BÀI TẬP NHÓM OLAP

Thành viên nhóm :

* Đoàn Văn Đoàn 14520168.
* Trần Hoàng Sơn.
* Đặng Hồng Hải.
* Nguyễn Trần Minh An.

Dataset : BC Schools - Foundation Skills Assessment (FSA).

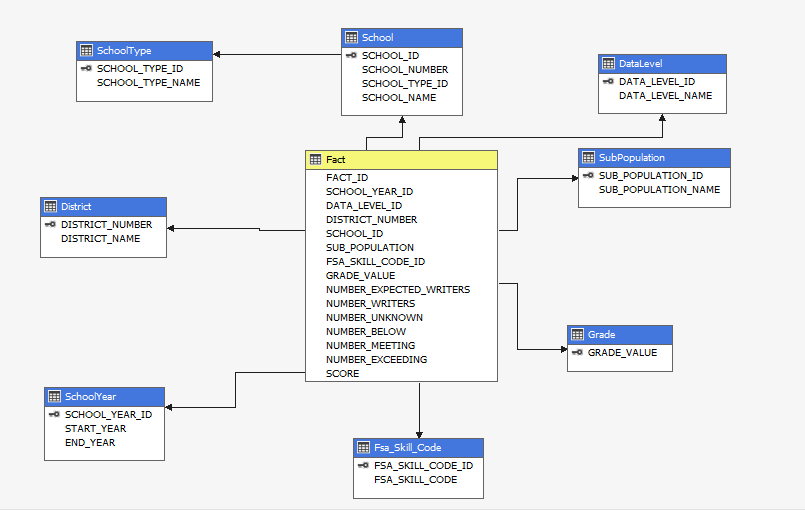
Bộ dữ liệu:Llà các kết quả của BC Foundation Skills Assessment (Đánh giá kỹ năng nền tảng) của lớp 4 và 7 BC về Tính toán, Đọc và Viết từ năm 2007/2008 đến 2015/2016 trên địa bàn tỉnh (bang) British Comlumbia, phía tây Canada.

Link dataset : <https://catalogue.data.gov.bc.ca/dataset/bc-schools-foundation-skills-assessment-fsa->

1. Phân tích bộ dữ liệu.
2. Phân tích.

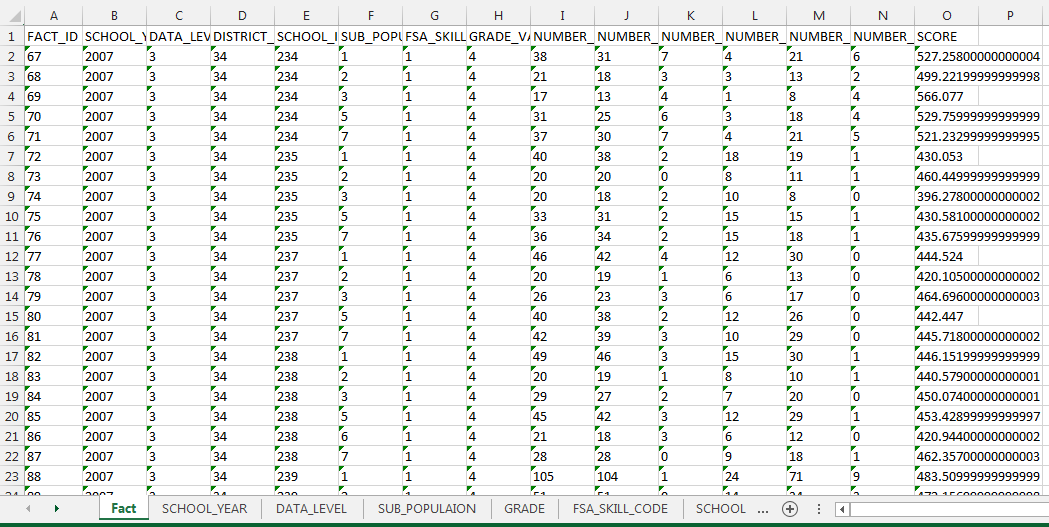
|  |  |
| --- | --- |
| Tên thuộc tính | Giải thích |
| SCHOOL\_YEAR | Năm học tiến hành khảo sát. |
| DATA\_LEVEL | Cấp độ dữ liệu, Thông tin về trường dữ liệu thuộc cấp độ nào (Tỉnh, quận, hay trường). |
| PULIC\_OR\_INDEPENDENT | Thông tin về loại trường của trường đó là trường công (pulic school) hay là trường tư (independent school). |
| DISTRICT\_NUMBER | Mã quận. |
| SCHOOL\_NUMBER | Mã trường. |
| SCHOOL\_NAME | Tên trường. |
| SUB\_POPULATION | Nhóm đối tượng thực hiện khảo sát. |
| FSA\_SKILL\_CODE | Kĩ năng được thực hiện khảo sát. |
| GRADE | Khối lớp. |
| NUMBER\_EXPECTED\_WRITERS | Số lượng người dự kiến tham gia. |
| NUMBER\_WRITERS | Số lượng người tham gia thực tế. |
| NUMBER\_UNKNOWN | Số lượng người không tham gia. |
| NUMBER\_BELOW | Số lượng người không đạt chuẩn. |
| NUMBER\_MEETING | Số lượng người đạt chuẩn. |
| NUMBER\_EXCEEDING | Số lượng người vượt chuẩn. |

1. Thiết kế lược đồ dữ liệu.

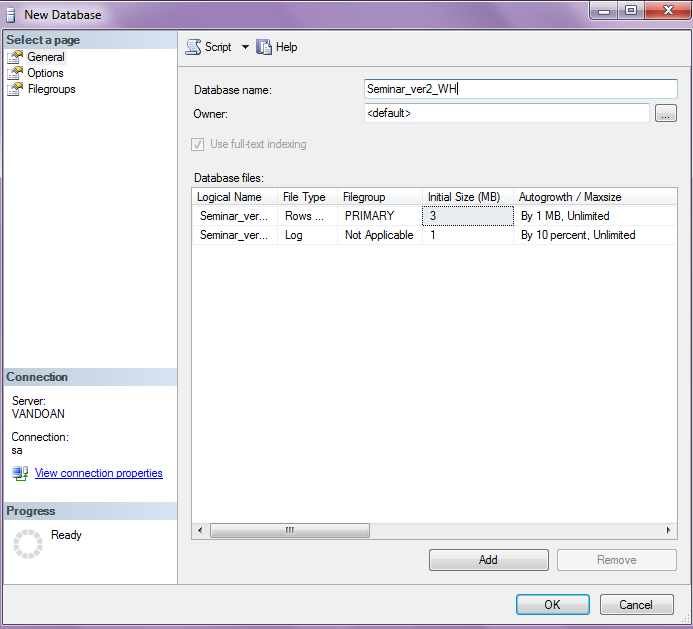


1. Quá trình Business intelligence.
2. Import với bảng fact.

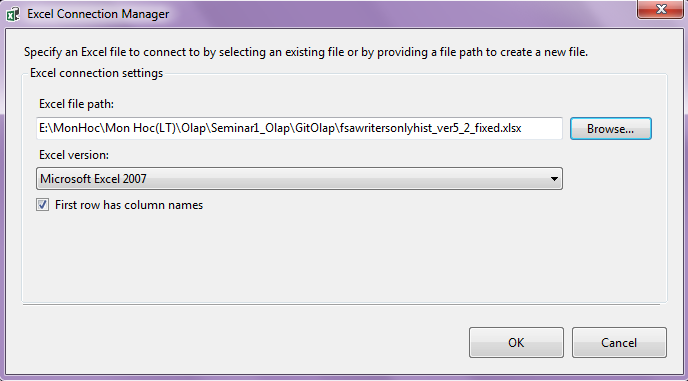
* Tiến hành xử dữ liệu, lược bỏ đi dữ liệu thừa, tách trích ra các dimension trên các sheet khác nhau để dễ dàng import dữ liệu cho quá trình BI.



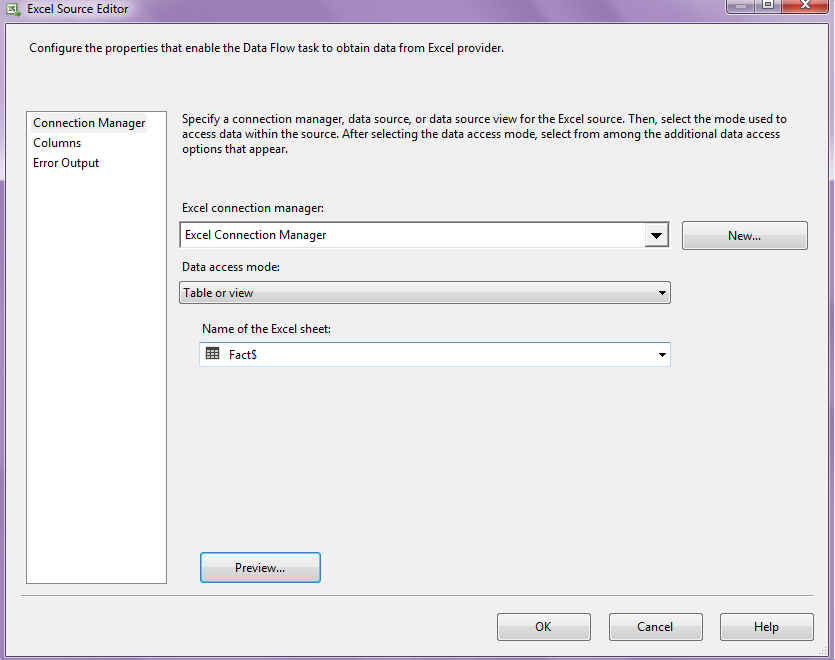
* Kết nối với hệ quản trị cơ sở dữ liệu SQL Server, tạo mới database.



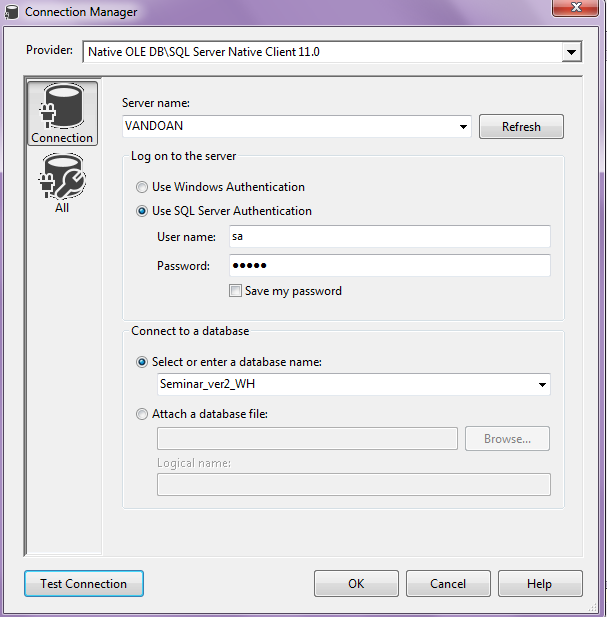
* Tạo project BI mới để import dữ liệu từ file excel ta đã xử lí từ trước.
* Chọn file excel để import.



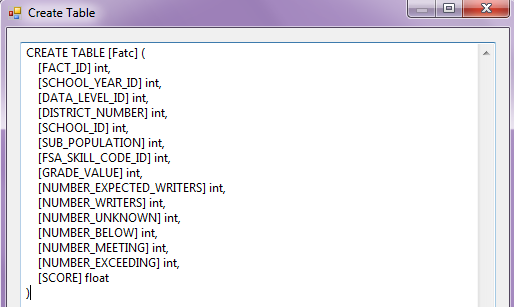
* Chọn sheet muốn import



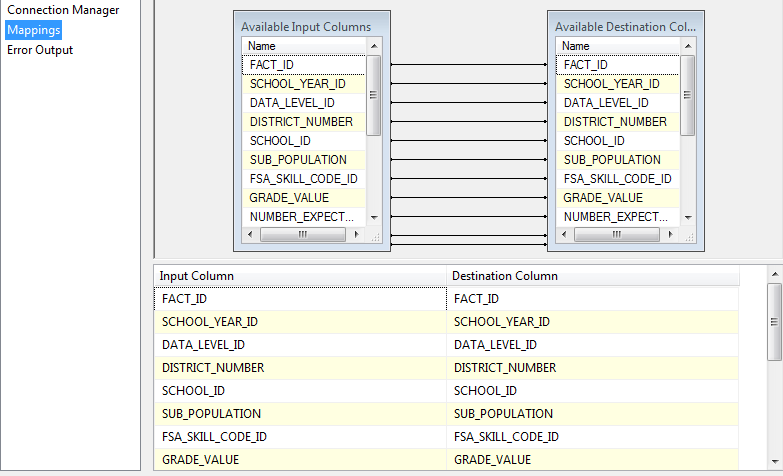
* Thực hiện chọn database để export data từ file excel



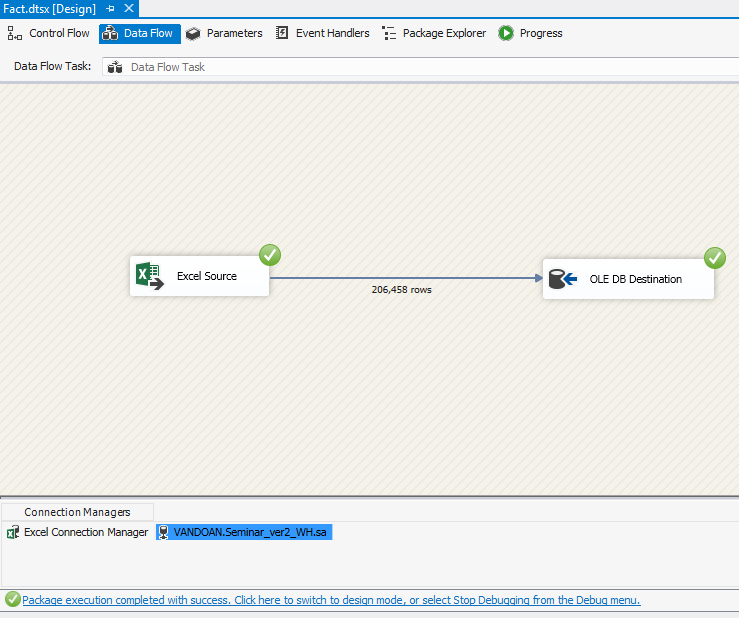
* Tùy chỉnh lại kiểu dữ liệu cho từng cột đối với bảng ta chuẩn bị tạo, ở đây là bảng Fact



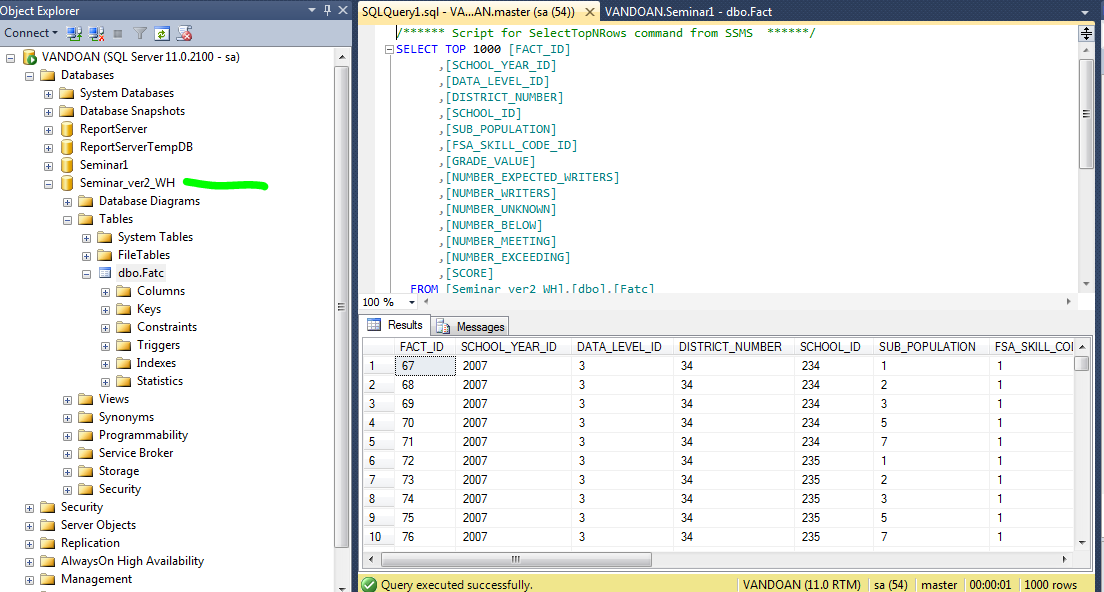
* Tiến hành mapping để đồng nhất giữa nguồn và đích của dữ liệu



* Tiến hành chạy package để import dữ liệu, kết quả được như sau

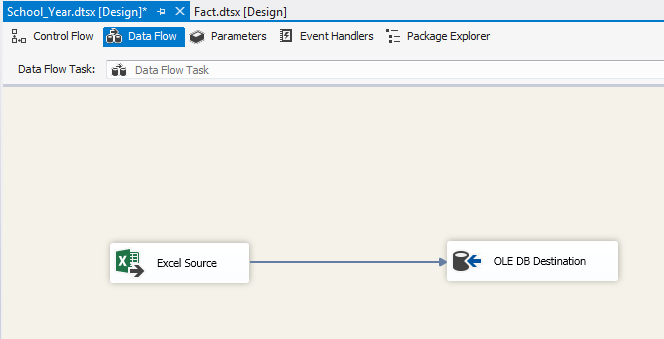


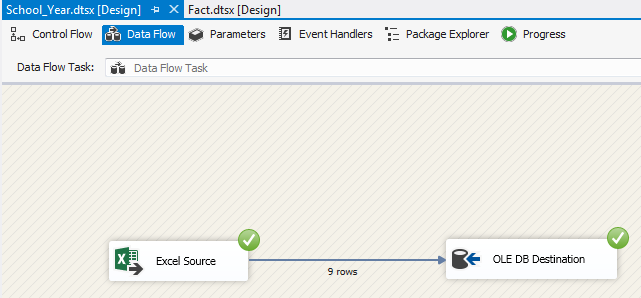
* Ta tiến hành kiểm tra lại trong SQL Server



* Ta cũng thực hiện tương tự với cái sheet khác.

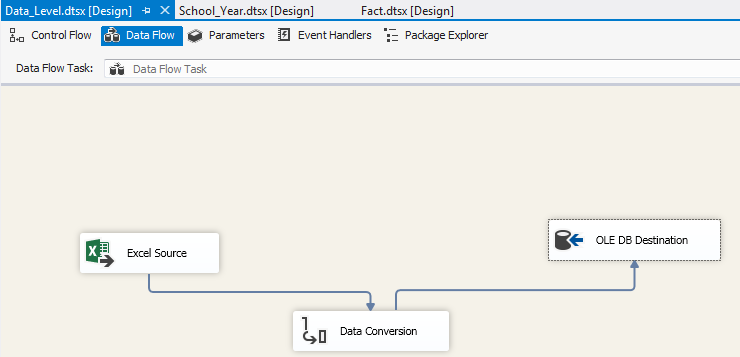
1. Import với bảng shool\_year

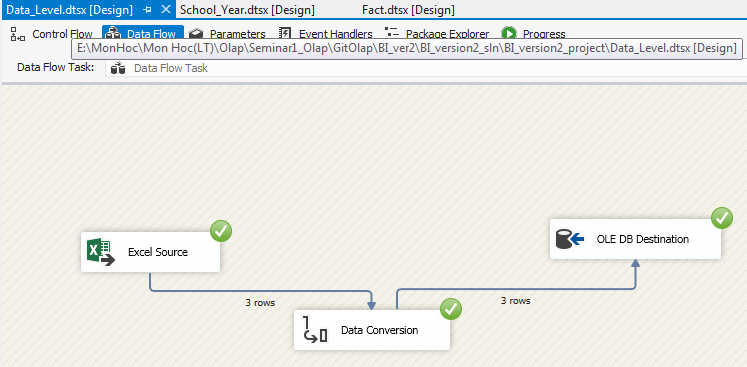




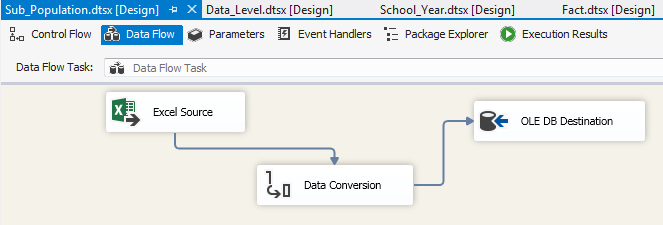
1. Import bảng Data\_level.

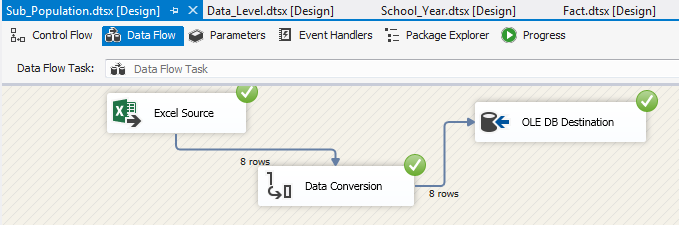
* Lưu ý : Đối với dữ liệu là kiểu varchar thì ta phải dùng bộ chuyển đổi để tránh gây ra lỗi trong lúc import dữ liệu.



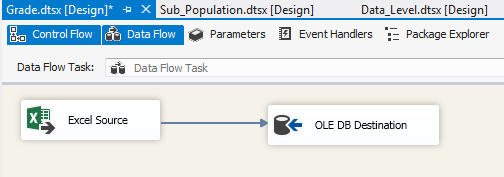


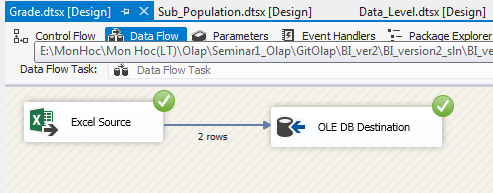
1. Import bảng sub population.



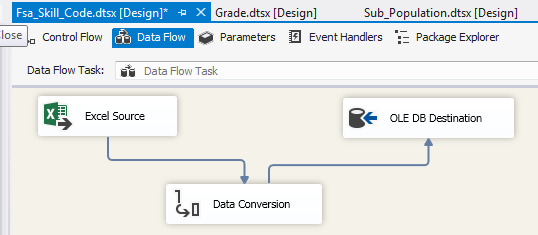


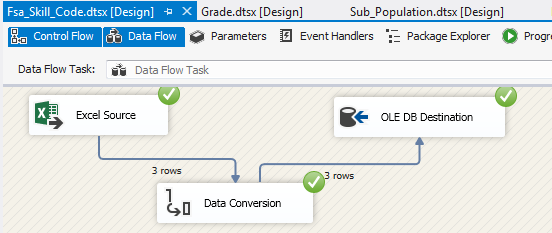
1. Import bảng Grade



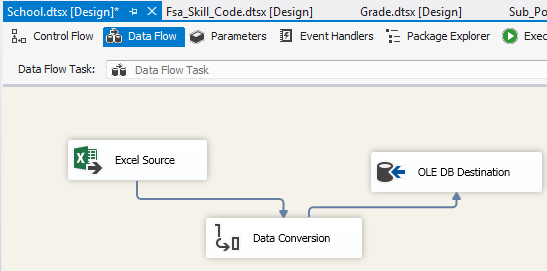


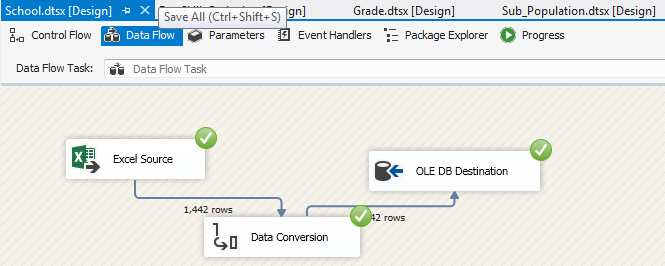
1. Import bảng Fsa skill code



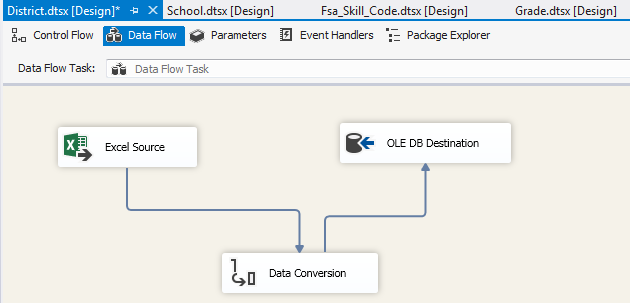


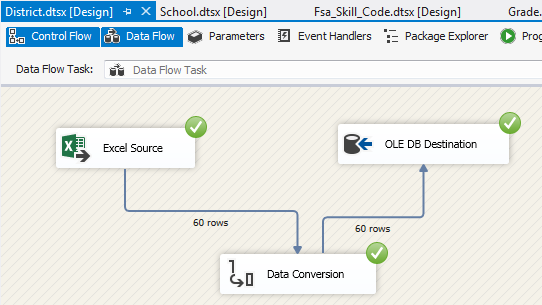
1. Import bảng school



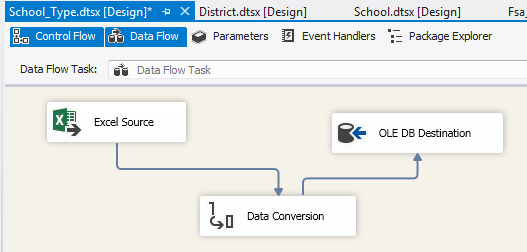


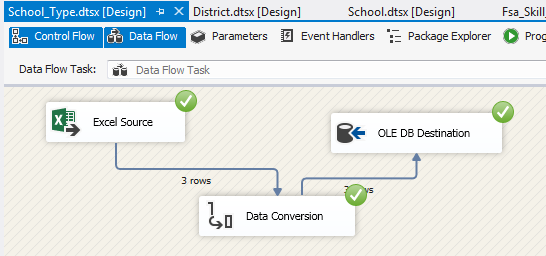
1. Import bảng district



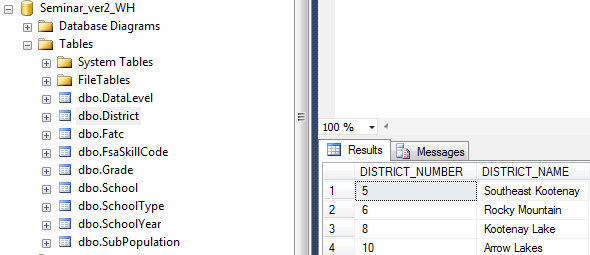


1. Import bảng school type.





Kiểm tra lại SQL Server và tất cả đã được import, sau đó ta tiến hành tạo khóa chính cho từng bảng.



1. Câu truy vấn.

Câu 1: Liệt kê 10 trường có đối tượng (SUB\_POPULATION) là “ENGLISH LANGUAGE LEARNER” và số lượng người (NUMBER\_WRITERS) lớn nhất đồng thời có hoạt động trong năm học “2007/2008”

Câu 2: Liệt kê 10 quận (DISTRICT\_NUMBER, DISTRICT\_NAME) sỡ hửu nhiều trường học nhất có hoạt động trong năm “2009/2010”

Câu 3: Số lượng người tham dự (NUMBER\_WRITERS) của khối (GRADE) 7 ở tất cả các trường trong năm học “2007/2008” tại quận (DISTRICT\_NAME) “Sea to Sky”

Câu 4: Trường có số lượng người tham gia lớn nhất (number\_writer) có kĩ năng (FSA\_SKILL\_CODE) là “Reading” có học tại năm học “2008/2009” và (SUB\_POPULATION) là “ABORIGINAL”.

Câu 5: Top 10 trường(DISTRICT\_NUMBER, DISTRICT\_NAME) có điểm số(SCORE) cao nhất trong quận “Central Okanagan” có hoạt động trong năm “2007/2008”

Câu 6: Trong năm 2007 – 2008 đối với cấp độ là cấp tỉnh (Data-level : Province level) hãy cho biết những đối tượng nào (sub\_population) có người vượt chỉ tiêu với kĩ năng là tính toán (numeracy).

Câu 7: Trong năm 2007-2008 với cấp độ là cấp quận (Data\_level : district\_level) và trường học thuộc trường công (PUBLIC\_OR\_INDEPENDENT: BC public school). Hãy cho biết 10 quận (district\_name) có số người đạt kĩ năng viết cao nhất.

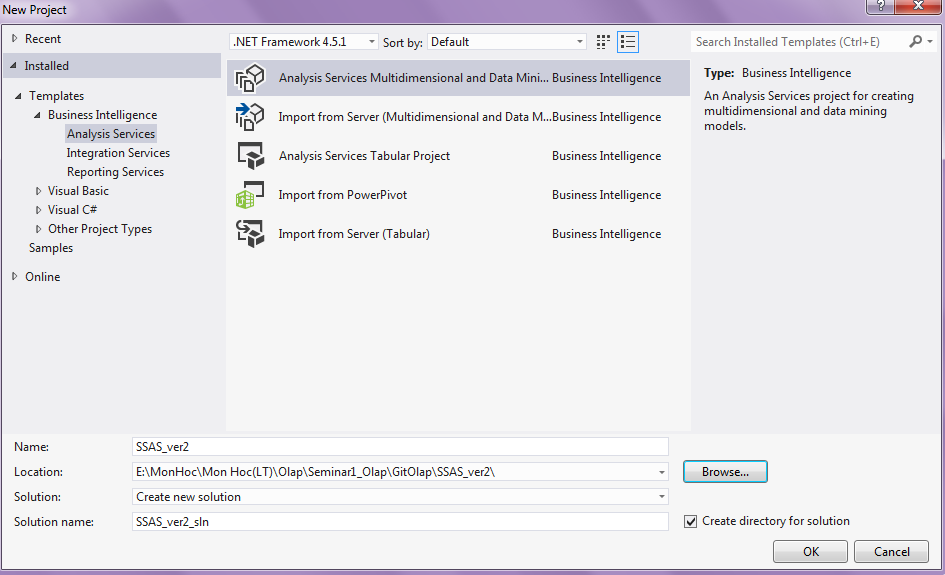
Câu 8 : Cho biết số lượng thí sinh thi không đạt với kĩ năng Reading, Writing ở tất cả các năm học.

Câu 9: Trong năm 2014-2015 (SCHOOL\_YEAR) với những trường (DATA\_LEVEL: SCHOOL LEVEL) thuộc khối trường tư (PUBLIC\_OR\_INDEPENDENT : BC Independent School) và với khối lớp 7 hãy cho biết 10 trường học (SCHOOL\_NAME) có số người có kĩ năng tính toán(Numeracy) cao nhất với những đối tượng là nam (SUB\_POPULATION: MALE).

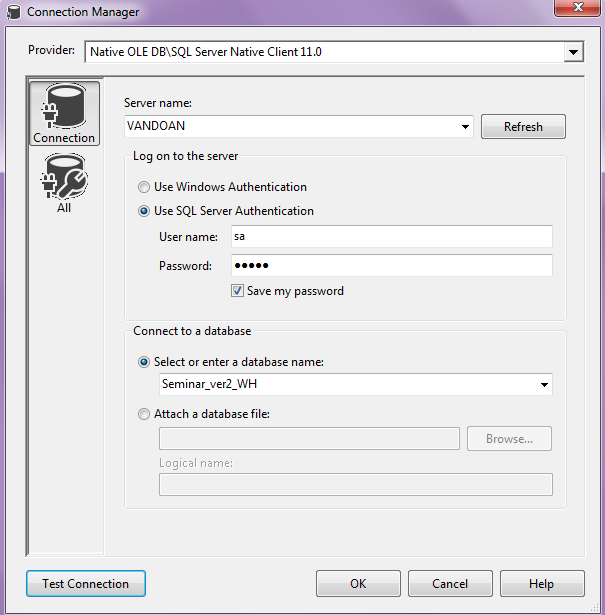
Câu 10: Liệt kê TOP 10 quận (DISTRICT\_NAME) có tổng điểm đánh giá (SCORE) lớn nhất trong năm học (SCHOOL\_YEAR) 2012/2013).

1. SQL Server Analysis Services (Áp dụng 10 câu truy vấn ở trên).
2. Tạo mới project SSAS

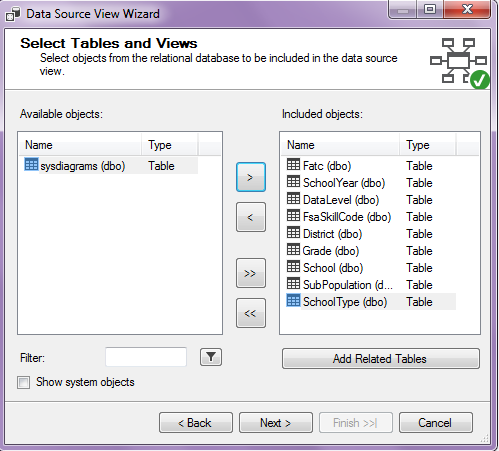
* Tạo mới project



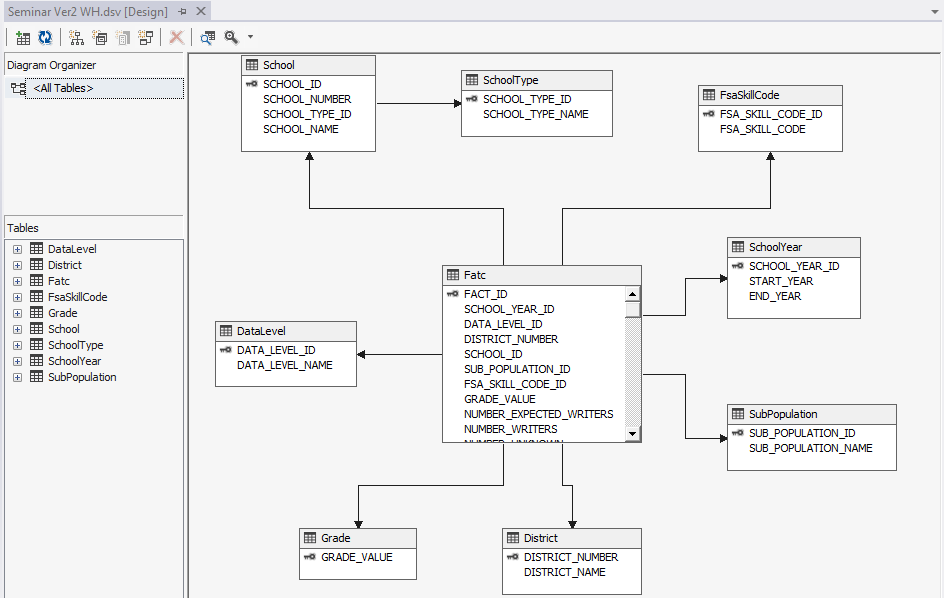
* Tạo data source.



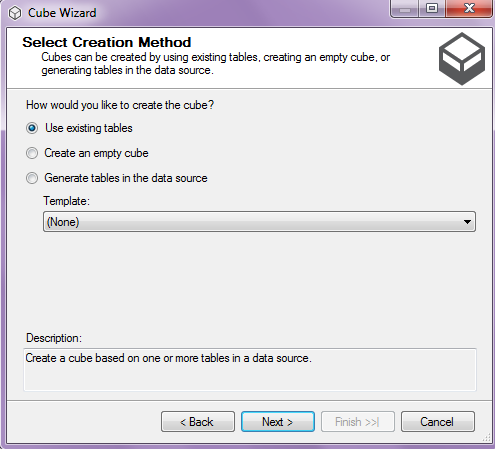
* Tạo New Data Source Views

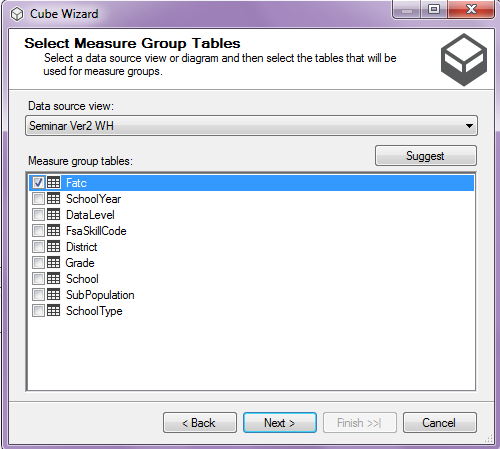


* Sau khi tạo xong thì ta có kết quả như hình dưới :

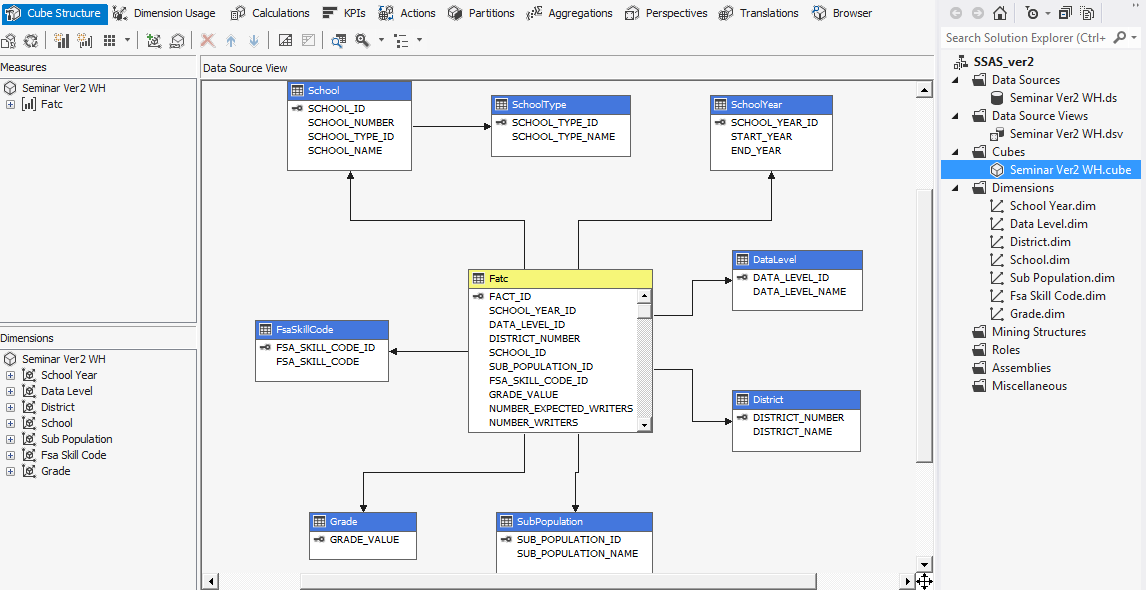


* Tạo mới Cube

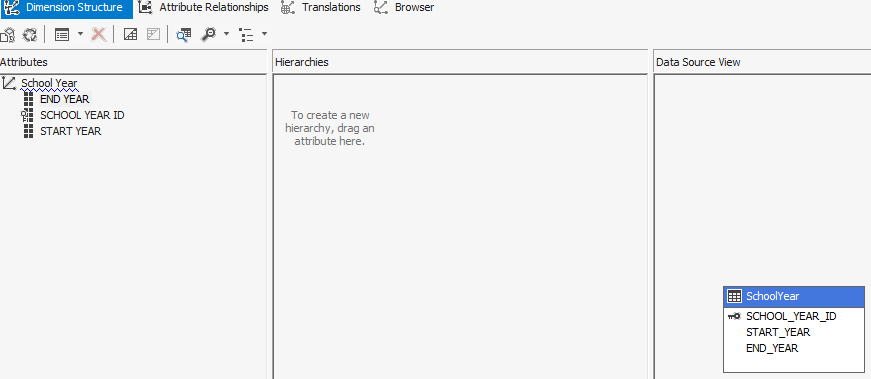


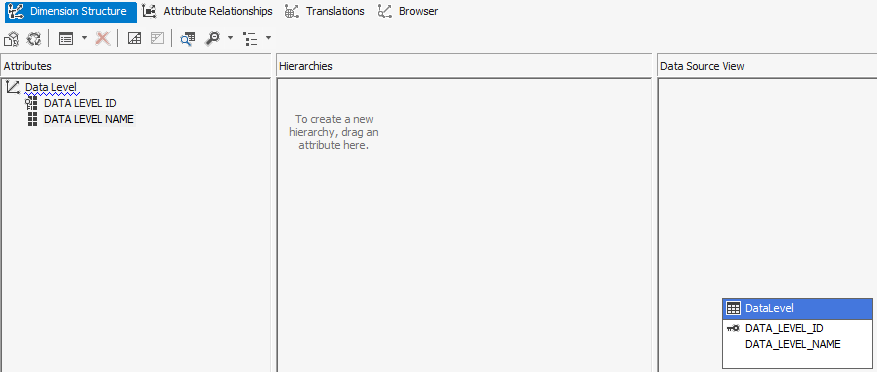


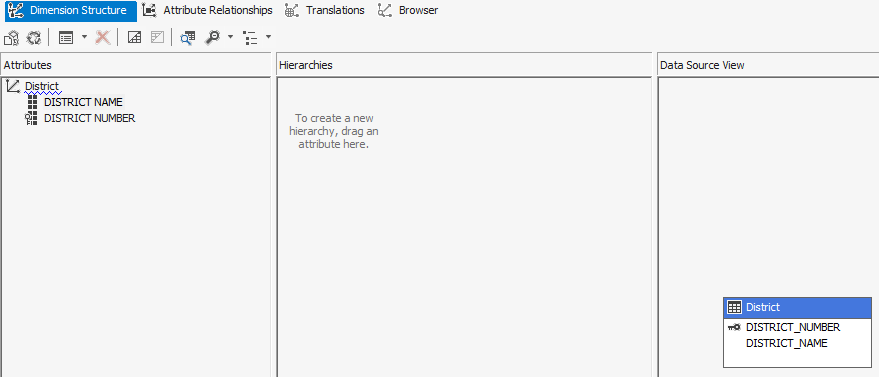
* Sau khi chạy xong ta có kết quả như hình bên dưới :

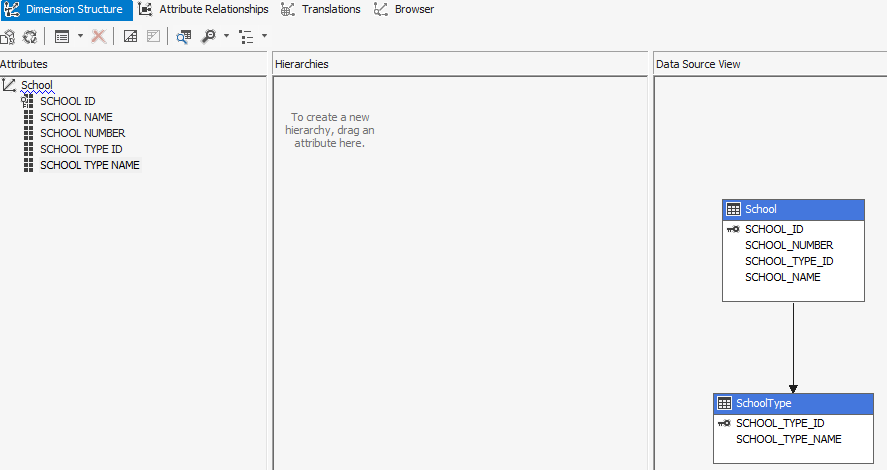


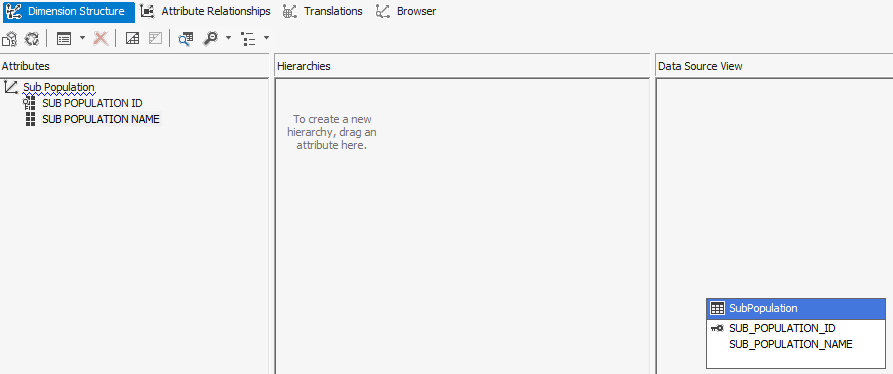
* Sau đó với từng dimension ta add thêm thuộc tính để thao tác xử lí, cụ thể ta nắm kéo thả như hình bên dưới, làm tương tự với các dimension còn lại :

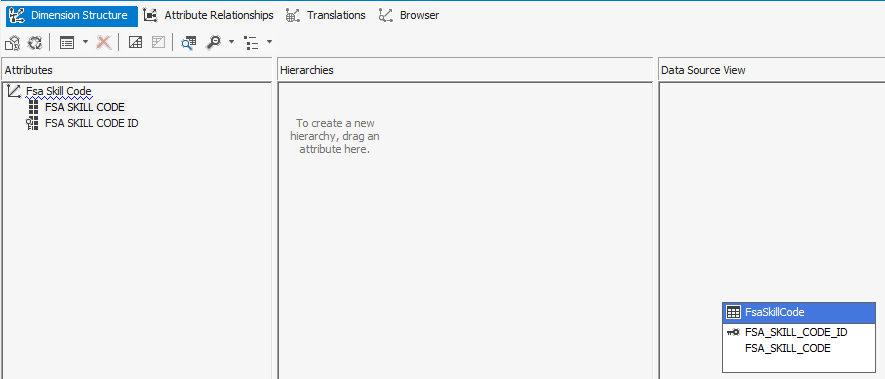












1. Áp dụng SSAS để giải quyết các câu truy vấn ở trên.