

2013 년 1 월

Technology Inside

LG CNS R&D Journal

In-Memory Computing 기술현황 및 전망

요약

- I. In-Memory Computing 개요
- II. In-Memory Computing 기술
- III. In-Memory Computing 적용
- IV. In-Memory Computing 전망
- V. 시사점

저자 : 구영수 선임연구원
(kys@lgcns.com)

 LG CNS 정보기술연구원

[요약]

인메모리 컴퓨팅(In-Memory Computing, 이하 IMC)이 2011 년에 이어 2012 년에도 가트너 10 대 기술에 2 년 연속 선정될 정도로 화두가 되고 있다. IMC 는 2000 년대 초반 기업에 도입되기 시작하였으나 비싼 메모리 가격과 모든 데이터를 메모리에 올려서 처리해야 할 정도로 빠른 처리가 필요한 비즈니스 어플리케이션 수요가 많지 않아 큰 관심을 받지 못했다. 최근 Big Data 가 이슈화되면서 하드디스크 기반 DBMS 에서 데이터 처리를 위해 메모리로 이동시키는 양이 많아지고 시간 주기가 빨라짐에 따라 기존 DBMS 로는 적시에 Big Data 를 처리하는데 한계를 드러냈다. 이로 인해 데이터를 메모리에 저장하여 빠르게 처리하는 IMC 가 다시 주목 받게 되었다.

IMC 는 하드 디스크에 저장하던 DB 데이터를 메모리에 저장하고 처리하는 컴퓨팅 기술을 말한다. 주요 IMC 기술에는 DRAM, Flash, SSD¹와 같은 하드웨어 기술, In-Memory DBMS, In-Memory Data grid 처럼 어플리케이션이 직접 메모리에 있는 데이터를 관리하는 부담을 덜어주는 In-Memory 데이터 관리 기술, 어플리케이션에서 직접 메모리 통신을 통해 메시지를 교환할 수 있는 메커니즘을 제공하는 In-Memory Messaging 영역, In-Memory 분석, 이벤트 처리와 같은 어플리케이션 플랫폼, In-Memory 데이터 관리 기술 또는 In-Memory Messaging 을 이용해 구축하는 어플리케이션 영역 등이 있다.

초기에는 증권사의 실시간 트레이딩, 통신사의 세션 관리 등 OLTP²성 데이터 처리에 사용되어 왔다. 최근에는 분석용 DBMS, 시각화된 데이터 탐색 툴, 하둡(Hadoop)³ 기반 Big Data 분석 등 분석 어플리케이션을 위한 기반 기술로 활용되고 있다.

IMC 는 단기적으로 Big Data 관련 실시간 분석을 위한 기반 기술로 적용되고 중장기적으로는 아키텍처 및 개발 역량의 변화를 가져올 것으로 전망된다.

¹ SSD(Solid-state Drive): DRAM 휘발성 메모리나 NAND 플래시 비휘발성 메모리에 정보를 저장하는 장치

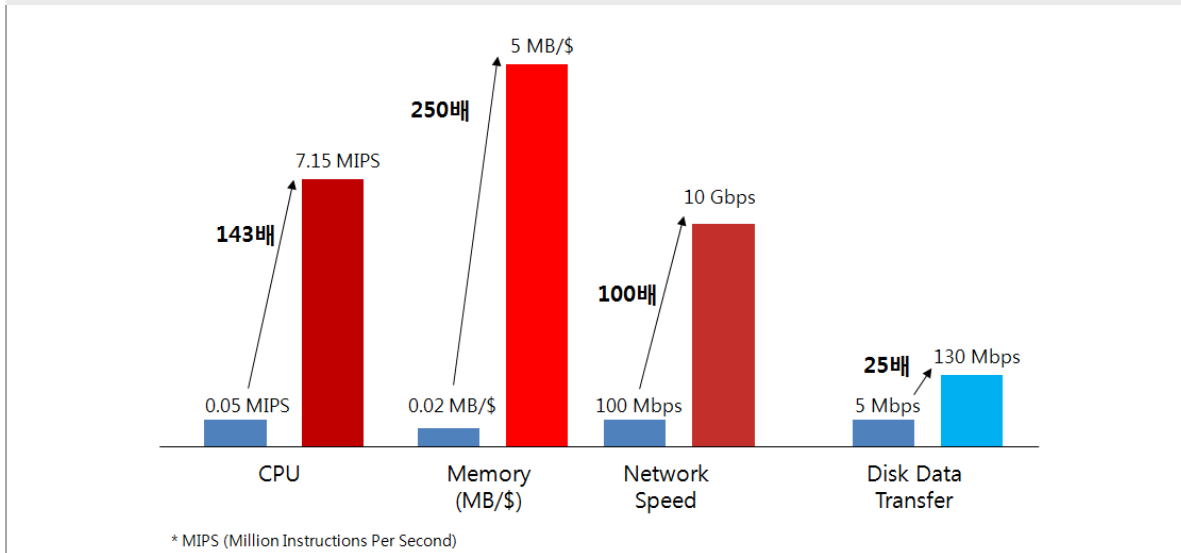
² OLTP(Online Transaction Processing): 온라인상으로 거래 등의 처리(트랜잭션)가 실시간으로 데이터베이스에 갱신되는 처리시스템에 대한 어플리케이션 또는 일련의 처리

³ Hadoop: 대량의 자료를 처리할 수 있는 큰 컴퓨터 클러스터에서 동작하는 분산 응용 프로그램을 지원하는 자유 자바 소프트웨어 프레임워크

I. In-Memory Computing 개요

□ 배경 : 컴퓨팅 환경의 변화가 진행 중임

그림-1. 1990 년 대비 2010 년 기술 발전 비교



- CPU, Memory, Network 속도에 비해 Disk 입출력 속도가 상대적으로 느리게 발전하여 전체 컴퓨팅 성능 향상 속도의 지연 요인으로 지적되고 있음 (그림-1)
- 메모리 가격의 지속적 하락에 따른 서버의 대용량 메모리 지원
 - 1TB⁴의 DRAM 을 지원하는 상용 블레이드 서버⁵ 등장
- 16 Exa⁶ byte 의 데이터까지 처리 가능한 64 비트 컴퓨팅의 보편화
 - 사용가능 메모리 주소 : 1990 년 2^{16} 에서 2010 년 2^{64} 로 2^{48} 배 증가
 - 64 비트 컴퓨팅에서는 CPU 가 메모리에 접근하는데 64 비트 주소를 사용하며 이는 2^{64} 개의 주소를 표현할 수 있음

⁴ TB (Tera byte): 10^{12} byte

⁵ 블레이드 서버(Blade Server): 서버가 차지하는 물리적 공간과 사용 전력을 최소화하기 위해 설계된 서버로 전통적인 박스 형태의 서버보다 공간을 덜 차지함. 고밀도 서버라고도 부름(위키피디아)

⁶ Exa: 10^{18}

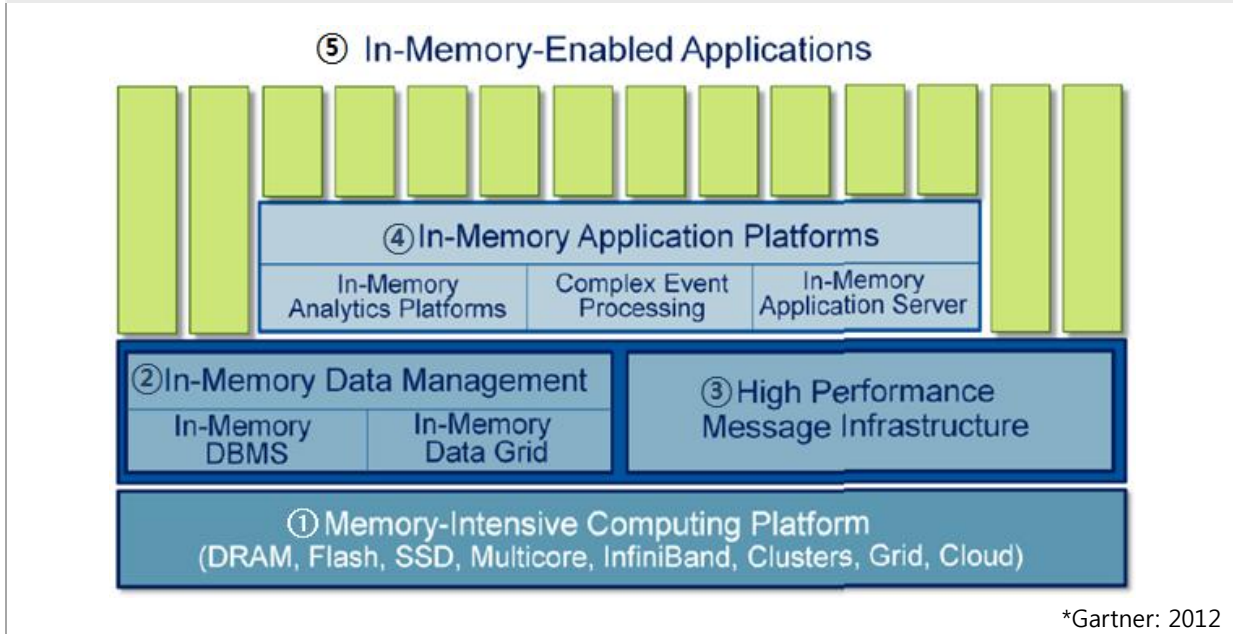
- 32 비트 컴퓨팅에서는 접근가능한 메모리 용량의 한계로 4GB 이상의 대용량 DB 를 메모리에 적재하기 어려웠음
- 64 비트 컴퓨팅에서는 이론상으로 16Exa byte 의 DB 데이터를 메모리에 적재 가능해짐에 따라 사실상 데이터나 어플리케이션의 크기에 상관없이 메모리에서 처리
- 서비스 차원에서 모바일 어플리케이션, 소셜 미디어, 온라인 게임에서 발생하는 실시간 대용량 데이터를 처리하고 분석하는 어플리케이션의 기반 기술로 In-Memory Computing 에 대한 요구가 증가하고 있음

□ In-Memory 컴퓨팅(In-Memory Computing, 이하 IMC)이란?

- 어플리케이션을 구동하는 컴퓨터의 메인 메모리에 DB 데이터와 같은 주요 데이터를 저장하고 처리하는 컴퓨팅 기술(Gartner)
- 분석과 트랜잭션 결과를 즉시 제공하기 위해 서버의 메인 메모리에서 초대용량 실시간 데이터를 처리하는 기술(SAP)
- 현재 디스크 기반 컴퓨팅에서는 데이터를 디스크에 저장하고 이를 필요에 따라 메모리에 적재하여 처리하고 백업을 위해 테이프를 사용함
- IMC 는 디스크 대신 메모리에 데이터를 저장하고 처리하여 백업에 디스크나 테이프를 사용함

II. In-Memory Computing 기술

그림-2. In-Memory Computing Taxonomy



1. Memory-Intensive Computing Platform

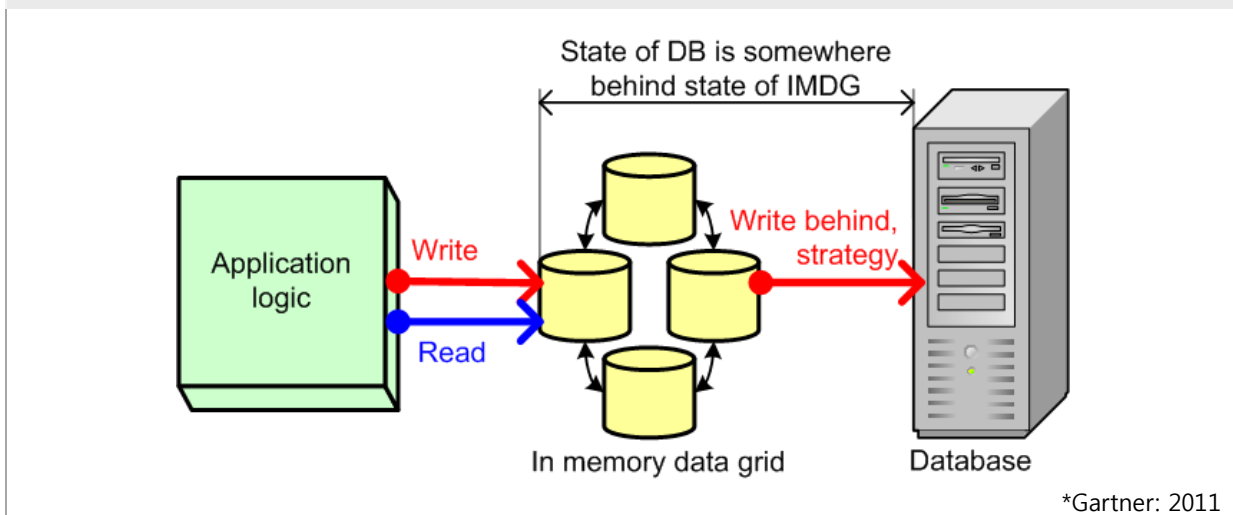
- 메모리 중심의 컴퓨팅을 제공하기 위한 기반이 되는 하드웨어 및 병렬/분산 컴퓨팅 기술
- 64 비트 멀티코어 프로세서, 초대용량 DRAM 및 NAND 플래시 메모리 기술, 플래시 메모리를 이용한 SSD, CPU와 주변기기 간 고속 통신을 위한 InfiniBand⁷ 등의 기술
- 여러 대의 서버를 하나로 묶어서 대용량 데이터를 병렬/분산 처리하여 서비스하기 위한 클러스터, 그리드 기술 및 클라우드 아키텍처

⁷ InfiniBand: CPU와 입출력 장치 간 데이터 흐름에 대한 규격으로 64,000개 주변장치 지원 가능. PCI를 대체할 것으로 예상되고 있음

2. In-Memory Data Management

- In-Memory DBMS : 데이터베이스 구조 전체를 컴퓨터 메인 메모리에 저장하고 어플리케이션이 메모리 위에서 완전히 동작할 수 있도록 입출력 명령 없이 메모리 상의 데이터베이스에 직접 접근하는 DBMS
- In-Memory Data Grid : 어플리케이션이 디스크 기반 DBMS 에 빈번히 접속함으로써 발생하는 병목현상을 줄이기 위해 어플리케이션에서 사용하는 데이터를 메모리에 저장하고 처리할 수 있도록 분산 객체 저장소를 제공하는 미들웨어
- 물리적으로 분리되어 있는 컴퓨터들의 메모리를 논리적으로 묶어 하나의 논리적 공간으로 제공
- 어플리케이션이 물리적 메모리를 할당하고 관리하는 수고를 하지 않도록 하기 위해 데이터 조작을 위한 API(Application Programming Interface)를 제공
- 어플리케이션에서 발생하는 데이터를 In-Memory Data Grid 에 읽고 쓴 후 In-Memory Data Grid 에서 비동기 방식으로 DB 에 저장하는 아키텍처를 보여줌 (그림-3)

그림-3. In-Memory Data Grid (Asynchronous DB writing)



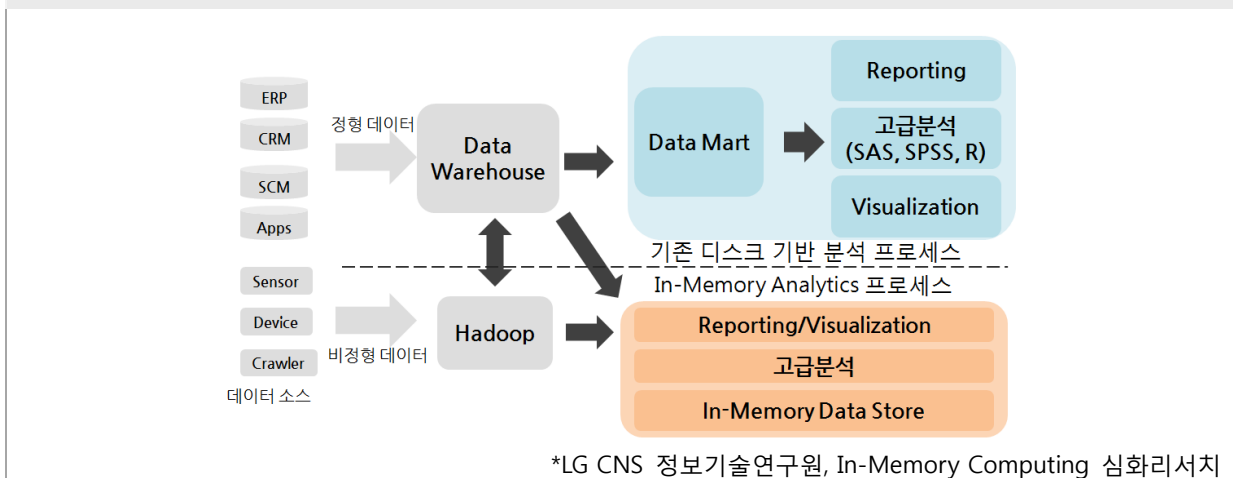
3. High-Performance Message Infrastructure

- 기업 및 기업 내 조직들의 실시간 분석 및 운영 환경에 대한 요구가 증가
- 어플리케이션에서 사용되는 데이터의 용량이 증가하며 필요한 데이터를 더 빠르게 제공할 필요성 제기
- 이전에는 주로 금융 거래를 위한 어플리케이션 또는 항공 및 디지털 제어 시스템에서 고성능 메시지 전달이 요구되었으나 웹 기반 게임 시스템, CRM, 소셜 컴퓨팅, 스마트 그리드, 스마트 시티 등에서 고성능 메시지 기반 어플리케이션에 대한 요구가 증가
- 기존의 MOM(Message Oriented Middleware)에 비해 더 큰 처리 용량, 보다 빠른 전달 속도, 더 많은 메시지 발신자 및 수신자를 지원
- 하나의 인스턴스 당 1 초에 10,000 개 메시지 또는 20MB 이상 전달
- 동일한 메모리 공간 내에서는 1 microsecond 이내에 메시지 전달이 가능하고, 클러스터 내에서는 10Gb 이더넷⁸이나 Infiniband 를 통한 경우 10 microseconds, LAN 을 통한 경우 50 microseconds 내에 메시지 전달 가능

4. In-Memory Application Platform

□ In-Memory Analytics

그림-4. In-Memory Analytics 프로세스와 기존 분석 프로세스 비교



⁸ 10Gb 이더넷 (10gigabit Ethernet): 초당 10 기가 비트의 데이터를 전송가능한 근거리통신망 기술

- 다양한 데이터 소스들로부터 데이터를 메모리에 적재하여 대용량 데이터에 대한 빠른 조회 및 연산 성능을 제공하는 기술
- 기존의 BI⁹에서는 분석 모델을 변경하기 위해 Data Mart의 DB 스키마나 SQL 쿼리 수정이 필요했음. 이를 위해 IT 부서에 요청하고 반영하는데 시간이 소요되므로 실시간 분석이 어려움
- In-Memory Analytics는 분석 모델 변경 시 별도의 Data Mart에 미리 모델을 반영하고 집계해놓을 필요 없이 In-Memory Data Store 상에서 실시간으로 반영하고 분석 가능함
- 시각화 및 텍스트 기반 데이터 탐색 툴¹⁰을 함께 사용하면 보다 직관적이고 자유롭고 빠른 데이터 쿼리를 수행할 수 있음

□ Complex-Event Processing

- 하나 이상의 출처에서 발생하는 이벤트 데이터 스트림을 처리하기 위해 알고리즘과 규칙을 이용해 복합 이벤트(Complex Event)를 처리함
- 대량의 이벤트를 메모리에 저장하고 연속적인 조회를 통해 새로운 데이터가 도착하자마자 작동함
- 전형적으로 개발과 관리 툴, 비즈니스 대시보드, 최종사용자 경고, 이벤트 데이터 소스와 출력 기기를 위한 어댑터가 포함됨

□ In-Memory Data Application Server

- Big Data 어플리케이션을 위한 In-Memory Data Store, In-Memory 클러스터링만의 탄력적 scale-out¹¹ 아키텍처, 이벤트 기반 프로그래밍 모델을 결합한 어플리케이션 플랫폼
- 기존 엔터프라이즈 어플리케이션 서버들도 속도와 확장성을 위해 메모리 캐시를 사용해 왔으나 메인 메모리는 비싸고 용량의 제한이

⁹ BI(Business Intelligence): 데이터, 정보 혹은 통찰에 기반한 의사결정을 지원하기 위한 데이터 분석 툴

¹⁰ 시각화 및 텍스트 기반 데이터 탐색 툴: 기존 BI에 비해 강화된 시각화/자연어 검색 기능을 통해 사용자 편의성을 높이고, In-Memory/인덱싱 기술을 이용해 신속한 분석을 가능케 하는 데이터 분석 툴

¹¹ Scale-out: 컴퓨터 시스템의 성능과 용량을 늘리기 위해 서버 혹은 노드를 추가하여 병렬적으로 연결하는 것을 말함(비교: scale-up은 CPU나 메모리, 하드 디스크 등을 추구함으로써 성능과 용량을 늘리는 것)

있다는 전제가 있었음. 따라서 클라우드 트랜잭션 처리와 클라우드 병렬 처리 등의 요구사항을 반영하지 못함

- 현재의 In-Memory Data Application Server 는 어플리케이션 데이터를 메모리에 모두 적재하고 SSD 나 HDD 등 외부 스토리지 기기는 경우에 따라서 제한적으로 사용한다는 전제하에 설계됨

5. In-Memory-Enabled Applications

- In-Memory Analytics, Event Processing Platform, In-Memory Application Server 와 같은 어플리케이션 플랫폼을 이용하거나 또는 직접 In-Memory Messaging 을 이용하여 구축된 어플리케이션

III. 산업별 In-Memory Computing 적용

1. 여행 : 고객 정보 통합 및 신속한 분석을 통한 맞춤형 서비스 실현

□ 정보의 통합과 신속한 분석의 필요성 대두

- 온라인 여행 서비스 영역의 급성장에 따라 기존의 기술로는 50 여 개국에 걸친 55,000 개의 호텔 네트워크, 크루즈 여행을 위한 해변 시설 및 서비스 데이터베이스를 이용해 적시에 적절한 가격에 적합한 서비스를 고객에게 제공하기 어려워짐
- 새로 등장하는 경쟁 업체들과의 경쟁 속에서 시장 선도 기업으로서의 위치를 유지하기 위해 비즈니스 성과, 고객 행동, 예약 취소율 등을 지속적으로 관찰하고 분석할 필요가 있음
- 기존 DBMS에서는 대용량 데이터 분석에 한계가 있어 레코드 크기를 줄이기 위해 고객 정보와 공급 정보를 분리하여 관리함으로써 고객 관점의 데이터 분석이 어려움이 있었음

□ IMC의 적용 및 효과

- 고객 정보와 공급 정보로 분리되어 있던 데이터를 고객 관점으로 통합하여 분석함으로써 고객 요구에 대한 이해도 향상
- In-Memory DBMS 와 In-Memory Analytics 기술 도입을 통해 데이터 적재 소요 시간 및 분석 시간 단축
 - 1 일 데이터 적재 시간 : 6 시간 → 15 분
 - 모든 데이터 재적재 시간 : 수 주일 → 1 일 이하
 - 수백만 건의 데이터를 분석하는데 수 시간이 소요되었으나 IMC 적용 후 수억 건의 데이터를 수 초 안에 분석할 수 있게 됨

2. 공공 : 교통정보 실시간 분석을 통한 필요한 데이터 추출 효율성 제고

□ 대용량 교통정보의 실시간 분석 필요

- 이동 중인 차량으로부터 들어오는 수많은 위치 및 속도 정보의 실시간 처리 및 분석이 어려워 도로에 설치된 한정된 수의 센서 정보만 분석

□ IMC 적용 및 효과

- 도로 상의 한정된 수의 센서 정보뿐만 아니라 운행 중인 택시의 위치 및 속도 센서에서 발생하는 데이터를 In-Memory DBMS 에 저장
- In-Memory DBMS 로부터 실시간 교통량 분석 및 경로 계산에 필요한 데이터 추출하여 BI 툴을 이용해 연산
- 약 1 만 3 천 대의 택시들로부터 들어오는 3 억 3,600 만 건의 데이터에서 분석에 필요한 600 건의 데이터를 추출하는데 기존에는 수 분이 걸렸으나 IMC 를 이용해 1 초 내로 단축됨

3. 통신 : 과금 데이터의 실시간 처리를 통한 시스템의 성능 향상

□ 과금 데이터의 처리속도 개선 필요

- 기존의 하드디스크 기반 DBMS 를 이용해 과금 업무를 처리할 경우 하드디스크에 과금 관련 데이터를 읽고 쓰는 부분에서 병목현상이 발생하여 과금 업무 처리가 지연됨. 예를 들어 청소년 요금제나 선불폰의 경우 실시간으로 사용가능한 데이터 용량을 계산하고 사용량을 모두 소진한 경우 서비스를 중지해야 하나 디스크 기반 DBMS 를 이용할 경우 실시간 처리에 어려움이 있음
- 기존 과금 시스템으로는 점점 복잡해지는 요금제와 부가 서비스를 반영하는데 한계가 있었음

□ IMC 의 적용 및 효과

- 빠른 DB 응답시간을 제공하는 In-Memory DBMS 도입을 통해 하루에 수억 건의 과금 데이터를 실시간 처리하거나 복잡한 요금 계산을 효과적으로 연산할 수 있게 됨
- 기존 시스템에 비해 40 배의 성능 향상을 보임

4. 마케팅 : 대용량의 고객정보 분석을 통한 맞춤형 마케팅 효과 제고

□ 대용량 고객 데이터의 신속한 분석 처리 필요

- 통신사업에서는 고객 이탈 방지 및 신규 고객 확보를 위한 맞춤형 마케팅이 매우 중요함
- 기존의 시스템으로는 매 9개월마다 2배씩 증가하는 정보 수집과 처리에 한계가 있었음
- 생산성 향상 차원에서 모바일 환경을 최대한 활용한 영업 및 고객 지원을 위해 신속한 분석처리 성능이 필요함

□ IMC의 적용 및 효과

- 기존시스템보다 5~60 배 이상의 분석 처리 성능을 가진 In-Memory DBMS 도입
- 5.5 배 이상의 데이터 압축을 통해 18개월 분량, 60억 건의 데이터 처리 가능
- 마케팅 팀의 고객 분석 소요 시간이 2주에서 1일로 단축됨
- 이를 통한 스마트폰 및 태블릿 고객에게 모바일을 통한 개인화된 맞춤형 마케팅 활동 가능해짐
- 향후 마케팅 외 재무, 인사 등 대용량 처리가 필요한 업무 영역으로도 IMC 적용 확대

IV. In-Memory Computing 전망

□ In-Memory DBMS 는 기존 Disk 기반 DBMS 의 보완 또는 대체용으로 적용이 확대될 전망이다

- SAP : 현재는 분석용 DBMS 용으로 사용되고 있음. 데이터를 휘발성인 DRAM 에 저장하고 처리하는 것이 특징. 향후 ERP 와 같은 OLTP 성 데이터까지도 메모리 상에서 처리하여 기존 디스크 기반 DBMS 를 In-Memory DBMS 로 대체하고자 함
- Oracle : 모든 데이터를 DRAM 에 처리하는 것은 아니지만 빈번히 사용되는 데이터를 DRAM 에 저장, 처리하여 디스크 입출력으로 인한 지연 시간을 줄이고 나머지는 플래시 메모리 및 디스크에 저장하는 방식을 사용함. 모든 데이터를 DRAM 에 적재하여 처리하지는 않고 기존에 보유하고 있는 DBMS 를 보완하기 위해 In-Memory DBMS 를 사용함

□ In-Memory Data Grid 가 기존 분석 어플리케이션에 통합될 전망이다

- 기업용 트랜잭션 시스템, 모바일 어플리케이션, 소셜 미디어 등에서 발생하는 실시간 대용량 데이터를 처리하기 위한 분석 어플리케이션의 기반 기술로서 중요성이 확대될 것임
- In-Memory Data Grid 는 단독 솔루션으로서 사용이 늘어나고 있으며 뿐만 아니라 순간적으로 증가하는 조회 요청으로 인해 확장성이 중요한 포털(Portal), 실시간 대용량 이벤트 처리 플랫폼, 어플리케이션 서버 등에 통합되어 제공될 전망이다

□ In-Memory Analytics 가 상용 통계분석 솔루션에도 적용될 전망이다

- In-Memory Analytics 기술은 수년 전부터 BI 시각화 영역에서 적용되어 왔으며 많은 기업들이 도입하여 시장에서 주류로 정착되었음

- 최근에는 BI 시각화 기능 외 상용 통계분석 솔루션에도 In-Memory 기술을 적용하는 제품이 출시되는 등 In-Memory Analytics의 영역이 확대되고 있으며 향후 오픈소스 통계분석 솔루션에도 In-Memory 기술이 적용될 전망

V. 시사점

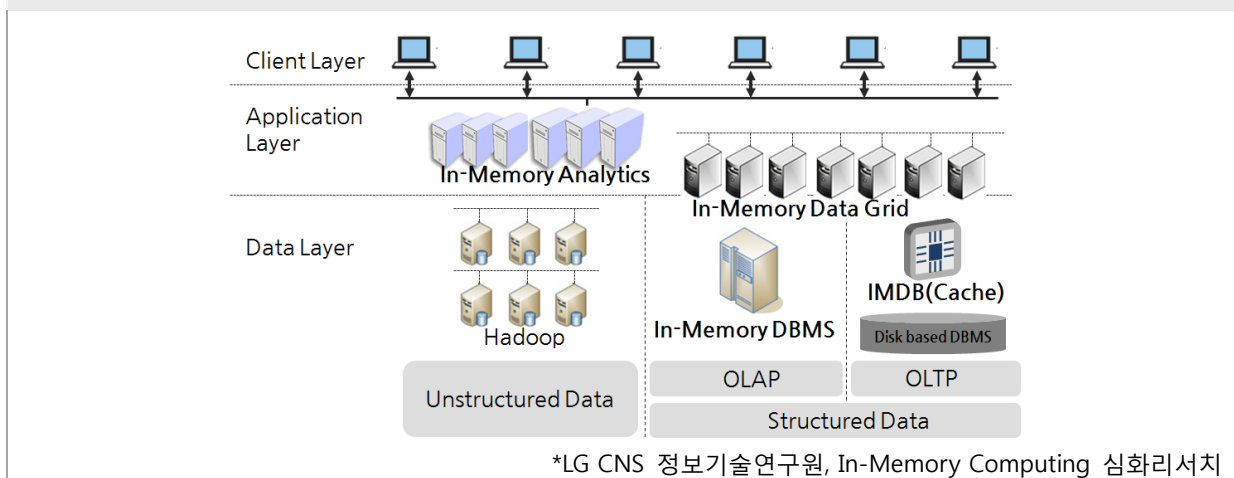
□ IMC 기술 이해와 기회 탐색을 통해 IMC 전략을 수립해야 함

- 금융, 광고, 미디어 등의 산업에 영향이 크고 물류, 유통, 제약, 자동차 산업 등에서 수요가 있을 것으로 전망됨
- 특히 인터넷 상거래, 온라인 게임, 금융, 소셜 네트워크, SaaS의 경우 정상적인 서비스 제공을 위해 IMC 기술이 필수적임
- IMC의 가장 큰 강점인 빠른 속도를 이용해 과거에는 구현 못했던 어플리케이션이 구현 가능해짐에 따라 유용성이 증가할 것임

□ 기존 기술적 안정성 문제를 극복할 수 있는 차세대 비휘발성 메모리 개발에 주목해야 함

- 현재는 휘발성이 있는 DRAM과 읽기 및 쓰기 횟수가 제한적인 플래시 메모리를 사용하기 때문에 안정성 및 기능에 대한 검증이 요구되고 있음
- 메모리 반도체 제공업체들은 차세대 비휘발성 메모리 개발에 박차를 가하고 있으며 개발 후 시장성을 갖춘 양산 제품이 출시되면 IMC의 기술적 안정성 문제가 많이 해소될 것임

그림-5. In-Memory Computing Architecture



□ 장기적인 관점에서 IMC 기반으로 아키텍처의 변화에 대응해야 함

- 디스크 기반 DBMS 이용을 전제로 한 현재의 어플리케이션에서 메모리 기반 어플리케이션으로 어플리케이션 아키텍처가 변화될 것. 예를 들어 현재는 회계, 고객, 공급망 정보 등이 발생하여 저장/처리되는 OLTP 시스템과 데이터를 분석하기 위한 DW/BI 시스템이 분리되어 있으나, 장기적으로는 발생하는 데이터를 실시간 분석하고 이를 의사결정에 반영하기 위한 OLTP와 분석 시스템이 통합을 지향하게 될 것임
- IMDB는 Data 계층, IMDG는 Data-Application 계층, In-Memory Analytics는 Application 계층에 해당하나 In-Memory Analytics는 Hadoop과 결합되고 있고, IMDB와 IMDG를 통합한 제품이 등장하는 등 Application 계층과 Data 계층의 경계가 흐려지며 3-Tier 아키텍처에서 2-Tier 아키텍처로의 변화가능성을 발견할 수 있음 (그림-5)

[참고문헌]

- Massimo Pezzini 외 7 인 (2012). 『Taxonomy, Definitions and Vendor Landscape for In-Memory Computing Technologies』 Gartner
- Massimo Pezzini (2011). 『Net IT Out : In-Memory Computing - Thinking the Unthinkable Applications』 Gartner Symposium 2011
- 구영수 (2012). 『In-Memory Computing 심화리서치』 LG CNS 정보기술연구원
- SAS White Paper (2011). 『In-Memory Analytics for Big Data』 SAS
- Technical White Paper (2011). 『VMware vFabric GemFire – High Performance, Distributed Main-Memory and Events Platform』 VMware
- SAP In-Memory Computing 소개 웹 페이지 : <http://www.sap.com/korea/platform/in-memory-computing/index.epx>
- Oracle In-Memory DBMS 제품 웹 페이지 : <http://www.oracle.com/technetwork/products/timesten/overview/index.html>
- IBM In-Memory DBMS 제품 웹 페이지 : <http://www-01.ibm.com/software/data/soliddb/>
- 그림-2 출처 : http://blog.ragozin.info/2009_04_01_archive.html