

# CHUYÊN ĐỀ

## Big Data với Excel Power Pivot

### Mục tiêu

- Phân tích, tính toán và ra quyết định từ tập dữ liệu lớn với Excel power pivot.

### Phần I Giới thiệu về Excel Power Pivot

Excel Power Pivot là một công cụ mạnh mẽ và hiệu quả được tích hợp vào Excel. Với Power Pivot, bạn có thể nạp hàng trăm triệu dòng dữ liệu từ các nguồn bên ngoài và quản lý dữ liệu hiệu quả với máy gia tốc mạnh theo một dạng nén cao. Điều này làm cho nó có thể thực thi tính toán, phân tích dữ liệu và đem đến một báo cáo để rút ra kết luận và các quyết định.

Các đặc điểm của Power Pivot:

1. Dữ liệu của Power Pivot được lấy từ nhiều nguồn khác nhau

Power Pivot có thể tập hợp dữ liệu từ nhiều nguồn dữ liệu khác nhau để thực thi tính toán theo yêu cầu.

2. Mô hình dữ liệu của Power Pivot

Sức mạnh của Power Pivot nằm trong cơ sở dữ liệu của nó đó chính là mô hình dữ liệu. Dữ liệu được lưu trữ theo khuôn dạng của các bảng dữ liệu. Bạn có thể tạo các mối liên kết giữa các bảng để kết hợp dữ liệu từ các bảng dữ liệu khác nhau cho mục đích phân tích và báo cáo.

3. Tạo các bảng và các lược đồ trong Power Pivot

Các bảng và các lược đồ trong Power Pivot cung cấp cho bạn một cách để phân tích dữ liệu để đi đến các kết luận và các quyết định.

#### 4. DAX (Data Analysis Expression) cơ bản

DAX là ngôn ngữ được sử dụng trong Power Pivot để thực thi các tính toán. Các biểu thức tính toán của DAX tương tự như các hàm của Excel, chỉ có khác là Excel thì tính toán dựa vào các ô riêng biệt còn DAX tính toán dựa vào các cột (các trường).

#### 5. Các kiến trúc

Bạn có thể định nghĩa các kiến trúc dữ liệu trong một bảng dữ liệu mà nó sẽ dễ dàng để điều khiển các trường dữ liệu liên quan với nhau trong các bảng của Power Pivot.

#### 6. Các chức năng của Power Pivot với Excel

- Khả năng để điều khiển dữ liệu lớn, nén vào các file nhỏ với tốc độ nhanh
- Lọc dữ liệu và đổi tên các cột, các bảng trong khi thêm dữ liệu
- Tổ chức các bảng vào trong các tab của các trang riêng biệt
- Tạo mối quan hệ giữa các bảng để phân tích dữ liệu theo tập hợp các bảng.
- Cung cấp ngôn ngữ biểu thức phân tích dữ liệu (DAX) để tạo các hàm phức tạp.
- Thêm các trường và các cột được tính toán vào trong các bảng dữ liệu
- Tạo các chỉ số thực thi khóa để sử dụng trong Pivot Table và các báo cáo.

#### 7. Các nguồn dữ liệu được hỗ trợ bởi Power Pivot

- Microsoft Access database
- SQL Server Analysis Services
- SQL Server Reporting Service (SQL 2008 R2)
- ATOM data feeds
- Text files

- Microsoft SQL Azure
- Oracle
- Teradata
- Sybase
- Informix
- IBM DB2
- Object Linking and Embedding Database/ Open Database Connectivity
- (OLEDB/ODBC) sources
- Microsoft Excel File
- Text File

## **Phần II Phân tích, tính toán với tập dữ liệu lớn**

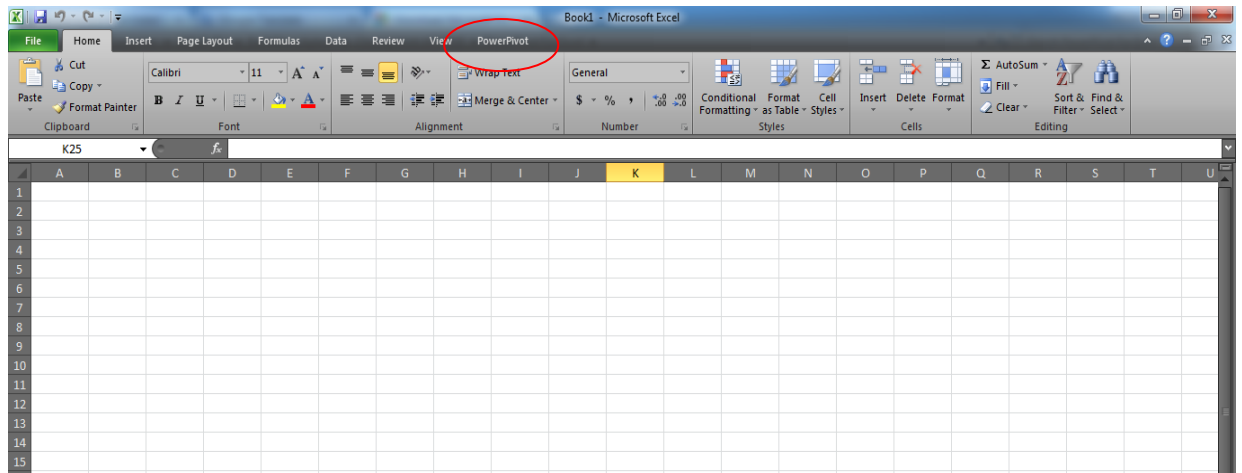
---

### **1. Tích hợp Power Pivot vào Excel**

- Download bộ cài Power Pivot đúng với phiên bản của Excel (thường download bộ cài 32 bit - x86)
- Nháy đúp vào file .msi vừa download về để cài đặt

### **2. Khởi động Power Pivot từ chương trình Excel**

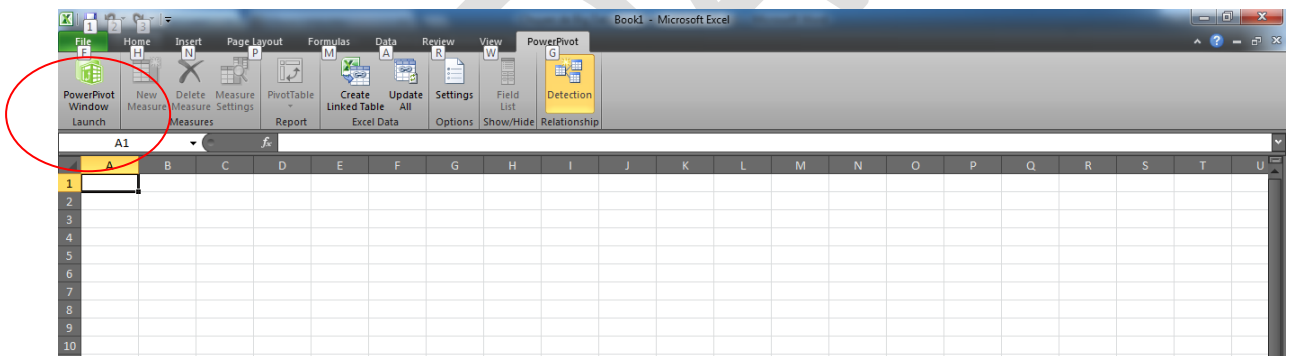
- Khởi động chương trình Excel.
- Trên thanh công cụ sẽ hiển thị biểu tượng của PowerPivot



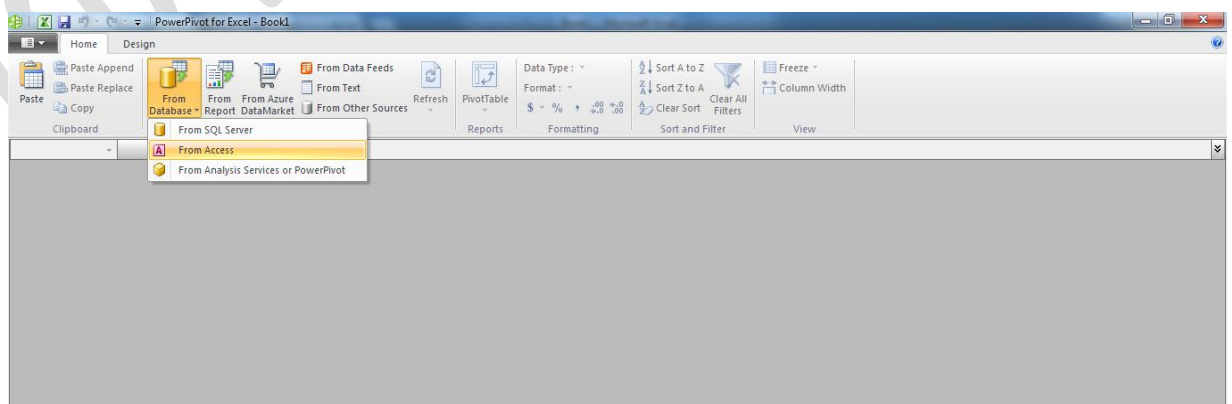
### 3. Nạp dữ liệu trực tiếp vào trong PowerPivot (Nạp dữ liệu từ Access)

Để nạp dữ liệu trực tiếp vào trong Power Pivot, thực thi theo các bước sau:

- Từ giao diện của PowerPivot chọn PowerPivot Window Launch



- Lựa chọn nguồn sẽ nạp dữ liệu vào cho PowerPivot (chọn from Database/ From Access). Bảng Table Import Wizard sẽ xuất hiện



- Chọn database name (nhấn vào nút Browse để tìm và chọn)

**Table Import Wizard**

**Connect to a Microsoft Access Database**  
Enter the information required to connect to the Microsoft Access database.

Friendly connection name: Access MyDatabase

Database name: D:\BachKhoa\_Aptech\Chương trình ACCPi18\2 Browse...

Log on to the database

User name:

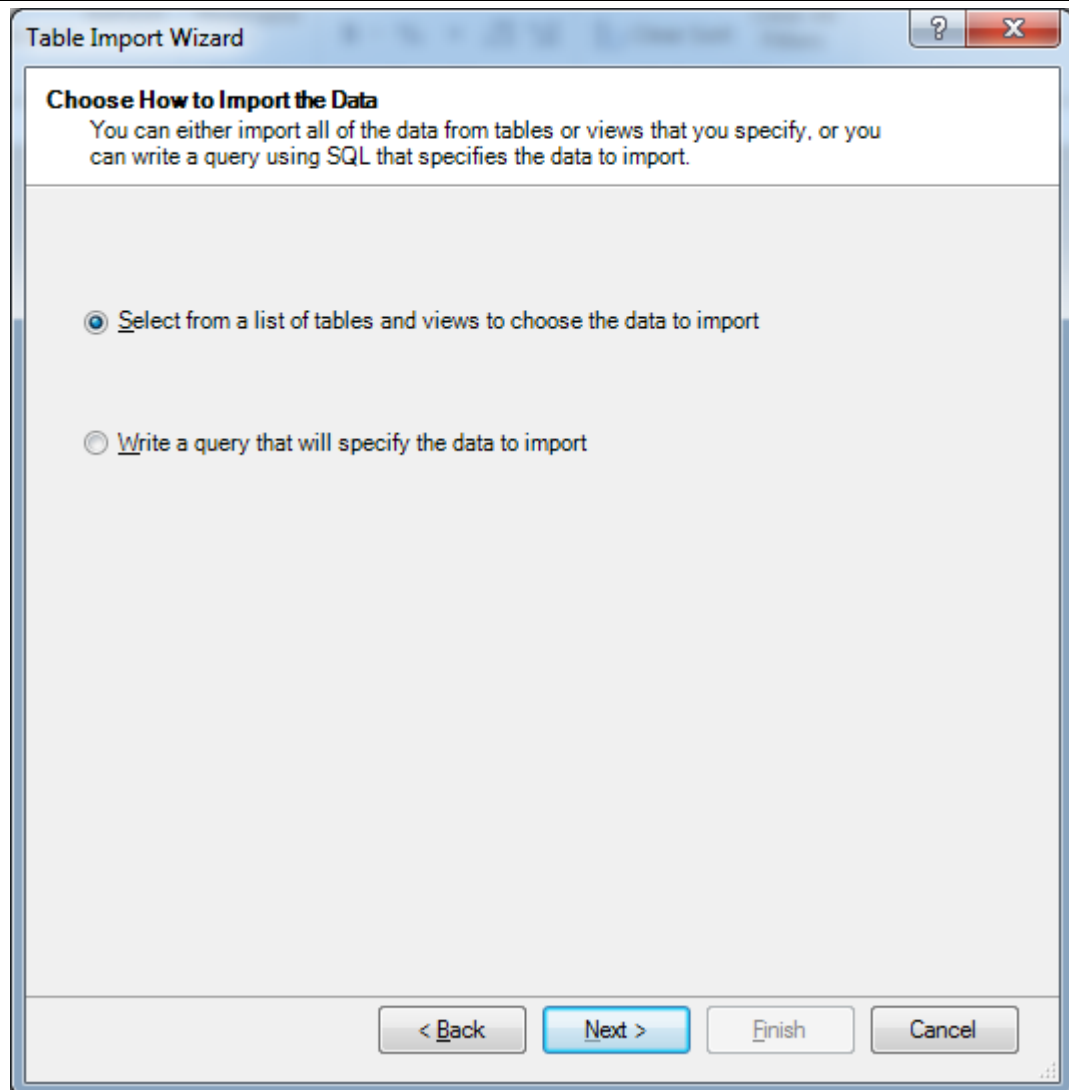
Password:

☐ Save my password

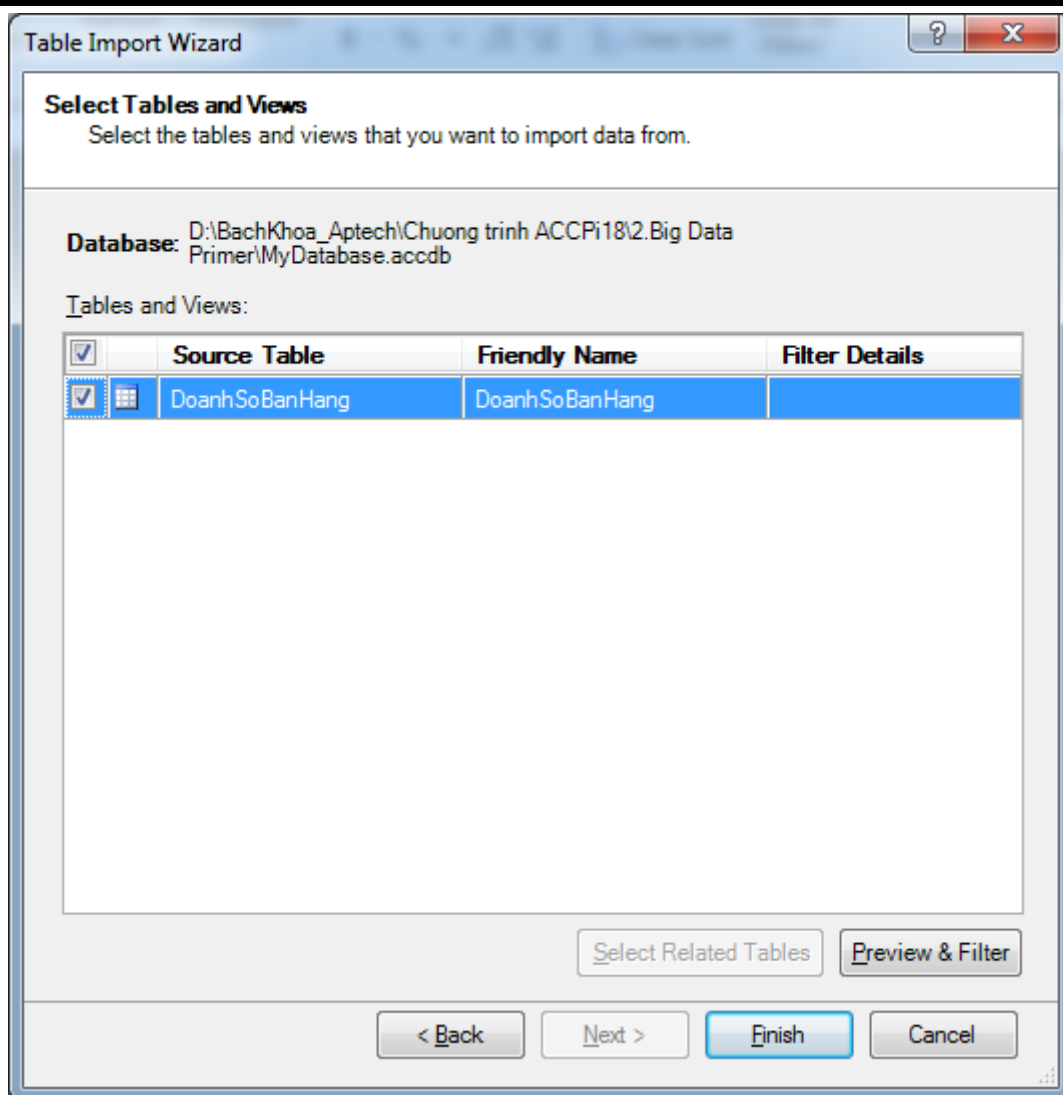
Advanced... Test Connection

< Back Next > Finish Cancel

- Chọn Next để lựa chọn cách import dữ liệu vào



- Lựa chọn các bảng để import vào PowerPivot và chọn Finish

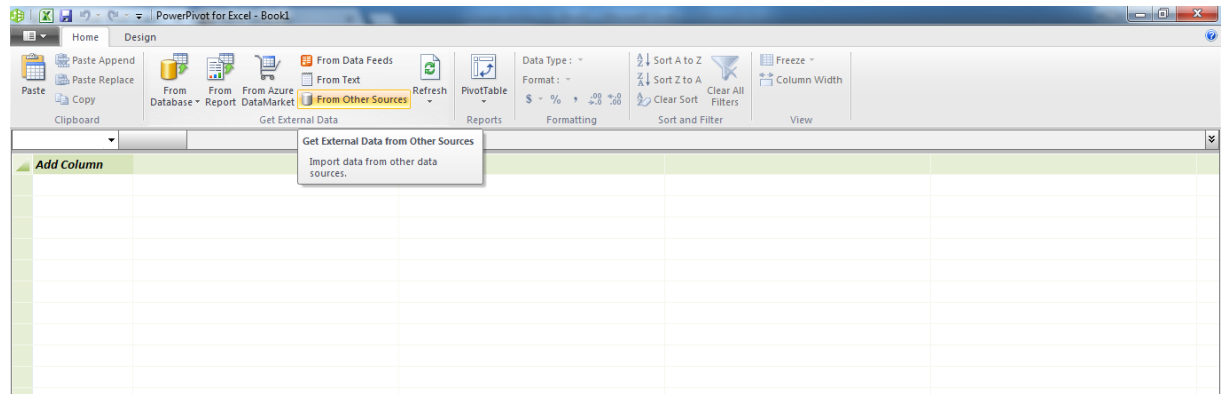


– Dữ liệu được nạp vào PowerPivot

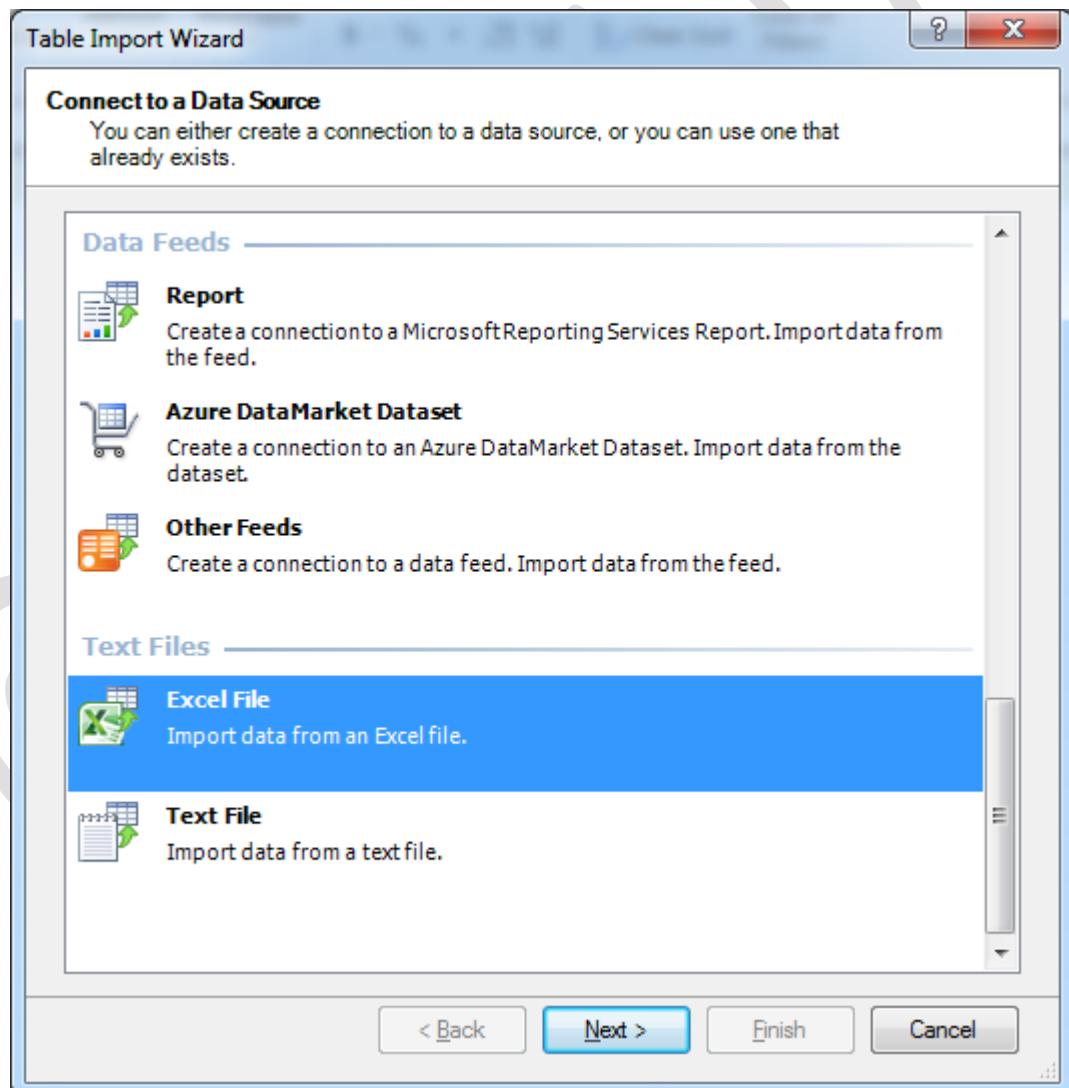
ID	Field13	Field1	Field2	Field3	Field4	Field5	Field6	Field7	Field8	Field9	Field10	Field11	Field12
12	Tháng 12	Họ tên	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11
13	33000000	Nguyễn ...	30000000	25000000	35000000	20000000	22000000	23500000	20000000	25000000	27000000	0	0
14	27000000	Nguyễn ...	8000000	7000000	10000000	12000000	10000000	15000000	12000000	15000000	20000000	22000000	23000000
15	8000000	Trần Thủ...	5000000	6000000	5500000	7000000	8000000	6000000	7000000	8000000	5000000	7000000	5000000
16	10000000	Ngô Thị ...	2000000	4000000	5000000	6000000	7000000	6000000	5000000	7000000	8000000	7000000	8000000
17	14000000	Đoãn Th...	5000000	3000000	5000000	6000000	8000000	6000000	7000000	5000000	7000000	8000000	12000000
18	15000000	Nguyễn ...	6000000	7000000	8000000	9000000	12000000	10000000	11000000	13000000	12000000	11000000	13000000

#### 4. Nạp dữ liệu từ file Excel

- Từ giao diện của PowerPivot chọn From Other Sources



- Chọn Excel file



- Chọn Next và Browse để tìm đến vị trí file excel
- Chọn Next và chọn các sheet trong file excel sẽ chứa dữ liệu



- Chọn Finish sau đó dữ liệu sẽ được import vào PowerPivot

	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11	Tháng 12
Họ tên	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11	Tháng 12
Nguyễn Thị Hằng	30000000	25000000	35000000	20000000	22000000	23500000	20000000	25000000	27000000	0	33000000	
Nguyễn Trà Giang	8000000	7000000	10000000	12000000	10000000	15000000	12000000	15000000	20000000	22000000	23000000	27000000
Trần Thủy Hương	5000000	6000000	5500000	7000000	8000000	6000000	7000000	8000000	5000000	7000000	5000000	8000000
Ngô Thị Thủy Linh	2000000	4000000	5000000	6000000	7000000	6000000	5000000	7000000	8000000	8000000	10000000	
Đoàn Thu Trang	5000000	3000000	5000000	6000000	8000000	6000000	7000000	5000000	7000000	8000000	12000000	14000000
Nguyễn Mạnh Đức	6000000	7000000	8000000	9000000	12000000	10000000	11000000	13000000	12000000	11000000	13000000	15000000

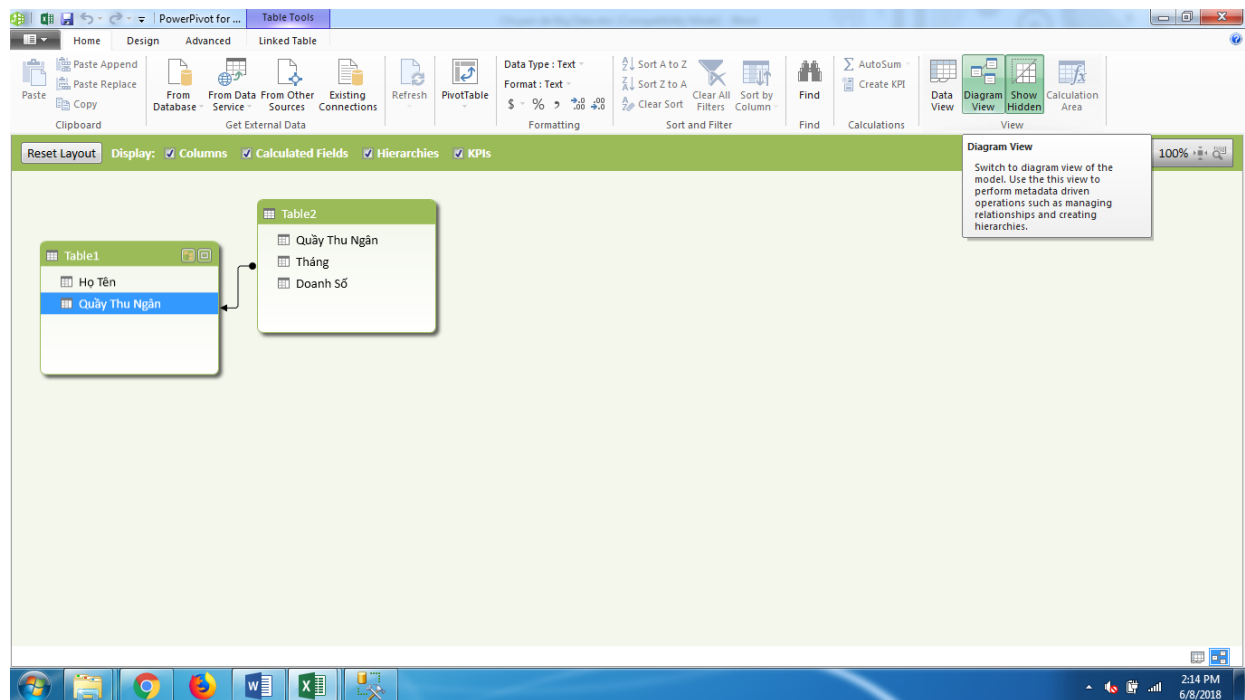
## 5. Thêm nhiều bảng của Excel vào mô hình dữ liệu của PowerPivot với PivotTable (thực hiện với Office 2013 trở lên)

- Đánh dấu bảng cần thêm vào rồi chọn PowerPivot
- Chọn Add to Data Model

	Họ Tên	Quầy Thu Ngân
1		
2	Họ Tên	Quầy Thu Ngân
3	Nguyễn Thanh Mai	Quầy 1
4	Nguyễn Thị Huyền	Quầy 2
5	Trần Thanh Vân	Quầy 3
6	Ngô Ngọc Hà	Quầy 4
7	Lê Thanh Hải	Quầy 5

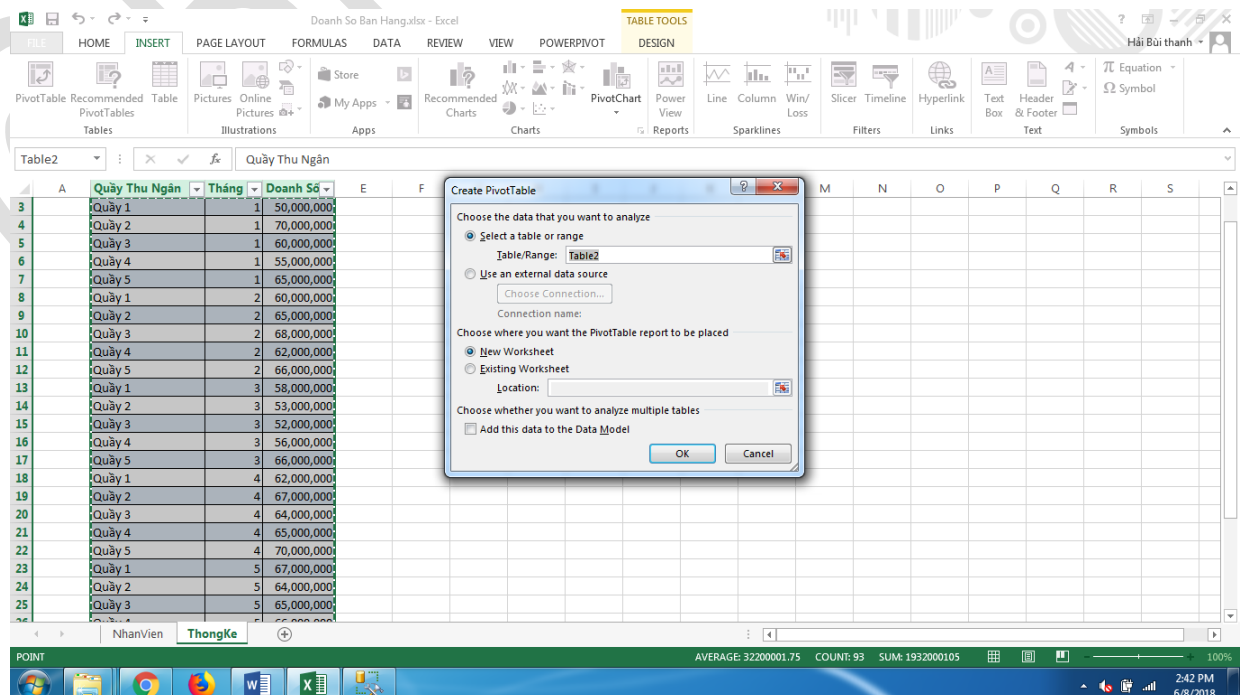
- Làm tương tự như vậy cho các bảng cần thêm vào, ta sẽ thêm được nhiều bảng vào Data Model.
- Chọn Diagram View để xem lược đồ các bảng và tạo liên kết giữa các bảng

- Kéo thả giữa các cột chung để tạo liên kết giữa các bảng này

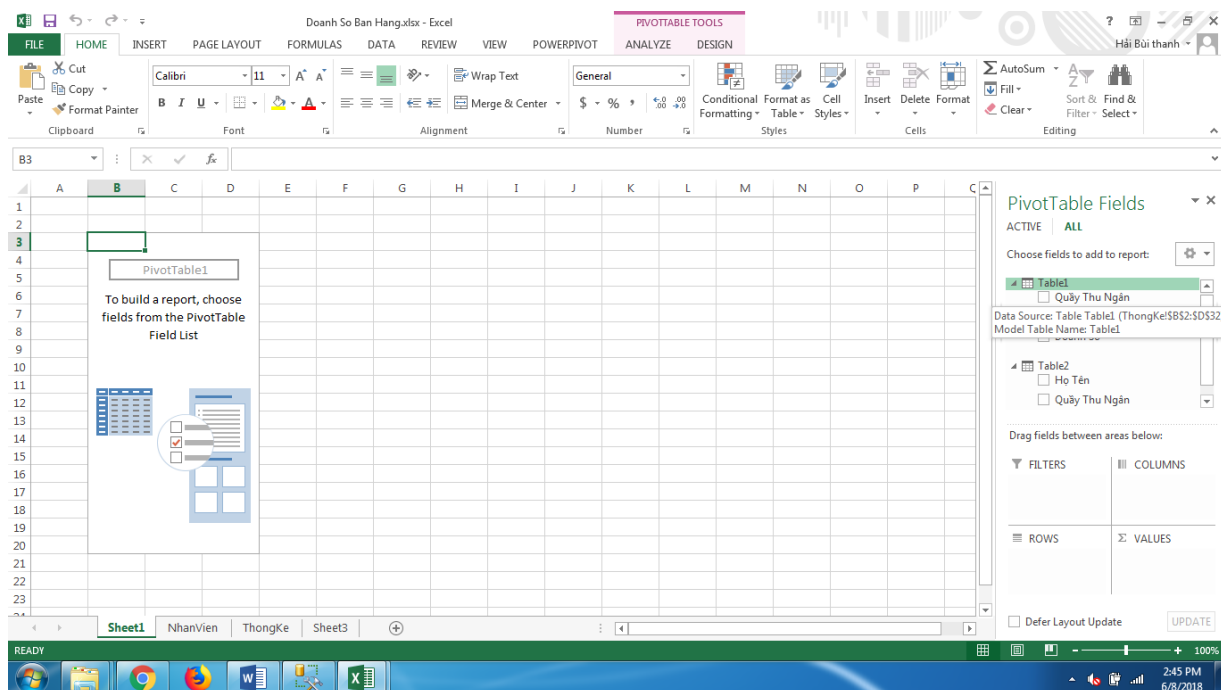


## 6. Tạo PivotChart để xem thống kê giữa các bảng

- Lựa chọn bảng Thống Kê (Bảng chi tiết)
- Click tab INSERT trên thanh menu
- Chọn PivotTable trên nhóm các bảng



- Chọn Ok một cửa sổ PIVOTTABLE TOOLS sẽ xuất hiện.
- Chọn MORE TABLES để thêm các bảng khác vào, cửa sổ nhắc có tạo một Pivot Table mới không, chọn Yes
- Hai bảng với các cột sẽ hiển thị ở bên phải của giao diện



- Gấp vào kéo cột Họ Tên của bảng thứ hai vào vùng ROWS
- Gấp vào kéo cột Tháng của bảng thứ nhất vào vùng ROWS
- Gấp vào kéo cột Doanh Số của bảng thứ nhất vào vùng VALUES
- Chúng ta sẽ được Doanh Số thống kê bán hàng của mỗi nhân viên.

Row Labels	Sum of Doanh Số
Lê Thanh Hải	410000000
1	65000000
2	66000000
3	66000000
4	70000000
5	68000000
6	75000000
Ngô Ngọc Hà	374000000
1	55000000
2	62000000
3	56000000
4	65000000
5	66000000
6	70000000
Nguyễn Thanh Mai	369000000
1	50000000
2	60000000
3	58000000
4	62000000
5	67000000
6	72000000

- Để xem thống kê có thêm thông tin về Quầy Thu Ngân của nhân viên đó, gấp vào kéo cột Quầy Thu Ngân của bảng thứ hai xuống vùng COLUMNS, ta được kết quả như sau:

Sum of Doanh Số	Column Labels	Quầy 2	Quầy 3	Quầy 4	Quầy 5	Grand Total
Lê Thanh Hải	Quầy 1					410000000
1						65000000
2						66000000
3						66000000
4						70000000
5						68000000
6						75000000
Ngô Ngọc Hà						374000000
1						55000000
2						62000000
3						56000000
4						65000000
5						66000000
6						70000000
Nguyễn Thanh Mai						369000000
1						50000000
2						60000000
3						58000000
4						62000000
5						67000000
6						72000000

- Chú ý: Để load được dữ liệu từ tất cả các nguồn mà Pivot Table hỗ trợ thì bạn nên sử dụng bộ office 2013 trở lên, các bộ office 2010 trở về trước đó sẽ không hỗ trợ lấy dữ liệu từ một số nguồn.

#### BÀI TẬP TỰ LÀM.

Hãy tạo một CSDL về điểm thi của lớp theo các môn và các học kỳ (có thể tạo bằng SQL Server). Sau đó sử dụng Pivot Table để thống kê điểm trung bình theo mỗi kỳ của từng sinh viên đó.