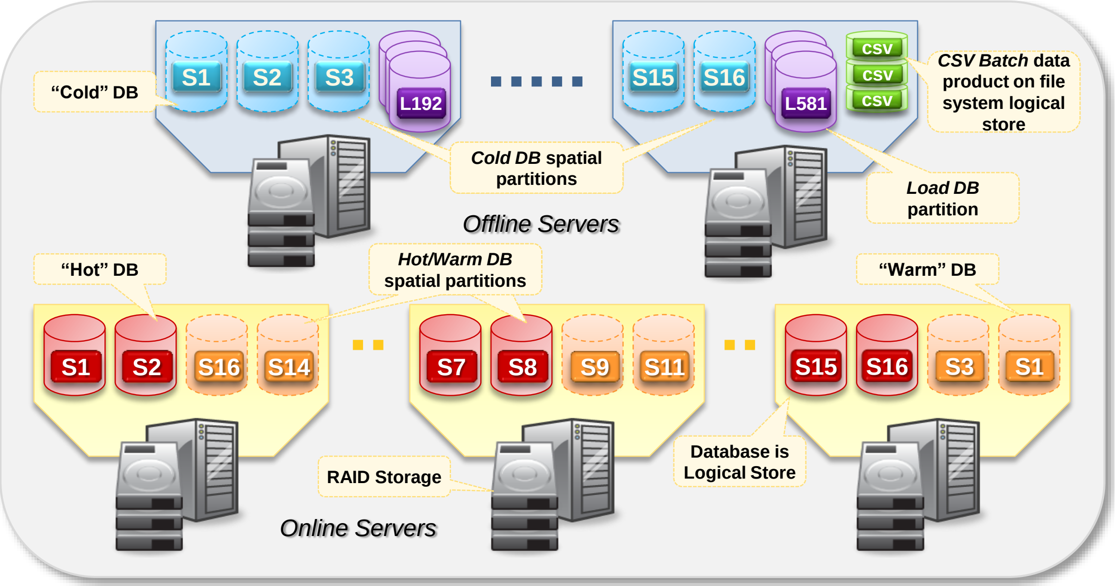
The Andromeda Galaxy, M31

Dome of PS1 telescope at Haleakala

@Microsoft Research with Johns Hopkins, UHawaii,

HW & DB Architecture

- HW/SW/DB layout co-design
- **GrayWulf** commodity cluster for scale out †
  - Amdahl's ratios: I/O BW= 0.5, Memory=1.04
- Distributed **MSSQL** Databases
  - CASJobs auto, query generation
  - "MyDB" local scratch DB of results

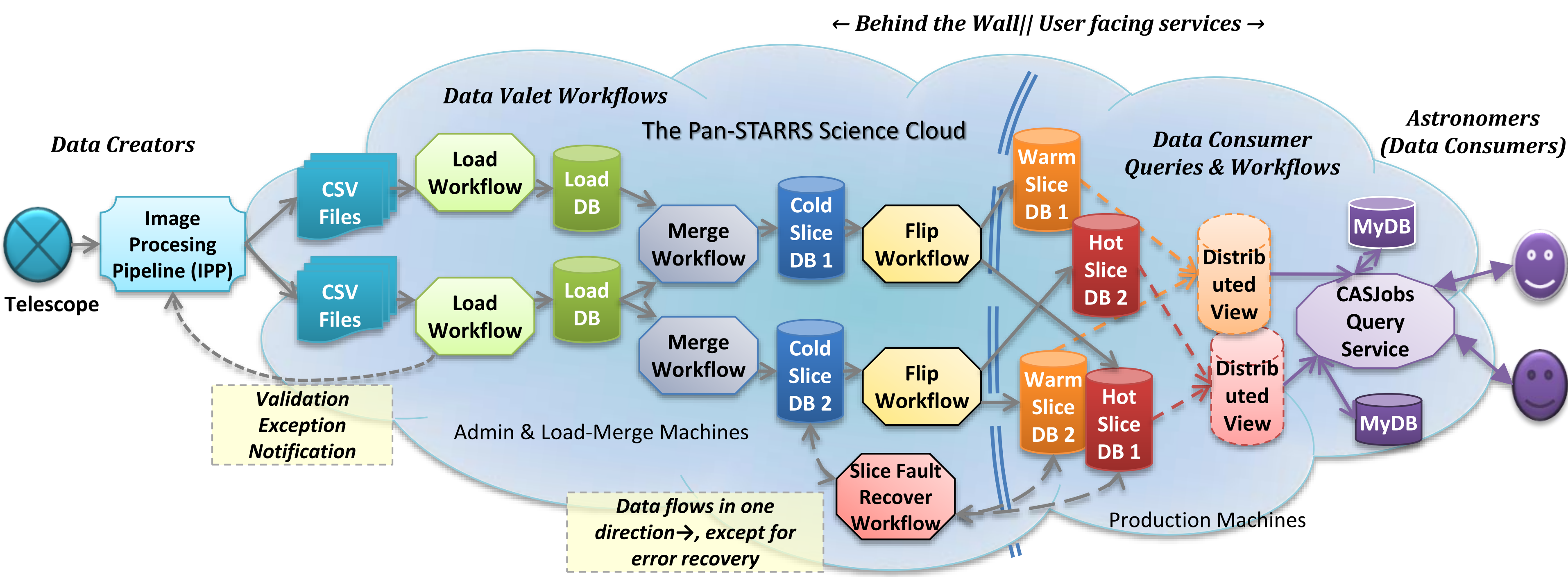


†SC 2008 Storage Challenge Award

Stargazing through a digital veil, Simmhan, van Ingen, Heasley, Szalay, *HPCDB*, 2011

Scientific Data Ingest Pipeline

- Reduce time to <u>science ready</u> data
  - Once every 6 months → once/week, 10x data
- Ensure performance: *Relax ACID* on distributed DB
- Ensure *resilience* & externalize *consistency*



Transactional ETL Workflows

- Well defined, Well tested workflows
  - Run repeatedly, impact cumulative
- Granular, Reusable workflows
  - Separate policy from mechanism
- Workflows as Data State Machines
  - *Data containers* have states
  - *Workflows* & *tasks* cause state transitions
- Leverage provenance as transaction log

Building Reliable Data Pipelines for Managing Community Data using Scientific Workflows, Simmhan, van Ingen, Barga, Szalay, Heasley, ***eScience***, 2008

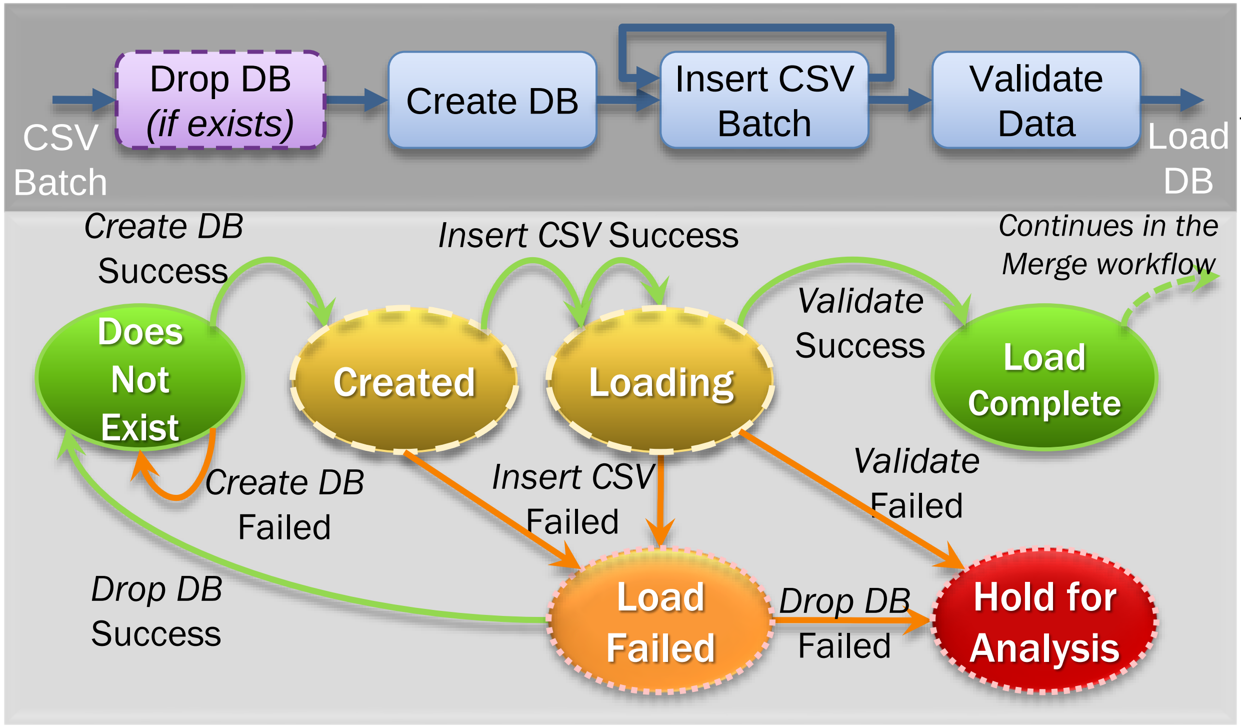
WF Recovery Baked into Design

- Faults are a fact of life in distributed sys.
  - Handling faults a *routine* action
  - Mitigate I/O cost, ease manageability

Load Workflow Activities

State Machine for Load DB

State transitions take place when activities execute.



Activity Type

Blue/Solid : Mainline Code

Purple/Dashed: Recovery Preamble

Success States

Green/Solid : Clean

Yellow/Dashed: In Flight

Failed States

Orange/Dotted: Recoverable

Red/Dotted: Unrecoverable

Using Provenance for Resilience

1. Re-Execute Idempotent Recovery
   - Rerun without side-effects
2. Resume Idempotent Recovery
   - Allow a “goto” at the start
3. Recover & Resume
   - Tasks to rollback to initial state. Reduce to #3
4. Independent Recovery
   - Complex faults, global sync, manual oversight

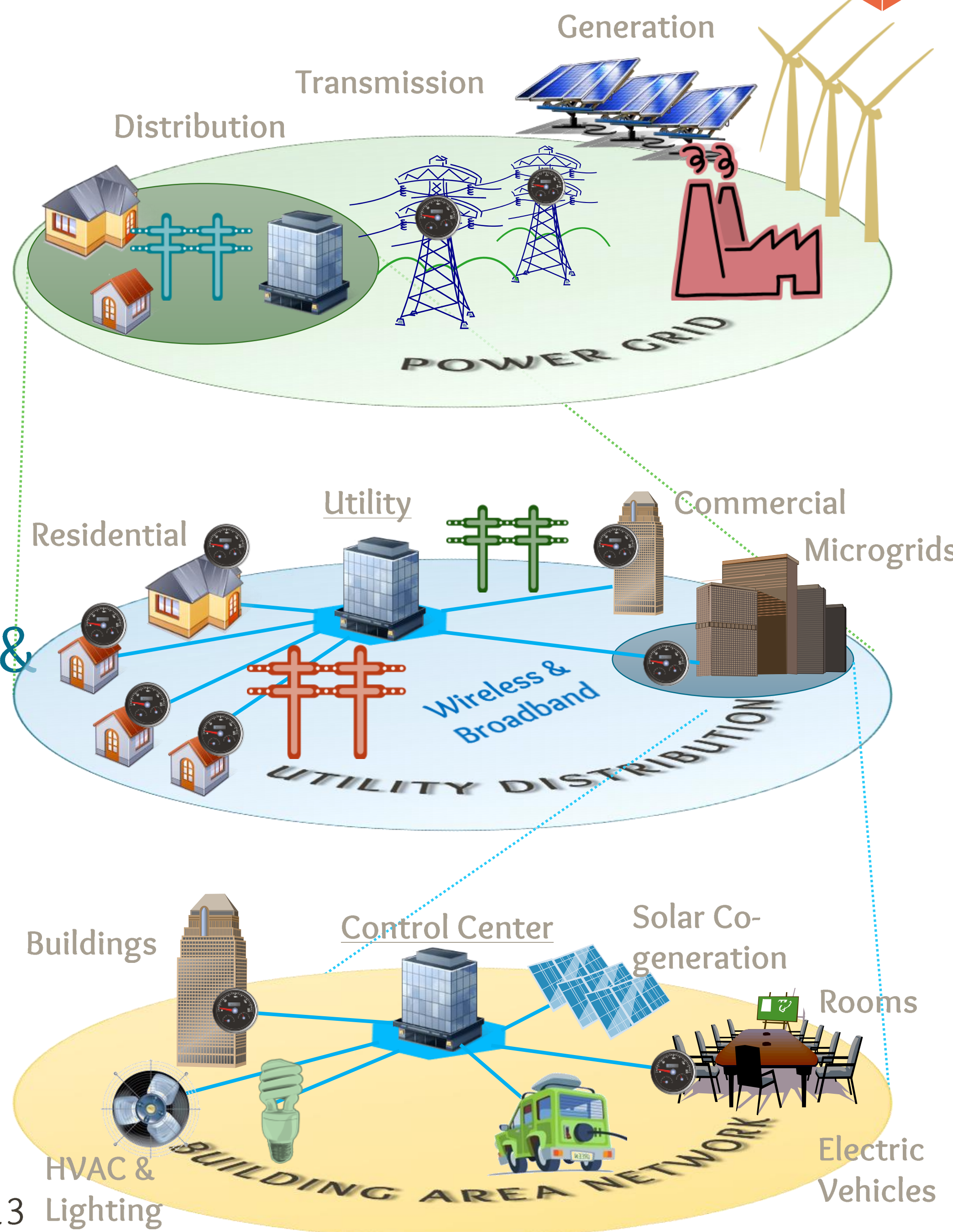
Velocity | The Los Angeles Smart Grid, *2011*

"Hmmm, lets examine this…"

/bit.ly/1etUwra

Smart Grids: *The* Cyber Physical Sys.

- Integration of Renewables
- Advanced Instrumentation
- Bi-directional communication
- Real-time data acquisition & control
- Self-contained 'Micro Grids'...*like USC*
- *LADWP: largest US public utility*

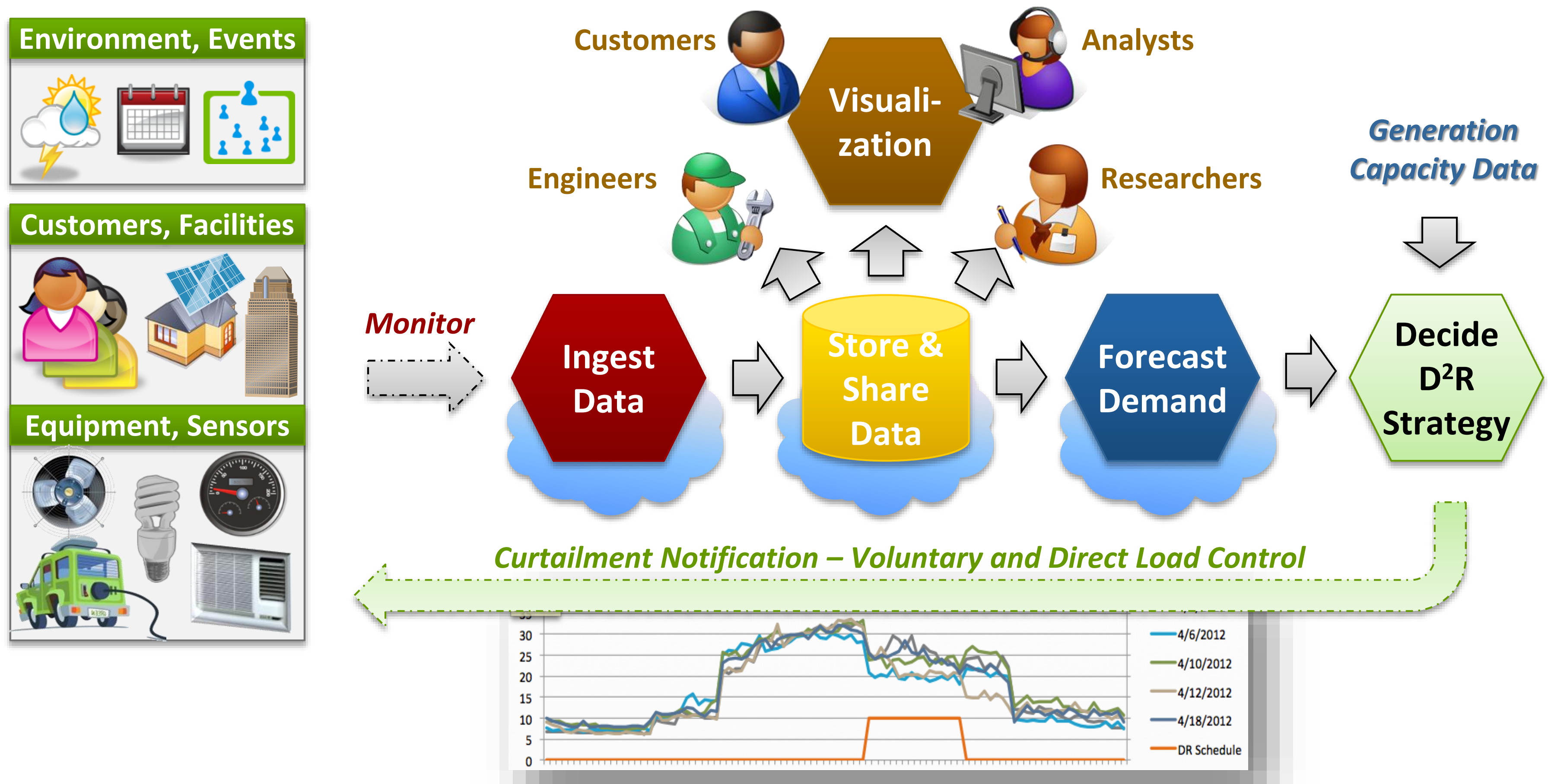


Cloud-based software platform for data-driven smart grid management, Simmhan, et al, *CiSE*, 2013

Dynamic Demand Response (D^2R)

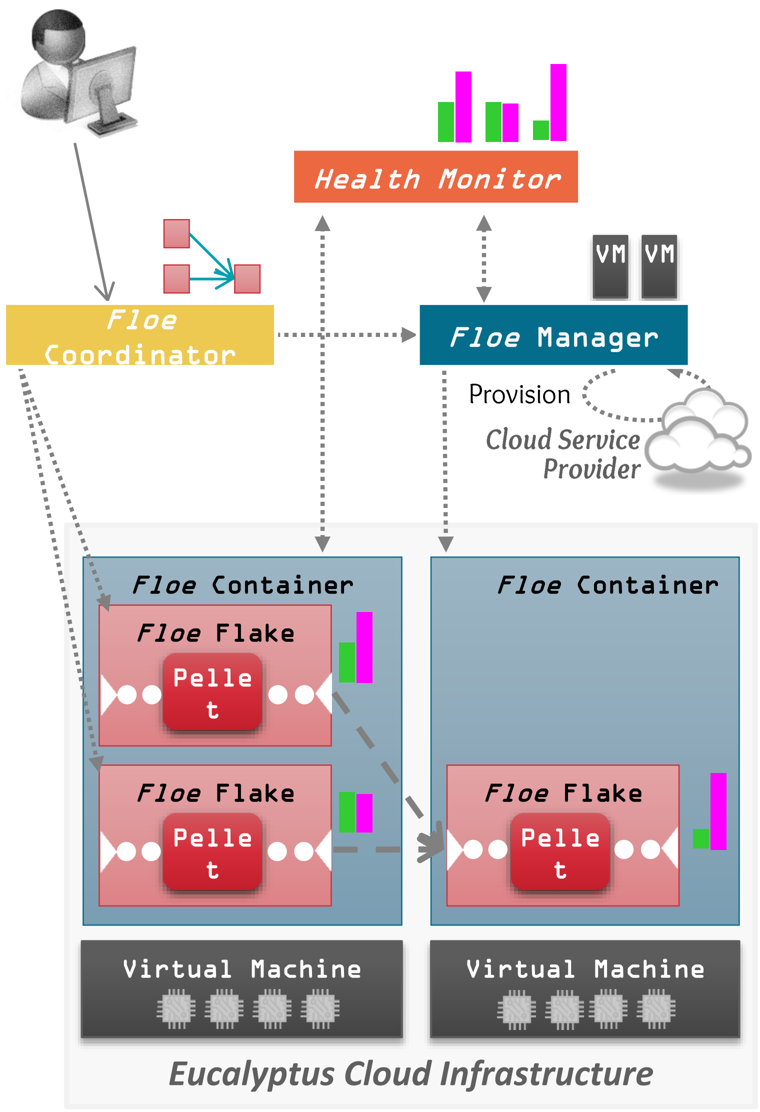
Reduce consumer demand for electricity during periods of peak usage to relieve stress on power grid

When → By How Much → How/Whom … ***Predict, Adapt, Evolve***



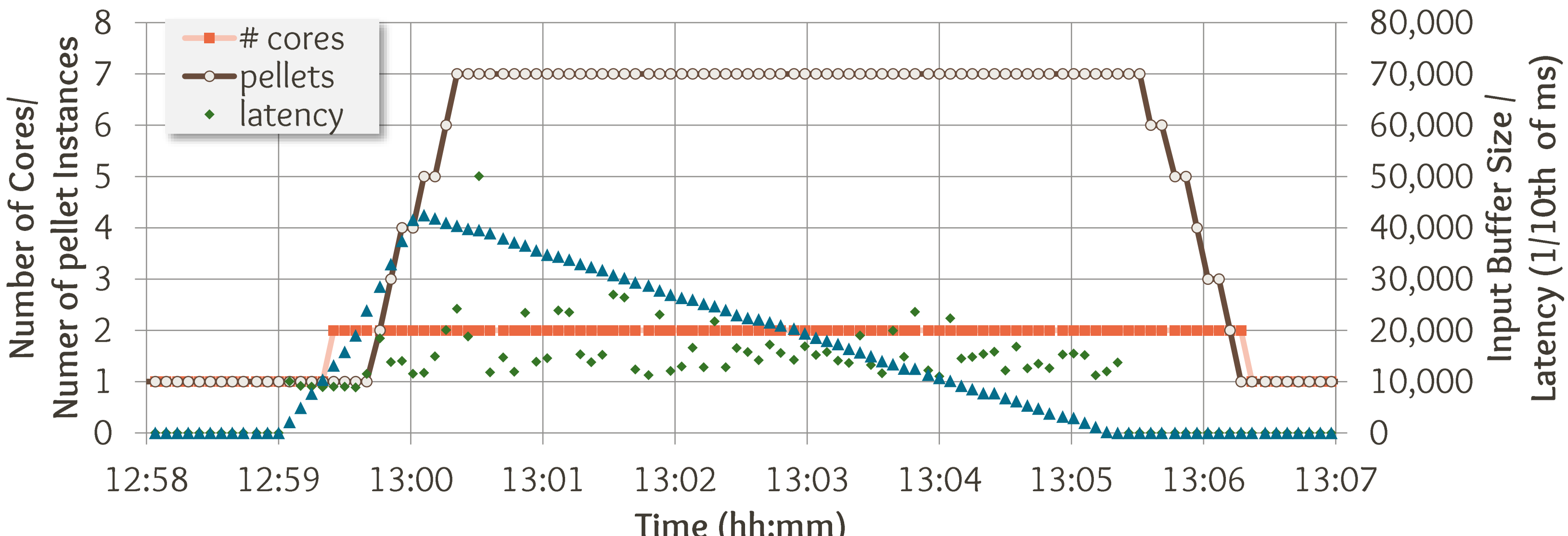
Information Integration of Big Data

- Real-time data streams
  - ~50,000 Sensors
  - 1/15min intervals
- Semantic Information Ingest Pipeline
  - Normalize Heterogeneous Data
  - Ease data access in a complex environment
- Scale to thousands of customers
  - ***Floe: Continuous Dataflow Engine for Elastic Execution on Clouds***



Elastic Scaling Up & Out on VMs

- Ensure latency target is met
  - Add/remove # of cores allocated per VM
  - Add/remove VMs allocated per dataflow
- Initial placement on independent VMs
- Decentralized VM-local scaling algorithm



IEEE SCALE Challenge. **First Place.** Adaptive Energy Forecasting and Information Diffusion for Smart Power Grids, Simmhan, et al. (2012)

Runtime Adaptation QoS Trade-off

- Allow alternate tasks with differential QoS
  - E.g. high rez model w/ high cost & utility *vs.* low rez model
  - Logically independent, no app. side effects
  - Meet throughput goal, Maximize value
- Heuristic runtime adaptation algorithm
  - Thru'put skew of ε triggers adaptation
  - Estimate local+downstream impact
  - Incremental +/- 1 core/VM per timestep

Exploiting Cloud Elasticity to Enhance the Value of Dynamic, Continuous Dataflows, Kumbhare, Simmhan and Prasanna, *SC* 2013

Semantic CEP for D2R

- **Complex Event Processing (CEP)**
  - Detect event patterns from data streams
- **Semantic CEP:** Use domain semantics for higher abstraction in pattern specification
  - E.g. Find *offices* with *airflow* greater than 200
  - Predict energy spikes, energy leaks
- Go forward and back in time

```
SELECT  ?event
FROM    OPCStream
WHERE {?event  evt:hasEventSource ?src .
?src  ee:hasLocation  ?loc .
?loc  rdf:type  bd:Office .
?src  rdf:type  ee:AirflowSensor .
?event.value > 200 }
```

DEBS 2014 Challenge

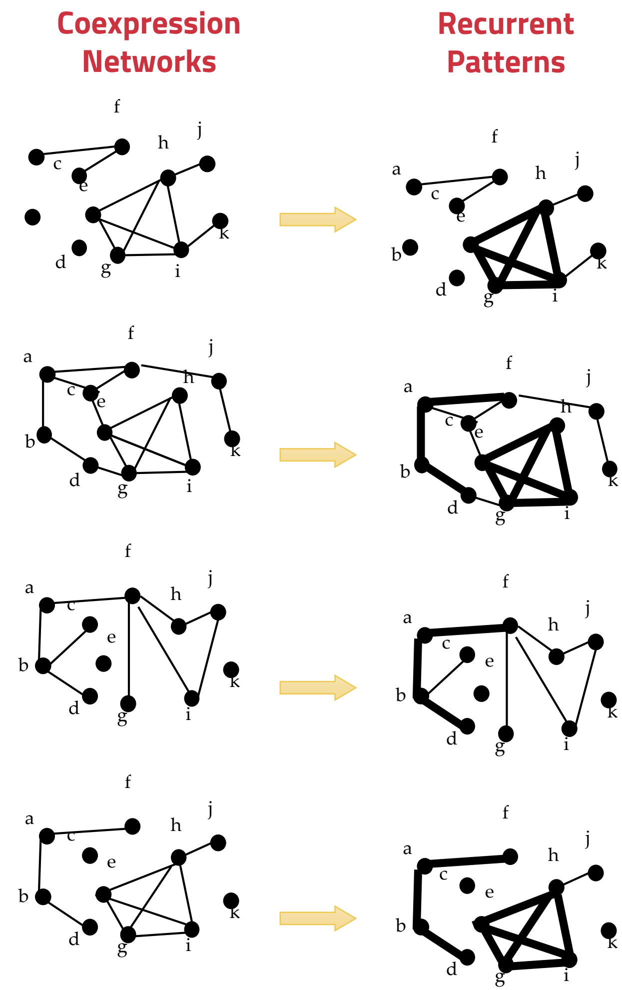
Variety | Computational Biology, etc., *2013*

"Hmmm, lets examine this…

/bit.ly/1etUwra

Graph Collections in Systems Bio.

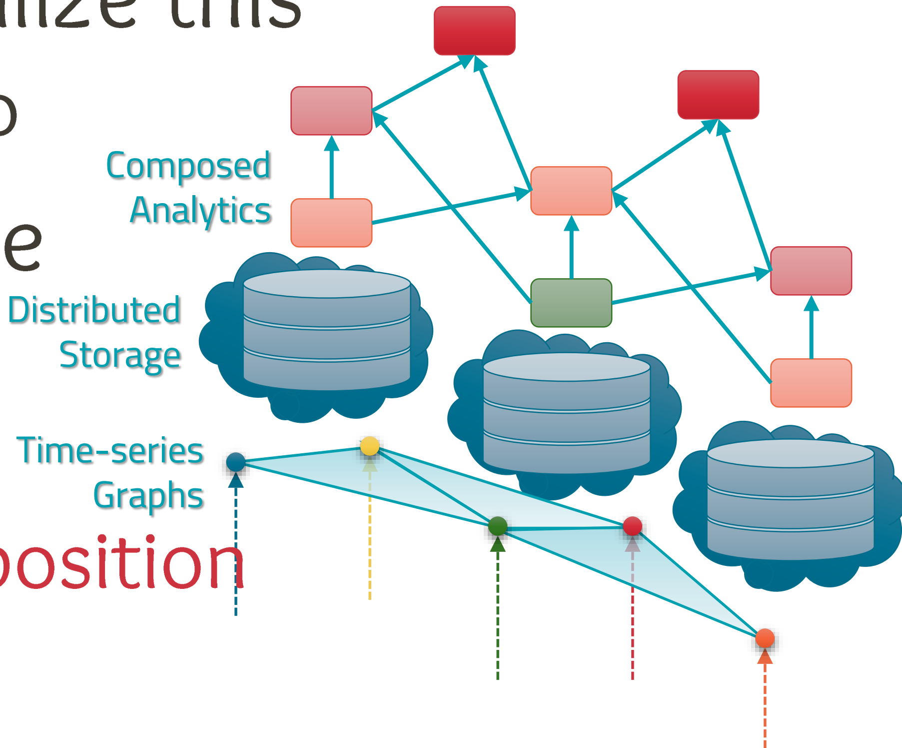
- Co-expression networks
  - Recurrent correlation behavior between gene
  - Over time (lifespan), Across space (cancer patients)
- Modelled as a graph series
  - Same topo, different values
- Find frequent clusters



A graph-based approach for the integrative analysis of gene expression data, Jasmine Zhou, USC, 2013

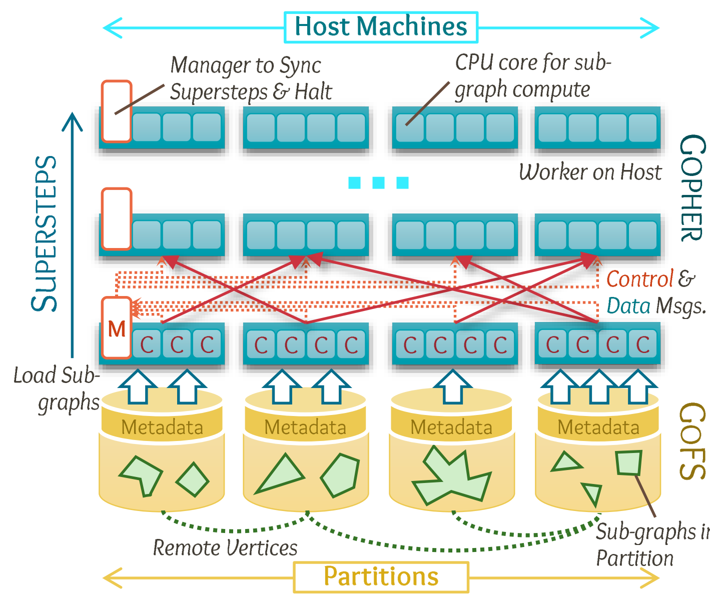
Dynamic & Timeseries Graphs

- Graph (time)series are common in CPS
  - Static Road N/W, Variable traffic flows in time
  - Power grid N/W, Power loads on vertices
- Dynamic graphs generalize this
  - Topology can change too
- Abstractions for scalable analytics on TS graphs
  - Efficient storage model
  - Intuitive & efficient composition



GoFFish Software Platform

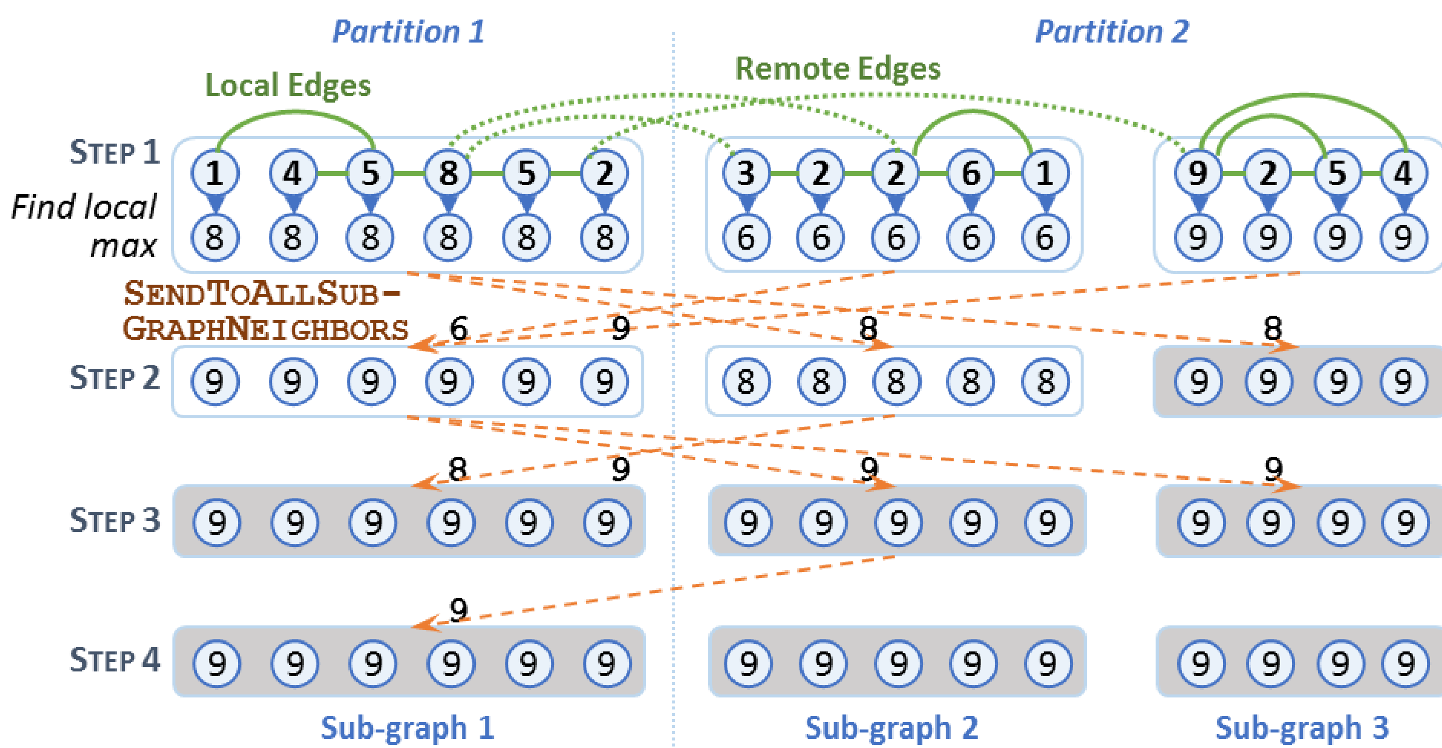
- GoFS: Distributed Graph-oriented File Sys.
- Gopher: Compose *sub-graph* centric analytics
- Targeted at distributed commodity H/W
- Sub-graph & TS aware distributed storage
  - APIs for SG *Iteration, Filtering* and *Projection*
  - Temporal Instance *Packing*
  - *SG Binning & Caching*



Scalable Analytics over Distributed Time-series Graphs using GoFFish, Simmhan, et al, (under review)

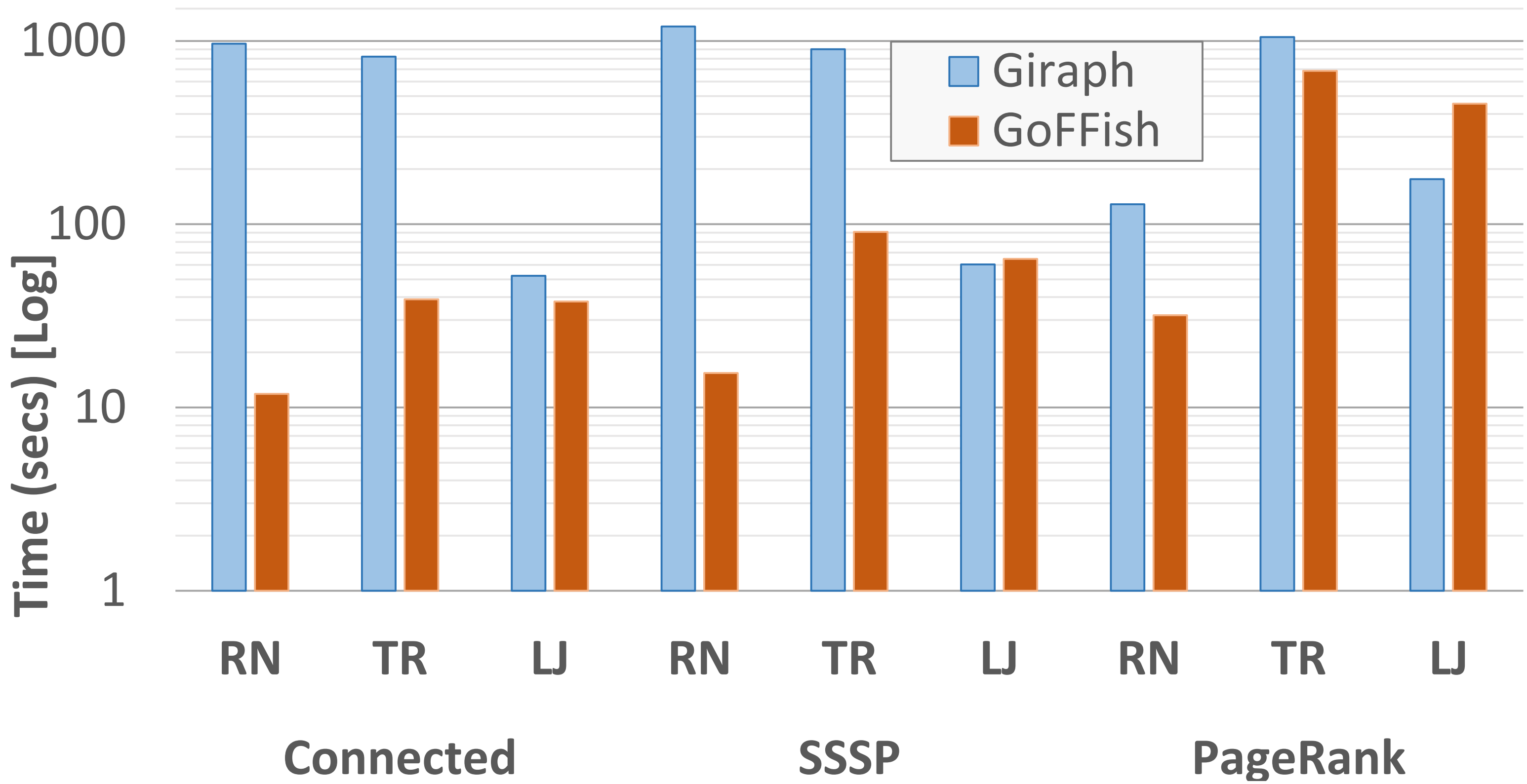
Sub-graph centric programming

- Logic defined for sub-graphs (> *Google Pregel*)
- Bulk Synchronous Parallel exec of supersteps
- Message passing between SG's in superstep



Results vs. Apache Giraph

CA Road (*2M/2.7M*), Traceroute (*19M/23M*), Live Journal (*5M/68M*)



GoFFish: A Sub-Graph Centric Framework for Large-Scale Graph Analytics, Simmhan, et al, *ArXiv* 2013

To Conclude

- eScience has been focussing on "Big Data" for a while
  - There is some credence to the hype
- Novel applications are coming up
  - Scientific apps are a vanguard
- Platforms for analytics on dynamic & interconnected data are vital
  - Internet of Things, *anyone?*
- ***We need you @ SERC, IISc!***
  - Application deadline for MSc/PhD is **Mar, 2014**

DREAM:Lab

DISTRIBUTED RESEARCH ON EMERGING APPLICATIONS & MACHINES

dream-lab.in

Indian Institute of Science

Thank You!

Questions?

Acknowledgements

Catharine van Ingen, Roger Barga, Alex Szalay, Jim Heasley, Viktor Prasanna, Alok Kumbhare, Charith Wickramaarachchi, Soonil Nagarkar, Santosh Ravi, Raghu Raghavendra, Shel Swenson & Jasmine Zhou