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).

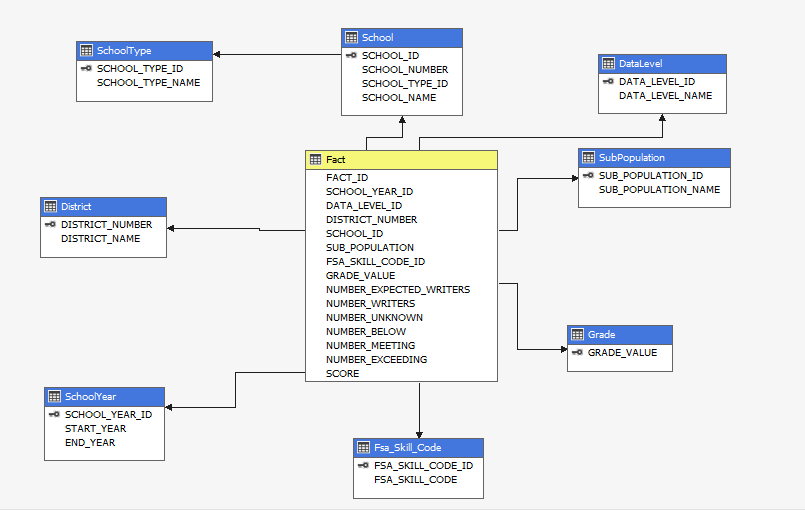
Bộ dữ liệu:Là các kết quả của BC Foundation Skills Assessment (Đánh giá kỹ năng nền tảng) của lớp 4 và 7 BC về Tính toán, Đọc và Viết từ năm 2007/2008 đến 2015/2016 trên địa bàn tỉnh (bang) British Comlumbia, phía tây Canada.

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

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

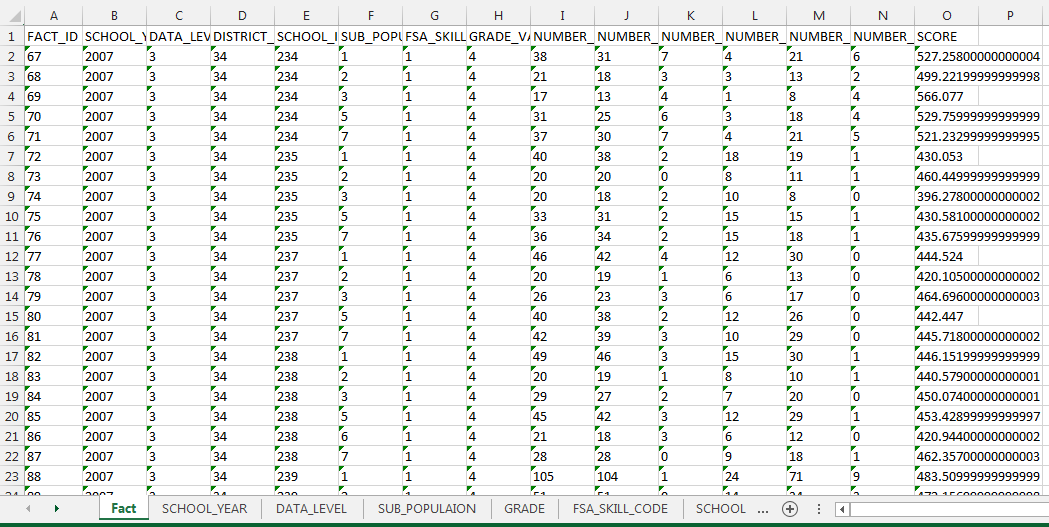
|  |  |
| --- | --- |
| Tên thuộc tính | Giải thích |
| SCHOOL\_YEAR | Năm học tiến hành khảo sát. |
| DATA\_LEVEL | Cấp độ dữ liệu, Thông tin về trường dữ liệu thuộc cấp độ nào (Tỉnh, quận, hay trường). |
| PULIC\_OR\_INDEPENDENT | Thông tin về loại trường của trường đó là trường công (pulic school) 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.

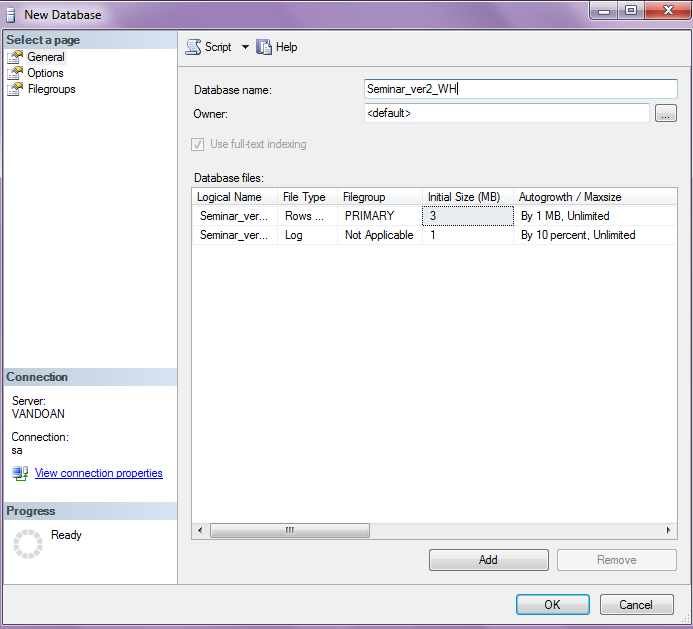


1. Quá trình Business intelligence.
2. Import với bảng fact.

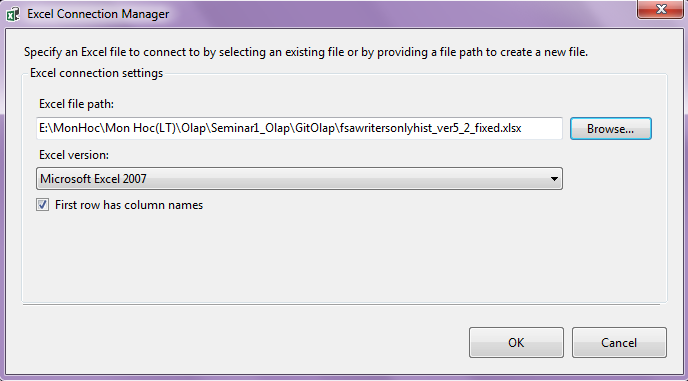
* Tiến hành xử dữ liệu, lược bỏ đi dữ liệu thừa, tách trích ra các dimension trên các sheet khác nhau để dễ dàng import dữ liệu cho quá trình BI.



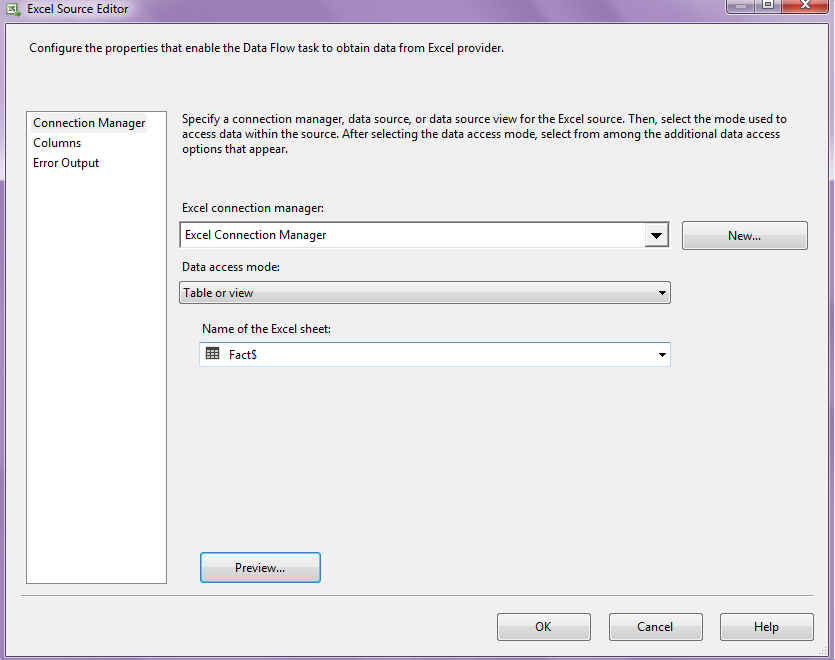
* Kết nối với hệ quản trị cơ sở dữ liệu SQL Server, tạo mới database.



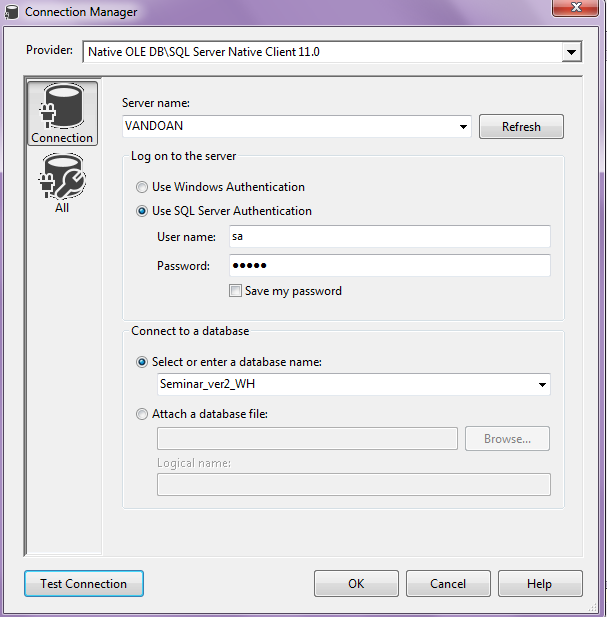
* Tạo project BI mới để import dữ liệu từ file excel ta đã xử lí từ trước.
* Chọn file excel để import.



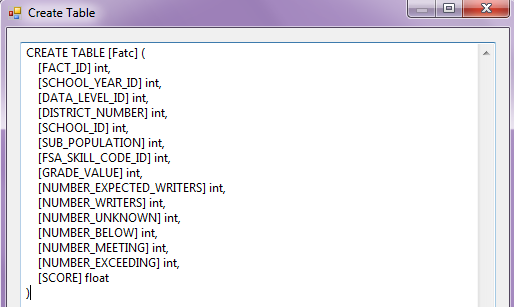
* Chọn sheet muốn import



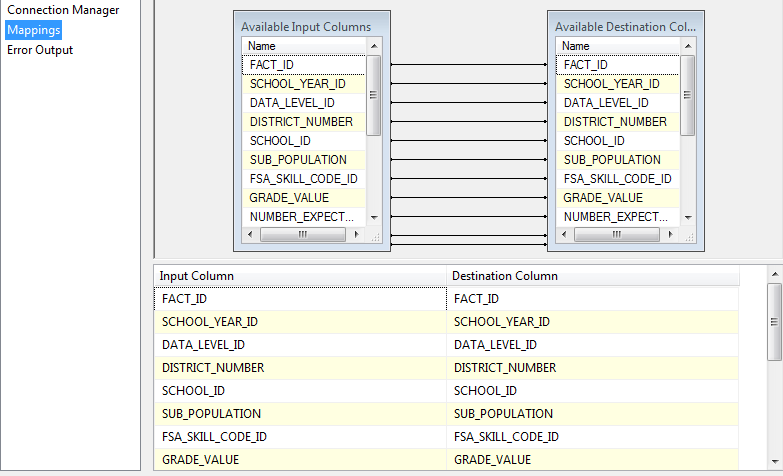
* Thực hiện chọn database để export data từ file excel



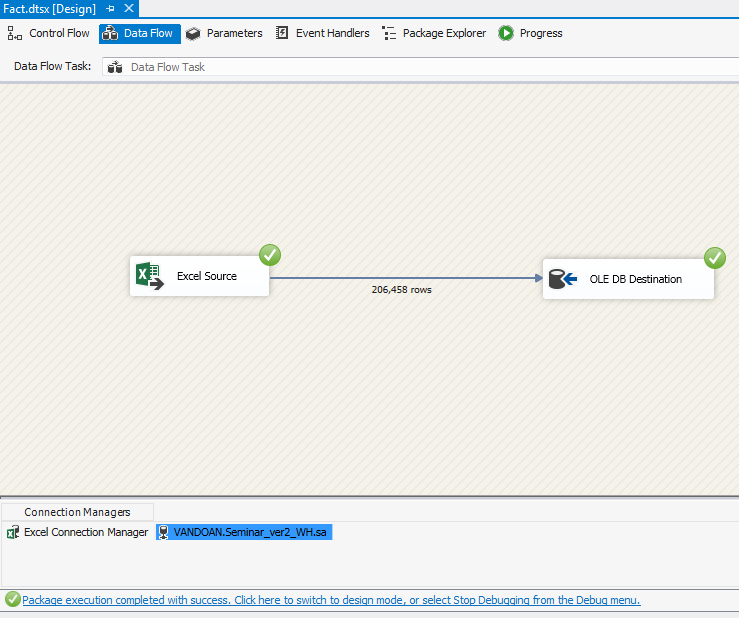
* Tùy chỉnh lại kiểu dữ liệu cho từng cột đối với bảng ta chuẩn bị tạo, ở đây là bảng Fact



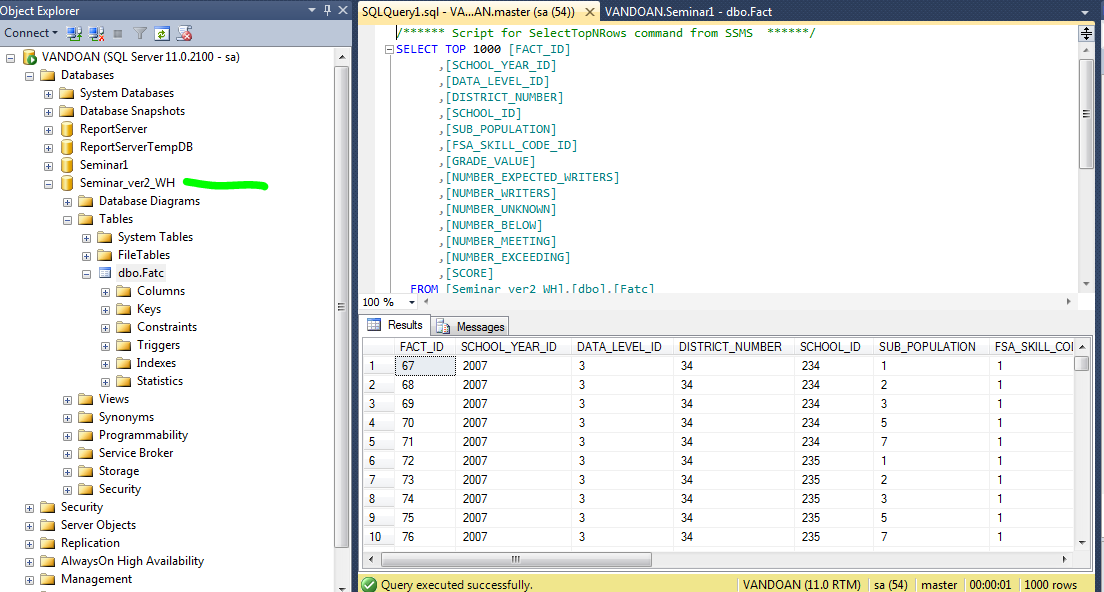
* Tiến hành mapping để đồng nhất giữa nguồn và đích của dữ liệu



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

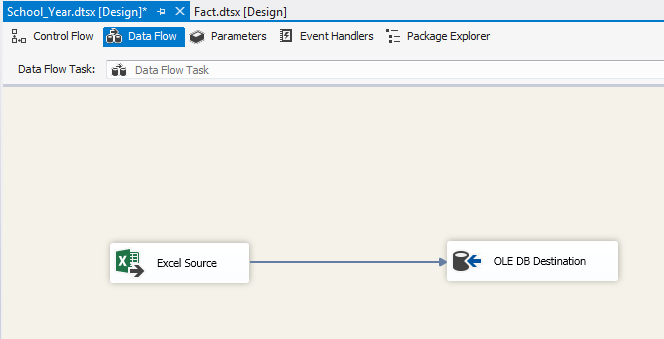


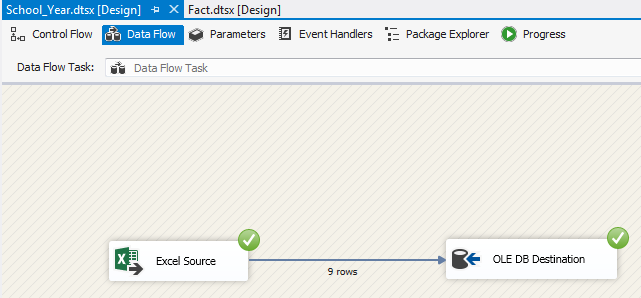
* Ta tiến hành kiểm tra lại trong SQL Server



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

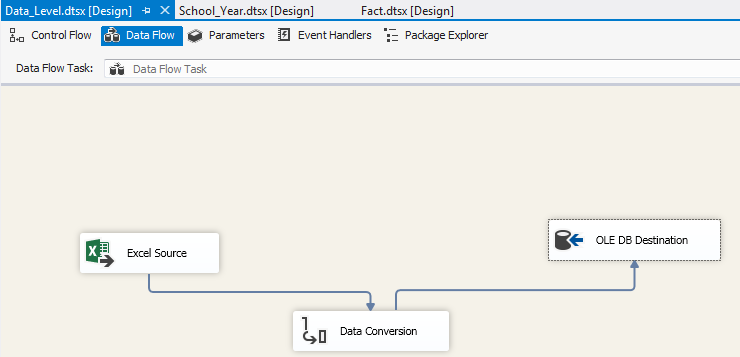
1. Import với bảng shool\_year

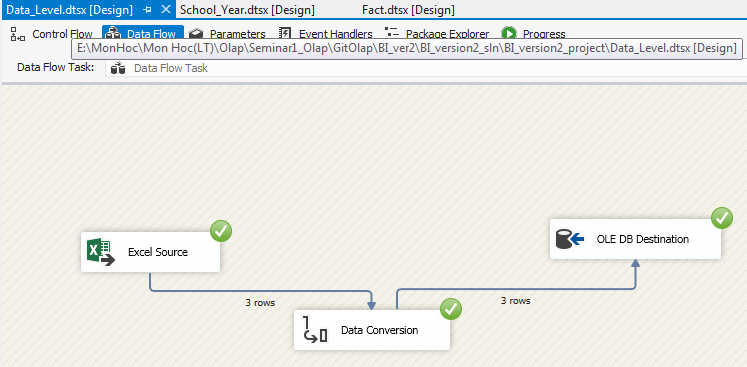




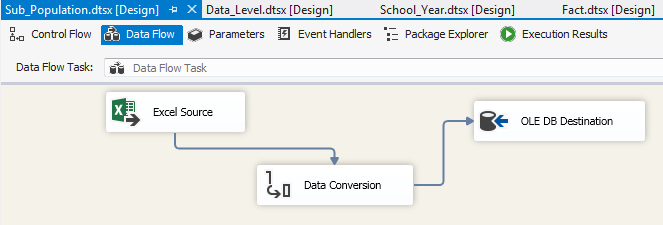
1. Import bảng Data\_level.

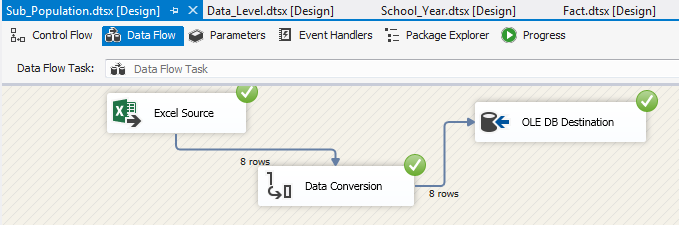
* Lưu ý : Đối với dữ liệu là kiểu varchar thì ta phải dùng bộ chuyển đổi để tránh gây ra lỗi trong lúc import dữ liệu.



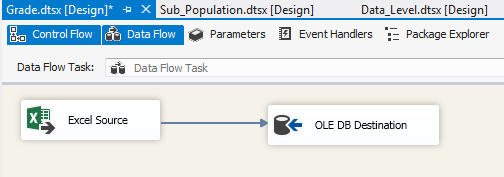


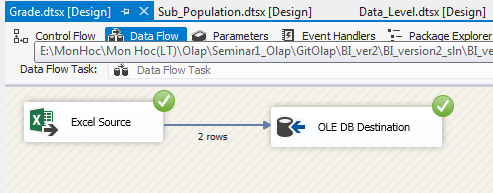
1. Import bảng sub population.



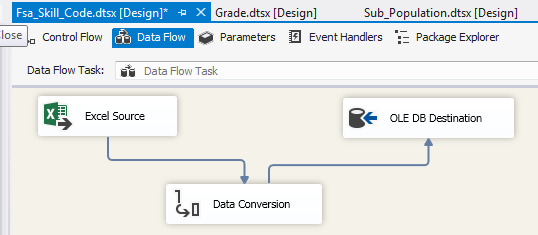


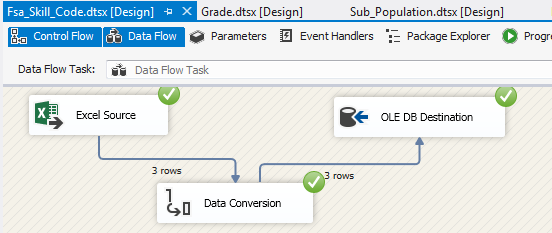
1. Import bảng Grade



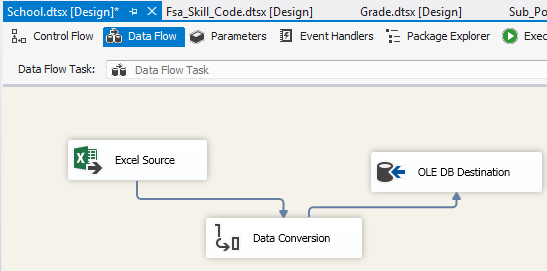


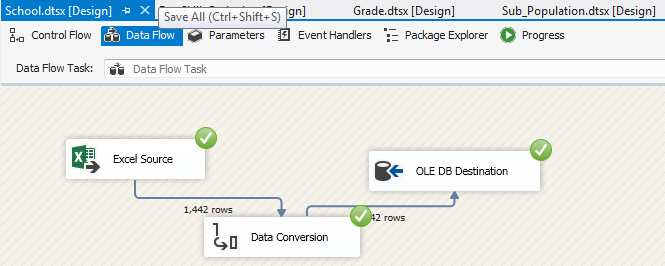
1. Import bảng Fsa skill code



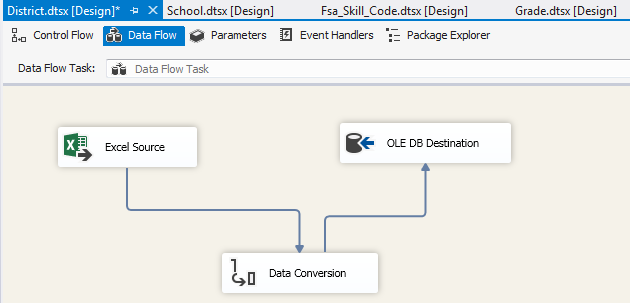


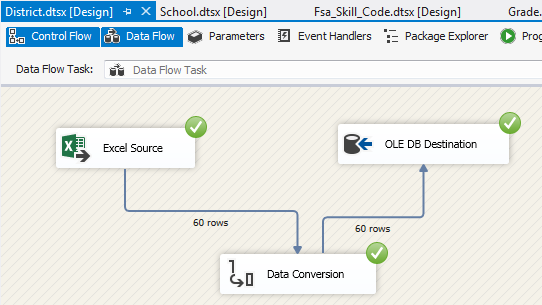
1. Import bảng school



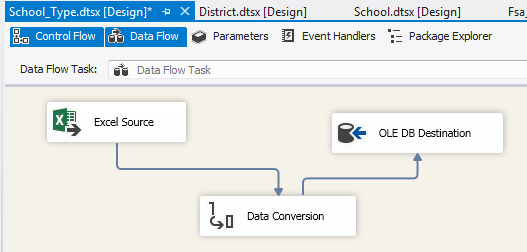


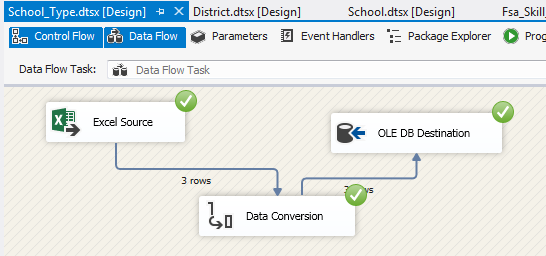
1. Import bảng district



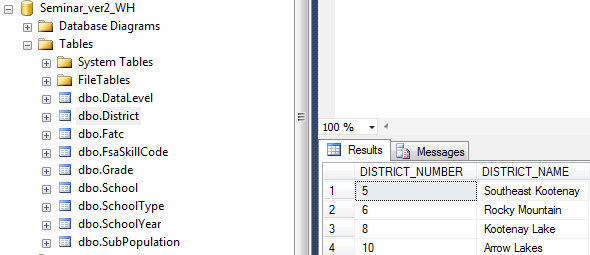


1. Import bảng school type.





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 trường học có đối tượng (SUB\_POPULATION) là “ENGLISH LANGUAGE LEARNER” và số lượng người tham dự (NUMBER WRITERS), số lượng người dự kiến (NUMBER EXPECTED WRITERS) trong năm học “2007/2008”.

Câu 2: Liệt kê số trường học của các quận ( DISTRICT\_NAME) trong năm học “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: Số lượng người tham gia (number\_writer), số lượng người đạt (NUMBER\_MEETING), số lượng người không đạt (NUMBER\_BELOW) của các trường học với đối tượng tham gia có kĩ năng (FSA\_SKILL\_CODE) là “Reading” có học tại năm học “2008/2009” và đối tượng (SUB\_POPULATION) là “ABORIGINAL”.

Câu 5: Tổng điểm số (SCORE) của tất cả các trường học (SCHOOL\_NAME) trong từng quận (DISTRICT\_NAME) với kĩ năng là tính toán (NUMERACY) và tại năm học “2007/2008”.

Câu 6: Số người vượt chỉ tiêu (NUMBER\_EXCEEDING) của từng đối tượng (sub\_population) về kĩ năng tính toán (numeracy) trong năm học “2007/2008” với những trường học có cấp độ là cấp tỉnh (Data-level : Province level).

Câu 7: Số người không đạt (NUMBER\_BELOW) và đạt (NUMBER\_MEETING) với kĩ năng là kĩ năng viết của những trường học là trường công (PUBLIC\_OR\_INDEPENDENT: BC PUBLIC SCHOOL) trong năm học “2007/2008” và cấp độ là cấp quận (DATA\_LEVEL : DISTRICT\_LEVEL).

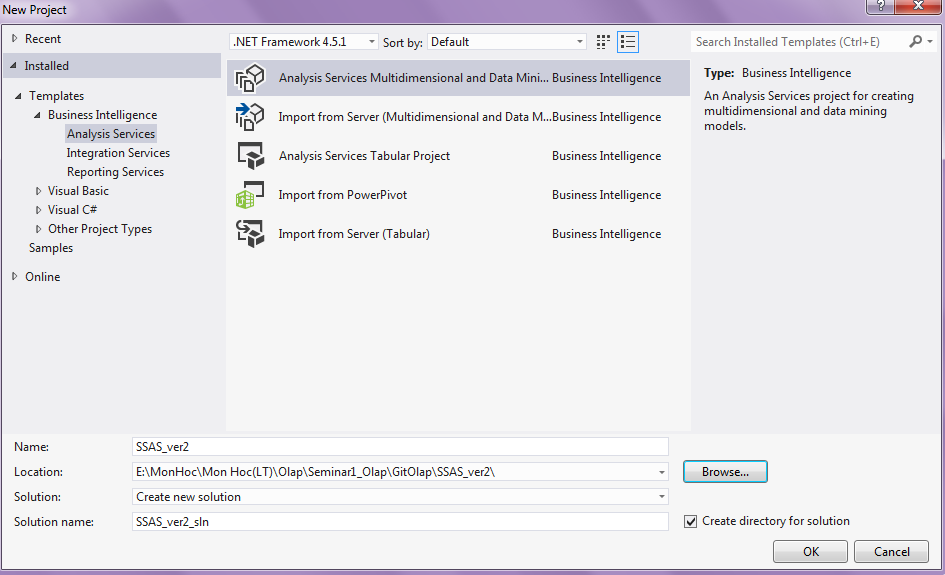
Câu 8 : Cho biết số lượng thí sinh thi không đạt với kĩ năng Reading, Writing ở tất cả các năm học.

Câu 9 Số lượng người đạt (NUMBER\_MEETING) với kĩ năng là tính toán (NUMERACY) của những trường học thuộc khối trường tư (PUBLIC\_OR\_INDEPENDENT : BC Independent School) ở khối lớp 7 và đối với người tham dự có giới tính là nam (SUB\_POPULATION: MALE) trong năm học “2014/2015”.

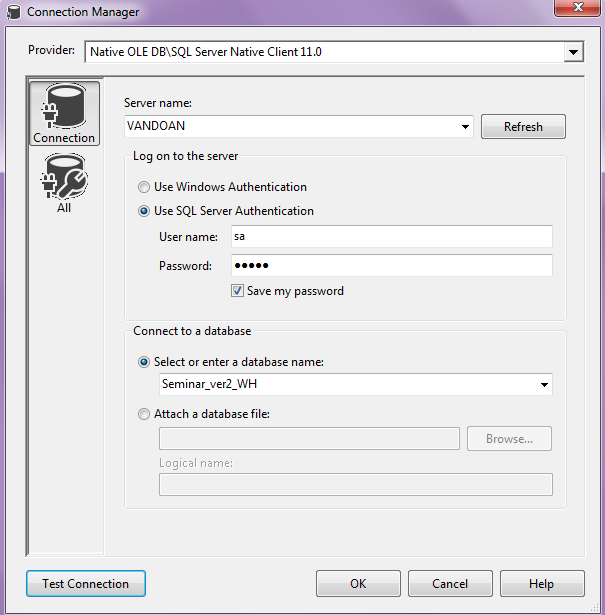
Câu 10: Điểm đánh giá của các trường học trong năm “2012/2013” ở quận New Westminster với đối tượng tham gia là ENGLISH LANGUAGE LEARNER.

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

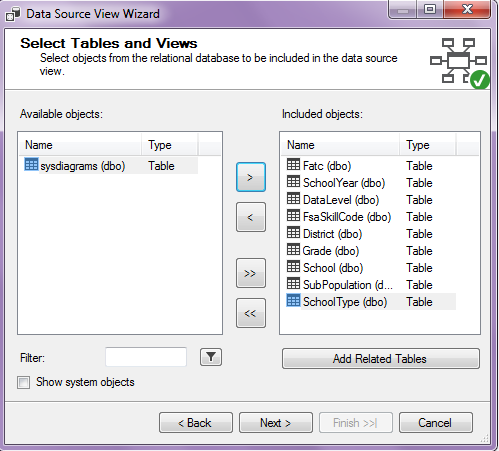
* Tạo mới project



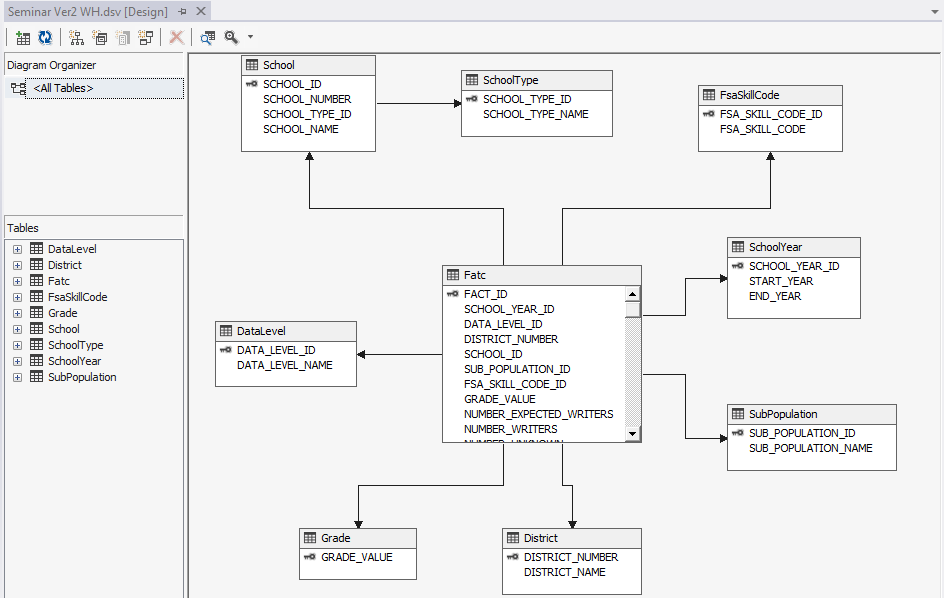
* Tạo data source.



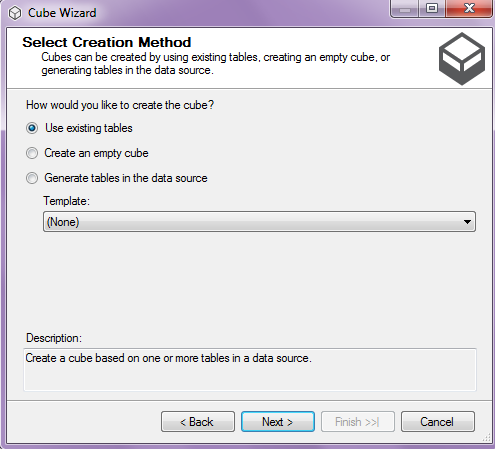
* Tạo New Data Source Views

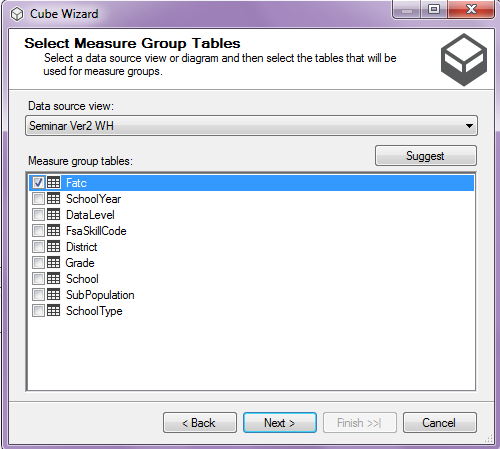


* Sau khi tạo xong thì ta có kết quả như hình dưới :

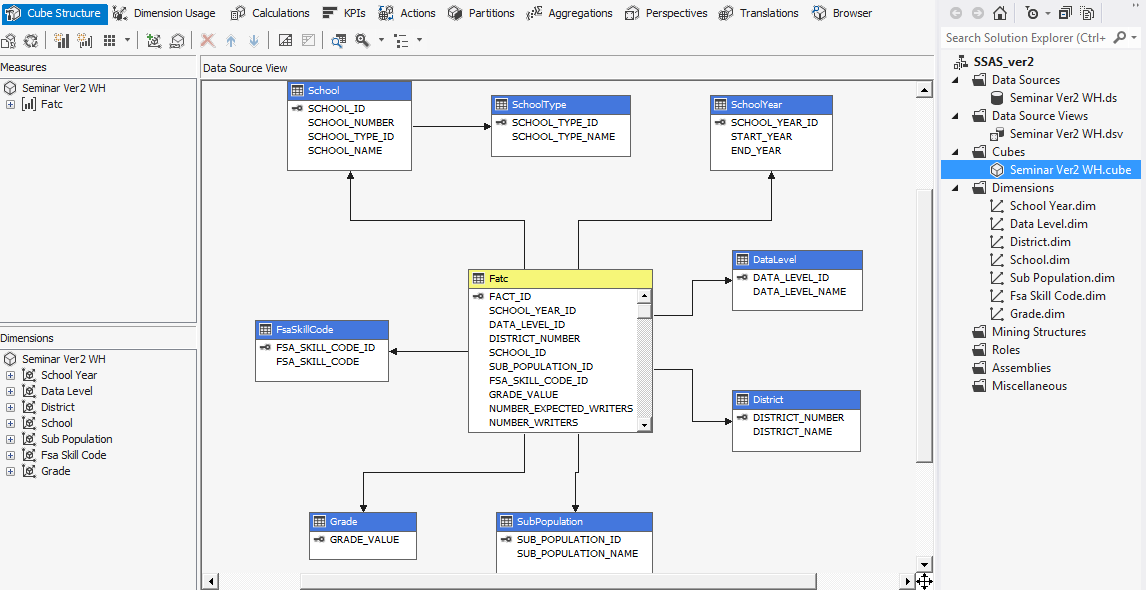


* Tạo mới Cube

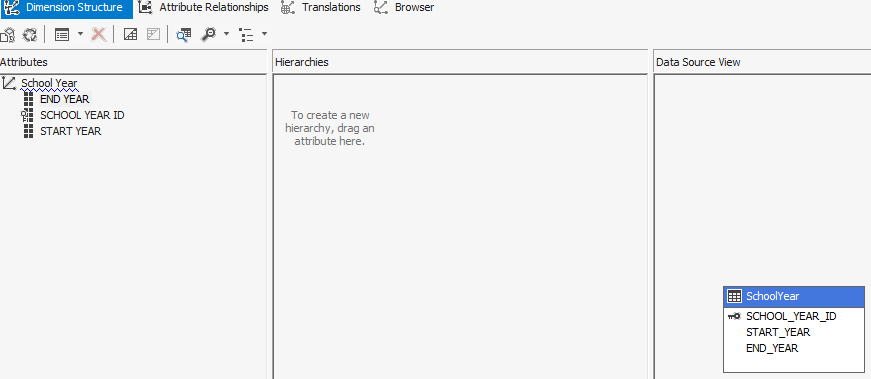


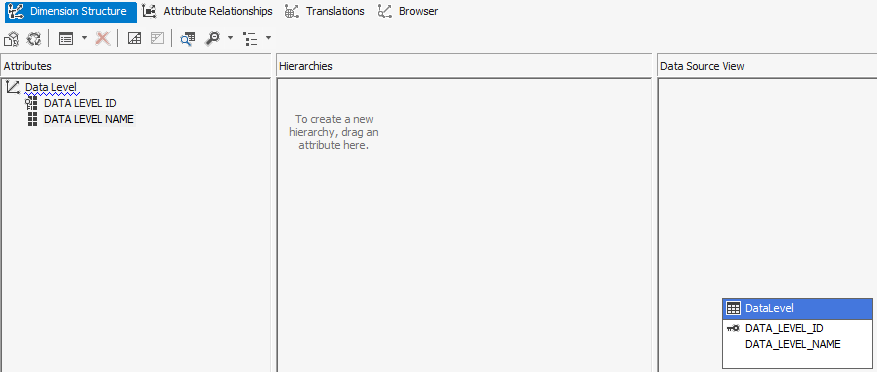


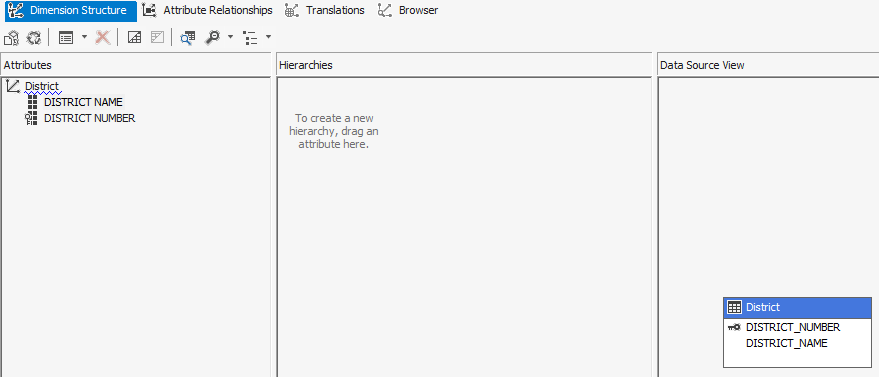
* Sau khi chạy xong ta có kết quả như hình bên dưới :

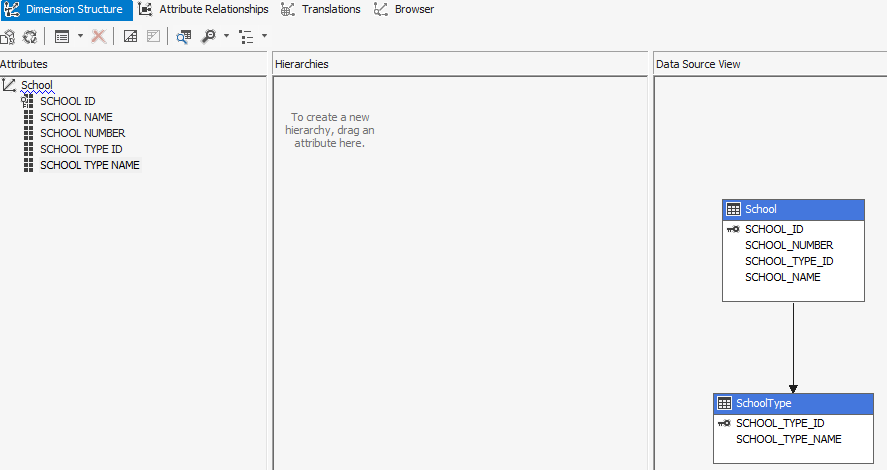


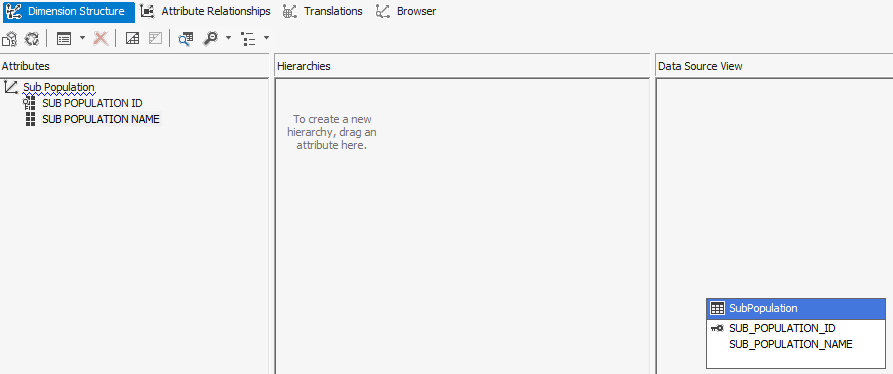
* Sau đó với từng dimension ta add thêm thuộc tính để thao tác xử lí, cụ thể ta nắm kéo thả như hình bên dưới, làm tương tự với các dimension còn lại :

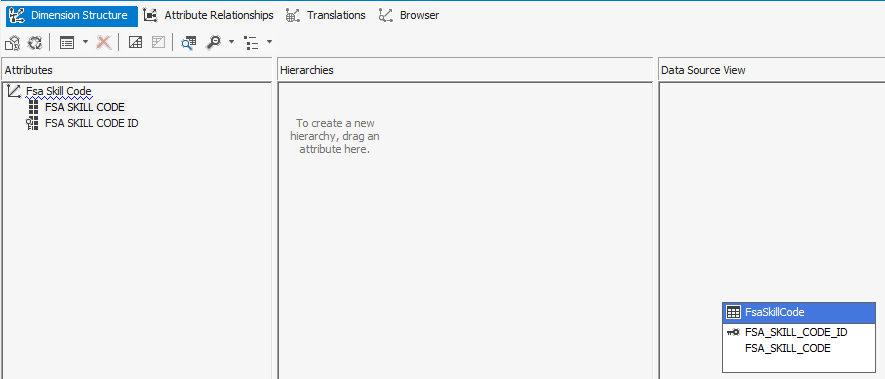






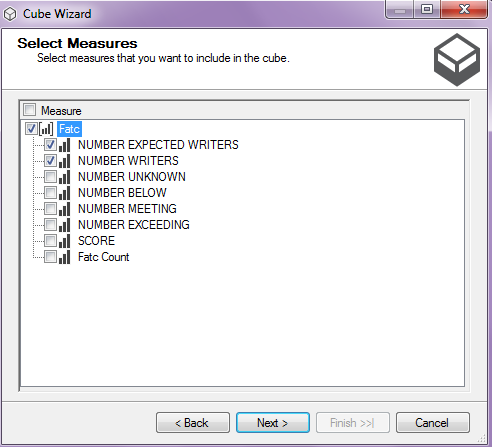




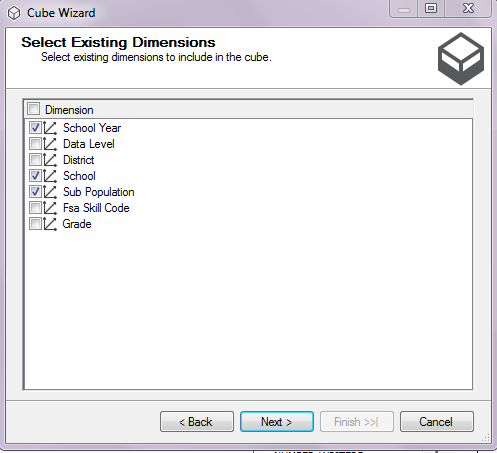


1. Áp dụng SSAS để giải quyết các câu truy vấn ở trên.
   1. Liệt kê trường học có đối tượng (SUB\_POPULATION) là “ENGLISH LANGUAGE LEARNER” và số lượng người tham dự (NUMBER WRITERS), số lượng người dự kiến (NUMBER EXPECTED WRITERS) trong năm học “2007/2008”.

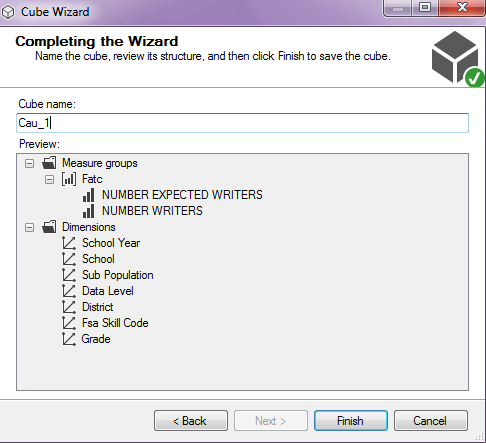
* Chọn những cột cần thiết.



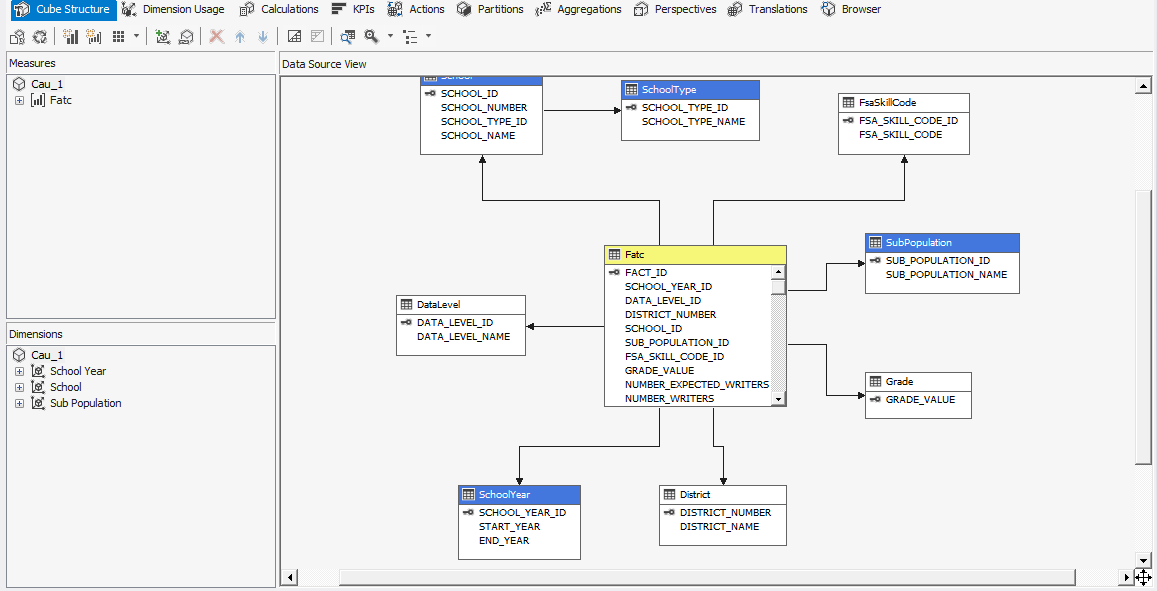
* Chọn các dimension cần thiết, ở đây là school\_year, sub\_population, school



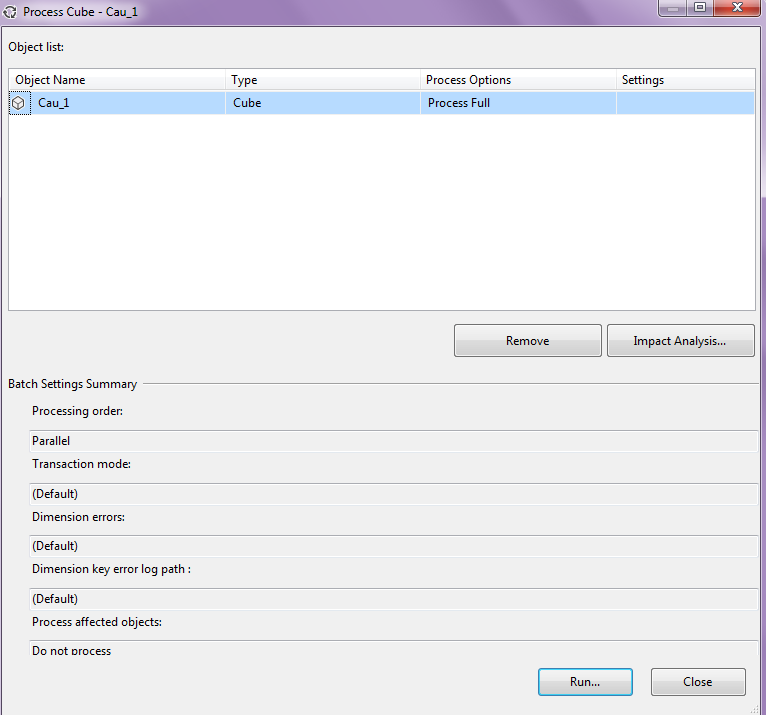
* Chọn finish để hoàn thành việc tạo cube xử lí cho câu truy vấn.



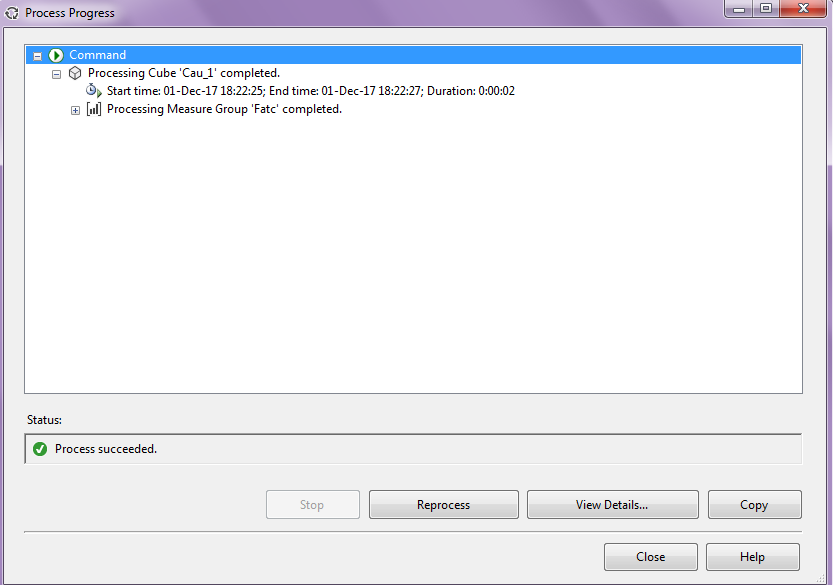
* Tiến hành xóa đi những dimension không cần thiết, kết quả title sẽ thành màu trắng.



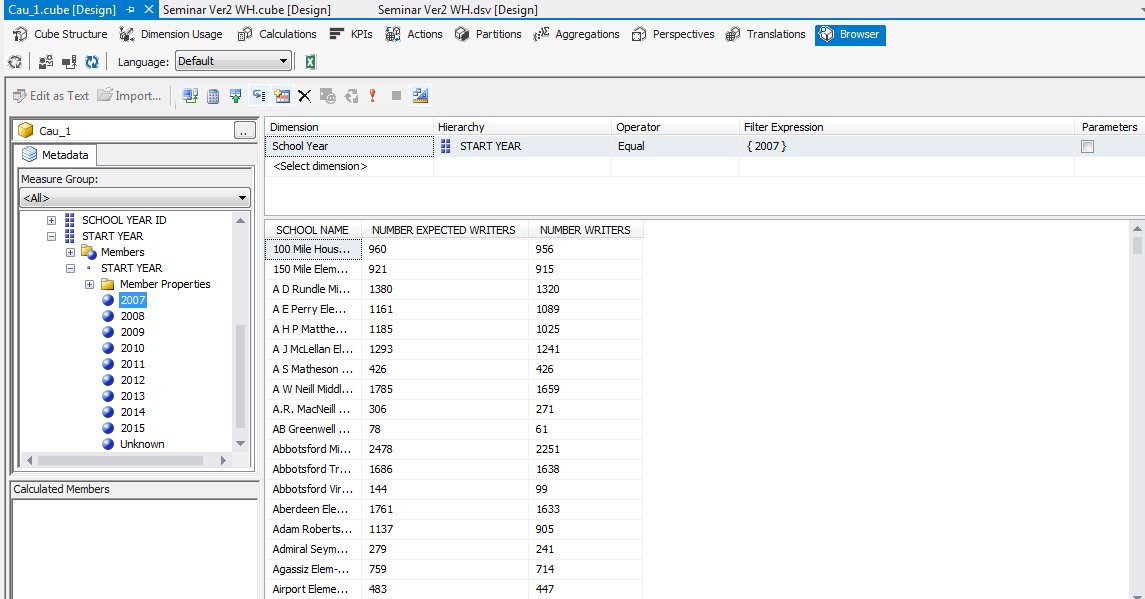
* Khởi chạy cube



* Màn hình thông báo thành công.

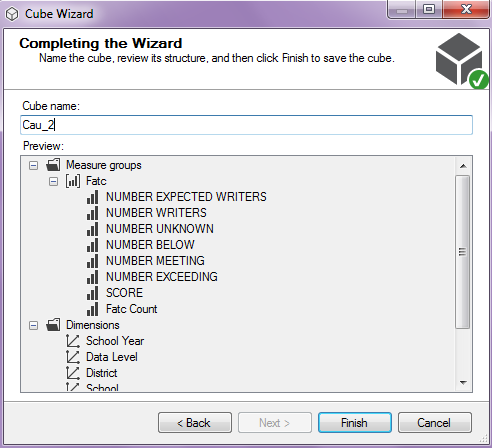


* Kết quả truy vấn câu 1 ở tab brower, trong đó bộ lọc theo năm là 2007 vì trong database thì “2007/2008” sẽ được chia ra thành năm bắt đầu và năm kết thúc : start\_year = 2007 and end\_year = 2008.

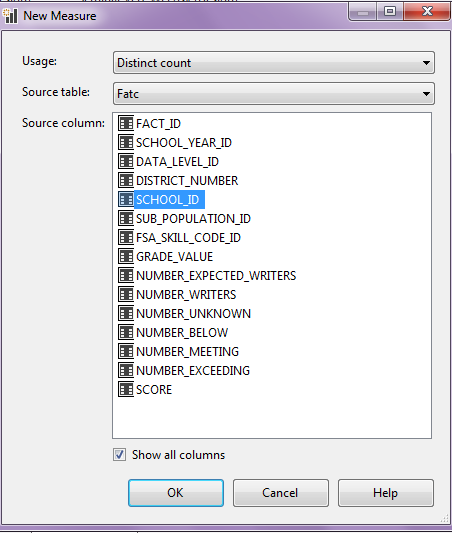


Câu 2. Liệt kê số trường học của các quận ( DISTRICT\_NAME) trong năm học “2009/2010”.

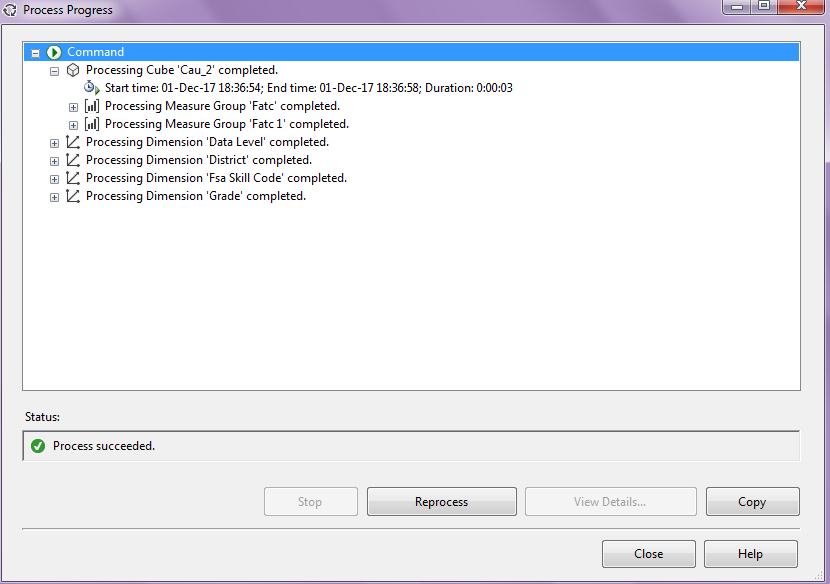
* Khởi tạo cube



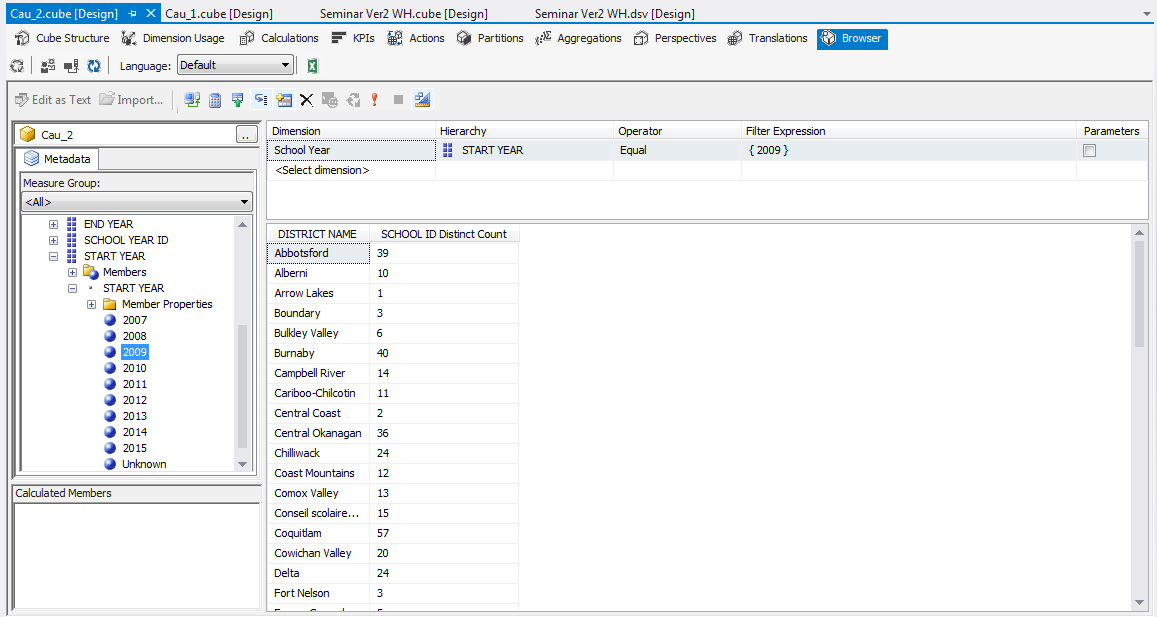
* Định nghĩa measure đếm số trường học, ở đây những trường học nào đã đếm rồi thì không đếm nữa.



* Chạy process

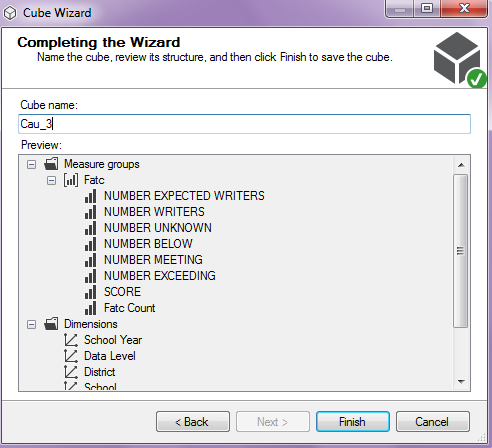


* Kết quả thực hiện truy vấn.

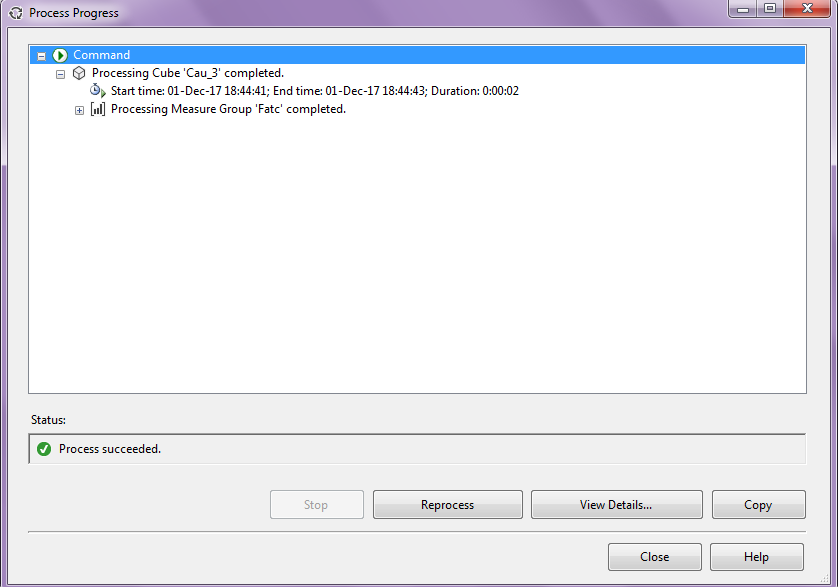


Câu 3. Số lượng người tham dự (NUMBER\_WRITERS) của khối (GRADE) 7 ở tất cả các trường trong năm học “2007/2008” tại quận (DISTRICT\_NAME) “Sea to Sky”.

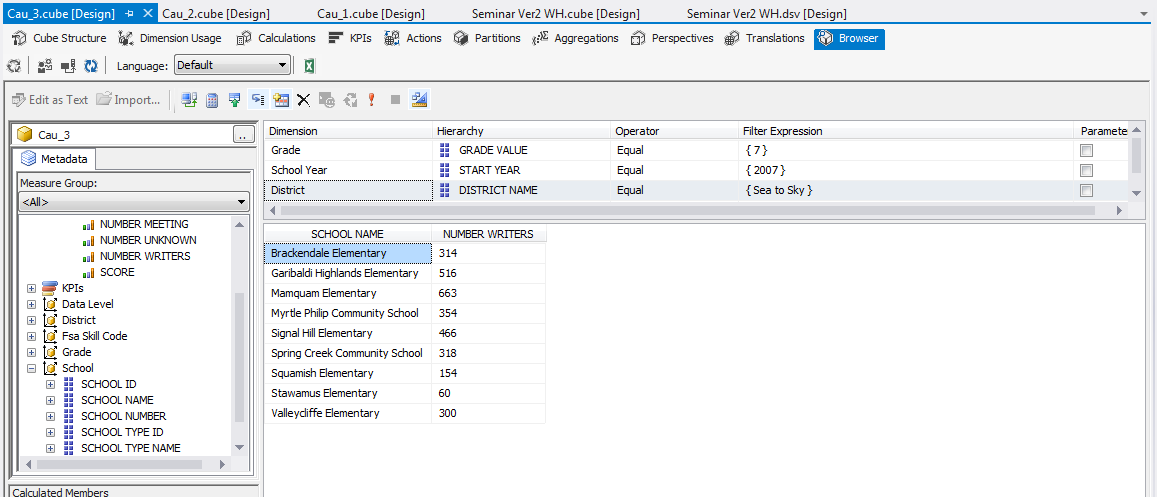
* Tạo mới cube



* Chạy process

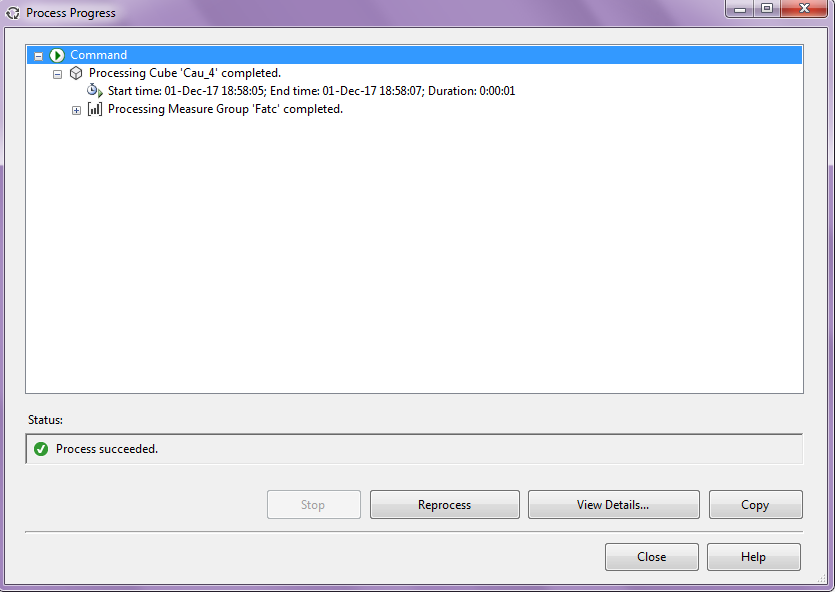


* Kết quả thực hiện truy vấn.

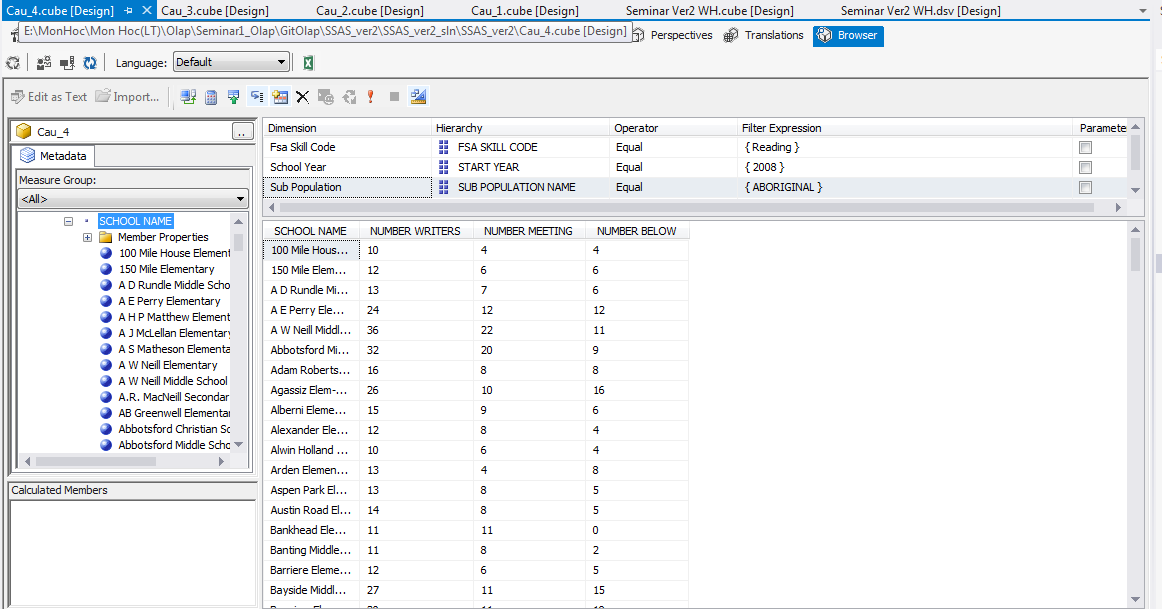


Câu 4 Số lượng người tham gia (number\_writer), số lượng người đạt (NUMBER\_MEETING), số lượng người không đạt (NUMBER\_BELOW) của các trường học với đối tượng tham gia có kĩ năng (FSA\_SKILL\_CODE) là “Reading” có học tại năm học “2008/2009” và đối tượng (SUB\_POPULATION) là “ABORIGINAL”.

* Khởi tạo cube

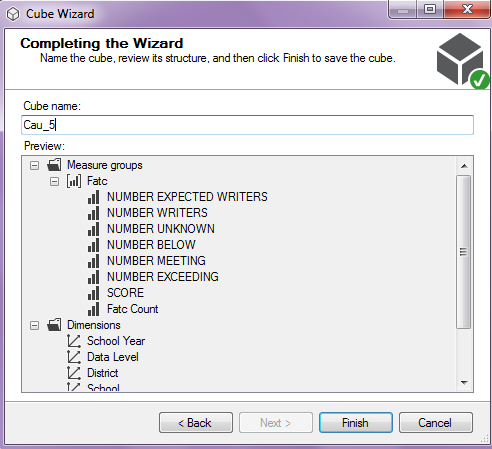


* Kết quả truy vấn.



Câu 5: Tổng điểm số (SCORE) của tất cả các trường học (SCHOOL\_NAME) trong từng quận (DISTRICT\_NAME) với kĩ năng là tính toán và tại năm học “2007/2008”.

* Khởi tạo cube. Ở đây ta có sẵn measure score hay là total\_score rồi nên ta không cần tạo mới measure nữa



* Chạy process và chờ báo kết quả thành công. Sau đó thực hiện truy vấn.

