

# 빅데이터 환경에서 MongoDB와 MySQL의 CRUD 연산의 성능 연구 설계

서정연\*, 진은광\*, 채민수\*, 이화민\*\*

\*순천향대학교 컴퓨터학과

\*\*순천향대학교 컴퓨터소프트웨어공학과

e-mail : busyppp@gmail.com, imdae11@naver.com, cmspr0@nate.com,  
leehm@sch.ac.kr

## Performance study design of CRUD operation of MongoDB and MySQL in big data environment 1)

Jung-Yeon Seo\*, Eun-Kwang Jeon\*, Min-su Chae\*, Hwa-Min Lee\*\*

\*Dept of Computer Science, Soonchunhyang University

\*\*Dept of Computer Software Engineering, Soonchunhyang University

### 요 약

최근 들어 모바일 디바이스의 발전으로 인해 생성되는 데이터의 종류는 다양해지고, 양은 방대해지고 있다. 이렇게 생성된 방대한 양의 데이터를 빅데이터라고 한다. 빅데이터들은 기존의 데이터 처리 방법과 다른 방법으로 처리되어야 한다. 빅데이터 처리의 대표적인 방법인 관계형데이터베이스시스템(RDBMS)과 NoSQL 방법 중 대표적인 방법인 MySQL과 MongoDB의 데이터를 모델링한다. 설계된 데이터를 바탕으로 보다 편하고 알맞게 데이터베이스시스템 성능평가를 수행한다.

### 1. 서론

스마트 디바이스와 IoT가 발전됨으로써 그에 따른 데이터들이 증가하고 있다. 이로 인해 생성되는 데이터들의 양은 무수히 많고 종류도 다양하게 생성되고 있다. 이러한 데이터들을 빅데이터라고 한다. 빅데이터 처리를 하기 위해서는 빅데이터 처리 방법이 필요하다. 빅데이터 처리 방법은 대표적으로 관계형데이터베이스시스템(RDBMS)과 NoSQL이 있다. 두 가지의 방식은 빅데이터를 처리 할 수 있다는 공통점이 있지만, 각각의 속성이 달라서 쓰이는 상황이 다르다.

이에 본 논문에서는 RDBMS와 NoSQL의 빅데이터 분석, 처리 성능을 확인하는 것에 주목한다. 각각의 환경에서 빅데이터 분석, 처리 실험환경을 설계한다. RDBMS는 MySQL을 사용하고 NoSQL은 MongoDB를 사용한다. 빅데이터 설계 과정을 통해 상황과 특징에 맞게 설계하는 법을 자세하게 연구한다.

### 2. 관련 연구

빅데이터에 대한 관심이 증가하면서 처리 방법에 대한 연구가 증가하고 있다. 특히 RDBMS와 NoSQL에 대한 비교 연구가 가장 활발하게 이루어진다. 빅데이터 환경에서

PostgreSQL과 MongoDB의 데이터입출력 성능개선에 관한 연구[1]에서는 PostgreSQL과 MongoDB를 대상으로 데이터의 입력, 조회, 갱신, 삭제 연산을 통해 성능 비교를 진행하였다. 그 결과 전체적으로 MongoDB의 연산 속도가 더 빠르다는 것을 확인하였다. 성능 개선 관점에서 MongoDB는 관계형 데이터 모델로 설계하는 것 보다 비정형데이터 모델로 설계하는 것이 더 나은 성능을 나타내는 것을 확인 할 수 있다.

MongoDB와 MySQL 대용량 데이터 처리 비교를 통한 nosql 활용 방안 연구[2]에서는 MongoDB와 MySQL의 대용량 데이터 처리 성능 측정을 비교하였다. 그 결과 대용량 데이터의 성능 측정 결과에서 MongoDB의 결과가 약 80% 이상 성능이 좋은 것으로 비교되었다. 대규모의 분산 환경에서는 결과치가 더욱 더 벌어질 것으로 예상된다. 하지만 아직 자료와 경험의 축적이 충분하지 않은 MongoDB 대신 데이터의 정합성, 데이터의 무결성 위주의 시스템을 운영한다면 MySQL을 사용하는 것이 바람직하다는 결과를 도출했다.

그 외에 소규모 마리아 운영 시스템에서의 MongoDB 적용 성능 평가[3], 구조를 개선한 MariaDB와 MongoDB의 성능 분석에 대한 연구[4], YCSB를 사용한 PostgreSQL과 MongoDB 성능 비교 분석[5] 등에 따르면 MongoDB는 RDBMS보다 삽입, 조회, 수정, 삭제 연산에서 성능이 우수하다는 결과가 나왔다. 시스템의 구조에 따라 필요한 데이터베이스 시스템이 다르다. 안정적인 시스템에서는 RDBMS가 요구되고, 속도가 중요한 시스템에서는

1) 이 논문은 2016년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2014R1A1A2057878).

MongoDB가 요구된다.

다음 [표 1]은 MySQL과 MongoDB의 특징을 비교한 표이다.[6]

[표 1] MySQL과 MongoDB의 특징 비교

특징	MySQL	MongoDB
풍부한 데이터 모델	X	O
동적 스키마	X	O
정형 데이터	O	O
데이터 지역성	X	O
필드 업데이트	O	O
쉬운 프로그래밍	X	O
복잡한 트랜잭션	O	X
감사 기능	O	O
자동 샤딩	X	O

### 3. 구현 결과

MongoDB와 MySQL의 성능을 비교하기 위해 다음과 같은 시스템 환경으로 구성 진행하였다. 시스템 환경을 구축하기 위한 컴퓨팅 디바이스는 가상 서버를 활용하여 진행하였다. 실제적으로 성능을 비교할 MySQL은 5.7버전, MongoDB는 3.2버전을 사용하였다. 가상 서버의 시스템 성능은 [표 2]와 같다.

[표 2] 가상 서버 시스템 성능

시스템	성능
CPU	Intel Core I5-35300 CPU @ 3.3.Hz
OS	Linux 3.10.0 (64bit)
RAM	4G Byte

RDBMS인 MySQL과 NoSQL중 하나인 MongoDB의 성능을 비교한다. MongoDB는 MongoDB의 쿼리를 보다 효과적으로 수행할 수 있게 해주는 index를 추가해서 한 번 더 비교한다. 성능 비교는 CRUD연산, 즉 삽입, 조회, 갱신, 삭제 연산을 실행 한 뒤 연산 시간을 비교한다. 이러한 성능 연산을 비교하기 위해 빅데이터 실험환경을 설계한다. MySQL과 MongoDB의 성능평가 실험을 위해 정형 데이터 모델을 동일한 모델로 적용한다.

데이터는 ASA(American Standards Association: 미국 규격협회)에서 제공하는 1987년부터 2008년까지의 미국 항공편 운항 통계 데이터 중 2008년의 데이터로 실험을 실시한다. 이는 미국 내 모든 상업 항공편에 대한 항공편 도착과

출발 세부 사항에 대한 정보를 제공한다. 2008년의 항공 데이터는 약 7만개가 존재한다.

미국 항공 데이터 테이블	
동계 번호	
연도	
월	
일	
요일	
실제 출발 시각	
예정 출발 시각	
실제 도착 시각	
예정 도착 시각	
항공사 코드	
항공편 번호	
항공기 등록 번호	
실제 경과 시간	
예정 경과 시간	
방송시간	
도착 지연 시간	
출발 지연 시간	
출발지 공항 코드	
도착지 공항 코드	
비행 거리	
비행기 바퀴가 지면에 닿아서 목적지 공항의 게이트에 도착할 때까지의 시간	
출발지공항의 게이트에서 출발해서 바퀴가 지면에서 떨어질 때까지의 시간	
비행 취소 여부	
비행 취소 코드	
우회 여부	
항공사 지연 시간	
기상 지연 시간	
NAS 지연 시간	
보안 지연 시간	
연착 항공기 지연 시간	

[그림 1] 데이터 논리 설계 모델

[그림 1]은 데이터 모델 논리 설계를 나타낸 것이고, [그림 2]는 데이터 모델 물리 설계를 나타낸 그림이다.

USA Flight Data Table	
ArrNum	VARCHAR(5)
Year	VARCHAR(5)
Month	VARCHAR(3)
DayofMonth	VARCHAR(3)
DayofWeek	VARCHAR(3)
DepTime	DATETIME
CRSDepTime	DATETIME
ArrTime	DATETIME
CRSArrTime	DATETIME
UniqueCarrier	VARCHAR(10)
FlightNum	VARCHAR(10)
TailNum	VARCHAR(10)
ActualElapsedTime	VARCHAR(5)
CRSElapsedTime	VARCHAR(5)
AirTime	VARCHAR(5)
ArrDelay	VARCHAR(5)
DepDelay	VARCHAR(5)
Origin	VARCHAR(10)
Dest	VARCHAR(10)
Distance	VARCHAR(5)
TaxiIn	VARCHAR(5)
TaxiOut	VARCHAR(5)
Cancelled	VARCHAR(5)
CancellationCode	VARCHAR(5)
Diverted	VARCHAR(5)
CarrierDelay	VARCHAR(5)
WeatherDelay	VARCHAR(5)
NASDelay	VARCHAR(5)
SecurityDelay	VARCHAR(5)
LateAircraftDelay	VARCHAR(5)

[그림 2] 데이터 논리 설계 모델

설계한 데이터 모델을 바탕으로 MySQL과 MongoDB의 성능연산을 실험한다. 모두 같은 데이터를 바탕으로 5,000건, 10,000건, 30,000건, 50,000건, 70,000건의 데이터를 순차적으로 삽입, 조회, 갱신, 삭제 연산 시에 걸리는 시간을 측정한다.

#### 4. 결론

설계한 데이터베이스의 내용을 바탕으로 MongoDB와 MySQL의 성능연산을 실험한다. 실험결과를 바탕으로 특

정 상황에 맞는 데이터베이스 시스템을 설정하고 사용한다. 이러한 데이터를 사용하면 데이터의 품질이 올라간다. 즉, 데이터의 활용가치가 높아진다. 또한 복잡한 정보 요구 사항에서 간결하게 표현이 가능하며, 데이터 구조변경을 막아주어 시스템 구축 작업에서 중요하게 차지한다.

향후 연구과제로는 정형 데이터 뿐만 아니라 비정형 데이터의 설계를 통해 여러 형식의 데이터를 처리, 분석할 수 있는 환경이 제공되어야 한다. 비정형 데이터 모델의 연산을 실행하고, NoSQL의 분산 가능성에 대한 연구를 추가하고 안전성을 보완시켜서 빅데이터 환경의 분석기술의 발전을 기여할 수 있을 것이다.

#### 5. 참고문헌

- [1]정민규, “빅데이터 환경에서 PostgreSQL과 MongoDB의 데이터 입출력 성능개선에 관한 연구”, 숭실대학교 정보과학대학원, 숭실대학교 도서관
- [2]김은기, “MongoDB와 MySQL 대용량 데이터 처리 비교를 통한 Nosql 활용 방안 연구”, 숭실대학교 대학원, 숭실대학교 도서관
- [3]성흥기, “소규모 마리아 운영 시스템에서의 MongoDB 적용 성능 평가”, 동명대학교 대학원, 동명대학교 도서관
- [4]차승훈, “구조를 개선한 MariaDB와 MongoDB의 성능 분석에 대한 연구”, 숭실대학교 정보과학대학원, 숭실대학교 도서관
- [5]김기성, “YCSB를 사용한 PostgreSQL과 MongoDB 성능 비교 분석”, 정보과학회논문지 제 43권 제 12호, 2016.12, pp.1385-1395
- [6]Cornelia Györödi, Robert Györödi, George Pecherle, Andrada Olah “ comparative study: MongoDB vs. MySQL”, Engineering of Modern Electric Systems (EMES 2015), 2015.7