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

DREAM:Lab

SE252:Lecture 1, *6 Jan 2015*

Introduction to Cloud Computing

Yogesh Simmhan

©DREAM:Lab, 2014
This work is licensed under a Creative Commons Attribution 4.0 International License

DREAM:Lab

SE252

Introduction to Cloud Computing, *Jan 2014*

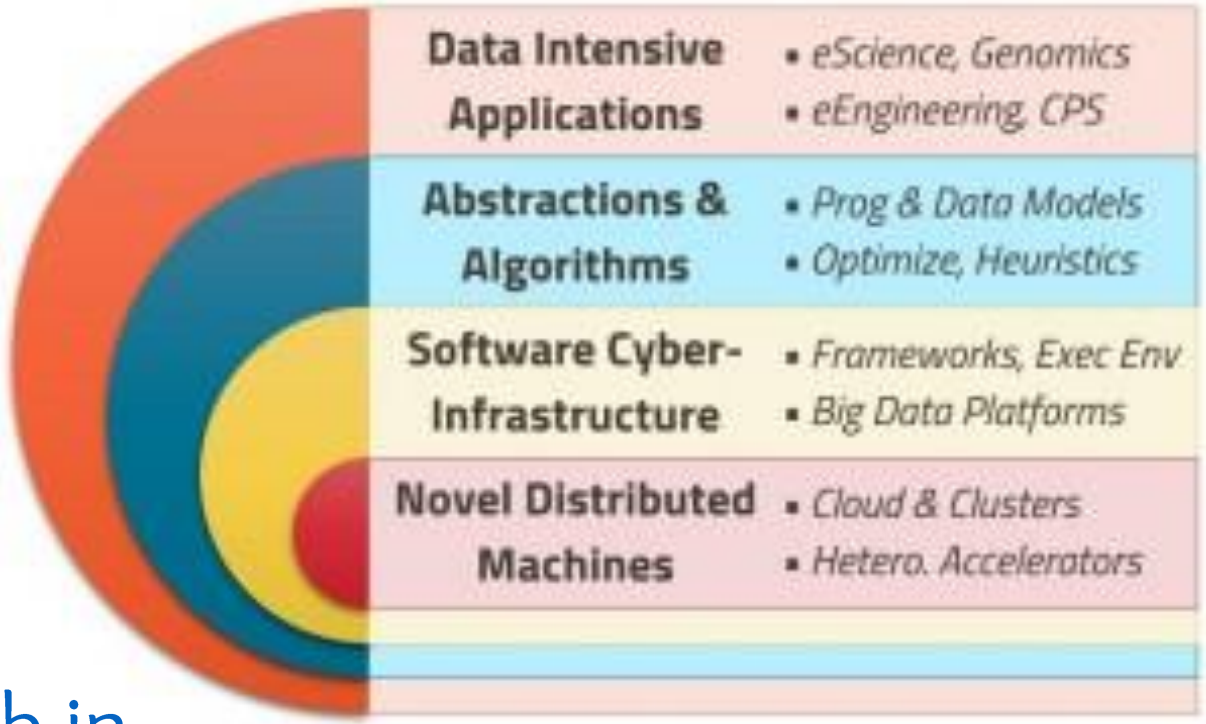
Who am I?

- "Yogesh"
  - ...or Dr. Simmhan
  - ~~Not Dr. Yogesh, "Sir", other honorifications~~ ☺
- Assistant Professor, SERC
- Former:
  - Research Assistant Professor, USC, Los Angeles
  - Postdoc, Microsoft Research, San Francisco
  - Indiana University graduate

The DREAM:Lab

- **D**istributed **R**esearch on **E**merging **A**pplications and **M**achines
  - Applied distributed systems research

Open for New Students!



- www.dream-lab.in

Who are you?

- Name
- Department
- Degree & Year
- Have you signed up?

What this course is

1. Cloud computing as a technology
   - How to use Cloud tools, APIs, SDKs?
   - *Access to Amazon AWS IaaS Cloud*
2. Cloud computing as a distributed systems environment
   - Why cloud computing works?
   - How to design applications for Clouds?
3. Cloud computing as a research topic
   - What are the gaps and emerging ideas?

This course is NOT for you if:

- You have not taken algorithms/data structures/OS/networking UG courses
- You're not comfortable with programming
  - Network programming
  - ✓ Java, Python
  - .NET
  - ~~× C/C++~~
  - ➢ *No TA, so you should be able to resolve programming problems on your own!*

Intended Learning Outcomes (ILO)

- **What** you need to be *able to do* after *learning* from the course…expectations

1. **Remembering** Define, describe, draw, find, identify, label, list, match, name, quote, recall, recite, tell, write
2. **Understanding** Classify, compare, exemplify, conclude, demonstrate, discuss, explain, identify, illustrate, interpret, paraphrase, predict, report
3. **Applying** Apply, change, choose, compute, dramatize, implement, interview, prepare, produce, role play, select, show, transfer, use
4. **Analysing** Analyse, characterize, classify, compare, contrast, debate, deconstruct, deduce, differentiate, discriminate, distinguish, examine, organize, outline, relate, research, separate, structure
5. **Evaluating** Appraise, argue, assess, choose, conclude, critique, decide, evaluate, judge, justify, predict, prioritize, prove, rank, rate, select, monitor
6. **Creating** Construct, design, develop, generate, hypothesise, invent, plan, produce, compose, create, make, perform, plan, produce

Source: Anderson and Krathwohl (2001); Biggs and Tang (2007)

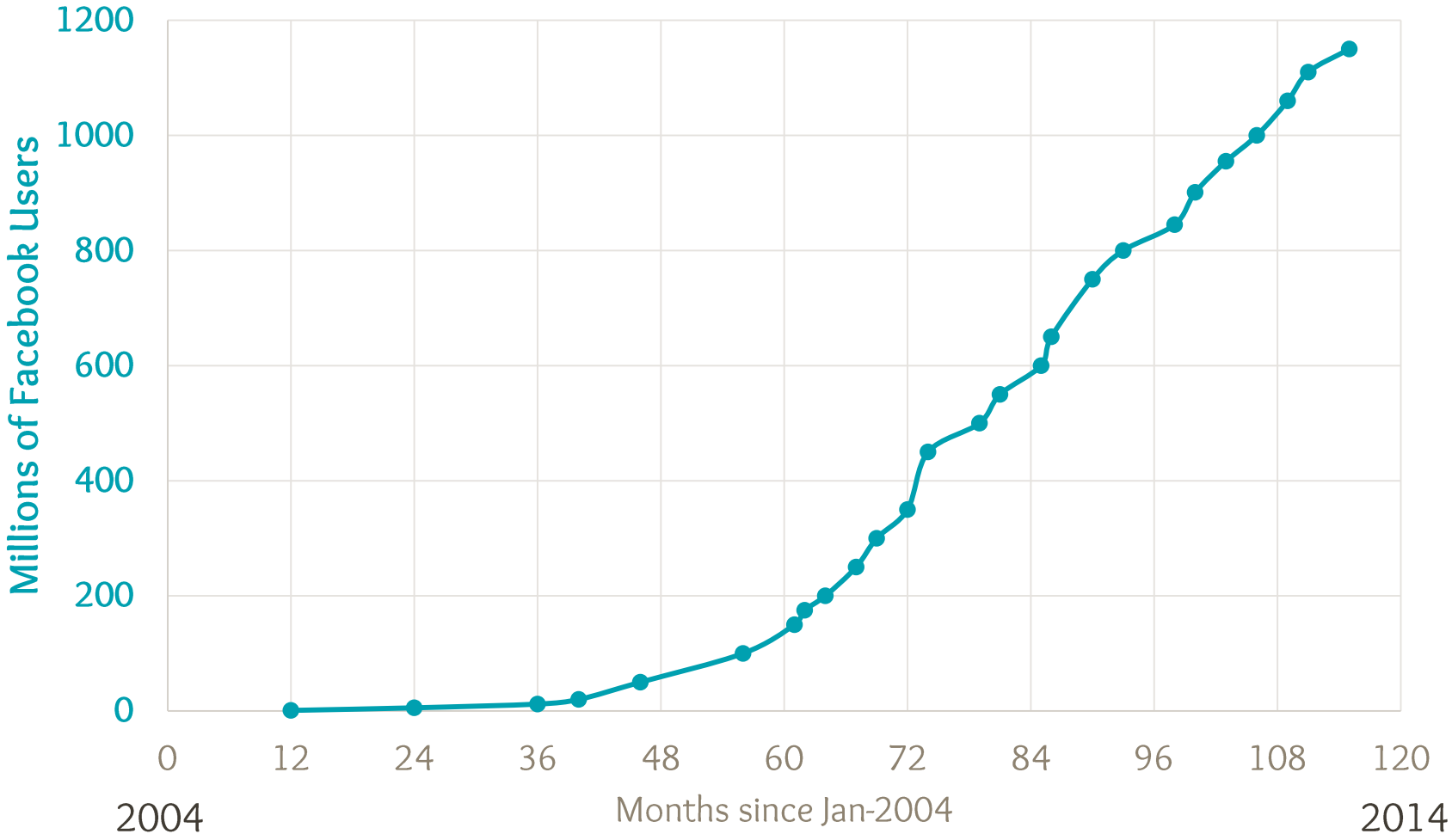
DREAM:Lab

Facebook: A *Canonical* Cloud Based Application

DREAM:Lab

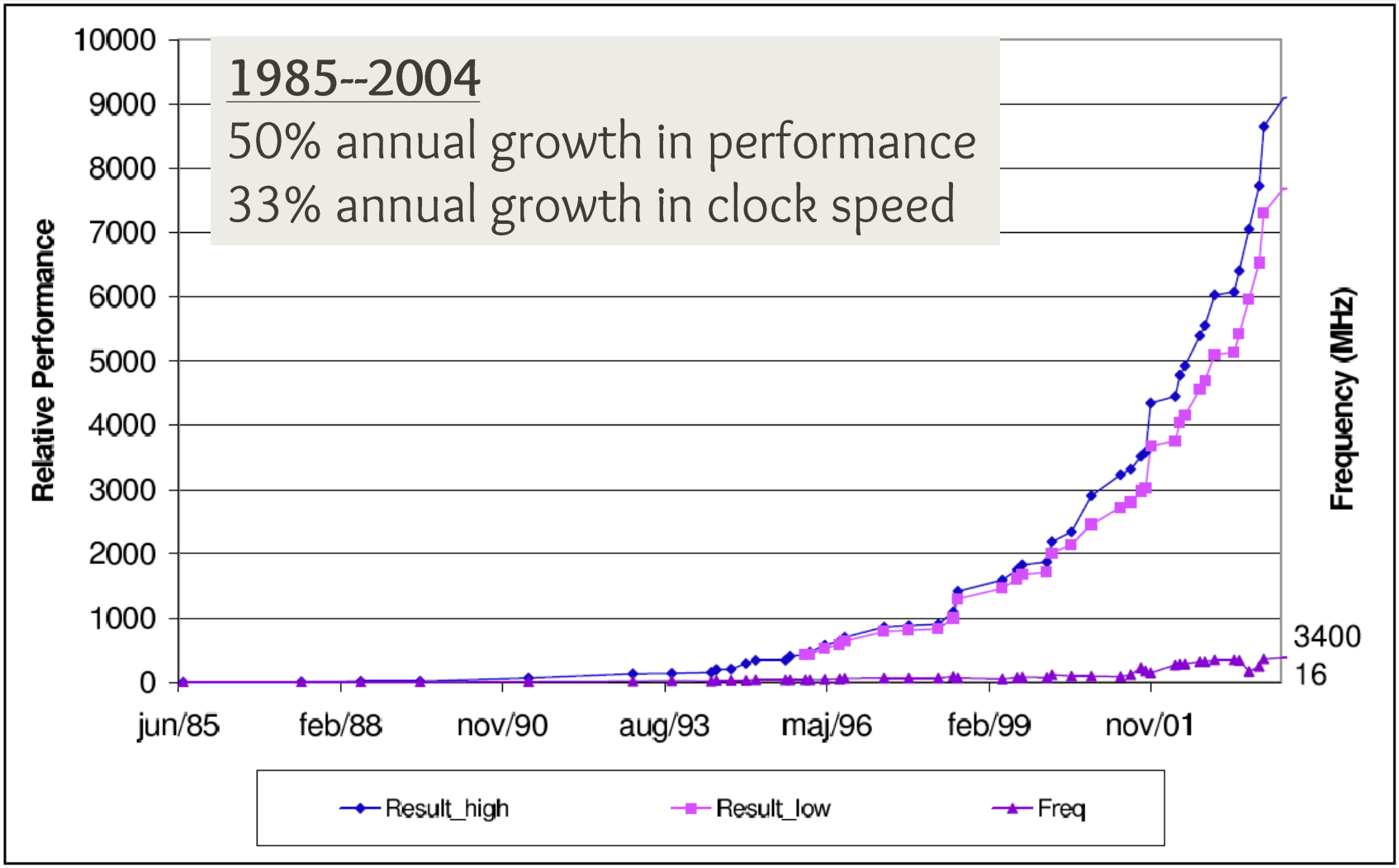
ILO1: Parallel and Distributed Systems Context

Growth of Facebook



http://www.benphoster.com/facebook-user-growth-chart-2004-2010/

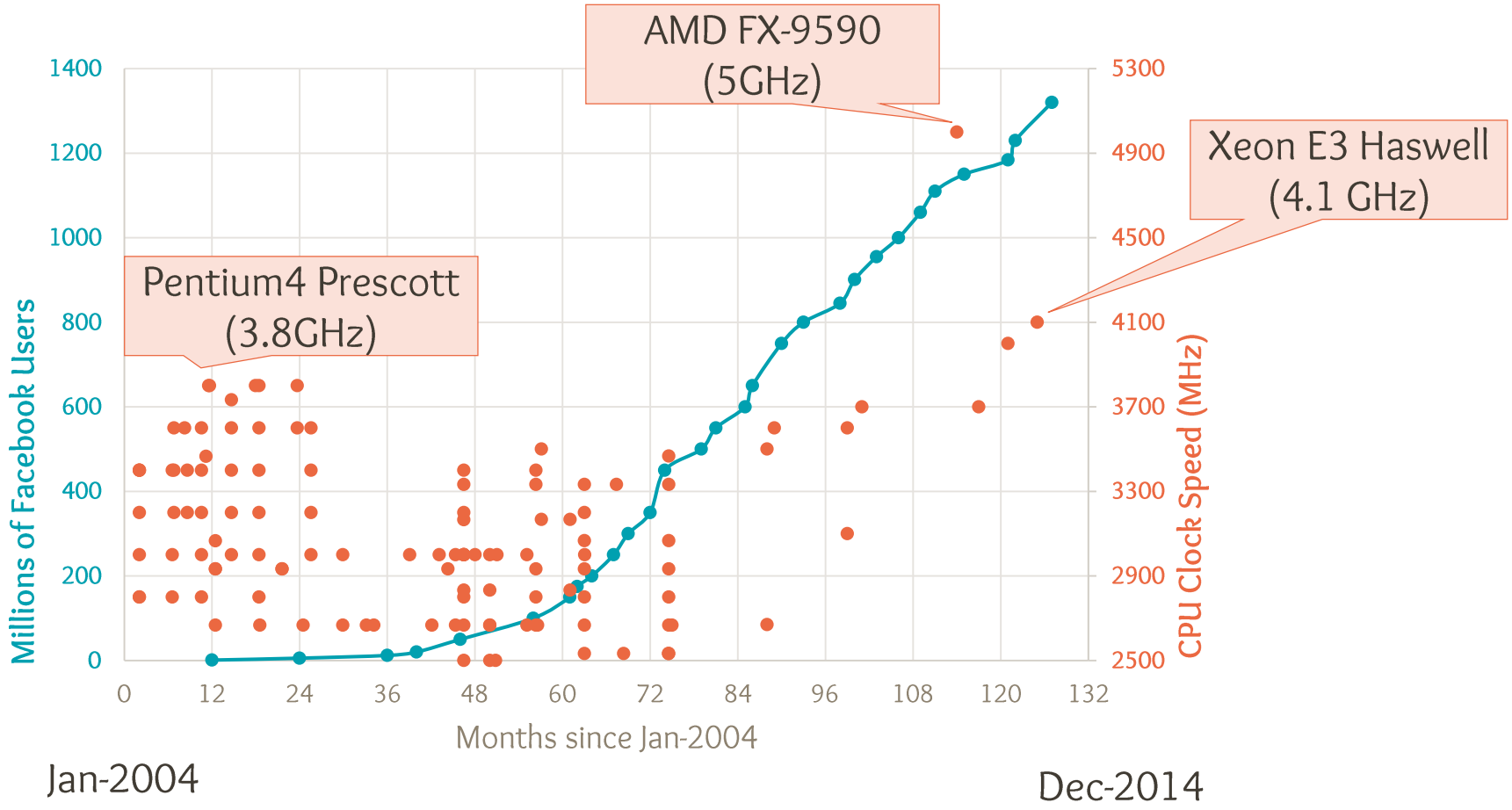
Single CPU Performance (Scale **UP**)



Computer performance 1985-2004, normalized to SPEC CINT89.

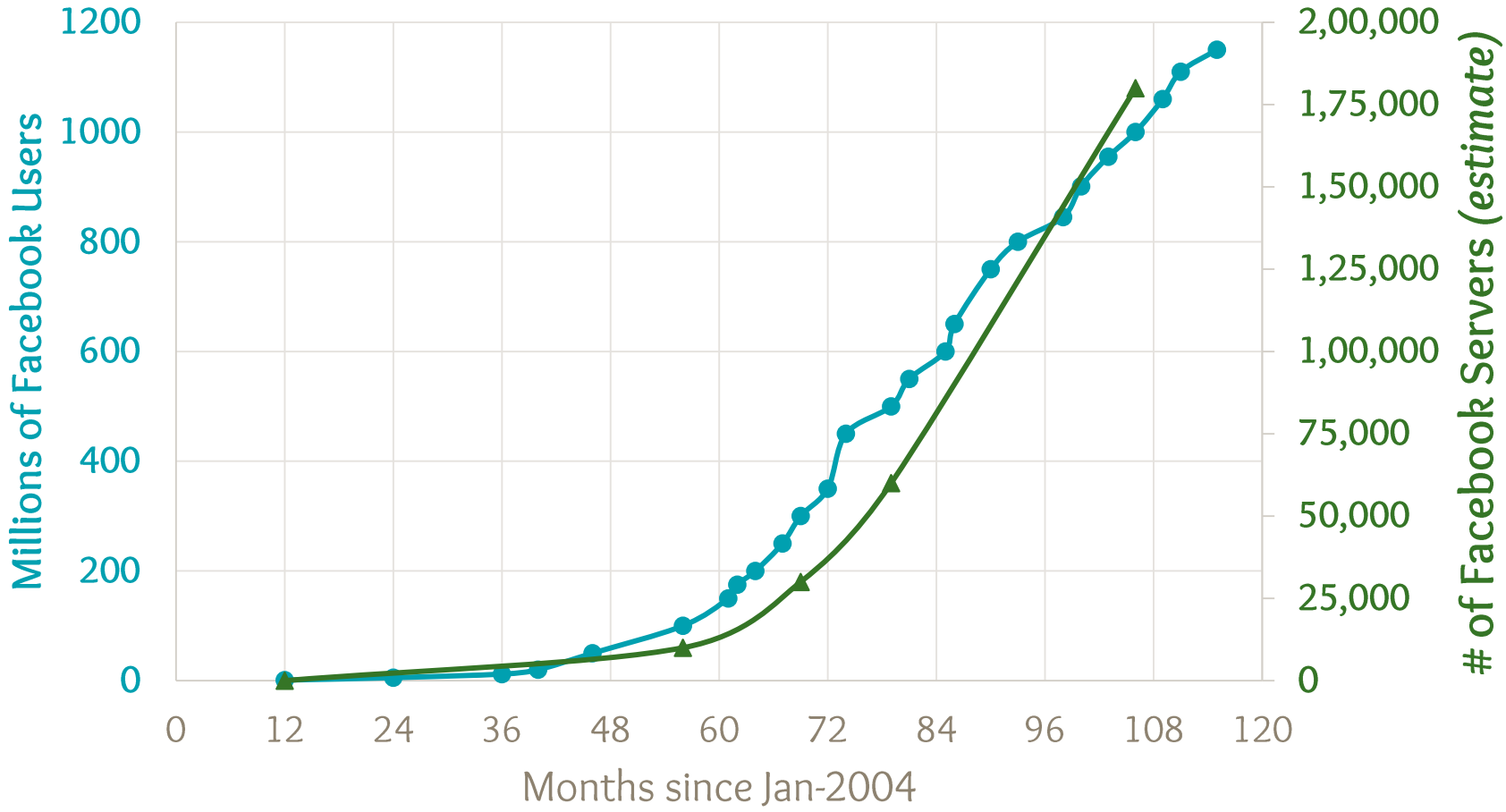
"An In-Depth Look at Computer Performance Growth", Ekman, et al, 2004

Facebook & Clock Speed Growth



http://www.r-bloggers.com/cpu-and-gpu-trends-over-time/

Facebook & Scale **OUT**



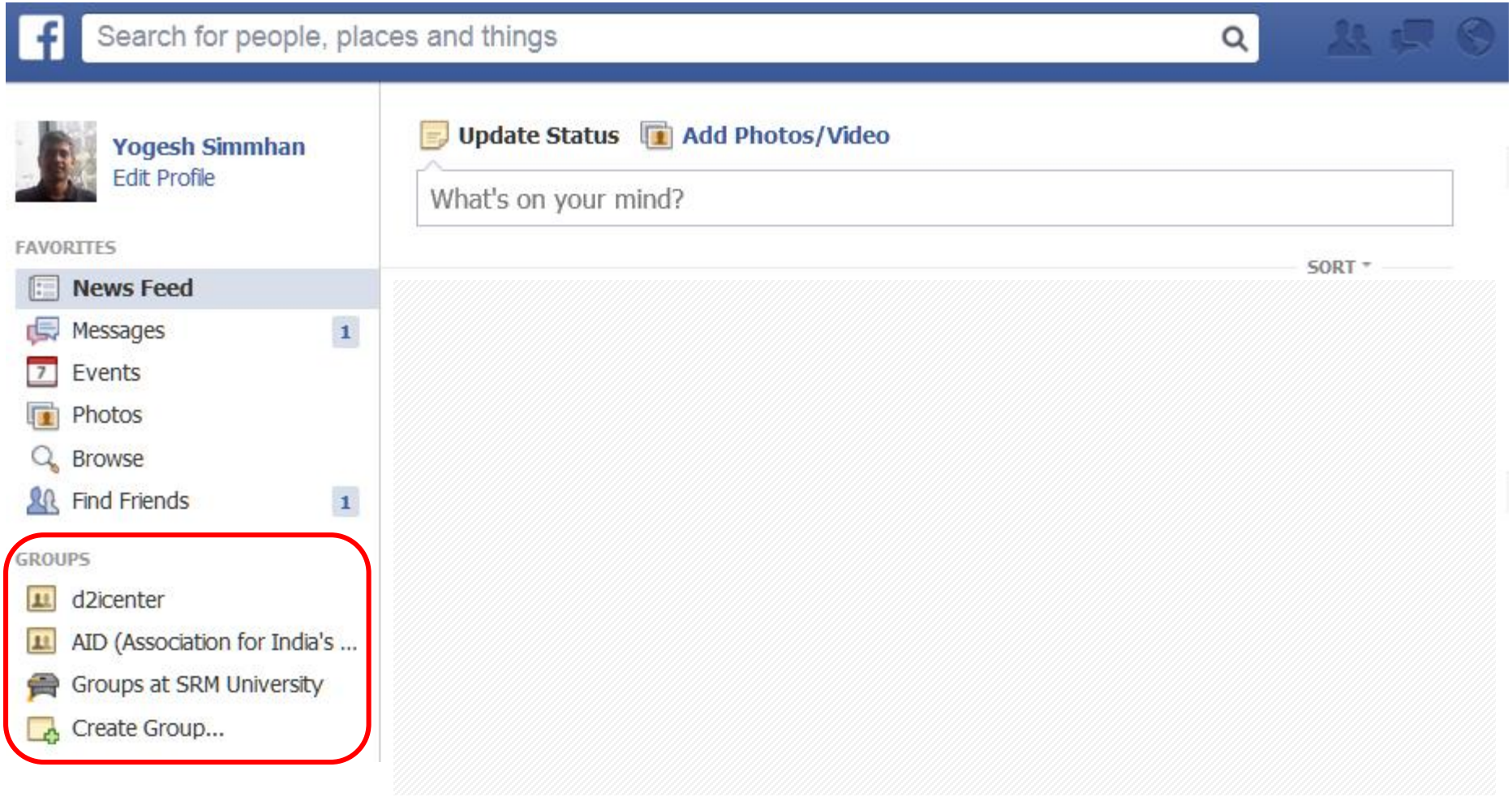
Scale **Up** vs Scale **Out**

- Understand parallel computing paradigms
  - Shared memory vs. Distributed memory
  - Parallel, Distributed, Cloud, …
  - Concurrency from micro to macro scale
  - Performance, scalability, efficiency
- Taxonomy, Pros & Cons, When to pick what, …
- ILO1: Parallel and Distributed Systems Context

DREAM:Lab

IL02: Cloud Virtualization, Abstractions and Enabling Technologies

My Facebook "Groups"



My Browser ⟷ Internet ⟷ "Facebook" Server

Client-Server Interactions

- Request-Response model
- Often ask for a static content
  - https://www.facebook.com/images/loaders/indicator_blue_medium.gif
  - https://fbstatic-a.akamaihd.net/rsrc.php/v2/yK/r/Rf0y2zAXKrZ.gif
  - Served by facebook or other Content Distribution Networks
- Sometimes perform a "Remote Procedure Call"
  - Get the list of groups that I belong to
- Serialization, deserialization

developers.facebook.com/tools/explorer

Facebook Graph Explorer

GET → /v2.2/6839096?fields=id,name,groups Submit

Learn more about the Graph API syntax.

Request: User ID & Fields to "GET"

Node: 6839096

- [x] id
- [x] name
- [x] groups →

+ Search for a field

+ Search for a field

Response: List of group names


```
{
  "id": "6839096",
  "name": "Yogesh Simmhan",
  "groups": {
    "data": [
      {
        "name": "AID Bangalore",
        "unread": 7,
        "bookmark_order": 2,
        "id": "536069206450363"
      },
      {
        "name": "d2icenter",
        "unread": 13,
        "bookmark_order": 1,
        "id": "268551846502548"
      },
      {
        "name": "AID (Association for India's Development) ",
        "administrator": true,
        "bookmark_order": 3,
        "id": "4848588894"
      }
    ],
    "paging": {
      "next": "https://graph.facebook.com/v2.2/6839096/groups?icon_size=16&limit=5000&offset=5
    }
  }
}
```


RPC using RESTFull Request Response using HTTP Standards

DEMO

https://graph.facebook.com/v2.2/6839096?fields=id%2Cname%2Cgroups&format=json&method=get

Service Oriented Architecture (SOA)

- Allows standard way for clients & services to
  - Exchange data Structures over the network
    - » Serialization & Deserialization
  - Invoke remote methods over the network
  - Defining QoS, discovering services, etc.
- SOA is an enabling *technology* for Clouds
  - E.g. SOAP, REST, Thrift*, Protocol Buffers, etc.
  - RPC *concept* has existed since 1980's

***Thrift: Scalable Cross-Language Services Implementation,** Mark Slee, Aditya Agarwal and Marc Kwiatkowski, 2007

SOA & Virtualization

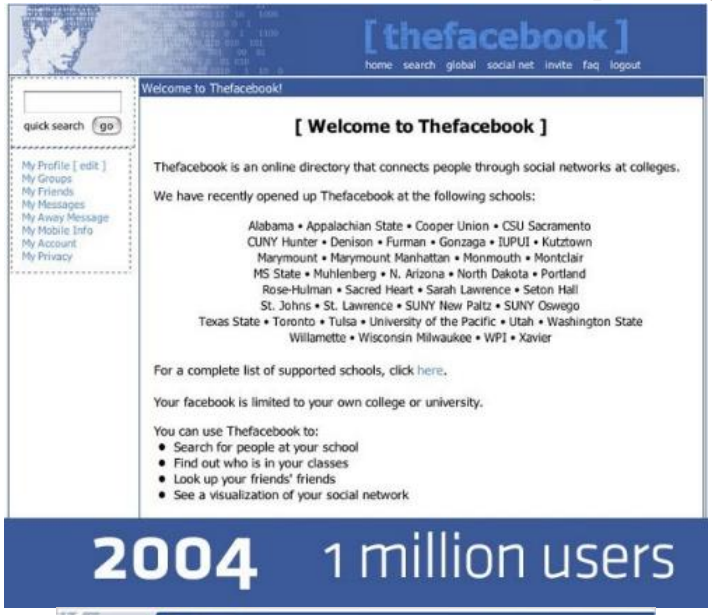
- All Cloud operations can be performed as web service calls by application
  - Storage services, VM Management services,…
- Virtualization is the other key enabler
- ILO2: Cloud Virtualization, Abstractions and Enabling Technologies

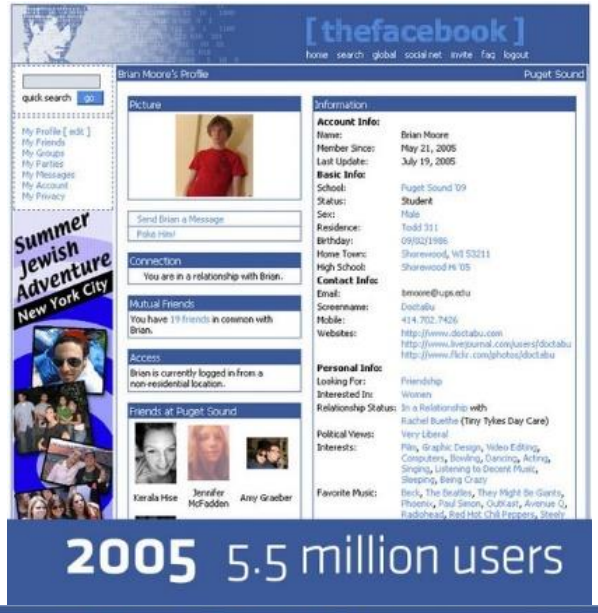
DREAM:Lab

ILO3: Algorithms and Programming Patterns for Cloud Applications

Facebook Webpage over the Years

http://mashable.com/2011/02/04/facebook-7th-birthday/





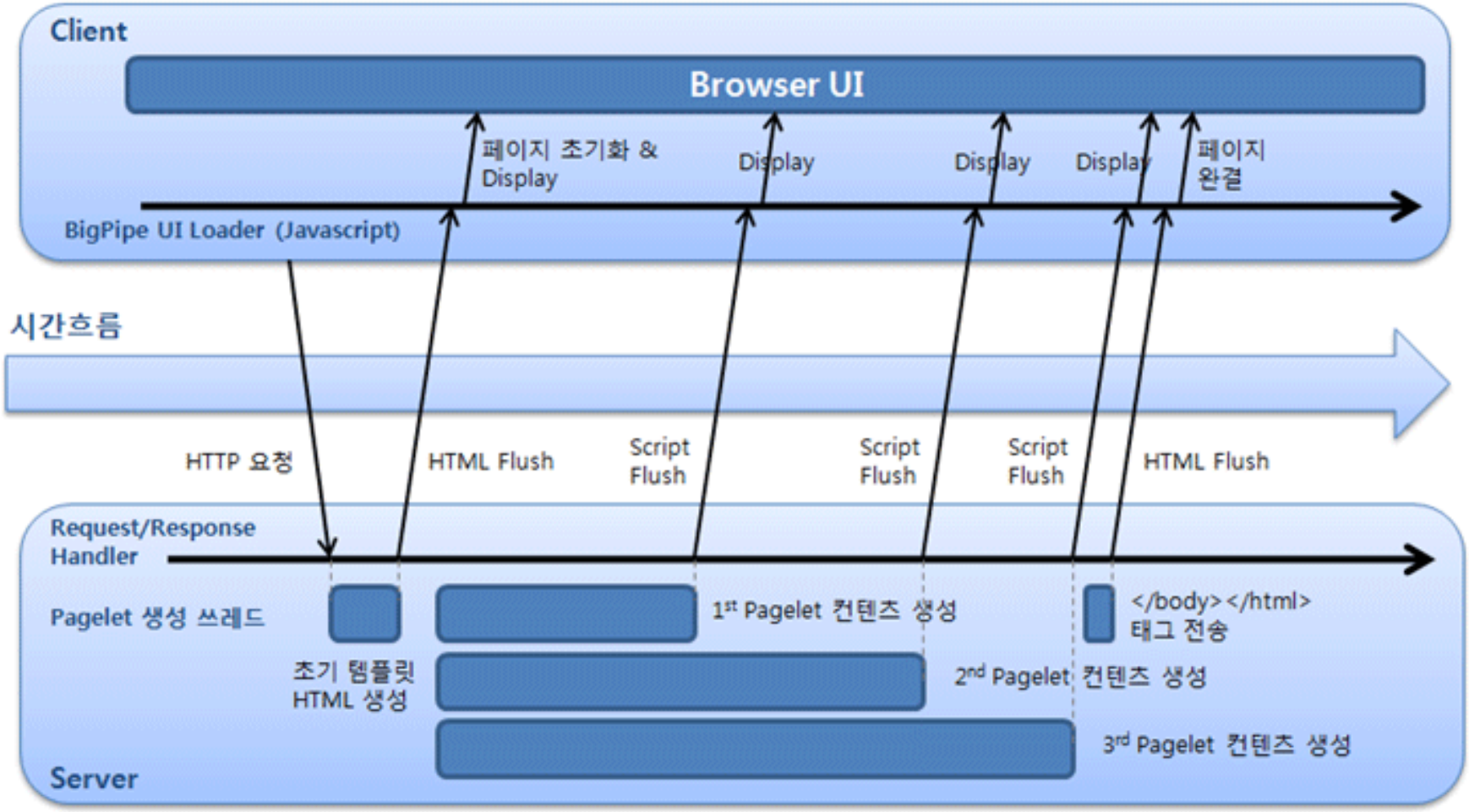
Facebook Webpage over the Years

http://mashable.com/2011/02/04/facebook-7th-birthday/

Facebook's BigPipe

https://www.facebook.com/note.php?note_id=38941403391 9

Pipelined vs Single Shot Page Loading

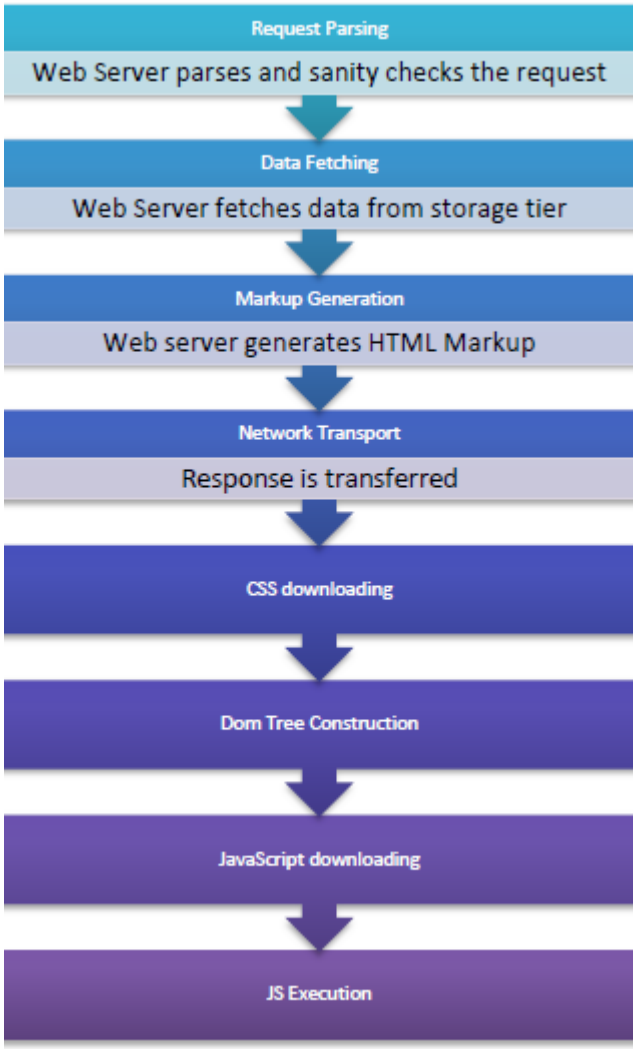


DEMO

http://www.cubrid.org/blog/dev-platform/faster-web-page-loading-with-facebook-bigpipe/

Data Parallel Execution

- Each *pagelet* is independent unit of rendering
- Same set of tasks executed for each pagelet (data), *in parallel*
  - CSS D/L, JS D/L, Render
- This is also a task "graph"
  - Sequence of tasks that execute one after the other



http://www.slideshare.net/AditiTechnologies/facebook-architecture-breaking-it-open

Data & Task Parallel Models

- Data parallel model is common in Clouds
  - E.g. Map-Reduce, Giraph
  - Helps exploit independent units of data on multiple machines/processors/threads
- Task parallel model help in composition
  - Allows tasks to operate concurrently on same/different data

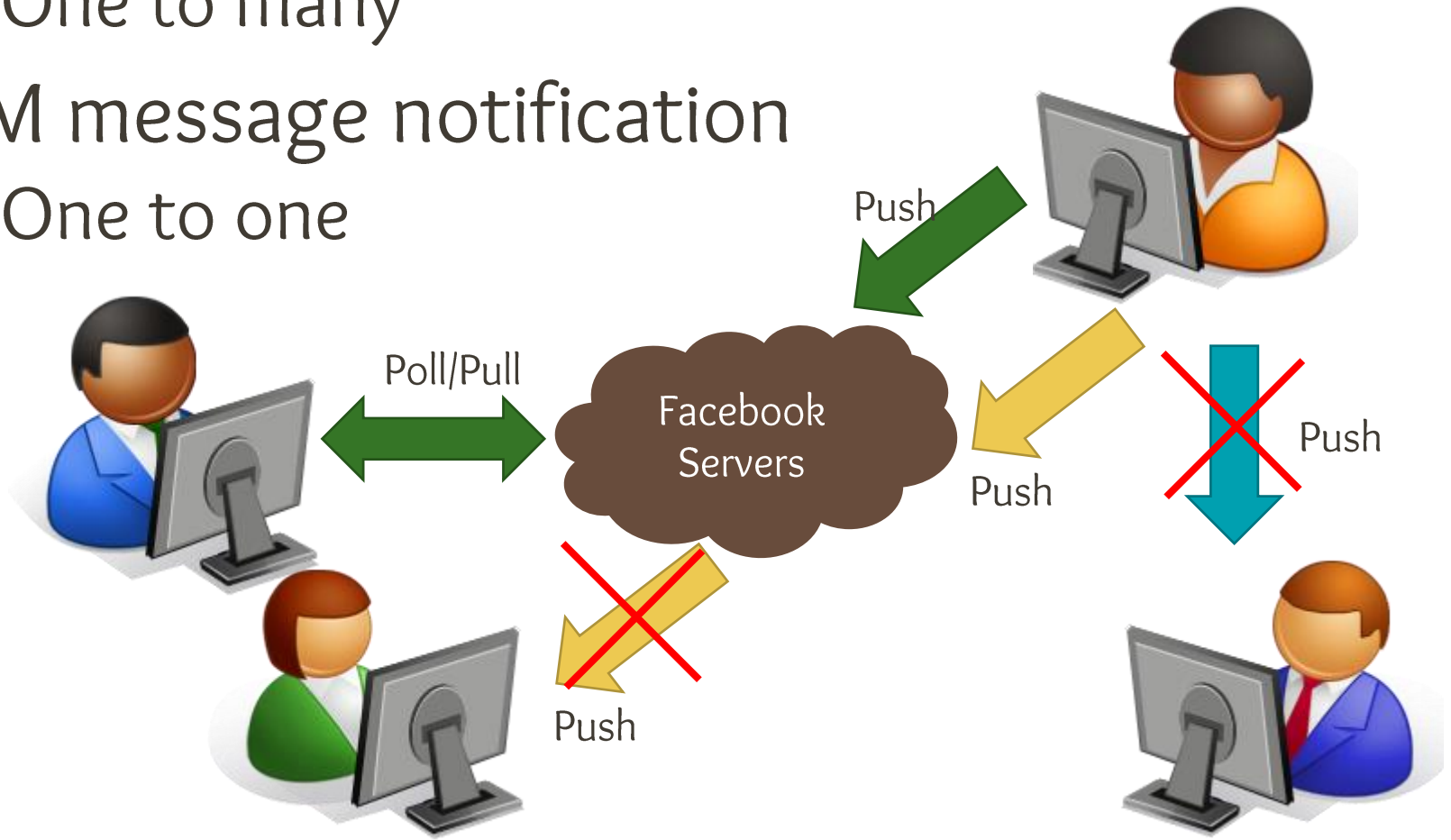
- ILO3: Algorithms and Programming Patterns for Cloud Applications

DREAM:Lab

ILO 4: Application Execution Models on Clouds

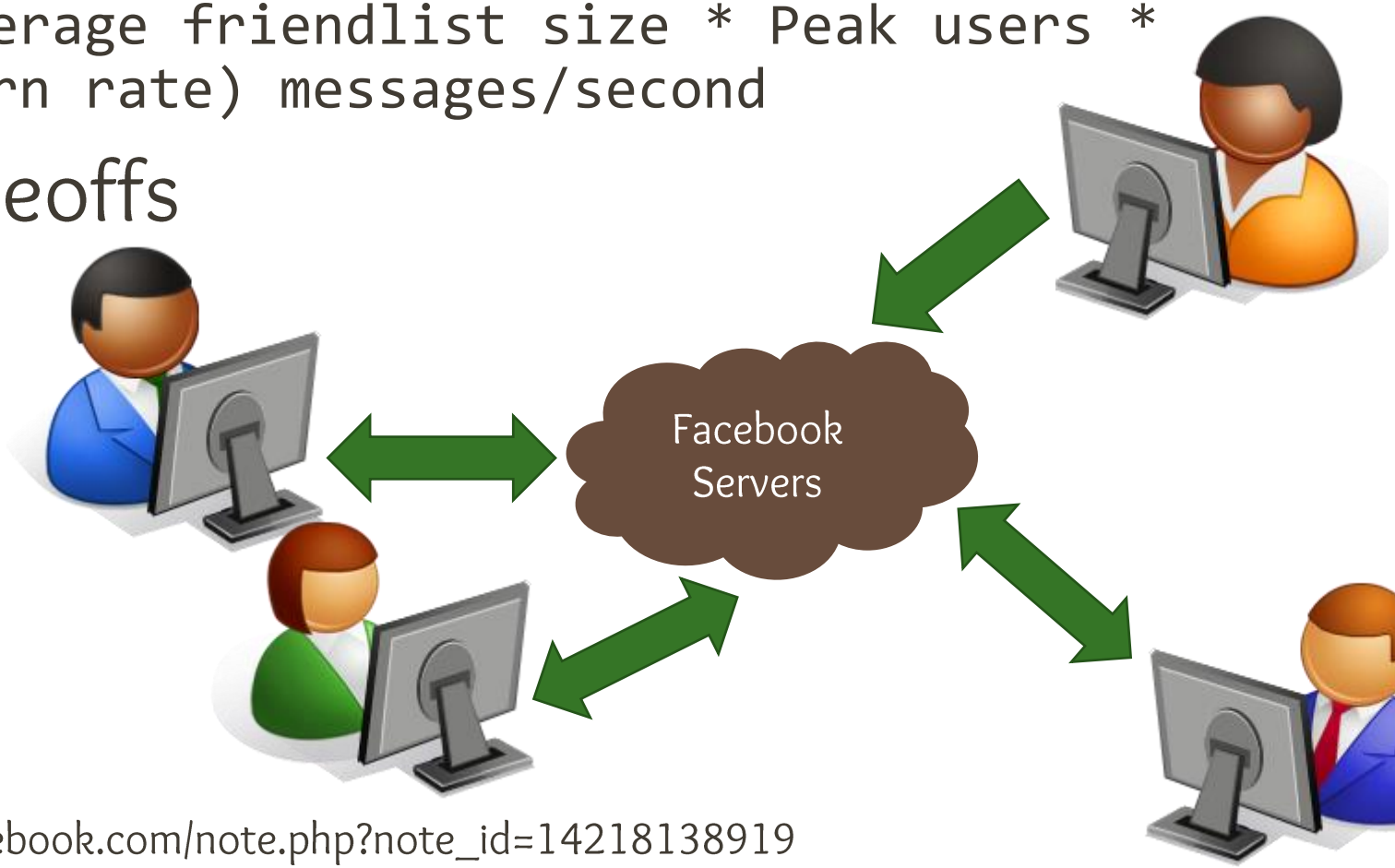
Facebook (Browser) Chat

- Online/offline status
  - One to many
- IM message notification
  - One to one



Synchronous vs Asynchronous

- Immediately or eventually delivered?
  - Cost of sync goes up with # of users
  - `(Average friendlist size * Peak users * Churn rate) messages/second`
- Tradeoffs



https://www.facebook.com/note.php?note_id=14218138919

Programming Cloud Apps

- Coordinate execution across VMs, clients
- Synchronization of activities
- Storage-based vs. In-memory operations
- Trade-off between guarantees

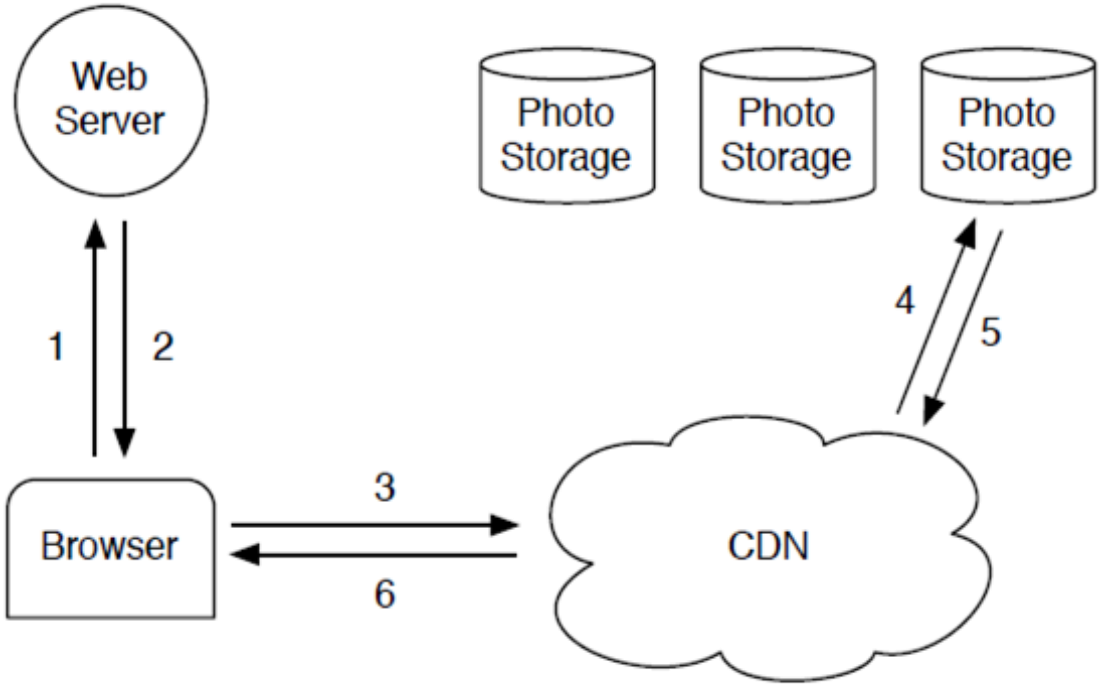
- ILO 4: Application Execution Models on Clouds

DREAM:Lab

IL05: Performance, scalability & consistency on Clouds

Data Locality for Facebook Photos

- Content Distribution Networks (CDN)
  - Host "Hot" data, Spatially close to the client
  - E.g. Akamai
- Low latency
- Freshness?
- Consistency?

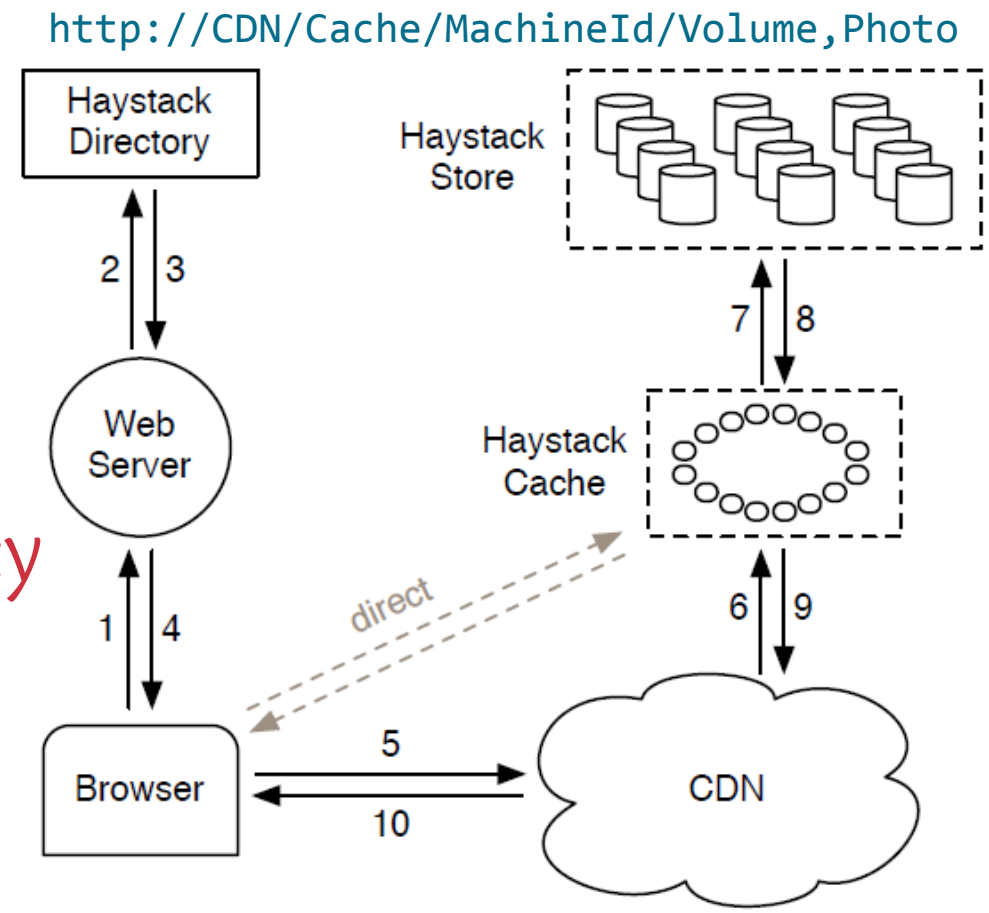


Finding a needle in Haystack: Facebook's photo storage
Doug Beaver, Sanjeev Kumar, Harry C. Li, Jason Sobel, Peter Vajgel

Performance, Availability, Consistency, Scalability, …

- CDNs are costly for long tail
  - And poor cache hits
- Haystack
  - Incremental URLs
  - Better caching
  - Low latency response

- ILO5: Performance, scalability & consistency on Clouds



Bottomline

Cloud is ~~a revolution~~ **an important evolution**

- Yet another distributed system
- Novel tools, technologies, applications
- Builds on fundamental concepts
  - Distributed systems
  - Programming models & algorithms
- Different applications, reliability, performance trade-offs

Teaching and Learning Activities (TLA)

- **How** we'll go about meeting the ILO's
  1. Lectures
  2. Guest Lectures
  3. Homework
  4. Research Reading & Summarization
  5. Project
  6. Exam

Biggs and Tang (2007)

Assessment *(1000 point scale)*

- **30% Homework** Three homework assignments *(100 points each)*.
- **10% Research Summary** Reading and summary report on one research paper *(100 points)*.
- **30% Project** One ungraded and two graded midterm *(100 points)* and final *(200 points)* project assignments.
- **25% Exams** One Mid-term *(100 points)* and one Final *(150 points)* exam.
- **5% Participation** Participation (i.e. not just "attendance") in classroom discussions and online forum for the course *(50 points)*.

Academic Integrity

- Students must uphold IISc's Academic Integrity guidelines. While these are common sense, it is helpful to review them since failure to follow them will lead to sanctions and penalties.
- **This includes a reduced or failing grade in the course. Severe cases of academic violations will be reported to the Institute and may lead to an expulsion.**

http://www.iisc.ernet.in/students-corner/existingstudents-academicintegrity.php

Academic Integrity

- Learning takes place both within and outside the class. Hence, **discussions** between students and reference to **online material** is **encouraged** as part of the course to achieve the intended learning objectives.
- **However**, while you may learn from any valid source, you **must form your own ideas** and **complete problems and assignments by yourself.**
- *All works submitted by the student as part of their academic assessment must be their own!*

Academic Integrity

- **Plagiarism**: *Verbatim* reproduction of material from external sources (web pages, books, papers, etc.) is **not acceptable**.
- If you are *paraphrasing* external content (or even your own prior work) or were otherwise *influenced* by them while completing your assignments, projects or exams, **you must clearly acknowledge them**.
- *When in doubt, add a citation!*

Academic Integrity

- **Cheating**: While *you may discuss* lecture topics and broad outlines of homework problems and projects with others, **you cannot collaborate** in completing the assignments, **copy** someone else's solution or **falsify results**.
- You **cannot use** notes or unauthorized resources during exams, or copy from others.
- The narrow exception to collaboration is between team-mates when competing the project, and even there, *the contribution of each team member* for each project assignment should be **clearly documented**.

Academic Integrity

- **Classroom Behaviour**: Ensure that the course atmosphere, both in the class, outside and on the online forum, is conducive for learning.
- *Participate* in discussions but *do not dominate* or be abusive. **There are no "stupid" questions** ☺ Be considerate of your fellow students and avoid disruptive behaviour.

Resources

- **Textbook:** ***Distributed and Cloud Computing: From Parallel Processing to the Internet of Things***, Kai Hwang, Jack Dongarra and Geoffrey Fox, *Morgan Kaufmann*, 2011 (Tata Book House)
  - Select sections from Part I and II (Ch. 1-6)
- **Official Mailing List:** se252.jan15@mailman.serc.iisc.in
  - Everyone **MUST** be on the list. All *official course announcements* will be posted on it.

How to use the mailing list

- Unmoderated
- Open for **everyone** to discuss
- Can suggest resources to others
- ***Do not post partial/full solutions!***
- Can request for official clarifications
- Official announcements will be posted
  - Your responsibility to check regularly

Teaching & Office Hours

- **Lecture:** TTh 2PM-330PM, SERC 202
  - *If you miss the lecture, its your responsibility to catch up from slides & talking to other students!*
- **Office Hours:** Fri, 4-5PM or by appointment *(send email)*, SERC 411
  - May have project discussions at this time, if necessary

Tentative Schedule *(Subject to change…See webpage)*

- http://www.serc.iisc.ernet.in/~simmhan/SE252/