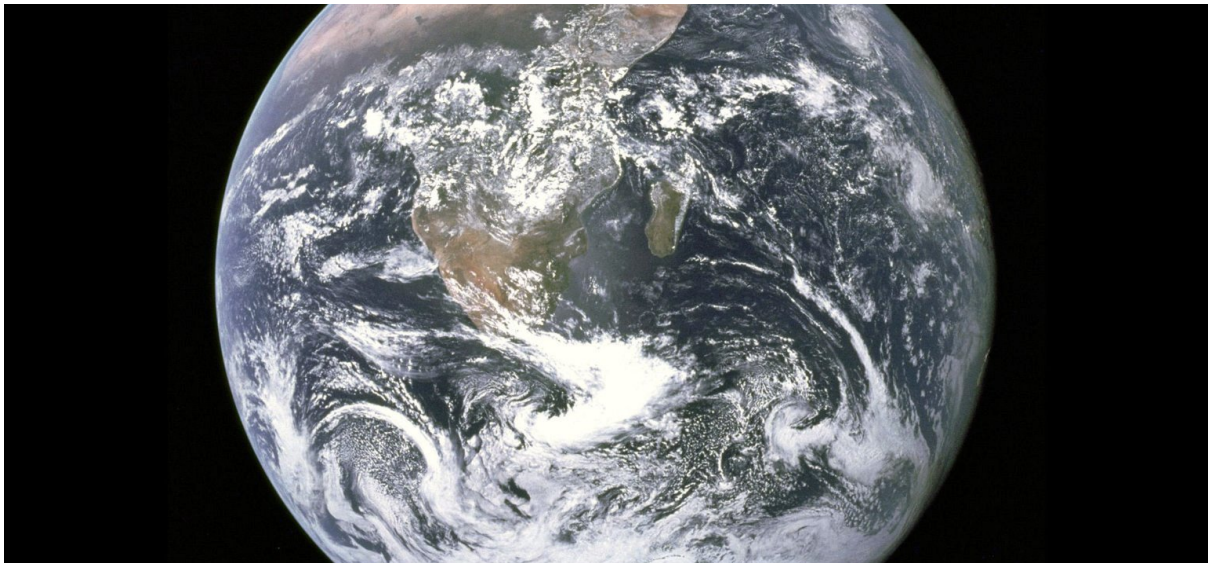

1. 몽고 DB 소개, 기초, 특징

● NoSQL?

. Non Relational Operational Database SQL로 '관계형 데이터베이스가 아닌 SQL'이라는 의미로 해석

. 클라우드 컴퓨팅 환경에서 발생하는 빅 데이터를 효과적으로 저장, 관리하는데 문제점을 보완하기 위한 새로운 데이터 저장 기술로 NoSQL이 대두됨.

- 1960 ~ 1980년대에는 Main-Frame에서 저장 관리하는 개발 환경을 사용하였기에 해당 서버의 CPU, Memory, Disk의 Resource가 많이 요구됨 → File System의 Database
- 1988년도 기점으로 국내 PC가 보급 되었지만 활용도는 낮아 이를 활용하기 위해 서버에는 데이터를 배치, 사용자의 PC에는 운영 프로그램을 배치하여 서버 관리 비용은 줄이고 PC 자원은 적극적으로 활용하는 클라이언트/서버 환경 구축 → 많은 사람이 서로 공유 하기 위해 DBMS 기술 발전
- 빅 데이터가 양산됨에 따라 운영프로그램들을 언제, 어디에서도 설치, 사용할 수 있어야 하며 새로운 버전의 프로그램들을 저렴하게 공급 받고 사용가능 해야함 (클라우드 컴퓨팅 환경) → 기업 환경에 맞게 최적의 데이터 베이스를 구축 할 수 있어야 하며 필요에 따라 DBMS 소스 프로그램을 수정할 수 있어야 하며 저렴한 비용으로도 구축 될 수 있는 Free-Ware 제품이 요구 → NoSQL 필요성 제기
- 2000 ~ 2015년 경까지의 제품들을 1세대 NoSQL이라 하며 Document DB, KeyValue DB, ColumnFamily DB, Graph DB로 크게 4가지 저장 유형으로 각각 제

품화 되어 있음

- 2016 ~ 최근까지 제품은 2세대 NoSQL이라 불리며 1세대의 유형을 통합한 Multi-Data Model로 제공되는 것이 특징이며 RDBMS의 경우 문자, 숫자를 기반으로 하기에 CPU기반 솔루션인 반면 NoSQL은 동영상, 음성, 사진, 좌표와 같은 바이너리 데이터을 기반으로 하는 GPU 기반 솔루션으로 볼 수 있음

● NoSQL의 장점

1. 클라우드 컴퓨팅 환경에 적합하다.
   . Open Source이다.
   . 기존 RDBMS에서 처리하기 힘든 비정형 데이터의 저장과 관리가 용이하다.
   . 무엇보다 RDBMS에 비해 분산 처리와 병렬 처리가 용이하며 운영 및 관리가 쉽다.

2. 유연한 데이터 모델이다.
   . 비정형 데이터 구조 설계로 설계 비용 감소
   . RDBMS의 Relationship과 Join 구조를 Linking과 Embedded로 구현하여 성능이 빠르다.

3. Big Data 처리에 효과적이다.
   . Memory Mapping 기능을 통해 Read/Write가 빠르다.
   . 전형적인 OS와 Hardware에 구축할 수 있다.
   . 기존 RDBMS와 동일하게 트랜잭션 제어가 가능하다.

4. 기존 RDBMS에 비해 비용이 적게 든다.
   . Open Source 사용 시 SW 구입 비용이 없다.
   . 기존 RDBMS에 비해 유지보수 비용이 적게 든다.
   . 유연한 데이터 모델을 제공하기 때문에 개발시간과 구축비용이든다.

구축 비용

Aa 제목	선행 비용(UpFront Cost)1	진행 비용(Ongoing Cost)1
제목 없음	초기 개발자 비용 (Application Coding 및 Data Store)	진행 개발자 비용 (사용자 요구에 따라 변경 등)

Aa 제목	선행 비용(UpFront Cost)1	진행 비용(Ongoing Cost)1
제목 없음	초기 관리자 비용 (설치, 환경 설정, Shard/Replication등)	진행 관리자 비용 (Data 유지보수 Health Check Backup & Recovery등)
제목 없음	Software Licenses	Software Maintenance & Support
제목 없음	Server Hardware	Server Maintenance & Support
제목 없음	Server Storage	Storage Maintenance & Support
제목 없음		기타 개발 비용

. 선행 비용 = 초기 구축 비용

. 진행 비용 = 운영 비용

. RDBMS 도입 보다 1/3비용으로 구축 가능

● NoSQL의 종류 및 제품군

1. Key-Value Database(1세대)

   . Amazon's Dynamo Paper
   . Data Model : Collection of K-V pairs
   . 제품 유형 : Riak, Voldemort, Tokyo

2. ColumnFamily Database(1세대)

   . Google's BigTable paper

   . Data Model : Column Families

   . 제품 유형 : Hbase, Casandra, Hypertable

3. Document Database(1세대)

   . Lotus Notes
   . Data Model : Collection of K-V collection

. 제품 유형 : MongoDB, Cough DB

4. Graph Database(1세대)

. Euler & Graph Theory

. Data Model : nodes, rels, K-V on both

. 제품 유형 : AllegroGraph, Neo4J

5. Multi-Model Database(2세대)

. Data Model :Document + KV + ColumnFamily + Graph Model

. 제품 유형 : Cosmos DB, Dynamo DB, Arango DB, Orient DB

● NoSQL EcoSystem

MongoDB EcoSystem/ Hadoop EcoSystem

● MongoDB란

1. Humongos라는 회사의 제품명이었으며 현재 MongoDB Inc.로 회사명 변경
2, JSON Type의 데이터 저장 구조 제공
3. Sharding(분산/병렬처리) / Replica(복제) 기능 제공
4. MapReduce(분산/병렬처리) 기능을 제공
5. CRUD 위주의 다중 트랜잭션 처리도 가능
6. Memory Mapping 기술을 기반으로 Big Data 처리에 탁월한 성능을 제공한다.

● 용어 설명(Terminology)

. Table → Collection

. Row → Document

. Column → Field

. Primary key → Object_ID field

. Relationship → Embedded & Link