

0강 개념이해





데이터의 이해



데이터의 종류

❖ 정형데이터

- 고정된 필드에 저장되는 데이터
- 학생정보, 성적정보, 게임캐릭터 정보 등.

❖ 반정형 데이터

- 고정된 필드를 사용하지 않지만 메타데이터와 같은 정보를 사용
- XML, JSON

❖ 비정형데이터

- 필드가 고정되지 않고 다양한 데이터 사용
- 사진, 동영상, 메신저, 위치정보 등

SQL vs NoSQL

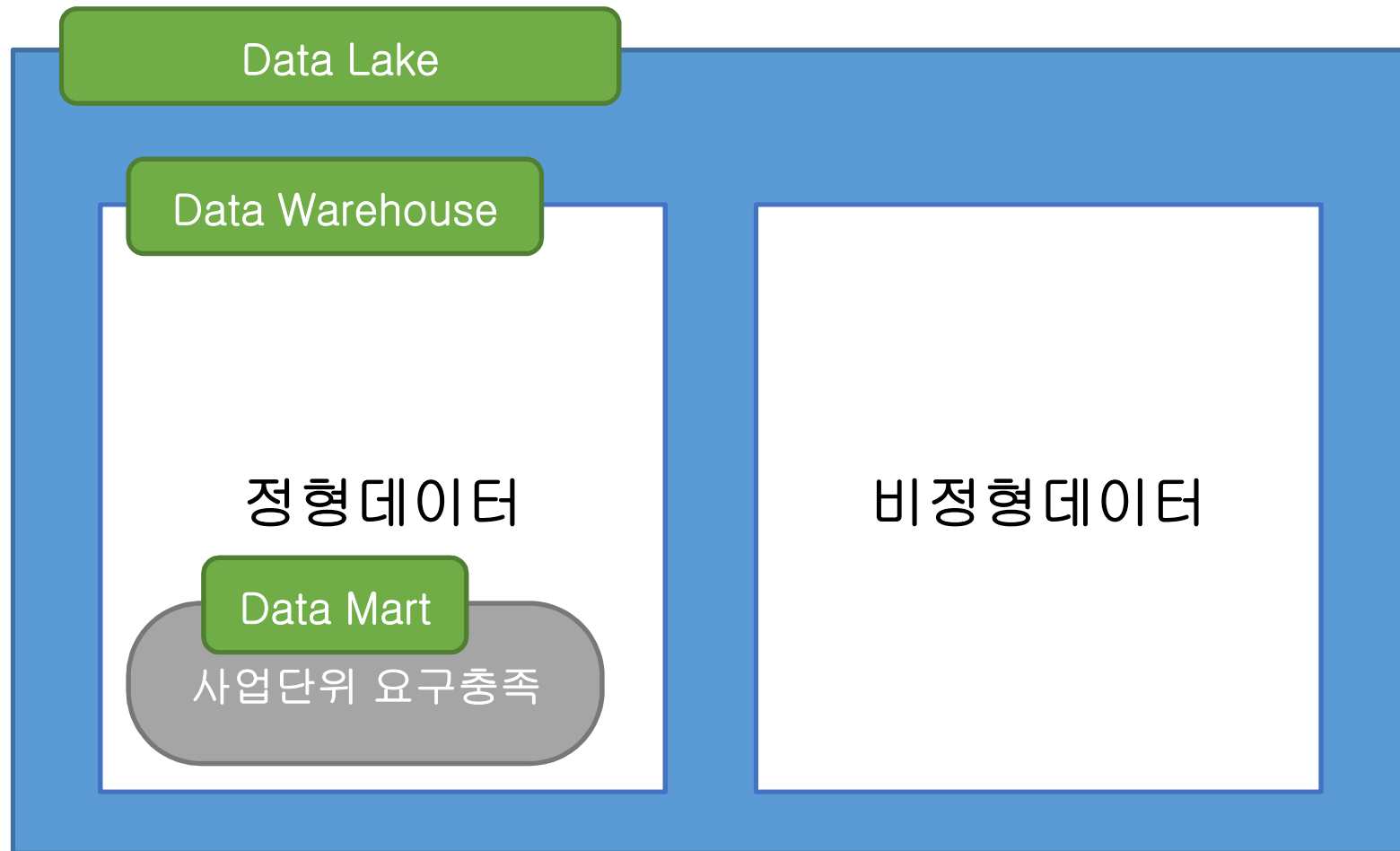
❖ SQL

- 기존의 RDBMS에서 테이블간의 관계를 표현하기 위해 사용되었던 기술
- Amazon aurora, mysql, MariaDB, PostgreSQL, Oracle, MS SQL

❖ NoSQL

- 비정형데이터를 처리하기 위한 기술
- 관계보다는 키와 값을 갖는 단순 처리
- 조인을 사용하지 않으며 샤딩을 활용
- MongoDB, Hbase,
- Amazon 지원 DB
 - 키값 기반 DynamoDB
 - 문서 기반 DocumentDB
 - 그래프 기반 amazon Neptune

데이터 저장 방식



데이터 저장 기법

❖ 파일스토리지

- 파일과 폴더로 이루어진 저장 방식

❖ 블록스토리지

- 블록 단위로 저장
- DB에서 Table을 구성하여 사용하는 방법

❖ 오브젝트 스토리지

- 논리적인 가상의 공간에 인덱싱을 이용하여 저장하는 방법
- 단일 저장소에 저장
- AWS의 경우 S3가 이 방식을 사용함.

데이터 저장 장치

❖ DAS (Direct Attached Storage)

- 서버에 직접 외장 저장장치 추가
- 속도가 빠르고 확장이 쉽다.
- 연결 수가 한계가 있다.

❖ NAS(Network Attached Storage)

- 네트워크를 통해 연결
- 중계 역할을 하는 파일 서버 사용
- 확장성과 유연성이 뛰어남
- 경제적이며 설치와 유지보수가 용이
- 접속 증가 시 성능이 저하됨
- 파일 전송 속도는 DAS보다 느림
- 파일 시스템 공유로 보안성 약화

❖ SAN(Storage-Area Network)

- Fiber Channel Switch로 연결하여 고속처리 가능



빅데이터

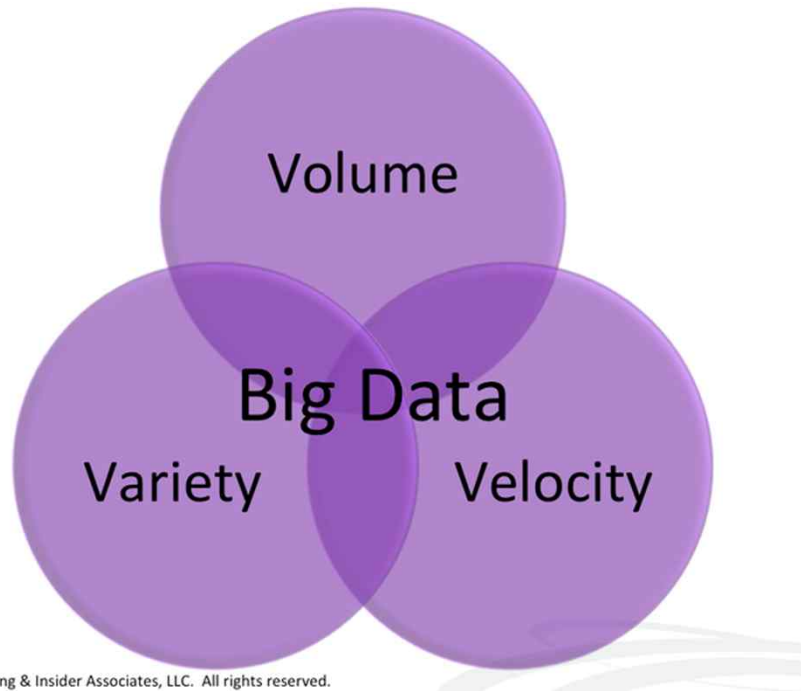


빅데이터란

- ❖ 데이터의 규모에 초점을 맞춘 정의
기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석 역량을 넘어서는 데이터 (맥킨지 2011년 5월)
- ❖ 업무 수행 방식에 초점을 맞춘 정의
다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 빠른 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처(IDC 2011년 6월)

빅데이터의 3대 요소(3V)

❖ BI/DW 리서치 기관인 TDWI 가 빅데이터의 3대 요소(3V)를 그림으로 표현



- Volume : terabyte 이상
- Velocity
 - 실시간 처리
 - 장기적인 접근
- Variety
 - 정형데이터
 - 반정형데이터
 - 비정형데이터

❖ 3대 요소 가운데 두 가지 이상의 요소만 충족된다면 빅데이터라고 볼 수 있음

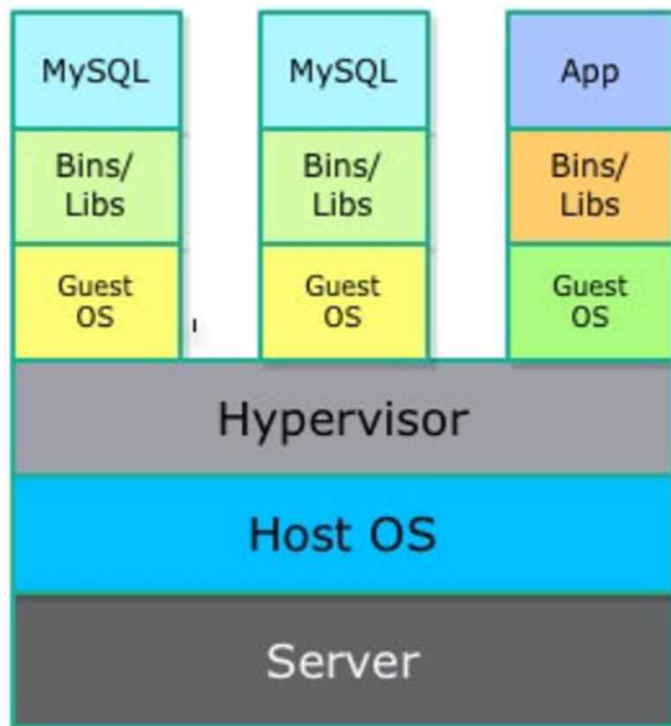


가상화

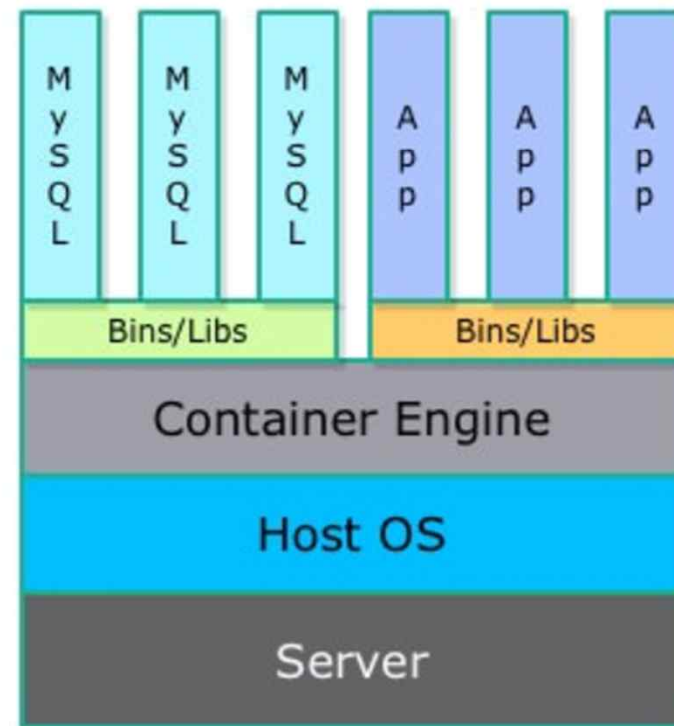


가상머신과 컨테이너

Virtual Machines



Containers



참조 : <http://blog.drakejin.me/Docker-araboza-1/>