Evaluation of Some Properties of Distributed Databases using Oracle

Alaa Faisal Saeed Al-Mukhtar
Computer and Informatics Department
General Directorate of North Electricity Distribution
Ministry of Electricity

Received on: 25/03/2012 **Accepted on: 28/06/2012**

ABSTRACT

In this study, query processing efficiency of the distributed database in the relational database Oracle9i was examined and compared. Our test data is based on a (Customer Type Definition), which is designed to represent customer information of bank. We generated 8 different sizes of test customers: 300, 600, 900, 2100,4500, 45000, 1360 Records.

DML language (SQL) as tools for any comparison can be achieved by the statistical performance parameter (Elapsed time, CPU time, Logical read , Physical read, Logical write, Physical write, Elapsed time, UGA) .statistical analysis using ANOVA TABLE is used to judge between comparisons.

All the queries were run on two computers for different sizes of vertical and horizontal fragments. Changes in performance metrics for each query were written in table after experiments. Finally, we presented the results, and discussed where enhancements are required.

Statistical analysis using ANOVA TABLE were used to distinguish between this comparison, it had been found that SELECT, UPDATE and DELETE have the most influence on the performance parameter respectively.

Keywords: distributed database , distributed management system , data fragmentation CPU time, Logical Read , Physical Read, Physical Write , Elapsed time , UGA.

تقييم بعض خواص قواعد البيانات الموزعة باستخدام اوراكل

آلاء فيصل سعيد المختار
قسم الحاسبة والمعلوماتية
المديرية العامة لتوزيع كهرباء الشمال
وزارة الكهرباء

تاريخ استلام البحث: 2012/03/25 **تاريخ قبول البحث: 2012/06/28**

الملخص

في هذا البحث تم فحص ومقارنة كفاءة تنفيذ جمل الاستفسارات لقواعد البيانات الموزعة في النظام العلائقي بلغة Oracle9i . البيانات المعتمدة تمثل المعلومات الأساسية لمشتركي الكهرباء وتم إدخالها برمجيا وبثمانية مستويات وبإحجام 300 , 600, 900, 2100, 4500, 45000, 136000 قيد . ان تضبيط تنفيذ لغة معالجة البيانات متمثلة بلغة الاستفسارات المهيكلة يمكن تقييمها إحصائيا بوصفها من خلال المعلمات التالية(وقت تنفيذ وحدة المعالج المركزي , القراءة المنطقية والقراءة الفيزيائية للكتل البيانية , الكتابة المنطقية والكتابة الفيزيائية لكتل البيانات , الوقت المستغرق في التنفيذ) . تم تجزئة جدول البيانات أفقيا وعموديا باستخدام قواعد التجزئة

واستخدام طريقة التجزئة المركزية . حيث لوحظ تأثير معنوي ملحوظ لعملية التجزئة بنوعيها ولنوع المعالجة وكذلك حجم البيانات على أداء عملية توزيع البيانات متمثلة بالصفات المعبرة عن أمثلية هذا الأداء.
بواسطة التحليل الإحصائي باستخدام جدول التباين مكننا من الحكم على عمليات المقارنة . وقد وجد أن هناك تأثيرا معنويا للتداخل ما بين ايعازات المعالجة ونوع التجزئة أي إن لكل منها تأثيرا على الآخر بضمان الفرق الإحصائي على جميع الصفات المقاسة .كما وجد أيضا تأثير معنوي للتداخل ما بين ايعازات المعالجة وحجم البيانات ومابين نوع التجزئة وحجم البيانات

الكلمات المفتاحية: قواعد البيانات الموزعة، نظام إدارة قواعد البيانات الموزعة، تجزئة البيانات، وقت المعالجة، القراءة المنطقية، القراءة الفيزيائية، الكتلة الفيزيائية، الوقت المستغرق بالتنفيذ، مساحة المستخدم العامة.

المقدمة

قواعد البيانات الموزعة (Distributed database) هي مجاميع من قواعد البيانات المترابطة منطقيا والموزعة على شبكة الحاسبات وتدار باستخدام نظام إدارة قواعد البيانات الموزعة (Distributed database Management system) الذي يوفر شفافية التوزيع بالنسبة للمستخدم[8]. تتكون قواعد البيانات الموزعة من مجموعة من المواقع التي تتصل ببعضها من خلال شبكة الاتصالات وكل موقع هو نظام لإدارة قاعدة البيانات ويتمكن المستخدم في أي موقع من الوصول للبيانات على الشبكة كما لو أن البيانات خزنت في موقع المستخدم نفسه [2] .

أصبح استخدام قواعد البيانات الموزعة أكثر شيوعا في الوقت الحالي لأسباب متعددة تتضمن المشاركة في البيانات، الوثوقية، الكفاءة والسرعة في عمليات التنفيذ والمعالجة. هناك مجموعة من العوامل التي تؤثر على كفاءة جمل الاستفسارات في تطبيقات أنظمة قواعد البيانات الموزعة وهي عدد مرات الوصول إلى القرص وكلفة نقل البيانات أو التأخير في الاتصال [1]. كفاءة تطبيق البيانات العلائقية الموزعة تعتمد على التصميم وهناك طريقتان يمكن استخدامها في تحسين كفاءة التطبيق من خلال التصميم وهما مخطط التجزئة،ومخطط التقسيم .
مخطط التجزئة ويمثل العلاقة ما بين قواعد البيانات الموزعة وبين أجزاء قاعدة البيانات الحالية ويمثل بعلاقة (ONE TO MANY) التي تستخدم لتجزئة كل جدول عام إلى أجزاء وهي قطع منطقية من الجداول العامة والتي توزع فيزيائيا في مواقع متعددة من الشبكة ، الهدف الأساسي لتجزئة العلاقات هو لتقليل كمية البيانات غير المتصلة بالموضوع من خلال الإجراءات transaction التي تصل إلى العلاقة. اما مخطط التقسيم فينضم العلاقة بين الأجزاء والمواقع الخاصة بخصائص قواعد البيانات الموزعة، الهدف الرئيسي من تقسيم و توزيع الأجزاء هو لتقليل كلفة تناقل البيانات عند معالجة الأجزاء أو التطبيقات [12].

وجد [6] معالجة ملايين القيود عبر القواعد الموزعة يمكن ان تزداد كفاءة ادائها بزيادة عدد النقاط باخذ حجم البيانات وازمنة التاخير بوصفها مقاييس للحكم على كفاءة الاداء. وقام [14] ببناء نماذج لمحاكاة اداء قواعد البيانات الموزعة وذلك لحل مشاكل تناقل هذه البيانات عبر نقاط التوزيع.
في لغة أوراكل تحول قواعد البيانات الموزعة في المواقع المحلية الى مجموعة من الاستفسارات البعيدة وترسل إلى المواقع البعيدة لتنفيذها ، تنفذ المواقع البعيدة الاستفسارات وتعيد النتائج إلى المواقع المحلية ، تنجز المواقع المحلية أي عملية في موقعها وترسل الناتج إلى المستخدم أو التطبيق [7].

توفر لغة أوراكل استفسارات موزعة وشفافة للوصول إلى البيانات من قواعد بيانات مختلفة. وتوفر أيضا مجموعة أخرى من خصائص التوزيع مثل الشفافية Transparentوشفافية الأجزاء الموزعة[9].

يهدف البحث إلى قياس تأثير أنواع الايعازات وعمليات المعالجة على الحجوم المختلفة للبيانات الموزعة بنوعي التجزئة الأفقي والعمودي وإيجاد تأثير كل من هذه العوامل على الآخر باستخدام المقاييس الإحصائية.

1- تقنية تصميم قواعد البيانات الموزعة

1-1 تجزئة البيانات (Data Fragmentation)

تجزئة البيانات هي طريقة للتصميم واتخاذ قرار لأي من ألاساليب ستستخدم [6]. الجدول (r) يجزء إلى مجموعة من ألا جزاء غير المترابطة (r1,r2,r3,…rn) والتي تحتوي على معلومات كافية من البيانات وبربطها مع بعضها يتم إعادة بناء الجدول الأساسي (r)[6].النظام يدعم التجزئة إذا كانت العلاقة تقسم إلى أجزاء صغيرة (fragments) لإغراض الخزن الفيزيائي أذ تخزن البيانات في المواقع الكثيرة الاستخدام . وتوفر هذه التجزئة كفاءة الأداء [6].

1-1-1 التجزئة الأفقية (Horizontal Fragmentation)

التجزئة الأفقية للنظام ألعلائقي هي مجموعة من القيود التي تنتمي إلى هذه العلاقة [7]. التجزئة الأفقية تجزء الجدول "أفقيا "من خلال تجميع القيود(tuples) التي لها معنى منطقي محدد لتكوين حزمة مشتركة منها على شكل أجزاء وبعدها توزع هذه الأجزاء على مواقع مختلفة في النظام الموزع ، القيود التي تعود إلى التجزئة الأفقية تحدد من خلال الشروط على واحدة أو أكثر من الصفات (attribute) وعند إجراء عملية التجميع (union) على الأجزاء الكلية في التجزئة الأفقية تعود العلاقة الأصلية [12]. تجزئة الصفوف للجدول الرئيسي إلى مجموعة من الأجزاء , الجزء r1 يتم الحصول عليه بتنفيذ إيعاز الاختيار (select) على الجدول الرئيسي باستخدام التعبير pi ويعاد بناء العلاقة r بإجراء عملية التجميع على كل الأجزاء [6].

1-1-2 التجزئة العمودية Vertical Fragmentation

التجزئة العمودية للنظام العلائقي تحافظ على الصفات الأساسية للجدول الرئيسي .أذ تقسم الجدول إلى مجموعة من الأعمدة ومن الضروري أن يضمن المفتاح الأساسي (row_id) في كل جزء من الأجزاء العمودية وهذا يمكن من إعادة بناء الجدول الرئيسي باستخدام عملية الربط (join) [5].

1-2 قواعد التجزئة (Rule of Fragmentations)

1- تصاغ الأجزاء بأجراء عملية الاختيار التي تسيطر على إجراءات قواعد البيانات,وتحدد قيم بيانات الصفات باستخدام عمليات الربط (AND) وعملية فك الربط (OR) التي تصيغ جمل الاختيار للقيود التي تحتوي على نفس القيم في الأجزاء.
2- الأجزاء يجب أن تكون منفصلة وعملية التجميع تعيد الجدول الأساسي ، التداخل في الأجزاء يشكل صعوبة في عمليات التمثيل والتحليل.
3- تمثل الأجزاء الكبيرة بالجدول بأكمله و تمثل الأجزاء الصغيرة بالقيد المفرد ، و الأجزاء يجب أن تصمم حتى تدعم التوازن مابين هذين الحجمين [15].

3-1 تصميم التجزئة وأنواعها (Distributed Design & type)

التجزئة مطلوبة لأنها تضع البيانات في مكان قريب لعمل المستخدم . الطاقة الكامنة في التجزئة هي لتقليل كلفة إرسال ونقل البيانات وتقليل حجم العلاقات المتصلة في استفسارات المستخدم . يمكن أن تكرر الاجزاء ويفضل استخدام هذه الطريقة عندما يصل التطبيق الى البيانات نفسها التي تنفذ في مجموعة من المواقع وفي هذه الحالة فان الحصول على اقل كلفة يكون من خلال تكرار البيانات بمجموعة من المواقع بدلا من استمرار نقلها بين هذه المواقع[7].

قواعد البيانات الموزعة تعرف المحددات وتحت أي واحدة من طرائق توزيع البيانات سيتم التنفيذ. وهذا يحدد من خلال هيكلية النظام وبرمجيات قواعد البيانات المتوفرة في الشبكة هناك ثلاث طرائق لتنفيذ التجزئة وهي [7]:

1- المركزية (centralized) تمتاز الطريقة المركزية لتوزيع قواعد البيانات بسهولتها أذ توضع كل البيانات في موقع واحد. حجم البيانات يحدد من خلال قابلية الخزن في الموقع المركزي إضافة إلى إن قواعد البيانات تكون غير متاحة من قبل أي جهة بعيدة ، يفشل نظام قاعدة البيانات كليا عند حدوث فشل في الموقع المركزي.

2- التجزئة (partitioned) في هذه الطريقة تجزء قاعدة البيانات إلى أجزاء غير مترابطة كل جزء يوضع في جهة (site) معينة ، هذه الطريقة مناسبة جدا عندما تكون كمية الخزن المتوفر في الموقع الثانوي محددة مقارنة بحجم قاعدة البيانات . أن الوثوقية من قاعدة البيانات المركزية غير كافية وكفاءة التشغيل تكتسب من خلال استغلال المرجعية الموقعية في الوصول لقاعدة البيانات.

3- التكرار (replicated) طريقة التكرار في قواعد البيانات الموزعة توزع نسخة كاملة من قاعدة البيانات في كل موقع على الشبكة وتكون هذه الطريقة مناسبة عندما تكون الوثوقية مطلوبة وقواعد البيانات صغيرة وتحديث قواعد البيانات غير مسيطر عليه[15] .

وتم اعتماد طريقة التجزئة في تصميم قواعد البيانات الموزعة في البحث ودراسة النتائج الخاصة بها.

المواد وطرائق العمل Material & Method

نفذت التجارب باستخدام حاسوبين متماثلين بنظام تشغيل Windows XP وأديرت قاعدة البيانات بواسطة Oracle9i تعمل بوصفها مخدماً Server في كلا الحاسوبين. تم تهيئة الاتصال بين الحاسوبين باستخدام أداة (oracle net manager) وتكوين وصلة الارتباط بين الحاسوبين بتنفيذ أيعاز الربط وبالصيغة التالية[7] .

CREATE DATABASE LINK custom.world.

نفذ البحث بعدة عوامل:-

1- عامل نوع التجزئة

تم استخدام عامل التجزئة للملف الرئيسي بمستويين عمودي وافقي، ونفذ نوع التجزئة العمودية بأسلوبين (Vertical Local) وتم فيه تنفيذ مجموعة من جمل DML في الحاسبة المحلية (Local) و (Vertical Remote) أذ نفذت الايعازات نفسها في الحاسبة البعيدة (Remote) باستخدام الرابط. كانت التجزئة الأفقية بمستويين (Full Horizontal) ونفذت فيها الايعازات على الحجم الكامل من الحجوم المستخدمة، (

Half Horizontal) ويمثل نصف حجم البيانات المستخدمة من الحجوم الكلية وفي كلا الحاسبتين المحلية والبعيدة [11].

– تم استخدام بيانات ممثلة بالجدول customer وفيه المعلومات الأساسية لمشتركي الكهرباء وجزء من هيكليته كالتالي :

```
CREATE TABLE customer
    (customer_id    varchar2 (5),
     last_name      varchar2(20),
     first_name     varchar2(14),
     street         varchar2(120),
     city           varchar2(30),
     state      char(2),
     location       varchar2(25));
```

وتم إدخال البيانات برمجيا إلى الملف بحجم 136000 قيد في كل موقع من مواقع المخدم site1 و sit2 بافتراض إن كل موقع يمثل فرعاً لمديرية توزيع كهرباء مع إضافة اسم الموقع لأحد حقول الملف وذلك لغرض استخدامه في حالة التجزئة الأفقية، كما تم تجزئة جدول customer بصورة عمودية للحصول على جدول customer1في sit1 و customer2 في sit2 وكان تركيبهما:

```
CREATE TABLE customer1
( customer_id  varchar2(5),
  first_name   varchar2(14),
  last_name    varchar2(20),
  country      varchar2(20),
  location     varchar2(25));

 CREATE TABLE customer2
( customer_id  varchar2(5),
  street       varchar2(120),
  city         varchar2(30),
  state        char(2),
  location     varchar2(25));
```

2– عامل حجم البيانات

ويمثل عدد القيود في الملف الرئيسي customer بثمانية مستويات تتضمن 300 ، 600 , 900 ، 2100 ، 3000 ، 4500 ، 45000 ، 136000 قيد والتي نفذت باستخدام جمل DML بثلاثة مكررات للحصول على النتائج الإحصائية.وتم الحصول عليها باستخدام حزمة برمجيات التحليل الاحصائي (SAS) ، واستخدم جدول تحليل التباين ANOVA[1] للحصول على اختبارات Duncan والتي تستخدم للتمييز بين معدلات القيم البيانية (Means) بواسطة الأحرف الأبجدية للصفات البيانية المقاسة التالية

– وقت المعالجة (CPU time) يمثل الوقت المستغرق من قبل cpu في الوصول الى القرص الصلب والتي تؤثر على كفاءة معالجة الاستفسارات وتقاس بجزء الثانية (ms).

[1] تحليل التباين (analysis of variance ANOVA) هو مجموعة من النماذج الإحصائية (statistical model) مع إجرائيات مرافقة لهذه النماذج تمكن من مقارنة المتوسطات لمجتمعات إحصائية مختلفة عن طريق تقسيم التباين variance الكلي الملاحظ بينهم إلى أجزاء مختلفة.

- القراءة المنطقية (Logical Read) تمثل عدد مرات الوصول ومعالجة البيانات من خلال الذاكرة خلال الثانية الواحدة للفترة المحددة من الزمن وتقاس بالعدد (No).
- القراءة الفيزيائية (Physical Read) وهي عدد مرات نقل كتل البيانات من القرص إلى الذاكرة خلال الثانية الواحدة للفترة المحددة من الزمن وتقاس بالعدد (No).
- الكتابة الفيزيائية (Physical Write) تمثل عدد مرات الكتابة الفيزيائية على القرص خلال الثانية الواحدة خلال الفترة المحددة من الزمن وتقاس بالعدد (No).
- الوقت المستغرق بالتنفيذ (Elapsed time) وتمثل الوقت الكلي المستغرق في إعراب جمل SQL وتقاس (ms).
- مساحة المستخدم العامة (User Global Area UGA) وتمثل حجم UGA المستخدمة بالتنفيذ خلال المدة المحددة من الزمن وتقاس بالبايت (Byte) [10].

في قاعدة البيانات الكبيرة الحجم تكتسب هذه المتغيرات أهمية عالية وذلك لان معظم الإجراءات تحتاج إلى استرجاع عدد كبير من القيود مما يسبب وصول متكرر للقرص وينتج عنه تناقل كبير بين القرص والذاكرة [13].

هذه المتغيرات تم قياسها باستخدام مجموعة من جمل SQL التالية والتي نفذت في اداة SQL PLUS والتي تم وضعها في مقدمة جمل الاستفسارات وفي نهايتها وحساب الفرق في القيم للحصول على القيم البيانية يدويا بدلا من استخدام Oracle SQL Analyze Package لعدم قابليته على تنفيذ جمل الاستفسارات الموزعة.

```
select value logic_read  from v$sysstat where name='session logical reads' ;
select value cpu  from v$sysstat where  name='CPU used by this session';
select value phy_wr from v$sysstat where name='physical writes';
select value phy_read  from v$sysstat where name='physical reads' ;
select s.value uga from V$sesstat s,v$statname n where n.name='session uga memory'
and n.statistic#=s.statistic# or n.name='db blocks gets' ;
select s.value elapsed_time from V$sesstat s,v$statname n where n.name='parse time
elapsed' and n.statistic#=s.statistic# or n.name='db blocks gets' ;
```

3- عامل جمل الاستفسارات

وتم استخدام أنواع مختلفة من جمل الاستفسارات التي تسترجع وتحدث البيانات في قاعدة البيانات بحالتي التجزئة العمودية والتجزئة الافقية مع عملية التجميع (union operation) والمتمثلة في الجمل التالية :

جمل الاختيار (Selection) : تقوم هذه الجمل با سترجاع البيانات من قاعدة البيانات ومعظم جمل الاستفسارات التي تستخدم مع قواعد البيانات هي جمل الاختيار ومن المهم إن تقاس كفاءة هذا النوع من الايعازات [11].

Select vertical

```
select a.customer_id , a.first_name,b.customer_id,b.city
   from customer1 a,customer2@custom.world b
  where  a.customer_id<3000 and    b.customer_id<3000  and
   (a.customer_id=b.customer_id);
```

Select horezantle

```
select a.customer_id , a.first_name
   from customer a  where a.customer_id<1500
```

```
   union
 select b.customer_id , b.city
   from customer@custom.world b where b.customer_id  <1500;
```

عمليات معالجة البيانات (Data Manipulation Operation) : هناك نوعان من العمليات التي استخدمت مع مجموعة الاستفسارات وهي عملية التحديث (Update) وعملية الحذف (Delete) تقوم هذه العمليات بتحديث محتويات قاعدة البيانات و تم قياس كفاءة هذه الايعازات أيضا [11] .

Update_Vertical Local

```
update customer1 a set a.location='al_rafedin'
 where a.customer_id<3000;
```

update Vertical Remot

```
update customer2@ custom.world b set b.location='al_rafedin'
 where b.customer_id<3000;
```

delete Vertical Local

```
delete customer1 a where a.customer_id<3000;
```

delete Vertical Remot

```
delete customer2@custom.world a where a.customer_id<=3000;
```

من الصعب الحكم على هذه القيم الكبيرة العدد للصفات المقاسة مع أنواع المعاملات المختلفة (الايعازات , المعالجة , نوع التجزئة وحجم البيانات) بمستوياتها المختلفة لذا فقد اظهر التحليل الإحصائي باستخدام جدول تحليل التباين ANOVA [14] , نتائج تبرز أهمية هذه العوامل ومستوياتها مع إعطاء الثقة الإحصائية في الحكم على مستوياتها والتي بدونها لا نستطيع الحكم على نتائج البحث بهذا اليسر والدقة.

3- النتائج والمناقشة

باستخدام جمل الاستفسارات المطبقة في SQL PLUS لقياس بعض الصفات المهمة لتنفيذ الايعازات تم فحص ومقارنة كل واحدة من طرائق التنفيذ من ناحية تجزئة البيانات عموديا وأفقيا وتقييم نتائج البحث من ناحية تناقل حجوم البيانات المختلفة وكان الحد الأعلى للبيانات هو 136000 قيد لحصول بطء ملحوظ في زمن التنفيذ عند مضاعفة هذا الرقم.

من نتائج التحليل الإحصائي باستخدام تحليل التباين (ANALYSIS OF VERIANCE) وجد إن لكل من أنواع المعالجات (UPDATE, SELECT , DELETE) ونوع التقسيم (الأفقي والعمودي) ولحجم البيانات المستخدم تأثير معنوي (تأثير مقبول من الناحية الإحصائية) على جميع الصفات المقاسة (Logical read و , Physical Read CPU time وPhysical Write) ماعدا Elapsed time, UGA لم تتأثر تأثيرا معنويا مع الحجم والجداول ذات الارقام 1,2,3 تبين الفروقات في تأثير معدلات مستويات كل هذه الصفات وقد وجد إن هناك تأثيرا معنويا للتداخل ما بين ايعازات المعالجة ونوع التجزئة أي إن لكل منها تأثيرا على الآخر بضمان الفرق الإحصائي على جميع الصفات المقاسة التالية (Logical read ، CPU time، Physical Read ، Physical Write، Elapsed time ، UGA). كما اثبت التحليل الإحصائي أيضا تأثيرا معنويا للتداخل ما بين ايعازات المعالجة وحجم البيانات (بضمان تأثير كل منها على الآخر بموثوقية إحصائية) لجميع الصفات المقاسة وهذا ينطبق أيضا على التداخل ما بين نوع التجزئة وحجم البيانات .

جدول (1) يبين تأثير عملية التجزئة الأفقية والعمودية على البيانات الموزعة

Elapsed time Ms	UGA No	Physical Write No	Physical Read No	CPU time ms	Logical read No	**Kind**
a 501.05	**b** 96011	**a** 9307. 1	**a** 15759	**a** 118556	**a** 452797	Vertical_Remot
b 284.04	**c** 77665	**b** 8295. 7	**b** 7379	**a** 97265	**b** 349381	Vertical_Local
a 539.44	**a** 113717	**b** 7703. 7	**c** 1265	**b** 20530	**c** 38465	Horizantl_Full
a 434.67	**ab** 103390	**c** 6050. 3	**d** 363	**b** 13814	**d** 36604	Horezantl_Half

* إن الأرقام ذات الأحرف المشتركة لا تفرق عن بعضها إحصائيا

من الجدول (1) تبين إن هناك فروقات معنوية ما بين التجزئة العمودية والأفقية حيث انها اخذت احرف مختلفة ولم يكن هناك فرق ما بين مستويات التجزئة الأفقية أو العمودية في زمن CPU time وقد زادت عدد مرات Logical read وPhysical Read وPhysical Write مع التجزئة العمودية Vertical_Remot مع وجود تباين كبير جدا ما بين مستويات Physical Read ، مع عدم وجود تأثير ملحوظ على Elapsed time وقد يعزى ذلك لعدم علاقته بالتجزئة.

جدول (2) يبين تأثير نوع عملية المعالجة على البيانات الموزعة

Elapsed time Ms	UGA No	Physical Write No	Physical Read No	CPU time ms	Logical read No	**PROC**
b 291.62	**b** 83612	**a** 11640. 9	**a** 23202	**a** 172250	**a** 756420	Delete
a 438.49	**b** 83159	**b** 6368. 9	**b** 1719	**b** 66723	**b** 114677	Update
a 472	**a** 103738	**b** 6708. 8	**c** 914	**c** 20178	**c** 65058	Select

* إن الأرقام ذات الأحرف المشتركة لا تفرق عن بعضها إحصائيا

من الجدول(2) نلاحظ إن هنالك فروقات ملحوظة بين كل من عمليات Select , Delete , Update في جميع الصفات المقاسة حيث أنها أخذت أحرفاً متباينة وكانت عدد مراتLogical read وPhysical Read وPhysical Write قليلة مع عملية Select وأعلاها عند عملية الحذف Delete وانعكس هذا التأثير مع UGA و Elapsed time وهذا يتفق مع طبيعة الصفات المقاسة حيث أنها تمثل المساحة المحجوزة للمستخدم وزمن التنفيذ على التوالي.

جدول (3) يبين تأثير حجم البيانات المستخدمة على معالجة البيانات الموزعة

Elapsed time Ms	UGA No	Physical Write No	Physical Read No	CPU time ms	Logical read No	**SIZE**
a 480.33	**a** 79411	**Abc** 9431	**h** 5903	**d** 48351	**b** 395842	300
a 318.33	**a** 84592	**d** 7424.6	**g** 6533	**d** 53156	**g** 259502	600
a 349.72	**a** 89003	**d** 7645.2	**d** 8302	**b** 99864	**f** 310145	3000
a 432.44	**a** 92025	**cd** 8384.2	**e** 8126	**b** 97024	**e** 312551	9000
a 320.17	**a** 92010	**Bcd** 8470.8	**c** 8484	**bc** 94504	**d** 330429	21000
a 469.83	**a** 95966	**d** 7822.5	**f** 6619	**cd** 65836	**h** 219939	45000
a 469.25	**a** 89508	**ab** 9497.9	**b** 12970	**a b** 112179	**c** 373199	103000
a 475.08	**a** 89155	**a** 9913.3	**a** 16590	**a** 132171	**a** 545725	136000

* إن الأرقام ذات الأحرف المشتركة لا تفرق عن بعضها إحصائيا

ويلاحظ في جدول (3) ان لزيادة حجم البيانات تأثيرا ملحوظا ومعنويا بمستوياته المختلفة والتي تميزت عن بعضها إحصائيا باختلاف الأحرف وظهرت زيادة في عدد مرات Logical read وPhysical Read و Physical Write وCPU time ولم يكن للحجم أي تاثير ملحوظ على UGA و Elapsed time أذ إن جميع الإحجام كانت لها نفس الأحرف والسبب قد يعزى لما ذكر سابقا حول عدم تعلق هذه الصفات بعمليات المعالجة المنفذة.

جدول (4) يبين تأثير التداخل مابين نوع التجزئة والمعالجة

Elapsed time Ms	UGA No	Physical Write No	Physical Read No	CPU time ms	Logical read No	TYPE
						DELETE
ab 467.76	**ab**103945	**a** 12710.3	**a** 30707	**a** 194799	**a** 844237	Vertical_Remot
c 115.48	**e** 63279	**b** 10571.5	**b** 15698	**b** 149701	**b** 668603	Vertical_Local
						UPDATE
a 534.33	**cd** 88076	**d** 5903.8	**cd** 811	**d** 42312	**d** 61357	Vertical_Remot
b 326.67	**d** 77421	**cd** 6911.6	**c** 2778	**c** 95203	**c** 176883	Vertical_Local
						SELECT
ab 468.87	**bc** 98099	**cd** 6770.5	**cd** 1255	**d** 26329	**cd** 109468	Vertical
a 539.44	**A** 113717	**c** 7703.7	**cd** 1265	**d** 20530	**d** 38465	Horizantl_Full
ab 434.67	**ab**103390	**d** 6050.3	**d** 363	**d** 13814	**d** 36604	Horezantl_Half

* إن الأرقام ذات الأحرف المشتركة لا تفرق عن بعضها إحصائيا

من تحليل التباين ظهر بان التداخل ما بين نوع التجزئة ونوع المعالجة كان معنويا لكل صفات القياس ومن الجدول(4) يلاحظ بان السلوك العام لايعازات المعالجة هو نفسه ما ذكر في جدول (2) وكان تأثير عملية التجزئة داخل كل أيعاز هي بنفس الترتيب حيث تفوق تأثير التجزئة العمودية على التجزئة الأفقية في عمليات Logical read وPhysical Read وPhysical Write مع عدم وضوح هذا التأثير على كل من UGA , Elapsed time . كما لم يظهر لنوع التجزئة أي فرق بين مستوياتها (أفقياً وعمودياً) مع أيعاز Select.

جدول (5) يبين تأثير التداخل مابين الحجوم ونوع الايعاز

						DELETE
Elapsed time Ms	UGA No	Physical Write No	Physical Read No	CPU time Ms	Logical read No	SIZE
Abcde 373.7	**bc** 80444	**abcd** 11240.8	**ab** 22143	**def** 122058	**a** 816603	600
cde 248.5	**bc** 83947	**cd** 9889	**ab** 23458	**abc** 203790	**a** 853214	3000
cde 249.3	**bc** 82447	**a** 12802	**ab** 23677	**ab** 220369	**a** 862435	9000
cde 250	**bc** 82447	**a** 12833.7	**ab** 24294	**a** 237481	**a** 881221	21000
ce 300.2	**bc** 86425	**bcd** 10687.3	**b** 17998	**defg** 113714	**bc** 524579	45000
ce 305.5	**bc** 84788	**abc** 11818.7	**ab** 23021	**ce** 143468	**b** 575479	103000
bcde 314.2	**bc** 84788	**ab** 12215	**a** 27825	**bcd** 164869	**a** 781413	136000

SELECT

Elapsed time Ms	UGA No	Physical Write No	Physical Read No	CPU time Ms	Logical read No	SIZE
abc 480.3	**bc** 79411	**de** 9431	**c** 5903	**ghij** 48351	**cd** 395842	300

Abcde 397.9	**abc** 98610	**fg** 6109.8	**c** 435	**j** 3670	**f** 36126	600
abc 464	**ab** 107263	**f** 7460.7	**c** 1360	**j** 12102	**f** 37010	3000
abc 476.9	**abc** 97979	**fg** 6608.8	**c** 451	**ij** 23562	**f** 38360	9000
abc 574.5	**a** 125201	**fg** 6359.3	**c** 87	**hij** 33474	**f** 38773	21000
abc 477.2	**ab** 107246	**fg** 5994.2	**c** 215	**ij** 20555	**f** 37442	45000

UPDATE

Elapsed time Ms	UGA No	Physical Write No	Physical Read No	CPU time Ms	Logical read No	SIZE
de 143.7	**c** 67713	**fg** 5580. 7	**c** 69	**fghij** 58484	**f** 37464	600
bcde 336.7	**bc** 75801	**fg** 5585.8	**c** 89	**efgh I** 83699	**f** 40211	3000
a 665.3	**abc** 93318	**g** 4875	**c** 47	**fghij** 70720	**f** 35354	9000
e 136	**c** 68383	**fg** 6219.3	**c** 1072	**j** 12557	**f** 71293	21000
ab 632.2	**abc** 94227	**fg** 6786	**c** 1645	**fghij** 63240	**f** 97798	45000
ab 633	**abc** 94227	**f** 7177.2	**c** 2920	**efghi** 80890	**ef** 170919	103000
ab 636	**abc** 93521	**ef** 7611.5	**c** 5355	**efgh** 99474	**de** 310036	136000

* إن الأرقام ذات الأحرف المشتركة لا تفرق عن بعضها إحصائيا

أظهرت جداول تحليل التباين أثرا معنويا للتداخل مابين تاثير حجم البيانات ونوع الايعازات على مستوى اقل من 1% من الخطأ. ومن الجدول (5) يتضح بأنه لا يوجد تأثير معنوي لحجم البيانات مع الايعازات Delete وSelect و Update ولكنها تفرق بشكل عام مع كل منها على حدة أذ إنها تأخذ نفس الأحرف في عمليات Logical readوPhysical Read وPhysical Write و CPU time و مع كل إيعاز على حدة ولكنها تأخذ نفس هذه الأحرف مع كل إيعاز وهذا يظهر بان للإيعاز تأثيراً اكبر من تأثير حجم البيانات على البيانات الموزعة.

جدول (6) يبين تأثير التداخل مابين الحجوم ونوع المعالجة

						Vertical_Local
Elapsed time ms	UGA No	Physical Write No	Physical Read No	CPU time ms	Logical read No	SIZE
abc 480.3	**bc** 79411	**a b**9431	**a** 5903	**bcde** 48351	**ab** 395842	300
abcd 318.7	**bc** 89388	**a** 7765	**a** 4612	**abcde** 88217	**ab** 289341	600
D 16.2	**c** 58374	**a** 8268	**a** 8168	**a** 183792	**ab** 453615	3000
bcd 252	**c** 67437	**a** 9190	**a** 8339	**abc** 141881	**ab** 459613	9000
Cd 180.8	**bc** 81977	**a** 8090	**a** 6244	**abcde** 107432	**ab** 340007	21000
abcd 347.2	**bc** 81838	**a** 6734	**a** 2451	**cde** 33496	**ab** 118199	45000
abc 384.5	**bc** 77187	**a** 8887	**a** 98878	**abcde** 79243	**ab** 283495	103000
abc 393.5	**bc** 76481	**a** 9718	**a** 17117	**abcde** 102578	**a** 628465	136000
						Vertical_Remot
Elapsed time ms	UGA No	Physical Write No	Physical Read No	CPU time ms	Logical read No	SIZE
abcd 350.3	**bc** 74090	**a** 8303	**a** 15303	**bcde** 50034	**ab** 438421	600
ab 569	**ab** 101374	**a** 7207	**a** 15379	**abcde** 103697	**ab** 439810	3000
Ab 569.7	**ab** 101374	**a** 9207	**a** 15385	**abcd** 128656	**ab** 439918	9000

abcd 354.7	**bc** 89752	**a** 10108	**a** 16059	**abcde** 112052	**ab** 642736	21000
ab 553	**ab** 101829	**a** 10109	**a** 16061	**abcd** 128572	**ab** 462810	45000
ab 554	**ab** 101829	**a** 10109	**a** 16063	**abc** 145115	**ab** 462903	103000
Ab 556.7	**ab** 101829	**a** 10109	**a** 16063	**ab** 161765	**ab** 462985	136000
						Horezantl_Half
Elapsed time ms	UGA No	Physical Write No	Physical Read No	CPU time ms	Logical read No	SIZE
Bcd 289.3	**bc** 87901	**a** 5325	**s** 31	**e** 3645	**b** 35816	600
Bcd 260.3	**bc** 87901	**a** 6754	**a** 1463	**e** 3707	**b** 35340	3000
bcd 283	**bc** 87901	**a** 5316	**a** 24	**e** 3708	**b** 35359	9000
a 669.3	**a** 126625	**a** 6341	**a** 55	**de** 20624	**b** 37083	21000
a 671.3	**a** 126625	**a** 6515	**a** 240	**ce** 37387	**b** 39420	45000
						Horizantl_Full
Elapsed time ms	UGA No	Physical Write No	Physical Read No	CPU time ms	Logical read No	SIZE
Bcd 282.3	**bc** 87901	**a** 6748	**a** 1256	**e** 3731	**b** 35830	600
a 667.7	**a** 126625	**a** 8167	**a** 1256	**e** 20497	**b** 38680	3000
a 668.3	**a** 126625	**a** 8195	**a** 1283	**ce** 37363	**b** 40884	9000

* إن الأرقام ذات الأحرف المشتركة لا تفرق عن بعضها إحصائيا

جدول (6) لم يكن للتداخل ما بين نوع التجزئة وحجم البيانات أي تأثير معنوي على الصفات التالية Physical Write، Physical Read فيما كان للتداخل تأثير ملحوظ على مستوى 1% على بقية الصفات ومن الجدول (6) لم يلاحظ أي فرق ما بين الحجوم ضمن أي نوع من انواع التجزئة مع صفة Logical read, Elapsed time بينما كان هناك فرق من التأثير مع صفة CPU time من خلال Horezantl_Half مع عدم وجود فروق ما بين حجوم البيانات المختلفة ضمن أنواع التجزئة.

الاستنتاجات Conclusion : تبين من نتائج هذه الدراسة ما يلي :-

1- ان الصفات Logical read وPhysical Read وPhysical Write ممكن استخدامها في تقييم أداء قواعد البيانات الموزعة باستخدام لغة أوراكل .

2- إن الصفات Elapsed time , UGA لم تتأثر بالمعاملات المختلفة للبيانات المستخدمة وهي بالتالي لاتصلح لتقييم أداء قواعد البيانات الموزعة .

3- ملائمة استخدام الأسلوب الإحصائي في تحليل التباين ANOVA واختبارات Duncan في تقييم أداء قواعد البيانات الموزعة وسهولة الحكم عليها مقارنة بالطرق الإحصائية الأخرى.

4- اختلاف كلا نوعي التجزئة الأفقي والعمودي عن بعضها في التأثير.

5- الحاجة إلى معرفة مدى تأثير كفاءة تناقل البيانات في قواعد البيانات الموزعة باختلاف حجم البيانات ليس من ناحية عدد القيود وإنما حسب نوع البيانات مثل بيانات Multi Media.

المصادر

[1] Ayse yasemin seydim.1998. An OverView Of Distributed Database Management,cse8343-distributed operating systems fall 1998 term project.

[2] Distributed Data base,2002, www.Databasesystem@EricLlewellyn.

[3] Christopher G. B.; Guster D.; Krzenski S. 2007. Journal of Information Technology Management Volume XVIII, Number 1.

[4] Katja Hose , Ralf Schenkel .2010. Distributed Database Systems Introduction, Max_planck_Institut for Informatik.

[5] Kazerouni L., K.Karlapalem.2011. stepwise Redesign of Distributed Relational Databases, "Department of Computer Science the Hong Kong University of Science & Technology , wenku.baidu.com/71c5c2be9605.html.

[6] Matthias Nicola and M. Jarkey. 2006. Performance Modeling of Distributed and Replicated Databases. Technical University of Aachen, Informatik V (Information Systems) Ahornstr. 55, 52056 Aachen, Germany.

[7] Note on Distributed Database Systems, www.spynos.so.ukerase.

[8] Oracle corporation .2001.Database Distributed Strategies, world headquarters 500 oracle parkway redwood shores , ca 94065 USA,www.oracle.com.

[9] Oracle9i Concepts,Database Replication,2002,release 2.(9.2),partno:A96531-02.

[10] Ozuu, M. ,V. Principles of Distributed Database Systems,2011, http://www.springer.com/978-14419-8833-1.

[11] Server / Client Database and Distributed database , 2008, printed by Permission of state of Tennessee department of safety.

[12] Tore Risch .Distributed Database Systems, , Dept. of Information Technology , Uppsala University , Sweden. Tore.Risch@it.uu.se.

[13] Tuning Poor performance SQL using oracle 10g Enterprise Managers Automatic SQL Tuning Advisor, 2006, EMC Developer network, http://developer_emc.com.

[14] Venables W. N. and B. D. Reply .2002. Modern Applied Stattistics with S. 4th Edition . Springer.

[15] Yagoub K.,.Belknap, P.,Dagevile P .2008. Oracle SQL Performance Analyzer, @oracle.com, copy right 2008 IEEE,Bulletin of the IEEE computer society technical committee on data Engineering.