

*Hanoi, 12/2009*

*FAIR Conference*



# **Giải pháp lưới dữ liệu quản lý tài liệu điện tử được lưu trữ phân tán trong tổ chức ảo**

*Lê Đức Tùng, Nguyễn Thanh Thủy*

*Các thành viên:*

***Tô Trọng Hiến, Nguyễn Hồng Thanh, Nguyễn Việt Phương, Nguyễn Duy Hoàng,  
Đào Quang Minh,***

*Trung tâm tính toán hiệu năng cao  
Đại học bách khoa Hà Nội*

# Nội dung trình bày

Đặt vấn đề

Mô hình quản lý tài liệu trong tổ chức ảo

Thử nghiệm đánh giá

Kết luận

- Tài liệu điện tử nằm phân tán trên mạng Internet
  - Tài nguyên động: cá nhân, tổ chức, dữ liệu
  - Yêu cầu: mềm dẻo, bảo mật, cộng tác chia sẻ
  - Giải quyết **yêu cầu** trên **tập tài nguyên động** là phức tạp
- Công nghệ tính toán phân tán hiện tại:
  - Công nghệ Internet:
    - Truyền thông và trao đổi thông tin giữa các máy tính
    - Không cung cấp khả năng cộng tác tài nguyên để tính toán
  - CORBA và Enterprise Java
    - Chia sẻ tài nguyên trong **một** tổ chức đơn
  - Dịch vụ lưu trữ (SSP) và dịch vụ ứng dụng (ASP)
    - Cho phép lấy về tài nguyên lưu trữ và tính toán từ đơn vị thứ 2
    - Môi hình kết nối là ràng buộc, cứng nhắc, không tự nhiên:
      - SSP: sử dụng VPN

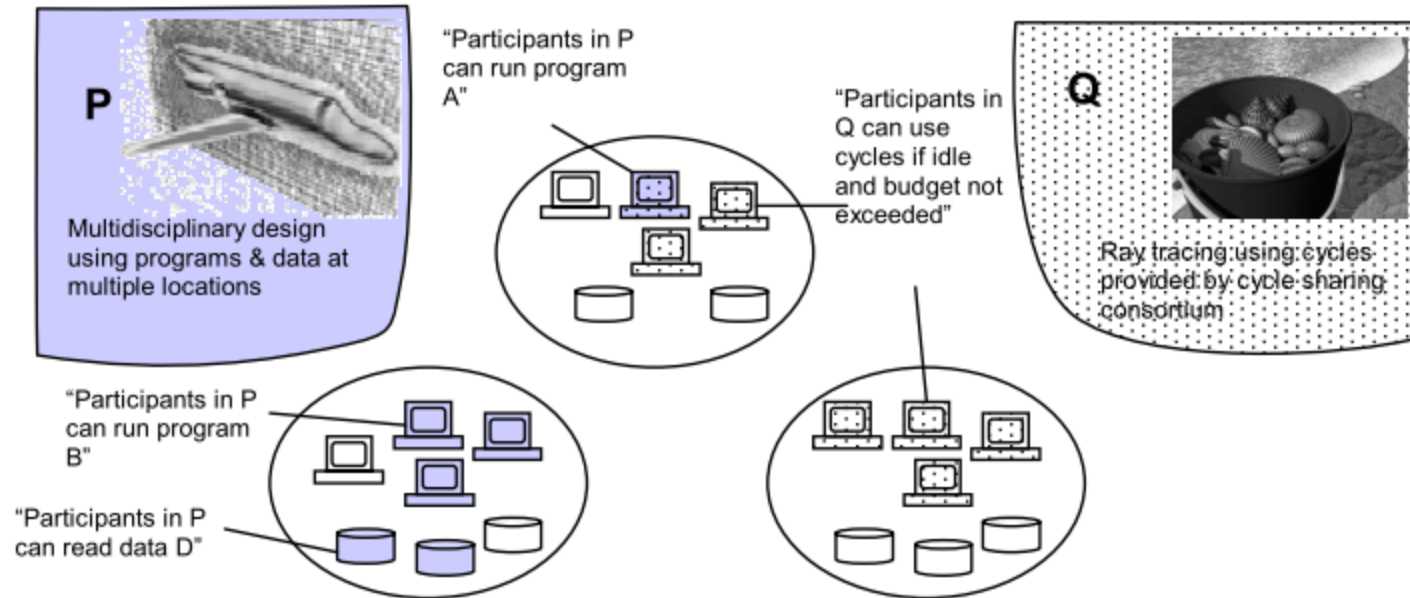
**Cần sự mềm dẻo trong quan hệ chia sẻ tài nguyên**

# Công nghệ lưới dữ liệu

- Phối hợp sử dụng tài nguyên phân tán mà không dựa trên điều khiển tập trung.
- Cung cấp hạ tầng lưới an toàn, bảo mật
- Tính dễ mở rộng của hệ thống

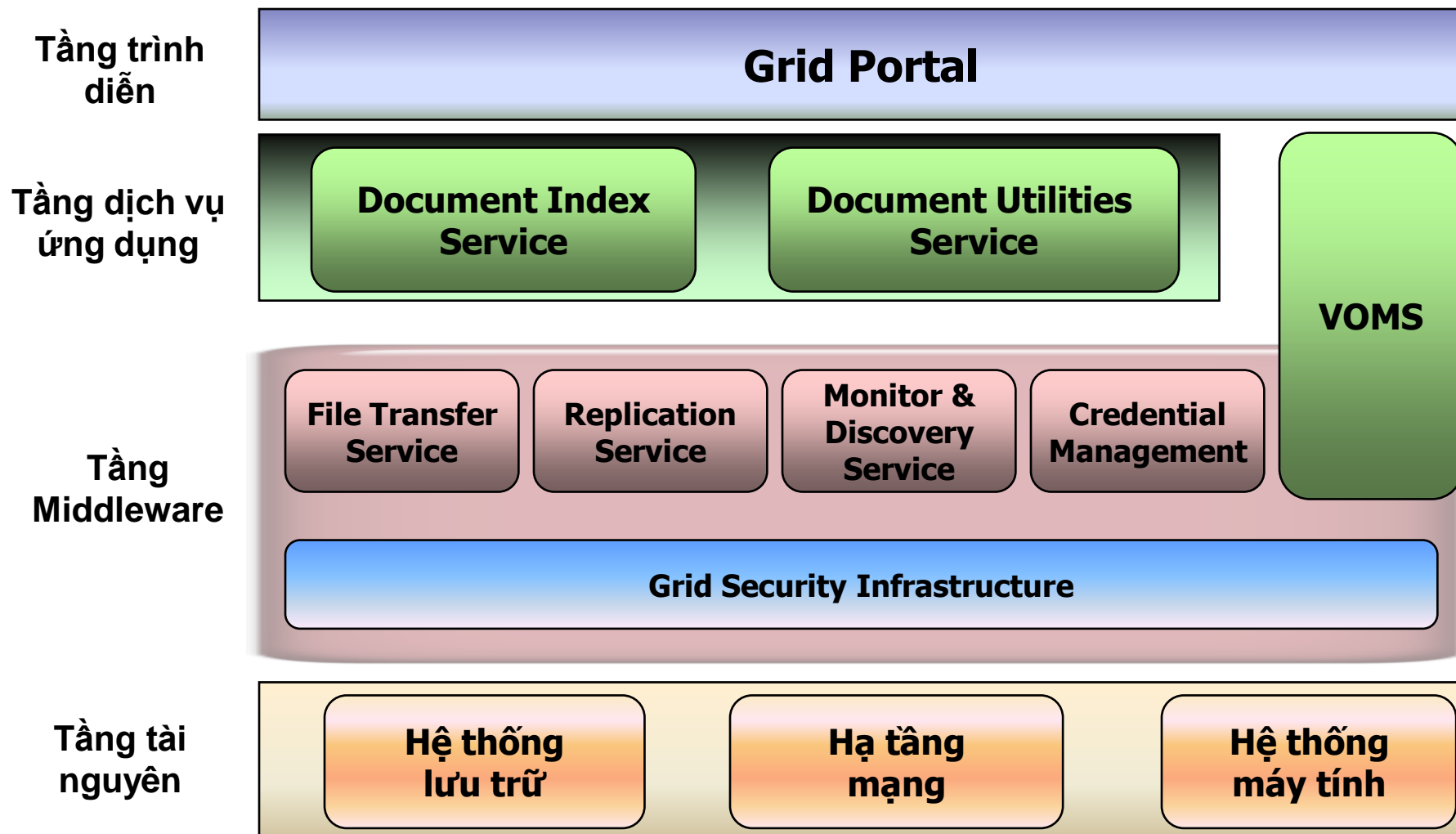


# Tổ chức ảo



- Một người dùng tham gia nhiều tổ chức ảo
- Mỗi tổ chức ảo:
  - Tài nguyên
  - Dịch vụ
- Tạo động các tổ chức ảo

# Đóng góp của bài báo



# Nội dung trình bày

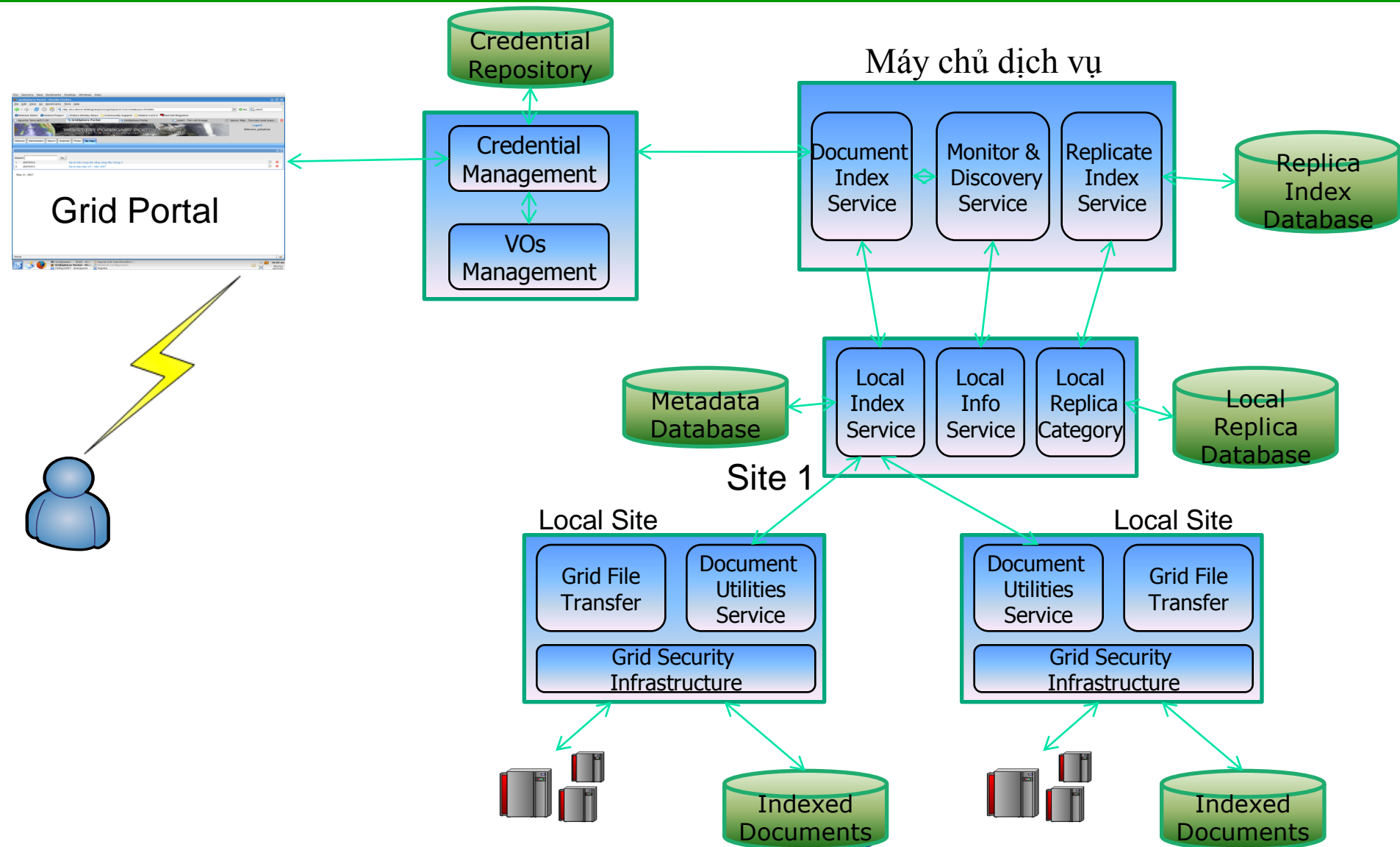
Đặt vấn đề

Mô hình quản lý tài liệu trong tổ chức ảo

Thử nghiệm đánh giá

Kết luận

# Mô hình quản lý tài liệu trong tổ chức ảo





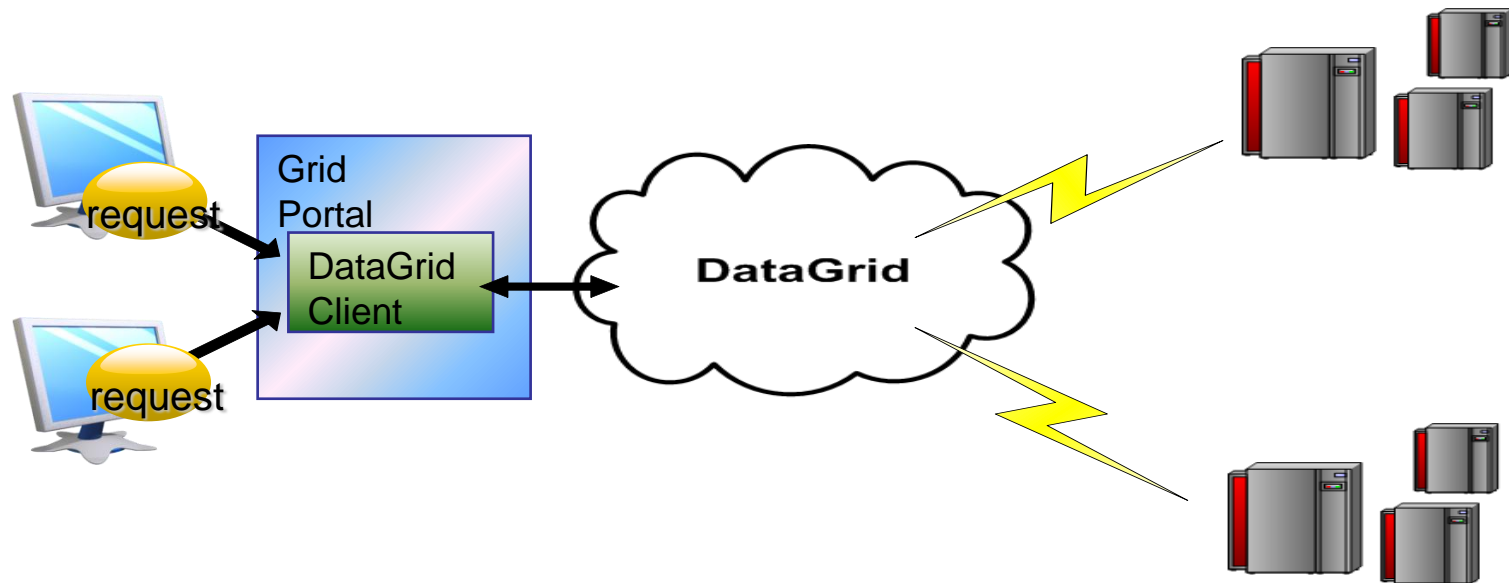
# Đảm bảo tính bảo mật cho mô hình

- Xác thực người dùng, duy trì bảo mật cho hạ tầng lưới nhưng phải đảm bảo trong suốt với người dùng
- Cung cấp các tài nguyên thông qua các dịch vụ & cho phép thiết lập các kiểm soát truy nhập
- Phân nhóm người dùng, gán vai trò, xác định quyền truy nhập đến các dịch vụ của hệ thống

Giải pháp xây dựng hệ thống theo mô hình Grid portal & ứng dụng tổ chức ảo (VO)

# Grid Portal

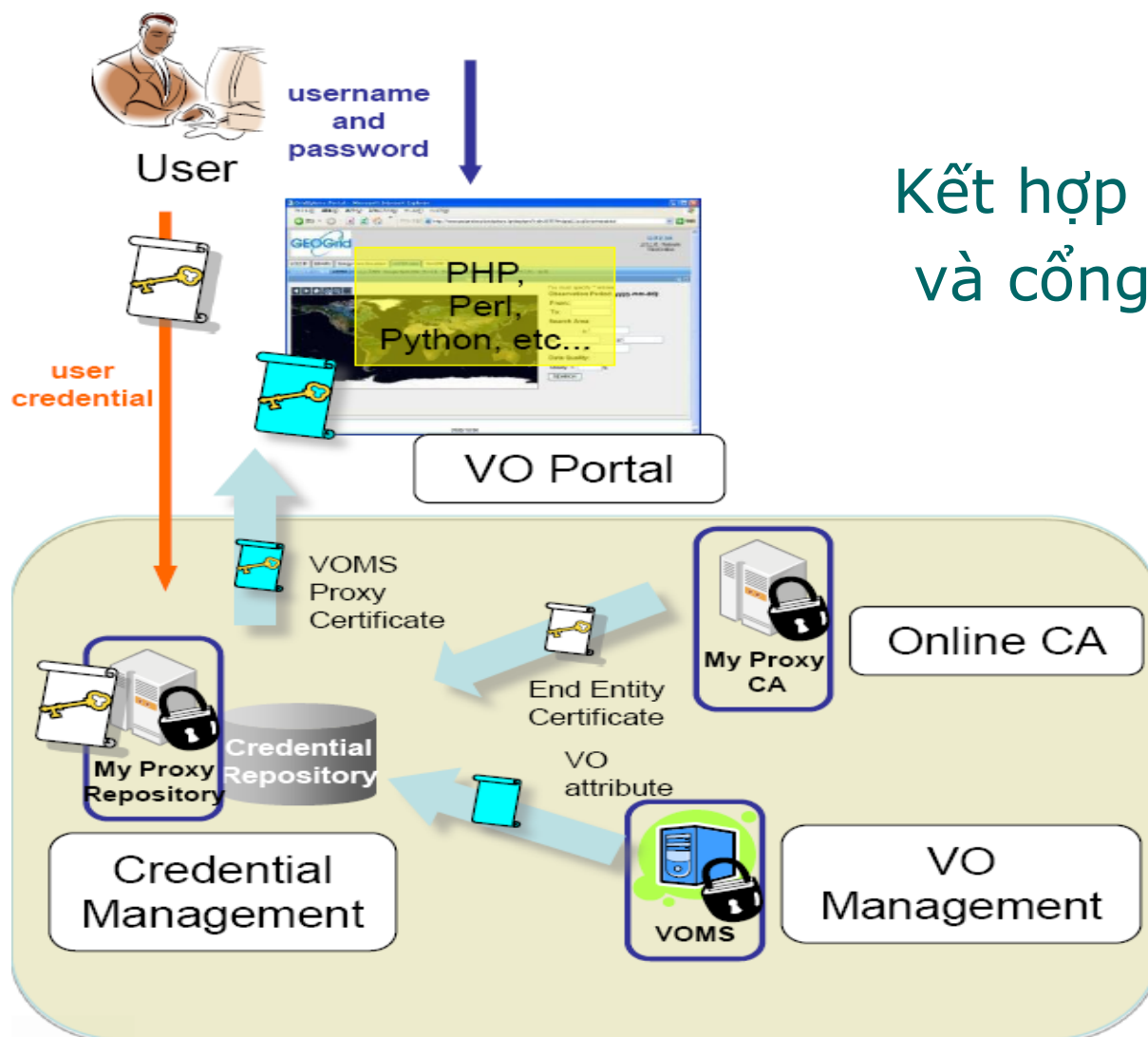
- Grid Portal: cung cấp các khả năng truy nhập tới các dịch vụ và tài nguyên lưới cho cộng đồng người dùng thông qua giao diện Web  
=> Grid portal làm trong suốt sự phức tạp của lưới tới người dùng



# VO Membership Service

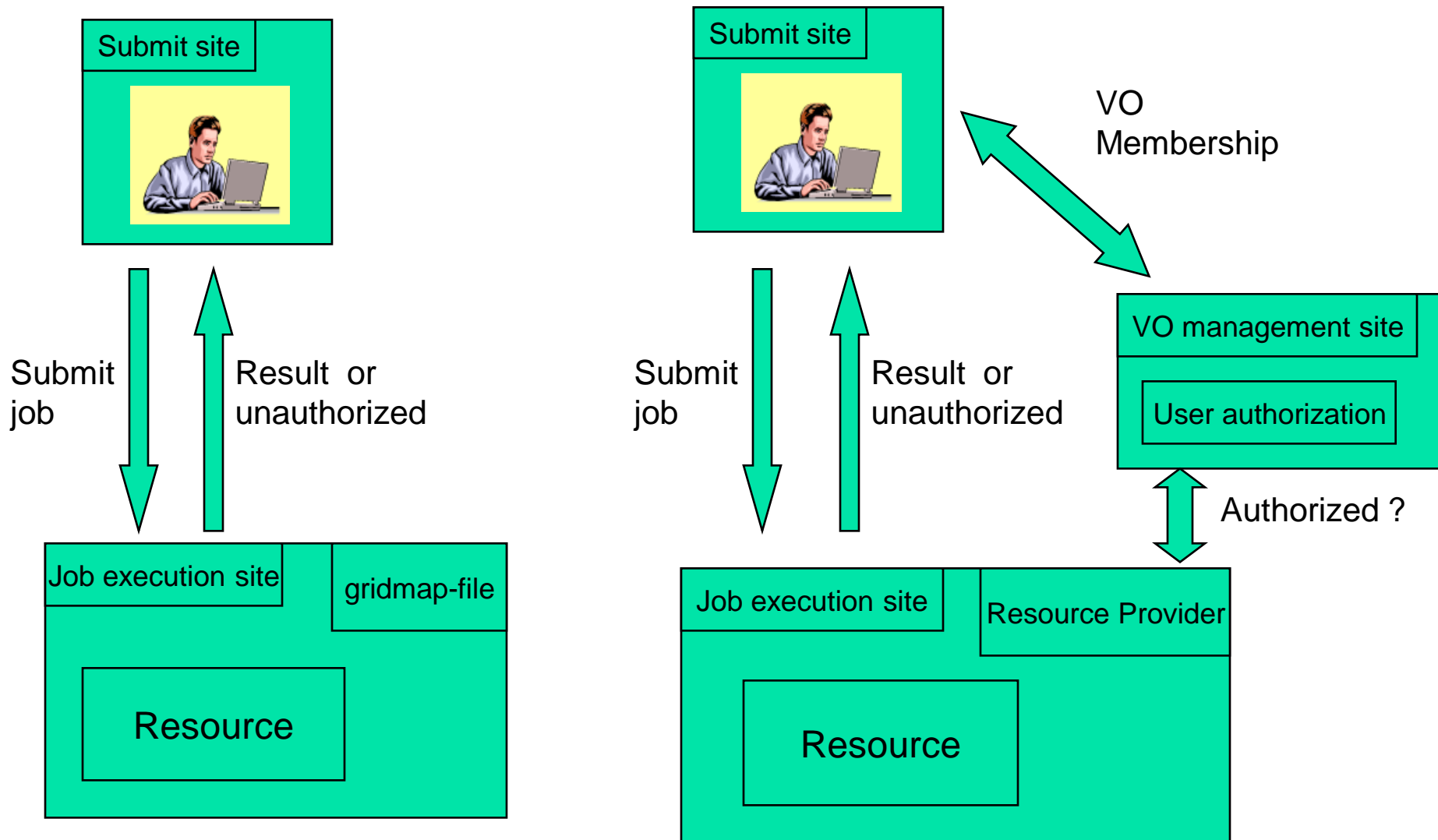
- Cung cấp các cơ chế ủy quyền dựa trên VO
- Người dùng có thể đăng ký với các VOs, gia nhập các Group & được gán các vai trò (Roles)
- Người cung cấp dịch vụ có thể cấu hình hệ thống để điều khiển truy nhập dựa trên:
  - Phạm vi Tổ chức (VO-based)
  - Nhóm người dùng (Group-based)
  - Nhóm & vai trò (Group&Role-based)

# Đảm bảo bảo mật cho mô hình (Tiếp)



Kết hợp tổ chức ảo  
và cổng thông tin

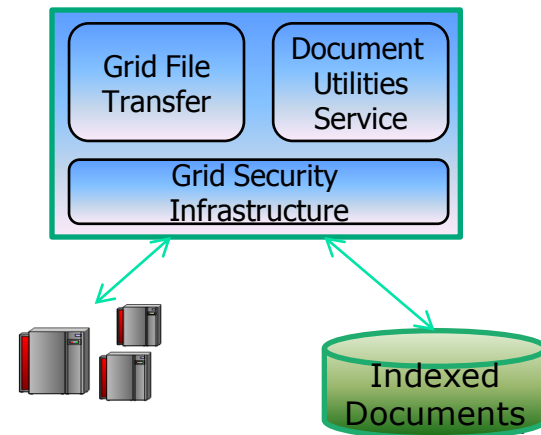
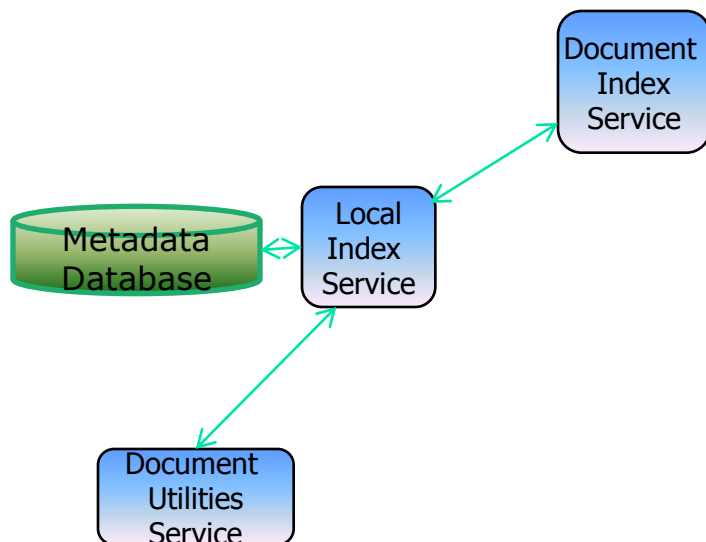
# Tích hợp VO vào Lưới dữ liệu



# Mô hình độc lập với dịch vụ

- Document Utilities Service:

- Viết theo chuẩn của Grid Service
- Liên kết với Local Index Service
- Thông tin service lưu trong Metadatabase



- Quản lý phân cấp
- Dễ dàng mở rộng

# Nội dung trình bày

Đặt vấn đề

Mô hình quản lý tài liệu trong tổ chức ảo

Thử nghiệm đánh giá

Kết luận

- Áp dụng mô hình quản lý tài liệu vào giải quyết một case study cụ thể: bài toán so khớp tài liệu trên môi trường phân tán.
  - *Bài toán*: Với mỗi tài liệu đầu vào, xác định mức độ trùng khớp (về ngữ nghĩa) so với các tài liệu trong kho hiện có
  - *Tính chất*: kho tài liệu thuộc nhiều lĩnh vực, nằm phân tán trên nhiều site, mỗi site có các chính sách truy cập & quản lý người dùng riêng.
  - Giải thuật so khớp được sử dụng: PLSA (Probabilistic Latent Semantic Analysis)
- Đánh giá hiệu quả của việc áp dụng mô hình vào giải quyết một vấn đề cụ thể



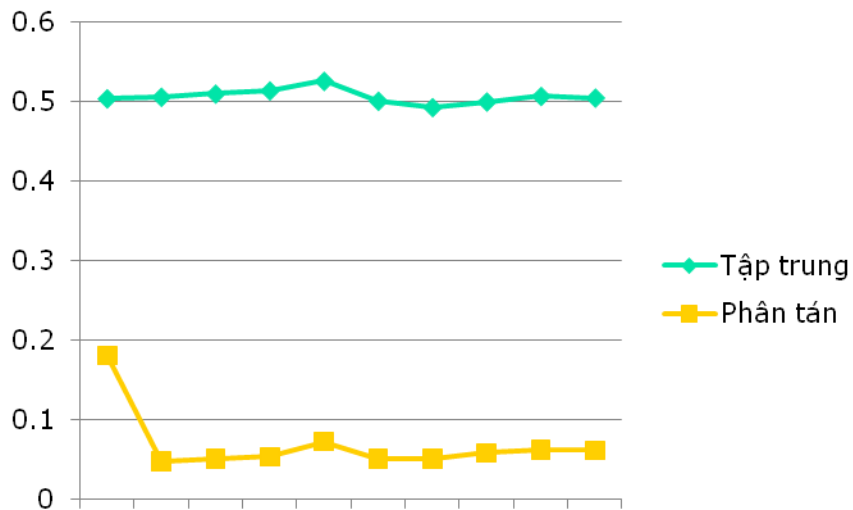
# Mô hình thử nghiệm

- Xây dựng lưới dữ liệu cho hệ thống, mỗi trường đại học, tổ chức nghiên cứu là một nút lưới
  - Trên mỗi nút lưới cài đặt các dịch vụ tìm kiếm, upload & so khớp tài liệu
- Sử dụng mô hình VO, phân nhóm người dùng theo các lĩnh vực & vai trò:
  - Nhóm lĩnh vực: IT, Literature, Math, Physic, Chemical...
  - Vai trò trong mỗi nhóm lĩnh vực: Nhà nghiên cứu, người dùng thường.

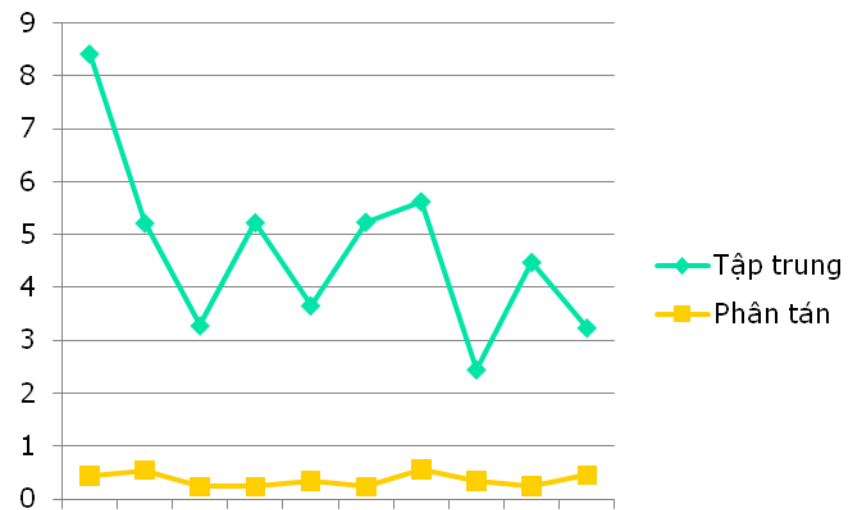
Vai trò	Tìm kiếm tài liệu	Upload tài liệu	So khớp tài liệu
Nhà nghiên cứu	x	x	x
Người dùng thường	x	o	o

# Kết quả thử nghiệm

- Hơn 2000 tài liệu được phân theo 10 nhóm lĩnh vực
- Kết quả xử lý được so sánh với việc xử lý tập trung



Thời gian xử lý (giây)



Số phần trăm lỗi (%)

# Nội dung trình bày

Đặt vấn đề

Mô hình quản lý tài liệu trong tổ chức ảo

Thử nghiệm đánh giá

Kết luận

- Kết luận:

- Mô hình giải quyết được bài toán dữ liệu lớn phân tán
- Mô hình là bảo mật
- Mô hình khả mở và mềm dẻo

- Hướng phát triển:

- Xây dựng giải pháp sao lưu máy chủ dịch vụ
- Phân cấp mô hình tổ chức ảo
  - Phân cấp logic
  - Phân cấp vật lý: mở rộng quản lý VO dùng nhiều máy tính theo mô hình phân cấp

Xin cảm ơn!