بسم الله الرحمن الرحيم

جامعة السودان للعلوم والتكنولوجيا
كلية الدراسات العليا

***

مقارنة قواعد البيانات المركزية والعنقودية في نظام MySQL بإستخدام إستعلامات select بسيطة

A Comparison of MySQL Centralized and Clustered DataBases using Simple SELECT Queries

بحث مقدم كأحد متطلبات نيل درجة الماجستير في علوم الحاسوب

مقدم البحث:
هبة عبدالمنعم المكي محمدأحمد

المشرف:
د.محمد عوض الشيخ

سبتمبر 2008

الــحمد لله

الحمد لله الملك المعبود الرحيم الودود ،،،

فتح بابه للطالبين وحث على دعائه في كتابه المبين ،،،

أحمده سبحانه وأشكره على فضله وعطائه وعلى جزيل نعمائه ،،،

وأحمده سبحانه أن هيأ لي كل من ساعدني لإكمال هذا البحث ،،،

فله الحمد أولا وأخيرا ،،،

أشهد أن لا إله إلا الله وحده لا شريك له ،،،

وأشهد أن محمدا عبده ورسوله الصادق الأمين

(صلي الله عليه وسلم وعلى آله و صحبه أجمعين)

،،،

الأية

(اللَّهُ نُورُ السَّمَاوَاتِ وَالْأَرْضِ ۚ مَثَلُ نُورِهِ كَمِشْكَاةٍ فِيهَا مِصْبَاحٌ ۖ الْمِصْبَاحُ فِي زُجَاجَةٍ ۖ الزُّجَاجَةُ كَأَنَّهَا كَوْكَبٌ دُرِّيٌّ يُوقَدُ مِنْ شَجَرَةٍ مُبَارَكَةٍ زَيْتُونَةٍ لَا شَرْقِيَّةٍ وَلَا غَرْبِيَّةٍ يَكَادُ زَيْتُهَا يُضِيءُ وَلَوْ لَمْ تَمْسَسْهُ نَارٌ ۚ نُورٌ عَلَىٰ نُورٍ ۗ يَهْدِي اللَّهُ لِنُورِهِ مَنْ يَشَاءُ ۚ وَيَضْرِبُ اللَّهُ الْأَمْثَالَ لِلنَّاسِ ۗ وَاللَّهُ بِكُلِّ شَيْءٍ عَلِيمٌ)

صدق الله العظيم
سورة النور – الأية 35

إهـــــــــــــداء

بكل ذرة جهد و نسمة فـــــكر ،،،

بكل الإخـــــلاص و المـــودة الخالصة ،،،

بكل المعـــــاناة في سبر بـــحر العلم ،،،

وبكل الجـــــد في التنـــاول البحثـــي

أهـــــــدي هذا الجهـــد على تـــواضعه

لـــــكل من أســـــهم في أن يـــرى هـذا البحـــث

الـــــــــــــــنور

شـــكر و عرفان

الحمد الله تعالى على ما أنعم علي به من نعمه ظاهرة وباطنة ،،،
أتقدم بالشكر أجزله لأستاذي الفاضل دكتور محمد عوض الشيخ لمعين عونه الذي لا ينضب أثاء إشرافه على هذا البحث ،،،
والشكر موصول لكل من
المهندس :: أحمد عبدالحليم
المهندس :: مصعب عبد المنعم
المهندس :: علاء الدين اسماعيل
الأخ :: مظفر هجو
لما قدموه لي من مساعدة قيمة في التعامل مع نظام التشغيل Linux و قاعدة البيانات MySQL وفي حل العديد من الصعوبات التقنية التي واجهت تقدم البحث .

المستخلص

تعتبر قواعد البيانات الدعامة الأساسية للأنظمة، بحيث يكاد لايوجد نظام لا يحتوي على قاعدة بيانات، كما أصبحت معظم هذه الأنظمة أنظمة تعتمد على الشبكات في عملها لخدمة أكبر عدد من العملاء، لهذا وجدت عدة طرق لتصميم قواعد البيانات. الطريقة الأولى هي تصميم قاعدة البيانات بصورة مركزية أي أن تكون قاعدة البيانات في جهاز واحد، ولكن أصبحت هذه الأنظمة تعاني من مشاكل في إستمرارية الخدمة وسرعتها وذلك عندما تخدم عدد كبير من العملاء، ومثال لذلك ما حدث لنظام نتيجة الشهادة السودانية للعام الدراسي 2007/2006 عبر المخدم الموجود بجامعة السودان للعلوم والتكنولوجيا والذي توقف عن العمل بعد فترة بسيطة من إتاحة الفرصة للطلاب للدخول عليه نتيجة لكثرة الطلبات في نفس الوقت والتي لم يستطع المخدم التعامل معها، لهذا تم الإتجاه للطريقة الثانية وهي تصميم قاعدة البيانات بصورة موزعة لتجنب المشاكل التي حدثت من الطريقة الأولى.

تعتبر قاعدة البيانات العنقودية Mysql cluster آلية لتطبيق قاعدة البيانات الموزعة. لقاعدة البيانات العنقودية مزايا عديدة أهمها الإتاحية العالية للنظام.

يقدم البحث إختباراً عملياً لمستوى أداء النظام المركزي والنظام العنقودي، والمقارنة بينهما لتحديد أفضل مستوى أداء حسب ظروف الأختبار والتي تتمثل في عدد العملاء وعدد الطلبات المرسلة. وقد تبين أن مستوى الأداء للنظام المركزي أفضل من العنقودي عندما يكون عدد العملاء قليلاً وفي مدى معين من الطلبات، وعند زيادة عدد الطلبات عن مدى معين وأيضا زيادة عدد العملاء يصبح أداء النظام العنقودي أفضل من المركزي.

Abstract

Database is considered as the basic support to the system. There is practically no such system which contains no database. Most systems have become dependent on networks whose function is to serve the biggest number of clients. Therefore, there are several ways of designing a database. The first way is designing database centrally; that means all the database is set in one central computer. Central database have many problems regarding availability and response time, specially when a big number of clients is served. For example, the failure of the system of the results of the Sudanese Certificate 2006/2007 whose server was in Sudan University of Science and Technology; failed right after the students requested to know their scores. That's because the requests were too many for the server. That is why the attention was focused on the second way, which is designing a distributed database in order to avoid the problems caused by the first approach (central database).

Mysql cluster is the mechanism to apply the distributed database. The most important advantage of mysql cluster is high availability.

The research presents a practical test for the performance of the central and cluster database. The objective is to compare the two approaches in order to find out the best performance with regard to the circumstances of the test: the number of clients and requests. The results show that, the central performance is better than the cluster when then the number of clients is small and requests is small within a certain range. When there is an increase in number of clients and the requests range, the performance of the cluster database will be more efficient.

المصطلحات

Database Cluster ……..	قاعدة البيانات العنقودية.....
Node ……	عقدة .....
Management node (MGM node).....	عقدة الإدارة .....
Data node.....	عقدة البيانات .....
Availability……..	إتاحة الخدمة .....
Performance.....	أداء النظام .....
Query……	إستعلام .....
Centralized Database.....	قاعدة بيانات مركزية .....
Distributed Database……..	قاعدة بيانات موزعة .....
Application.....	تطبيق .....
local application.....	تطبيق محلي .....
global application.....	تطبيق عام.....
Scalable……..	التوسع.....
transaction.....	المعاملات.....
Asynchronous data replication……	التكرار غير المتزامن للبيانات.....
synchronous data replication……..	التكرار المتزامن للبيانات.....
storage engine…….	محرك التخزين.....
shared-nothing architecture……	معمارية عدم المشاركة.....
Distributed DataBase Management System(DDBMS)…….	نظام إدارة قاعدة البيانات الموزعة.....

فهرس الأشكال

رقم الشكل	موضوع الشكل	رقم الصفحة
شكل (1.2)	الشكل العام للأنظمة المركزية	7
شكل (2.2)	الشكل العام للأنظمة الموزعة	8
شكل (3.2)	معمارية عدم المشاركة	9
الشكل (4.2)	التكرار القياسي في Mysql	13
الشكل (5.2)	معمارية Mysql	15
الشكل (1.3)	معمارية النظام العنقودي Cluster	19
الشكل (2.3)	تقسيم البيانات و تكرارها في مجموعتين من العقد.....	22
الشكل (3.3)	نموذج لتكرار مكونات النظام العنقودي Cluster	23
الشكل (1.4)	وصف جدول City	26
الشكل (2.4)	وصف جدول Country	26
الشكل (3.4)	وصف جدول CountryLanguage	27
الشكل(4.4)	وصف جدول time_query	27
الشكل (5.4)	وصف ملف config.ini	28
الشكل (6.4)	وصف ملف my.cnf	30
الشكل (7.4)	وصف دالة select()	31
الشكل (8.4)	خوارزمية عمل كل من المخدم والعميل في حالة النظام المركزي...	33
الشكل (9.4)	خوارزمية عمل كل من المخدم والعميل في حالة النظام العنقودي...	34
الشكل(1.5)	متوسط الزمن لقاعدة البيانات المركزية عند إستخدام (1 ، 3، 5) عملاء في المدي من 100 – 1000000 طلب	37
الشكل(2.5)	وصف زمن تنفيذ النظام المركزي	38
الشكل(3.5)	متوسط الزمن لقاعدة البيانات العنقودية عند إستخدام (1 ، 3، 5) عملاء في المدي من 100 – 1000000 طلب	39
الشكل(4.5)	وصف زمن تنفيذ النظام العنقودي	40
الشكل(5.5)	مقارنة بين متوسط الزمن لقاعدة البيانات المركزية و العنقودية عند إستخدام عميل واحد في المدي من 100 – 1000000 طلب	41
الشكل (6.5)	وصف زمن تنفيذ النظام المركزي و العنقودي لعميل واحد	41
الشكل(7.5)	مقارنة بين متوسط الزمن لقاعدة البيانات المركزية و العنقودية عندإستخدام 3عملاء في المدي من 100 – 1000000 طلب	43
الشكل (8.5)	وصف زمن تنفيذ النظام المركزي و العنقودي لثلاثة عملاء	43
الشكل (9.5)	مقارنة بين متوسط الزمن لقاعدة البيانات المركزية و العنقودية عند إستخدام 5عملاء في المدي من 100 – 1000000 طلب	45
الشكل (10.5)	وصف زمن تنفيذ النظام المركزي و العنقودي لخمسة عملاء.....	46
الشكل (11.5)	مقارنة بين متوسط الزمن لقاعدة البيانات المركزية و العنقودية عند إستخدام (1 و 3 و 5) عملاء المدي من 100 – 1000000 طلب	47

فهرس الجداول

رقم الجدول	موضوع الجدول	رقم الصفحة
جدول (1.4)	الخيارات الأساسية لتهيئة عقدة الأدارة	28
جدول (2.4)	الخيارات الأساسية لنهيئة عقدة البيانات و عقدة التطبيق	29
جدول(1.5)	يوضح عدد الجداول أثناء التجربة	36
جدول(2.5)	متوسط الزمن لكل التجارب	36
جدول(3.5)	تقريرالأداء الأفضل في حالة عميل واحد	42
جدول(4.5)	تقرير الأداء الأفضل في حالة ثلاثة عملاء	44
جدول(5.5)	تقرير الأداء الأفضل في حالة خمسة عملاء	46

فهرس المحتويات

الباب - الفصل	الموضوع	الصفحة
الباب الأول	**المقدمة**	1
1.1	مقدمة	2
2.1	المشكلة	2
3.1	فكرة البحث	2
4.1	أهداف البحث	3
5.1	معوقات البحث	3
6.1	بيئة العمل	4
7.1	الشكل العام للبحث	4
الباب الثاني	**قواعد البيانات الموزعة و MySQL Server**	5
1.2	قاعدة البيانات المركزية	7
2.2	قاعدة البيانات الموزعة	7
1.2.2	معمارية الأنظمة الموزعة Architecture of DDBMS	8
2.2.2	مزايا وعيوب الأنظمة الموزعة	9
1.2.2.2	مزايا الأنظمة الموزعة	9
2.2.2.2	عيوب الأنظمة الموزعة	10
3.2.2	تصميم قاعدة البيانات الموزعة Design of DDB	10
1.3.2.2	التجزئة Fragmentation	10
2.3.2.2	التسكين Allocation	11
3.3.2.2	التكرار Replication	12
3.2	قاعدة البيانات MySQL Server	13
1.3.2	التكرار في Mysql (Mysql Replication)	13
الباب الثالث	**قاعدة البيانات العنقودية Mysql Cluster**	16
1.3	مقدمة عن قاعدة البيانات العنقودية Mysql Cluster …	17
2.3	بنية و معمارية Mysql Cluster	17
3.3	مزايا قاعدة البيانات العنقودية Mysql Cluster	20
4.3	التكرار في Mysql Cluster	21
1.4.3	مقدمة	21
2.4.3	آلية التكرار Replication mechanism	21
الباب الرابع	**التـطبيق العمـــلي**	24
1.4	وصف لبيئة العمل	25
2.4	وصف قاعدة البيانات المستخدمة	25
3.4	شرح لبعض التفاصيل العملية	27
1.3.4	عقد لنظام العنقودي	27
2.3.4	قاعدة البيانات	30
4.4	برنامج إختبار أداء Cluster	30

5.4 تجربة النظام 35

الباب الخامس **النتائج و التحليل** 35
1.5 النتائج 36
1.1.5 عدد الجداول 36
2.1.5 متوسط الزمن 36
2.5 التحليل..... 37
1.2.5 أداء النظام المركزي عند تغيير عدد العملاء 37
2.2.5 أداء النظام العنقودي عند تغيير عدد العملاء 39
3.2.5 مقارنة أداء النظام العنقودي مع النظام المركزي بإستخدام عميل واحد 41
4.2.5 مقارنة أداء النظام العنقودي مع النظام المركزي بإستخدام ثلاثة عملاء 42
5.2.5 مقارنة أداء النظام العنقودي مع النظام المركزي بإستخدام خمسة عملاء 45

الباب السادس **الخاتـــمة والتوصيات** 48
1.6 الخاتمة 49
2.6 التوصيات والمهام المستقبلية..... 49
المراجع 50
الملاحق 52
ملحق (أ) 53
ملحق (ب) 54
ملحق (ت) 55
ملحق (ث) 56
ملحق (ج) 57
ملحق (ح) 58
ملحق (خ) 63
ملحق (د) 69