

# Grid Asia 2009, Singapore.

## University Relationships in India

Chid Kollengode

Grid Engineering, Cloud Computing and Data Infrastructure Group  
Yahoo India R & D, Bangalore, INDIA



# Let us begin with the goal!

- Leverage momentum for Hadoop
  - Establish Yahoo! as the technology leader in Cloud Computing
- Get students excited about distributed computing, about Hadoop and about Yahoo
- Evangelize open-source
- Increase odds of a “eureka” by assembling right ingredients
  - Alternative is to “buy” the recipe for success (if there is one!)
- **Provide leadership Off-Network**
  - I40: Provide support, training and define research collaborations
  - Conferences: HiPC, Software Engineering Conference, Hadoop India seminar, Hadoop India User Group (incubating), International Conferences
  - Blogs, paper publications (2 published, 1 – in peer review as of Apr 2009)

# Motivation

- Remove bias towards alternatives that perpetuate status quo
- Together we win!
- Accelerate development of insights by Research
  - Invent-it-ourselves and diminishing returns
- Showcase the cutting-edge work in Yahoo!
- Provide opportunities for our architects to interact with academia
- Stretch target – growing talent is the best way to beat competition (Alternative is trimming ambition to match resources)

**Think Universities!**



## How did we get to where we are

- Opportunistically taking chances that come up vs. target specific institutions and knock on their doors
- Invest time to identify professors with passion for Cloud Computing research and tap into student base
- Diversity is strength - willing to tailor agreements that are specific to each University
- Time and effort to keep the engagement active (avoid out-of-sight is out-of-mind syndrome)
- Shared vision and clear understanding of deliverables from both sides

# Current academic partners

- PSG Tech, Coimbatore
- IISc - Bangalore
- IIT – Mumbai
- IIIT – Hyderabad

- Hadoop cluster purchased and installed
- Hadoop as a case study in Distributed Systems course of BE (CSE), ME(CSE) and ME(SE) programmes.
- “Data Discovery in Grid using Content Based Searching Technique” - 2nd India Software Engineering Conference
- To implement Map reduce programming model for genome sequence alignment (First Version delivered)
- Exploring virtualization on Hadoop (full stack)
- Research assistants and students working on Hadoop projects

# IISc, Bangalore

- Simulation model was built by the professor
- Approaches to increasing efficiencies of map and reduce tasks
- Next step – in collaboration with College of Engineering, Madurai
  - Develop heuristics and optimal techniques based on dynamic analysis of output patterns of map tasks
  - Enroll students to do this research



# IIT-Mumbai

- 40 node cluster order placed, IIT installing it
- Joint research – Grid and Search, areas identified
- Four professors signed on for research, 100 core usage load already identified
- Course work in B.Tech -Two day seminar on Hadoop
- Working with Univ relations in SNV to provide dated crawl data corpus for research (Large data set so far)



# IIIT - Hyderabad

- Latest MoU (signed Jan '09) – national press coverage
- Hadoop used to crawl and index India web sites
- Hadoop core – scheduling and user based access privileges
- Cross Lingual image search – ranking is done on Hadoop
- Image processing tasks in parallel on a number of scanned documents using Hadoop (MR paradigm)
- Hadoop added to syllabus - Grid Computing: Hadoop as DFS environment – (Industry Partner: Yahoo)



# Where are we in the journey?

- Strong branding, we are recognized as the best of the companies working with Universities now
- Already a big part of mindshare of students
- Enough buzz and awareness – we get requests for collaboration
- Nurturing the right talent base by active engagement
- Getting people to contribute to open source (IIIT-H)
- Aid industry-institution interaction and lay strong foundation for Yahoo to build on.

## What's remaining to do?

- Scaling University relations beyond CCDI (to Yahoo Blr)
- Setting up a lab in Yahoo Blr for Univ access
- Hadoop Summit in India – Q3 2009
- Strong User group in Bangalore – Q4 2009
- Stronger International collaborations for Yahoo Hadoop
- CRL collaboration – expanding research collaborations

# Q & A

For any deficiencies, make me the spokesman for defense.

For everything wonderful in this program, the Cloud Computing team (Blr and US) made it possible.

