ĐẠI HỌC THỦY SẢN
KHOA CÔNG NGHỆ THÔNG TIN

CƠ SỞ DỮ LIỆU PHÂN TÁN
(Distributed database)

Ng Duc Thuan

ThS. Ng ĐứcThuần
BM Hệ Thống Thông Tin

Mùa Xuân 2006

TÀI LIỆU THAM KHẢO

- [1] M. Tamer Ozsu and Patrick Valduriez, "Principles of Distributed Database Systems," Second Edition, Prentice Hall 1999.
- [2] Stefano ceri, Giuseppe Pelagatti :
  Distributed Databases Principles & Systems.
  McGraw-Hill International Editions. 1984
- [3] P. Bernstein, E. Newcomer, "Principles of Transaction Processing,"
  Morgan Kaufmann, 1997.
- [4]. Đỗ Trung Tuấn :
  Cơ sở Dữ liệu. NXB. Giáo dục, 1997
- [5] Ng Bá Tường
  Nhập môn CSDL phân tán, NXBKH&KT 2004
- [6] Slide bài giảng của Anand Rajaraman (http://www.stanford.edu/class/cs347)
- [7] Bài giảng Cơ sở Dữ liệu II - của các giảng viên ĐHBK Hà Nội

Ng Duc Thuan

CHƯƠNG I : TỔNG QUAN CSDLPT

- *CSDLPT là tập CSDL có quan hệ logic với nhau và được trãi trên 1 mạng máy tính. Mỗi trạm của mạng có khả năng xử lý tự quản và có thể thực hiện các ứng dụng cục bộ, mỗi một trạm cũng có thể tham gia vào ít nhất 1 ứng dụng toàn cục, có yêu cầu truy xuất dữ liệu tại nhiều trạm.*

Ng Duc Thuan

CHƯƠNG I : TỔNG QUAN CSDLPT

Chú ý: có thể dữ liệu được phân tán (vật lý) trên nhiều máy. Tuy nhiên, nếu việc truy xuất dữ liệu chỉ tại duy nhất một máy tích hợp dữ liệu của các máy trên. Trường hợp này, không thuộc loại hệ thống có CSDL phân tán được xem xét.

Ng Duc Thuan

CHƯƠNG I : TỔNG QUAN CSDLPT

- Tại sao cần CSDL-PT?
- Ví dụ: Công ty Big Corp. Có văn phòng tại London, New York, và Hong Kong.

Dữ liệu Employee :

- EMP(ENO, NAME, TITLE, SALARY, …)

Chúng ta nên đặt dữ liệu EMP này ở đâu?

Ng Duc Thuan

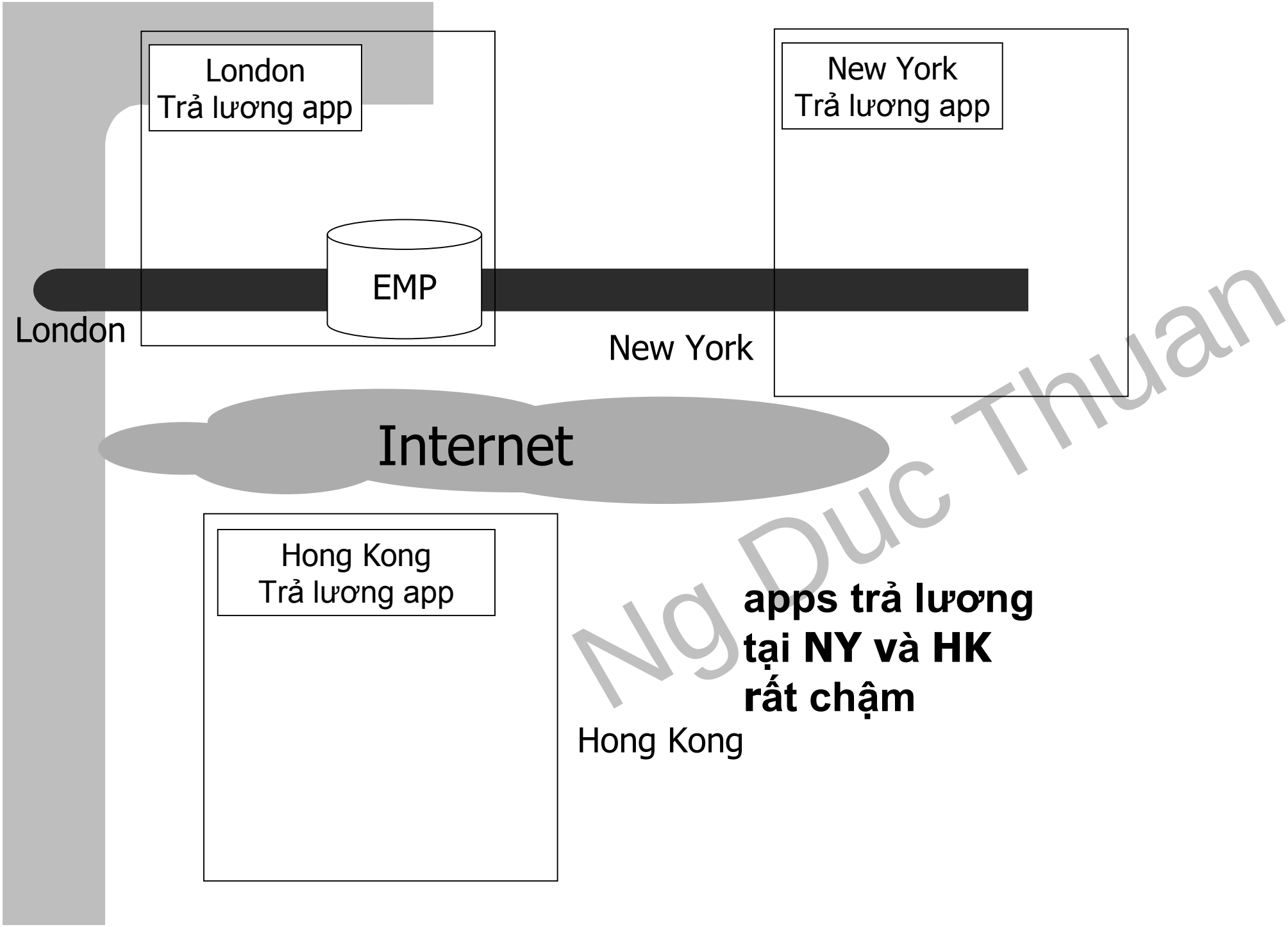
Big Corp. Dữ liệu mẫu

- Hầu như, dữ liệu nhân viên được quản lý tại văn phòng nơi mà nhân viên làm

  Ví dụ: Lương, phụ cấp, tiền thuê..

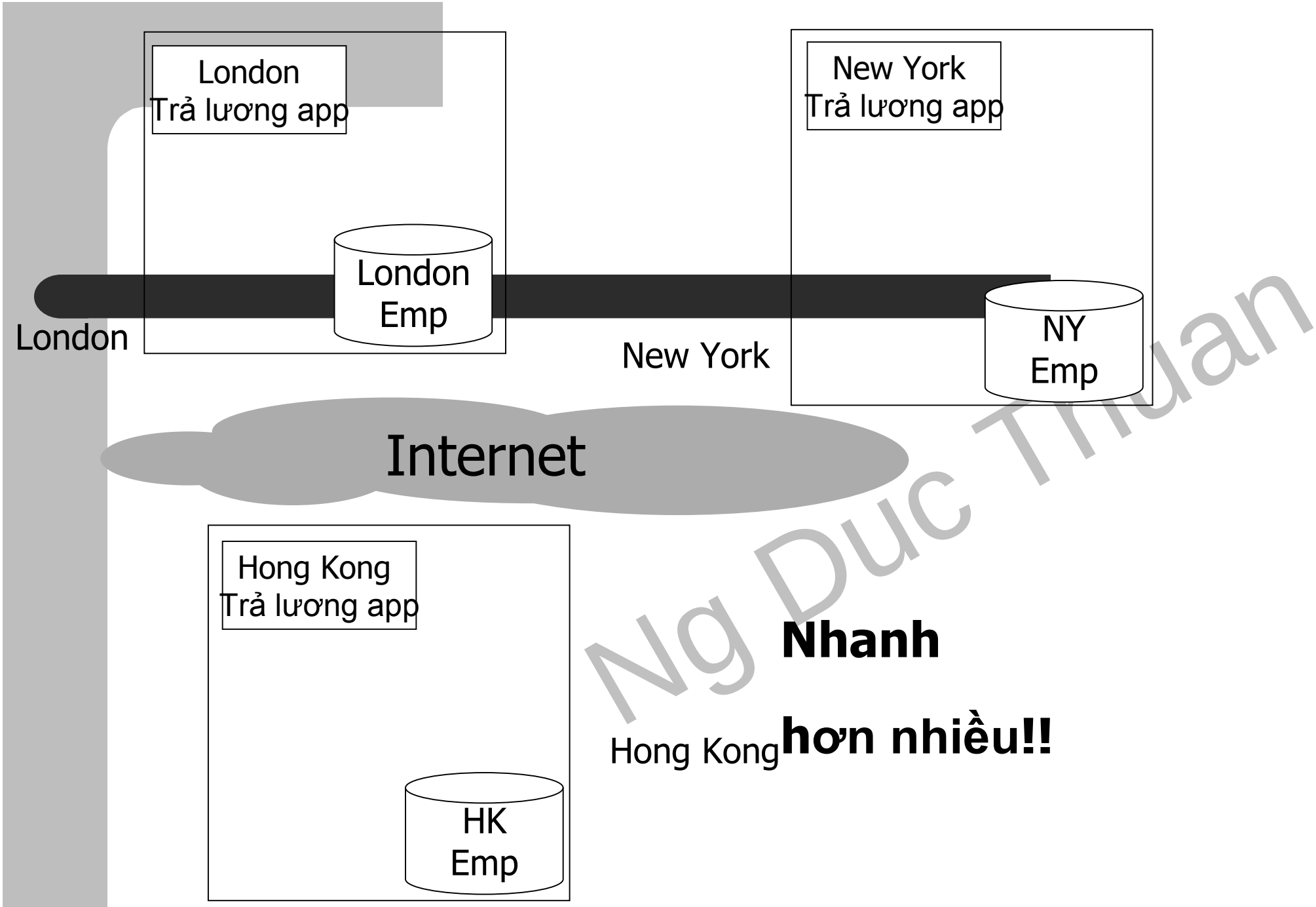
- Định kỳ, Big Corp cần hợp nhất truy xuất đến dữ liệu nhân viên
  - Ví dụ: Tính tông số lương cần trả cho bảng cân đối lương
  - Ví dụ: Thưởng thêm hàng năm phụ thuộc lợi ích mạng toàn cục

Ng Duc Thuan

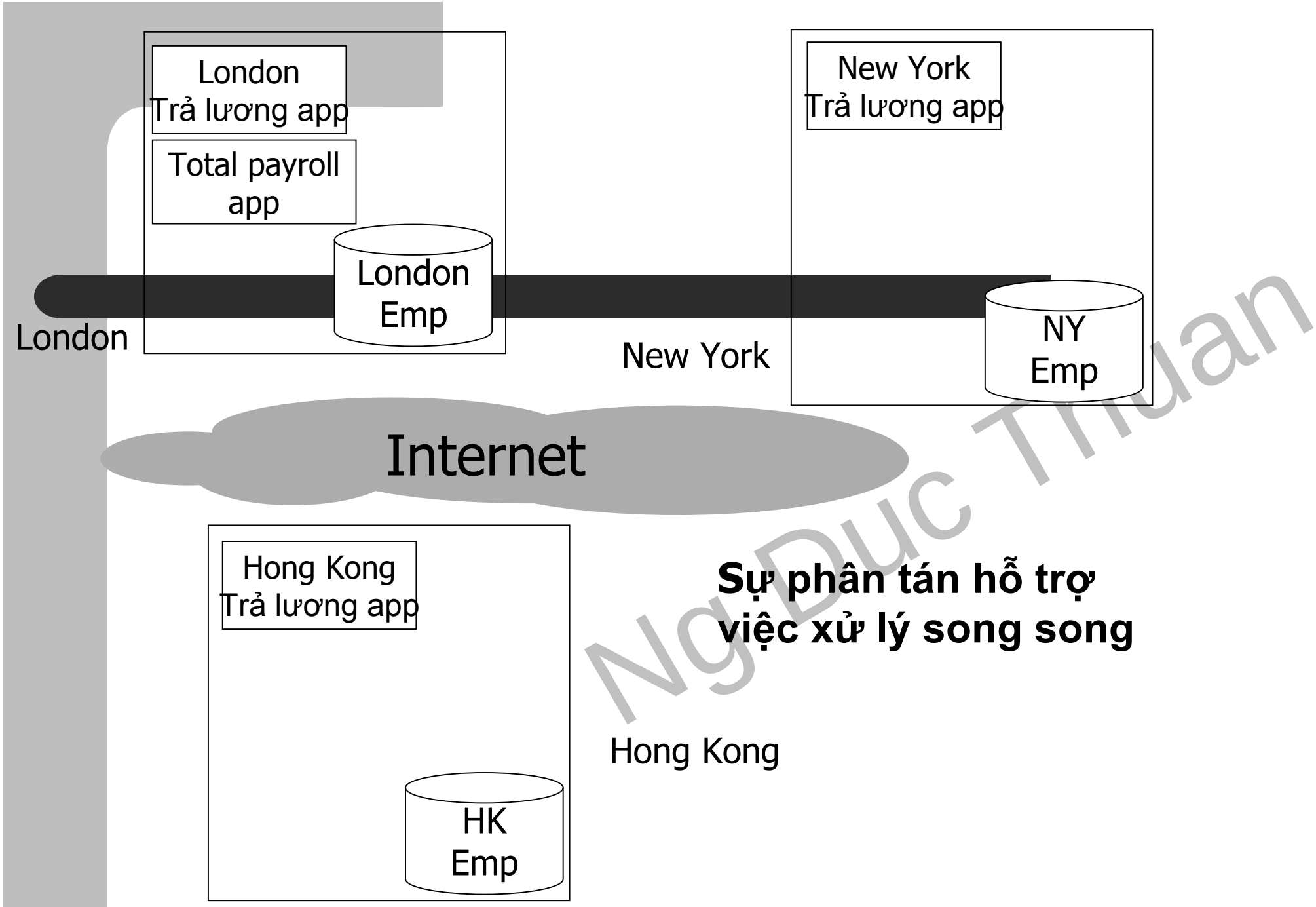


apps trả lương tại NY và HK rất chậm

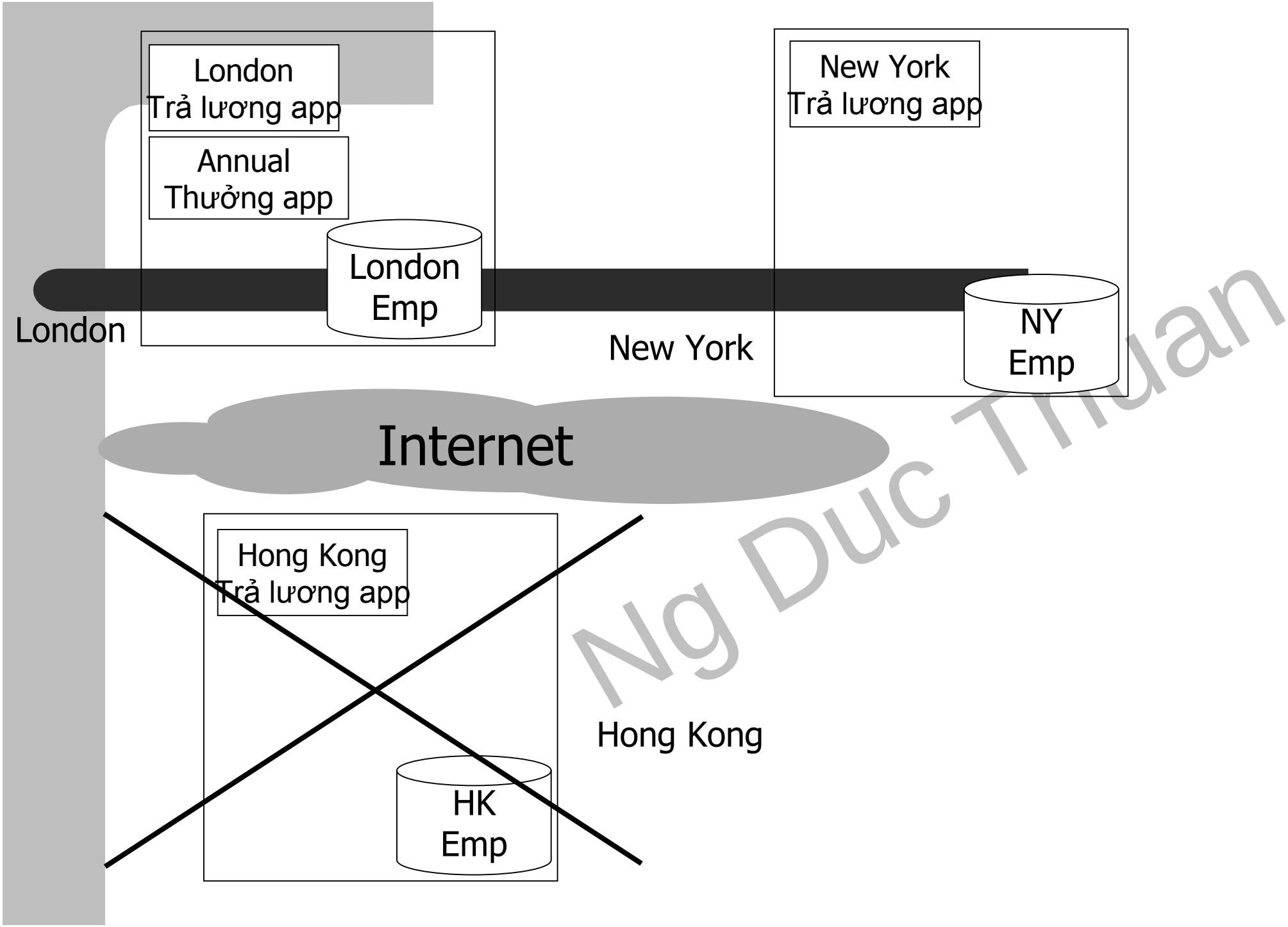
Ng Duc Thuan


London
Trả lương app
London
Emp
New York
Trả lương app
NY
Emp
London
New York
Internet
Hong Kong
Trả lương app
HK
Emp
Nhanh
Hong Kong hơn nhiều!!

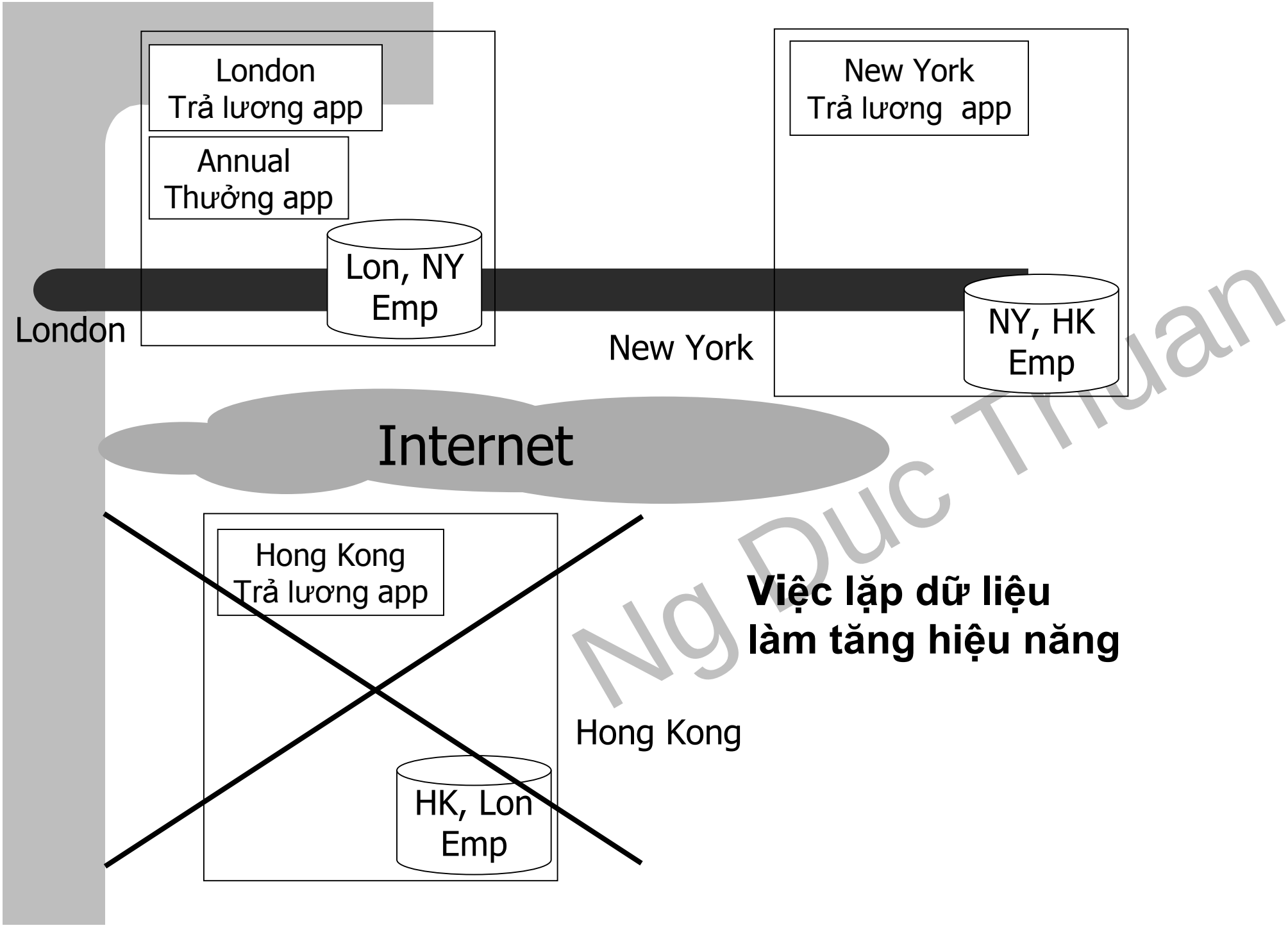
Ng Duc Thuan


London
Trả lương app
Total payroll app
London Emp
London
New York
Trả lương app
NY Emp
New York
Internet
Hong Kong
Trả lương app
HK Emp
Hong Kong
Sự phân tán hỗ trợ việc xử lý song song

Ng Duc Thuan

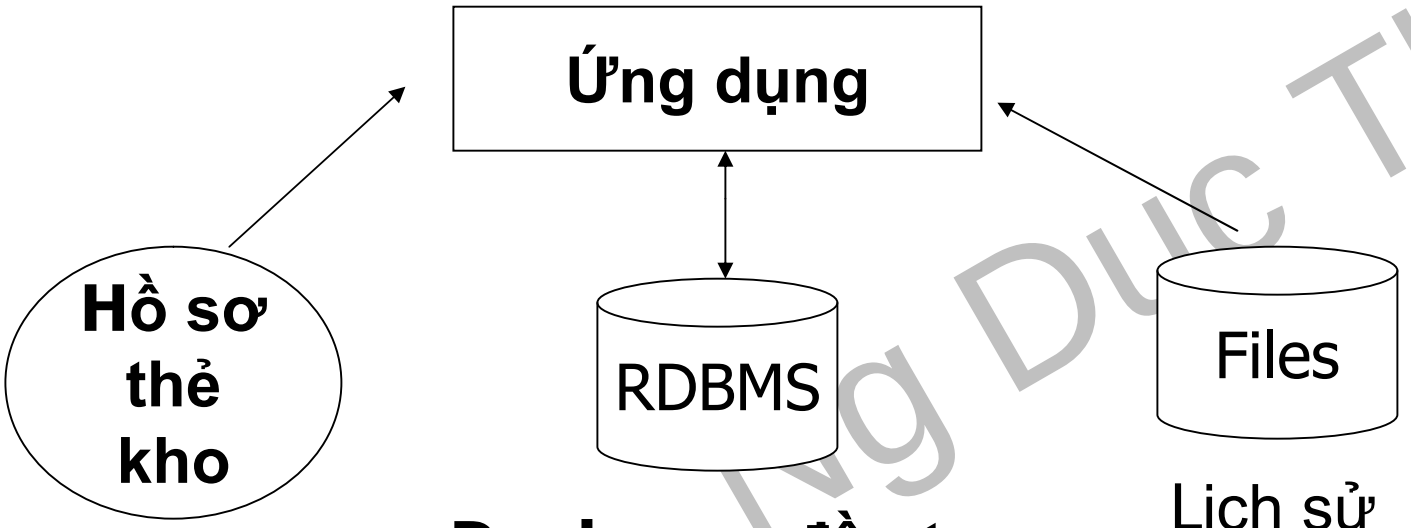

London
Trả lương app
Annual
Thưởng app
London
Emp
London
New York
Trả lương app
NY
Emp
New York
Internet
Hong Kong
Trả lương app
HK
Emp
Hong Kong

Ng Duc Thuan


London
Trả lương app
Annual
Thưởng app
Lon, NY
Emp
London
New York
Trả lương app
NY, HK
Emp
New York
Internet
Hong Kong
Trả lương app
HK, Lon
Emp
Hong Kong
Việc lặp dữ liệu
làm tăng hiệu năng
Ng Duc Thuan

Tính không đồng nhất & sự tự quản

Lựa chọn
Đầu tư mới



Ng Duc Thuan

NHỮNG ĐẶC ĐIỂM CỦA CSDL PHÂN TÁN SO VỚI CSDL TẬP TRUNG

(FEATURES OF DISTRIBUTED VERUS CENTRALIZED DATABASE)

- ***Điều khiển tập trung (Centralize control)***
- **Tự quản trạm** (*site autonomy*).
- ***Độc lập dữ liệu (Data independence)***
- **Tính trong suốt p.tán CSDL *(transparent)***
- ***Rút gọn dư thừa (Reduction of redundancy)***
- **Tính lặp dữ liệu** (*replication*)
- **Cấu trúc vật lý phức và truy xuất hiệu suất *cao* *(Complex physical structure & efficent access)***
- ***Chương trình chỉ đường*** (*navigational*)

Ng Duc Thuan

NHỮNG ĐẶC ĐIỂM CỦA CSDL PHÂN TÁN SO VỚI CSDL TẬP TRUNG

(FEATURES OF DISTRIBUTED VERUS CENTRALIZED DATABASE)

- ***Sự ràng buộc toàn vẹn, phục hồi và điều khiển tương tranh (Intergrity, recovery, and concurrency control)***
- ***Tính riêng rẽ & tính an toàn: (Privacy and security)***

Ng Duc Thuan

TẠI SAO CSDLPT CẦN THIẾT & PHÁT TRIỂN

- *Những lý do tổ chức và kinh tế : (Organizational and economic reasons)*
- *Liên kết với dữ liệu đã có (Interconnection of existing database)*
- *Sự tăng trưởng (Incremental growth)*
- *Giảm chi phí truyền thông (Reduced communication overhead)*
- *Cân đối hiệu năng (Performance consideration)*
- *Độ tin cậy và tính sẵn sàng (Reliability and availability)*

Ng Duc Thuan

MỘT SỐ NHƯỢC ĐIỂM KHI SỬ DỤNG HỆ PHÂN TÁN

- ***Sự thiếu kinh nghiệm***: (*Lack of Experience*)
- ***Sự phức tạp*** (Complexity)
- ***Chi phí*** : (*Cost*)
- ***Tính an toàn***: (*Security*)
- ***Khó khăn khi chuyển đổi*** (*Difficulty of change*) từ tập trung sang phân tán

Ng Duc Thuan

NHỮNG VẤN ĐỀ CẦN GIẢI QUYẾT KHI TIẾP CẬN CSDL PHÂN TÁN
(PROBLEM AREAS)

- **Thiết kế CSDL phân tán**

(Distributed Database Design)

- **Xử lý câu hỏi phân tán**

(Distributed Query Processing)

- **Quản trị thư mục phân tán**

(Distributed Directory Management)

- **Điều khiển tương tranh phân tán**

(Distributed Concurrency Control)

Ng Duc Thuan

NHỮNG VẤN ĐỀ CẦN GIẢI QUYẾT KHI TIẾP CẬN CSDL PHÂN TÁN
(PROBLEM AREAS)

- **Quản trị khóa chết phân tán**
  (Distributed Deadlock Management)
- **Độ tin cậy của hệ quản trị CSDL phân tán**
  (Reliability of Distributed DBMS)
- **Hỗ trợ của hệ điều hành**
  (Operating System Support)
- **Cơ sở dữ liệu không đồng nhất**
  (Heterogeneous Database)

Ng Duc Thuan

HỆ QUẢN TRỊ CSDL - PT

Là phần mềm cho phép tạo lập CSDL phân tán, điều khiển mọi truy cập đối với CSDL-PT đảm bảo trong suốt đối với người sử dụng.

- Chương trình quản trị DL địa phương
- Chương trình quản trị DL phân tán
- Chương trình quản trị truyền thông DL
- Chương trình quản trị tự điển DL

- **Các chức năng đặc thù của Hệ QTCSDL-PT**
  - Chương trình quản trị tự điển DL
  - Đánh giá câu hỏi phân tán
  - Quản trị giao dịch phân tán
  - Các chức năng phục hồi & toàn vẹn DL

Ng Duc Thuan

PHÂN LOẠI HỆ QTCSDL-PT

- **THUẦN NHẤT(homogeneous)**

  -DDBMS với hệ quản trị CSDL như nhau tại mỗi trạm , ngay cả khi máy tính và/hoặc hệ điều hành là không như nhau

- **KHÔNG THUẦN NHẤT (Heterogeneous)**

  - DDBMS với hệ quản trị CSDL khác nhau tại mỗi trạm

  - Có chức năng chuyển đổi giữa các mô hình dữ liệu khác nhau của các hệ quản trị CSDL cục bộ khác nhau

Ng Duc Thuan

PHÂN LOẠI HỆ QTCSDL-PT

- *Theo Bochmann, các loại hệ xử lý phân tán có thể phân loại dựa theo các tiêu chuẩn :*

Độ kết dính : (*degree of coupling*)là số đo mức độ nối kết chặt chẽ giữa các phần tử được xử lý, được tính bằng tỷ số giữa lượng dữ liệu trao đổi với lượng xử lý tại chỗ, được thực hiện trọng việc hoàn thành 1 giao tác

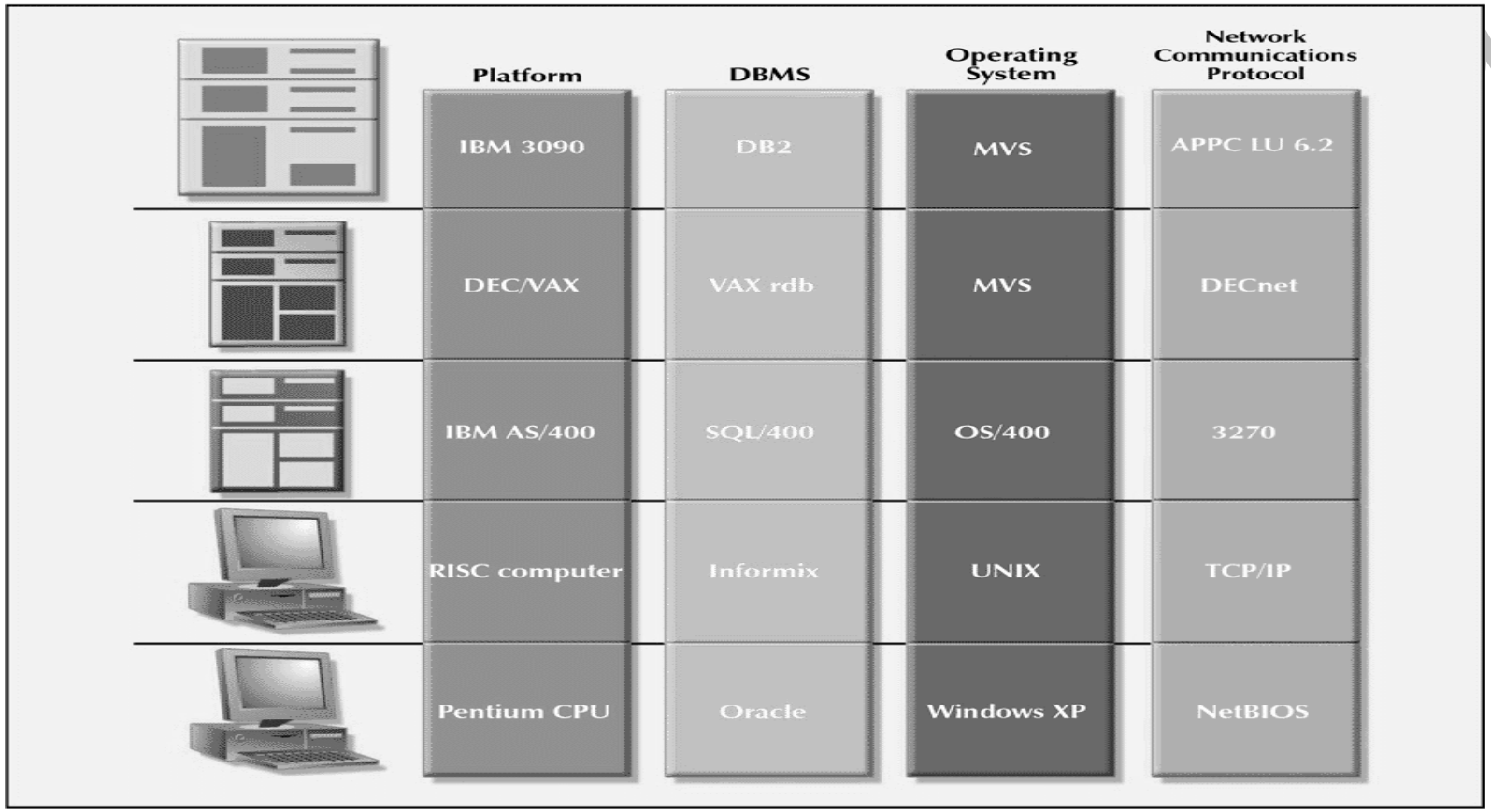
Cấu trúc nối kết (*Interconnection structure*) Nối kết điểm - diểm (*point-to-point*) giữa các phần tử xủ lý hay dùng kênh truyền chung

Sự đồng bộ hóa: (*Synchronization*) Sự đồng bộ hóa giữa các phần tử xử lý có thể được duy trì bởi các phương cách đồng bộ hay bất đồng bộ

Ng Đức Thuan

PHÂN LOẠI HỆ QTCSDL-PT

FIGURE 10.8 HETEROGENEOUS DISTRIBUTED DATABASE SCENARIO



TÀI LIỆU THAM KHẢO

- Slide bài giảng của Anand Rajaraman (http://www.stanford.edu/class/cs347)
- M. Tamer Ozsu and Patrick Valduriez, "Principles of Distributed Database Systems," Second Edition, Prentice Hall 1998.
- Đỗ Trung Tuấn :

  Cơ sở Dữ liệu. NXB. Giáo dục, 1997

Ng Đức Thuan