

**UNIVERSITY OF SCIENCE**

Ho Chi Minh City

**INFORMATION OF TECHNOLOGY**

High quality

# **SEMINAR PROJECT**

Under the course of

**CSC12107 – Business Intelligence**

Semester I

*Submitted by:*

*Group: 19HTT1.BI01*

- |    |                       |          |
|----|-----------------------|----------|
| 1. | Trần Ngọc Lam         | 19127040 |
| 2. | Nguyễn Phương Vy      | 19127088 |
| 3. | Nguyễn Kim Thị Tố Nga | 19127219 |

**Tiết Gia Hồng**

**Hồ Thị Hoàng Vy**

---

*Academic year*  
*(2019-2023)*

## Mục lục

<b>I. Tóm tắt</b>	<b>4</b>
A. Nêu vấn đề	4
B. Monitoring	4
<b>II. Các ý tưởng để monitoring ETL</b>	<b>5</b>
A. Giám sát các tác vụ vận chuyển dữ liệu bên ngoài	5
B. Giám sát các tác vụ ETL	5
C. Chủ động giám sát	6
D. Kiểm tra sự bất thường trong dữ liệu	6
E. Giám sát các thông số của cơ sở dữ liệu	7
<b>III. Kết hợp monitoring với AI Machine Learning</b>	<b>7</b>
A. Phát hiện và cảnh báo thông qua các chỉ số của ETL	7
B. Xác định các tắc nghẽn cụ thể	8
C. Định lượng các tác động của thay đổi trong việc quản lý dữ liệu	8
D. Giảm chi phí vận hành	8
<b>IV. Demo Monitoring ETL</b>	<b>9</b>
A. Logging	9
1. SSIS log provider for SQL Text files	13
2. SSIS log provider for SQL Server	16
3. SSIS log provider for XML files	19
B. Built-in Reporting	23
1. Package report	23
2. Dashboard report	26
C. Views	28
<b>V. Tài liệu tham khảo</b>	<b>29</b>

## Thông tin thành viên và Phân Công

**Nhóm: 19HTTT1.BI01**

MSSV	Họ và tên	Phân công	Đóng góp
19127040	Trần Ngọc Lam	<ul style="list-style-type: none"><li>- Tìm hiểu monitoring</li><li>- Tìm hiểu demo</li><li>- Làm slide thuyết trình</li></ul>	33.33%
19127088	Nguyễn Phương Vy	<ul style="list-style-type: none"><li>- Tìm hiểu monitoring</li><li>- Tìm hiểu demo</li><li>- Làm báo cáo</li></ul>	33.33%
19127219	Nguyễn Kim Thị Tố Nga	<ul style="list-style-type: none"><li>- Tìm hiểu monitoring</li><li>- Tìm hiểu demo</li><li>- Quay clip demo</li></ul>	33.33%

## I. Tóm tắt

### A. Nêu vấn đề

Các quy trình ETL đóng vai trò như một lớp ngăn cách trước khi dữ liệu đi vào data mart hoặc data warehouse. Quy trình ETL rất phức tạp do phải xử lý các định dạng dữ liệu đầu vào khác nhau, khối lượng dữ liệu khác nhau,...

VD: Giả sử chúng ta đang thực hiện chuyển đổi dữ liệu từ 2 cơ sở - một cơ sở ở Mỹ, một cơ sở ở Việt Nam. Cơ sở ở Mỹ sử dụng định dạng ngày là MM/DD/YYYY, và cơ sở ở Việt Nam sử dụng định dạng ngày là DD/MM/YYYY. Sau khi tập dữ liệu đang được chuyển đổi, định dạng ngày thay đổi. Khi đó công cụ ETL sẽ không biết được rằng đã có một sự thay đổi, và tất cả dữ liệu của chúng ta sẽ bị xáo trộn.

#### ❖ *Các vấn đề có thể gặp phải khi ETL xảy ra sai sót:*

- Dữ liệu không được tải vào cơ sở dữ liệu
- Dữ liệu được tải vào không chính xác
- Dữ liệu được sử dụng để đưa ra report hoặc quyết định không chính xác do hệ thống nguồn có vấn đề hoặc dữ liệu vận chuyển không hoàn chỉnh và chưa chính xác.

### B. Monitoring

Quy trình ETL bị lỗi có thể gây ra thông tin bị nhiễu hoặc khiến cho quá trình chuyển đổi bị trì trệ. Đã có rất nhiều hệ thống quản lý cơ sở dữ liệu đưa ra nhiều cách khác nhau để monitor các quy trình ETL, như các cơ chế để kiểm tra từng bước riêng lẻ của ETL.

Tuy nhiên, để có thể quan sát và theo dõi đầy đủ tất cả các vấn đề về phân phối dữ liệu từ hệ thống bên ngoài, hoặc là dữ liệu không chính xác, chúng ta cần một công cụ monitor trong các quy trình ETL.

Ở trong bài tìm hiểu này, nhóm sẽ thực hiện monitor quá trình ETL bằng SQL Server:

#### ❖ **Logging**

Event logs rất cần thiết cho người quản trị cơ sở dữ liệu. Nó giúp chúng ta khắc phục sự cố hoặc hiểu rõ hơn về các quy trình đang chạy.

**Integration Services** gồm các loại log providers sau:

- **Text File log provider:** ghi các log vào file văn bản ASCII với định dạng csv. Phần mở rộng tên tệp có đuôi **.log**.
- **SQL Server Profiler log provider:** ghi dấu và có thể xem bằng SQL Server Profiler. Phần mở rộng tên tệp có đuôi là **.trc**.

- **SQL Server log provider:** ghi các log vào bảng *sysssislog* trong một cơ sở dữ liệu SQL Server
- **Window Event log provider:** ghi các mục vào log của ứng dụng trong log của Window Event trên máy tính cục bộ.
- **XML File log provider:** ghi các log vào một file XML. Phần mở rộng của tên tệp có đuôi là *.xml*.

#### ❖ Build-in Reporting

Reporting sẽ được tích hợp vào SSIS trong SQL Server. Ta không cần tự tạo bất kì bản report nào cả, trừ khi ta muốn tạo một bản report tùy chỉnh từ SSIS hoặc ta muốn một cơ sở dữ liệu SSIS auditing trung tâm.

Việc audit được tích hợp vào SSIS sẽ nằm trên máy chủ SSIS mà SSIS đang chạy, do đó nếu ta có 10 máy chủ khác nhau, ta phải xem report của từng máy chủ một.

#### ❖ Views

Truy vấn các Integration Services view để quan sát các đối tượng, các cài đặt, và dữ liệu thực thi được lưu trữ trong SSISDB catalog.

## II. Các ý tưởng để monitoring ETL

Trước khi đi sâu vào các ý tưởng này, thì thông thường để kiểm tra các quy trình ETL ta sẽ: kiểm tra luồng dữ liệu giữa 2 cơ sở dữ liệu hoặc 2 bảng, và thực hiện kiểm tra tính chính xác của dữ liệu. Hai nguyên tắc này có thể được thực hiện theo nhiều cách khác nhau, tùy thuộc vào kiến trúc hoặc môi trường dữ liệu.

### A. Giám sát các tác vụ vận chuyển dữ liệu bên ngoài

Do dữ liệu đến từ nhiều nguồn khác nhau nên việc giám sát các tác vụ trong các nguồn dữ liệu này là vô cùng cần thiết.

Các mục tiêu giám sát bao gồm:

- Thời gian thực hiện vận chuyển dữ liệu
- Cách hoạt động và hiệu quả hoạt động của các tác vụ tại nguồn.

### B. Giám sát các tác vụ ETL

Tất nhiên, ta cũng nên giám sát các tác vụ ETL. Cũng giống như đối với các tác vụ bên ngoài, ta có thể đo thời gian thực hiện và việc tác vụ có chạy hay không ở một thời điểm cụ thể.

### C. Chủ động giám sát

Chúng ta cũng có thể dự đoán các vấn đề xảy ra trong tương lai bằng cách dành thời gian quan sát, phân tích kiến trúc và mô hình, cũng như xác định được các dấu hiệu tiềm ẩn như sau:

- **Khối lượng dữ liệu được xử lý bởi một công việc cụ thể đang tăng lên theo thời gian.**

VD: Giả sử số lượng hồ sơ đến từ hệ thống CRM đang ngày càng lớn hơn trong vài tháng vừa qua. Nếu số lượng cứ tiếp tục tăng như vậy, phần cứng mà cơ sở dữ liệu đang chạy có thể cần phải được cải thiện hoặc công việc xử lý dữ liệu cần được xem xét phân chia.

- **Một công việc luôn mất nhiều thời gian để hoàn thành hơn so với trước đây**

Đây cũng có thể là một dấu hiệu cho thấy khối lượng dữ liệu đang tăng lên hoặc khả năng xử lý không đủ mạnh.

- **Một cơ sở dữ liệu đang phát triển lớn dần gây ra mất tương xứng theo thời gian**

Nếu phát hiện ra vấn đề này,, cần đưa ra các giả định về tình hình sẽ xảy ra trong tương lai gần và thực hiện các bước nếu cần thiết. Hơn nữa, ta có thể phát hiện các sự cố tiềm ẩn trước khi nó xảy ra và chủ động khắc phục nếu thường xuyên theo dõi.

### D. Kiểm tra sự bất thường trong dữ liệu

Sự bất thường của dữ liệu sẽ xảy ra ở tình huống như:

Khi ban quản lý bất ngờ nhận được báo cáo từ hệ thống trí tuệ kinh doanh cho thấy doanh số bán hàng tăng đột biến mà không giải thích được, đây có thể là một điều đáng mừng hoặc đã đến lúc kiểm tra lại tính đúng đắn của dữ liệu. Thông thường, những thay đổi lớn trong các báo cáo không có lời giải thích cho thấy có thể đã xảy ra sự cố với dữ liệu. Điều này sẽ trở nên rõ ràng hơn nếu ai đó nhận thấy rằng các số liệu có vẻ không chính xác.

Một vài công cụ giám sát như **PRTG Network Monitoring**, cung cấp phương tiện giám sát số lượng bản ghi hoặc tập dữ liệu. Ta có thể đặt cảnh báo ngưỡng cho các giá trị tối thiểu hoặc tối đa dựa trên dữ liệu lịch sử. Nếu số lượng bản ghi vượt quá ngưỡng này thì một cảnh báo sẽ được tạo ra. Dữ liệu sau đó có thể được xem xét kỹ hơn để đảm bảo rằng dữ liệu là chính xác.

## E. Giám sát các thông số của cơ sở dữ liệu

Ngoài theo dõi các quy trình ETL, việc theo dõi các chỉ số thông thường của cơ sở dữ liệu như thời gian phản hồi của cơ sở dữ liệu, khả năng kết nối,... là vô cùng cần thiết vì đây đều là những chỉ số quan trọng và cho ta cái nhìn toàn diện của vấn đề.

## III. Kết hợp monitoring với AI Machine Learning

Ta có thể chỉ định ai đó thực hiện việc giám sát quá trình ETL nhưng thật ra điều đó không đơn giản do một số vấn đề như:

- Dữ liệu xấu do lỗi của quy trình xảy ra quá nhanh đến mức không thể nhận thấy chúng trong thời gian thực.
- Kết quả của quy trình ETL bị hỏng không khác với dữ liệu chính xác.
- Ngay cả khi lỗi được thể hiện rõ ràng, vấn đề khiến quá trình bị lỗi có thể không dễ dàng theo dõi.

Đó là lý do ta cần một công cụ giám sát để hợp lý hóa cũng như tự động hóa quy trình.

### ❖ Một số lợi ích khi kết hợp giám sát quy trình ETL với AI Machine Learning:

- Máy học AI thì không bao giờ nghỉ ngơi, nghĩa là chúng liên tục theo dõi dữ liệu đi vào thông qua các công cụ ETL một cách tự động.
- Khi nó phát hiện ra những điểm bất thường trong thông tin, nó có thể được lập trình để cảnh báo cho các bộ phận cụ thể hoặc kiến trúc sư của hệ thống về những vấn đề cần phải khắc phục.
- Với việc giám sát được thực hiện tại chỗ, dữ liệu có thể trở nên chính xác hơn.

**Một số chức năng của AI trong việc quan sát và theo dõi quá trình ETL:**

### A. Phát hiện và cảnh báo thông qua các chỉ số của ETL

Mặc dù dữ liệu là một quá trình thay đổi liên tục, quá trình ETL vẫn phải tạo ra các giá trị nhất quán với tốc độ nhất quán. Khi những chỉ số này thay đổi, nó là nguyên nhân gây ra các cảnh báo.

Con người có thể nhận ra những thay đổi lớn trong dữ liệu và phát hiện ra lỗi, nhưng máy học có thể nhận ra các lỗi nhỏ hơn và nhanh hơn. Hệ thống máy học có thể cung cấp khả năng phát hiện bất thường theo thời gian thực và cảnh báo trực tiếp cho bộ phận công nghệ thông tin, cho phép họ tạm dừng quy trình và khắc phục sự cố mà không cần phải bỏ hàng giờ nỗ lực tính toán.

## **B. Xác định các tác nghẽn cụ thể**

Mặc dù các kết quả của quá trình chính xác nhưng quá trình diễn ra có thể quá chậm để được đưa vào sử dụng thực tế. Khoảng 80% thông tin thu được từ phân tích sẽ được khai thác để tạo ra các giá trị tiền tệ, điều này có thể do các nhà lãnh đạo doanh nghiệp không thể nhìn thấy các thông tin chi tiết kịp thời để tận dụng nó. Công nghệ máy học có thể cho biết hệ thống đang chậm lại ở đâu và cung cấp cho câu trả lời ngay lập tức, giúp ta có được dữ liệu tốt hơn, nhanh hơn.

## **C. Định lượng các tác động của thay đổi trong việc quản lý dữ liệu**

Các hệ thống sửa đổi và nâng cấp dữ liệu và phân tích một cách liên tục. Đôi khi, những điều này cũng có thể ảnh hưởng đến quy trình dữ liệu được tạo ra và giải thích dữ liệu, dẫn đến kết quả không chính xác. AI có thể đánh dấu các thay đổi trong kết quả và theo dấu nó tới các máy hoặc ứng dụng cụ thể tương ứng

## **D. Giảm chi phí vận hành**

Hoạt động phân tích bị đình trệ và thời gian dùng để tìm ra và giải quyết vấn đề có thể gây ra tổn thất lớn trong tiền tệ cũng như thời gian và nhân lực.

Công nghệ máy học giúp đi vào trọng tâm vấn đề bằng cách chỉ cảnh báo cho các nhóm có thể chịu trách nhiệm với các loại sự cố cụ thể, để những người còn lại trong bộ phận công nghệ thông tin có thể thực hiện tiếp tục các công việc khác. Ngoài ra, máy học sẽ giúp loại bỏ các thông tin sai, giảm số lượng cảnh báo đồng thời tăng mức độ chi tiết của thông tin mà chúng có thể cung cấp.



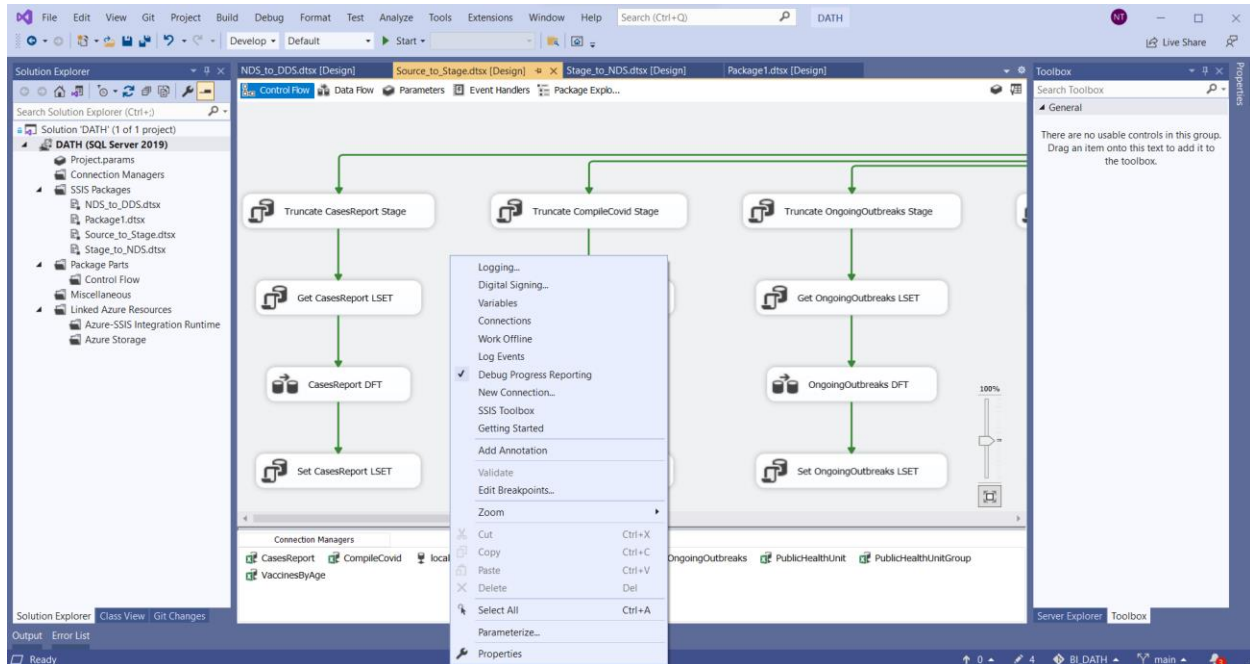
## IV. Demo Monitoring ETL

### A. Logging

Ở phần này, nhóm sẽ demo cách giám sát quá trình ETL thông qua Visual Studio.

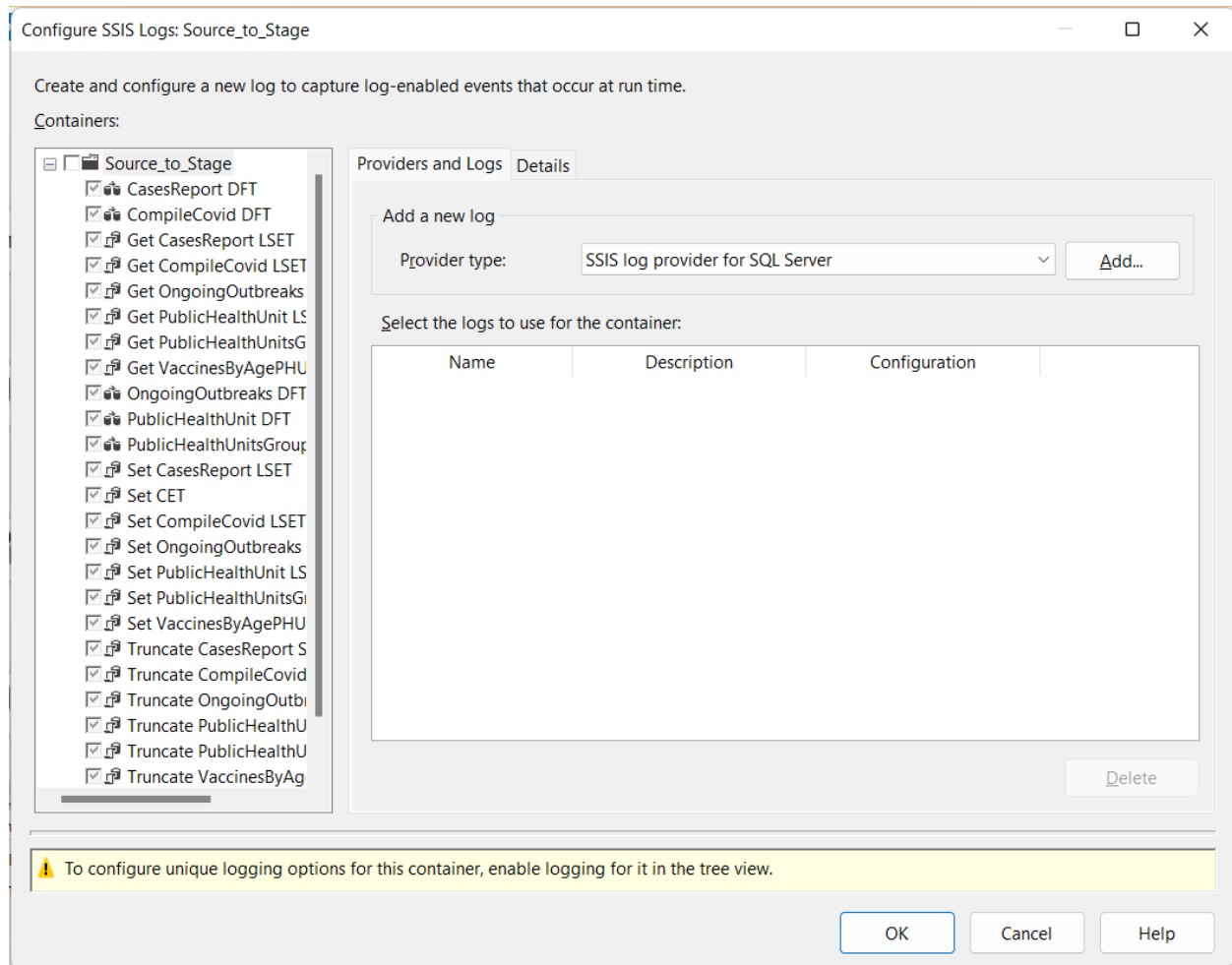
Tại màn hình của package cần giám sát, nhấp chuột phải vào vùng **control flow**, visual studio sẽ hiện các lựa chọn.

Nhấp chọn **Logging**, cửa sổ **Configure SSIS Logs** sẽ hiện ra.

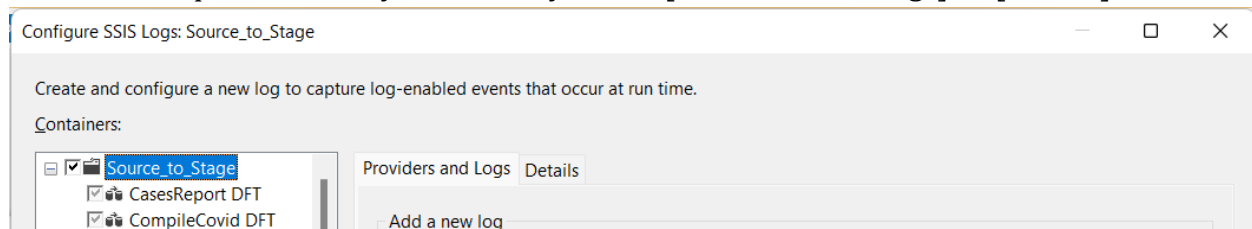


Tick vào project checkbox và enable logging cho cái SSIS package này.

Ở thanh trạng thái phía dưới, có thông báo là: ***"To configure unique logging options, we need to enable logging for it in the tree view"***

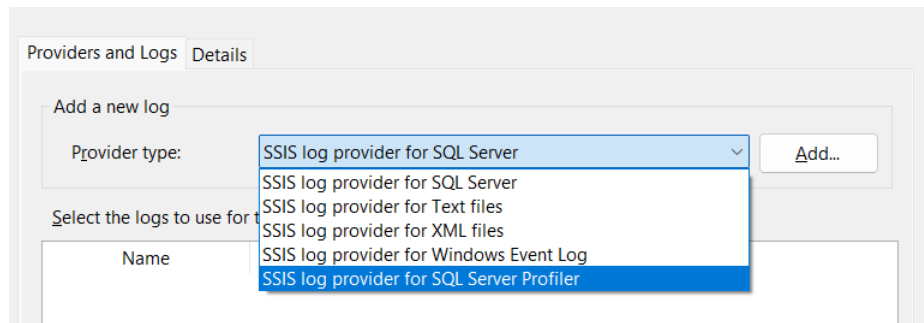


Ở bên phải, ta sẽ thấy 2 tab để tùy chỉnh: [Providers and Logs] và [Details]

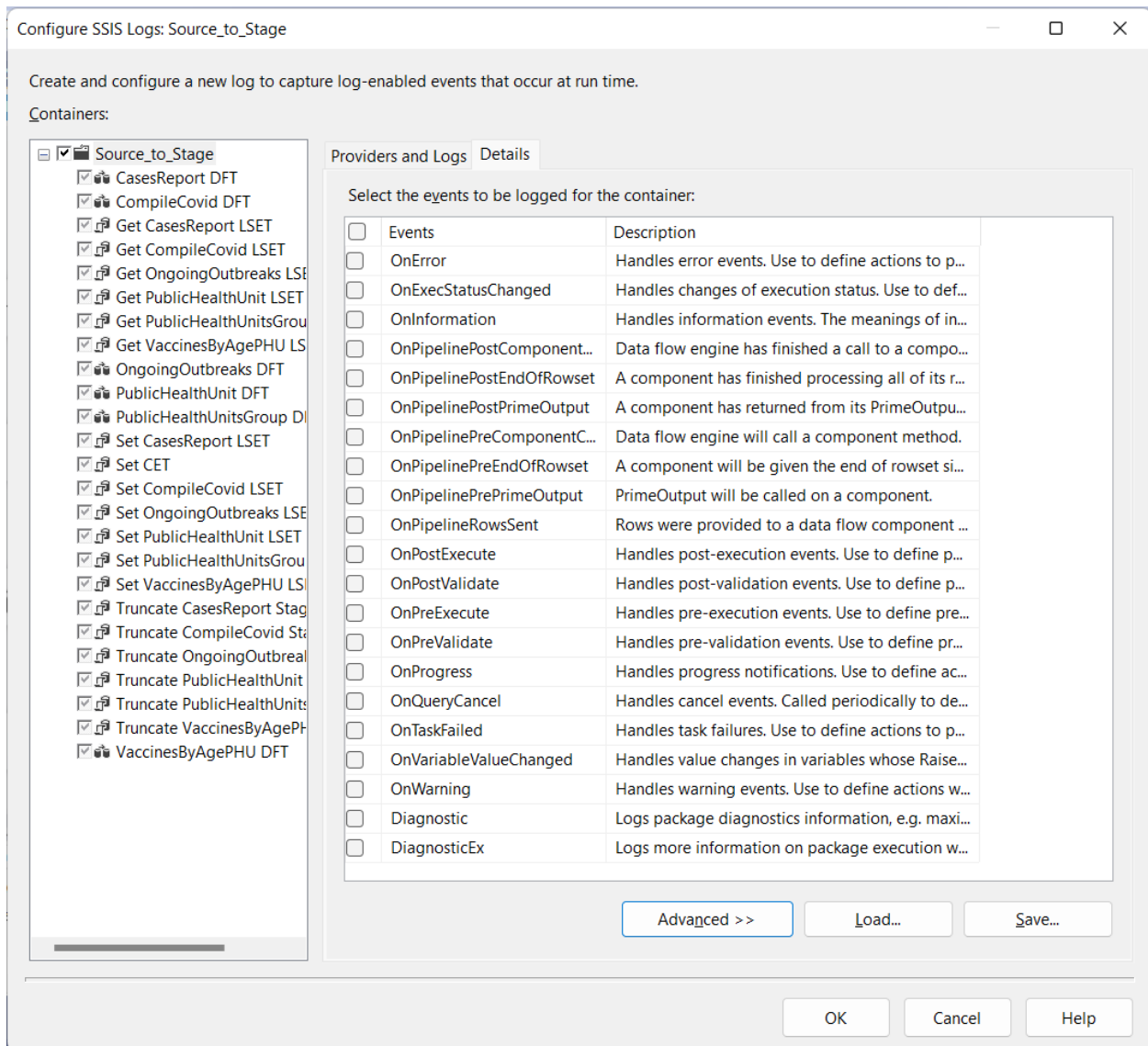


Trong tab **[Providers and Logs]**, chúng ta chọn Provider type và tùy chỉnh.

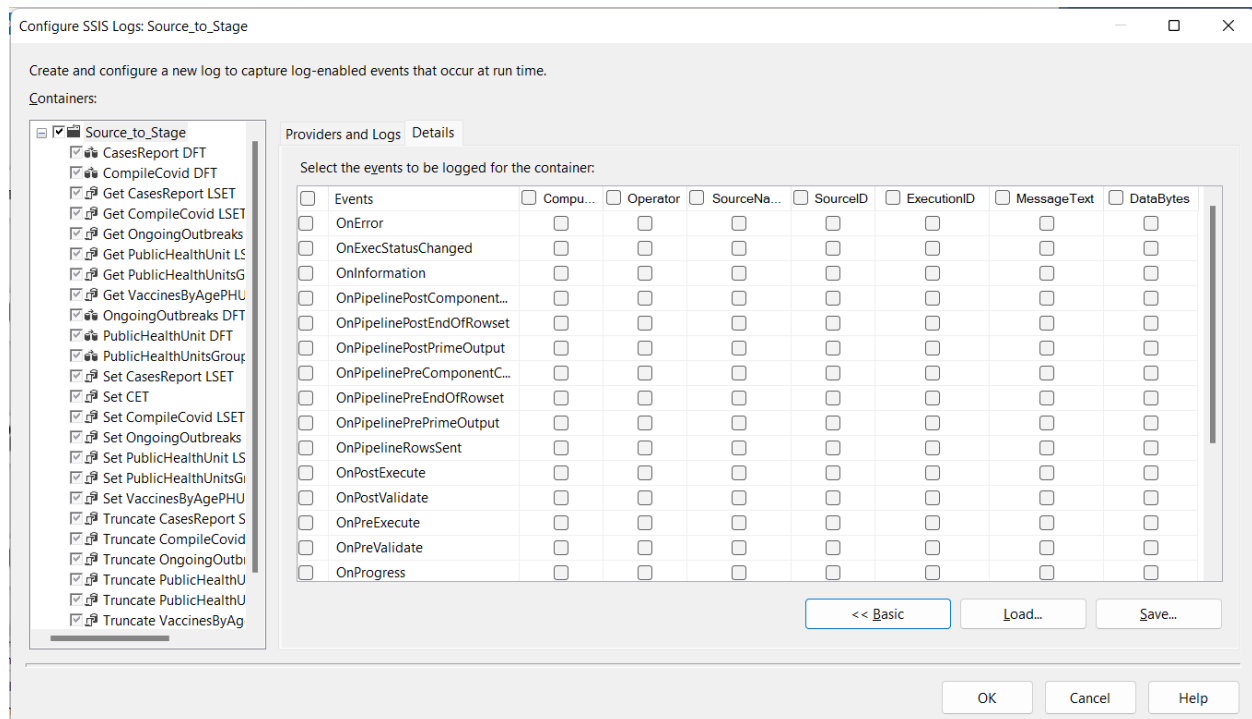
Các loại log provider đã được đề cập ở mục B.



Trong tab **[Details]**, ta sẽ lựa chọn những gì muốn ghi vào log



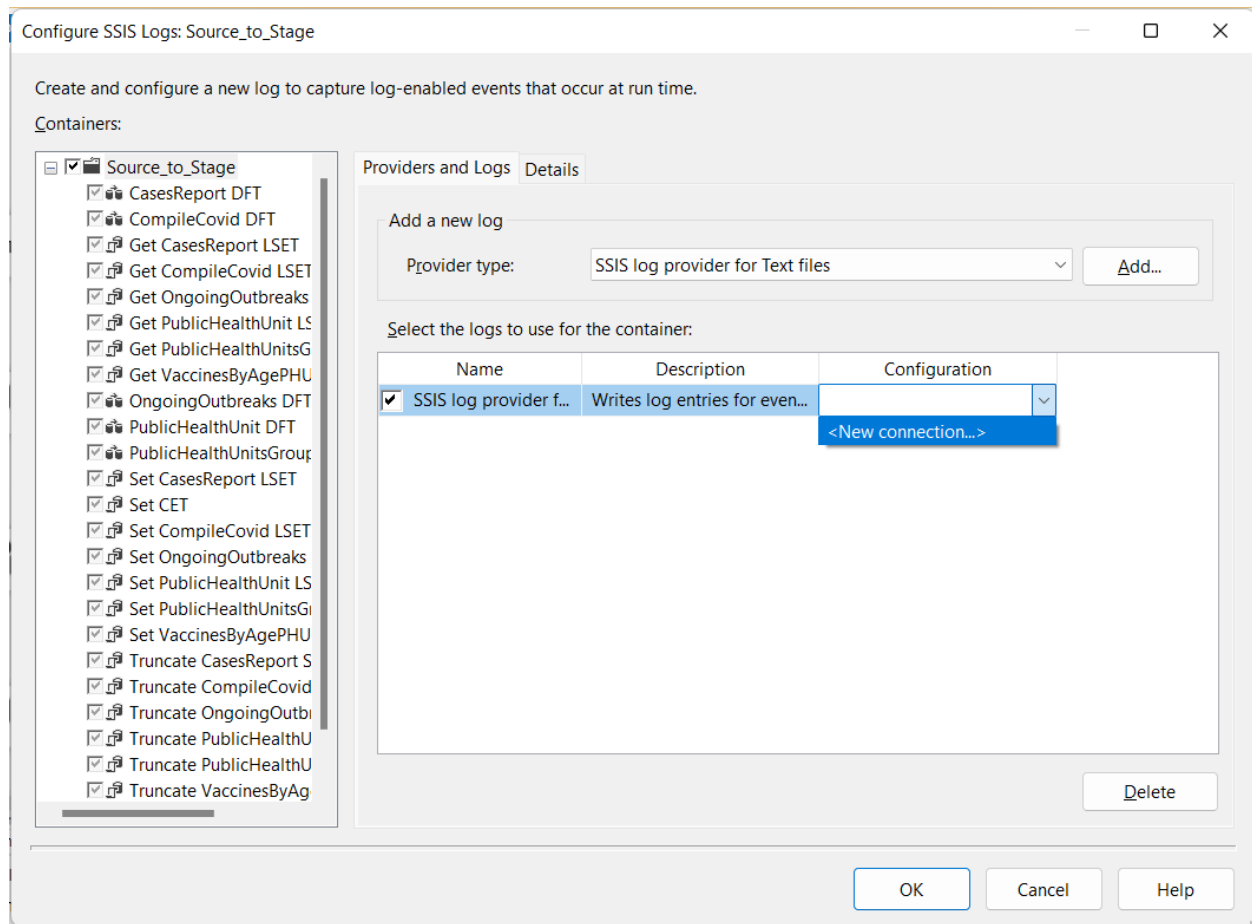
Chọn tab Advanced >> ở phía dưới trang details, cho phép ta chọn các cột bổ sung để ghi log lại.



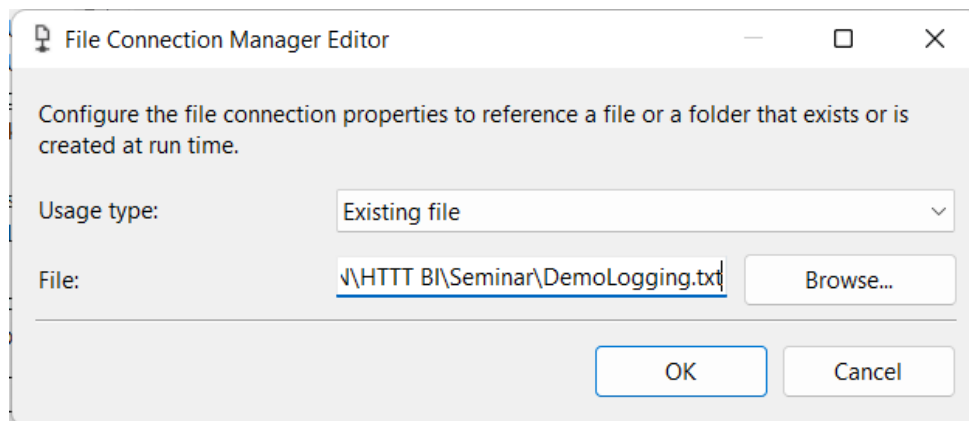
Ở đây nhóm sẽ demo 3 loại logging:

## 1. SSIS log provider for SQL Text files

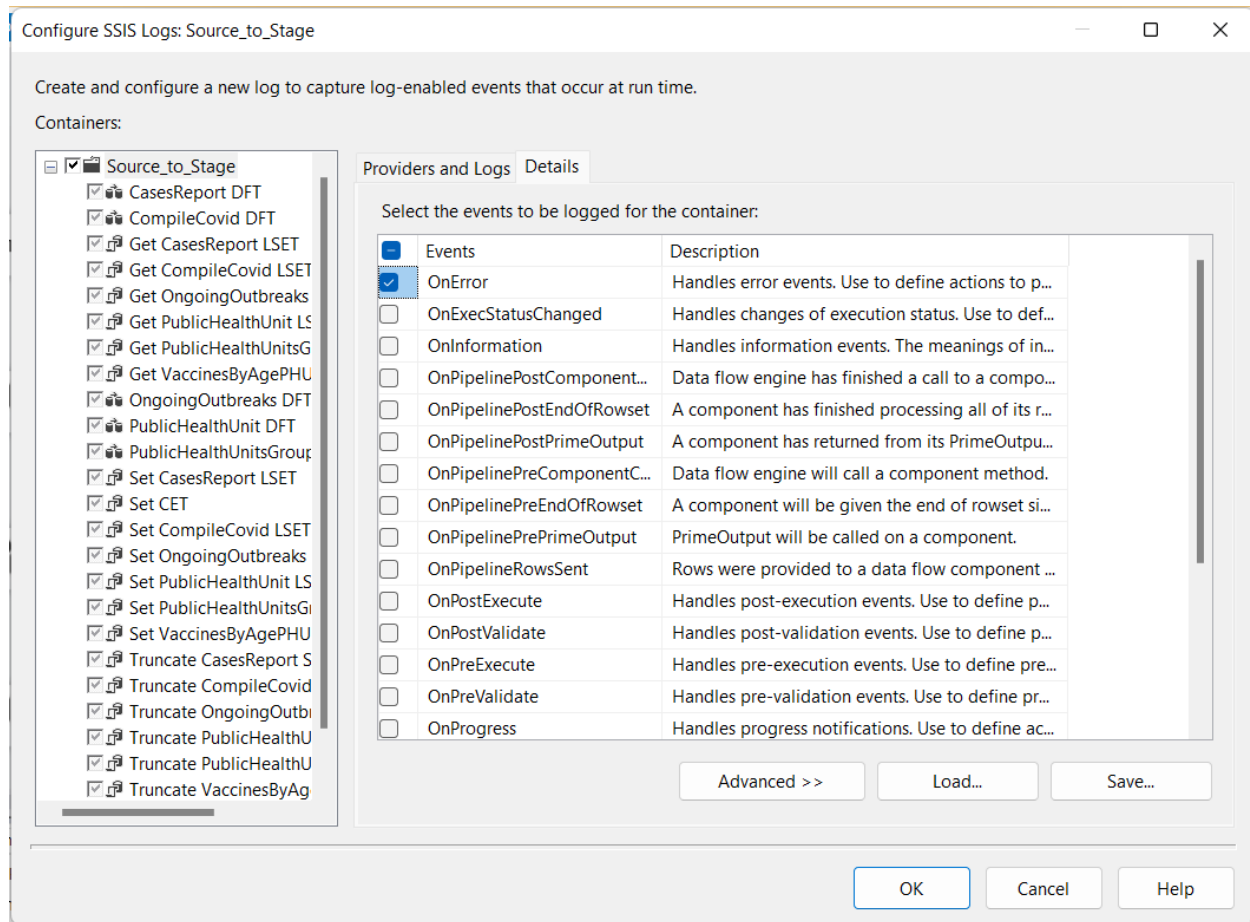
Chọn SSIS log provider for SQL Text files và chọn Add.



Ở cột Configuration, chọn file dùng để ghi log. Usage type có thể chọn **Create file** hoặc **Existing file**.

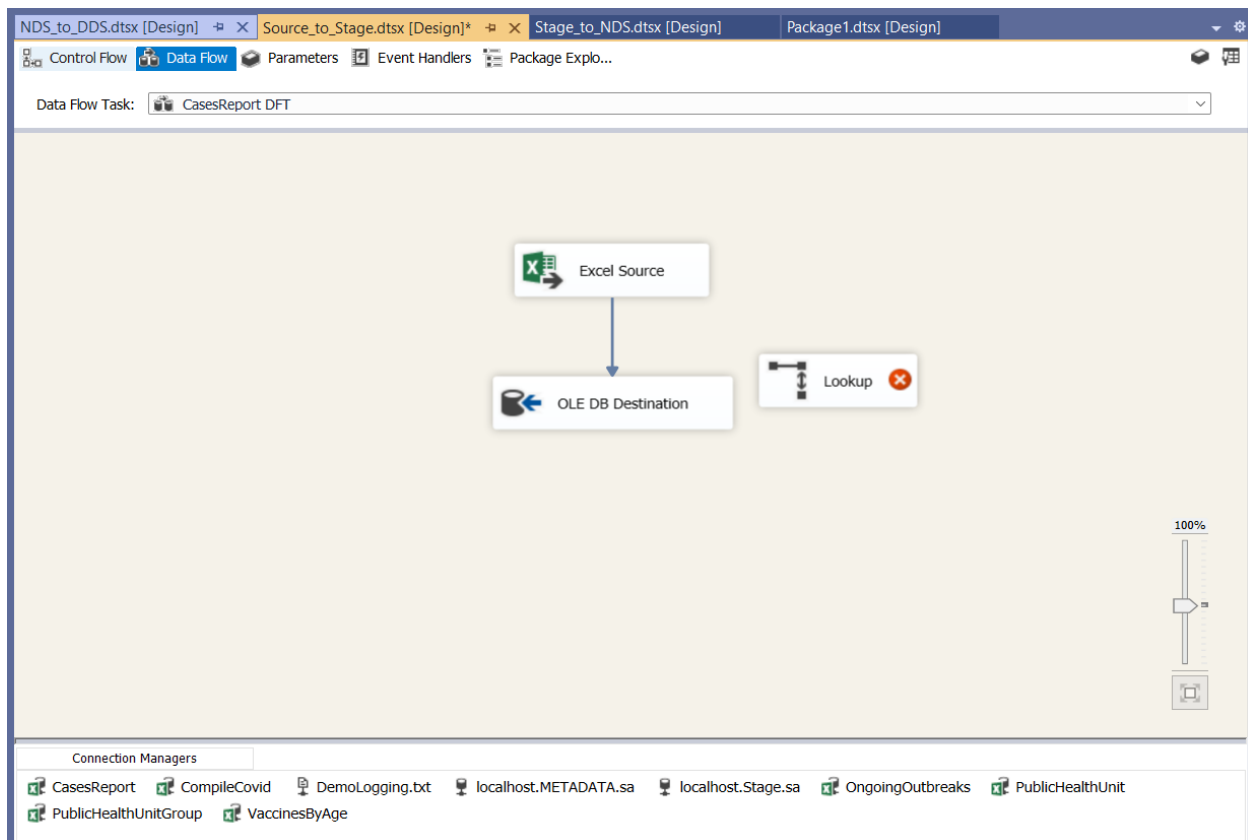


Giả sử chúng ta chỉ muốn ghi lại những error thì ở tab Details tick chọn OnError.



Chọn OK và chạy package.

Giải sử package có lỗi:



Mở file DemoLogging.txt:

```
DemoLogging.txt - Notepad
File Edit View

OnError,DESKTOP-8ADCC67,DESKTOP-8ADCC67\ADMIN,CasesReport
DFT,{89BC3EE6-3A11-4537-98F4-09E6EDD70455},{E9FDD849-8B96-4773-9803-3A9003EF6D5F},12/8/2022 11:23:37
AM,12/8/2022 11:23:37 AM,-1073446884,0x,The "Lookup.Connections[OleDbConnection]" in the connection manager
collection, Connections, of "Lookup" does not have a value for the ID property. Verify that the
ConnectionManagerID property of the runtime connection object has been set for the component.

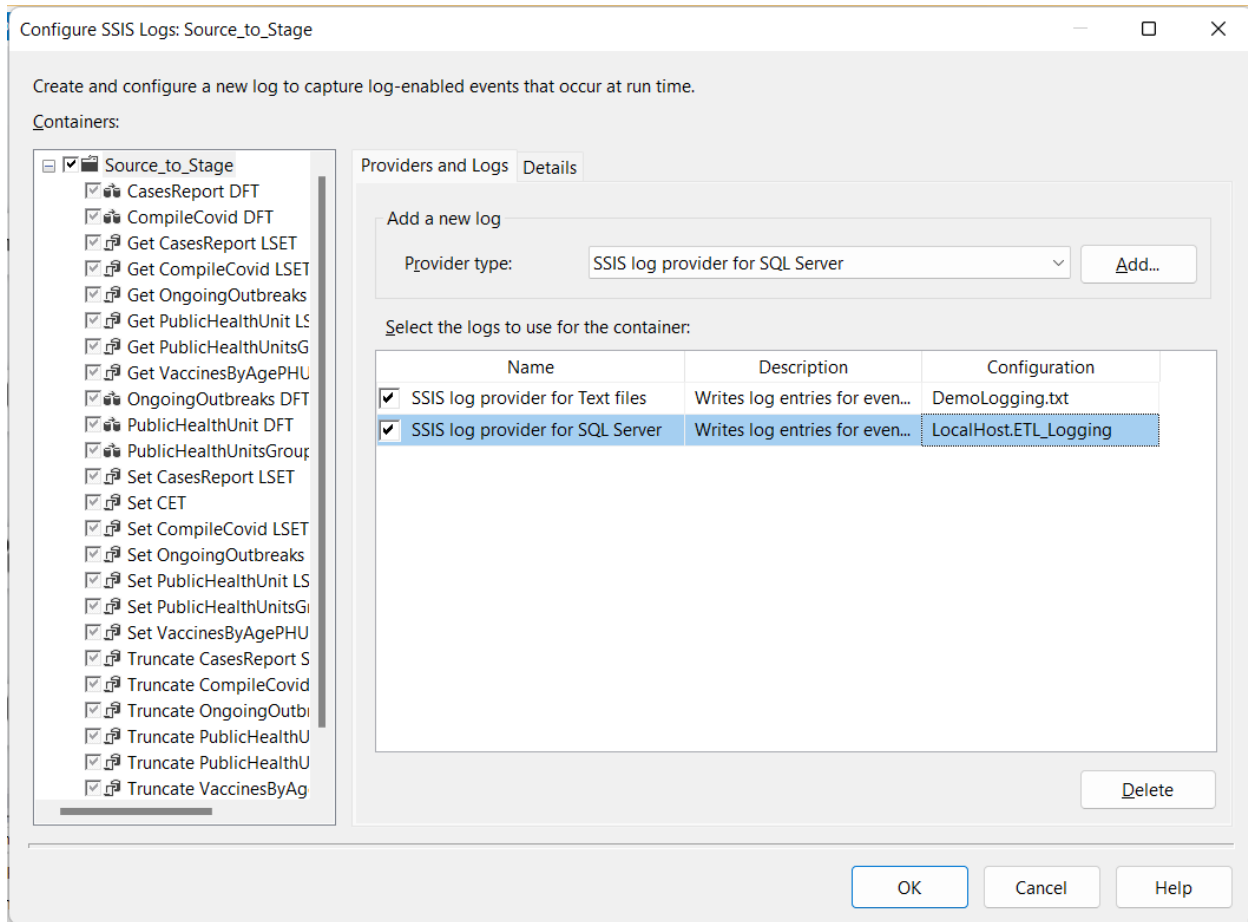
OnError,DESKTOP-8ADCC67,DESKTOP-8ADCC67\ADMIN,CasesReport
DFT,{89BC3EE6-3A11-4537-98F4-09E6EDD70455},{E9FDD849-8B96-4773-9803-3A9003EF6D5F},12/8/2022 11:23:37
AM,12/8/2022 11:23:37 AM,-1073450985,0x,Lookup failed validation and returned error code 0xC004801C.

OnError,DESKTOP-8ADCC67,DESKTOP-8ADCC67\ADMIN,CasesReport
DFT,{89BC3EE6-3A11-4537-98F4-09E6EDD70455},{E9FDD849-8B96-4773-9803-3A9003EF6D5F},12/8/2022 11:23:37
AM,12/8/2022 11:23:37 AM,-1073450996,0x,One or more component failed validation.

OnError,DESKTOP-8ADCC67,DESKTOP-8ADCC67\ADMIN,CasesReport
DFT,{89BC3EE6-3A11-4537-98F4-09E6EDD70455},{E9FDD849-8B96-4773-9803-3A9003EF6D5F},12/8/2022 11:23:37
AM,12/8/2022 11:23:37 AM,-1073594105,0x,There were errors during task validation.
```

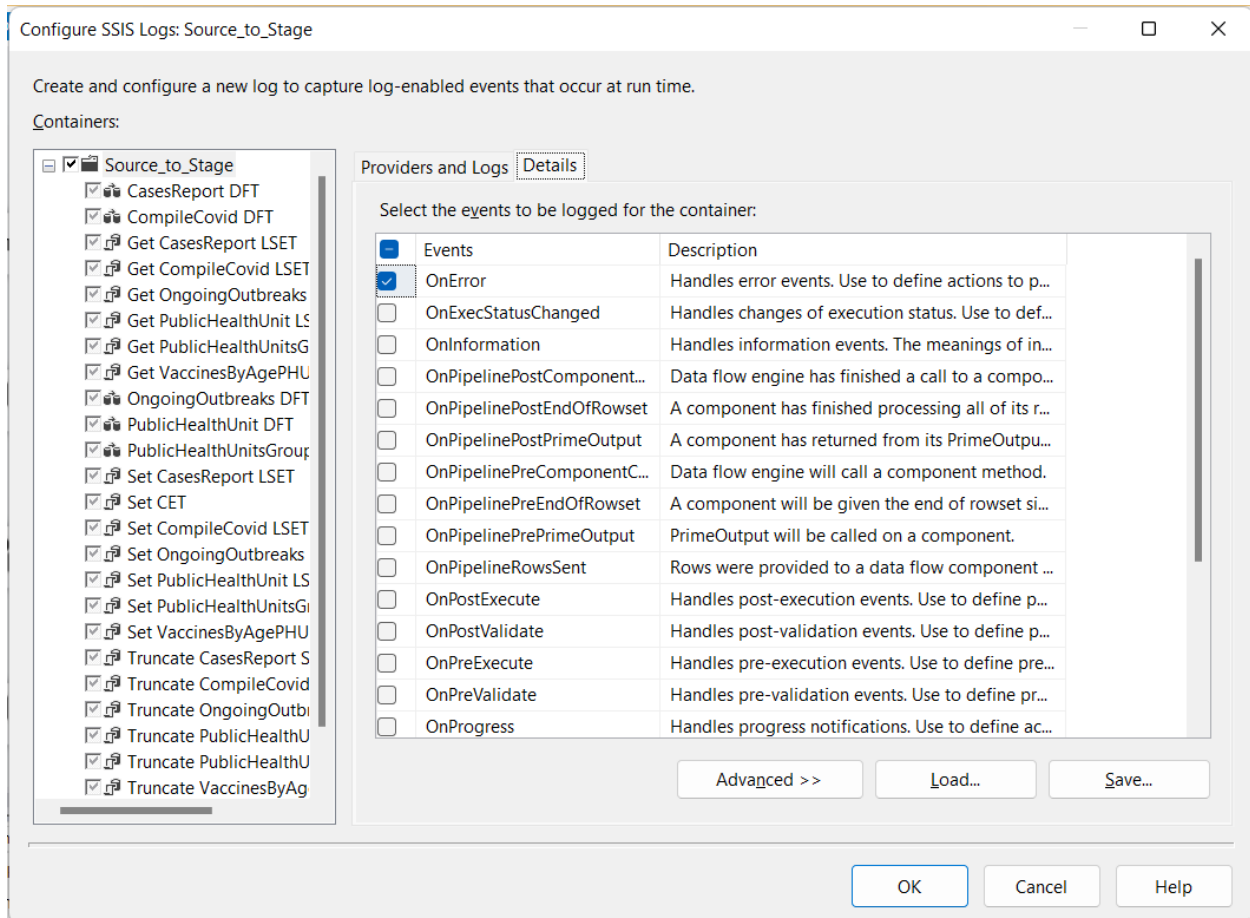
## 2. SSIS log provider for SQL Server

Chúng ta có thể tùy chỉnh nhiều loại log trong cùng một package. Nhấp chuột phải vào vùng control flow và chọn Logging. Chọn SSIS log provider for SQL Server và chọn Add. Sau đó chọn connection tới database mong muốn.

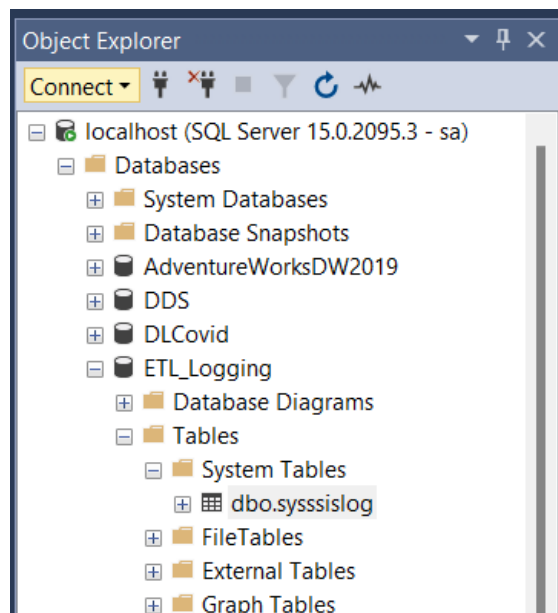


Trong SQL Server logs ta cũng sẽ ghi lại error.





Chọn OK và chạy package. Sau đó tìm table **dbo.sysssislog** trong System tables của database vừa chọn lúc này.



Vì package này lỗi ở bước validation nên chỉ có 4 dòng OnError

SQLQuery1.sql - loc...ost.master (sa (63))

```
select * from ETL_Logging.dbo.sysssislog
```

100 %

Results Messages

	id	event	computer	operator	source	sourceid	executionid	starttime	endtime	datacode	databytes
1	1	OnError	DESKTOP-8ADCC67	DESKTOP-8ADCC67\ADMIN	CasesReport DFT	89BC3EE6-3A11-4537-98F4-09E6EDD70455	E9FDD849-8896-4773-9803-3A9003EF6D5F	2022-12-08 12:50:41.000	2022-12-08 12:50:41.000	-1073446884	0x
2	2	OnError	DESKTOP-8ADCC67	DESKTOP-8ADCC67\ADMIN	CasesReport DFT	89BC3EE6-3A11-4537-98F4-09E6EDD70455	E9FDD849-8896-4773-9803-3A9003EF6D5F	2022-12-08 12:50:41.000	2022-12-08 12:50:41.000	-1073450985	0x
3	3	OnError	DESKTOP-8ADCC67	DESKTOP-8ADCC67\ADMIN	CasesReport DFT	89BC3EE6-3A11-4537-98F4-09E6EDD70455	E9FDD849-8896-4773-9803-3A9003EF6D5F	2022-12-08 12:50:41.000	2022-12-08 12:50:41.000	-1073450996	0x
4	4	OnError	DESKTOP-8ADCC67	DESKTOP-8ADCC67\ADMIN	CasesReport DFT	89BC3EE6-3A11-4537-98F4-09E6EDD70455	E9FDD849-8896-4773-9803-3A9003EF6D5F	2022-12-08 12:50:41.000	2022-12-08 12:50:41.000	-1073594105	0x

Ví dụ ở trường hợp khác, log sẽ có 3 trạng thái theo thứ tự là:

- PackageStart
- OnError
- PackageEnd

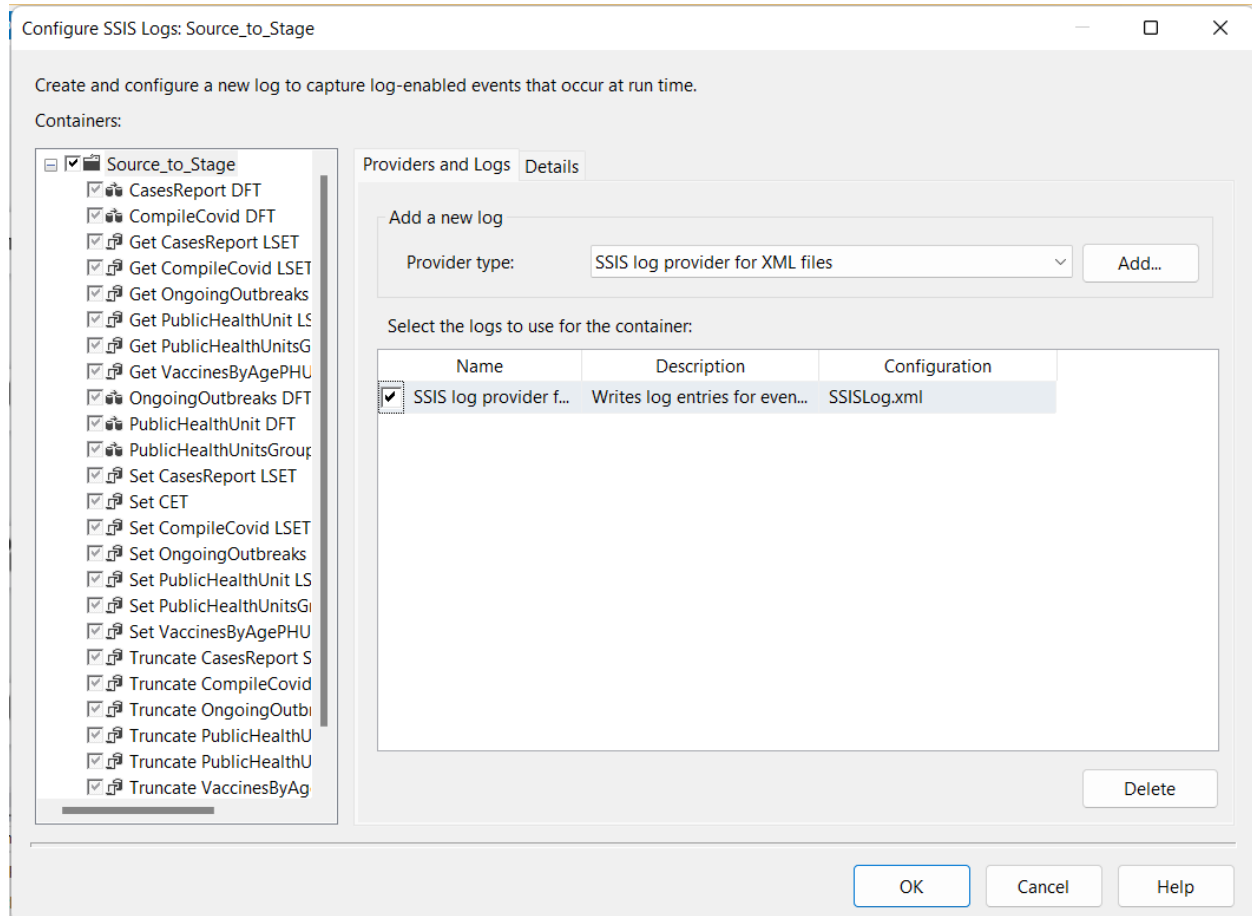
	id	event	computer	operator	source	sourceid	executionid
1	1	PackageStart	DESKTOP-8ADCC67	DESKTOP-8ADCC67\ADMIN	Source_to_Stage	10FCEC86-5341-4875-9194-2E7250ABAD1A	66271FA5-D2F
2	2	OnError	DESKTOP-8ADCC67	DESKTOP-8ADCC67\ADMIN	Get CasesReport LSET	9935180B-0B20-423A-B4B5-B24955C883F2	66271FA5-D2F
3	3	OnError	DESKTOP-8ADCC67	DESKTOP-8ADCC67\ADMIN	Source_to_Stage	10FCEC86-5341-4875-9194-2E7250ABAD1A	66271FA5-D2F
4	4	PackageEnd	DESKTOP-8ADCC67	DESKTOP-8ADCC67\ADMIN	Source_to_Stage	10FCEC86-5341-4875-9194-2E7250ABAD1A	66271FA5-D2F

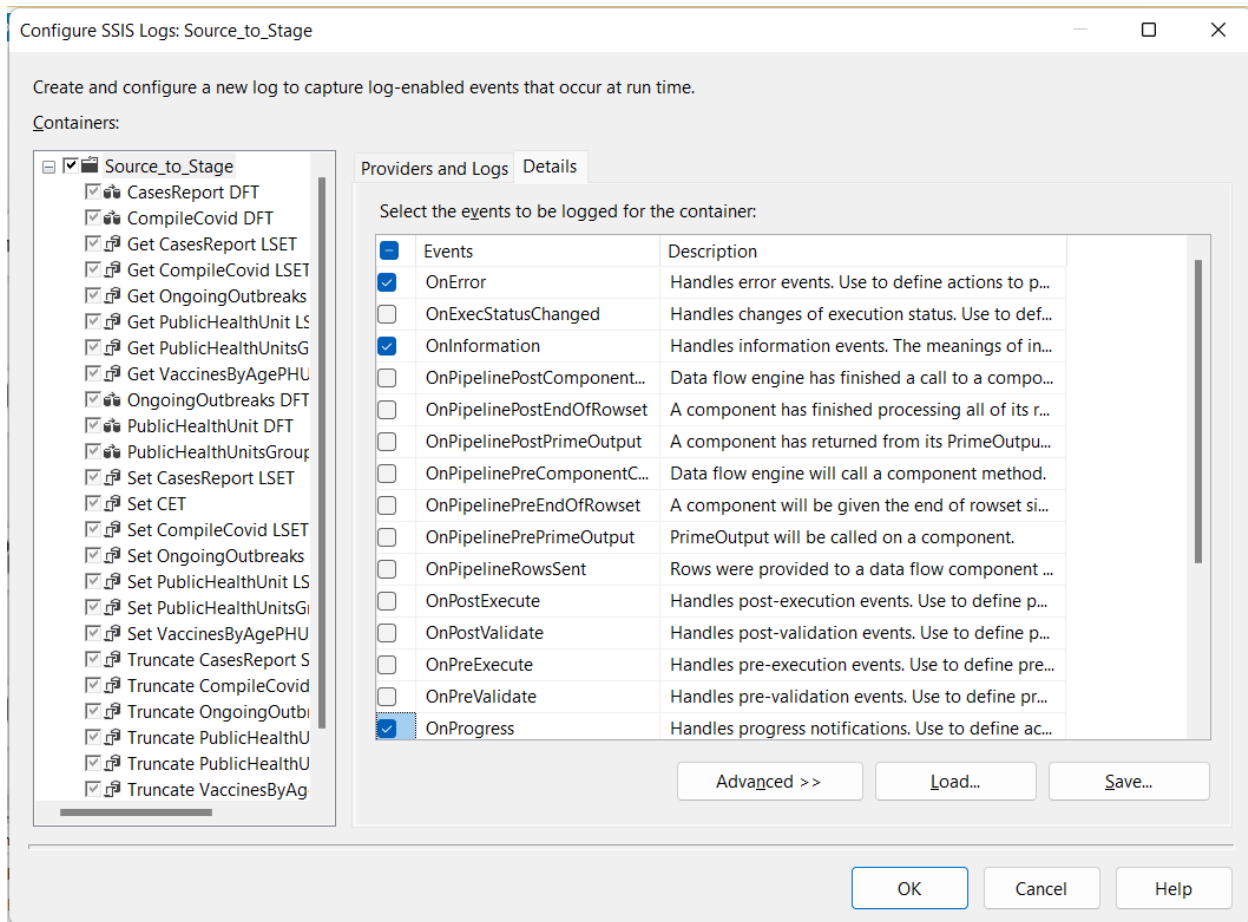
		starttime	endtime	datacode	databytes	message
1	?DE-4753-A636-7228BFE52173	2022-12-08 19:31:20.000	2022-12-08 19:31:20.000	0	0x	Beginning of package execution.
2	?DE-4753-A636-7228BFE52173	2022-12-08 19:31:20.000	2022-12-08 19:31:20.000	-1073548535	0x	An error occurred while assigning a value to var...
3	?DE-4753-A636-7228BFE52173	2022-12-08 19:31:20.000	2022-12-08 19:31:20.000	-1073548535	0x	An error occurred while assigning a value to var...
4	?DE-4753-A636-7228BFE52173	2022-12-08 19:33:46.000	2022-12-08 19:33:46.000	1	0x	End of package execution.

### 3. SSIS log provider for XML files

XML cũng là một cách khá phổ biến để ghi SSIS logs. Chọn SSIS log provider for XML files và chọn file cần ghi log.



Ở lần ghi log này, chọn thêm thông tin để ghi lại, ví dụ như OnInformation, OnProgress.



Chọn OK và chạy package.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
<dtslogs>
  <dtslog>
    <record>
      <event>OnInformation</event>
      <message>Validation phase is beginning. </message>
      <computer>DESKTOP-8ADCC67</computer>
      <operator>DESKTOP-8ADCC67\ADMIN</operator>
      <source>CasesReport DFT</source>
      <sourceid>{89BC3EE6-3A11-4537-98F4-09E6EDD70455}</sourceid>
      <executionid>{8C73C4E7-4D97-4CF9-B74A-5CC324BCAEF9}</executionid>
      <starttime>12/8/2022 9:22:24 PM</starttime>
      <endtime>12/8/2022 9:22:24 PM</endtime>
      <datacode>1074016266</datacode>
      <databytes>0x</databytes>
    </record>
    <record>
      <event>OnInformation</event>
      <message>Validation phase is beginning. </message>
      <computer>DESKTOP-8ADCC67</computer>
      <operator>DESKTOP-8ADCC67\ADMIN</operator>
      <source>CasesReport DFT</source>
      <sourceid>{89BC3EE6-3A11-4537-98F4-09E6EDD70455}</sourceid>
      <executionid>{8C73C4E7-4D97-4CF9-B74A-5CC324BCAEF9}</executionid>
      <starttime>12/8/2022 9:22:24 PM</starttime>
      <endtime>12/8/2022 9:22:24 PM</endtime>
      <datacode>1074016266</datacode>
      <databytes>0x</databytes>
    </record>
  </dtslog>
  <dtslog>
    <record>
      <event>OnProgress</event>
      <message>Validating</message>
      <computer>DESKTOP-8ADCC67</computer>
      <operator>DESKTOP-8ADCC67\ADMIN</operator>
      <source>VaccinesByAgePHU DFT</source>
      <sourceid>{d8fd355b-0b53-420e-a58a-5b4b5e13a386}</sourceid>
      <executionid>{62622D3B-31F0-4D7F-801F-2431FD543A92}</executionid>
      <starttime>12/8/2022 9:24:20 PM</starttime>
      <endtime>12/8/2022 9:24:20 PM</endtime>
      <datacode>100</datacode>
      <databytes>0x</databytes>
    </record>
  </dtslog>
  <dtslog>
    <record>
      <event>PackageStart</event>
      <message>Beginning of package execution. </message>
      <computer>DESKTOP-8ADCC67</computer>
      <operator>DESKTOP-8ADCC67\ADMIN</operator>
      <source>Source_to_Stage</source>
      <sourceid>{10FCEC86-5341-4875-9194-2E7250ABAD1A}</sourceid>
      <executionid>{62622D3B-31F0-4D7F-801F-2431FD543A92}</executionid>
      <starttime>12/8/2022 9:24:20 PM</starttime>
      <endtime>12/8/2022 9:24:20 PM</endtime>
      <datacode>0</datacode>
      <databytes>0x</databytes>
    </record>
  </dtslog>
</dtslogs>

```

```

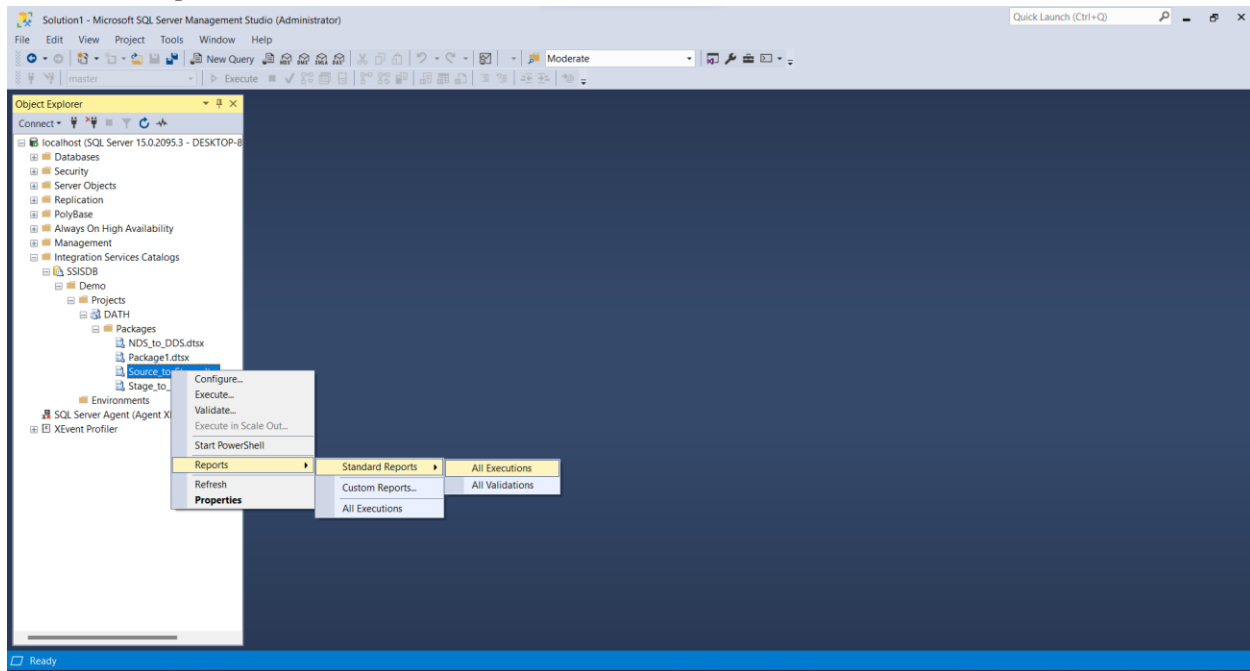
<databytes>0x</databytes>
</record>
▼<record>
  <event>OnError</event>
  <message>An error occurred while assigning a value to variable "LSETResult": "Single Row result
set is specified, but no rows were returned.". </message>
  <computer>DESKTOP-8ADCC67</computer>
  <operator>DESKTOP-8ADCC67\ADMIN</operator>
  <source>Get CasesReport LSET</source>
  <sourceid>{9935180B-0B20-423A-B4B5-B24955C883F2}</sourceid>
  <executionid>{62622D3B-31F0-4D7F-801F-2431FD543A92}</executionid>
  <starttime>12/8/2022 9:24:20 PM</starttime>
  <endtime>12/8/2022 9:24:20 PM</endtime>
  <datacode>-1073548535</datacode>
  <databytes>0x</databytes>
</record>
▼<record>
  <event>OnError</event>
  <message>An error occurred while assigning a value to variable "LSETResult": "Single Row result
set is specified, but no rows were returned.". </message>
  <computer>DESKTOP-8ADCC67</computer>
  <operator>DESKTOP-8ADCC67\ADMIN</operator>
  <source>Source_to_Stage</source>
  <sourceid>{10FCEC86-5341-4875-9194-2E7250ABAD1A}</sourceid>
  <executionid>{62622D3B-31F0-4D7F-801F-2431FD543A92}</executionid>
  <starttime>12/8/2022 9:24:20 PM</starttime>
  <endtime>12/8/2022 9:24:20 PM</endtime>
  <datacode>-1073548535</datacode>
  <databytes>0x</databytes>
</record>
▼<record>
  <event>OnProgress</event>
  <message>Executing query "select LSET from Data Flow where TenBang = 'Ongoing...". </message>

```

## B. Built-in Reporting

### 1. Package report

Để xem report của một package click chuột phải vào package cần xem, chọn Reports -> Standard Reports -> All Execution.



All Executions report gồm lịch sử chạy của package. Bao gồm Execution ID, kể cả package có succeeded hoặc failed, folder chứa package, tên project, tên package, thời gian bắt đầu chạy và thời gian kết thúc, khoảng thời gian mà package chạy tính bằng giây.

All Executions - 12/9/2022 12:23 AM - DESKTOP-8ADCC67

SQL Server

### All Executions

on DESKTOP-8ADCC67 at 12/9/2022 12:23:05 AM

This report provides information about the Integration Services package executions that have been performed on the connected SQL Server instance.

Filter: Start time range: 12/3/2022 - 12/9/2022. Status: All (4 more)

Execution Information

0 Failed 0 Running 1 Succeeded 0 Others

ID	Status	Report	Folder Name	Project Name	Package Name	Start Time	End Time	Duration (sec)
5	Succeeded	<a href="#">Overview</a> <a href="#">All Messages</a> <a href="#">Execution Performance</a>	Demo	DATH	Stage_to_NDS.dtsx	12/9/2022 12:22:55 AM	12/9/2022 12:22:59 AM	

1 / 1

Một trong các tính năng của report được tích hợp trong SSIS đó là các đường dẫn cho phép xem chúng ta truy cập vào một report chi tiết hơn. Đường dẫn đầu tiên của All Execution report là Overview Report.

Overview - 12/9/2022 12:32 AM - DESKTOP-8ADCC67

### Overview

on DESKTOP-8ADCC67 at 12/9/2022 12:32:44 AM

This report provides an overview of the package tasks and parameters, including execution or validation information.

[View Messages](#)  
[View Performance](#)

Execution Information

Operation ID	5
Package	Demo\DATA\Stage_to_NDS.dtsx
Environment	-
Status	Succeeded
Machine	DESKTOP-8ADCC67

Duration (sec)	4.416
Start Time	12/9/2022 12:22:55 AM
End Time	12/9/2022 12:22:59 AM
Caller	DESKTOP-8ADCC67\ADMIN

Parameters Used

Name	Value	Data Type
CALLER_INFO		String
DUMP_EVENT_CODE	0	String
DUMP_ON_ERROR	False	Boolean
DUMP_ON_EVENT	False	Boolean
LOGGING_LEVEL	1	Int32
SYNCHRONIZED	False	Boolean

Property Overrides

Property Path	Property Value
---------------	----------------

Execution Overview

Filter: Result All (3 more)

Result	Duration (sec)	Package Name	Task Name	Execution Path
Succeeded	1.031	Stage_to_NDS.dtsx	Stage_to_NDS	<a href="#">\Stage_to_NDS</a>
Succeeded	0.141	Stage_to_NDS.dtsx	DFT AgeGroup	<a href="#">\Stage_to_NDS\DFT AgeGroup</a>
Succeeded	0.109	Stage_to_NDS.dtsx	DFT CasesReport	<a href="#">\Stage_to_NDS\DFT CasesReport</a>
Succeeded	0.141	Stage_to_NDS.dtsx	DFT CaseStatus	<a href="#">\Stage_to_NDS\DFT CaseStatus</a>
Succeeded	0.172	Stage_to_NDS.dtsx	DFT CompileCovid	<a href="#">\Stage_to_NDS\DFT CompileCovid</a>

Trong Overview report có 2 đường dẫn giúp ta có thể drill down để có được các thông tin chi tiết hơn. Đầu tiên là View Messages report, bao gồm các chi tiết về package cho đến cấp độ sự kiện. View Messages report rất hữu ích để khắc phục sự cố khi package



có lỗi. Đường dẫn khác là View Performance. Đường dẫn này sẽ mở View Performance report, hiển thị xu hướng của package về thời lượng.

Messages - 12/9/2022 12:33 AM - DESKTOP-8ADCC67

Messages

on DESKTOP-8ADCC67 at 12/9/2022 12:33:21 AM

This report displays messages for the executions or validations.

[View Overview](#)

Execution Information

Operation ID	5 (Execution)	Duration (sec)	4.416
Name	Demo\DATA\Stage_to_NDS.dtsx	Start Time	12/9/2022 12:22:55 AM
Environment	-	End Time	12/9/2022 12:22:59 AM
Status	Succeeded	Caller	DESKTOP-8ADCC67\ADMIN

☐ Error Messages

Message Type	Action	Message Time	Message	Message Source Name	Subcomponent Name	Execution Path
--------------	--------	--------------	---------	---------------------	-------------------	----------------

All Messages

[Filter](#) Message Type: All; (4 more)

Message Type	Message Time	Message	Message Source Name	Subcomponent Name	Execution Path
OnPostExecute	12/9/2022 12:22:59 AM	Stage_to_NDS Finished, 12:22:59 AM, Elapsed time: 00:00:01.031.	Stage_to_NDS		\Stage_to_NDS
OnPostExecute	12/9/2022 12:22:59 AM	DFT CompileCovid Finished, 12:22:59 AM, Elapsed time: 00:00:00.172.	DFT CompileCovid		\Stage_to_NDS\DFT CompileCovid
OnInformation	12/9/2022 12:22:59 AM	DFT CompileCovid Information: Cleanup phase is beginning.	DFT CompileCovid	SSIS Pipeline	\Stage_to_NDS\DFT CompileCovid
OnInformation	12/9/2022 12:22:59 AM	DFT CompileCovid Information: "Insert dữ liệu" wrote 0 rows.	DFT CompileCovid	SSIS Pipeline	\Stage_to_NDS\DFT CompileCovid

Execution Performance - 12/9/2022 9:55 AM - DESKTOP-8ADCC67

Execution Performance

on DESKTOP-8ADCC67 at 12/9/2022 9:55:23 AM

This report provides performance information for an execution.

[View Overview](#)

Execution Information

Operation ID	8	Duration (sec)	36.035
Package	Demo\DATA\Stage_to_NDS.dtsx	Start Time	12/9/2022 9:54:46 AM
Environment	-	End Time	
Status	Running	Caller	DESKTOP-8ADCC67\ADMIN

Past Executions Information

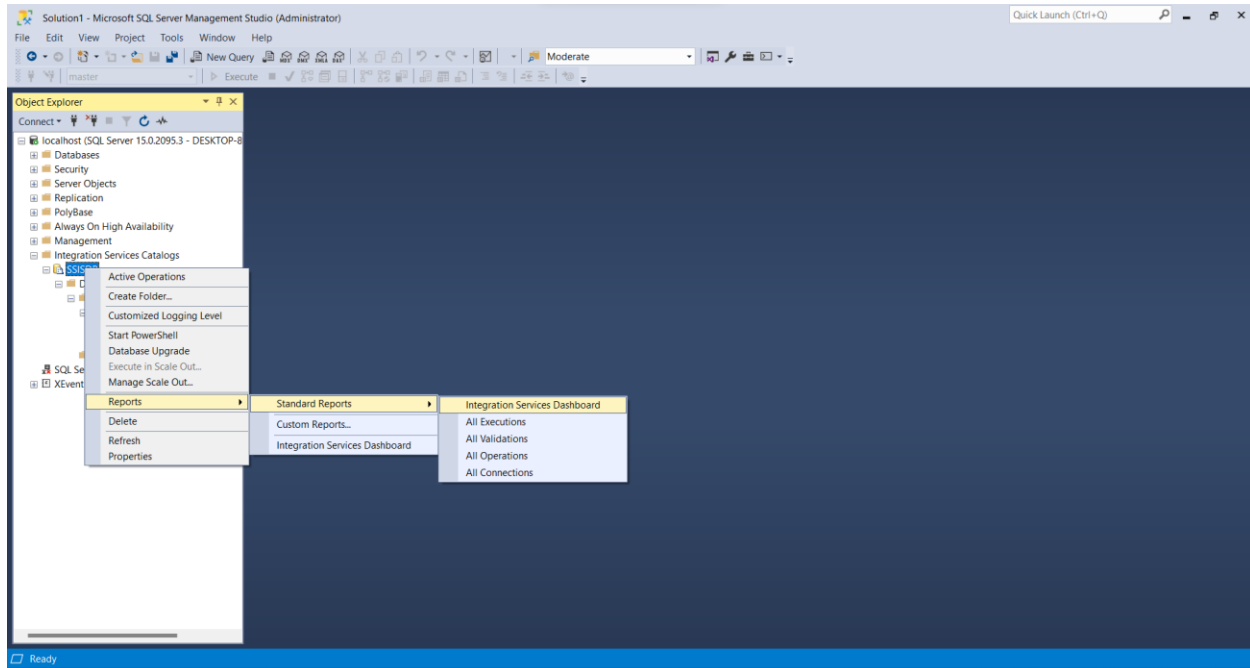
This chart displays the duration of past 10 successful executions.

3-month average duration (sec): 4.001  
Standard deviation of duration (sec): 0.588

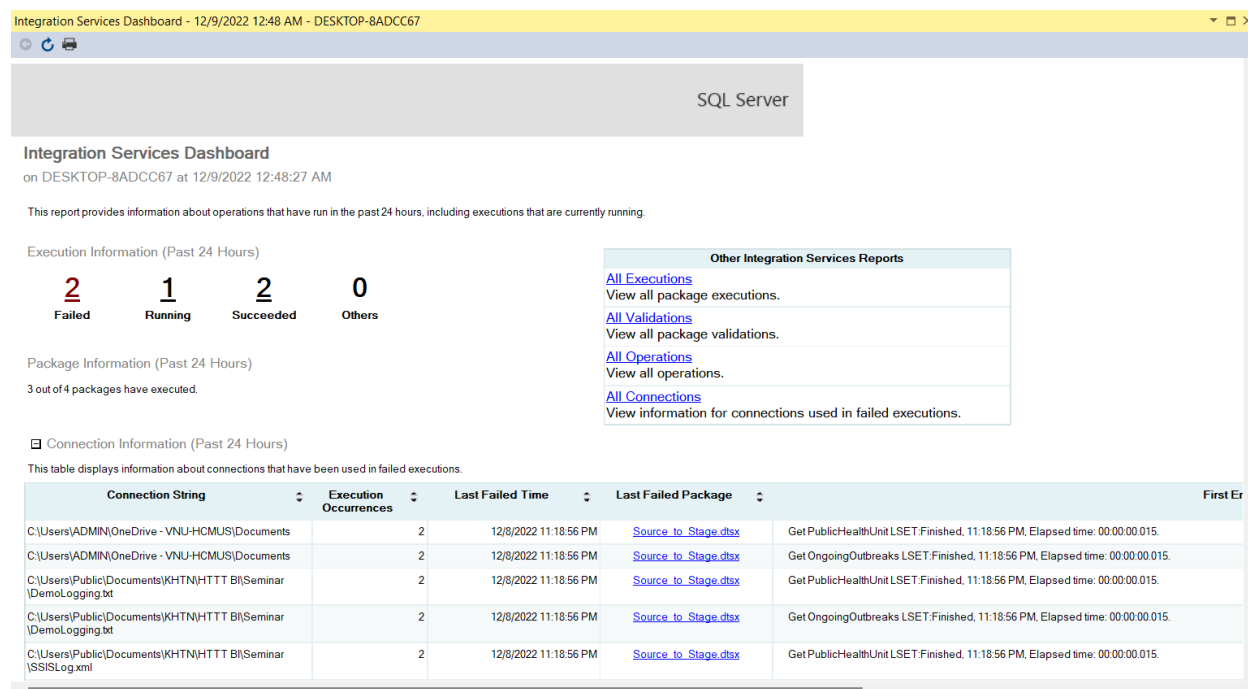
ID	Start Time	Duration (sec)
7	12/9/2022 12:48:25 AM	3.585
5	12/9/2022 12:22:55 AM	4.416

## 2. Dashboard report

Chọn SSIS Catalog và điều hướng tới Reports -> Standard Reports -> Integration Services Dashboard



Dashboard report cung cấp một báo cáo tổng quan ở mức high-level về toàn bộ SSIS catalog. Sử dụng report này, ta có thể nhanh chóng xem số lượng package Failed, Running hoặc Succeeded và các thông tin trạng thái khác. Bạn có thể xem chi tiết các báo cáo bằng cách nhấp vào liên kết ở bên phải, như minh họa trong ảnh chụp màn hình bên dưới.



### Các lựa chọn của dashboard report là:

- All Executions: cung cấp thông tin của tất cả các package đã chạy trên SQL Server instance.
- All Validations: cung cấp thông tin của tất cả các validations thực hiện trên SQL Server instance.
- All Operations: cung cấp thông tin của tất cả hoạt động.
- All Connections: cung cấp thông tin của ngữ cảnh kết nối của những lần chạy lỗi đã xảy ra.

Ngoài ra mỗi report đều có filter để lọc các thông tin dựa trên ngày, trạng thái, tên package, tên project, ID và tên thư mục.

**Filter Settings**

**General**

Folder Name  
ID  
Package Name  
Project Name

**Operation**

Status: All

**Start Time Range**

From: 12/3/2022  
To: 12/9/2022

**Folder Name**  
Display operations in which the Folder Name contains the specified string.

OK Cancel

## C. Views

Truy vấn các Integration Services view để quan sát các đối tượng, các cài đặt, và dữ liệu thực thi được lưu trữ trong SSISDB catalog.

Tên mặc định của catalog là SSISDB. Các đối tượng được lưu trữ trong catalog bao gồm project, package, environment, và lịch sử thực thi.

**Một số view của catalog:**

- `catalog.catalog_properties` (SSISDB Database): hiển thị tính chất của Integration Services catalog.
- `catalog.execution_parameter_values` (SSISDB Database): hiển thị tham số được dùng bởi các Integration Service package trong một thực thể thực thi.
- `catalog.executions` (SSISDB Database): hiển thị một thực thể thực thi package trong Integration Services catalog. Các package được thực thi với tác vụ Execute Package chạy cùng một bản thực thi được gọi là parent package.
- `catalog.folders` (SSISDB Database): hiển thị các thư mục trong Integration Services catalog.
- `catalog.packages` (SSISDB Database): hiển thị tất cả các package trong Integration Services catalog.
- `catalog.projects` (SSISDB Database): hiển thị tất cả các project trong Integration Services catalog.

Ví dụ dưới đây là view `catalog.executions`.

SQLQuery6.sql - L:\DCC67\ADMIN (53)\*

```

select execution_id, folder_name, project_name,
package_name, executed_as_name, [status], start_time, end_time
from [SSISDB].catalog.executions

```

	execution_id	folder_name	project_name	package_name	executed_as_name	status	start_time	end_time
1	2	Demo	DATH	Source_to_Stage.dtsx	DESKTOP-8ADCC67\ADMIN	4	2022-12-08 23:16:24.1099938 +07:00	2022-12-08 23:17:14
2	3	Demo	DATH	Source_to_Stage.dtsx	DESKTOP-8ADCC67\ADMIN	4	2022-12-08 23:18:41.1523772 +07:00	2022-12-08 23:19:32
3	5	Demo	DATH	Stage_to_NDS.dtsx	DESKTOP-8ADCC67\ADMIN	7	2022-12-09 00:22:55.2749482 +07:00	2022-12-09 00:22:56
4	6	Demo	DATH	NDS_to_DDS.dtsx	DESKTOP-8ADCC67\ADMIN	7	2022-12-09 00:47:59.8159320 +07:00	2022-12-09 00:48:00
5	7	Demo	DATH	Stage_to_NDS.dtsx	DESKTOP-8ADCC67\ADMIN	7	2022-12-09 00:48:25.0981486 +07:00	2022-12-09 00:48:28
6	8	Demo	DATH	Stage_to_NDS.dtsx	DESKTOP-8ADCC67\ADMIN	7	2022-12-09 09:54:46.5246715 +07:00	2022-12-09 09:55:27
7	10	Demo	DATH	Stage_to_NDS.dtsx	DESKTOP-8ADCC67\ADMIN	7	2022-12-09 14:32:15.6149406 +07:00	2022-12-09 14:32:18
8	11	Demo	DATH	Stage_to_NDS.dtsx	DESKTOP-8ADCC67\ADMIN	7	2022-12-09 14:33:40.4107946 +07:00	2022-12-09 14:33:43
9	12	Demo	DATH	Stage_to_NDS.dtsx	DESKTOP-8ADCC67\ADMIN	7	2022-12-09 14:34:43.9097953 +07:00	2022-12-09 14:34:46
10	13	Demo	DATH	Source_to_Stage.dtsx	DESKTOP-8ADCC67\ADMIN	4	2022-12-09 14:35:55.1977720 +07:00	2022-12-09 14:36:35
11	14	Demo	DATH	Source_to_Stage.dtsx	DESKTOP-8ADCC67\ADMIN	4	2022-12-09 14:36:14.9979508 +07:00	2022-12-09 14:36:18

Query executed successfully. localhost (15.0 RTM) DESKTOP-8ADCC67\ADMIN ... master 00:00:00 11 rows

Với status của executions là created (1), running (2), canceled (3), failed (4), pending (5), ended unexpectedly (6), succeeded (7), stopping (8), và completed (9).

## V. Tài liệu tham khảo

<https://blog.paessler.com/monitoring-etl-processes>

[https://docs.oracle.com/cd/E49961\\_01/doc.10/e38589/monitoringetl.htm](https://docs.oracle.com/cd/E49961_01/doc.10/e38589/monitoringetl.htm)

<https://mindmajix.com/ssis/monitoring-package-executions>

<https://www.anodot.com/blog/etl-monitoring/>

<https://www.tidydata.io/etl-elt-monitoring.html>

<https://www.techopedia.com/4-ways-ai-driven-etl-monitoring-can-help-avoid-glitches/2/33969>

<https://www.sqlshack.com/overview-of-ssis-package-logging/>

<https://learn.microsoft.com/en-us/sql/integration-services/system-views/views-integration-services-catalog?view=sql-server-ver16>