

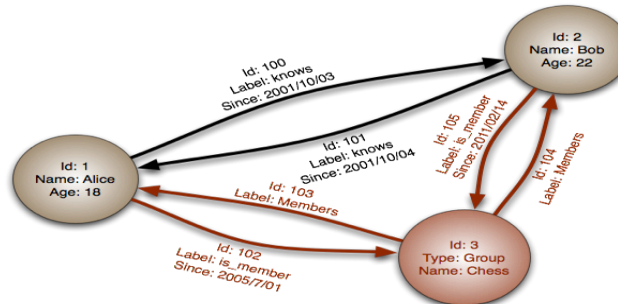
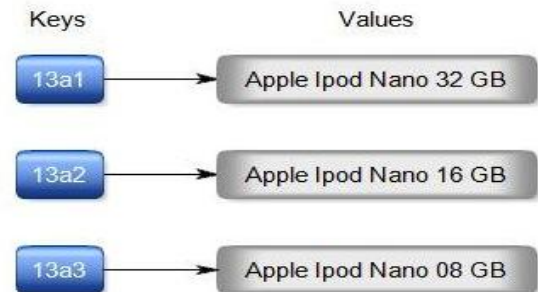
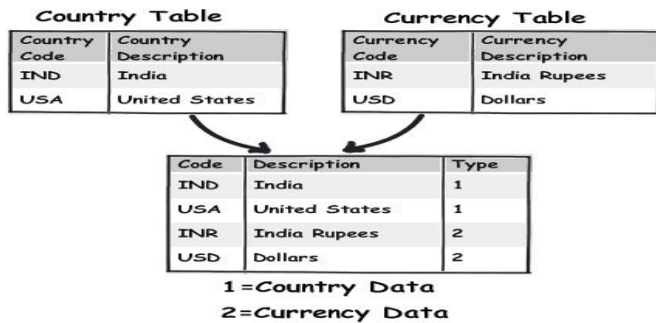


Multi-representational Data Processing: Leveraging Replication to tackle Variety of Big Data

Vaibhav Arora
University of California, Santa Barbara

BIG DATA – Variety

- ▶ Structured data, semi-structured data and Un-structured data
- ▶ Variety in structure as well as access methodologies



Fall of 'One Size Fits all'

- ▶ Multitude of Systems to tackle **Variety**

- ▶ OLTP



- ▶ OLAP



- ▶ Graph

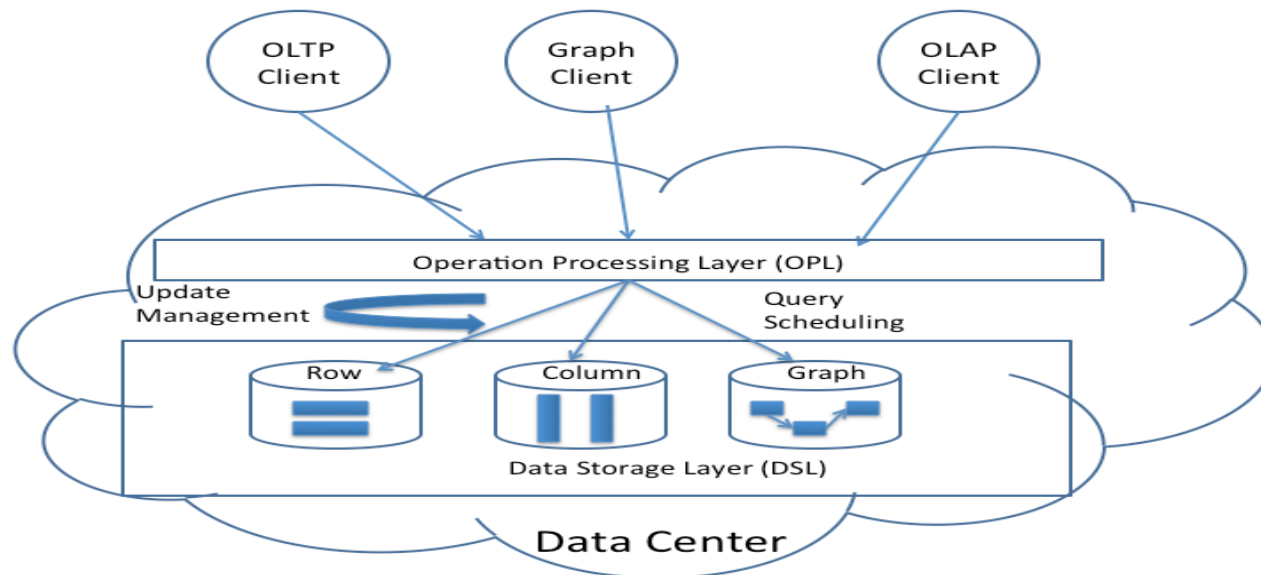


- ▶ **Drawbacks**

- Stale Analytical Results
- Maintenance of data pipelines like **ETL**
- Application Developer must choose “right solution”

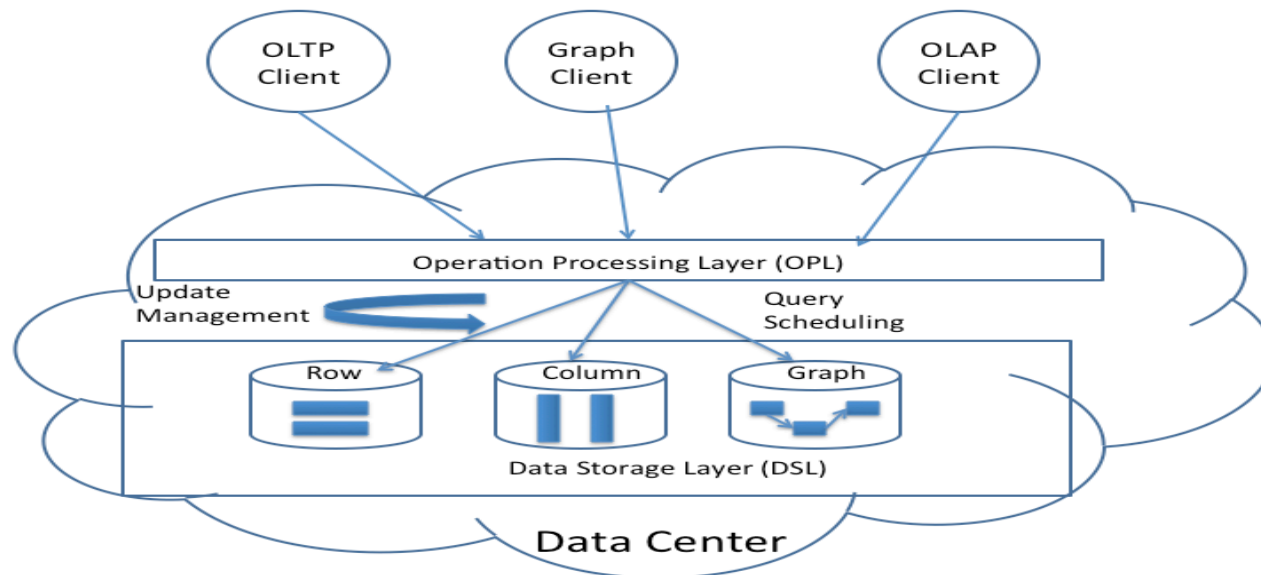
Replication driven solution

- ▶ Leverage Replication by storing replicas in **different representations**
- ▶ **Single interface** to the application developer
- ▶ Scheduling queries to the most suited representation
- ▶ **Cross representation** query processing



Replication driven solution

- ▶ Various replica classification and update models for **synchronizing** among different representations.
- ▶ Future research opportunities in designing **Update protocols** and **Query Analysis** models



Hybrid Partitioned Row Column Store

