



빅데이터 처리 및 저장관리 기술동향 및 전망

이미영* · 최완*

*한국전자통신연구원 클라우드컴퓨팅연구부

목 차

| | |
|---------------|------------------|
| I. 서론 | IV. 빅데이터 처리 기술 |
| II. 빅데이터 활용 | V. 빅데이터 저장 관리 기술 |
| III. 빅데이터 플랫폼 | VI. 결론 |

I. 서론

모바일 환경, 센싱 장치 발달은 데이터의 생산, 유통, 소비 체계에 큰 변화를 주고 있다. 데이터가 경제적 자산이 되는 빅데이터 시대를 맞이하고 있으며, 빅데이터의 활용 및 이를 위한 새로운 IT 기술에 대한 요구가 커지고 있다. 빅데이터는 기업의 마케팅, 전략 수립 등 경제 활동뿐만 아니라 사회, 과학 연구 등 전 분야의 활동에 영향을 미칠 것으로 예측하고 있다.

빅데이터 조사 분석에 의하면, 2020년 전 세계 디지털 데이터량은 2009년 대비 44배 증가한 35제타바이트에 달할 것으로 예측하고 있으며, 생성 데이터의 90% 이상이 텍스트, 영상, 음성, 동영상 등 비정형 데이터이며, 센서 데이터, 트윗 등 실시간으로 발생하는 스트림 데이터가 폭증하고 있다.

데이터의 폭증, 비정형 데이터의 급증 및 빠른 데이터 생성은 기존의 전형적인 데이터 처리, 저장 관리 및 분석 기술에 큰 변화를 요구하고 있다. 데이터를 중심으로 프로그래밍할 수 있는 환경 제공과 데이터 규모에 대한 확장성을 제공하기 위한 분산 컴퓨팅 기술, 고속의 실시간 처리를 위한 인메모리 컴퓨팅 기술들이 빅데이터 시대의 주요 기반 기술로 대두되고 있다.

본 고에서는 빅데이터 시대에 대처하기 위한 빅데이터 처리 및 저장 관리 기술 동향에 대해 소개한다. 2장에서 빅데이터 속성 및 활용 방법의 변화에 대해 알아보고, 3장에서는 빅데이터 시대가 빅데이터 플랫폼에 미치는 영향에 대해 알아보고, 4장에서는 빅데이터 처리 기술 동향, 5

장에서는 빅데이터 저장 관리 기술 동향에 대해 기술하고, 결론을 맺는다.

II. 빅데이터 활용

빅데이터는 기존의 IT 방식으로는 처리/저장/관리/분석이 어려울 정도로 데이터의 생성 속도가 빠르고, 다양한 종류로 구성된 큰 규모의 데이터를 의미한다. IBM 정의에 따르면 빅데이터의 속성은 크게 데이터 규모, 데이터 다양성, 데이터 생성/처리 속도로 규정할 수 있다.

● 데이터 규모(Volume)

트위터의 1일 생성 데이터가 약 7 테라바이트, 페이스북은 약 10 테라바이트 등 대규모 데이터가 매일 생성되고 있어, 2020년에는 전 세계 데이터량이 35제타바이트에 달할 것으로 예측하고 있다[1].

● 데이터 다양성(Variety)

정형 데이터뿐만 아니라 텍스트, 영상, 음성, 동영상 등 비정형 데이터가 증가하고 있으며, 전체 데이터량의 90% 이상이 비정형 데이터일 것으로 예측하고 있다.

● 데이터 생성 및 처리 속도(Velocity)

자동으로 데이터를 생성하는 센싱 장치의 확산으로 데이터 생성 속도가 빨라지고 있으며, 지속적으로 생성되는 RFID, 음성 패킷 등 작은 단위의 스트림 데이터가 전체 생성 데이터 건수의 99% 이상이라고 한다 [1]. 빠른 데이터 생성 속도는 이를 처리, 활용하는 속

도에도 큰 영향을 미치고 있다.

빅데이터 시대가 되면서 정보를 분석하여 활용하는 방법에 있어서도 사전에 예측 및 대응하는 방향으로 발전하고 있다.

기존에는 이미 발생한 현상에 대한 정보를 저장 관리하고, 이로부터 왜 발생했는지 이유를 확인하는 방식이었지만, 지금은 현재 어떤 일이 발생하고 있으며, 왜 발생하는지를 실시간으로 파악, 활용하고자 하는 요구가 커지고 있다. 앞으로는 미래에 발생할 일에 대해 미리 예측하고 대응 방안을 마련하기 위한 분석 방법으로 발전할 것이다[2].

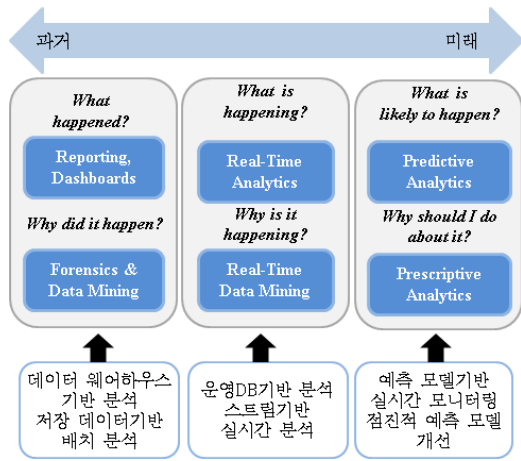


그림 1. 분석 서비스 발전 방향

분석 서비스의 변화는 이를 지원하는 데이터 처리 및 저장 관리 기술도 대규모 데이터 및 실시간 발생 데이터에 대한 분석을 지원하도록 요구하고 있다.

- 과거 상황 분석: 데이터 웨어하우스나 대규모 파일에 데이터를 저장 관리하고 이를 기반으로 대규모 데이터 배치 분석 처리
- 현재 상황 분석: 데이터가 동적으로 변화하는 운영 DB기반 온라인 분석 및 실시간 발생 스트림기반 분석 처리
- 미래 예측 대응: 예측 모델을 기반으로 발생하는 데이터를 실시간으로 모니터링하고, 최신 데이터를 활용하여 예측 모델을 지속적으로 검증 및 개선할 수 있는 환경 지원

III. 빅데이터 플랫폼

빅데이터 3V 속성과 빅데이터 활용 방법의 변화는 데이터를 처리 및 저장 관리하는 빅데이터 플랫폼 기술에 다음과 같은 영향을 준다.

데이터 규모 증가에 대처하기 위한 확장성을 위해서 분산 컴퓨팅 기술 적용이 필수적이다. 범용 서버로 구성된 클러스터 시스템을 기반으로 데이터 증가시 노드 증설을 통해 대처가 가능한 분산 데이터 처리 및 저장 관리 기술이 필요하다

데이터 다양성은 정형 데이터뿐만 아니라 비정형 데이터 처리를 쉽게 구축할 수 있는 환경을 제공하여야 한다. 사용자가 자유롭게 데이터 처리 로직을 작성하여 쉽게 통합하여 활용할 수 있는 유연한 데이터 중심의 프로그래밍 환경이 필요하다. 또한 다양한 데이터를 수집 저장시 일반적으로 데이터 구조가 미리 정의되지 않은 경우가 많이 발생하므로 기존의 관계형 데이터베이스 시스템 외에도 다양한 데이터 관리 시스템, 즉 noSQL 기술이 필요하다.

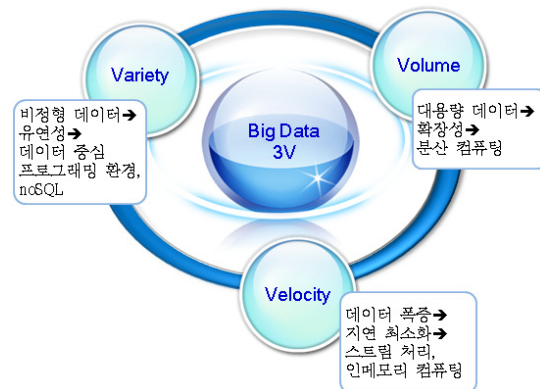


그림 2. 빅데이터 특성과 주요 기술

데이터 생성 및 처리 속도는 최신 데이터의 적시 활용이 가능하도록 데이터 처리 지연을 최소화하기 위한 스트림 처리 및 인메모리 컴퓨팅 기술이 중요하다.

스트림 처리는 데이터를 접수하는 즉시 처리하는 기술로, 기존에 데이터를 먼저 저장하고, 저장되어 있는 데이터를 기반으로 처리하는 방식에 비해 데이터 처리 지연이 줄어든다.

응용에서 사용하는 데이터의 주 저장소로 메모리를

활용하는 인메모리 컴퓨팅 기술은 64bit 범용 서버의 확산, DRAM, NAND Flash 가격 하락 및 실시간 서비스 요구 확산에 의해 활용이 확대되고 있다. 가트너 조사에 의하면 2011년 GB당 DRAM은 약 10달러, NAND Flash는 1달러가 소요되지만 2015년에는 DRAM 2달러, NAND Flash는 0.25 달러가 될 것으로 예측하고 있다[3]. 또한 엔터프라이즈 서버 시장에서 인메모리 컴퓨팅 채택이 2011년 7%에서 2020년에는 65%로 늘어날 것으로 예측하고 있다[4].

IV. 빅데이터 처리 기술

빅데이터에서 유용한 정보 및 숨어 있는 지식을 찾아내기 위한 데이터 가공 및 분석 과정을 지원하는 빅데이터 처리 기술은 대규모 데이터 처리를 위한 확장성, 데이터 생성 및 처리 속도를 해결하기 위한 처리 시간 단축 및 실시간 처리 지원, 비정형 데이터 처리 지원 등이 가능하여야 한다.

데이터가 늘어나도 허용되는 시간내에 처리가 가능한 대규모 데이터 처리 확장성을 위해 분산 컴퓨팅 기술이 활용되고 있으며, 다양한 데이터 처리 지원을 위해 사용자가 정의하는 처리 로직의 유연한 통합을 제공하는 데이터 중심의 프로그래밍 모델이 중요해 지고 있다.

빅데이터 처리 기술은 저장되어 있는 대량의 데이터를 일괄 처리하는 분산 데이터 처리 기술이 제안되어 널리 활용되고 있으며, 반복적으로 처리하는 업무를 효과적으로 지원하기 위해 업무 처리 환경을 설정하고 나면 반복적으로 사용할 수 있는 빅데이터 반복 처리 기술이 제안되고 있다. 최근에는 지속적으로 발생하는 데이터를 실시간으로 연속 처리하는 스트림 처리 기술로 발전하고 있으며[5], 각 기술에 대한 주요 사례 시스템은 그림 3과 같다.

| | 빅데이터 배치 처리 | 빅데이터 반복 처리 | 폭증 스트림 처리 |
|--------|--|---|--|
| 사례 시스템 | <ul style="list-style-type: none"> Google MapReduce Hadoop MapReduce MS Dryad | <ul style="list-style-type: none"> Twister HaLoop | <ul style="list-style-type: none"> Yahoo! S4 Twitter STORM HStreaming |

그림 3. 빅데이터 처리 기술 분류

4.1. 빅데이터 배치 처리 기술

빅데이터 배치 처리 기술은 대량의 데이터를 분할하여 각 데이터 파티션을 동시에 분산 처리함으로써 처리 시간을 단축하는 기술이다.

대표적인 빅데이터 배치 처리 기술인 구글의 MapReduce 기술은[6] 오픈 소스 Hadoop[7]의 성공으로 분산 데이터 처리 기술의 사실 표준이 되고 있다. MapReduce 기술은 Map과 Reduce 태스크로 서비스를 구성하며, key-value 기반의 입력 데이터에 대해 Map 태스크를 처리하고, 그 결과에 대해 Reduce 태스크를 수행하는 모델이다. MapReduce 모델을 지원하는 프레임워크는 입력 데이터 집합을 분할하여 여러 노드에 분배하고, 각각의 분할된 데이터 집합에 대해 Map 태스크를 동시에 분산 처리하며, 그 결과를 통합 및 분할하여 Reduce 태스크에 전달, 이를 기반으로 Reduce 태스크가 동시에 분산 처리할 수 있도록 지원한다. 즉, 입력 데이터 분할 처리 및 처리 결과 통합, Job 스케줄링, 태스크 분산 배치 및 실행 환경 제공, 장애 대처를 위한 태스크 재수행 기술 등이 통합된 분산 데이터 처리 기술이다.

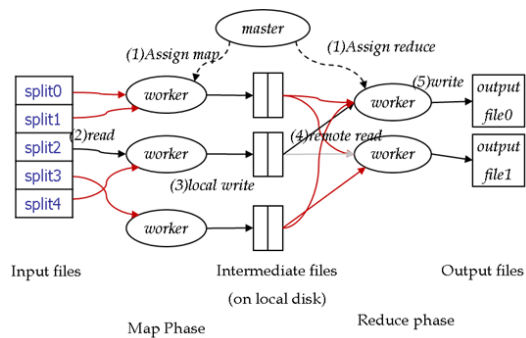


그림 4. MapReduce 분산 데이터 처리

Map과 Reduce의 2 단계로 서비스를 표현하는 단순한 MapReduce 모델외에도 태스크간의 복잡한 관계를 표현할 수 있는 DAG 기반의 프로그래밍 모델을 제공하는 빅데이터 배치 처리 시스템으로 Microsoft의 Dryad[8]가 있다. Dryad는 데이터 처리 로직을 표현하는 Vertex와 입출력 데이터나 Vertex 간의 데이터 전송 채널을 표현하는 Edge로 구성된 그래프 기반으로 서비스를 정의하며, 다중 코어기반 단일 컴퓨터부터 클러스터 시스템기반 수행을 지원한다.

4.2. 빅데이터 반복 처리 기술

데이터를 가공, 분석하는 업무는 많은 경우에 서로 다른 데이터 집합에 대해 동일한 로직을 적용하여 처리한다. 이와 같이 데이터만 달리하여 반복적으로 처리하는 업무를 위해 실행 환경을 설정하고 나면 반복적으로 사용할 수 있도록 확장함으로써, 프로세스 재구동 등 수행 환경 설정에 소요되는 시간 및 자원 낭비를 막을 수 있다. 또한 데이터가 변경되어도 변화하지 않는 데이터를 재활용할 수 있도록 지원함으로써 데이터 처리 성능 개선이 가능하다. 이와 같은 반복 데이터 처리 환경을 제공하는 MapReduce기반의 데이터 반복 처리 기술에 대한 주요 연구로 Twister[9], HaLoop[10] 등이 있다.

4.3. 폭증 스트림 처리 기술

주식 거래 데이터, 센서 데이터처럼 지속적으로 발생하는 정형 데이터 스트림에 대한 실시간 처리를 위해, 관계형 모델, XML 모델 등 처리 대상이 되는 데이터 형식에 맞는 데이터 모델에 따라 처리 연산을 제공하는 CEP(Complex Event Processing) 기술[11]이 주로 활용되고 있었으나, 트윗 단문 메시지, CCTV 영상 데이터 등 비정형 스트림 데이터 처리를 위해 처리 로직을 자유롭게 정의하여 사용 가능한 분산 스트림 처리 기술이 나오기 시작했다.

스트림 처리 기술은 입력되는 데이터 스트림에 대해 사용자가 정의한 처리 로직을 적용, 결과 스트림을 생성하고, 이 결과 스트림이 다른 처리 로직의 입력 스트림으로 제공되거나, 외부 응용으로 제공되며, 데이터 입력에 의해 지속적인 처리를 제공하는 기술이다.

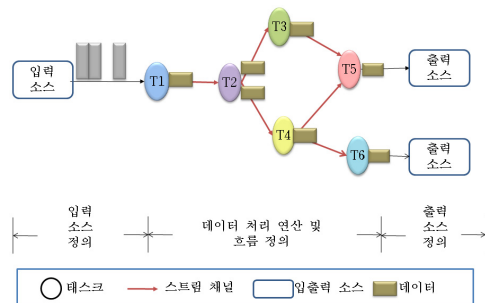


그림 5. 폭증 스트림 처리 기술 개념

분산 스트림 처리 기술은 처리 지연을 방지하기 위해 메모리 기반 처리를 하고 있으며, 서비스를 구성하

는 태스크들이 동시에 처리되도록 태스크 병렬 처리를 제공하며, 폭증 스트림 처리를 위하여 태스크를 복제 구동하여, 입력 스트림을 복제 태스크들에 분배함으로써 데이터 병렬 처리를 제공한다.

대표적인 오픈 소스 시스템으로 트윗 실시간 처리를 위해 만들어진 트위터의 STORM[12], 초당 수 천 건의 질의 요청을 받는 웹 검색 서비스에서 사용자의 검색 요구를 실시간으로 분석하는 환경 구축을 위해 만들어진 Yahoo!의 S4(Simple Scalable Streaming System)[13] 등이 있다. 또한 HStreaming[14]에서는 Hadoop기반으로 배치 처리 및 실시간 처리를 모두 지원하는 시스템을 내고 있다.

앞으로 스트림 처리 기술의 활용이 확대되면 더 큰 메모리에 대한 요구 및 메모리 소비 전력 절감에 대한 필요성이 커질 것으로 예상되므로, 요구되는 처리 속도 및 처리 로직 등을 고려하여 DRAM뿐만 아니라 차세대 비휘발성 메모리 활용에 대한 고려가 필요하다.

V. 빅데이터 저장 관리 기술

빅데이터가 온라인 트랜잭션 데이터 및 온라인 분석 데이터 저장 관리를 제공하는 데이터베이스 기술에 미치는 영향을 보면 확장성을 제공하기 위한 분산 컴퓨팅 기술, 데이터 발생 속도에 대응하여 고성능으로 저장 관리를 제공하는 인메모리 컴퓨팅 기술, 다양한 데이터 및 응용 환경에 맞게 새롭게 등장하는 noSQL 기술들이 융합되어 발전하고 있다[15]. 데이터 저장 관리 기술에 따른 주요 사례 시스템은 그림 6과 같다.

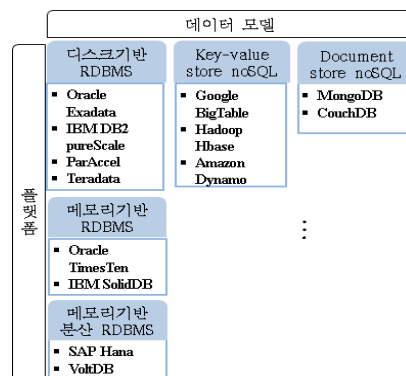


그림 6. 빅데이터 저장 관리 기술 분류

5.1. 디스크기반 RDBMS 기술

디스크기반 RDBMS는 ParAccel, Teradata처럼 대량의 데이터 분석을 위해 shared nothing 클러스터 시스템을 기반으로 확장성을 제공하는 시스템과 DB2 pureScale, Oracle Exadata처럼 공유 디스크를 제공하는 클러스터 시스템을 기반으로 온라인 트랜잭션 처리와 온라인 분석을 동시에 지원하면서 확장성을 제공하는 시스템으로 발전하고 있다[15].

디스크기반 RDBMS는 대량의 데이터 저장 관리 기능을 제공하지만 디스크 접근 속도의 한계로 실시간성을 제공할 수 없다. 데이터베이스 서비스 성능을 향상시키기 위해 대량의 메모리를 버퍼로 활용하고 있으며, 최근에는 Flash SSD를 추가로 버퍼로 활용, 성능을 향상시키고 있다.

5.2. 메모리기반 RDBMS 기술

통신이나 금융 분야에서 고성능 데이터 저장 관리를 위해 TimesTen이나 SolidDB와 같은 메모리 상주 RDBMS를 활용해 왔다. 메모리기반 RDBMS 기술은 데이터 용량에 대한 확장성을 제공하기 위해 분산 컴퓨팅 기술을 적용한 메모리기반 분산 RDBMS로 발전하고 있고, 대표적인 시스템으로 SAP의 Hana[16]와 VoltDB[17]가 있다. SAP Hana는 온라인 트랜잭션 처리와 온라인 분석을 동시에 지원하기 위하여 행기반 저장과 열기반 저장을 제공하는 분산 DBMS이다. VoltDB는 shared nothing 구조로 대규모 온라인 트랜잭션 처리를 고속으로 지원하기 위해 데이터베이스를 파티션하여 각 파티션을 병렬로 관리할 수 있게 제공한다.

메모리상주 DBMS는 DRAM의 휘발성 특성 때문에 데이터의 durability를 위해 주기적으로 디스크에 저장 및 로깅 장치로 디스크를 활용하므로 트랜잭션 처리 성능 향상에 제약을 받는다. 엔터프라이즈 서버에 NAND Flash 활용이 확산되면서 최근에는 Flash SSD를 로깅 장치로 활용하여 데이터베이스 시스템의 성능 향상을 얻고 있다[16].

NAND Flash기반 저장 장치는 접근 성능 향상을 위해 PCI 카드 인터페이스를 제공하고 있으며, 최근에는 한 단계 더 발전하여 기존 스토리지 프로토콜 사용 오버헤드까지 제거하여 고속 메모리처럼 작동하는 제품

이 나오고 있다[18].

NAND Flash 외에도 PRAM, MRAM 등 바이트 단위의 접근이 가능한 비휘발성 메모리는 표 1 처럼[19] DRAM에 근접하는 빠른 접근 속도와 비휘발성이라는 특성 때문에 로깅 장치로 활용, 버퍼로 활용 및 데이터 저장소로 활용 등 데이터베이스 시스템에서 활용이 기대된다.

BPRAM(Byte addressable Persistent RAM) 기반의 데이터베이스 기술은 비휘발성이라는 특성과 쓰기 회수가 제한되어 있다는 특성을 고려하여야 하며, 현재 글로벌 DBMS 업체에서도 이에 대한 연구를 수행하고 있으므로[20, 21], 앞으로 몇 년안에 비휘발성 메모리 기반의 데이터 저장 관리 기술 활용이 가시화될 것으로 보인다.

표 1 메모리 특성 비교

| 분류 | DRAM | Flash | PRAM | MRAM |
|-------|-----------|----------|-----------|-----------|
| 쓰기 시간 | 50ns | 10,000ns | 100 ns~ | 10-50ns |
| 일기 시간 | 50ns | 50ns | 20-80ns | 10-50ns |
| 쓰기 회수 | 10^{15} | 10^6 | 10^{12} | 10^{16} |
| 소비 전력 | 300mW | 30mW | ~30μW | ~30μW |
| 대기 전력 | ~100μA | ~1μA | ~1μA | ~1μA |

5.3. noSQL 기술

noSQL 기술은 구글의 Bigtable[22] 기술이 나오면서 확산되기 시작한 기술이다. noSQL 기술은 기존 RDBMS가 풍부한 데이터 모델을 기반으로 다양한 응용에서 요구하는 기능을 지원하는 반면, 비용이 높고 확장성이 떨어지는 문제점을 해결하기 위해, 응용 요구에 맞는 최적화된 다양한 데이터 관리 기술이 필요함을 인지하면서 관심을 받기 시작한 기술이다. 또한 사전에 스키마 구조를 정의할 수 없는 데이터에 대한 저장 관리 요구가 늘어나면서 RDBMS 이외의 다른 데이터 저장 관리 기술이 요구되면서 관심을 받고 있다.

대표적인 noSQL로는 대규모 인터넷 서비스를 위해 key-value 모델 기반으로 분산 저장 관리를 제공하는 구글의 Bigtable, Bigtable과 유사한 개념으로 만들어진

오픈 소스인 Hadoop의 Hbase[23], 아마존의 Dynamo [24] 등이 있다.

대표적인 noSQL인 key-value 모델 기반의 저장 관리를 제공하는 Hbase는 Master/Slave 구조의 분산 데이터 관리 시스템이다. 테이블의 행 키 기반으로 파티션을 논리적으로 분할하고, 각 파티션은 별도의 파일로 저장 관리되며, 파티션별로 서비스 노드를 지정함으로써 분산 데이터 관리를 제공한다. 또한 컬럼 기반 저장 모델을 제공, 컬럼 그룹별로 별도의 파일에 저장함으로써 특정 컬럼 기반의 전체 데이터 분석 작업이 빠른 접근을 제공한다. Hbase는 Hadoop의 분산 파일 시스템인 HDFS에 데이터를 저장함으로써 데이터 파일이 3중 복제 관리되므로 데이터가 손실되는 것을 방지한다. 또한 특정 파티션을 담당하던 노드에 장애가 발생하면 해당 파티션 서비스 노드를 재지정함으로써 서비스 노드 장애에 대처한다.

key-value 기반 데이터 저장 관리 기술외에도 대규모 문서 관리를 위한 MongoDB, CouchDB 기술이 개발되어 활용되고 있으며, 그래프 데이터 관리 기술도 연구되고 있다.

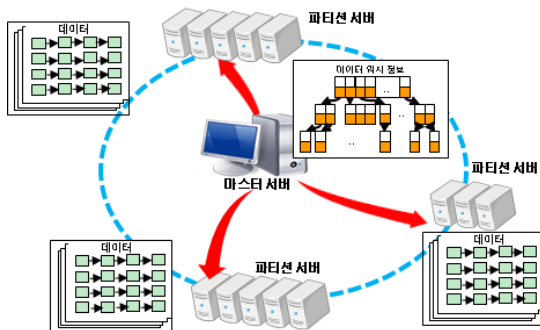


그림 7. Hbase 분산 데이터 관리

VI. 결론

빅데이터 활용은 개인, 기업, 정부 등 스마트한 사회를 이루기 위한 필수적인 요건이 되고 있다. 빅데이터를 얼마나 잘 다룰 능력이 있느냐에 국가의 경쟁력이 좌우되는 시대를 맞이하여, 이를 지원하는 빅데이터 플랫폼 기술 확보가 필수적이다.

이에 따라 빅데이터의 속성 및 활용 방식의 변화 방향에 대해 알아보고, 빅데이터 분석을 지원하는 기반 기술인 빅데이터 처리 및 저장 관리 기술의 연구 동향에 대해 알아보았다. 데이터 중심의 프로그래밍 환경 제공과 분산 컴퓨팅 기술, 인메모리 컴퓨팅 기술이 빅데이터 문제를 해결하는데 있어 주요 기반 기술이다.

빅데이터 3V 속성의 정도가 가속화될수록 현재의 IT 기술이 비용대비 효과나 성능 제약 문제가 대두될 것으로 보인다. 차세대 비휘발성 메모리, 매니코어 등 하드웨어 발전 방향을 고려한 빅데이터 처리 및 저장 관리 기술 개발을 통해 국내 기술의 경쟁력 확보가 필요한 시기이다.

참고문헌

- [1] John F. Gantz, "The Diverse and Exploding Digital Universe," IDC, 2008
- [2] Yvonne Genevese, Ian Bertram, "Information 2020: Senario for Business Intelligence & Information Management," Gartner, 2011
- [3] Massimo Pezzini, Carl Claunch, Joseph Unsworth, "Top 10 Strategic Technology Trends: In memory Computing," Gartner, 2012
- [4] Massimo Pezzini, "Innovation Insight: Invest in In-Memory Computing for Breakthrough Competitive Advantage," Gartner, 2011
- [5] 이미영, 최완, "빅데이터 분석을 위한 빅데이터 처리 기술 동향," 정보처리학회지, 2012.4
- [6] Jeffrey Dean, Sanjay Ghemawat, "MapReduce: Simplified Data processing on large clusters," OSDI, 2004.
- [7] Hadoop MapReduce, <http://hadoop.apache.org/mapreduce>
- [8] M. Isard, Mihai Budiu, Yuan Yu, Andrew Birrell, Dennis Fetterly, "Dryad: Distributed Data-Parallel Programs from Sequential Building Blocks," EuroSys, 2007
- [9] Jaliya Ekanayake, Hui Li, Bingjing Zhang, Thilina Gunarathne, Seung-Hee Bae, "Twister: A Runtime for Iterative MapReduce," HPDC, 2010

[10] Yingyi Bu, Bill Howe, Magdalena Balazinska, Michael D.Ernst, "HaLoop: Efficient Iterative Data Processing on Large Clusters," VLDB, 2010

[11] Gianpaolo Cugola, Alessandro Margara, "Processing Flows of Information: From Data Stream to Complex Event Processing," ACM Journal, 2010

[12] <https://github.com/nathanmarz/storm/wiki/>

[13] Leonard Neumeyer, Bruce Robbins, Anish Nair, Anans Kesari, "S4: Distributed Stream Computing Platform," KDCLOUD, 2010

[14] <http://www.hstreaming.com/>

[15] Carl W. Olofson, "The Big Deal About Big Data," IDC, 2011

[16] Gereon Vey, Ilya Krutov, "SAP In-Memory Computing on IBM eX5 Systems," IBM, 2012

[17] VoltDB Technical Overview, VoltDB, Inc. 2011

[18] Fusion-io, <http://www.fusionio.com/>

[19] 유병곤, 류상욱, 윤성민, "유비쿼터스용 유니버설 메모리 기술(MRAM, FeRAM, PRAM)," 전자통신 동향분석, 2005

[20] Ru Fang, Hui-I Hsiao, Bin He, C. Mohan, Yun Wang, "High Performance Database Logging using Storage Class Memory," ICDE, 2011

[21] Shimin Chen, Phillip B. Gibbons, Suman Nath, "Rethinking Database Algorithms for Phase Change Memory," CIDR, 2011

[22] Fay Chang 외 8인, "Bigtable: A Distributed Storage System for Structured Data," OSDI, 2006

[23] Hadoop Hbase, <http://hadoop.apache.org/hbase>

[24] Giuseppe DeCandia 외 8명, "Dynamo: Amazon's Highly Available Key-value Store," SOSOP, 2007

저자소개



이미영(Miyoung Lee)

'81년 서울대 식품영양학과(학사)
'83년 서울대 계산통계학과(석사)
'05년 충남대 컴퓨터공학과(박사)
'83~'85년 한국전기통신연구소

'88년~현재 한국전자통신연구원/책임연구원
※관심분야 : 데이터베이스 기술, 분산 처리 기술



최완(Wan Choi)

'81년 경북대 전자공학과(학사)
'83년 KAIST 전산학과(석사)
'88년 정보처리 기술사
'85년~현재 한국전자통신연구원/부장

※관심분야 : 클라우드 컴퓨팅, 슈퍼 컴퓨팅 기술