ROZPROSZONE PRZECHOWYWANIE DANYCH
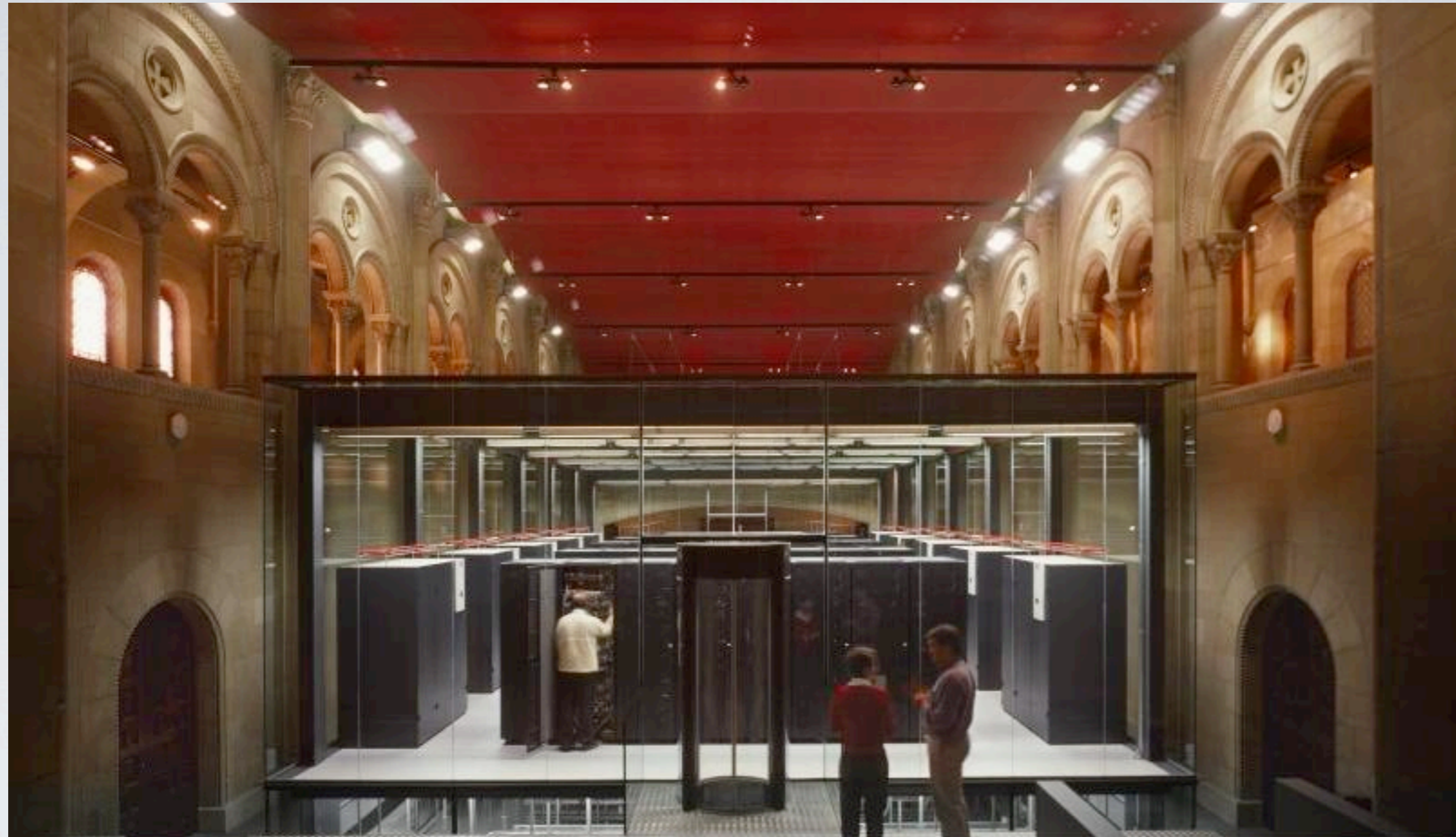
ZARZĄDZANIE ZASOBAMI SUPERKOMPUTERÓW

dr inż. Krzysztof Rządca

krzadca@mimuw.edu.pl

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

INNOVATIVE ECONOMY
NATIONAL COHESION STRATEGY

FNP
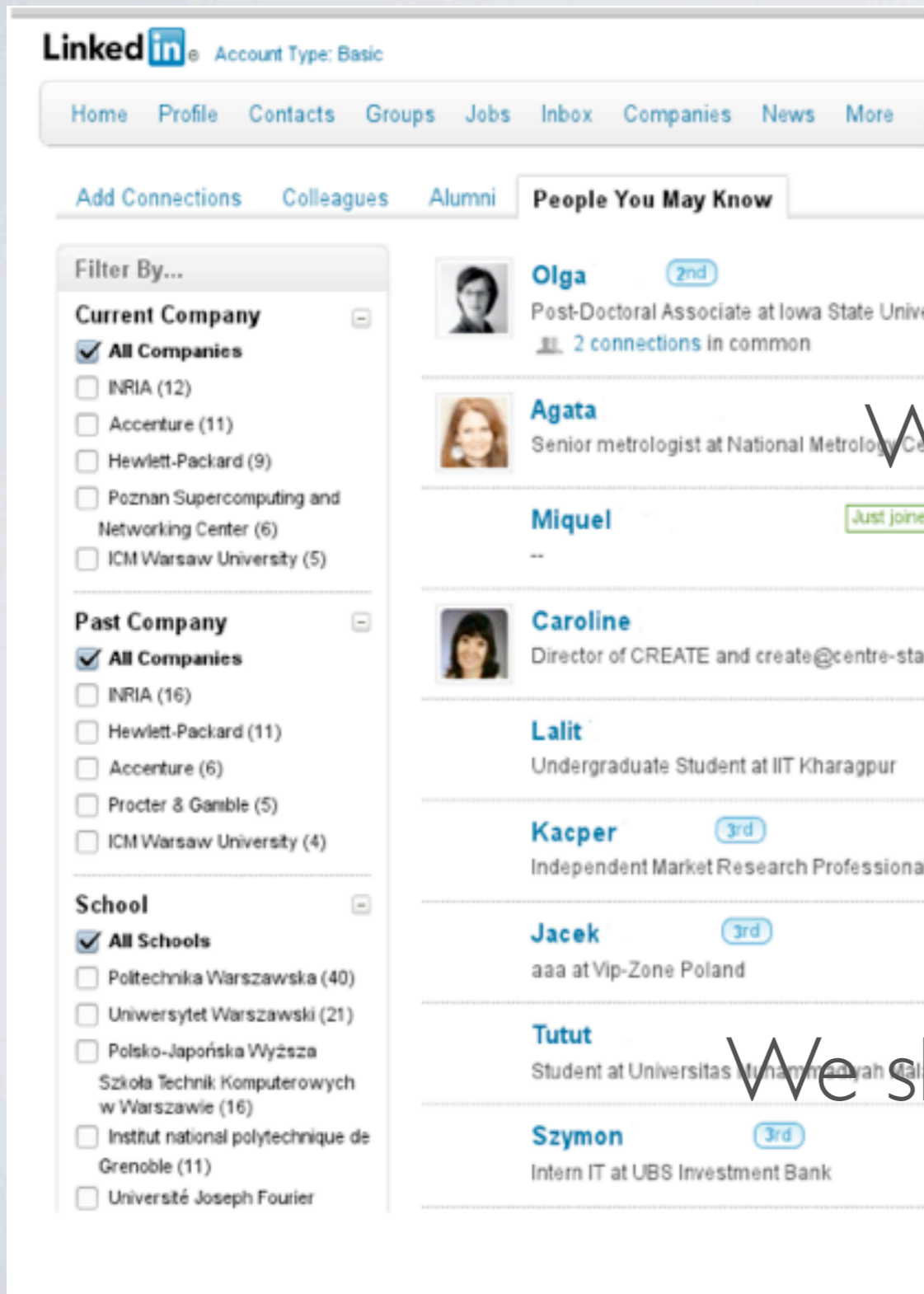Foundation for Polish Science

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND

LINKEDIN&FACEBOOK ARE FUN!

I met her husband during my Phd

We met at yoga classes in Singapore

My wife's cousin

It's scary!

We shared a bus in Jogjakarta, Indonesia
my Master student

DISTRIBUTED ONLINE SOCIAL NETWORKS: A P2P FACEBOOK

PeerSoN

- 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
krzadca@mimuw.edu.pl, mimuw.edu.pl/~krzadca