

오픈소스 DBMS의 성능 품질 평가

민미경[†]

서경대학교 컴퓨터학과

Evaluating the Performance Quality of Open Source Database Management Systems

Min, Meekyung[†]

Dept. of Computer Science, Seokyeong University

ABSTRACT

Purpose: The purpose of this paper is to evaluate the performance quality of the open source DBMSs. Performance quality is defined as processing time for Join queries. Query processing time is measured and compared in the most widely used open source DBMSs and commercial DBMS.

Methods: By varying the number of tuples of two relations to be joined, the average processing time(seconds) of a Join query in each DBMS was obtained experimentally. ANOVA and Tukey HSD test were used in order to compare the performance quality of DBMSs.

Results: There was a significant difference between the performance qualities of the three DBMSs at all experimental levels where the number of tuples was 100, 1,000, 2,000, 10,000, and 50,000. As a result of the Tukey HSD test, two open source DBMSs (MariaDB, MySQL) were classified in the same group only at the tuple level of 100. The commercial DBMS (MS-SQL Server) belonged to another group. At level of more than 1,000 tuples, all three DBMSs belonged to different groups.

Conclusion: Within the open source DBMS group, MariaDB showed the better performance quality except for a small number of tuples. Thus the results show that MariaDB can be the alternative to MySQL which is currently most widely used. Between open source DBMS and commercial DBMS groups, MS-SQL Server always shows the best performance quality, but the less number of tuples, the less the difference.

Key Words: Open Source Software, DBMS, Join Query, Time Performance, Tukey HSD

● Received 9 November 2017, 1st revised 12 November, accepted 13 November 2017

† Corresponding Author(mkmin@skuniv.ac.kr)

© 2017, The Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

* This Research was supported by Seokyeong University in 2017.

1. 서 론

오픈소스 소프트웨어(Open Source Software)는 개인용 소프트웨어에서부터 기업용 소프트웨어에 이르기까지 고가의 상용 소프트웨어에 대한 대안으로 점차 널리 사용되고 있다. 이러한 추세에 따라 오픈소스 소프트웨어의 품질 평가의 필요성이 대두되어 적합한 품질 측정 모델이나 품질 측정 도구 구현에 관한 연구가 진행되어 왔다. 이들은 ISO/IEC 표준에서 제시하는 소프트웨어 품질 특성을 기반으로 오픈소스에 적합한 특성 모델을 제시하거나, 실제 오픈소스 코드를 분석해서 품질을 측정하는 연구 등이다.

오픈소스 소프트웨어에는 다양한 종류가 있으며, 서버용 오픈소스 소프트웨어로 분류될 수 있는 데이터베이스 관리 시스템(DBMS: Database Management System)은 이 중 하나이다. DBMS는 사용자와 데이터베이스간의 소통을 위해 데이터베이스를 관리하는 소프트웨어 시스템으로서 데이터를 효율적으로 저장·관리하고 검색·사용하도록 지원하는 것이 주요한 목적이다. 최근에는 전통적인 상용 DBMS 뿐만 아니라 오픈소스 DBMS도 많이 사용되고 있다. 일반적인 DBMS의 품질에는 여러 가지 특성이 있을 수 있으나, 이 중 DBMS의 품질을 결정하는 가장 중요한 특성은 효율성, 즉 성능품질이다. 성능품질을 구성하는 시간성능과 공간성능 중에서는 DBMS의 처리시간을 나타내는 시간성능이 보다 중요시 되고 있다. 따라서 오픈소스 DBMS가 기존 상용 DBMS의 대체재로서 적합한지 알기 위해서는 시간성능 품질을 연구할 필요가 있다.

본 논문에서는 오픈소스 DBMS의 성능품질을 측정하여 이를 비교한다. 성능품질은 검색 쿼리에 대한 처리시간으로 정의한다. 윈도우즈 운영체제 하에서 HeidiSQL을 프론트엔드로 사용하며, 3개의 DBMS를 대상으로 성능을 측정한다. 오픈소스 DBMS로는 가장 널리 사용되는 MySQL과 MariaDB를 대상으로 하며, 상용 DBMS로는 가장 대표적인 윈도우즈용 DBMS인 MS-SQL Server를 대상으로 하여 서로 비교한다. 쿼리는 검색 쿼리 중 가장 처리시간이 오래 걸리는 조인 쿼리(Join query)를 사용하고, 튜플 수를 변화시켜 가면서 이에 따른 처리시간을 측정한다. 또한 분산분석을 통해 3개 DBMS의 쿼리 처리시간 간에 차이가 존재하는가 검정을 실시하고, 차이가 나는 그룹을 찾기 위해 Tukey HSD 검정을 실시한다.

다음 2장에서는 관련연구로 오픈소스 소프트웨어와 소프트웨어의 품질 평가 방법 및 오픈소스 DBMS의 현황을 살펴본다. 3장에서는 DBMS의 성능 측정 실험 결과를 보인다. MariaDB, MySQL, MS-SQL Server가 조인 쿼리를 처리하는 시간성능을 측정하고, 이를 비교분석한다. 4장에서는 결론을 맺는다.

2. 관련 연구

2.1 오픈소스 소프트웨어와 품질

오픈소스 소프트웨어(Open Source Software)는 소스코드에 대한 접근, 자유로운 재배포, 파생 저작물의 작성, 제한 없는 사용 등을 허용하는 라이선스와 함께 배포되는 소프트웨어이다(Androutsellis-Theotokis et al. 2011). 우리나라 공개 소프트웨어 포털에서는 널리 사용되는 오픈소스 소프트웨어(OSS)를 OS, DBMS, WEB, WAS와 같은 서버용 OSS, 데스크탑 용 OSS, 비즈니스 용 OSS, 임베디드 OSS, 클라우드 OSS 등으로 분류하고 있다. 오픈소스 애플리케이션들은 운영체제와 데이터베이스 같은 시스템 인프라, 소프트웨어 개발 도구, 교육, 과학, 엔지니어링 등

개인용 소프트웨어에서 기업용 소프트웨어까지 거의 모든 분야에서 그 활용도가 높아지고 있다.

소프트웨어 품질 평가를 위한 국제표준인 ISO/IEC 9126, ISO/IEC 25010 에서는 소프트웨어 품질특성 및 메트릭을 정의하고 있다. 이 품질 모델은 소프트웨어 품질을 측정, 평가하기 위해 소프트웨어의 품질을 이루는 품질 특성을 정의하고 이 품질 특성을 다시 부특성들로 세분한다. 또한 소프트웨어 제품의 품질을 측정하거나 평가하는데 필요한 방법과 절차를 여섯 개 부분으로 나누어 정의하여 표준으로 제시하고 있다(ISO/IEC 2001, 2003, 2011). 국내에서는 NIPA와 TTA에서 표준 지침을 제공하고 있다(Lim et al. 2012).

ISO/IEC 9126의 품질 특성과 부특성은 다음과 같다.

- . 기능성(Functionality): 부특성-합목적성, 정확성, 상호운영성, 표준적합성, 보안성
- . 신뢰성(Reliability): 부특성-성숙성, 장애허용성, 회복성
- . 사용성(Usability): 부특성-이해성, 습득성, 운용성
- . 효율성(Efficiency): 부특성-시간효율성, 자원효율성
- . 이식성(Portability): 부특성-해석성, 변경성, 안정성, 시험성
- . 유지보수성(Maintainability): 부특성-환경적응성, 설치성, 규격적합성, 치환성

이러한 국내의 표준을 바탕으로 다양한 응용분야의 소프트웨어에 대한 품질 평가 모델 연구가 이루어져 왔다. 빅데이터 분산처리시스템의 품질평가 모델(Choi et al. 2014), IoT 소프트웨어의 품질평가모델(Chung et al. 2016), 모바일 클라우드 소프트웨어의 품질평가모델(Lee et al. 2015) 등의 연구가 있다.

소프트웨어 제품 자체의 품질만이 고려되는 일반 소프트웨어와 다르게 오픈소스 소프트웨어는 제작, 배포과정과 커뮤니티, 라이선스 등이 중요한 요소로 작용한다. 따라서 별도의 품질 평가 방법이 필요하며, 오픈소스 소프트웨어의 품질평가에 관한 연구가 진행되어 왔다(Spinellis 2006, Spinellis et al. 2009, Samoladas et al. 2008). 이 중 Samoladas et al.(2008)에서는 소스코드가 개방되어 있는 오픈소스 소프트웨어들의 사례를 코드 수준에서 분석하였는데, 품질평가 특성 중에서는 오픈소스 소프트웨어에 적합한 신뢰성, 보안성, 효율성, 이식성, 유지보수성 특성을 선택하여 분석하였다. Spinellis(2006)과 Spinellis et al.(2009)에서는 ISO/IEC 9126 품질 모델을 오픈소스 소프트웨어에 맞게 변형한 모델을 제시하였으며, 이 모델을 기반으로 품질 평가를 자동화하는 프로그램을 개발하였다.

2.2 오픈소스 DBMS

DBMS(Database Management System)란 데이터베이스에 대한 모든 접근을 관리하는 소프트웨어이다. DBMS가 필수적으로 갖추어야 할 기능에는 데이터 정의기능, 데이터 조작기능, 최적화, 데이터 보안 및 무결성 유지기능, 데이터 회복 및 동시성 제어 기능, 데이터 사전 유지 기능이 있다(Date 2004). DBMS의 주요 역할은 데이터베이스의 데이터에 대한 검색, 삽입, 삭제, 갱신의 연산이다. 이 중 가장 많이 사용되는 것은 검색 연산으로, DBMS의 가장 중요한 선택 기준은 데이터베이스에 대한 빠른 검색이다. 이에 관한 연구는 데이터베이스가 등장한 이래로 수십 년간 계속되어 왔다. 따라서 DBMS의 품질에서 가장 중요한 품질 특성은 시간 효율성, 즉 시간성능이며, 이는 상용 DBMS를 대체할 오픈소스 DBMS의 선택에서도 마찬가지로 가장 중요한 품질 기준이 된다.

가장 널리 사용되는 상용 DBMS에는 오라클사의 Oracle과 Microsoft사의 MS-SQL Server 등이 있다. 오픈소스 DBMS는 서버용 오픈소스 소프트웨어로서, 대표적인 오픈소스 DBMS에는 MySQL, PostgreSQL, MariaDB, CUBRID, SQLite, MongoDB 등이 있다.

이 중에서 MySQL은 세계에서 가장 많이 사용하는 오픈소스 DBMS이다. 다중 스레드, 다중 사용자 형식의 구조

질의어 형식의 관계형 데이터베이스 관리 시스템으로서 MySQL AB가 관리 및 지원하고 있다. MariaDB도 오픈소스 관계형 DBMS로서 MySQL과 동일한 소스 코드를 기반으로 한다. MariaDB는 현재 오라클 소유의 불확실한 MySQL의 라이선스 상태에 반발하여 만들어졌으며, 배포자는 몬티 프로그램 AB(Monty Program AB)와 저작권을 공유하도록 되어있다. MariaDB는 MySQL과의 완벽한 호환성으로 인해 오픈소스 소프트웨어로서 그 위상이 매우 높다. 마이크로소프트사의 MS-SQL Server는 마이크로소프트가 1989년에 사이베이스(Sybase)를 기반으로 개발한 관계형 데이터베이스 관리시스템이며, 윈도우즈 운영체제 하의 상용 DBMS를 대표한다. 국내에서는 시장점유율 약 17%로 Oracle과 더불어 전 세계에서 가장 높은 비율로 DBMS 시장을 점유하고 있다.

3. 성능 측정 및 품질 분석

3.1 실험 환경

본 논문에서 3개의 DBMS를 대상으로 성능을 측정하였다. 오픈소스 DBMS로는 가장 널리 사용되는 MySQL과 MariaDB를 택하였다. 윈도우즈 환경 하에서 성능을 측정하였으므로 상용 DBMS 중에는 윈도우즈 DBMS의 대표인 MS-SQL Server를 택하였다.

동일 환경, 동일 인터페이스 하에서 3개 DBMS의 성능을 측정하기 위해 HeidiSQL을 프론트엔드로 사용한다. HeidiSQL은 자유 소프트웨어이며 오픈 소스 클라이언트로서 MySQL 프론트엔드 제품이다. 윈도우즈 환경 하에서 MySQL, MariaDB, MS-SQL Server를 모두 지원한다.

성능을 측정한 실험 환경은 다음과 같다. CPU i7-4790 3.6GHz, 8GB RAM, Intel HD Graphics 4600을 장착한 PC를 사용하였으며, 운영체제는 64비트 Windows 10, HeidiSQL은 버전 9.4를 사용하였다. DBMS는 MariaDB-10.2.9-winx64, MySQL Ver 14.14 Distrib 5.6.13 for win64, Microsoft SQL Server 2014를 사용하였다.

3.2 성능 측정 결과

DBMS의 성능을 측정하기 위해서 검색 쿼리 중에서 시간이 가장 많이 소요되는 조인 쿼리를 생성하였다. 2개의 릴레이션이 1개의 외래 키에 의해 연결된다고 가정하고 쿼리 Where 절의 조인 조건은 1개로 단순화하였다. 조인 선택율은 10%라고 가정하였다. 튜플은 무작위로 생성하였으며, 릴레이션에 포함된 튜플의 수를 변화시키면서 변화된 튜플 수에 따라 조인 쿼리를 수행하는 처리시간을 측정하였다. 튜플은 무작위로 생성하였다. 조인하는 튜플의 수를 2개의 릴레이션 모두에서 100개, 1000개, 2000개, 10,000개, 50,000개로 변화시키면서, 작업로드가 적은 100개 부터 작업로드가 큰 5만개의 경우까지 DBMS의 조인 쿼리 처리시간을 보았다. 릴레이션의 튜플 수에 따라 각각의 경우에 3개의 DBMS에서의 처리시간을 각 70회씩 측정한 후 이들의 평균값을 데이터로 사용하였다.

조인 쿼리의 실험 결과 릴레이션의 튜플 수의 변화에 따른 조인 쿼리 처리시간의 평균과 표준편차는 Table 1과 같다.

Table 1. Average processing time and STD for Join query

No	DBMS		Number of tuples in each relation				
			100	1000	2000	10,000	50,000
1	MariaDB	Average time (sec.)	0.043708	0.089174	0.209087	4.067886	94.40719
		STD (sec.)	0.008822	0.011251	0.014337	0.063534	0.586415
2	MySQL	Average time (sec.)	0.044513	0.098148	0.230953	4.562182	104.1560
		STD (sec.)	0.012983	0.017162	0.013316	0.057183	0.109484
3	MS-SQL Server	Average time (sec.)	0.035946	0.037140	0.039855	0.047543	0.093398
		STD (sec.)	0.008135	0.009895	0.009535	0.009011	0.011668

Table 1의 데이터에서 릴레이션의 튜플 수에 따른 조인 쿼리의 평균 처리시간을 그래프로 나타내면 Figure 1과 같다. 가로축은 튜플의 수이며, 세로축은 처리시간(초)이다. 좌측(LHS) 세로축의 눈금은 MariaDB와 MySQL을, 우측(RHS) 세로축의 눈금은 MS-SQL Server를 나타낸다. MS-SQL Server는 튜플의 수에 관계없이 항상 가장 좋은 성능을 보이며, 튜플 수에 따른 처리시간의 증가속도가 작다. 즉 튜플이 많아질수록 상용 DBMS인 MS-SQL Server와 오픈소스 DBMS들 간에 큰 성능차이를 보이며, 1,000개 이상일 때부터 그 차이가 매우 커짐을 볼 수 있다. 또한 오픈소스 DBMS들 간에는 튜플의 수가 100개로 작을 때에는 MySQL과 MariaDB 간 성능의 차이가 잘 나타나지 않으나, 튜플의 수가 1,000개 일 때부터는 MariaDB의 성능이 MySQL보다 우수하다는 것을 알 수 있다.

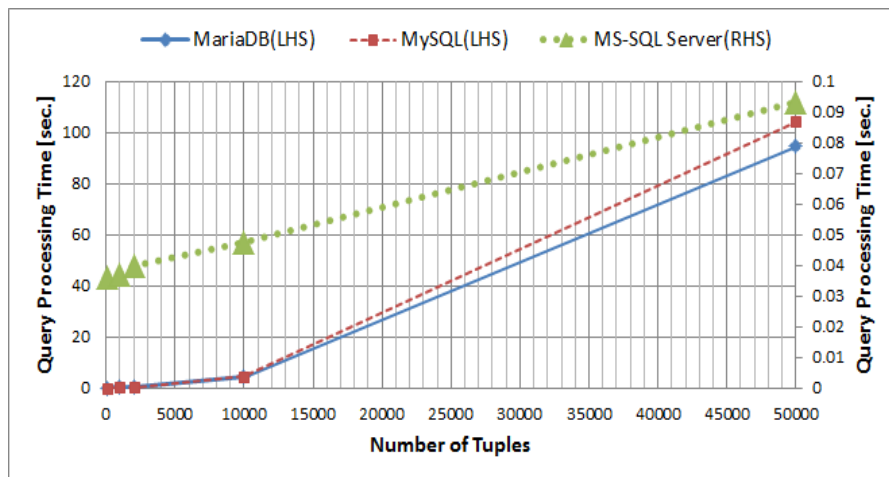


Figure 1. Graph of average processing time for Join query

3.3 성능품질 분석

DBMS의 성능품질은 조인 쿼리의 처리속도로 정의하였으므로, 3개 DBMS의 성능품질을 비교하기 위해서 각 DBMS 간에 튜플의 개수에 따른 처리시간 평균에 차이가 존재하는가를 분석하였다. 이를 위해 Ree(2013), Ryu(2014)에서와 같은 분산분석을 실시하였다. 분산분석 결과 모든 경우에 유의한 차이를 보였으며, 차이가 나는 그룹을 찾기 위해 Tukey의 HSD 검정을 실시하였다.

3.3.1 튜플의 개수가 100인 경우

튜플의 개수가 100인 두 릴레이션을 조인하는 경우 Table 1의 실험데이터의 분산분석을 실시한 결과는 Table 2와 같다. 분산분석 결과 유의확률이 <0.001 이므로 평균이 같다는 가설을 기각할 수 있다. 즉 평균이 동일하다는 가설을 99.999%의 확률로 기각할 수 있으므로 평균은 동일하지 않다고 말할 수 있다. 따라서 MariaDB, MySQL, MS-SQL Server 3개의 DBMS 간에는 처리시간의 평균들이 차이를 보이고 있다.

Table 2. ANOVA of DBMS response time at number of tuples = 100

Factor	Sum of Squares	DF	Mean Square	F ratio	P-value
Between Groups	0.003	2	0.002	15.037	0.000
Within Groups	0.022	207	0.000		
Total	0.025	209			

3개 DBMS의 처리시간에서 차이를 보이는 그룹을 찾기 위해 Tukey HSD 검정을 실시하였다. 결과는 Table 3과 같다. 튜플의 수가 100개인 경우에는 MariaDB와 MySQL은 서로 같은 그룹으로 유의한 차이가 없으며 MS-SQL Server와는 다른 성능품질을 가진 두 그룹으로 구분된다.

Table 3. Tukey HSD at number of tuples = 100

No	DBMS	Subset for alpha = 0.05	
		1	2
1	MariaDB		0.4370791
2	MySQL		0.4451280
3	MS-SQL Server	0.3594583	
	Sig.	1	0.887

3.3.2 튜플의 개수가 1,000인 경우

튜플의 개수가 1,000인 두 릴레이션을 조인하는 경우 Table 1의 실험데이터의 분산분석을 실시한 결과는 Table 4와 같다. 분산분석 결과 유의확률이 <0.001 이므로 평균이 같다는 가설을 기각할 수 있다. 즉 평균이 동일하다는 가설을 기각할 수 있으므로 평균은 동일하지 않다. 이 경우에서도 MariaDB, MySQL, MS-SQL Server 3개의 DBMS 간에는 처리시간에 있어서 차이를 보인다.

Table 4. ANOVA of DBMS response time at number of tuples = 1,000

Factor	Sum of Squares	DF	Mean Square	F ratio	P-value
Between Groups	0.152	2	0.076	438.972	0.000
Within Groups	0.036	207	0.000		
Total	0.188	209			

3개 DBMS의 성능품질에서 차이를 보이는 그룹을 찾기 위해 Tukey HSD 검정을 실시하였다. 결과는 Table 5와 같다. MariaDB, MySQL, MS-SQL Server는 서로 다른 성능품질을 가진 세 그룹으로 구분된다. 튜플이 100개인 경우와 달리 오픈소스인 MariaDB와 MySQL에서도 처리시간에서 차이가 나는 결과를 보였다. MySQL이 가장 느렸고, 다음이 MariaDB, MS-SQL server가 가장 빠른 처리시간을 보였다.

Table 5. Tukey HSD at number of tuples = 1,000

No	DBMS	Subset for alpha = 0.05		
		1	2	3
1	MariaDB	0.08917399		
2	MySQL		0.09814764	
3	MS-SQL Server			0.03714000
Sig.		1	1	1

3.3.3 튜플의 개수가 2,000인 경우

튜플의 개수가 2,000인 두 릴레이션을 조인하는 경우 Table 1의 실험데이터의 분산분석을 실시한 결과는 Table 6과 같다. 분산분석 결과 유의확률이 <0.001 이므로 평균이 같다는 가설을 기각할 수 있다. 즉 평균이 동일하다는 가설을 기각할 수 있으므로 평균은 동일하지 않다. 이 경우에서도 MariaDB, MySQL, MS-SQL Server 3개의 DBMS 간에는 처리시간에 있어서 차이를 보인다.

Table 6. ANOVA of DBMS response time at number of tuples = 2,000

Factor	Sum of Squares	DF	Mean Square	F ratio	P-value
Between Groups	1.532	2	0.766	4848.620	0.000
Within Groups	0.033	207	0.000		
Total	1.564	209			

3개 DBMS의 성능품질에서 차이를 보이는 그룹을 찾기 위해 Tukey HSD 검정을 실시하였다. 결과는 Table 7과 같다. MariaDB, MySQL, MS-SQL Server는 서로 다른 성능품질을 가진 세 그룹으로 구분된다. 튜플이 1,000 개인 경우와 마찬가지로 2,000개인 경우에도 3개의 DBMS 간 처리시간에서 차이가 나는 결과를 보였다. MySQL이 가장 느렸고, 다음이 MariaDB, MS-SQL server가 가장 빠른 처리시간을 보였다.

Table 7. Tukey HSD at number of tuples = 2,000

No	DBMS	Subset for alpha = 0.05		
		1	2	3
1	MariaDB	0.20908737		
2	MySQL		0.23095346	
3	MS-SQL Server			0.03985500
Sig.		1	1	1

3.3.4 튜플의 개수가 10,000인 경우

튜플의 개수가 1만 개인 두 릴레이션을 조인하는 경우 Table 1의 실험데이터의 분산분석을 실시한 결과는 Table 8과 같다. 분산분석 결과 유의확률이 <0.001 이므로 평균이 같다는 가설을 기각할 수 있다. 즉 평균이 동일하다는 가설을 기각할 수 있으므로 평균은 동일하지 않다. 이 경우에서도 MariaDB, MySQL, MS-SQL Server 3개의 DBMS 간에는 처리시간에 있어서 차이를 보인다.

Table 8. ANOVA of DBMS response time at number of tuples = 10,000

Factor	Sum of Squares	DF	Mean Square	F ratio	P-value
Between Groups	858.420	2	429.210	174295.3	0.000
Within Groups	0.510	207	0.002		
Total	858.930	209			

3개 DBMS의 성능품질에서 차이를 보이는 그룹을 찾기 위해 Tukey HSD 검정을 실시하였다. 결과는 Table 9와 같다. MariaDB, MySQL, MS-SQL Server는 서로 다른 성능품질을 가진 세 그룹으로 구분된다. 튜플이 1,000 개인 경우, 2,000개인 경우와 마찬가지로 3개의 DBMS 간 처리시간에서 차이가 나는 결과를 보였다. MySQL이 가장 느렸고, 다음이 MariaDB, MS-SQL server가 가장 빠른 처리시간을 보였다.

Table 9. Tukey HSD at number of tuples = 10,000

No	DBMS	Subset for alpha = 0.05		
		1	2	3
1	MariaDB	4.067886		
2	MySQL		4.562182	
3	MS-SQL Server			0.04754298
Sig.		1	1	1

3.3.5 튜플의 개수가 50,000인 경우

튜플의 개수가 5만 개인 두 릴레이션을 조인하는 경우 Table 1의 실험데이터의 분산분석을 실시한 결과는 Table 10과 같다. 분산분석 결과 유의확률이 <0.001 이므로 평균이 같다는 가설을 기각할 수 있다. 즉 평균이 동일하다는 가설을 기각할 수 있으므로 평균은 동일하지 않다. 이 경우에서도 MariaDB, MySQL, MS-SQL Server 3개의 DBMS 간에는 처리시간에 있어서 차이를 보인다.

Table 10. ANOVA of DBMS response time at number of tuples = 50,000

Factor	Sum of Squares	DF	Mean Square	F ratio	P-value
Between Groups	462447.2	2	231223.592	1948482	0.000
Within Groups	24.564	207	0.119		
Total	462471.7	209			

3개 DBMS의 성능품질에서 차이를 보이는 그룹을 찾기 위해 Tukey HSD 검정을 실시하였다. 결과는 Table 11과 같다. MariaDB, MySQL, MS-SQL Server는 서로 다른 성능품질을 가진 세 그룹으로 구분된다. 튜플이 5만 개인 경우에도 3개의 DBMS 간 처리시간에서 차이가 나는 결과를 보였다. MySQL이 가장 느렸고, 다음이 MariaDB, MS-SQL server가 가장 빠른 처리시간을 보였다.

Table 11. Tukey HSD at number of tuples = 50,000

No	DBMS	Subset for alpha = 0.05		
		1	2	3
1	MariaDB	94.40719		
2	MySQL		104.1560	
3	MS-SQL Server			0.09339764
Sig.		1	1	1

4. 결 론

본 논문에서 윈도우즈 환경 하에서 가장 널리 사용되는 오픈소스 DBMS인 MariaDB와 MySQL, 그리고 상용 DBMS인 MS-SQL Server의 시간성능을 실험에 의해 평가하였다. DBMS의 성능품질을 쿼리의 처리시간으로 정의하고, 릴레이션을 튜플의 수를 변화시키면서 두 릴레이션 간의 조인 쿼리의 처리시간을 측정하였다. 동일한 실험 환경을 위해 HeidiSQL을 프론트엔드로 사용하였다.

튜플을 100개에서 1,000개, 2,000개, 10,000개, 50,000개까지 증가시키면서 오픈소스 DBMS 간의 성능품질 비교 결과를 얻을 수 있었으며, 오픈소스 DBMS와 상용 DBMS 간의 성능품질 비교 결과도 얻을 수 있었다. 분산분석을 통해 쿼리의 처리시간 간에 차이가 존재하는가에 대한 검정을 실시하였다. 모든 실험에서 차이가 존재하였으며,

차이가 나는 그룹을 찾기 위해 Tukey HSD 검정을 실시하였다. 분석 결과 튜플의 개수에 관계없이 상용 DBMS인 MS-SQL Server의 성능이 가장 우수했다. 특히 튜플의 개수가 5만개로 주어진 스트레스 상황에서는 상용 DBMS와 오픈소스 DBMS간에 매우 큰 처리시간 차이를 보였다. 그러나 튜플의 수가 적어질수록 그 차이는 작아졌다. 오픈소스 DBMS인 MySQL과 MariaDB 간의 성능품질에 대해서는 튜플이 가장 적은 100개의 경우에만 두 DBMS간에 유의한 차이를 보이지 않았으나, 튜플이 1,000개 이상에서는 MariaDB가 MySQL 보다 평균 처리시간이 유의하게 빨라서 성능품질이 좋다는 사실을 밝혔다. 따라서 오픈소스 DBMS 중에서도 가장 많이 사용되는 MySQL 보다는 MariaDB를 사용하는 것이 더 바람직하다는 것을 알 수 있다.

본 연구결과는 오픈소스 DBMS와 상용 DBMS 간의 선택기준으로 활용될 수 있다. 앞으로 쿼리의 종류를 삽입, 삭제, 갱신의 경우로 확대하고 검색 쿼리에서도 조인 조건을 보다 다양화하면서, 동일한 실험을 수행 하는 것이 향후 과제이다.

REFERENCE

- Androutsellis-Theotokis, S., Spinellis, D., Kechagia, M., and Gousios, G. 2011. Open Source Software: A Survey from 10,000 Feet. Foundations and Trends in Technology Information and Operations Management and Publish.
- Choi, Seung-jun, Park, Jae-Won, Kim, Jong-Bae, and Choi, Jae-Hyun. 2014. "A Quality Evaluation Model for Distributed Processing Systems of Big Data." *Journal of Digital Contents Society* 15: 533-545.
- Chung, Su-min, Choi, Jae-hyun, and Park, Jae-won. 2016. "Design of Software Quality Evaluation Model for IoT." *Journal of the Korea Institute of Information and Communication Engineering* 20: 1342-1354.
- Date, C. J. 2004. *An Introduction to Database Systems* 8th Edition. Pearson.
- ISO/IEC TR 9126-1. 2001. *Software Engineering-Product Quality-Part 1: Quality Model*.
- ISO/IEC TR 9126-2. 2003. *Software Engineering-Product Quality-Part 2: Internal Metrics*.
- ISO/IEC TR 25010. 2011. *System and software engineering - Systems and software Quality Requirements and Evaluation*.
- Lee, Chung-Hyun, Choi, Jae-Hyun, and Park, Jea-Won. 2015. "A Practical Quality Evaluation Model of Mobile Cloud Software." *Journal of Korean Institute of Information Technology* 13: 95-110.
- Lim, Kwang-Hyeon, Ryu, Kwang-Sum, Shon, Ho-Sun, and Ryu, Keun-Ho. 2012. "The Study of quality measurement plan for software reliability." *Journal of the Korea Society of Computer and Information* 17: 187-198.
- Ree, Sangbok. 2013. "Study on the result changes with the size of the variance in Taguchi method and factor experimental." *Journal of Quality Management* 41: 119-134.
- Ryu, Gui-Yeol. 2014. "A Study on Quality Portals Based on Probability Distributions of Response Time." *Journal of Quality Management*: 42: 33-42.
- Samoladas, I., Gousios, G., Spinellis, D., and Stamelos, I. 2008. "The SQO-OSS quality model: measurement-based open source software evaluation." *Open source development, communities and quality*, 237-248.
- Spinellis, D. 2006. *Code Quality: The Open Source Perspective*. Addison Wesley Professional.
- Spinellis, D., Gousios G., Karakoidas, V., Louridas, P., and Adams, P. 2009. "Evaluating the Quality of Open Source Software." *Electronic Notes in Theoretical Computer Science* 233: 5-28.