Chương 8:
Bảo mật các dịch vụ cơ sở dữ liệu thuê ngoài
(Outsourcing Database Services Security)

Khoa Khoa học và Kỹ thuật Máy tính
Đại học Bách Khoa Tp.HCM

Nội dung

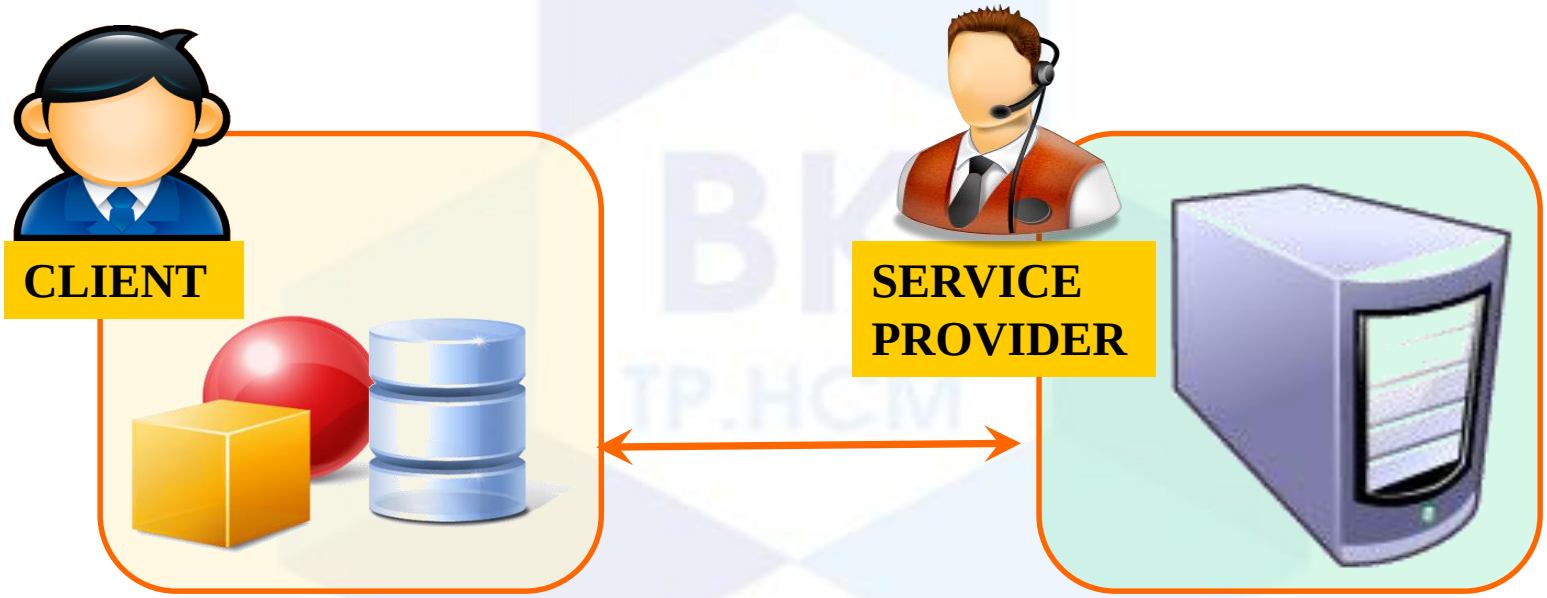
1. Giới thiệu dịch vụ CSDL thuê ngoài
2. Bảo mật với dịch vụ CSDL thuê ngoài
3. Mô hình nhà cung cấp dịch vụ

Giới thiệu về dịch vụ CSDL thuê ngoài

- Mô hình truyền thống:
  - Khách (Client) sở hữu và tự quản lý server CSDL của mình
  - Lợi ích: Có toàn quyền trên server
  - Bất lợi: chi phí ban đầu và chi phí bảo trì

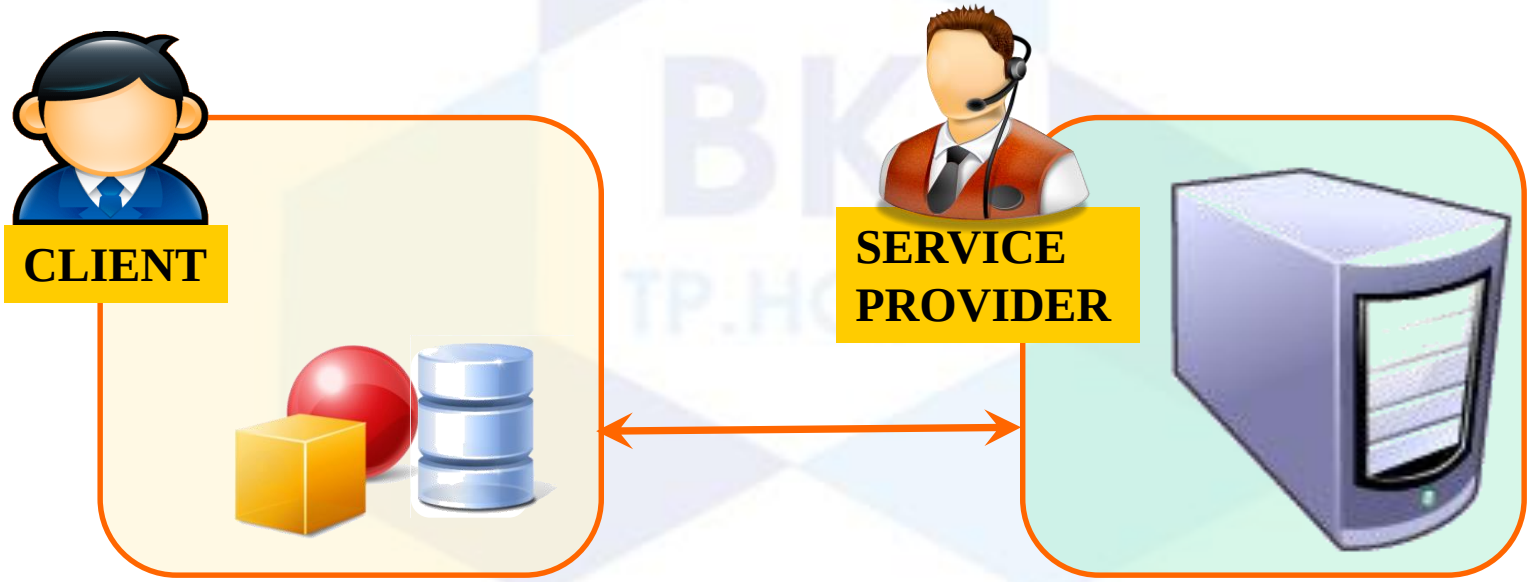
Giới thiệu về dịch vụ CSDL thuê ngoài

- Dịch vụ CSDL thuê ngoài (Outsourcing Database Services - OBDS)
  - Client thuê một nhà cung cấp dịch vụ (service provider) bên ngoài hỗ trợ việc lưu trữ và quản lý dữ liệu của mình



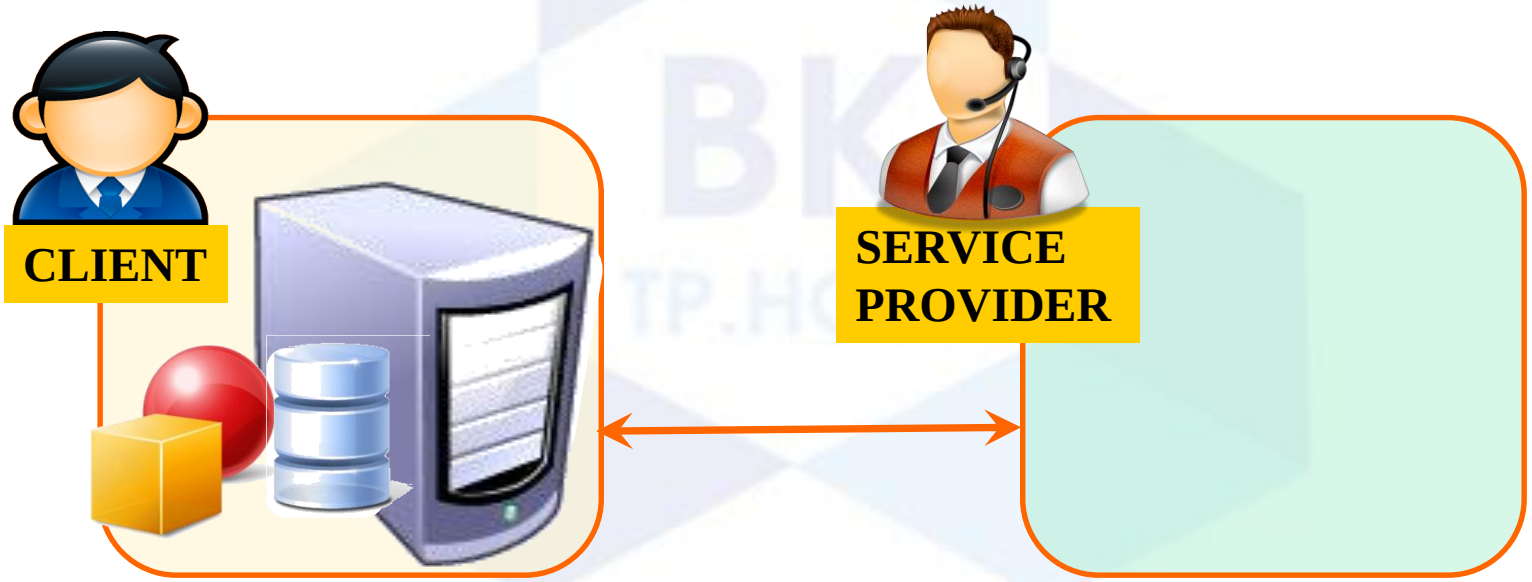
Giới thiệu về dịch vụ CSDL thuê ngoài

- Có 2 loại: hosting service và housing service
- **Hosting service:**
  - Nhà cung cấp dịch vụ có server CSDL
  - Client thuê một phần không gian trong server CSDL của nhà cung cấp dịch vụ để lưu trữ dữ liệu



Giới thiệu về dịch vụ CSDL thuê ngoài

- **Housing service:**
  - Nhà cung cấp dịch vụ có cơ sở vật chất tốt để bảo quản server CSDL
  - Client thuê một chỗ để đặt server CSDL



Ưu điểm của mô hình CSDL thuê ngoài

- Tập trung vào công việc kinh doanh
- Tiết kiệm chi phí:
  - Chi phí ban đầu: phần cứng, phần mềm, cơ sở vật chất và nhân viên kỹ thuật
  - Chi phí bảo trì
- Tiết kiệm thời gian thiết lập hệ thống CSDL
- Nhanh chóng có được những tài nguyên/dịch vụ không có sẵn
- Sử dụng dịch vụ từ những nhà cung cấp dịch vụ chuyên nghiệp
- Môi trường lưu trữ và quản lý CSDL ổn định, ít thay đổi

Khuyết điểm của mô hình CSDL thuê ngoài

- Thời gian đáp ứng chậm (turnaround time)
- Các chi phí ẩn khi yêu cầu những dịch vụ cao cấp
- Chất lượng dịch vụ
- Vấn đề giao tiếp giữa client và nhà cung cấp dịch vụ
- Khó can thiệp sâu khi giải quyết vấn đề
- Không thể kiểm soát hết việc điều khiển truy cập

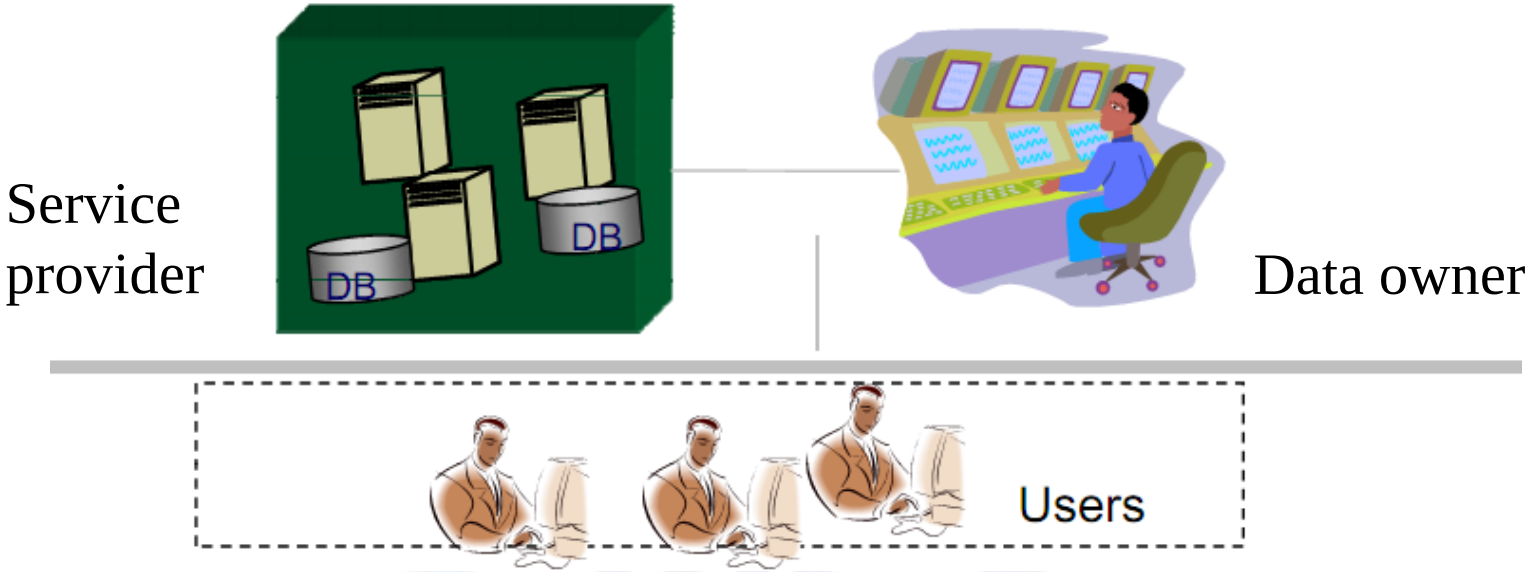
Một số nhà cung cấp dịch vụ thuê ngoài

- IBM
- Oracle
- EDS
- DbaDirect
- Ntirety
- Pythian
- TCS
- Satyam
- Wipro

Nội dung

1. Giới thiệu dịch vụ CSDL thuê ngoài
2. Bảo mật với dịch vụ CSDL thuê ngoài
3. Mô hình nhà cung cấp dịch vụ

Mô hình CSDL thuê ngoài



- **Chủ dữ liệu (Data owner):** cơ quan/tổ chức tạo ra dữ liệu
- **User (hoặc client):** khách hàng đưa ra yêu cầu (câu truy vấn) dữ liệu cho hệ thống
- **Nhà cung cấp dịch vụ:** tổ chức nhận dữ liệu từ chủ dữ liệu và phân phối dữ liệu đến user khi có yêu cầu

Các vấn đề bảo mật và giải pháp

- Bảo mật dữ liệu (Data confidentiality)
- Tính riêng tư của khách hàng (User privacy)
- Tính riêng tư của dữ liệu (Data privacy)
- Tính đảm bảo truy vấn (Query Assurance)

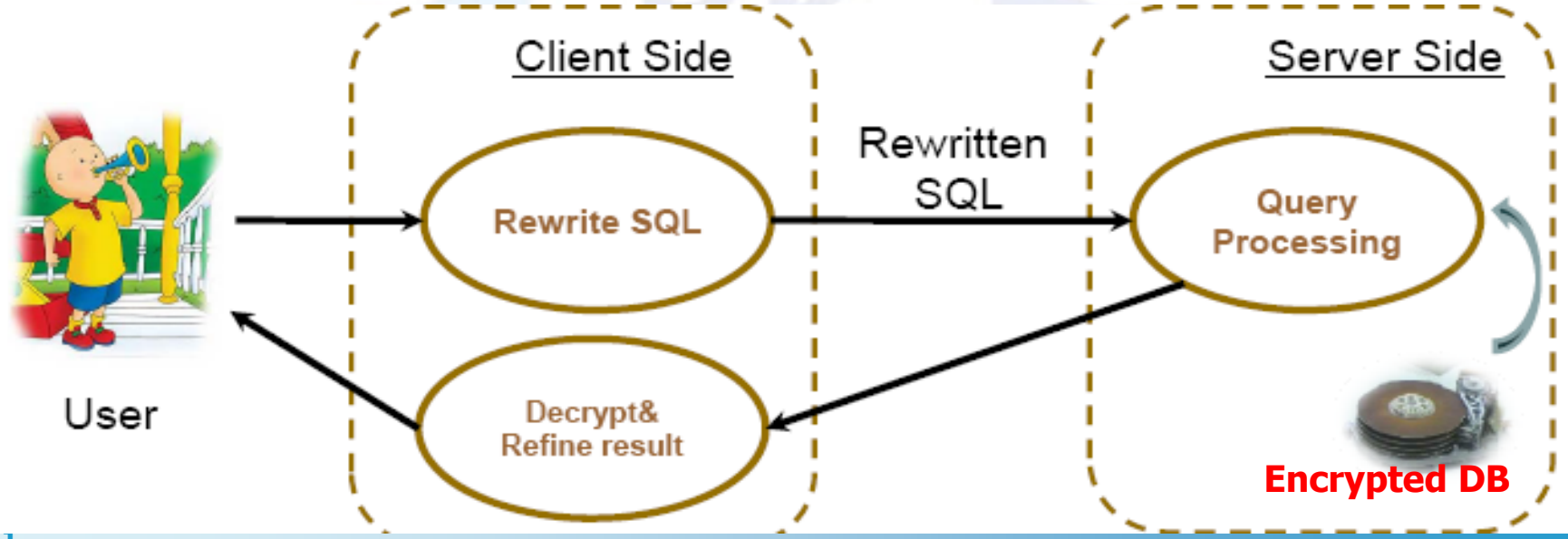
Bảo mật dữ liệu

- **Bảo mật dữ liệu (Data confidentiality):** những người ngoài và kể cả nhà cung cấp dịch vụ cũng không được phép xem dữ liệu thuê ngoài của khách hàng
- Vấn đề: nhà cung cấp dịch vụ không tin cậy
- Giải pháp:
  - Mã hóa ở phía server
  - Giải mã ở phía client
  - Viết lại câu SQL

  → Hacıgümüs Scheme

Bảo mật dữ liệu

- Hacıgümüs Scheme
  - Hàm phân vùng (Partition function)
  - Hàm xác định (Identification function)
  - Tách câu truy vấn: phía server và phía client
- Chủ dữ liệu và user giữ bí mật về hàm phân vùng và hàm xác định

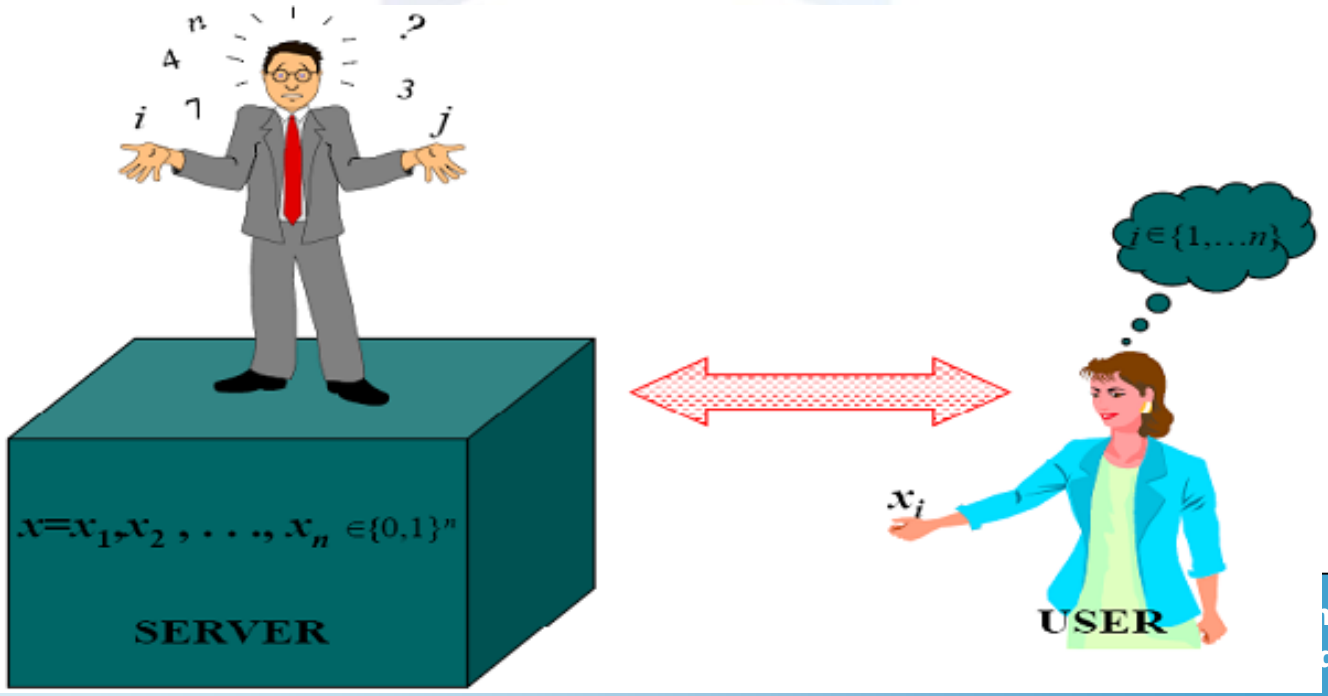


Tính riêng tư của khách hàng

- **Tính riêng tư của khách hàng (User privacy):** Khách hàng không muốn nhà cung cấp dịch vụ và chủ dữ liệu biết điều mà họ đang quan tâm (câu truy vấn và kết quả trả về)
- Giải pháp:
  - Giao thức PIR (Private Information Retrieval)
  - Giao thức RIR (Repudiative Information Retrieval)
  - Giao thức PIS (Private Information Storage)

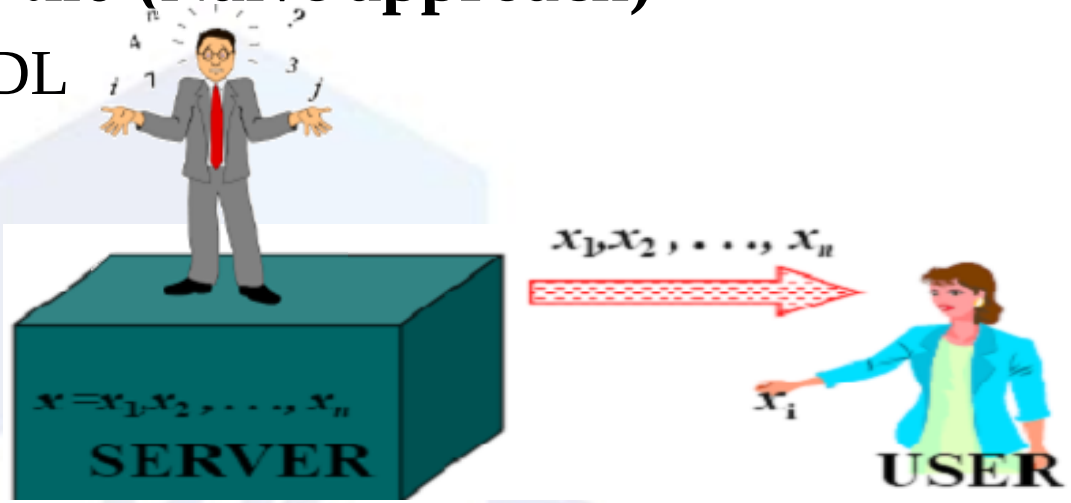
Giao thức PIR (Private Information Retrieval)

- Do Chor và các cộng sự giới thiệu
- Giao thức PIR cho phép khách hàng truy cập CSDL mà không bị lộ câu truy vấn và kết quả
- Khách hàng truy vấn *record thứ i* trong số *n record* từ CSDL mà ***không bị lộ giá trị i*** với server



Giao thức PIR (Private Information Retrieval)

- **Cách tiếp cận ngây thơ (Naïve approach)**
  - Tải về toàn bộ CSDL
  - Chi phí lớn



- **Kỹ thuật nặc danh (Anonymization techniques)**
  - User gửi câu truy vấn nặc danh và cũng nặc danh khi nhận kết quả trả về
  - Có thể dùng kỹ thuật khai phá dữ liệu (data mining) trên các CSDL thống kê và kết hợp với các yếu tố ngữ cảnh khác để phá vỡ tính riêng khách hàng của kỹ thuật nặc danh.

Giao thức RIR (Repudiative Information Retrieval)

- Giống với giao thức PIR
- Cho phép từ chối các truy cập thông tin nào.
- Giảm chi phí so với PIR
- *Không có nhiều thông tin* về danh định (identity) của record trong câu query bị lộ

  → Dù bị lộ một ít thông tin, nhưng người quan sát bên ngoài không cũng có cơ sở để quyết định chắc chắn rằng khách hàng có truy vấn record thứ *j* nào đó hay không.
- Dùng bộ đồng xử lý an toàn (secure coprocessor) hoặc bên thứ 3 đáng tin cậy

Giao thức PIS (Private Information Storage)

- Giống với giao thức PIR
- Cho phép từ chối các truy cập thông tin nào.
- Hỗ trợ thao tác ghi bí mật

Tính riêng tư của dữ liệu (Data privacy)

- **Tính riêng tư của dữ liệu (Data privacy):** Khách hàng không được phép nhận nhiều thông tin hơn so với câu truy vấn của họ.
- Giải pháp
  - Điều khiển truy cập
  - Giao thức SPIR (Symmetrically Private Information Retrieval)
  - Untrusted 3rd parties, Secure coprocessors

Xác thực và toàn vẹn dữ liệu

- ***Xác thực và toàn vẹn dữ liệu (Authentication and data integrity)***: Khách hàng phải được đảm bảo rằng dữ liệu trả về từ server là đúng và không bị thay đổi
- Giải pháp: Đảm bảo truy vấn an toàn (Query assurance)
  - Truy vấn đúng (Query correctness): kết quả câu truy vấn phải có trong CSDL
  - Truy vấn đủ (Query completeness): không có kết quả nào bị bỏ bớt trong nội dung trả về cho khách hàng
  - Truy vấn mới (Query freshness): kết quả trả về phải là kết quả mới nhất/hiện tại của CSDL

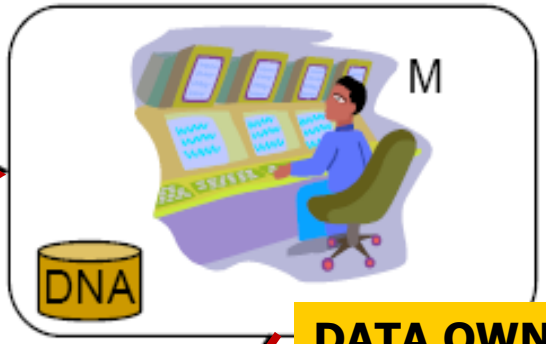
Nội dung

1. Giới thiệu dịch vụ CSDL thuê ngoài
2. Bảo mật với dịch vụ CSDL thuê ngoài
3. Mô hình nhà cung cấp dịch vụ

Một mô hình CSDL thuê ngoài thực tế

USER

Trả tiền để lấy thông tin



DATA OWNER

Gửi thông tin về DNA của user

User privacy
Data privacy

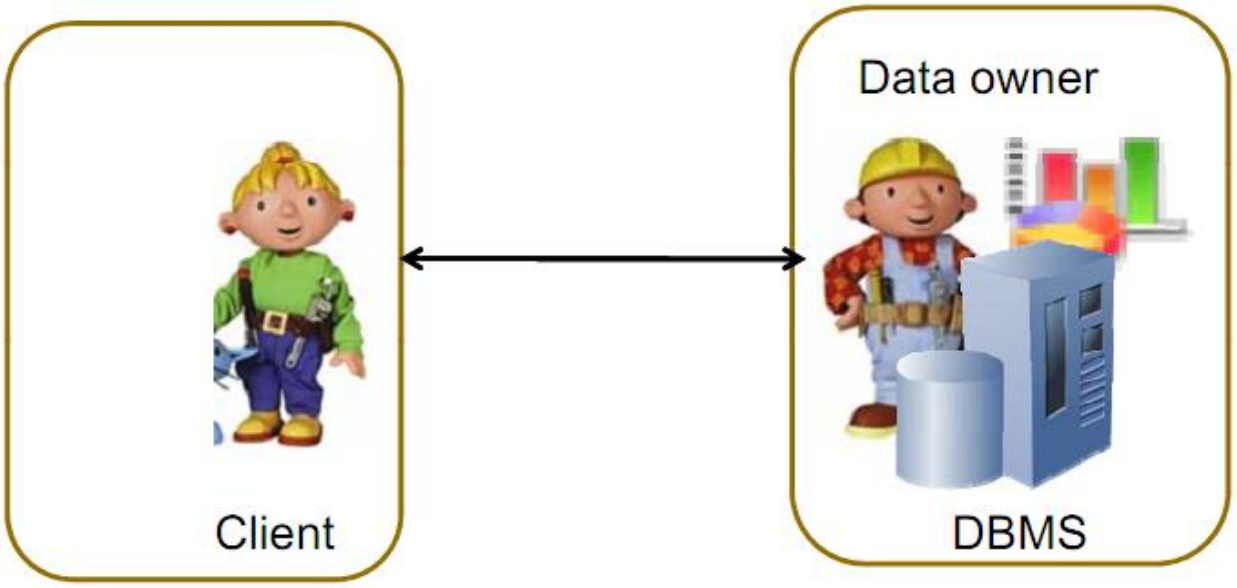
SERVICE PROVIDER

Lưu trữ CSDL DNA

Data confidentiality

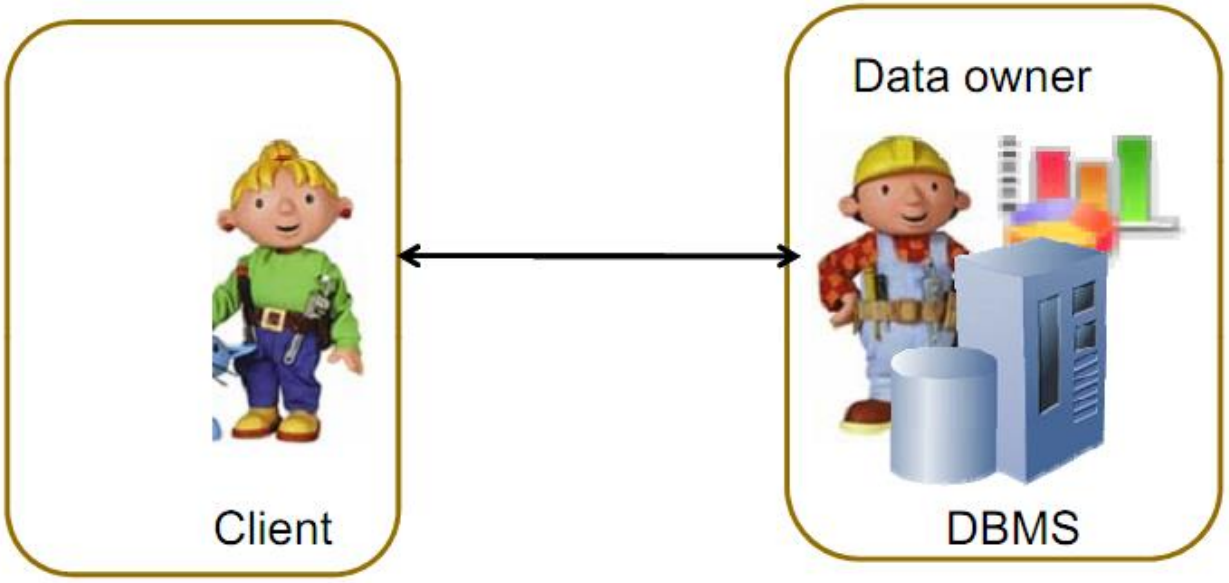
Mô hình nhà cung cấp dịch vụ

- **Mô hình UP-DP (User Privacy – Data Privacy)**
  - Chủ dữ liệu ≡ Nhà cung cấp dịch vụ
  - Nhà cung cấp dịch vụ quan tâm đến **tính riêng tư của dữ liệu.**
  - Khách hàng quan tâm đến **tính riêng tư của khách hàng.**



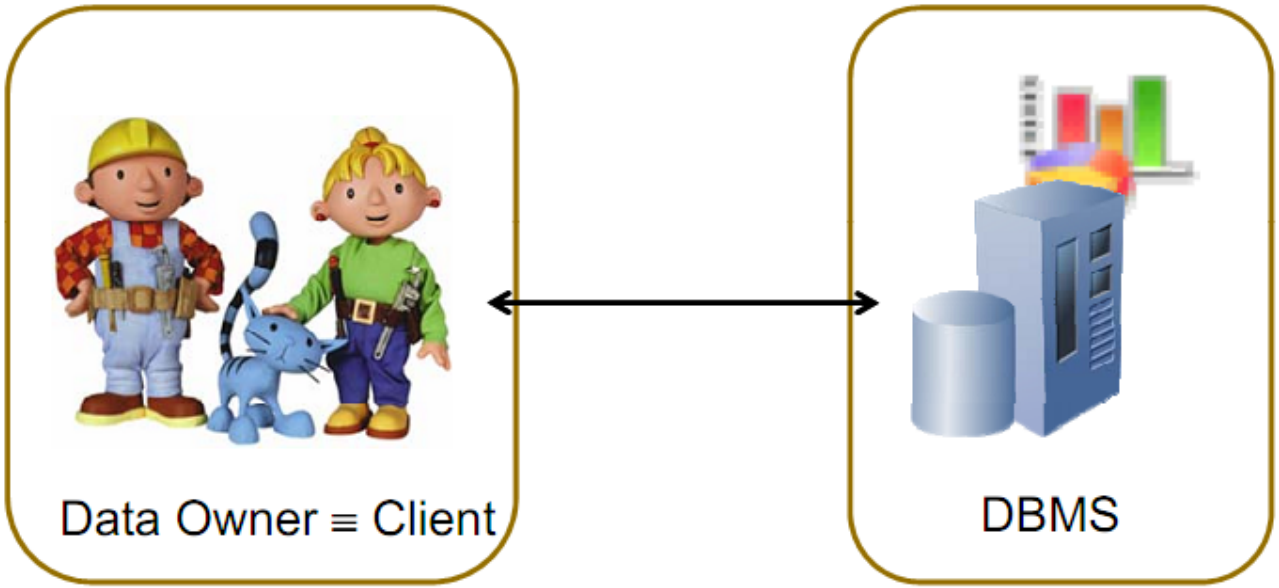
Mô hình nhà cung cấp dịch vụ

- **Mô hình UP-nDP (User Privacy – *not* Data Privacy)**
  - Chủ dữ liệu ≡ Nhà cung cấp dịch vụ
  - Dữ liệu công khai (public data) → Nhà cung cấp dịch vụ **không** quan tâm đến **tính riêng tư của dữ liệu.**
  - Khách hàng quan tâm đến **tính riêng tư của khách hàng.**



Mô hình nhà cung cấp dịch vụ

- **Mô hình DC-UP (Data Confidentiality– User Privacy)**
  - Khách ≡ Chủ dữ liệu
  - Khách hàng (chủ dữ liệu) quan tâm đến **tính bí mật của dữ liệu** và **tính riêng tư của khách hàng**



Mô hình nhà cung cấp dịch vụ

- **Mô hình DC-UP-DP (Data Confidentiality– User Privacy – Data Privacy)**
  - Khách hàng <> Chủ dữ liệu <> Nhà cung cấp dịch vụ
  - Khách hàng quan tâm đến **tính riêng tư của khách hàng**
  - Chủ dữ liệu quan tâm đến **tính bí mật của dữ liệu** và **tính riêng tư của dữ liệu**

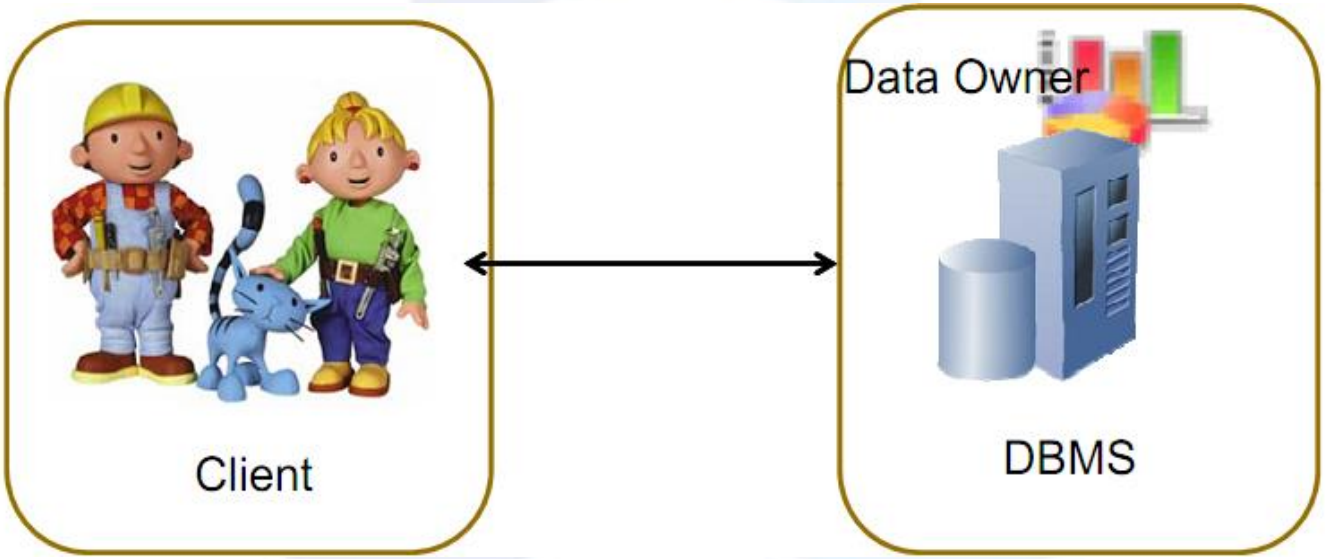
Nội dung

1. Giới thiệu dịch vụ CSDL thuê ngoài
2. Bảo mật với dịch vụ CSDL thuê ngoài
3. Mô hình nhà cung cấp dịch vụ


Question ?

Mô hình nhà cung cấp dịch vụ

- **Mô hình UP-DP (User Privacy– Data Privacy)**
  - Khách hàng ≡ Chủ dữ liệu
  - Chủ dữ liệu quan tâm đến **tính bí mật của dữ liệu**
  - Khách hàng quan tâm đến **tính riêng tư của khách hàng**



Mô hình nhà cung cấp dịch vụ

- **Mô hình UP-nDP (User Privacy – *not* Data Privacy)**
  - Nhà cung cấp dịch vụ ≡ Chủ dữ liệu
  - Dữ liệu công khai (public data) → Nhà cung cấp dịch vụ **không** quan tâm đến **tính riêng tư của dữ liệu.**
  - Khách hàng quan tâm đến **tính riêng tư của khách hàng**