Needs of data management and multi-model databases

Database Systems Lab

Bongki Moon

Database Systems Lab – Mission Statement

- Understand the need of data management and scalable strategies for a variety of traditional and emerging database applications.

- Development of new indexing and query processing algorithms

- Disruptive storage paradigms with new storage media

- Scale-out platforms for data services

Why Database Systems?

- Present at the core of most contemporary infrastructure for banking and commercial acts.

- Relational database systems(RDBMS's) serve as the engine and repositories of records for online transactions(OLTP) and online analytics(OLAP)

- However, explosion in unstructured data shows traditional RDBMS is not enough.
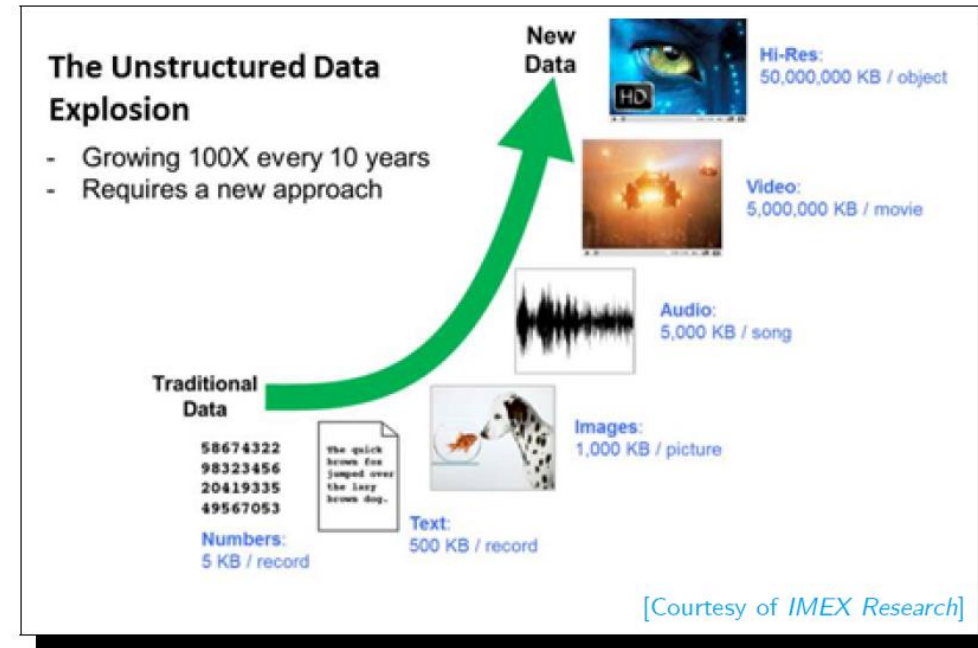
Traditional bigdata analysis and its limits

- Data scientists want to find the upcoming bottleneck with current traffic data, road network data and the navigation application data.
  - Traffic data is in a traditional relational model
  - Navigation application data is in a graph model
 - Road network data is in the JSON format
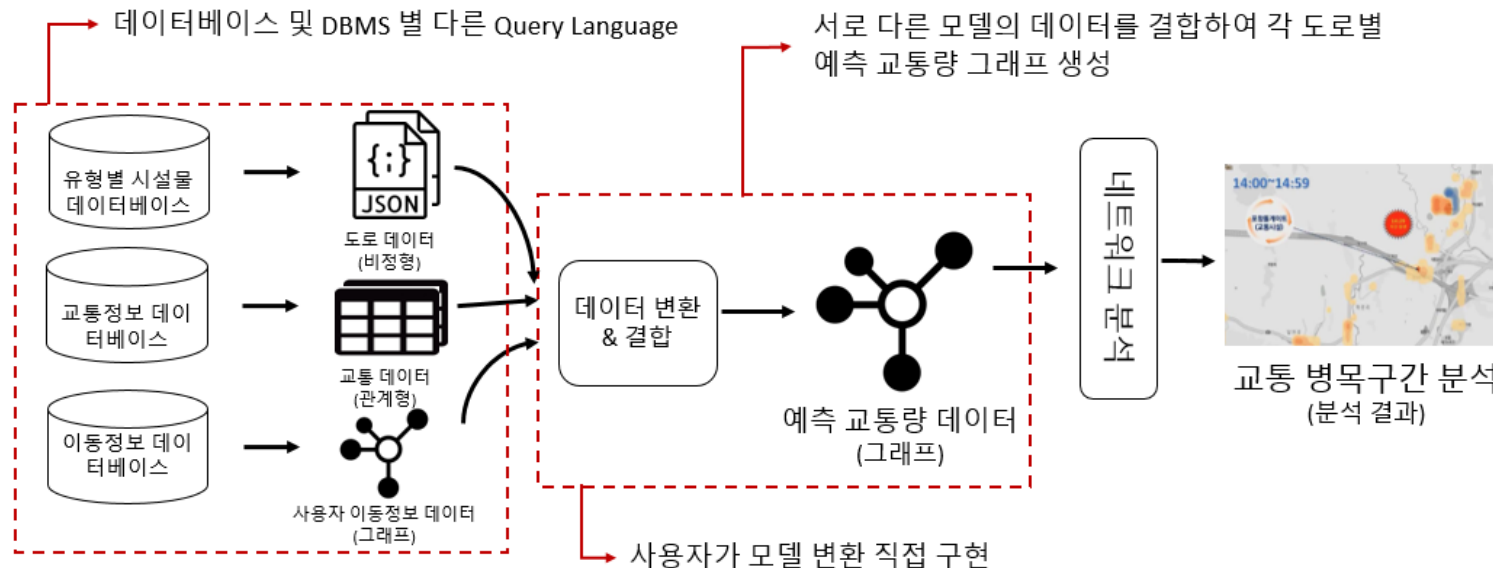
MxMDB Project (2020.03 ~ 2024.02)

- Single DBMS that can use multiple data models (Relational, JSON, Array, Graph) with multiple storage engines.

- Supports multi-model query processing for casual users.
  - Inter model foreign key to manage multiple models efficiently.

- Intuitive API to maintain database and enable effective analysis.
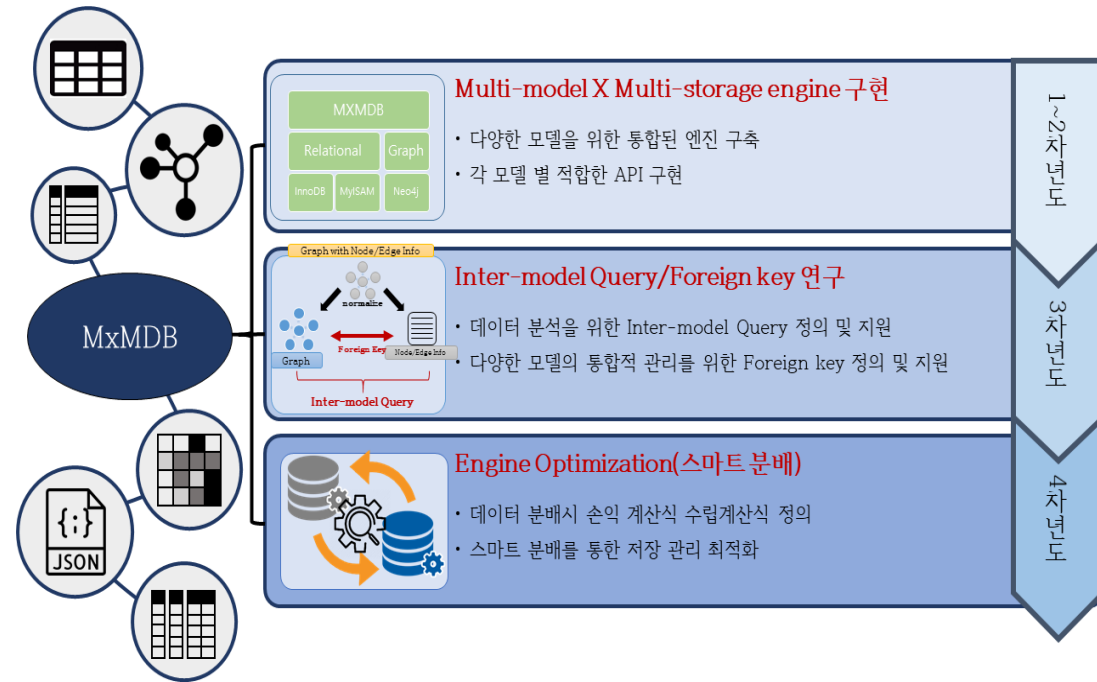
Expected benefits from MxMDB
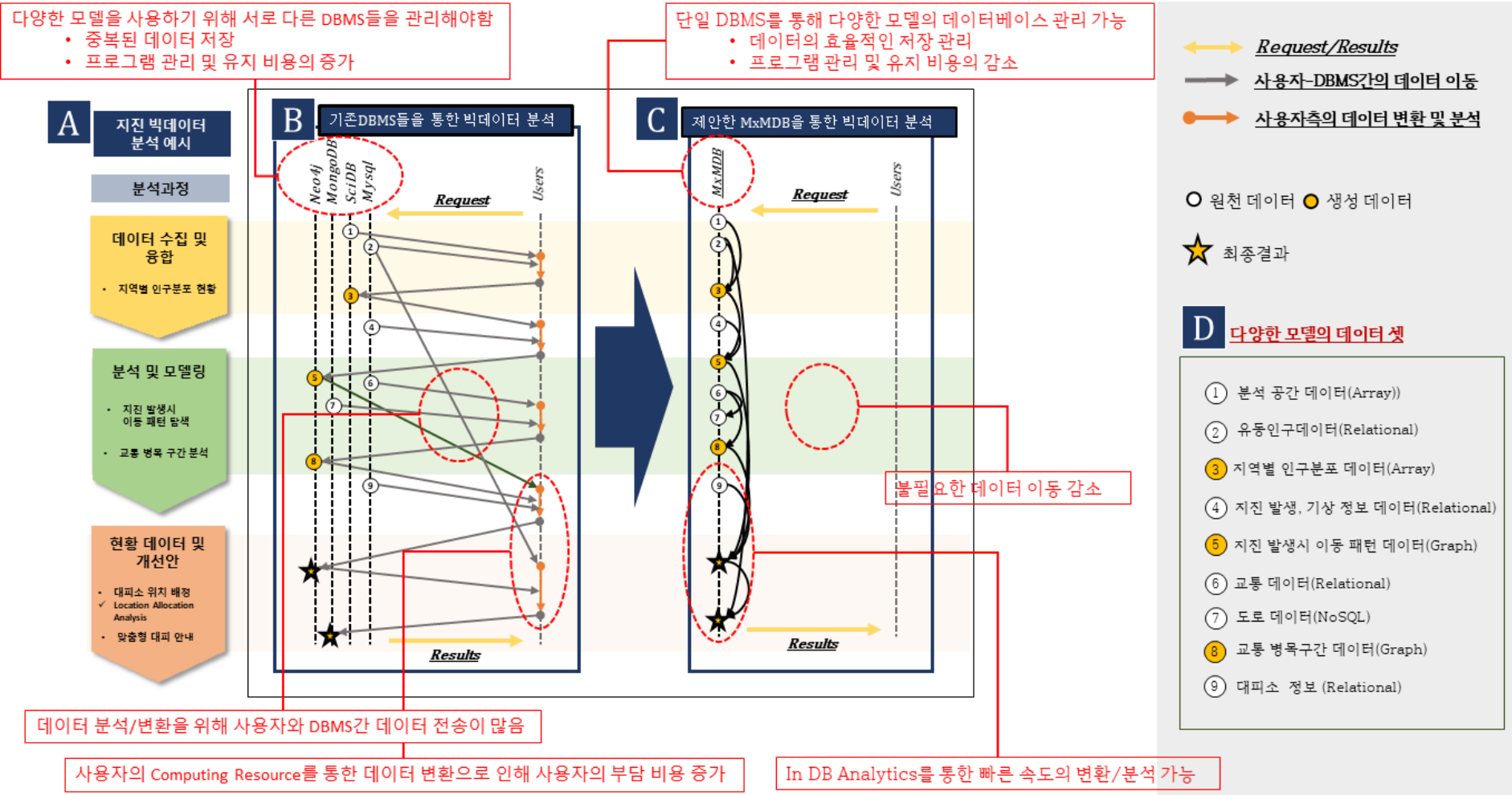
MxMDB – Challenges

- Multi-model query and inter-model foreign key are not well or formally defined.

- Different models across multiple storage engines make it difficult to maintain consistency.

- Not all storage engines are open-sourced and each uses different language.

Storage Engine	Database Model	Concurrency Control	Physical Layout	Feature	Language
InnoDB	Relational	Lock	variable-length record	acid full text search	C
HBase	Multi-dimensional data	MVCC	column family store	lsm-tree	JAVA
WiredTiger	key-value(Table) JSON	MVCC	column, row store	lsm-tree b+tree	C
SciDB	Multi-dimensional data	X	array, chunk	linear algebra	C++
Neo4j	graph	Lock	fixed size length	graph traversal	JAVA

Other projects

- **Construction and Demonstration of Fine Particle Monitoring System using Scanning LIDAR (2019 ~)**
  - Visualization and monitoring system with an efficient storage for multi-dimensional and spatial data on fine particles.

- **Predicting Potential Drug-Drug Interactions via Big Data Analysis and Machine Learning (2019 ~)**
  - System for predicting potential drug-drug interactions(DDI)

- **Big Data Big Computing Engine for High-Performance Computer (2016 ~)**
  - Big Data Big Computing (BDBC) engine for massive application programs based on high-performance computing.
  - Overcome the limitations of large-scale computations such as long-running time and massive memory consumption.