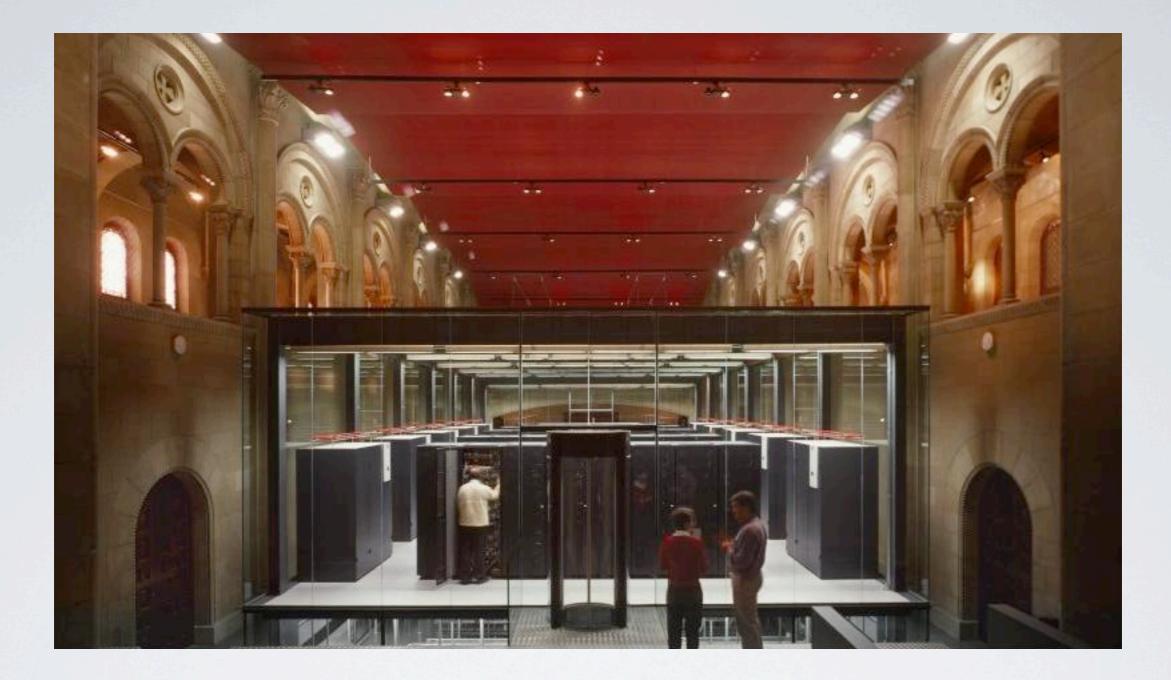
#### ROZPROSZONE PRZECHOWYWANIE DANYCH

#### ZARZĄDZANIE ZASOBAMI SUPERKOMPUTERÓW

dr inż. Krzysztof Rządca <u>krzadca@mimuw.edu.pl</u> mimuw.edu.pl/~krzadca

## ONGOING PROJECTS

- resource management for datacenters: complex applications, CPU/GPU clusters
- nebulostore: a p2p storage system for distributed Internet applications



#### RESOURCE MANAGEMENT IN DATACENTERS

# BIG DATA REQUIRES BIG COMPUTERS

- huge amount of digitalized data: webpages, web activity, shopping trends, location, ...
- to infer information, data must be analyzed and correlated
- cloud computing: for-hire datacenters with thousands of machines

image: google



#### MANAGING RESOURCES OF A DATACENTER: ASSIGN APPLICATIONS TO NODES

- large scale: thousands of nodes and applications
- failures: hard disks, CPUs, power supply, air-conditioning, network
- many applications share a single node
- applications have imprecise resource requirements

image: google



# RESOURCE MANAGEMENT: PROFILE

- MSc thesis profile
  - theoretical work: modeling, developing algorithms
  - validation through simulation on real traces
  - implementation on real resources

- what we offer:
  - cooperation with
    Google
    ICM
    Grenoble
    Rutgers
  - access to supercomputers

## EXAMPLE MSC THESIS TOPICS

- Scheduling jobs with fuzzy (stochastic) resource requirements Overload resources with a certain probability (Robert Rosołek)
- failure-aware scheduling
  Ensure application survivability on failing resources
- scheduling short-running jobs and long-running services
  Balance resource consumption of two classes
- fair scheduling algorithms in multi-user environment
- implementation of OStrich algorithm in SLURM (Filip Skalski)

# NEBULOSTORE: A P2P STORAGE SYSTEM

- privacy problems in organizationally-centralized systems
- distributed on-line social networks
- universal Internet-scale storage system

Project operated within the Foundation for Polish Science "Homing Plus" Programme co-financed by the European Regional Development Fund (Innovative Economy Operational Programme 2007-2013) European Funds — Grants for innovation



INNOVATIVE ECONOMY NATIONAL COHESION STRATEGY



EUROPEAN REGIONAL DEVELOPMENT FUND



#### LINKEDIN&FACEBOOK ARE FUNILinked in . Account Type: Basic Home Profile Contacts Groups Jobs Inbox Companies News More Add Connections Colleagues Alumni People You May Know Filter By .... (2nd) Olga I met her husband during my Phd Post-Doctoral Associate at Iowa State Unive Current Company 1 2 connections in common All Companies INRIA (12) Agata We met at yoga classes in Singapore Accenture (11) Senior metrologist at National Hewlett-Packard (9) Poznan Supercomputing and Just joine Miguel Networking Center (6) My wife's cousin ICM Warsaw University (5) Past Company Ξ Caroline Director of CREATE and create@centre-sta All Companies INRIA (16) Hewlett-Packard (11) Lalit It's scary! Undergraduate Student at IIT Kharagpur Accenture (6) Procter & Gamble (5) (3rd) Kacper ICM Warsaw University (4) Independent Market Research Professional School (3rd) Jacek All Schools aaa at Vip-Zone Poland Politechnika Warszawska (40) Uniwersytet Warszawski (21) Tutut Student at Universitas When shared a bus in Jogjakarta, Indonesia Polsko-Japońska Wyższa Szkoła Technik Komputerowych w Warszawie (16) Szymon 3rd Institut national polytechnique de my Master student Grenoble (11) Intern IT at UBS Investment Bank Université Joseph Fourier

# DISTRIBUTED ONLINE SOCIAL NETWORKS: A P2P FACEBOOK

• wall, messaging, file sharing, likes



- p2p infrastructure under the hood
- fine-grained privacy

# NEBULOSTORE: A ORGANIZATIONALLY-DECENTRALIZED STORAGE FOR DECENTRALIZED APPS

- bilateral storage contracts: I store your data if you store mine
- redundancy for availability and survivability
- prototype implementation

#### MSC THESIS TOPICS

- Privacy-aware data query language (Marcin Walas)
- Efficient downloads and subscriptions by BitTorrent (Rafał Hryciuk)
- Distributed lists for Distributed Online Social Networks (Jadwiga Kańska)

- Cloud-assisted storage
- Crypto utils for privacy & security

# NEBULOSTORE: PROFILE & BENEFITS

- ideal profile:
  - systems / algorithms
  - modeling

- our offer
  - training in distributed systems
  - tests on planet-lab (1000+computers)
  - collaboration with partners from KTH (Stockholm), TU Darmstadt, NTU (Singapore)

#### SUMMARY: RESOURCE MANAGEMENT IN LARGE-SCALE SYSTEMS

- scheduling in supercomputers: complex applications, large-scale infrastructure
- nebulostore: a p2p storage system for distributed Internet applications



#### DZIĘKUJĘ! dr inż. Krzysztof Rządca pok. 4280 <u>krzadca@mimuw.edu.pl</u>, mimuw.edu.pl/~krzadca