DISTRIBUTED RESEARCH ON EMERGING APPLICATIONS & MACHINES
dream-lab.in | **Indian Institute of Science**, Bangalore

SE252:Lecture 8+9, Jan 30/Feb 3

ILO1:Parallel and Distributed Systems Context

Yogesh Simmhan

ILO 1

- Parallel and Distributed Systems Context
  - *Classify* and *describe* the architecture and taxonomy of
    - » parallel and distributed computing,
    - » Shared and distributed memory, and
    - » data and task parallel computing.
  - *Explain* and *contrast* the role of Cloud computing within this space.

DREAM:Lab

Lecture 8+9

Computer Organization

- Diverse forms of computing models
- Complex architectures, systems, interactions

- How can we take a step back and examine their properties?
  - *Why do we need to classify?*

Flynn's Taxonomy

- Categorization of processor architectures
  - "Macroscopic" view: Process + Data
  - Generalizable to models of computing
- Instruction Stream
  - Discrete stream of operations to perform
- Data Stream
  - Discrete datum on which to perform ops

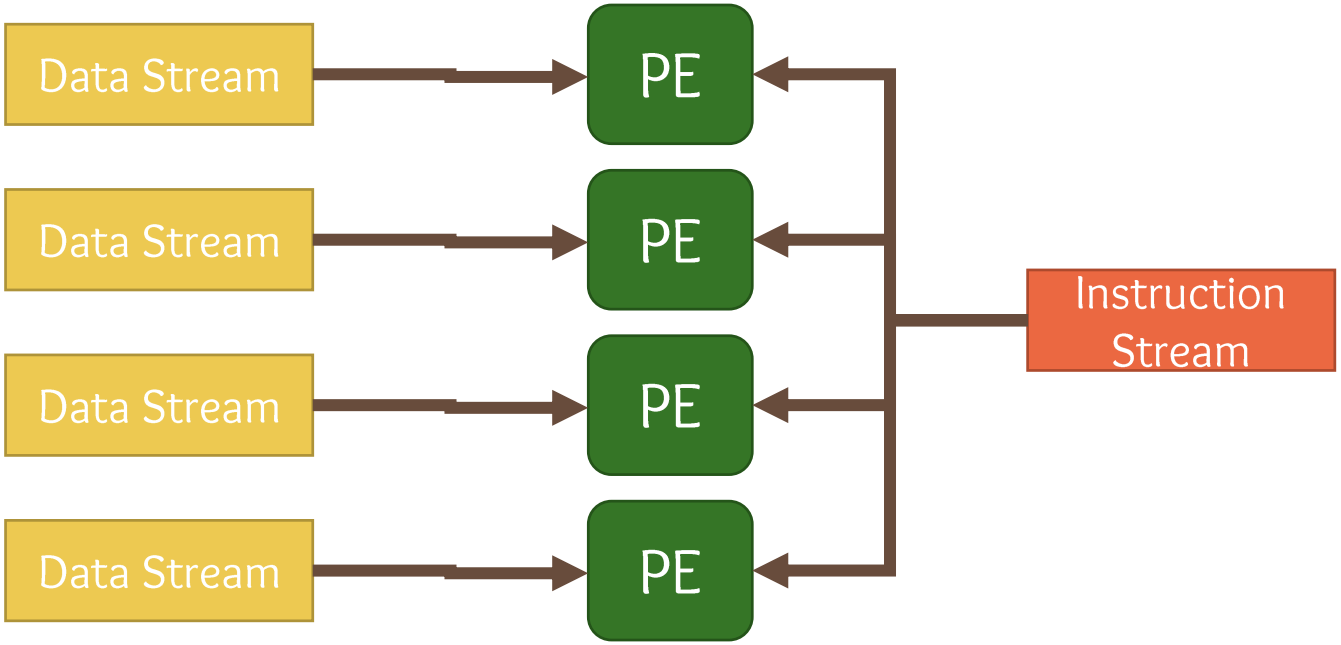
Some Computer Organizations and Their Effectiveness, MICHAEL J. FLYNN, *IEEE TRANSACTIONS ON COMPUTERS*, 1972

Single Instruction, Single Data

- Sequential, single processor machines
- No parallelism

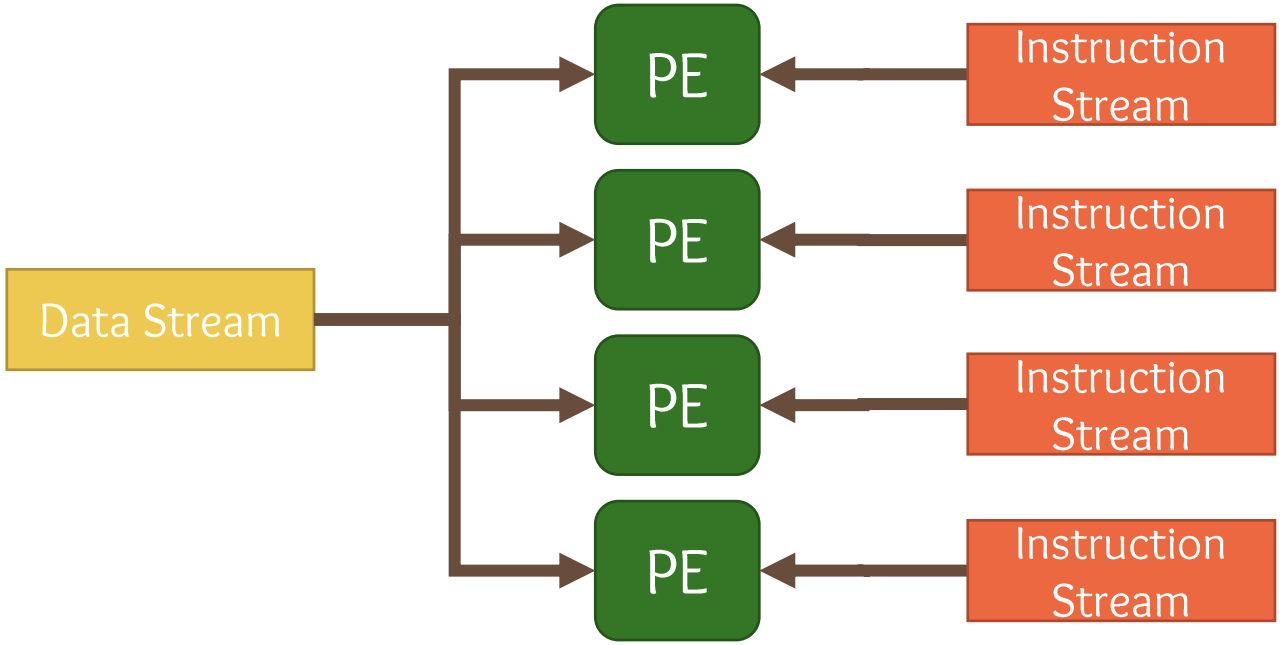
Single Instruction, Multiple Data

- Multiple processors, execute same instruction
- Different data streams
  - Data parallelism, large datasets
- *Think GPGPUs*



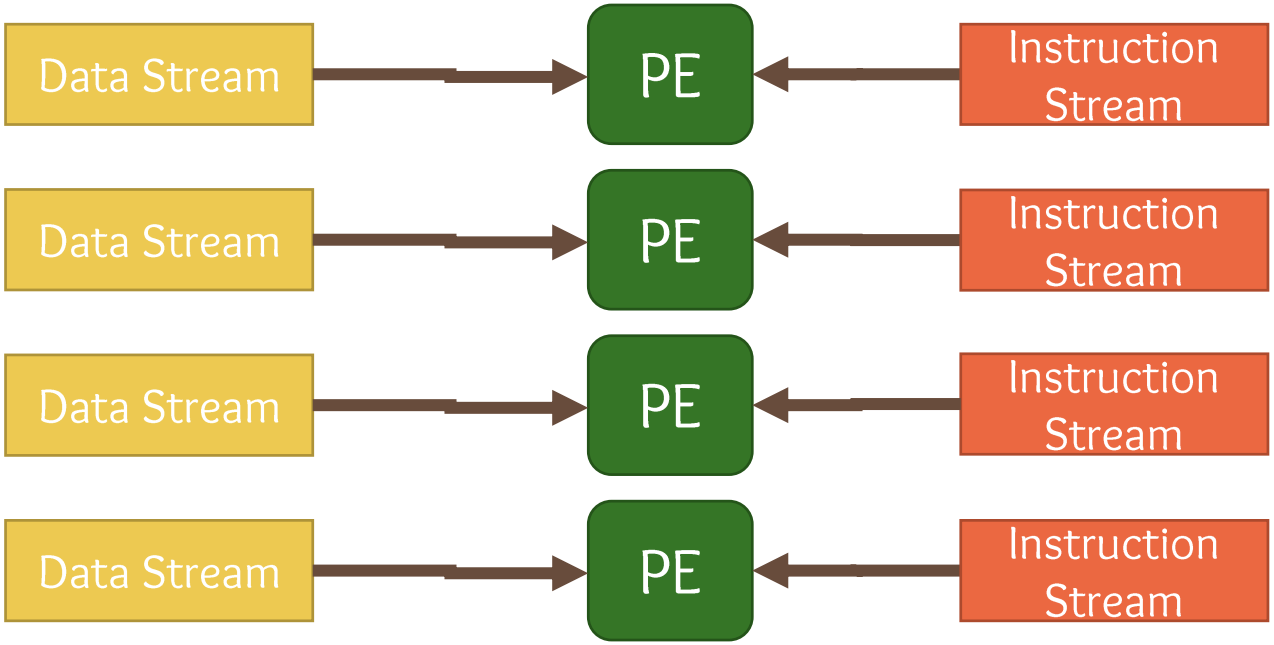
Multiple Instruction, Single Data

- Multiple processors operate on same data
  - Execute different instructions
- Rare. E.g. for redundancy



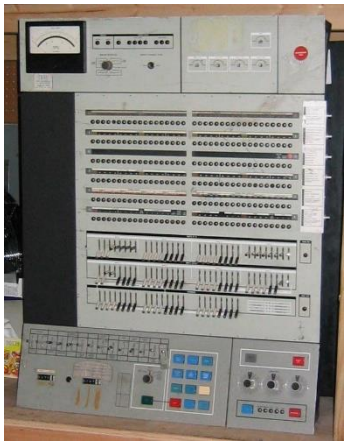
Multiple Instruction, Multiple Data

- Multiple processors
  - Operate on different data
  - Execute different instructions
- Multi-processor

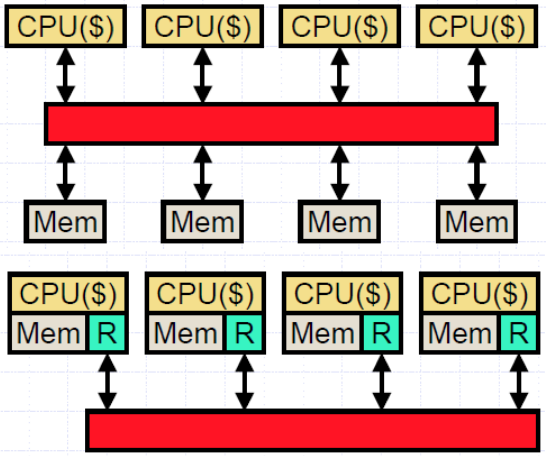


Shared vs Distributed Memory MIMD

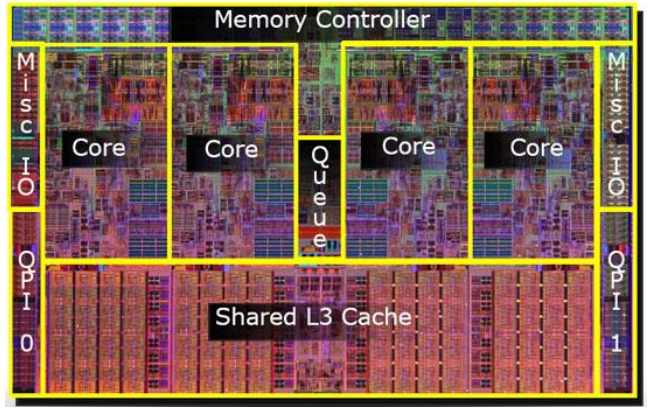
- Shared memory multi processors (SMP)
  - Collection of processors on same machine
  - Share memory through common bus
  - Thread level parallelism
  - Symmetric Multi-Processing
  - Non-uniform Memory Access (NUMA)



IBM System 360/Model 65, 2 processors
http://www.cs.kent.edu/~rothstei/10051/history/IBM-360-65.htm



© CIS 501 (Martin/Roth)



Intel Core i7

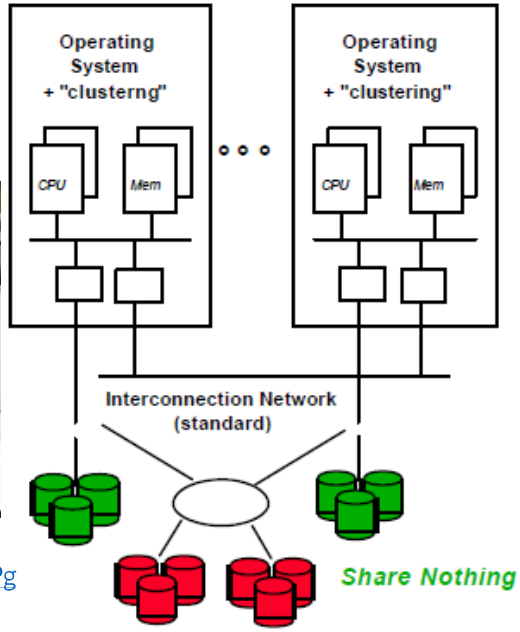
© www.legitreviews.com

Shared vs Distributed Memory MIMD

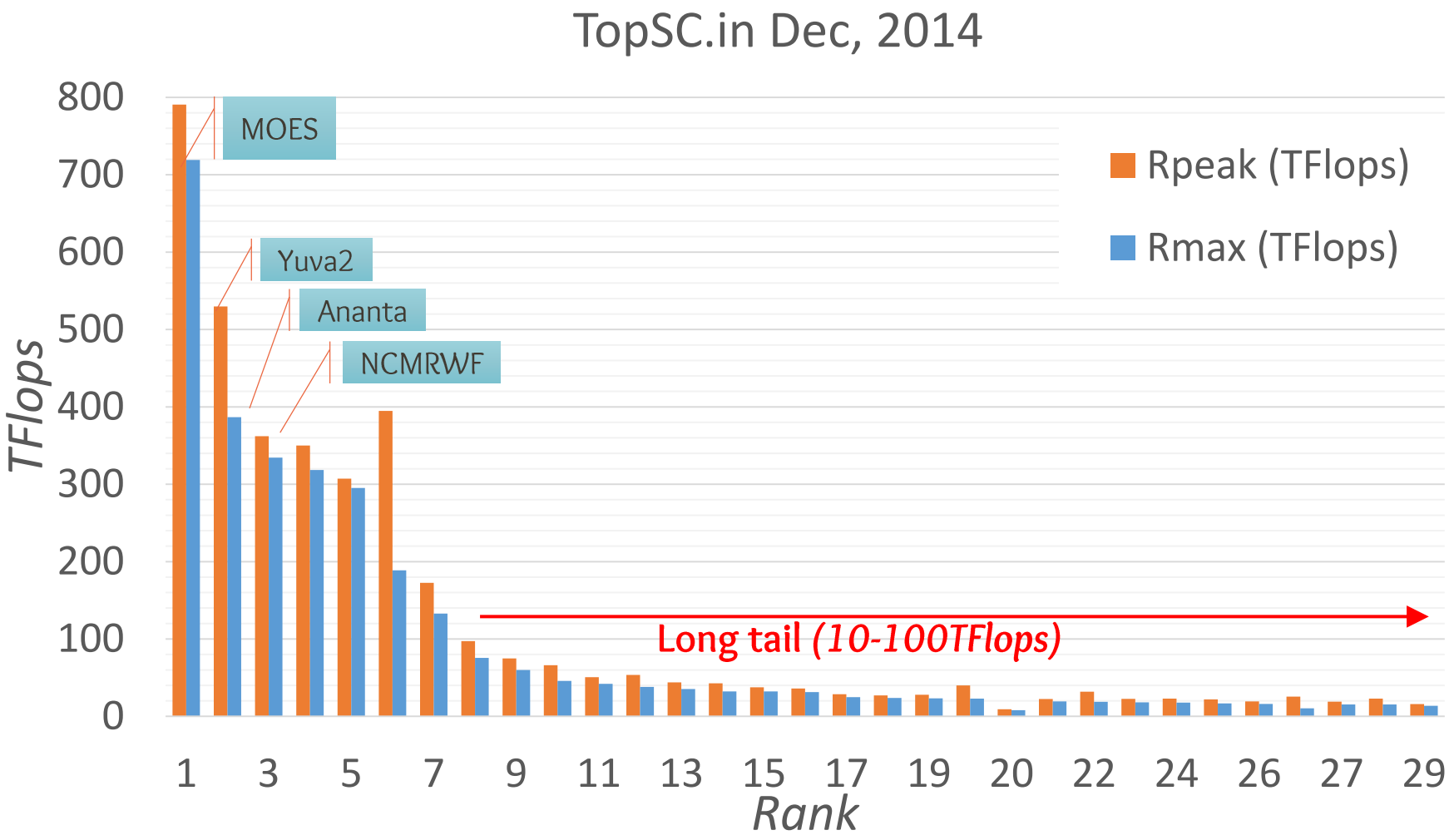
- Distributed memory multi processing
  - Processors on different machines
  - Shared resources through network
  - Process level parallelism
  - Massively Parallel Processing (MPP)
  - Cluster Computing

IBM iDataPlex/
MOES (Cluster)

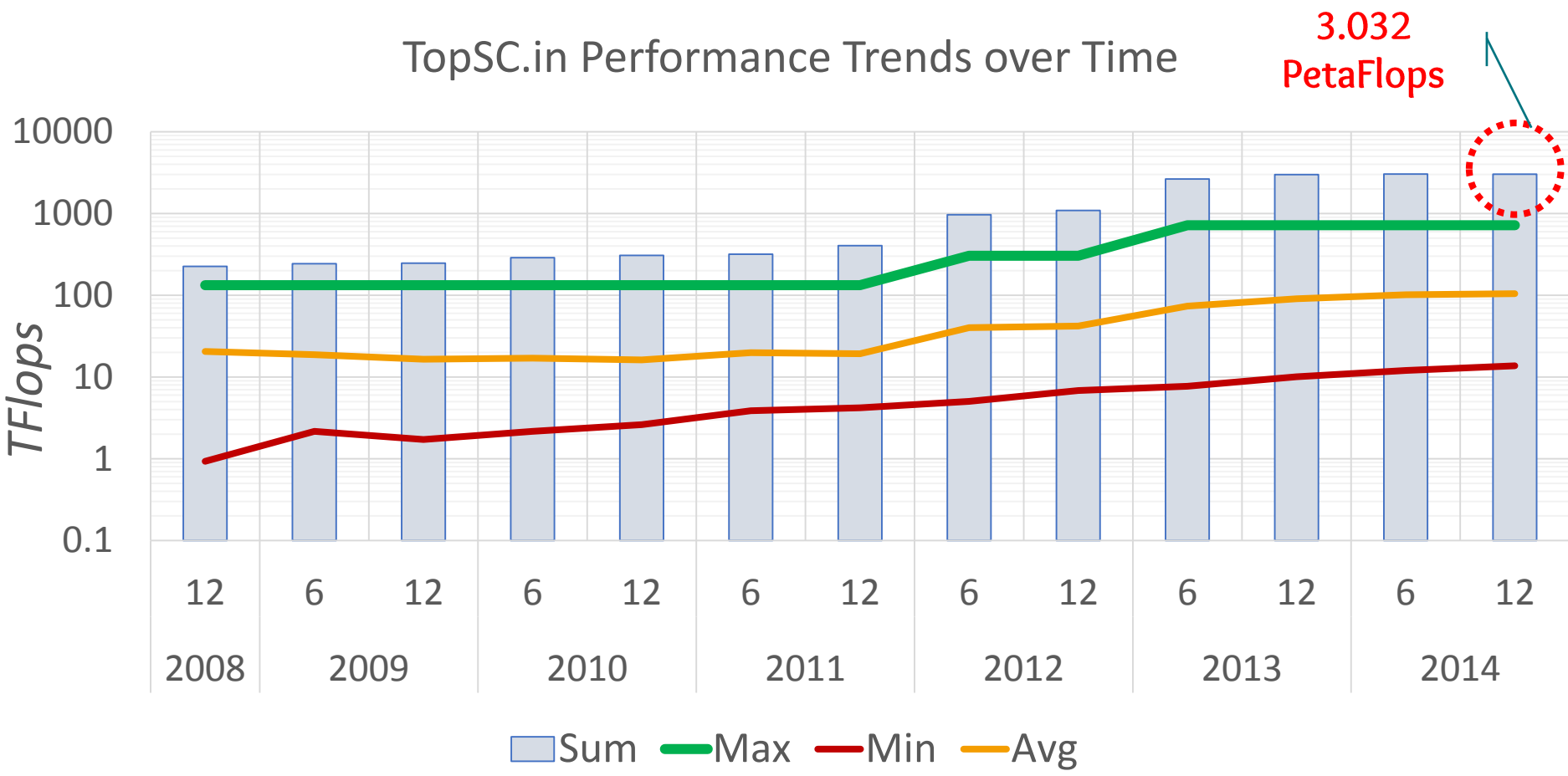
SC13 Student Cluster Competition, Bentley/NEU
http://www.youtube.com/watch?v=9wpjx8qMtPg



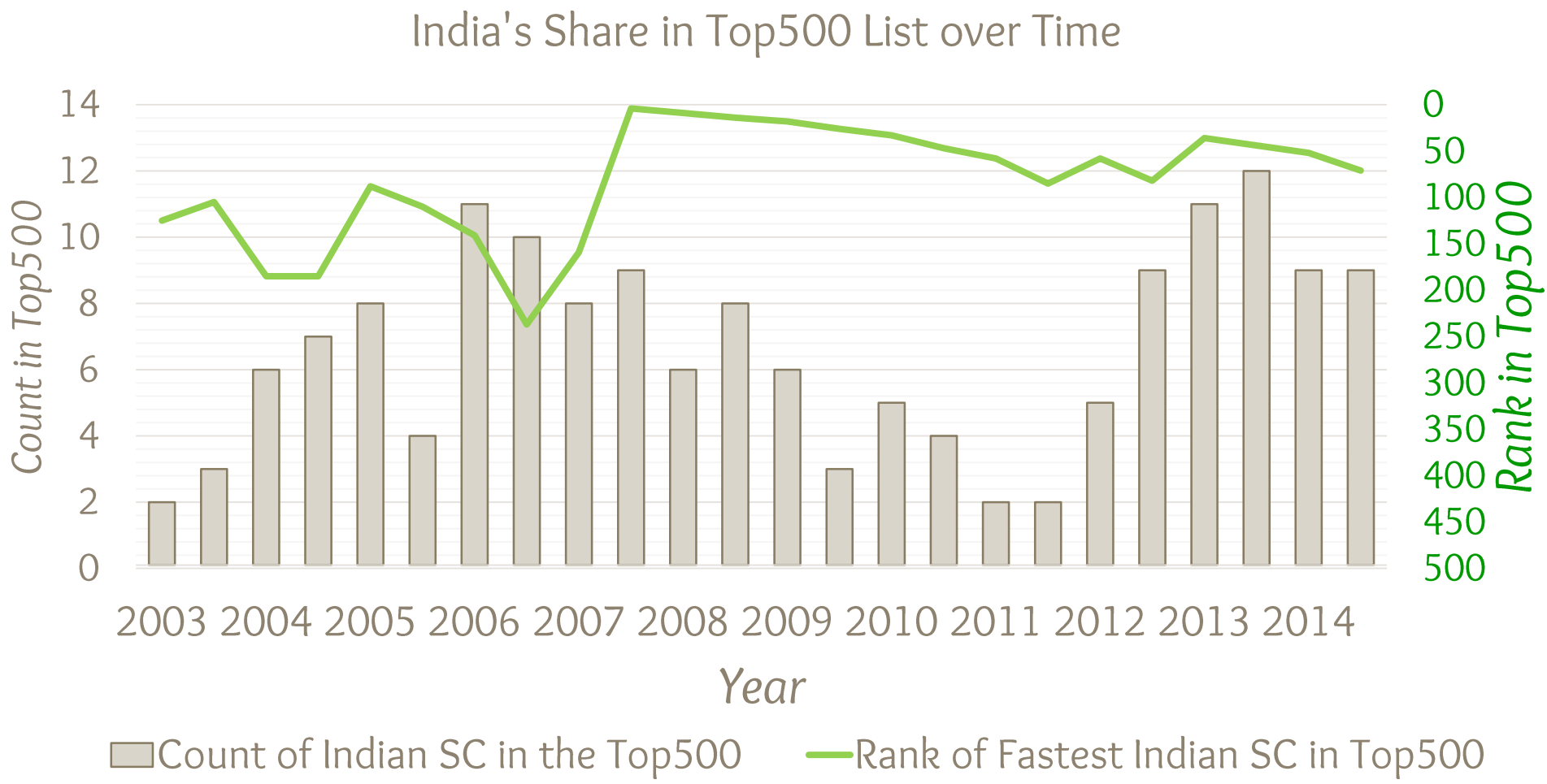
TFlop Distribution for Dec/2014



Cumulative Perf over Time



Performance in Top500



More Machines in the Pipeline

Computing Paradigms

- Centralized computing
  - Resources physically centralized at a location
  - Tightly coupled environment
  - May allow distributed access
- Parallel computing
  - SMP, MPP and clusters with high speed N/W
  - MPI, OpenMP. High Performance.
- Distributed computing
  - Distributed memory, N/W communication
- Cloud computing
  - Distributed computing in centralized/ distributed location(s)

Concurrent Computing

HPC, HTC, OLTP, Big Data

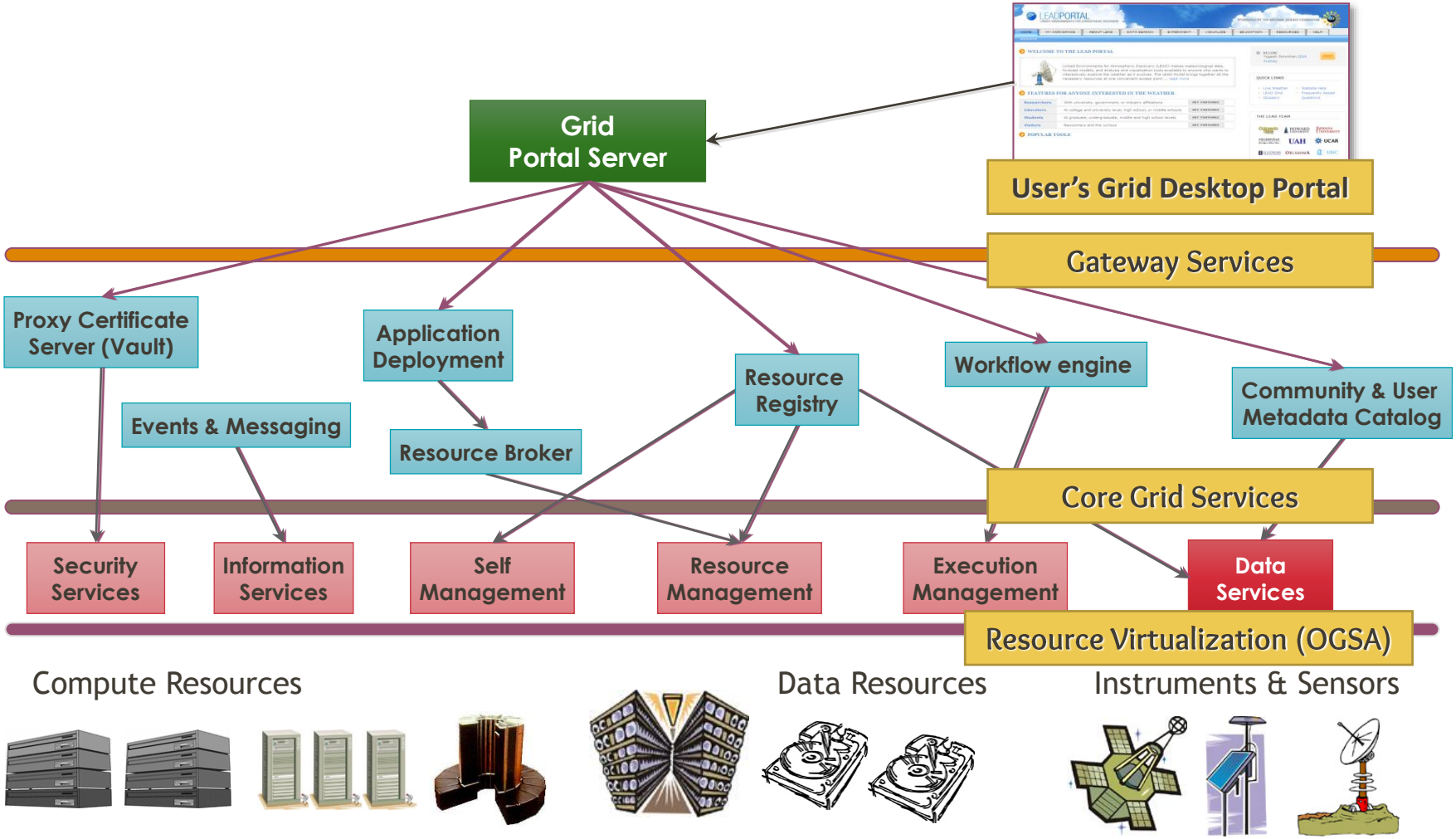
- High Performance Computing
  - Large # of processors for a few hours. Batch ops.
  - Weather simulation models
- High Throughput Computing*
  - Large # of processors for a few months
  - High energy physics. Parameter sweeps.
- Online Transaction Processing
  - Web based biz transactions. Loosely coupled, millions per second.
- Big Data Analytics
  - Large scale data processing. I/O bound. Batch ops(?)
  - Web log processing, Data & pattern mining

*http://research.cs.wisc.edu/htcondor/HPCwire.1

Distributed Computing Sys Models

- Cluster Computing on LAN
  - Beowulf Commodity/Ethernet), HPC/Infiniband
  - ***Middleware*** for coordination across machines
- Grid Computing
  - WAN ***Virtual Orgs (VOs)*** across diverse clusters
  - Seamless computing as a utility across orgs
  - Shared middleware for critical services: security, job submit. *E.g. Globus Toolkit*
  - "Virtualization" of physical layer

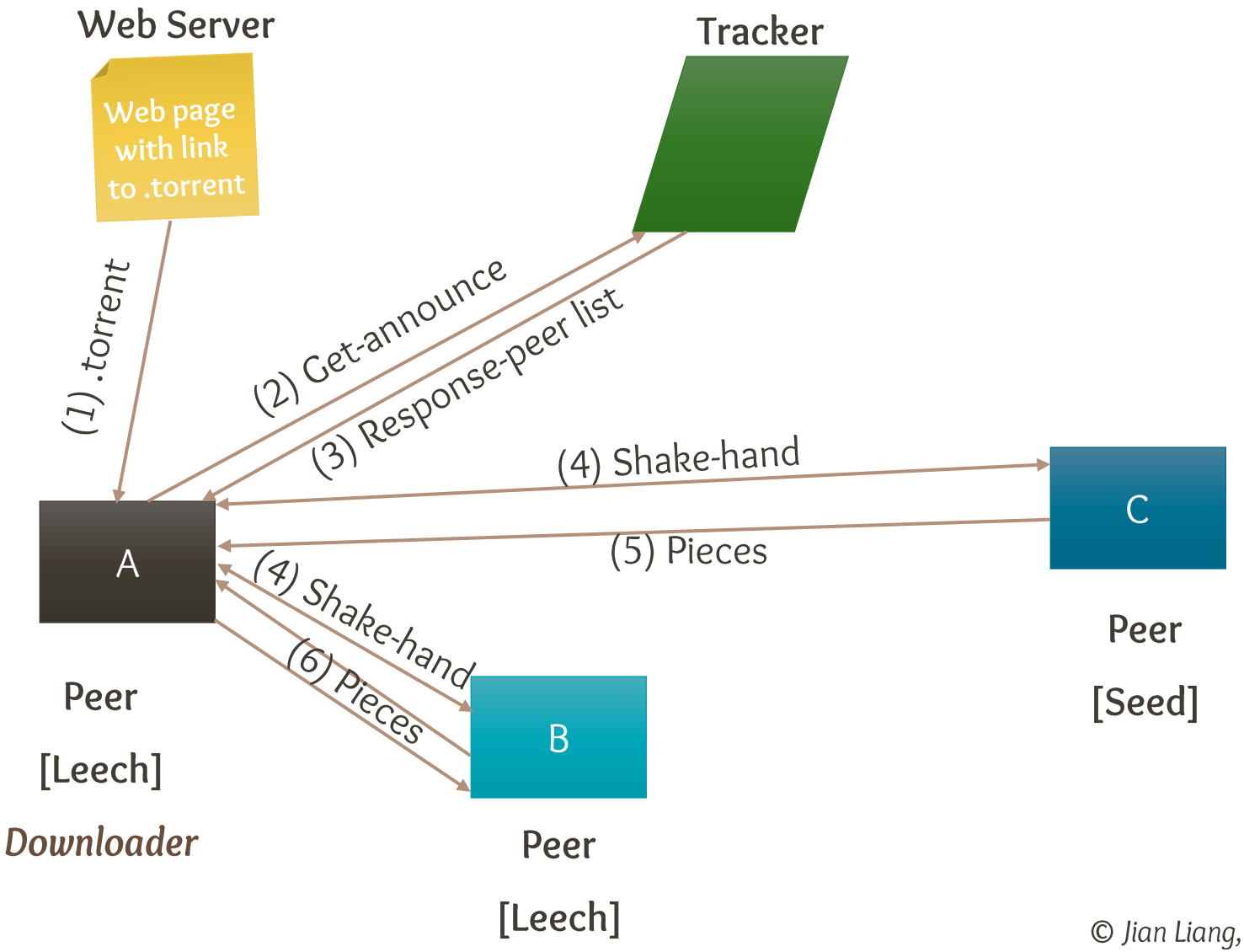
Grid Computing for Science



Distributed Computing Sys Models

- Peer to Peer Computing
  - Client-oriented computing. Overlay N/W.
  - Autonomous, self-organizing, distributed control, heterogeneity.
  - *E.g. Bit Torrent*
- Cloud Computing
  - Client-service SOA model. Cloud fabric.
  - Utility computing: on-demand, elastic.

Torrent P2P Protocol



Internet of Things (IoT)

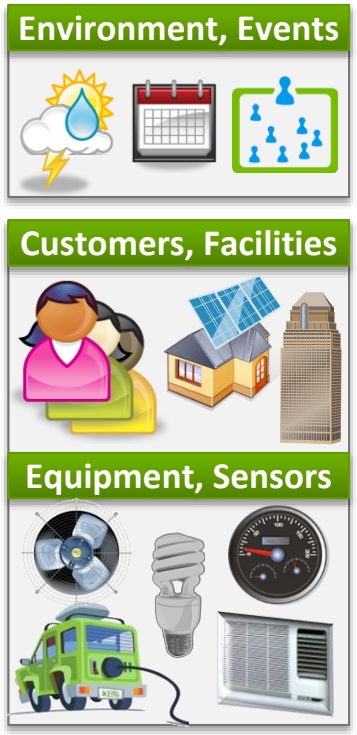
- Network connected devices
  - Sensors, smart phones, instruments
- Both Sensing & Control of systems
  - Observe, Orient, Decide, Act (OODA)
- "Smart" infrastructure (Cyber-Physical Systems)
  - Smart Power Grid/Transport/Water
  - Can we make the system efficient? Can they serve society better?
- Human-centric: Internet of Things & Humans
- Wide-area distributed computing!

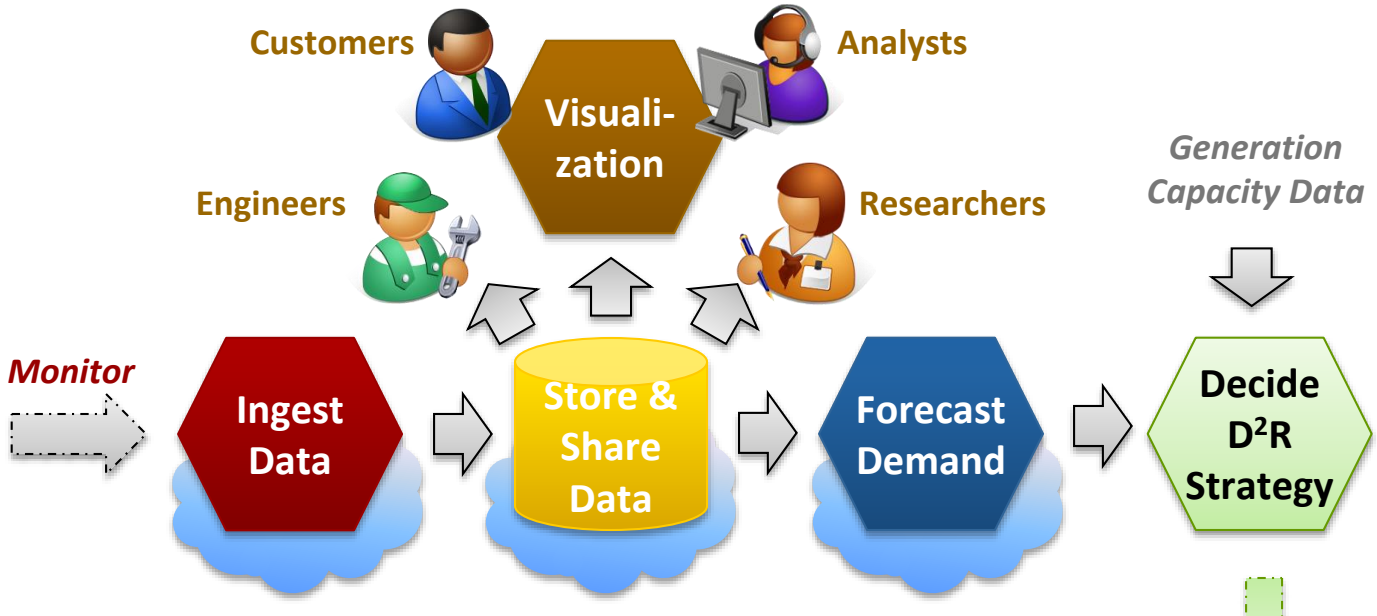
http://radar.oreilly.com/2014/04/ioth-the-internet-of-things-and-humans.html

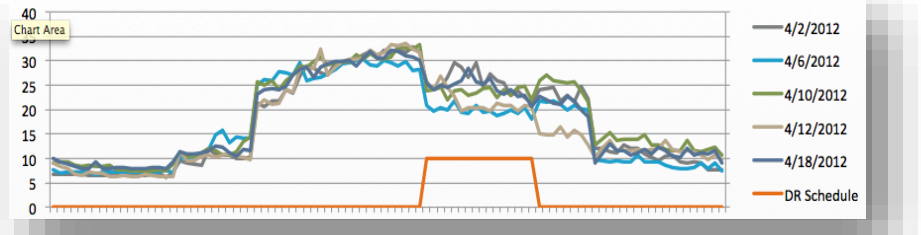
Smart Power Grids: Demand Response Optimization

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***







IoT: Edge+Cloud

- Distributed Processing
  - Use both edge devices (P2P) & central devices (Clouds)
- Leverage capabilities of mobile & edge platforms
- Reduce latency for decision making
- Ensure privacy of data
- *Challenge: How do we intelligently distribute the processing? How do we adapt to changing situations, reliability?*

5 min peer discussion

Map app types to sys models

HPC, HTC, OLTP, Big Data →

Cluster, Grid, P2P, Cloud

Utility vs Exclusive Computing

- Exclusive computing
  - Captive compute resources
  - Can own the h/w or lease it long term
    - » Reserved instances
- Utility computing
  - Computing as a “pay as you go” service
  - On-demand, on tap; multi-tenancy, multi-inst.
  - Grids recently pioneered it, but were complex
  - Public Clouds are a continuation
- *Is your “personal” computer obsolete?*

Reading Assignment

- Chapter 1 of text book

Assignments

- Homework A
  - Now available; due by class on *Fri Feb 6*
  - Project & team selection due on Feb 5.
- Research reading list
  - Available Feb 3. Review and select one by *Tue Feb 10.*