مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (6) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (6) 2018

تصميم مستودع بيانات لتحليل بيانات نظم صيانة الاليات

د. علي سليمان*

(تاريخ الإيداع 27 / 9 / 2018. قُبِل للنشر في 21/ 11 / 2018)

□ ملخّص □

تم في هذه الورقة تصميم نظام يعتمد على تقنية مستودعات البيانات والنمذجة البعدية لقواعد البيانات قادر على استقبال البيانات المتعلقة بصيانة الاليات وتحليلها بهدف الوصول إلى نتائج تساهم في اتخاذ القرارات التي قد تتعلق بحالة الاليات او انواع الاعطال او زمن وقوعها، تم الحصول على البيانات من محطة حاويات اللاذقية الدولية.
يعتمد النظام على مخدم من نوع *Compaq* ومحرك قواعد البيانات *Oracle 11* و*Oracle Developer* لإنشاء التقارير وعمليات البحث.
بينت النتائج ان استخدام تقينات مستودعات البيانات في تحليل بيانات نظم صيانة الاليات المعتمدة على الحاسب يخفض من الزمن اللازم لانجاز الاستعلامات على البيانات كما يسمح بتوليد العديد من التقارير التي يحتاج توليدها باستخدام قاعدة البيانات التقليدية الى كتابة عدد كبير من الاستعلامات يدويا.

الكلمات المفتاحية: مستودع بيانات – OLAP – صيانة الاليات باستخدام الحاسب

* أستاذ مساعد – قسم الهندسة الطبية– كلية الهندسة الميكانيكية والكهربائية– جامعة تشرين – اللاذقية – سورية

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (6) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (6) 2018

Design of Data warehouse to Analyze CMMS Data

Dr. Ali Suleiman*

(Received 27 / 9 / 2018. Accepted 21 / 11 / 2018)

□ ABSTRACT □

In this paper, a system based on data warehouse technology and remote database modeling is designed to receive data related to the maintenance and analysis of the machines in order to reach results that contribute to the decision making that may relate to the state of machines or types of faults or the time of occurrence. Lattakia International. The system relies on a Compaq server, an Oracle 11 database engine, and Oracle Developer to create reports and searches.

The results showed that the use of data warehouse technologies in the analysis of computer maintenance systems data reduces the time required for the completion of queries on the data and allows generation of many reports that need to be generated using the traditional database to write a large number of queries manually.

Keywords: data warehouse – OLAP – CMMS

* **Associate Professor, Department of Medical Engineering, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria.** suleiman_a@hotmail.com

مقدمة

ترى الجمعية الهندسية الأسترالية للصيانة Maintenance Engineering Society of Australia MESA أن الصيانة هي الوصول إلى القدرات المطلوبة من التجهيزات ضمن ظروف العمل أو الظروف الاقتصادية. كما تتضمن الصيانة أيضاً القرارات الهندسية والأفعال المرافقة الضرورية من أجل الحصول على القدرات المثلى للتجهيزات،أي أن إنجاز وظيفة محددة ضمن مجال من مستويات الأداء مرتبط بالاستطاعة، النوعية، الأمان والمسؤولية.[1] تحتاج عمليات الصيانة المعقدة في المؤسسات إلى دعم قرار عمليات الصيانة من خلال أنظمة حاسوبية تدير عمليات الصيانة تسمى هذه الأنظمة بنظم إدارة الصيانة باستخدام الحاسب CMMS، و تستثمر الشركات الآن الملايين من الدولارات والجنيهات في تكنولوجيا المعلومات، ويتمثل الهدف الذي تسعى إلى تحقيقه كل هذه الشركات في جمع المعلومات و استغلالها بشكل فعال مما يوفر لها قرارات أسرع و أفضل فيما يتعلق بأعمالها، ويتم ذلك بمعونة نظم دعم القرار والتي ازدادت أهميتها مع ازدياد حجم البيانات في المؤسسة و الحاجة الماسة إلى وضع أداة مناسبة في يد متخذ القرار تسمح بالتعامل مع الكميات الهائلة من المعلومات المتاحة[2].

مشكلة البحث

مع ازدياد تعقيد الأعمال في المؤسسات وازدياد كمية البيانات التي تتعامل معها هذه المؤسسات لم تعد أنظمة قواعد المعطيات العادية قادرة على تلبية متطلبات متخذ القرار من حيث السرعة و الدقة وحجم البيانات الذي تستطيع التعامل معه، لذلك ظهرت فكرة مستودعات البيانات والتي تعد من أقوى أدوات نظم دعم القرار DSS، ويعد بيل انيمون Bill Inmon رائدا لهذه الفكرة،[3] فقد برز تحدّ جديد في كيفية تحويل قواعد البيانات من قواعد تخزين وبحث عن المعلومة إلى مخازن للمعلومات تستنتج المعرفة وتساعد في اتخاذ القرار، لذلك أصبح من الضروري وجود أنظمة معلوماتية جديدة تتعامل مع هذه البيانات من حيث التخزين والاسترجاع والعرض بهدف المساعدة في التخطيط وبناء الرؤى المستقبلية وخصوصاً في مجال مثل صيانة الاليات في المنشات التي تعتمد على الاليات في عملها بشكل رئيسي وتحتاج الى النظر الى البيانات المتوفرة من مناظير مختلفة لاتخاذ القرار على وجه السرعة فيما يخص التعامل مع الوضع الراهن للاليات مقارنة مع نظم صيانة تقليدية، تعتمد في إدارتها لهذه البرامج على الخبرة الذاتية المتراكمة لمهندسي وفني الصيانة وعلى توقعاتهم الشخصية وبالرغم من وجود بيانات حول أوقات التشغيل و العطل وكذلك توفر وسائل تقنية حديثة كالحاسوب التي تسهم في مساعدة الإدارة في المحطة قيد الدراسة على تبني برامج صيانة أكثر فاعلية لها الأثر البالغ في تعزيز وتحسين الأداء الوظيفي لآليات المحطة.

أهمية البحث

تعد مستودعات البيانات بأسلوب بنائها و تمثيل بياناتها أسهل بالنسبة للمستخدم من ناحية الاستعلام و الفهم أكثر من قواعد البيانات العلاقاتية التي صممت للتعامل مع العديد من العمليات، كما أن الاستعلامات التي سوف تكون معقدة جدا في قواعد البيانات التي خضعت لعملية التقييس Normalized Database سوف تصبح أسهل في التصميم و الاستخدام في مخازن البيانات التي لا تخضع فيها قاعدة البيانات للعملية السابقة، وخصوصا في مجال صيانة الاليات التي تتدخل فيها العديد من العوامل لانجاز العمل المطلوب.

أهداف البحث

يهدف البحث الى وضع تصميم لمستودع بيانات يخدم نظام صيانة الاليات باستخدام الحاسب ويسمح بالحصول على نتائج الاستعلامات المختلفة بكفاءة عالية مقارنة مع الاستعلام من قواعد البيانات التقليدية.

طرائق البحث ومواده

تمّ اتباع المنهج التجريبي *Experimental Method* للتحقق من فرضيات البحث، حيث تمّ تثبيت المتغيرات المستقلة، ودراسة القيم التي تمّ الحصول عليها للمتغيرات التابعة، كما تمّ اتباع المنهج الوصفي *Descriptive Method* من خلال استخدام التحليل الإحصائي، وذلك للتحقق من زمن تنفيذ الاستعلامات.

مجتمع وعينة البحث

تم تطبيق البحث على محطة حاويات مرفأ اللاذقية الدولية *Lattakia International Container Terminal(LICT)* نظراً لأنها تلعب دوراًكبيراً في مجال النقل البحري والبري، وقد حصلت مجموعة *CMA- CGM* الفرنسية العالمية على عقد إدارة وتشغيل محطة حاويات مرفأ اللاذقية بعد المنافسة مع عدة شركات عالمية بتاريخ 2009/10/1، لما يحظى موقع سورية الجغرافي من أهمية عالمية، فأعطت الإدارة الحديثة اهتماماً كبيراً لصيانة الآليات وخصصت لها ورشة صيانة بمساحة كبيرة مع مفهوم حديث للصيانة بإدخال المعلوماتية في برامجها.

نظم دعم القرار

يمكن تعريف نظام دعم القرار *Decision Support System (DSS)* بأنه تكنولوجيا متقدمة للمعلومات معتمدة على الحاسب وتستخدم لدعم اتخاذ القرارات المركبة وحل المشكلات واختبار السياسات ومحاكاة السيناريوهات والتخطيط الاستراتيجي.*[4]*

ظهرت بدءاً من أوائل التسعينيات أربعة أدوات قوية لبناء أنظمة دعم القرار (بجانب أنظمة النماذج وأنظمة إدارة قواعد النماذج)، وكانت أول أداة جديدة لدعم القرار هي مستودع البيانات*(Data Warehouse)*، وأعقب ظهور مستودع البيانات أداتان جديدتان هما المعالجة التحليلية الآنية أو المباشرة (*On-Line Analytical Process (OLAP*) والتنقيب في البيانات*(Data Mining)*، أما مجموعة الأدوات الرابعة الجديدة فكانت التكنولوجيا المرتبطة بشبكة المعلومات العالمية – الانترنت.*[5]*

الوكالات الحكومية، المعاهد العلمية والأعمال خصصت موارد ضخمة لجمع وتخزين البيانات، وفي الحقيقة فإن جزءا صغيرا فقط من هذه البيانات سوف يستخدم في العديد من الحالات، لأنه سوف يكون من الصعب إدارة الحجوم الكبيرة من البيانات، أو أن المعطيات ستكون معقدة بشكل كبير من أجل عملية تحليل فعالة لهذه المعطيات.

والملامح المميزة لهذه التكنولوجيا الجديدة الناشئة والذي يفصلها عن غيرها من الأنظمة المعتمدة على الحاسب مثل أنظمة المعلومات الإدارية*(Management Information Systems (MIS)* أو الأنظمة الخبيرة *Expert Systems*، هو اعتمادها على النماذج الرياضية وأساليب المحاكاة وغير ذلك من الأدوات الكمية التي تعد وسيلة تجريبية مناسبة لعملية لدعم القرار، حيث أنها قد صممت على وجه خاص لتقييم القرارات البديلة واختبار السياسات المقترحة وتقييم أثر قواعد تشغيل النظم وإعداد إسقاطات الأداء المستقبلي للمؤسسات.

مستودعات البيانات

تعرّف مستودعات البيانات بأنها أنظمة تستعمل تقنيات جديدة في تخزين كميات كبيرة من المعلومات المتكاملة و الغير متشابهة بهدف استعمالها في سرعة اتخاذ القرار لذلك تسمى أنظمة تقريرية في مقابل الأنظمة الإنتاجية والتي منها قواعد البيانات، حتى أن الهيكليات*Architecture* المستخدمة صممت على أساس استخراج المعرفة للمساعدة في اتخاذ القرار. *[6]*

برزت الحاجة الملحة إلى مستودعات البيانات في الشركات الكبرى حيث أن كل قسم من هذه الشركات يدير قواعد بيانات مستقلة خاصة به (تسويق، مالي...) والتي تتضمن كثيراً من البيانات المشتركة وتتطلب حاجة الشركة إيجاد كامل البيانات المتعلقة بموضوع محدد كالزبائن مثلاً من خلال البحث في قاعدة واحدة بدلاً من البحث في القواعد المختلفة، لذا ظهرت فكرة تجميع البيانات في قاعدة واحدة تسمى مخزن البيانات مع الاحتفاظ بالقواعد الخاصة، على أن يصار إلى تغذية المخازن دورياً في حال حصل تعديل أو تبديل في القواعد.

تختلف طريقة معالجة البيانات بين قواعد البيانات ومخازن البيانات، فالأولى تستعمل المعالجة الآنية *On (OLTP) Line Transaction Processing* حيث أنه يمكن تعديل، زيادة أو البحث عن بيانات بسرعة مع الاحتفاظ بتكامل قاعدة البيانات *Database Integrity* أما مخازن البيانات فتستخدم طريقة تحليل البيانات *On Line Analysis processing (OLAP)* من خلال معالجة كميات ضخمة وتحليلها وإخراجها بشكل معرفة تسمح باستنتاجات وتحديد اتجاهات جديدة بدون أية إمكانية لتعديل البيانات، ويمكن استخدام أسئلة غير متوقعة تأخذ أحياناً وقتا طويلاً لإصدار النتائج.

لا يمكن تنفيذ وظائف *OLAP* بشكل جاهز على أنظمة *OLTP* ويعود السبب في ذلك إلى أن أنظمة *OLTP* تحتفظ فقط بلقطة لحظية لأحدث المعلومات المتوفرة بشكل مباشر فيما تتطلب وظائف *OLAP* تاريخ التعاملات السابقة عبر فترات زمنية طويلة. وتمتاز أنظمة *OLTP* بالتحديث المستمر وتحتوي في بعض الأحيان على بيانات ناقصة أو مغلوطة بينما تتطلب أنظمة *OLAP* بيانات ساكنة و كاملة و خالية من الأخطاء، كما هو مبين في الجدول (1). *[7]*

الجدول (1) مقارنة بين نظامي OLAP،OLTP

	OLTP	OLAP
User	Clerck,IT Proffesional	Knowledge Worker
Function	Day to Day Operrations	Desision Support
Db Design	Application Oriented(E-R Based)	Subject Oriented
Data	Current,Isolated	Historical,Consolidated
View	Detailed,Flat Relational	Summarized,Multidimentional
Unit of Work	Short,Simple Transactions	Complex Query
Access	Read/Write	Read Mostly
Record Accesses	Tens	Millions
Users	Thousands	Hundreds
Db Size	100 MB-GB	100 GB-TB
Metric	Trans.throughput	Query throuput,response

القضايا التي يجب أخذها بعين الاعتبارعند إنشاء مستودع البيانات

1. استخلاص البيانات من عدة مصادر قد تكون غير متجانسة.
2. تهيئة البيانات لضمان تلاؤمها داخل مستودع البيانات.
3. تنظيف البيانات لضمان شرعيتها (إلغاء التكرار – التصحيح الشكلي – معالجة البيانات الناقصة...) .
4. مراقبة وضبط حجم مستودع البيانات بعد تحميله بالبيانات.
5. الفترات الفاصلة بين التحديثات المتتالية لمستودع البيانات.
6. الزمن اللازم لبناء مستودع البيانات وجدواه الاقتصادية.

7. هل يجب أن يكون مستودع البيانات مركزيا أو موزعا.
8. الاهتمام بالبعد الزمني.
9. حفظ كامل البيانات من الأقدم حتى الأحدث.

الوظائف داخل مستودع البيانات

Roll_up :تلخيص البيانات بشكل متزايد (أسبوعيا ،سنويا،...) أو يمكن التعبير عنها في قواعد البيانات متعددة الأبعاد بأنها الانتقال عبر تسلسل الأبعاد من الأدنى (الأبناء) إلى الأعلى (الآباء).

Drill_down :بعكس Roll_up تتم هنا زيادة مستوى تفصيل البيانات.

Pivot(Rotation) : الانتقال من بعد إلى بعد آخر (تغيير وجهات النظر إلى البيانات).

Sorting :ترتيب البيانات ضمن مستودع البيانات .

Slice :أخذ شريحة من مستودع البيانات ويتم ذلك بناء على قيمة مفردة لأحد الأبعاد كما هو مبين في الشكل (1 و2 و 3).

Dice : اختيار مكعب جزئي من مكعب اكبر .[8]

:

الشكل (1) ***Sales Slice of the Database***

النتائج والمناقشة

نمذجة مستودع البيانات المطلوب

سوف يتم استخدام النمذجة البعدية لقواعد البيانات في المستودع المطلوب Multidimensional Database Modeling، إذ سوف يتم تنظيم عملية الصيانة باستخدام أبعاد مختلفة مثل معلومات نوع العطل، زمن العطل، سبب العطل، ويتم تمثيل النموذج البعدي كمكعب افتراضي Hypercube بحيث يمكن للمستخدمين تقسيم و تقليب البيانات بالشكل الذي يختارونه.

تمثل كل كتلة في المكعب متعدد الطبقات قيمة من القيم. و يمكن إيجاد القيم عن طريق أي بعد من الأبعاد أو عبر مجموعة من الأبعاد مثل تجميع المبيعات عبر البعد الزمني أو غيره.

تخزين النموذج متعدد الأبعاد يشمل نوعين من الجداول:

جداول الأبعادDimension table: وحقوله تسمى سمات البعد Dimension Attributes.

جدول الحقيقة Fact Table:ويتضمن مقاييس أداء العمل كما يحتوي على مؤشرات لجداول الأبعاد.

تحقيق النموذج البعدي

يمكن تحقيق النموذج البعدي لقواعد البيانات بعدة طرق. أكثرها استخداما:

- **ROLAP** اختصار ل Relational OLAP يستطيع المستخدمين أن يروا بياناتهم منظمة في مكعب و لكنها في الحقيقة تكون مخزنة في قاعدة بيانات علاقاتية ومتحكم بها باستخدام برنامج إدارة قواعد بيانات علاقاتية RDBMS مثل ORACLE، وتخزن البيانات بنوع من التفصيل و يكون زمن الاستجابة بطئ نسبيا مقارنة مع تقنيات اكثر تقدما مثل
- **MOLAP** اختصار ل Multidimensional OLAP المستخدمون يرون بياناتهم مخزنة في مكعب ذو أبعاد و البيانات تخزن فعليا في قاعدة بيانات متعددة الأبعاد أي أن هيئة التخزين تكون مبنية على أساس مصفوفاتArray based storage structure و يمكن الوصول إلى عناصر المصفوفات مباشرة ويتحكم بها باستخدام برنامج إدارة قواعد بيانات متعددة الأبعاد MDBMS مثل ORACLE EXPRESS SERVER يكون عامل الزمن في هذا النموذج حاسما و زمن الاستجابة سريعا. الجدير بالذكر انه عند استخدام MOLAP تظهر مشاكل في التعامل مع قواعد البيانات الكبيرة جدا وذات عدد الأبعاد الكبير، فبزيادة عدد الأبعاد يزداد عدد الخلايا الفارغة في مكعب البيانات والتي تحتاج إلى متطلبات فهرسة و تخزين بدون فائدة. يمكن تخفيف أثر هذه المشكلة باستخدام تقنيات ضغط معينة، على الرغم من أن ذلك سوف يؤدي إلى تدمير البنية والفهرسة الطبيعية من أجل هذا النوع من التخزين.
- **HOLAP(Hybrid OLAP)** وهي مجموعة من الأدوات التي تجمع بين MOLAP و ROLAP.
- **WOLAP(Web-Enabled OLAP)**.

سوف يتم استخدام نموذج **ROLAP** كون قاعدة البيانات الأساسية لنظام الصيانة هي قاعدة بيانات علائقية.

العمليات الأساسية على الأبعاد

- التجميع Aggression(Roll up):

بالنسبة لبعد واحد: مجموع أزمان الصيانة حسب بعد نوع الالية (يتم حساب مجموع أزمان الصيانة لكل الية في بعد نوع الاليات بغض النظر عن الأبعاد الأخرى)

بالنسبة لعدة أبعاد: مجموع ازمنة الصيانة حسب الاليات و أرباع السنة (سيكون هناك أربع مجاميع لكل مدينة لكل ربع من السنة بغض النظر عن الأبعاد الأخرى و التي قد تكون بعد نوع العطل مثلا)

- الاختيار Selection (Slice): مثلا اختيار عمليات الصيانة بالنسبة لبعد نوع الية معين ويكون بناء على قيمة مفردة لهذا البعد(الية معينة).
- الانتقال إلى بيانات أكثر تفصيلا(Drill down).
- عمليات مرئية مثل التدوير(Pivot).

تخزين النموذج متعدد الأبعاد

كما ذكر سابقا فان تخزين النموذج متعدد الأبعاد يشمل نوعين من الجداول:

جداول الأبعاد:وحقوله تسمى سمات البعد Dimension Attributes.

جدول الحقيقة Fact Table:ويتضمن مقاييس أداء العمل كما يحتوي على مؤشرات لجداول الأبعاد.

وتوجد بنيتين شائعتين لتعريف هذه الجداول:

Star Schema : يعد من أبسط أشكال تحقيق مستودعات البيانات ويدعى بهذا الاسم لشبه شكل المخطط بالنجمة، يتكون مركز المخطط من جدول حقيقة واحد أو أكثر و يكون هناك جدول وحيد لكل بعد كما هو مبين في الشكل (2)، يعيبه الحاجة لمساحة كبيرة للتخزين، و مع ذلك فان شركة اوراكل تنصح باستخدام هذا النموذج نظرا لأنه يوفر الأداء الأحسن فهو يحافظ على الاستعلامات بسيطة و لا يحتاج لأكثر من عملية جداء واحدة بين جدول الحقيقة و جداول الأبعاد.

يمتلك جدول الحقيقة نوعين من الأعمدة. الأول يحتوي حقائق رقمية تدعى القياسات Measurement والنوع الثاني هو مفاتيح خارجية إلى جداول الأبعاد.

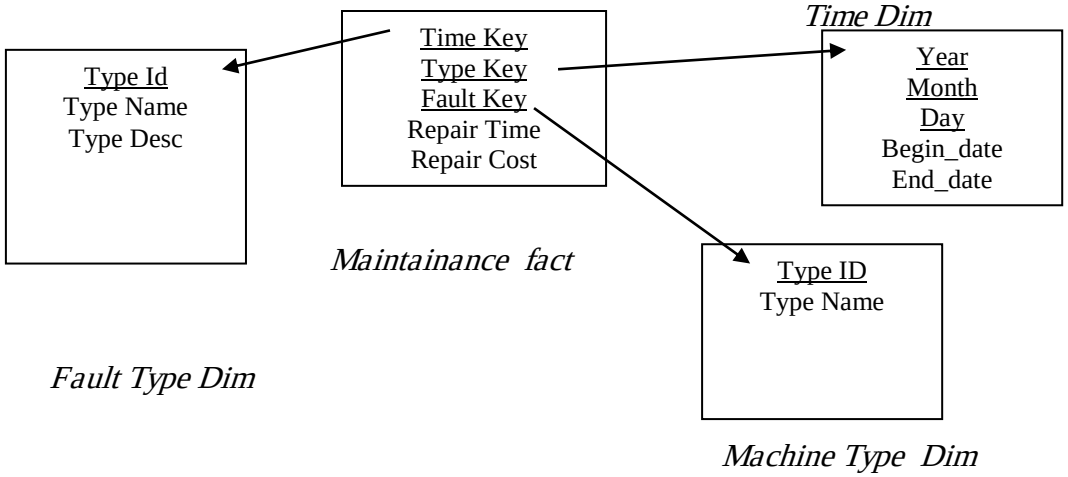
يمكن أن يحتوي جدول الحقيقة قيم مجمعة Summarized table وعندها يدعى بجدول الخلاصة Summery table وعادة تكون البيانات مجمعة على نفس المستوى في جدول الحقيقة.

المفتاح الرئيسي لجدول الحقيقة مكون من مجموع المفاتيح الخارجية لجداول الأبعاد.

البيانات الخاصة بجداول الأبعاد تجمع على أصغر مستوى ثم تتم عملية تجميع للقيم في مستويات أعلى لتصبح أكثر فائدة في عمليات التحليل.

مزايا استخدام المخطط النجمي:

- يسقط متطلبات العمل بشكل مباشر على التصميم.
- يوفر الأداء الأفضل.
- الاستعلامات الخاصة به سهلة نسبيا.
- بوجد العديد من الأدوات التي تدعم هذا المخطط.
- يحتاج لعمليات صيانة اقل.
- يمكن استخدامه في مستودعات البيانات الصغيرة و الكبيرة على حد سواء.

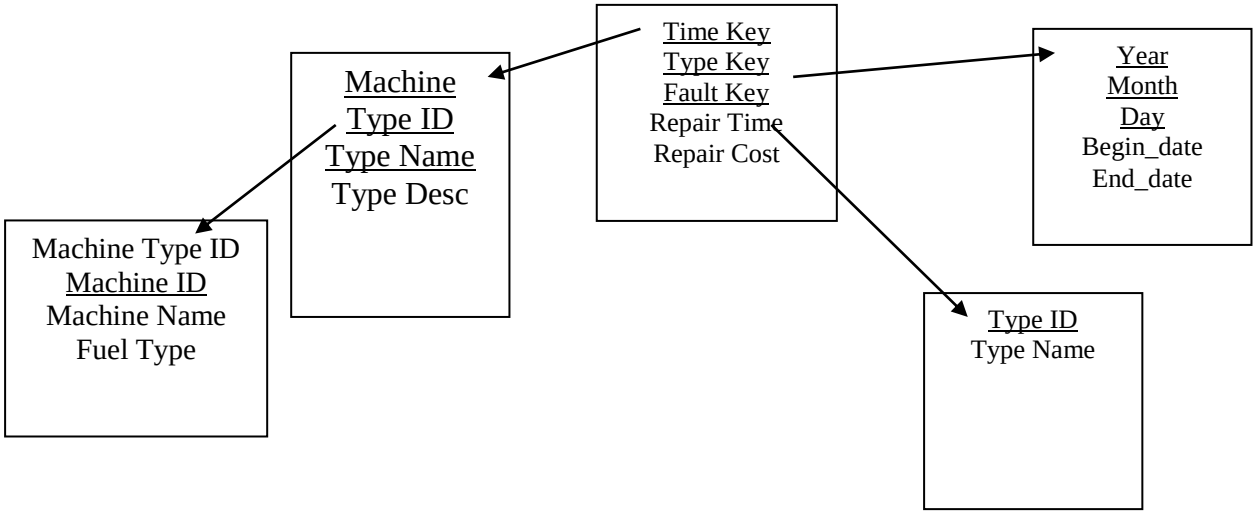


الشكل (2) الشكل التالي مخطط نجمي لمستودع البيانات

Snowflake Schema :تعديل ل Star Schema وذلك بتطبيق قوانين التطبيع على الأبعاد. مثلا الجدول الخاص ببعد الاليات يمكن أن يقيّس فيصبح جدول نوع الالية و جدول الالية كما هو مبين في الشكل (3)، وبينما توفر هذه العملية من المساحة اللازمة للتخزين فانه يعيبها تضييعها لطاقة الجهاز المضيف في عمليات الربط بالإضافة إلى أن الاستعلامات تصبح أكثر تعقيدا.

ميزات استخدام المخطط Snowflake :

- يوفر المساحة اللازمة للتخزين.
- يبين هيكلية و تسلسلية المعلومات.



الشكل (3) مثال عن المخططات من النوع Snowflake المستخدمة في مستودع البيانات المصمم

مراحل معالجة البيانات في مستودع البيانات

تخليص البيانات Data Extraction

تخلّص البيانات من مصدر من اجل استخدامها لاحقا في بيئة مستودع البيانات، وهي أول خطوة في عملية ETL، بعد تخليص البيانات يتم نقلها و تحميلها في مستودعات البيانات.

يعد تصميم و إنشاء عملية الاستخلاص من أكثر المراحل استهلاكا للوقت في العمليات الخاصة بمستودع البيانات. بما أن النظام المصدر قد يكون معقدا جدا وموثق بشكل غير جيد فان تحديد البيانات التي يجب استخلاصها قد يكون صعبا، وقد تكمن المشكلة في أن النظام المصدر غير مصمم بحيث يكون قابلا للتعديل ليلائم مستودع البيانات.

هناك اعتبارات أساسية في هذا المجال يجب ملاحظتها:

ما هي الطريقة التي يجب استخدامها لاستخلاص البيانات؟

كيف يمكن تزويد وتحضير البيانات لعملية المعالجة اللاحقة؟

تم استخدام الطرق التالية لاستخلاص البيانات:

- الاستخلاص الكامل.
- الاستخلاص التزايدي.

الاستخلاص الكامل:يتم استخلاص البيانات بشكل كامل من مستودع البيانات. وهذا الاستخلاص سوف يعكس البيانات الموجودة حاليا في المصدر، وبذلك لا توجد حاجة لمعرفة البيانات الجديدة التي أضيفت إلى مصدر البيانات في وقت سابق. قد يكون الاستخلاص الكامل للبيانات عن طريق استيراد ملف ما أو عن طريق تعليمة sql تمسح الجدول كاملا.

الاستخلاص المتزايد: تستخلص البيانات التي تغيرت فقط منذ حدث محدد زمنيا، مثلا قد يكون هذا الحدث هو آخر عملية استخلاص. يجب أن يكون الحدث الزمني قادرا و كافيا لتحديد و حصر جميع التغيرات التي أضيفت إلى البيانات التشغيلية. قد يكون المحدد عبارة عن عمود يحتوي زمن إدخال البيانات ويعمل في هذه الحالة كبصمة زمنية.

التحميل و التحويل Loading and Transformation

بما أن مستودع البيانات يتعامل مع مصادر مختلفة للبيانات (اليات – مشتريات – قوى بشرية–..) و لكل من هذه المصادر بنيتها الخاصة بها، فان عملية تحويل لتنسيق و شكل البيانات يجب أن تتم قبل تخزينها في المستودع ويوضح ذلك المثال التالي:

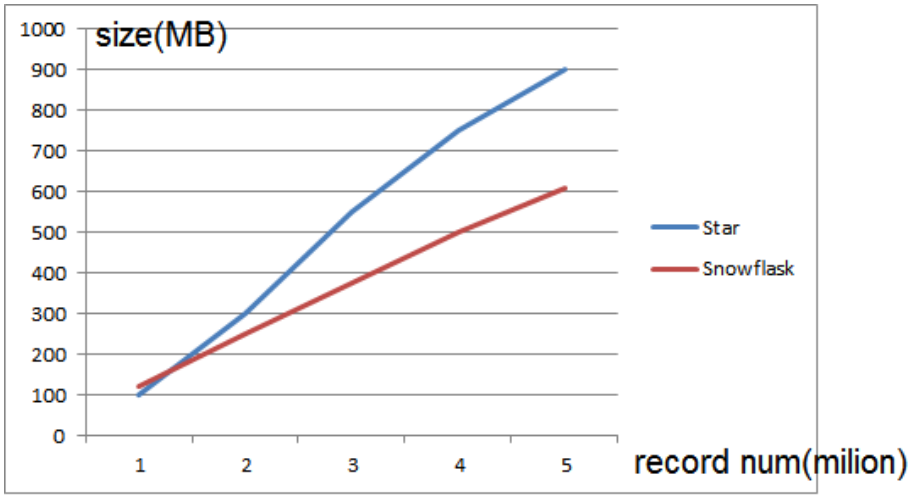
مثلا في النظام الفرعي الذي يخزن أسعار قطع التبديل، لكل منتج سعر أدنى و سعر أعلى و بنية schema جدول الأسعار كالتالي:

رقم المنتج	السعر الأعلى	السعر الأدنى
100	1500	1000

في مستودع البيانات يتم اعتماد البنية التالية:

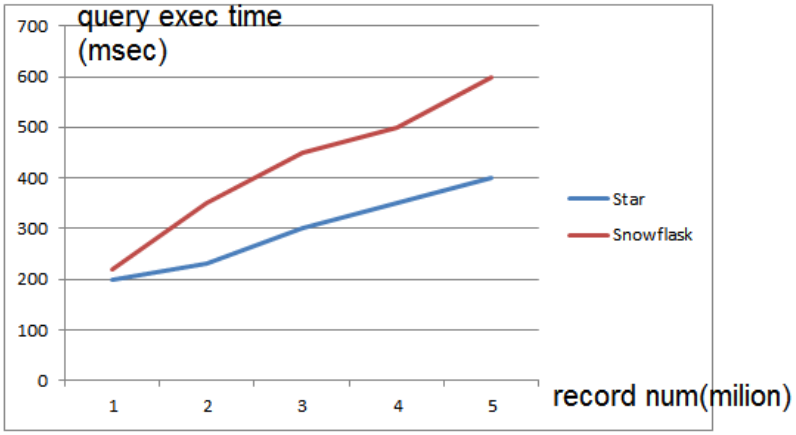
رقم المنتج	نوع السعر	السعر
100	الأعلى	1500
100	الأدنى	1000

تم حساب المساحة التخزينية لنموذجي Snowflake و النموذج النجمي في مستودع البيانات المصمم مقارنة مع عدد السجلات الإجمالي في قاعدة البيانات، كما هو مبين في الشكل (4).



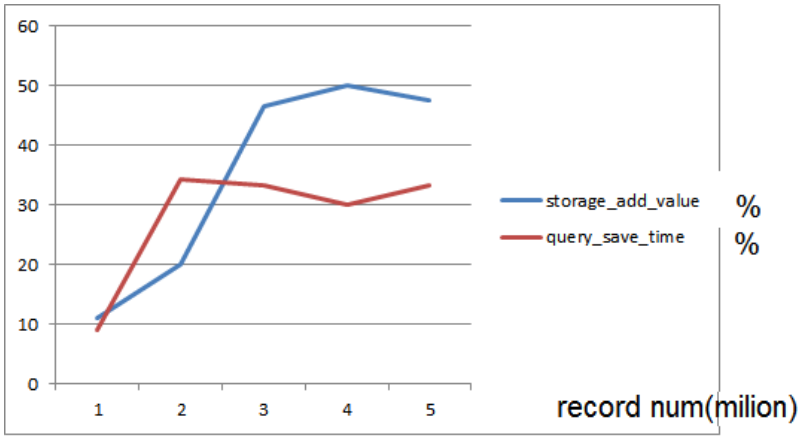
الشكل (4) الحجم التخزيني لقاعدة البيانات وفق نموذجي التخزين في مستودع البيانات المصمم

الشكل (5) يبين مقارنة بين عدد السجلات ووسطي زمن تنفيذ الاستعلام لكل من نموذجي التخزين.



الشكل (5) وسطي زمن تنفيذ الاستعلام لكل من نموذجي التخزين

تم مقارنة الوفر في زمن تنفيذ الاستعلام لدى استعمال كل من النموذجين مقارنة مع النسبة المئوية لزيادة حجم التخزين في قاعدة البيانات نتيجة استخدام النموذج النجمي الذي لاتتم فيه أي عملية تقييس كما هو مبين في الشكل (6).



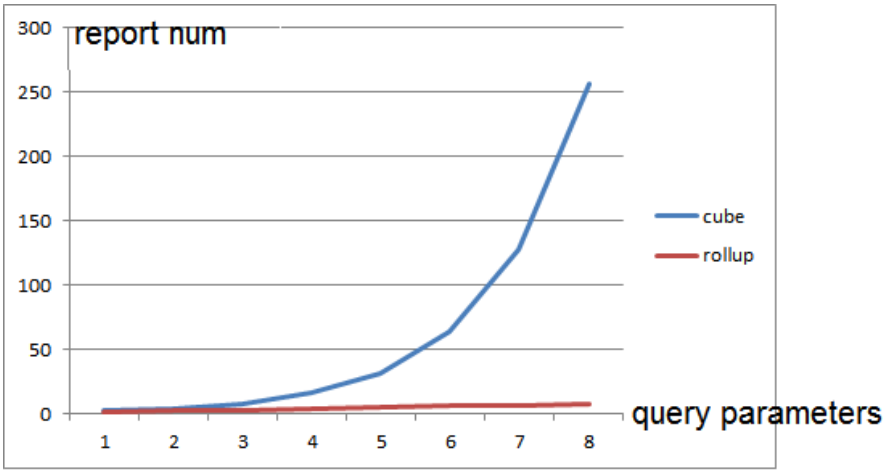
الشكل (6) النسبة المئوية من زيادة حجم قاعدة البيانات مقارنة مع زيادة زمن تنفيذ الاستعلام

تعد عمليات التجميع من العمليات الأساسية في مستودع البيانات، ولتحسين أداء عمليات التجميع زودتنا اوراكل بمجموعة من العبارات الخاصة التي تعد امتدادا لعبارة Group By الخاصة بتجميع البيانات وفق مناظير معينة.

- Cube and Rollup
- Grouping Functions

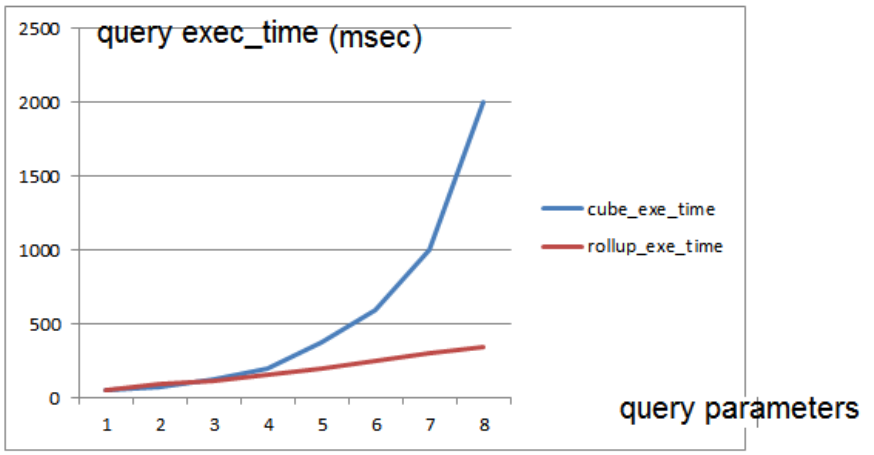
إن الامتدادات السابقة تجعل من عملية تجميع البيانات أسهل وأسرع، تقوم عبارة Rollup بحساب توابع التجميع(SUM, COUNT, MAX, MIN, and AVG) بشكل تزايدي وفق تسلسلية معينة تبدأ من المستوى الأكثر تفصيلا إلى المستوى الأقل تفصيلا، تسمح عبارة Cube بحساب المجاميع و فق جميع المناظير المحتملة للبيانات.

تم مقارنة زمن تنفيذ الاستعلامات لكل من الاستعلامات باستخدام توابع التجميع التقليدية وباستخدام توابع التجميع الخاصة بمستودع البيانات ونبين في الشكل (7) عدد التقارير التي يمكن توليدها باستخدام كل من عبارتي cube وعبارة rollup مقارنة مع عدد بارامترات التجميع.



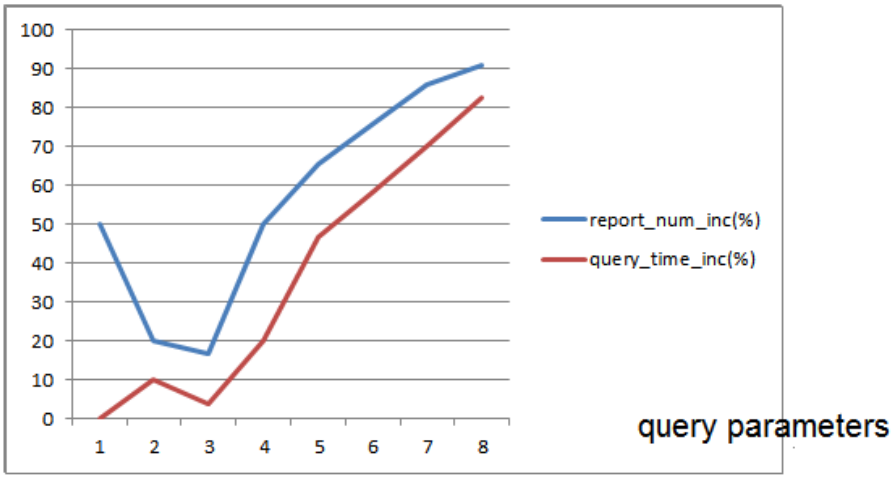
الشكل (7) عدد التقارير التي تولدها كل من عبارة rollup وعبارة cube مقارنة مع عدد البارامترات

الشكل(8) يبين مقارنة عدد بارامترات الاستعلام مع زمن تنفيذ الاستعلام لكل من عبارتي rollup و cube.



الشكل (8) مع زمن تنفيذ الاستعلام لكل من عبارتي rollup و cube

تم مقارنة زيادة عدد التقارير التي يقدمها استعلام cube مقارنة مع الزيادة في زمن الاستعلام وكانت النتيجة كما هو مبين في الشكل (9).



الشكل (9) النسبة المئوية لزيادة زمن الاستعلام مقارنة مع زيادة عدد التقارير التي يقدمها الاستعلام

نلاحظ من الشكل 9 وجود زيادة ملحوظة لزمن الاستعلام باستخدام عبارة ***cube by*** group مقارنة مع عدد البارامترات ومقارنة مع زمن تنفيذ الاستعلام باستخدام عبارة ***rollup*** وعليه نرى استخدام عبارة ***rollup*** بدلا من عبارة cube طالما يمكن حصر التقارير المطلوبة.

الاستنتاجات والتوصيات

الاستنتاجات

- استخدام النموذج النجمي سوف يؤدي الى زيادة جدم قاعدة البيانات بشكل ملحوظ مقارنة مع النموذج ***snowflake***.
- مع ازدياد حجم قاعدة البيانات نتيجة استخدام النموذج ***snowflake*** يحصل زيادة في تأخير زمن الاستعلام بشكل ملحوظ، إلا أنه بعد حجم معين تستقر نسبة الوفر الحاصل في زمن تنفيذ الاستعلام مقارنة مع ازدياد الحجم وبالتالي ينصح باستخدام النموذج ***snowflake***.
- عبارة ***cube*** مكلفة زمنيا مقارنة مع عبارة ***rollup***.

التوصيات

- استخدام مستودعات البيانات في نظم دعم القرار.
- استخدام النموذج ***snowflake*** في تصميم مستودع البيانات.
- استخدام عبارة ***rollup*** بدلا من عبارة ***cube*** طالما يمكن حصر التقارير المطلوبة في مستودع البيانات.

المراجع

[1] Jitendra Kumar, V.K. Soni,Geeta Agnihotri , MAINTENANCE PERFORMANCE METRICS FOR MANUFACTURING INDUSTRY, IJRET: International Journal of Research in Engineering and Technology, ISSN: 2319-1163

[2] YiHua Philip Sheng, INFORMATION TECHNOLOGY , INVESTMENT AND FIRM PERFORMANCE: A PERSPECTIVE OF DATA QUALITY , Proceedings of the Seventh International Conference on Information Quality (ICIQ-02)

[3] G. M. Faruk Ahmed, Md Shoriful Islam, Molla Md Rezaul Karim, Comparison Between Inmon and Kilball Methodology for the Purpose of Designing, Constructing and Testing of a Commercial BIDW Project, International Journal of Computer Graphics Vol.8, No.1(2017), pp.11-20

[4] Business Rules Definition for Decision Support System Using Matrix Grammar Acta Informatica Pragensia, 2016, 5(1):72–81 DOI: 10.18267/j.aip.86

[5] Amandeep Kour,Data Warehousing, Data Mining, OLAP and OLTP Technologies Are Indispensable Elements to Support Decision-Making Process in Industrial World , International Journal of Scientific and Research Publications, Volume 5, Issue 5, May 2015, ISSN 2250-3153

[6] Luna Rajbhandari, Jason Schober, Data Warehouse and Analytics Discussion Paper, Director of Business Intelligence Solutions , 3/1/2013

[7] A. K. Hamoud1,Dr. Tale, b A.S. Obaid, A.K. Ha Moud et al , Using OLAP with Diseases Registry Warehouse for Clinical Decision Support ,International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4, April-2014, pg. 39-49

[8] Dr. Katherine Jones , An Introduction to Data Warehousing: What Are the Implications for the Network?, INTERNATIONAL JOURNAL OF NETWORK MANAGEMENT, VOL. 8, 42–56 (2018)Int. J. Network Mgmt., vol. 8, 42–56 (2018)