فعالية استخدام SSIS كأداة لاستخراج وتحويل وتحميل البيانات (ETL) عند عدم تجانس المصادر

طالب دراسات عليا (ماجستير): م. وسيم هندي
كلية الهندسة المعلوماتية، أكاديمية الأسد للهندسة العسكرية
اشراف الدكتور: محمد سلامي + د. عبد الله يوسف،

الملخص

يعتبر اختيار أداة ETL المناسبة لتحقيق مستودع البيانات أمراً بالغ الأهمية، ويُعلل ذلك بأن الاختيار الخاطئ للأداة قد يؤدي إلى سلسلة من الخسائر المالية والوقتية، ناهيك عن مقدار الجهد الشاق الذي سيتم بذله. تُواجه عملية استخراج وتحويل وتحميل البيانات الى المستودع صعوبات كثيره، أهمها عدم تجانس مصادر البيانات.

يمكن لمصمم مستودع البيانات الاختيار من بين مجموعة متنوعة من أدوات الاستخراج والتحويل والتحميل (ETL Tools)، ولكن بدون استكشاف أو معرفة ميزاتها، سيؤدي ذلك إلى عملية اتخاذ قرار غير صحيح. في هذا البحث تمَّ تقييم فعالية الأداة (SSIS-Microsoft SQL Server Service) في استخراج البيانات من مصادرها الغير مُتجانسة وتخزينها في مستودع البيانات.

الكلمات المفتاحية: مستودع البيانات، أدوات الاستخراج والتحويل والتحميل، تكامل البيانات، تدفق البيانات.

ورد البحث للمجلة بتاريخ / /2021
قبل للنشر بتاريخ / /2021

Effectiveness of using SSIS as a tool to extract, transform and load data (ETL) when data sources are heterogeneous

Mohamad Salami, Abdalla Yousef, Waseem Hindi*
AL- Assad Academy, For Military Engineering
Postgraduate Student (MSc.)

Abstract

Choosing the right ETL tool for the realization of the data warehouse is critical, reasoning that the wrong choice of tool can lead to a series of financial and time losses, not to mention the amount of hard work that will be put in. The process of extracting, transforming and uploading data to the repository faces many difficulties, the most important of which is the heterogeneity of data sources.

The data warehouse designer can choose from a variety of ETL tools, but without exploring or knowing their features, it will lead to an inaccurate decision-making process. In this paper, the effectiveness of the tool (SSIS-Microsoft SQL Server Service) will be evaluated in extracting data from its heterogeneous sources and storing it in a data warehouse.

Keywords: Data Warehouse, ETL Tools, Data Integration, Data Flow.

Received / /2021
Accepted / /2021

1- مقدمة

نتيجة تضخم الحجوم المخزنة في قواعد البيانات بصورة كبيرة جدا أصبح من الصعب معالجة هذه البيانات والاستفادة منها ضمن القيود الزمنية المسموحة، مما أدى إلى بروز تحدي جديد في كيفية تحويل قواعد البيانات من قواعد بيانات تشغيلية إلى مستودعات للبيانات (Data Warehouses)[1].

تدعم مستودعات البيانات عمليات اتخاذ القرار والوصول إلى البيانات بهدف اجراء المعاجة التحليلية المباشرة (OLAP)، والمساعدة في اتخاذ القرار والتخطيط والرؤية المستقبلية[2]. من أجل الحصول على نتائج تحليلية للبيانات وإعداد تقارير صحيحة لها، من المهم جداً أن تكون هذه البيانات المُحملة ضمن المستودع متكاملة (Integrated)، ومن أجل تحقيق هذا التكامل في ظل عدم تجانس مصادر البيانات لا بدّ من استخدام أدوات لاستخراج البيانات من مصادرها وتحويلها إلى الشكل المناسب لتحميلها في مستودع البيانات، هذه الأدوات هي أدوات استخراج وتحويل وتحميل البيانات (ETL Tools) [3,4].

يختصر مصطلح ETL الكلمات الثلاثة التالية: Extract, Transformation, Load. تُعرف أدوات ETL على أنها جزء من البرنامج المستخدم يتم من خلاله استخلاص البيانات من المصادر غير المتجانسة (قاعدة بيانات، جدول بيانات، ملف، وغيرها)، ثمّ تحويلها (حذف القيم الخاطئة والمكررة، تطبيق الحسابات، ضم الحقول، وغير ذلك) لكي تتوافق مع نموذج البيانات الخاص بمستودع البيانات، وتحميلها أخيرا في مستودع البيانات[5].

قامت الشركات البرمجية بتطوير عدد كبير من أدوات (ETL) يُقدر عددها بحولي 75 أداة، تُصنف هذه الأدوات تبعاً للوظيفة وبيئة التطوير. إنّ 80% من أدوات (ETL) مُطورة حسب الطلب، وهذه تُمثل صعوبة اضافية اذ أن تعدد وتنوع أدوات (ETL) تبعاً لطلب المستخدم يجعل منها أدوات غير معيارية ولا يمكن استخدامها بشكل واسع في مشاريع مستودعات البيانات[6].

2- أهمية البحث وأهدافه

الغرض من البحث هو تقييم فعالية أداة SSIS لاستخلاص وتحويل وتحميل البيانات من مصادرها الغير مُتجانسة إلى مستودع البيانات، وذلك بغية توفير معرفة مسبقة للمستخدمين حول معايير استخدام هذه الأداة وتوافقها مع المصادر الغير مُتجانسة للبيانات. لا شك أن هناك أدوات أخرى متاحة للمستخدم أيضاً، لكننا اخترنا أكثر الأدوات شهرةً وفقاً لتقرير Gartner 2021 [7]. تمّ في هذا البحث تقييم آخر اصدار من أداة SSIS بعد تضمينها في بيئة Visual Studio 2019. تمّ استخراج وتحويل البيانات من ثلاثة مصادر (قاعدة بيانات تشغيلية Oracle، وملف اكسل Excel File، وملف نصي Text File) وتحميلها الى مستودع البيانات وذلك باستخدام الأداة SSIS.

3- دراسات سابقة

إن العديد من الأبحاث في هذا المجال ركزت على مقارنات نظرية بين أداة SSIS مع أدوات أُخرى مثل Talend, PID-Pentaho Data Integrated [9][8]، في حين تمحورت بعض الأبحاث حول مقارنة عدة أدوات ETL من خلال معايير معينة مُستندة في مُقارنتها إلى تحليل التقارير والمقالات والمجلات [6]، قدم البحث [10] في عام 2018 مقارنة عملية بين أداة SSIS وأداة PID-Pentaho Data Integrated، غير أن هذه المقارنة اقتصرت على استخلاص وتحويل وتحميل عدد قليل من سجلات البيانات (1000) سجل، بالاضافة إلى أن الاصدار المُستخدم للأداة يعود إلى عام 2006 مع استخدام SQL Server 2008. ناقش البحث [11] كيفية بناء مستودع بيانات يُساعد صُناع القرار في اتخاذ القرارات المُناسبة ضمن مؤسسة افتراضية، وناقش البحث حلول نظرية فقط وشرح كيفية استخراج وتحويل وتحميل البيانات إلى مستودع البيانات باستخدام أداة SSIS.

4- مواد البحث وطرائقه

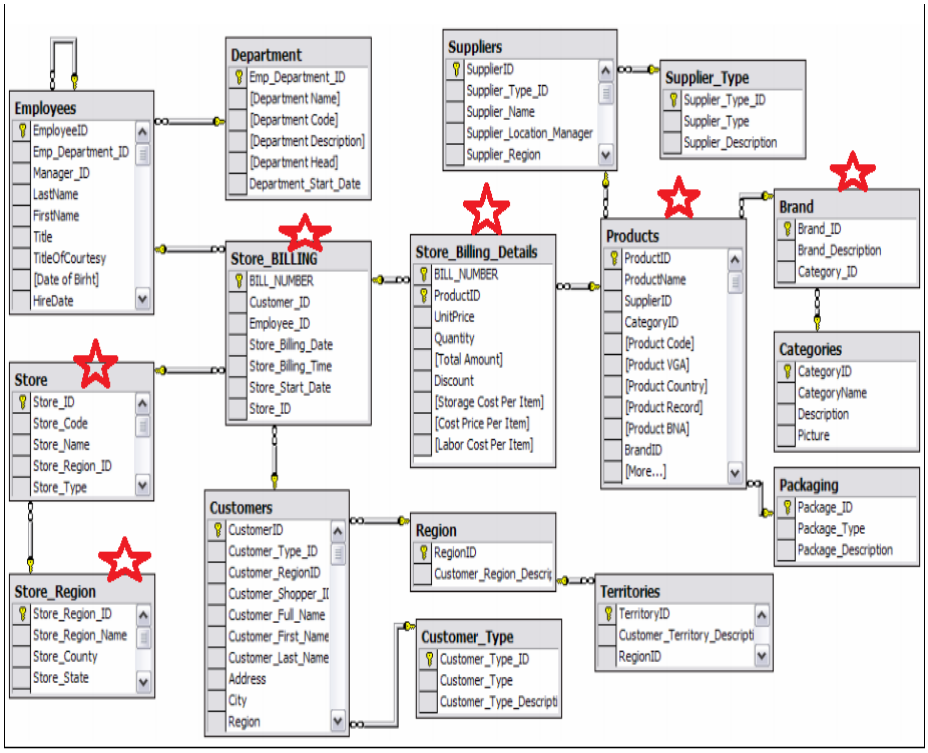
في هذا البحث تمّ اشتقاق نموذج بيانات لشركة تجارية افتراضية من قاعدة بيانات معيارية (Database Benchmark)[3]، ثمّ بناء مستودع البيانات لهذه الشركة، وبعد ذلك تمّ اجراء عملية ETL لاستخراج البيانات من قاعدة البيانات التشغيلية للشركة بالإضافة إلى مصدرين آخرين هما (Excel File) و(Text File) من خلال

الأداة SSIS، ثم تحويل البيانات إلى الصيغة المُناسبة لنموذج مستودع البيانات وتحميلها إلى المستودع.

4-1 بناء قاعدة البيانات المعيارية

تمَّ بناء قاعدة بيانات تشغيلية (Operational Database) موافقة لمبيعات شركة تجارية افتراضية لها عدة متاجر في المحافظات، وذلك بناءً على قاعدة بيانات معيارية، واستُخدم لتحقيق قاعدة البيانات المذكورة نظام إدارة قواعد المعطيات Oracle 11g.

يوضح الشكل (1) قاعدة البيانات المعيارية وإشارة فوق كل جدول تمّ استخدامه ضمن قاعدة البيانات التشغيلية المُقترحة.



الشكل (1) قاعدة بيانات معيارية لشركة مبيعات افتراضية[3]

يوضح الشكل (2) قاعدة البيانات المُقترحة والتي تحوي الجداول التالية:

1- جدول العلامة التجارية (Brand Table).

2- جدول المنتجات (Products Table).

3- جدول تفاصيل الفاتورة (Store_Billing_Details Table).

4- جدول فاتورة_متجر (Store_Billing Table).

5- جدول المتجر (Store Table).

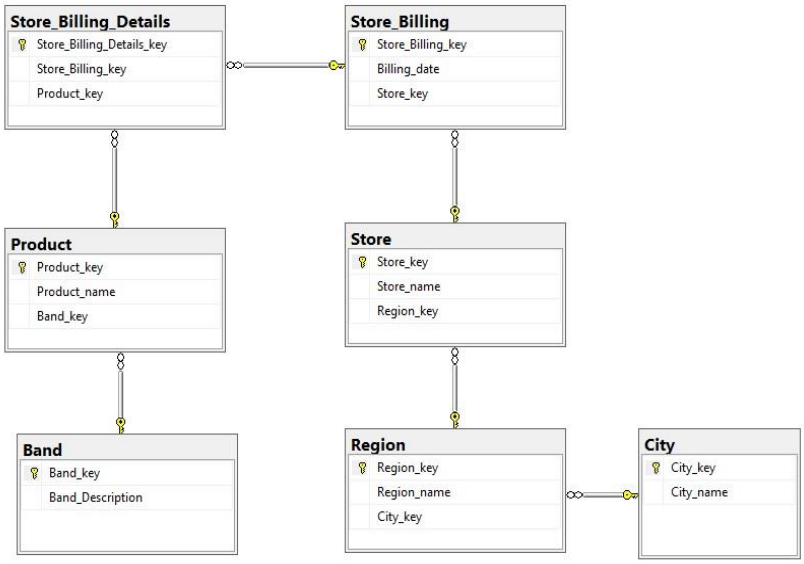
6- جدول المنطقة (Region Table).

7- جدول المدينة (City Table).

4-2 نموذج مستودع البيانات

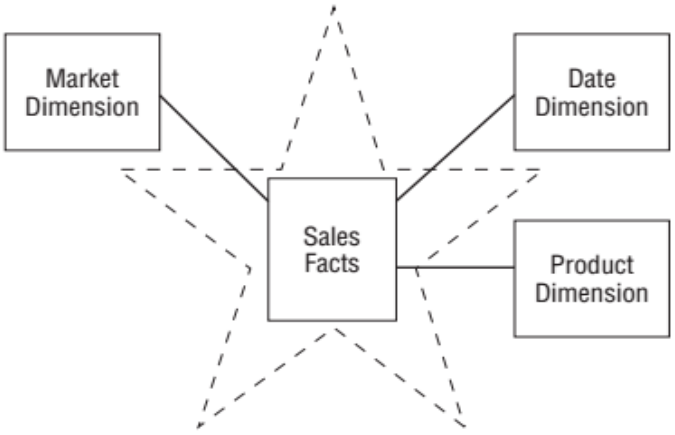
تمّ إعداد النموذج النجمي (Star Schema) لمستودع البيانات باستخدام نمذجة الأبعاد، وتمّ تحقيق هذا النموذج في بيئة SQL Server 2019،

قبل البدء بإعداد النموذج النجمي كان لا بدّ من التعرف على المفاهيم الأساسية لهذا النموذج، واستيعابها بشكل جيد. يستخدم نموذج الأبعاد ثلاثة مفاهيم أساسية لتنظيم البيانات هي: الحقائق (Facts)، الأبعاد(Dimensions)، المقاييس (Measures).



الشكل(2) نموذج ER مُقترح لمبيعات شركة تجارية افتراضية

تقوم نمذجة الأبعاد بتنظيم المفاهيم السابقة من خلال مخطط يسمى بالمخطط النجمي، يبين الشكل (3) مثال مبسط عن مخطط نجمي لمبيعات شركة تجارية:



الشكل (3) مخطط نجمي لمبيعات شركة

3-4 خطوات بناء مستودع البيانات وفق نموذج الأبعاد

يتم بناء نموذج الأبعاد باتباع أربع خطوات هي:

1- تحديد عملية الأعمال (Choose the business process)

يتم في هذه المرحلة وصف المهام التي يجب أن ينفذها النموذج، وبمعنى آخر دراسة المتطلبات وإجراء التحليلات اللازمة.

2- تحديد مستوى التمثيل الأدنى (Declare the grain)

يتم في هذه المرحلة تعريف الحقائق المطلوب تنفيذها من قبل النموذج، بمعنى آخر ماذا يمثل كل سطر بيانات يتم إدخاله إلى جدول الحقيقة، وهذا ما يُسمى بتحديد أصغر مستوى تمثيل.

الخطأ الأكثر شيوعاً هو عدم تحديد أصغر مستوى تمثيل في بداية عملية التصميم، إذ أنّ التصميم بأكمله يعتمد على ذلك، قد يتم إكتشاف أخطاء في الخطوة 3 أو 4 من عملية التصميم ناتجة عن التحديد الخاطئ لأصغر مستوى تمثيل. لذلك يتحكم على المصمم العودة إلى الخطوة 2، وإعادة ضبط ذلك بشكل صحيح، ثمّ إعادة زيارة الخطوتين 3 و 4 مرة أخرى.

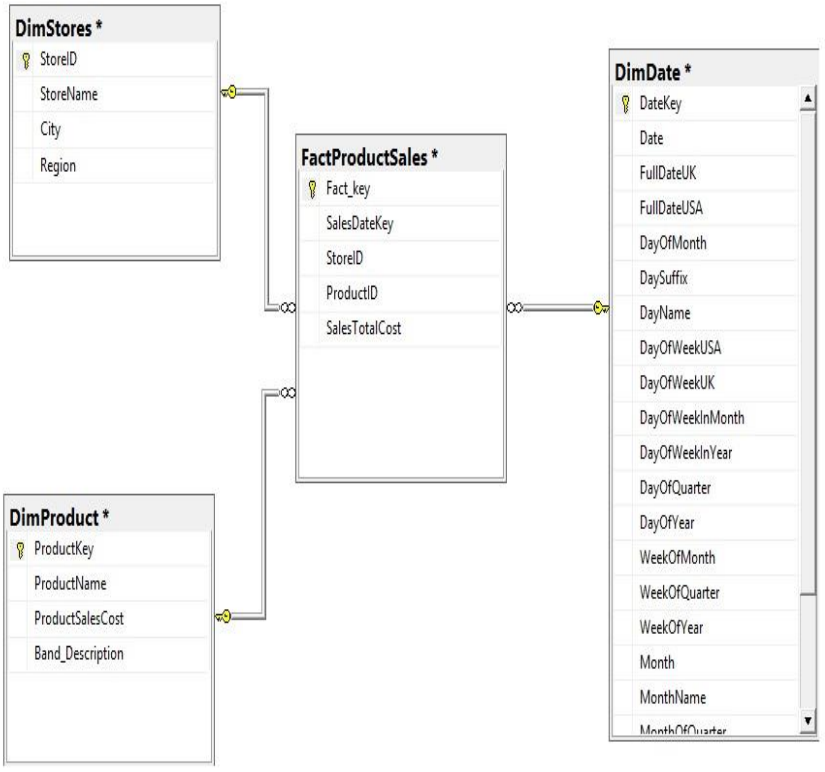
3- تعريف الأبعاد (Identify the dimensions) وبناء جداولها.

بعد أن تمَّ تعريف المتطلبات و توحيد قياسات جدول حقائق النموذج، يجب إختيار الأبعاد التي تمثل جميع الوصوفات المحتملة التي تأخذ قيماً مفردة في سياق كل قياس.

4- تعريف الحقائق (Identify the fact) وبناء جدولها.

تُعتبر المرحلة الأخيرة من بناء نموذج الأبعاد هي بناء جدول الحقائق. يجب أن تكون جميع الحقائق المرشحة للإستخدام في التصميم مطابقة لمستوى التمثيل الأصغر المُعرف في الخطوة 2.

يوضح الشكل(4) نموذج الأبعاد الموافق لمبيعات شركة تجارية، حيث تمَّ بناء النموذج بعد تطبيق المراحل الأربعة آنفة الذكر.



الشكل (4) نموذج الأبعاد لمبيعات شركة تجارية

4-4 إدخال البيانات الى قاعدة البيانات التشغيلية

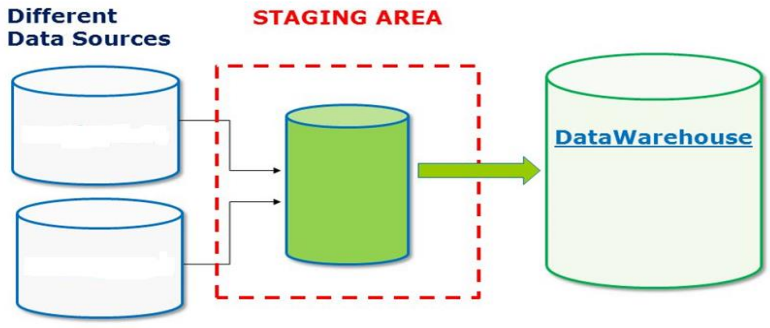
تم إدخال بيانات تجريبية عشوائياً إلى قاعدة البيانات التشغيلية بهدف شحنها بالبيانات التي سيتم لاحقا إجراء عمليات الاستخراج والتحويل والتحميل لها إلى مستودع البيانات الهدف، سيتم ذلك من خلال الأداة SSIS. تمّ ادخال البيانات إلى الجداول Region, Band, Store, City, Product من خلال عبارة insert، أمّا الجدولين Store_Billing, Store_Billing_Details تمّ استخدام إجرائية (procedureGenerateRandomData) وذلك لأن هذين الجدولين يحويان سجلات كثيرة ممكن أن تصل بياناتها الى مرتبة الملايين.

4-5 الخطوات الأساسية لعملية ETL

هناك ثلاثة خطوات أساسية لإجراء عملية ETL وهي:

1- استخلاص البيانات (Data Extraction)

هي الخطوة الأولى في عملية جلب البيانات إلى مستودع البيانات. في هذه المرحلة يتم قراءة البيانات من المصادر الغير متجانسة إلى منطقة تُسمى منطقة تقديم البيانات (Data staging area)، وهي منطقة تخزين متوسطة تقع بين أنظمة المصدر و مستودع البيانات، وهي عادةً منطقة تخزين مُؤقتة حيث تُحذف منها البيانات بعد تحميلها إلى المستودع بنجاح. لانه قد يؤدي نقل البيانات مباشرة إلى المستودع إلى أخطاء في نقل البيانات. تُستخدم منطقة تقديم البيانات لتنقية البيانات وتنظيمها. يوضح الشكل (5) عملية اسخلاص البيانات من مصادرها الغير مُتجانسة إلى منطقة تقديم البيانات.



الشكل (5) استخلاص البيانات (Data Extraction)

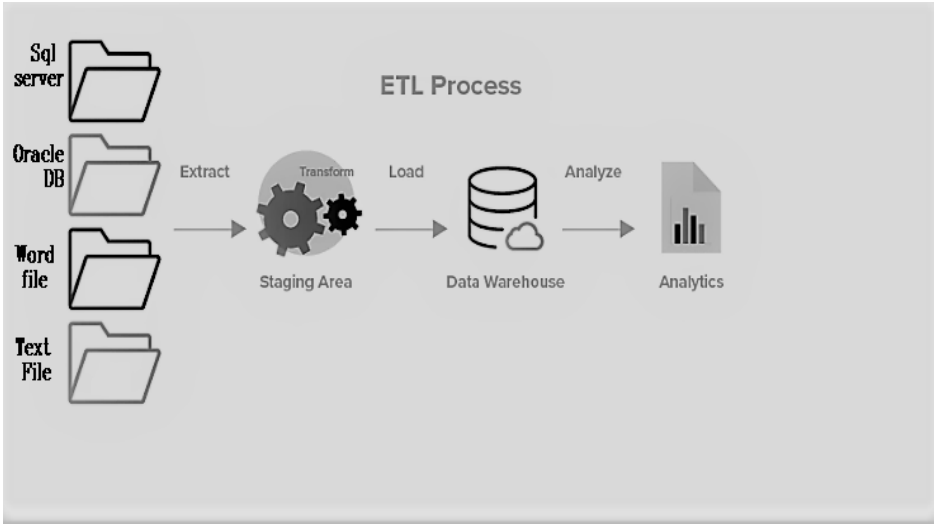
2- تحويل البيانات (Data Transformation)

يتم في هذه المرحلة حذف البيانات الغير مُنمذجة في مستودع البيانات، وإزالة البيانات المكررة، وتغيير التنسيق لبعض البيانات عند اللزوم لكي يتوافق مع التنسيق الموجود ضمن المستودع، وإجراء العمليات الحسابية اللازمة على البيانات.

إن الإجراءات المذكورة سابقاً تُؤمن تكامل البيانات (Data integration) ضمن مستودع البيانات، هذا التكامل يُعزز التحليل الفعال للبيانات.

3- تحميل البيانات (Data Loading)

في هذه المرحلة يتم تحميل البيانات التي تم استخراجها في المرحلة السابقة إلى مستودع البيانات، يجب مراعاة تحميل البيانات عندما يكون مستودع البيانات معطل عن العمل (Offline). يوضح الشكل (6) الخطوات الأساسية لعملية ETL الآنفة الذكر.



الشكل (6) الخطوات الأساسية لعملية ETL

6-4 أداة SSIS

هي أداة مرنة وسريعة من انتاج شركة (Microsoft)، تُستخدم من أجل استخراج البيانات من مصادر بيانات مُتعددة وتحويلها إلى الشكل المُناسب لتحميلها في مستودع البيانات. يتم تضمين هذه الأداة في بيئة SQL Server 2019.

أهم مكونات أداة SSIS التي تمّ استخدامها في هذا البحث هي:

1- Data Flow Task

هي أهم مكون من مكونات الأداة، حيث يتم من خلال هذه الأداة التحكم بتدفق البيانات من المصدر الى المآل.

2- ADO NET Source

يتم من خلال هذا المكون الربط مع مصدر بيانات بغرض الوصول إلى البيانات في قاعدة بيانات مثل Oracle Database، Microsoft SQL Server، MySQL، PostgreSQL وغيرها الكثير، في البحث نحتاج لهذا المكون من أجل الربط مع قاعدة البيانات التشغيلية التي تمّ تصميمها في نظام ادارة قواعد بيانات Oracle.

3- OLE DB Destination

يتم من خلال هذا المكون الربط مع المآل الذي تمّ بناؤه في نظام ادرة قواعد بيانات Microsoft SQL Server 2019.

4- Excel Source

من خلال هذا المكون يتم الربط مع مصدر بيانات Excel.

5- Flat File Source

من خلال هذا المكون يتم الربط مع مصدر بيانات Text File.

6- Sequence Container

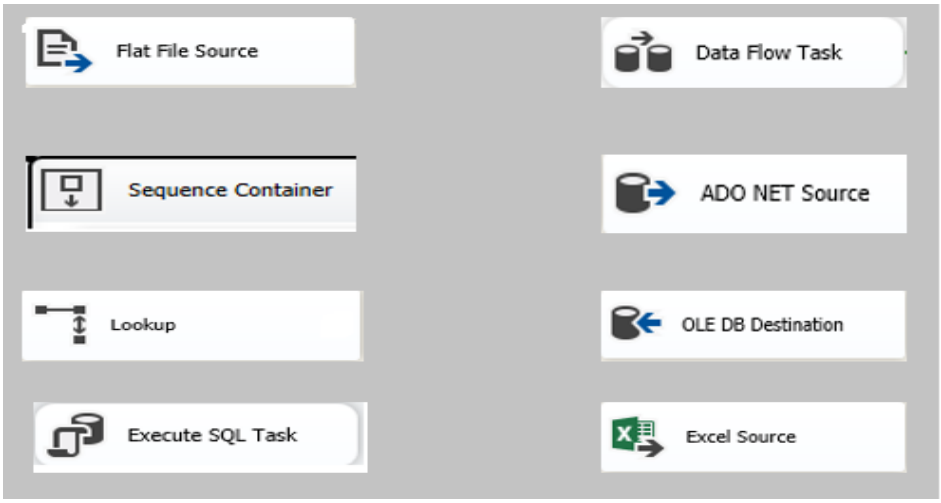
إن تحميل البيانات الى مستودع البيانات يتم من خلال تسلسل مُعين للإدخال إلى جداول البيانات، حيث لا يمكن إدخال البيانات الى جدول الحقيقة قبل الادخال إلى كافة الأبعاد، لان جدول الحقيقة يحوي مفاتيح أجنبية لكافة الأبعاد في المستودع.

7- Look up

يقوم هذا المكون بعملية بحث في الجدول المآل عن القيم الغير موجودة في المصدر، حيث يتم إضافة السجلات الغير موجودة فقط في جدول المآل.

8- Execute SQL

يُعتبر من أهم مكونات أداة SSIS حيث يقوم هذا المكون بتنفيذ استعلام من بيانات المصدر بحيث تُصبح البيانات موافقة لجدول البيانات الموجود في المستودع. يوضح الشكل (7) المكونات آنفة الذكر.



الشكل (7) مكونات SSIS التي تمّ استخدامها في البحث

7-4 استخراج وتحويل البيانات من المصادر الغير مُتجانسة

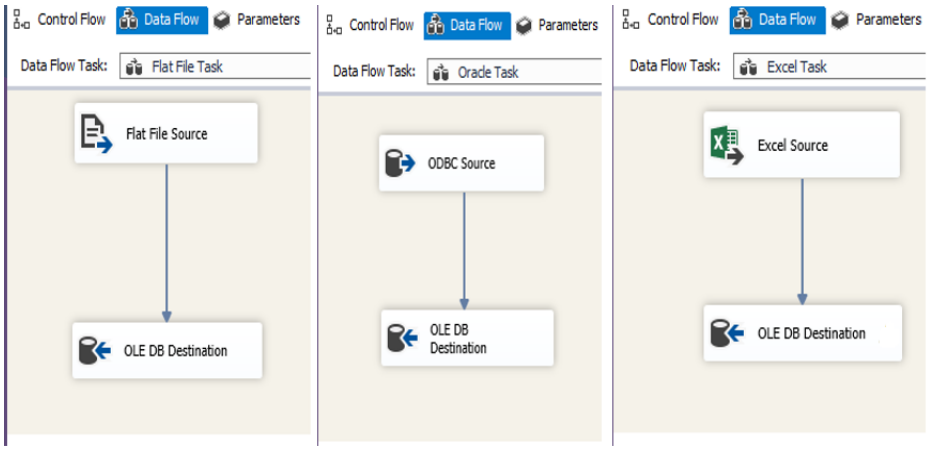
سيتم استخراج البيانات من مصادرها الغير مُتجانسة إلى مستودع البيانات باستخدام الأداة SSIS، يوجد ثلاثة مصادر للبيانات وهي:

1- قاعدة بيانات تشغيلية لمبيعات الشركة التجارية الافتراضية موضحة بالشكل (2).

2- مبيعات أحد متاجر الشركة يتم تخزين بياناتها ضمن ملف نصي Text File.

3- مبيعات لمتجر آخر للشركة يتم تخزين بياناتها ضمن ملف إكسل Excel File.

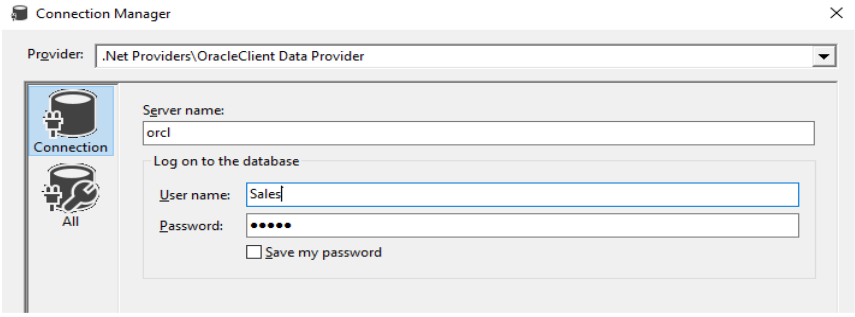
وبالتالي لدينا ثلاثة مهام يتوجب على الأداة تنفيذها متوافقة مع المصادر الثلاثة للبيانات يوضحها الشكل (8).



الشكل (8) مهام SSIS

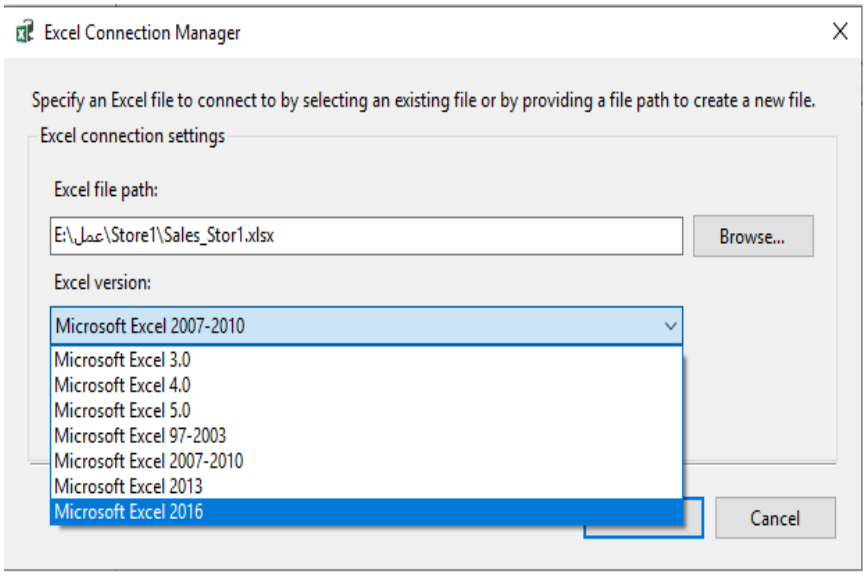
يتم تنفيذ هذه المهام وفق المراحل التالية:

1- تأمين الربط مع مصدر البيانات: يتم ذلك من خلال الربط مع مخدم قاعدة البيانات التشغيلية لمبيعات الشركة التجارية كما هو مُوضح بالشكل (9)، حيث يتم تحديد اسم المخدم واسم قاعدة البيانات التشغيلية وكلمة المرور الخاصة بالقاعدة.

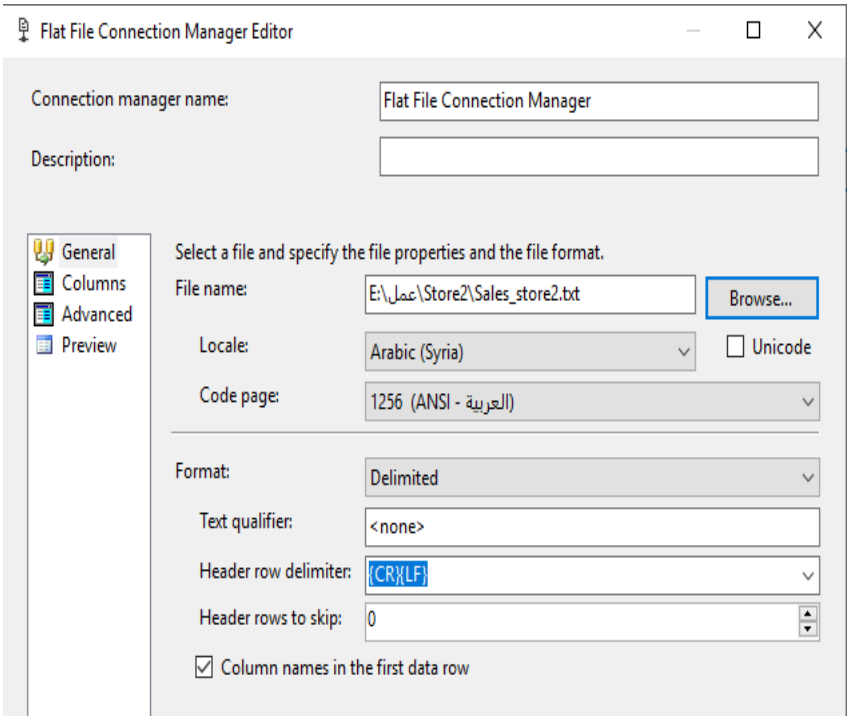


الشكل (9) الربط مع قاعدة البيانات التشغيلية Oracle

يتم الربط مع الملف النصي (Text file) من خلال تحديد مسار الملف كما هو مُبين بالشكل(10)، وكذلك الأمر مع ملف الإكسل (Excel File) يتم الربط من خلال تحديد المسار وإصدار الإكسل المُتوفر كما هو موضح بالشكل(11).



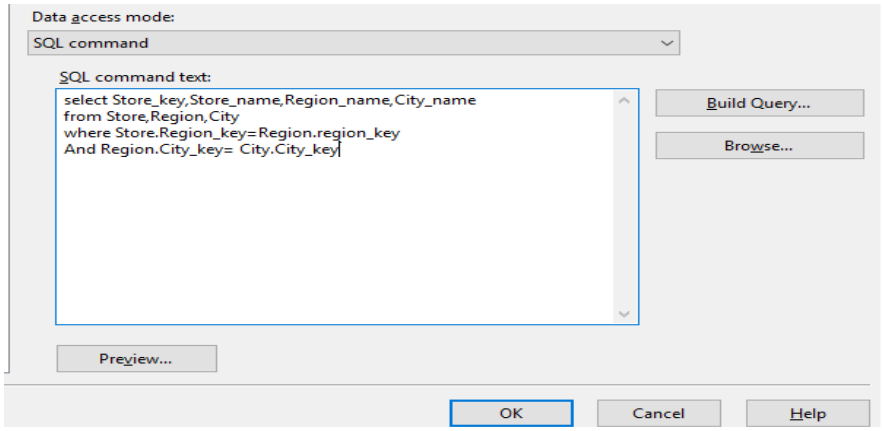
الشكل (10) الربط مع مصدر بيانات Excel File



الشكل (11) الربط مع مصدر البيانات Text File

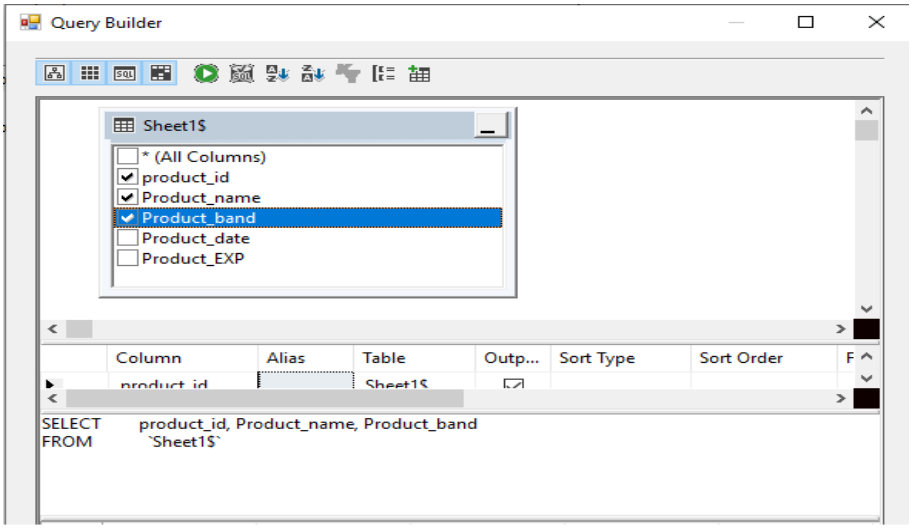
2- استخراج وتحويل البيانات إلى الصيغة الموافقة لنموذج مستودع البيانات: بعد أن تمّ تأمين الربط مع مصادر البيانات يتم استخراج البيانات من مصادرها وتحويلها إلى الصيغة الموافقة لنموذج البيانات الخاص بمستودع البيانات المُوضح بالشكل(4). تُؤمن أداة SSIS استخراج البيانات من المصادر بصيغ مُختلفة أهمها صيغة جدول كامل Full Table وعلى شكل أمر استعلام SQL Command، يتم الاختيار وفق نموذج مستودع البيانات، فبعد المتجر Dim_Store يتم استخراج بياناته من ثلاثة جداول لقاعدة البيانات التشغيلية وفق أمر استعلام كما هو مُوضح بالشكل(12).

بالنسبة لمصدر البيانات Text File تُؤمن أداة SSIS إمكانية تحديد فواصل الأعمدة، وتحديد فيما اذا كان السطر الأول من الملف يعبر عن أسماء الأعمدة، وأنواع البيانات Data Type لتوافق نوع البيانات لمستودع البيانات كما هو موضح في الشكل(11).



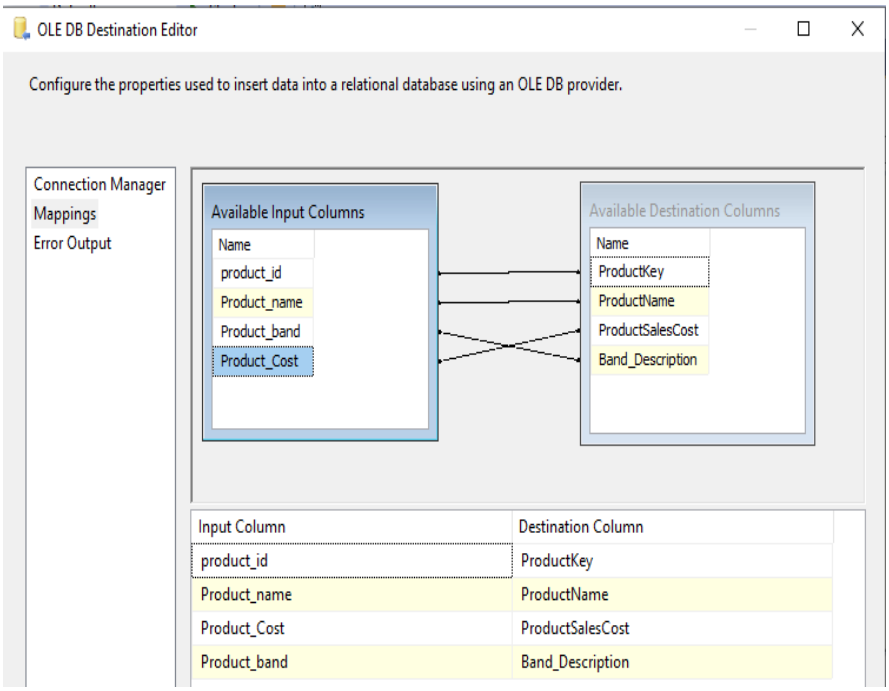
الشكل (12) استخراج وتحويل البيانات وفق SQL Command

يتم استخراج البيانات من مصدر البيانات Excel على شكل جدول بيانات، في حال كانت الصيغة موافقة لنموذج البيانات يتم ذلك بصيغة جدول، أما في حال وجود اختلاف فإن أداة SSIS تدعم إمكانية استخراج البيانات من الملف إكسل على شكل استعلام قاعدة بيانات، تُسمى هذه العملية ببناء الاستعلام Query Builder، يُوضح الشكل (13) عملية بناء الاستعلام لملف إكسل يحوي بيانات المنتجات التي يتوجب تخزينها في بعد المنتج DimProduct، من خلال ذلك يتم جلب البيانات التي نحتاج لها في عملية التحليل فقط.



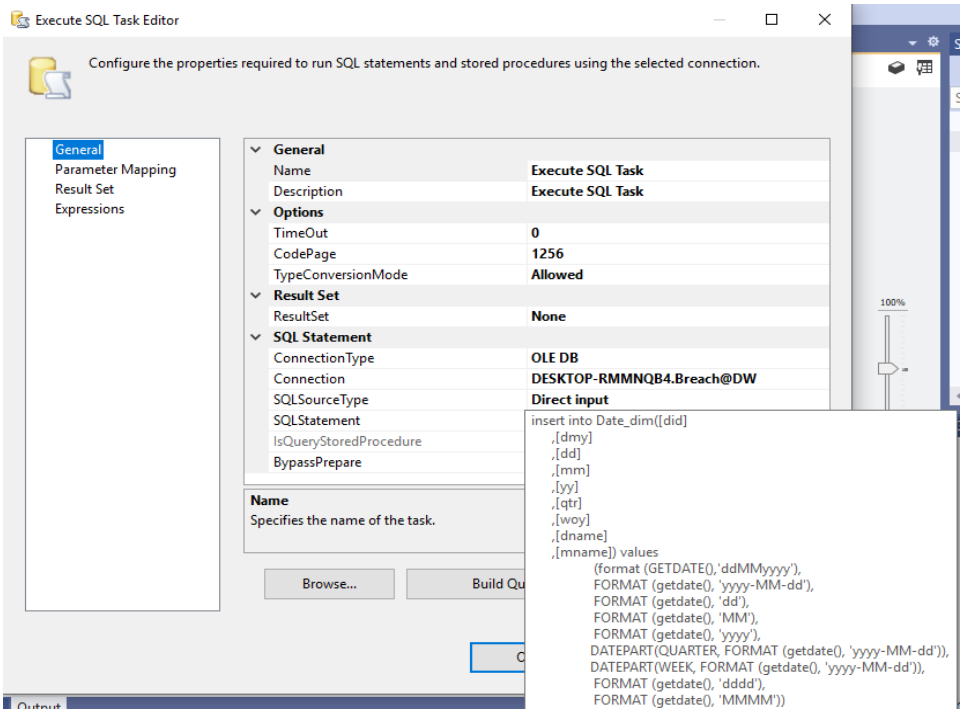
الشكل (13) بناء الاستعلام في ملف Excel File

3- الربط بين مصادر البيانات ومستودع البيانات وتحميل البيانات من مصادرها إلى المستودع: بعد أن تم استخراج البيانات وتحويلها إلى الشكل المُوافق لنموذج مستودع البيانات يتوجب ربط نتائج التحويل مع مستودع البيانات، تُؤمن أداة SSIS هذا الربط من خلال عملية Mapping كما هو مُوضح بالشكل (14)، حيث يتم ربط كل عمود من المصدر مع العمود الموافق له في مستودع البيانات. تُنفذ هذه العملية لربط كافة نتائج الاستخراج والتحويل مع مستودع البيانات.



الشكل (14) عملية الربط Mapping

إن الخطوة الأولى لتحميل البيانات إلى المستودع تتمثل بإضافة سجل تاريخ اليوم إلى البعد DimDate، يتم ذلك باستخدام المكون Execute SQL في أداة SSIS كما هو مُبين بالشكل (15).



الشكل (15) اضافة سجل تاريخ اليوم الى DimDate

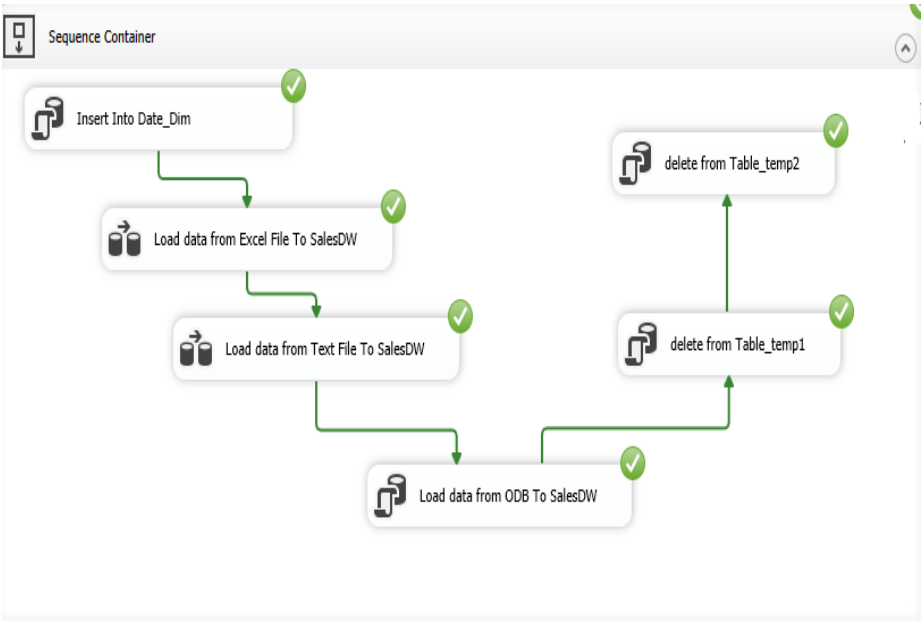
بعد إضافة سجل تاريخ اليوم يتم تحميل البيانات من مصادرها الغير مُتجانسة إلى مستودع البيانات بتنفيذ الحزمة Package التي تعمل على تحميل البيانات في المستودع وفق تسلسل يوافق نموذج مستودع البيانات المُوضح بالشكل (4)، حيث يتم تحميل البيانات الى الأبعاد أولاً ثمّ إلى جدول الحقائق، يتم ذلك باستخدام المكوّن Sequence Container، يوضح الشكل (16) تنفيذ الحزمة التي تقوم بتحميل البيانات من مصادرها إلى مستودع البيانات بعد إضافة سجل تاريخ اليوم إلى بعد الزمن.

عند تنفيذ حزمة SSIS يتم وبشكل متسلسل تنفيذ ما يلي:

1- تحديث بعد التاريخ.

2- استخراج وتحويل البيانات من ملف إكسل Excel File وتحميلها إلى مستودع البيانات.

3- استخراج وتحويل البيانات من ملف نصي Text File وتحميلها إلى مستودع البيانات.



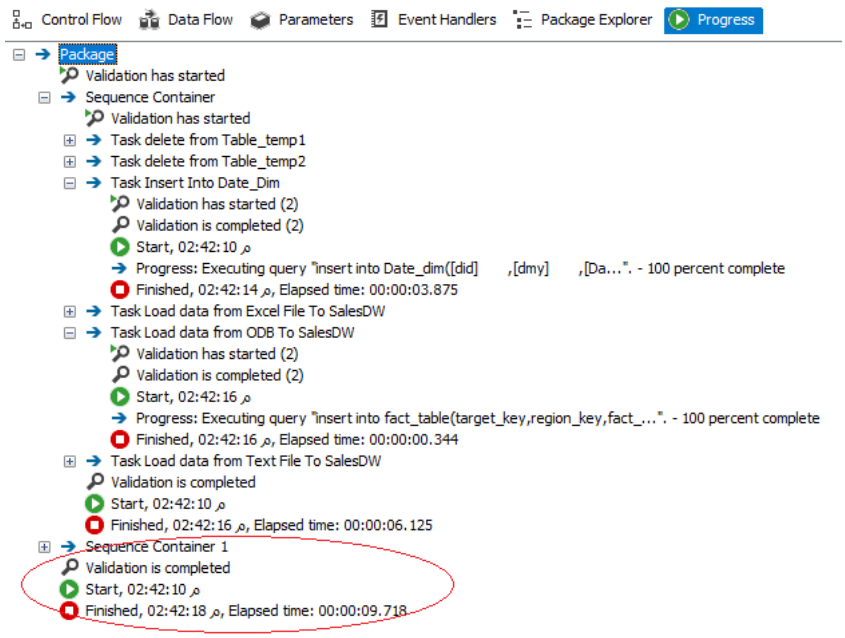
الشكل (16) حزمة SSIS لتحميل البيانات من مصادر البيانات المستودع البيانات

4- استخراج وتحويل البيانات من قاعدة البيانات التشغيلية وتحميلها إلى مستودع البيانات.

5- حذف البيانات من منطقة تقديم البيانات Data Stage Area، وهي المنطقة المتوسطة بين مصادر البيانات ومستودع البيانات، تُخزن فيها البيانات اللازمة لعملية التحويل بشكل مؤقت، حيث يتم حزفها بعد التحميل الناجح للبيانات إلى مستودع البيانات.

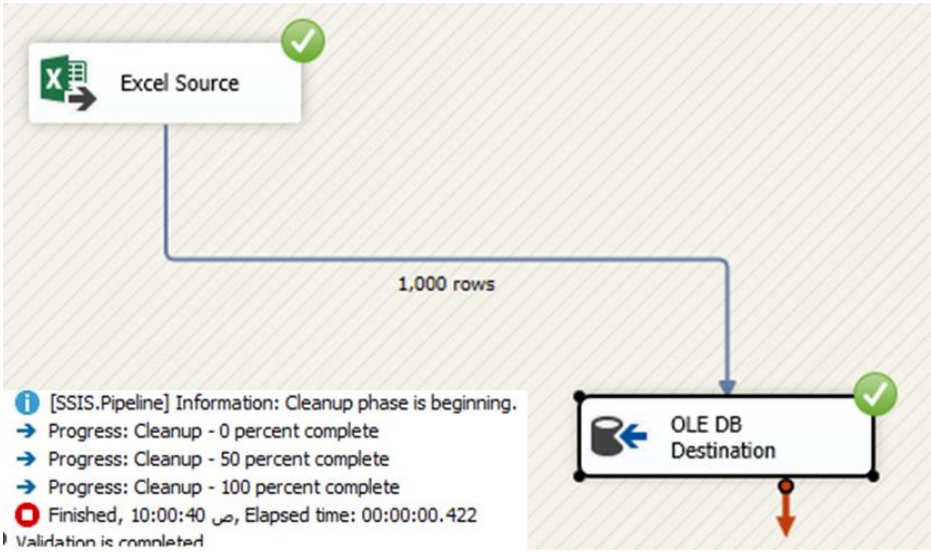
يوضح الشكل (17) الوقت اللازم لتنفيذ الحزمة بشكل مُفصل، حيث يُمكن معرفة زمن تنفيذ كل مهمة تدفق للبيانات ضمن الحزمة. احتاج استخراج وتحويل البيانات من مصادرها الغير متجانسة ثمّ تحميلها الى مستودع البيانات زمن قدره 9.718 ثانية فقط.

بالمقارنة مع نتائج البحث [10] الذي استخدم أداة SSIS بنسختها 2008، فإن استخراج وتحويل وتحميل 1000 سجل من مصدر Excel File إلى مستودع بيانات مبني في بيئة SQL Server 2008 احتاج الى 0.655 ثانية.



الشكل (17) زمن تنفيذ حزمة SSIS

في ظروف تجريبية مُشابهة وباستخدام أداة SSIS قيد البحث ومستودع بيانات مبني في بيئة SQL Server 2019 احتاج استخراج وتحويل وتحميل 1000 سجل بيانات من مصدر Excel File إلى مستودع البيانات 0.422 ثانية، يُبين الشكل (18) نتيجة تنفيذ عملية الاستخراج والتحويل والتحميل مع زمن التنفيذ.



الشكل (18) استخراج وتحويل وتحميل 1000 سجل من مصدر Excel File الى مستودع البيانات

5- مناقشة النتائج

تمّ في هذا البحث تقييم أداة SSIS في استخراج وتحويل وتحميل البيانات من مصادرها الغير مُتجانسة إلى مستودع البيانات والتوصل إلى النتائج التالية:

1- أبدت أداة SSIS فعالية ومرونة وسهولة في الاستخدام نسبياً عند التعامل مع مصادر غير مُتجانسة للبيانات.

2- إنّ أداة SSIS تُمكن المستخدم من تحويل البيانات إلى تنسيقات وأنواع مُختلفة بحيث يُؤمن التُوافق الضروري مع تنسيق ونوع البيانات المُستخدم في نموذج مستودع البيانات، بالإضافة الى إمكانية عرض نتائج التحويل للتأكد من موافقتها لنموذج البيانات قبل تحميلها إلى المستودع، وبذلك تحمي المستخدم من الوقوع في الخطأ.

3- يمكن للمستخدم عند استخدام أداة SSIS التحكم بتدفق البيانات من مصادر البيانات إلى مستودع البيانات.

المراجع

[1] KIMBALL R., 2013 –**The Data Warehouse Toolkit**. 3rd Ed, USA, pp: 37–109.

[2] INMON W.H, 2002 – **Building the Data Warehouse**. 3rd Ed, USA, pp 81–145.

[3] BALLARD C.; FARRELL D.; GUPTA A.; MAZUELA C.; VOHNIK S., 2006 – **Dimensional Modeling in a Business Intelligence Environment**. 1st Ed, USA, pp 47–73.

[4] **https://www.tutorialspoint.com/etl_testing/index.htm.**

[5] **https://www.alooma.com/blog/etl-tools-comparison**.

[6] Runtuwene J P A., Tangkawarow I R H T., 2018 – **A Comparative Analysis of Extract, Transformation and Loading (ETL) Process**. IOP Conference Series: Materials Science and Engineering, 306(12066).

[7] **https://www.gartner.com/reviews/market/data-integration-tools 2021**.

[8] SOUIBGUI M.; ATIGUI F.; ZAMMAI S.; CHERFI S.; BEN YAHIA S., 2019 – **Data quality in ETL process: A preliminary study**. ScienceDirect Procedia Computer Science, 159(676).

[9] TEJNANI P., 2017 – **Study of ETL Tools: Talend Open Studio and SQL Server Integration Services (SSIS).** International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2321(9653).

[10] AMINE A.; AIT DAOUD R.; BOUIKHALENE B., 2021 – **Performance and analyses using two ETL extraction**

software solutions. International Journal on Optimization and Applications IJOA, 1(2).

[11] YANGENG W,. LEI W,. 2011 - **The Research of Enterprise Data Center Based on SSIS**. International Conference on Business Management and Electronic Information, 12074510.