Advanced Data Management (CSCI 490/680)

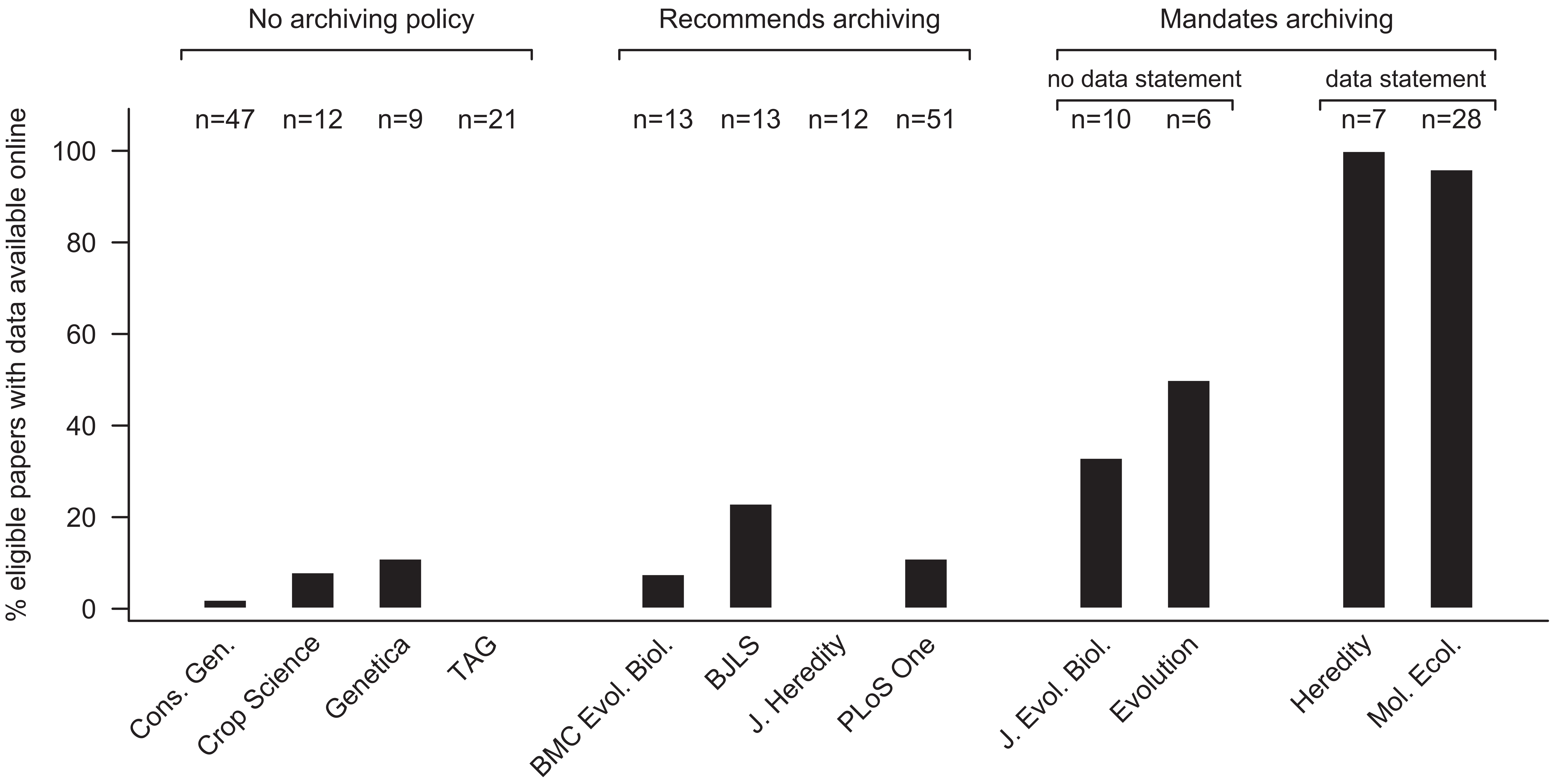
Scalable Databases

Dr. David Koop

Studying Data Availability

- Who **mandates** data sharing, and what is the impact?
  - Government
  - Funding agencies
  - Institutions
  - **Journals**
- How does the **age** of a publication/data item affect availability?
  - If not curated, how to locate?
  - What factors influence this?

Data Availability by Journal Policy



[T. Vines et al., 2013]

Data Availability by Year

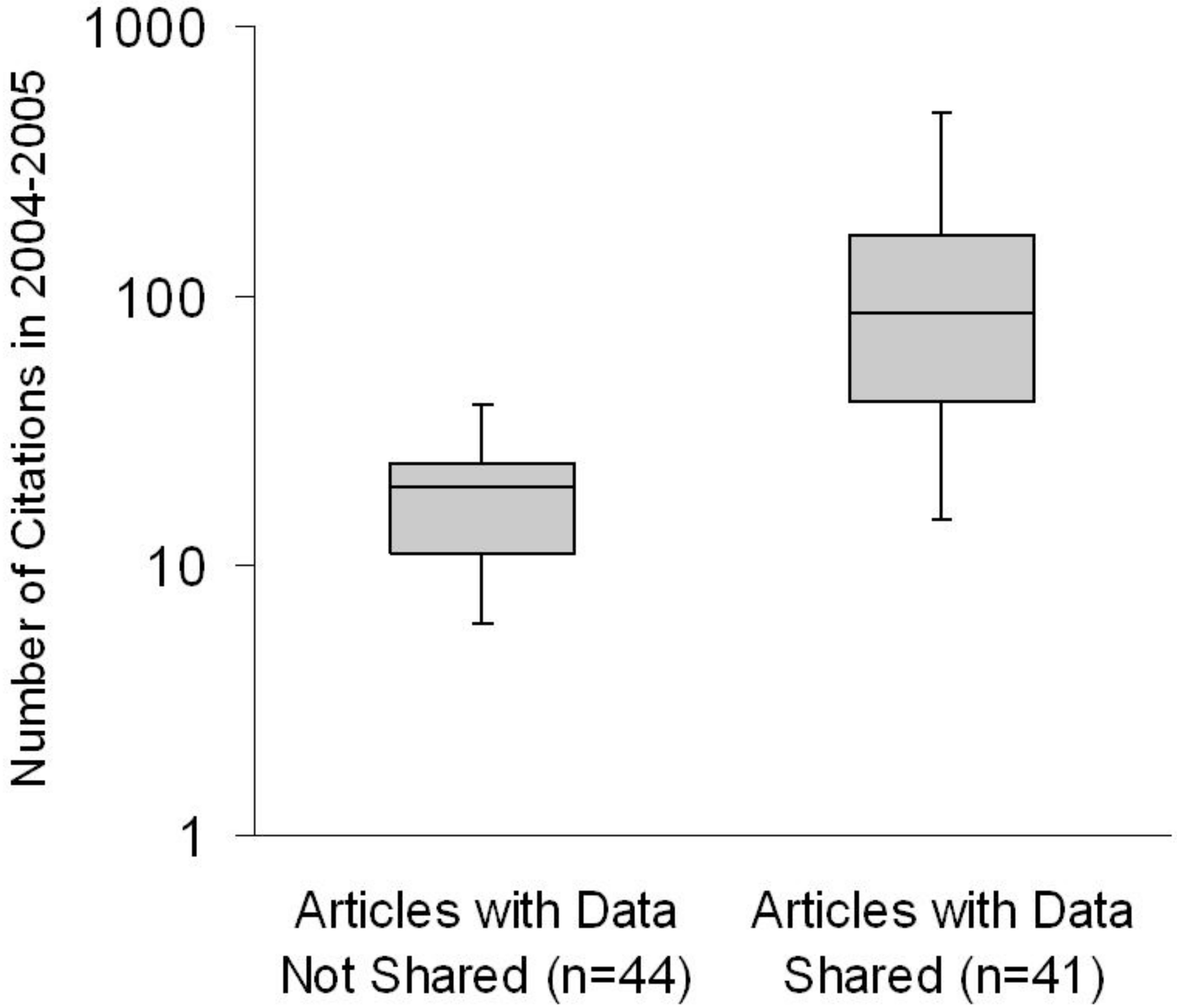
Table 1. Breakdown of Data Availability by Year of Publication

Year	No Working E-Mail	No Response to E-Mail	Response Did Not Give Status of Data	Data Lost	Data Exist, Unwilling to Share	Data Received	Data Extant (Unwilling to Share + Received)	Number of Papers
1991	9 (35%)	9 (35%)	2 (8%)	4 (15%)	1 (4%)	1 (4%)	2 (8%)	26
1993	14 (39%)	11 (31%)	3 (8%)	7 (19%)	0 (0%)	1 (3%)	1 (3%)	36
1995	11 (31%)	9 (26%)	0 (0%)	7 (20%)	2 (6%)	6 (17%)	8 (23%)	35
1997	11 (37%)	9 (30%)	1 (3%)	2 (7%)	3 (10%)	4 (13%)	7 (23%)	30
1999	19 (48%)	13 (32%)	1 (2%)	1 (2%)	0 (0%)	6 (15%)	6 (15%)	40
2001	13 (30%)	15 (35%)	3 (7%)	4 (9%)	0 (0%)	8 (19%)	8 (19%)	43
2003	9 (20%)	20 (43%)	4 (9%)	2 (4%)	0 (0%)	11 (24%)	11 (24%)	46
2005	11 (24%)	14 (31%)	6 (13%)	1 (2%)	0 (0%)	13 (29%)	13 (29%)	45
2007	12 (18%)	31 (47%)	2 (3%)	4 (6%)	1 (2%)	16 (24%)	17 (26%)	66
2009	9 (13%)	34 (49%)	3 (4%)	5 (7%)	6 (9%)	12 (17%)	18 (26%)	69
2011	13 (16%)	29 (36%)	8 (10%)	0 (0%)	7 (9%)	23 (29%)	30 (38%)	80
Totals	131 (25%)	194 (38%)	33 (6%)	37 (7%)	20 (4%)	101 (19%)	121 (23%)	516

Data are displayed as n (%); the percentages are calculated by rows.

[T. Vines et al., 2014]

Why Share Data? Increased Citations



[H. Piwowar, 2013]

What Factors Impact Sharing?

Funder
- funded by NIH?
- size of grant
- sharing plan req'd?
- funded by non-NIH?

Journal
- impact factor
- strength of policy
- open access?
- number of microarray studies published

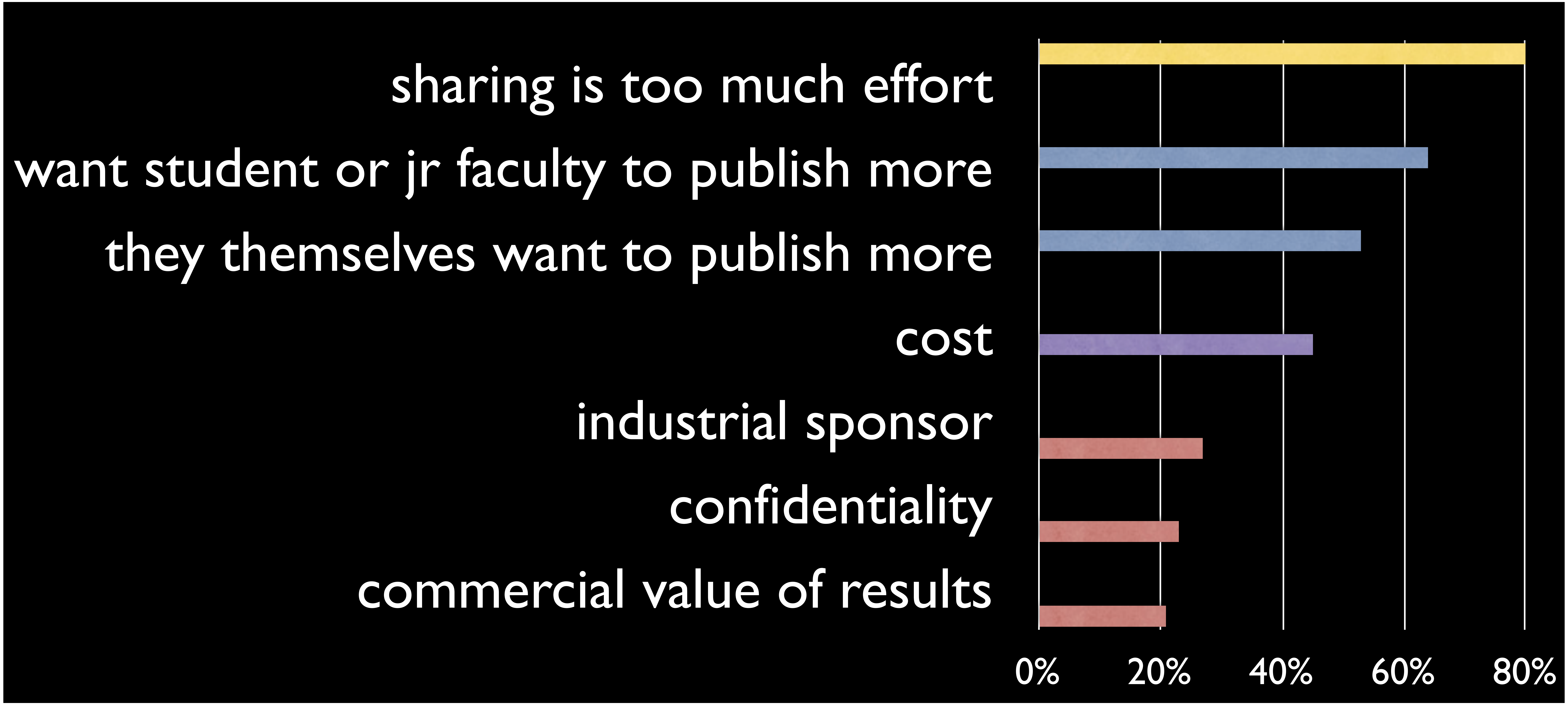
Investigator
- years since first paper
- # pubs
- # citations
- previously shared?
- previously reused?
- gender

Institution
- sector
- size
- impact rank
- country

Study
- humans?
- mice?
- plants?
- cancer?
- clinical trial?
- number of authors
- year

[H. Piwowar, 2013]

Why not data sharing? (self-reported)



[Campbell et al., 2002 via Piwowar, 2013]

Joint Declaration of Data Citation Principles

- Precursor to FAIR
- Importance: data is legitimate, citations should have importance
- Credit and Attribution: scholarly credit to all contributors
- Evidence: when data is relied on, it should be cited
- Unique Identification: machine-actionable, globally unique, and widely used
- Access: data, metadata, etc. is findable and usable
- Persistence: identifiers, metadata persist regardless of whether data does
- Specificity and Verifiability: provenance, fixity, granularity
- Interoperability and Flexibility: allow for variability across communities

Generic Data Citation

- Author(s), Year, Dataset Title, Global Persistent Identifier, Data Repository or Archive, version or subset
- Authors, repository → Principle 2
- Year and title → not related to principle but consistent with other citations
- Global Persistent Identifier: Principle 4 and 6

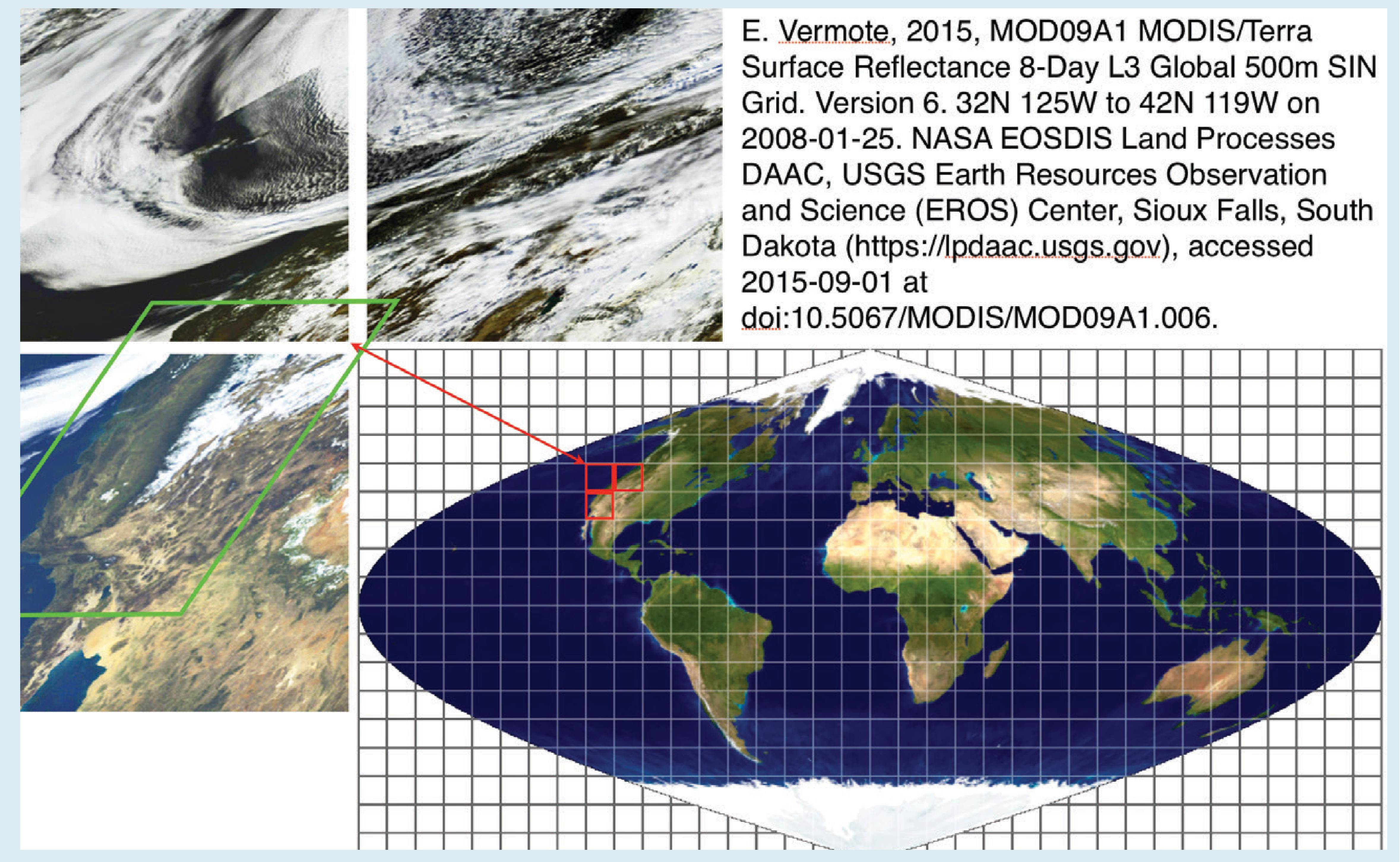
Computational Data Citation

- Given a database D and a query Q, generate an appropriate citation.
- Automatic Citation requires the answers to two questions:
  - Does the citation depend on both Q and D or just on the data Q(D) extracted by Q from D?
  - If we have appropriate citations for some queries, can we use them to construct citations for other queries?
- If the data is an image or numbers, cannot expect the citation to live in that data
- If the query returns an empty dataset, we still may wish to cite that
- People know how to cite certain parts of a dataset but not all…

[Buneman et al., 2016]

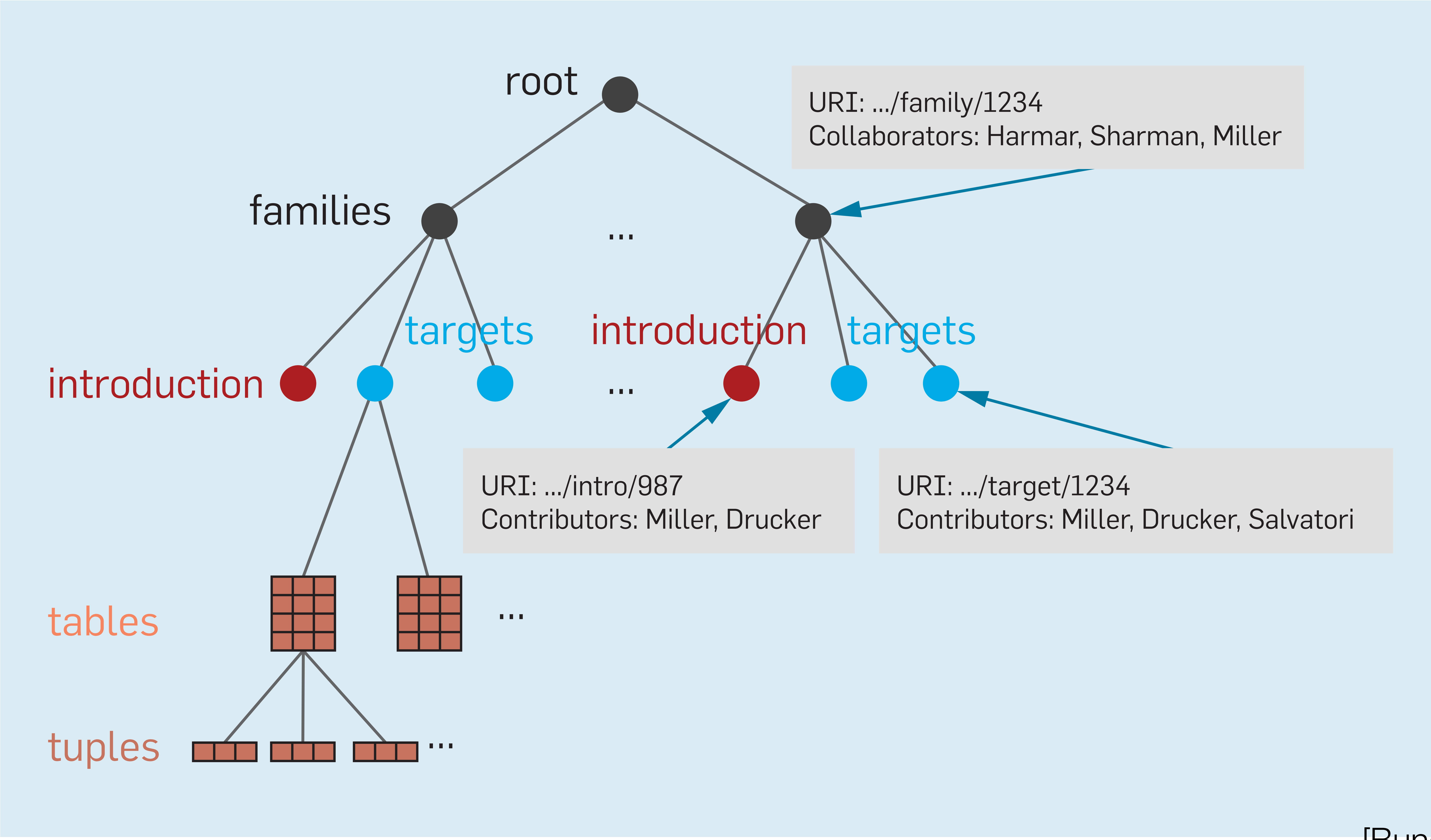
Views and Citable Units

- Views describe "areas of responsibility" for parts of a database
- Use views to create "citable units"
- Determine which view V answers a particular query Q and generate a citation for the view
- What happens if two different views can answer the same query?



[Buneman et al., 2016]

Citable Views and Partial Citations



[Buneman et al., 2016]

Next Class's Reading Response

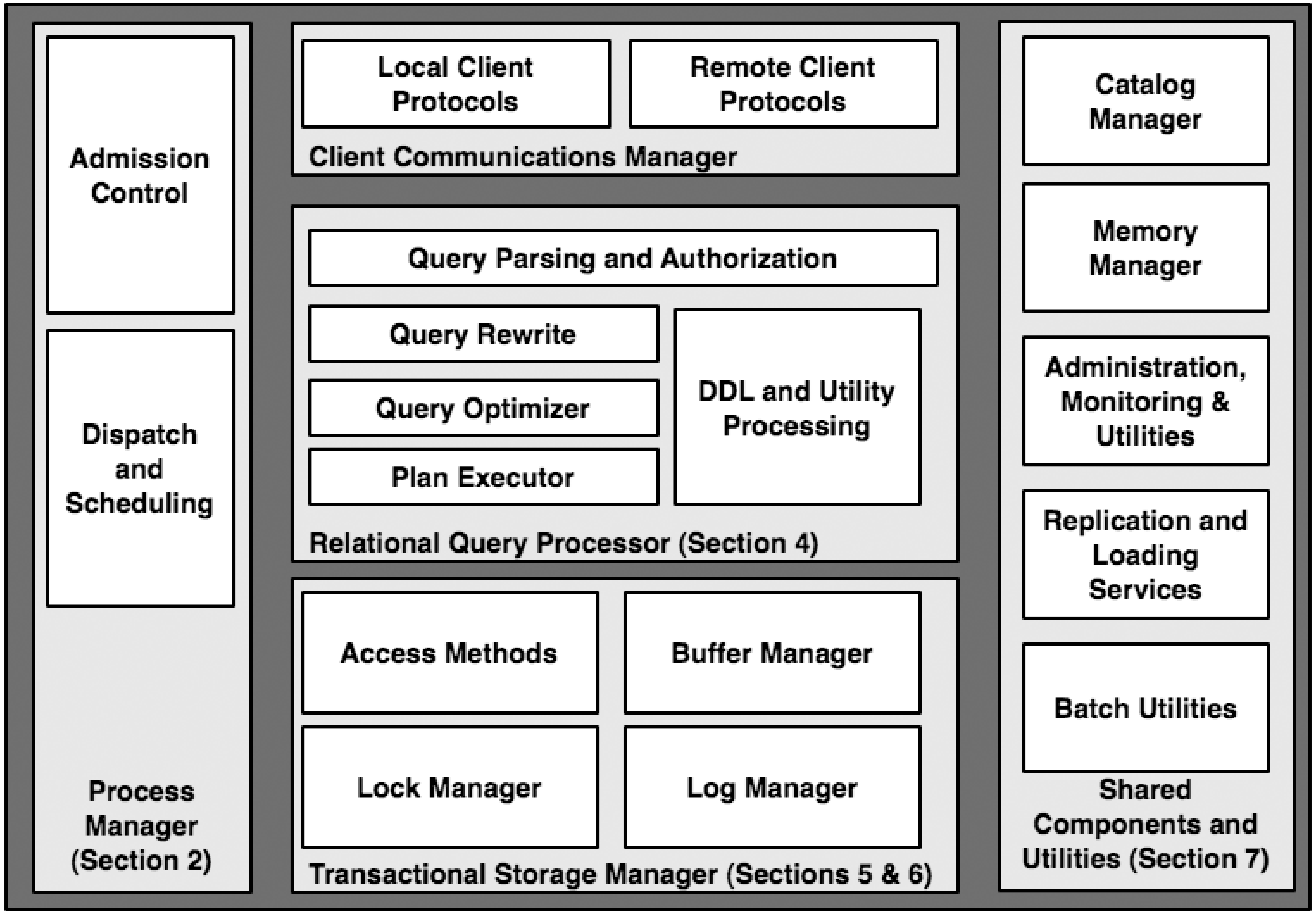
- Spanner: Google's Globally-Distributed Database
- Reading Response for Monday:
  - Focus on main concepts in the paper
  - Submit to Blackboard

Assignment 4

- World Education Data
- Collected/collated by UNESCO, World Bank, and OECD
- Transform World Bank Data
- Impute missing year data
- Integrate teacher and student numbers
- Fuse three datasets

Scalable Database Systems

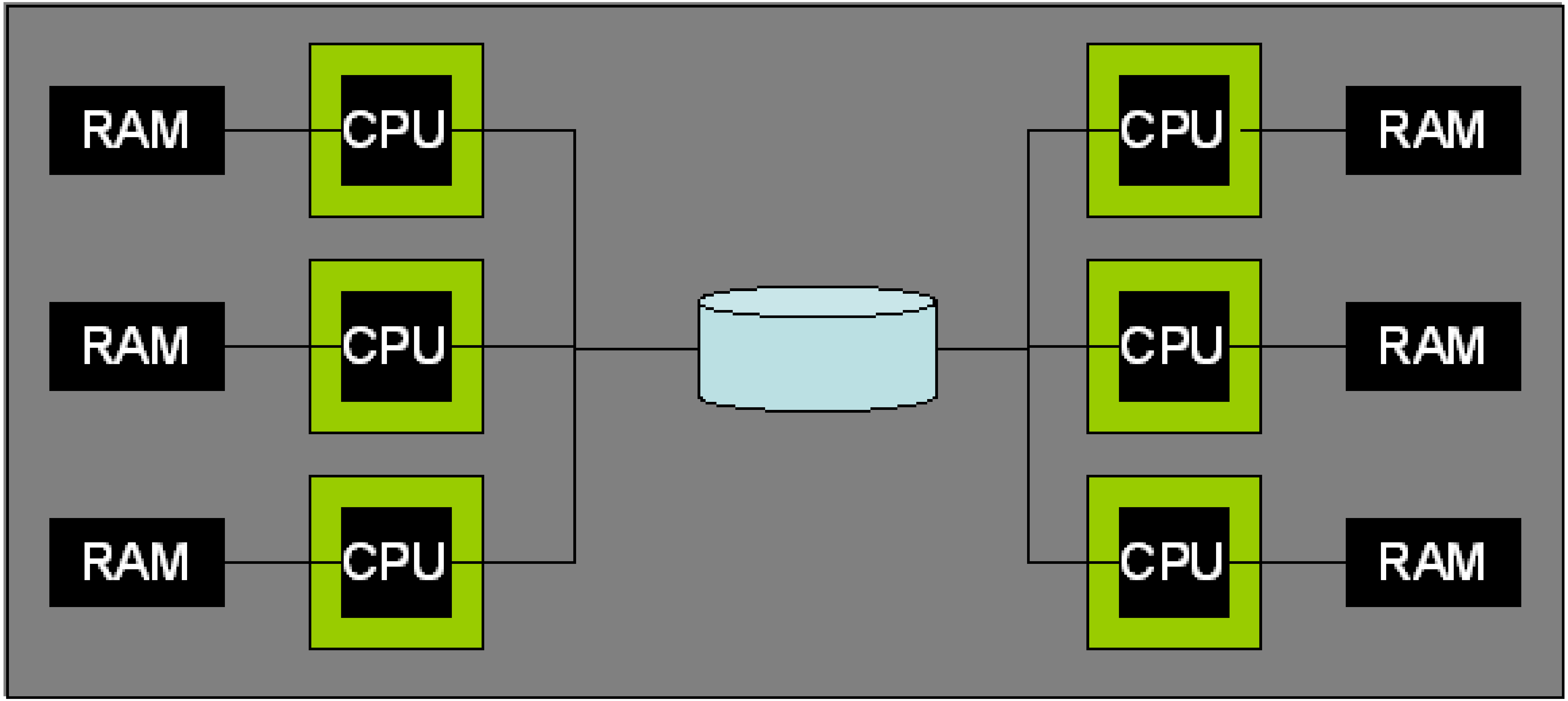
Relational Database Architecture



[Hellerstein et al., Architecture of a Database System]

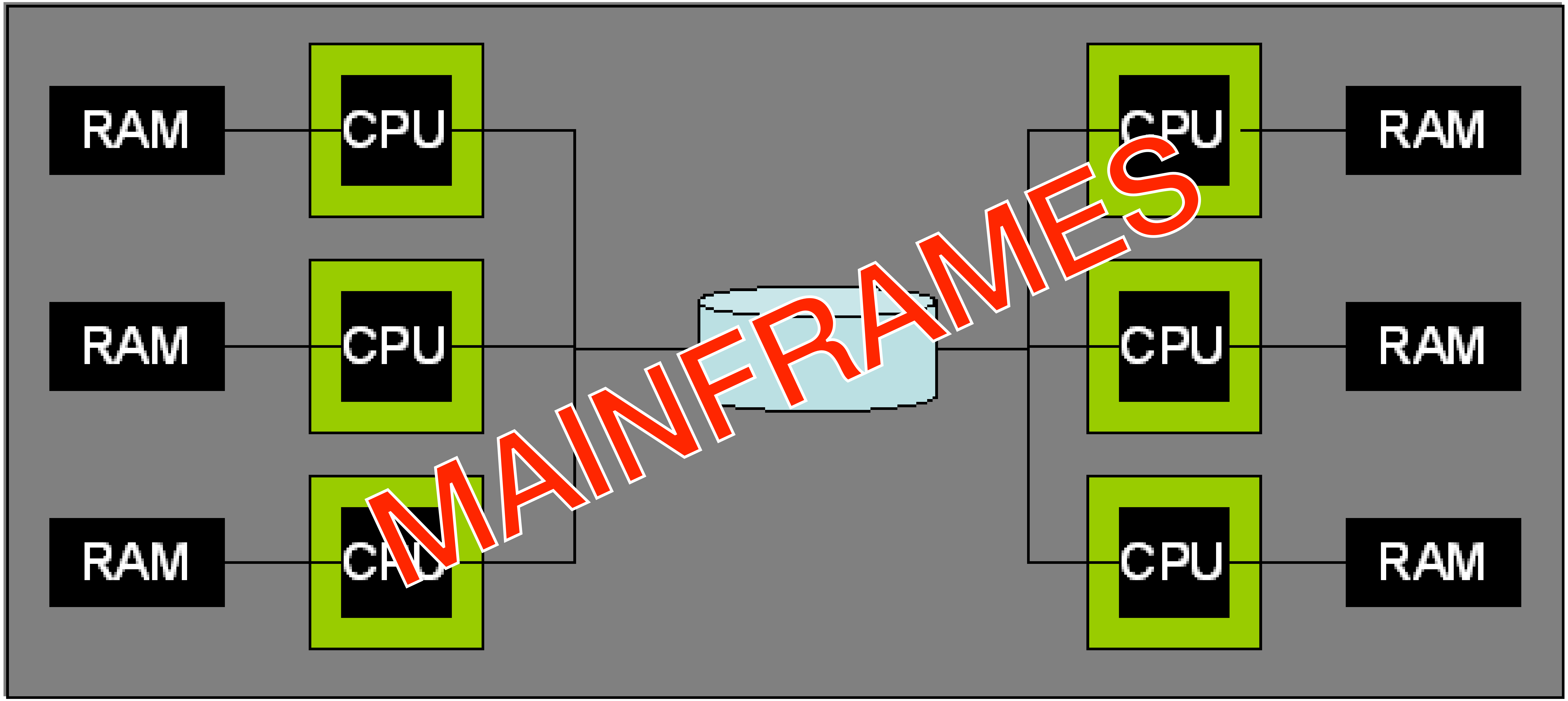
How to Scale Relational Databases?

Parallel DB Architecture: Shared Disk



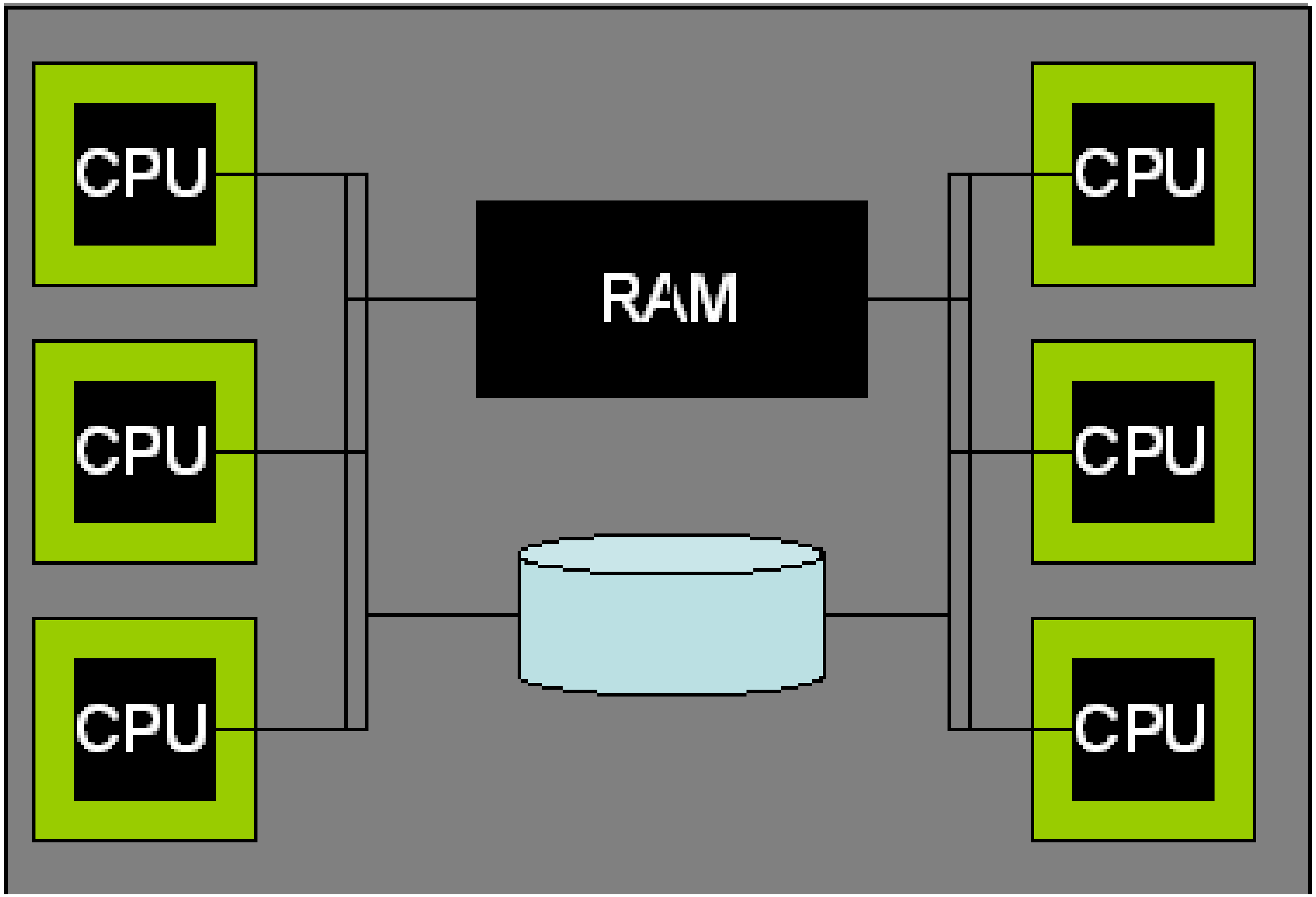
[Hellerstein et al., Architecture of a Database System]

Parallel DB Architecture: Shared Disk



[Hellerstein et al., Architecture of a Database System]

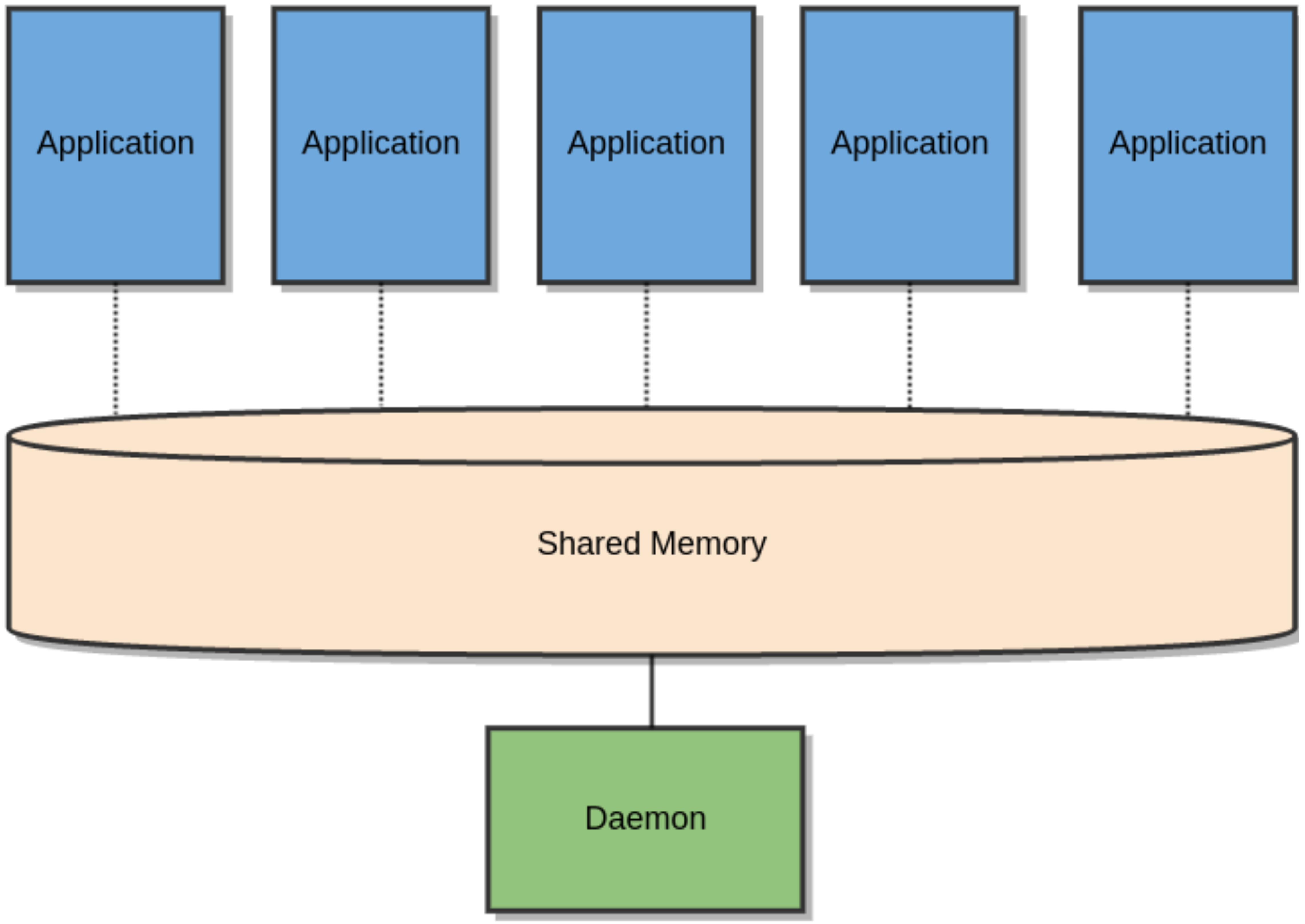
Parallel DB Architecture: Shared Memory



[Hellerstein et al., Architecture of a Database System]

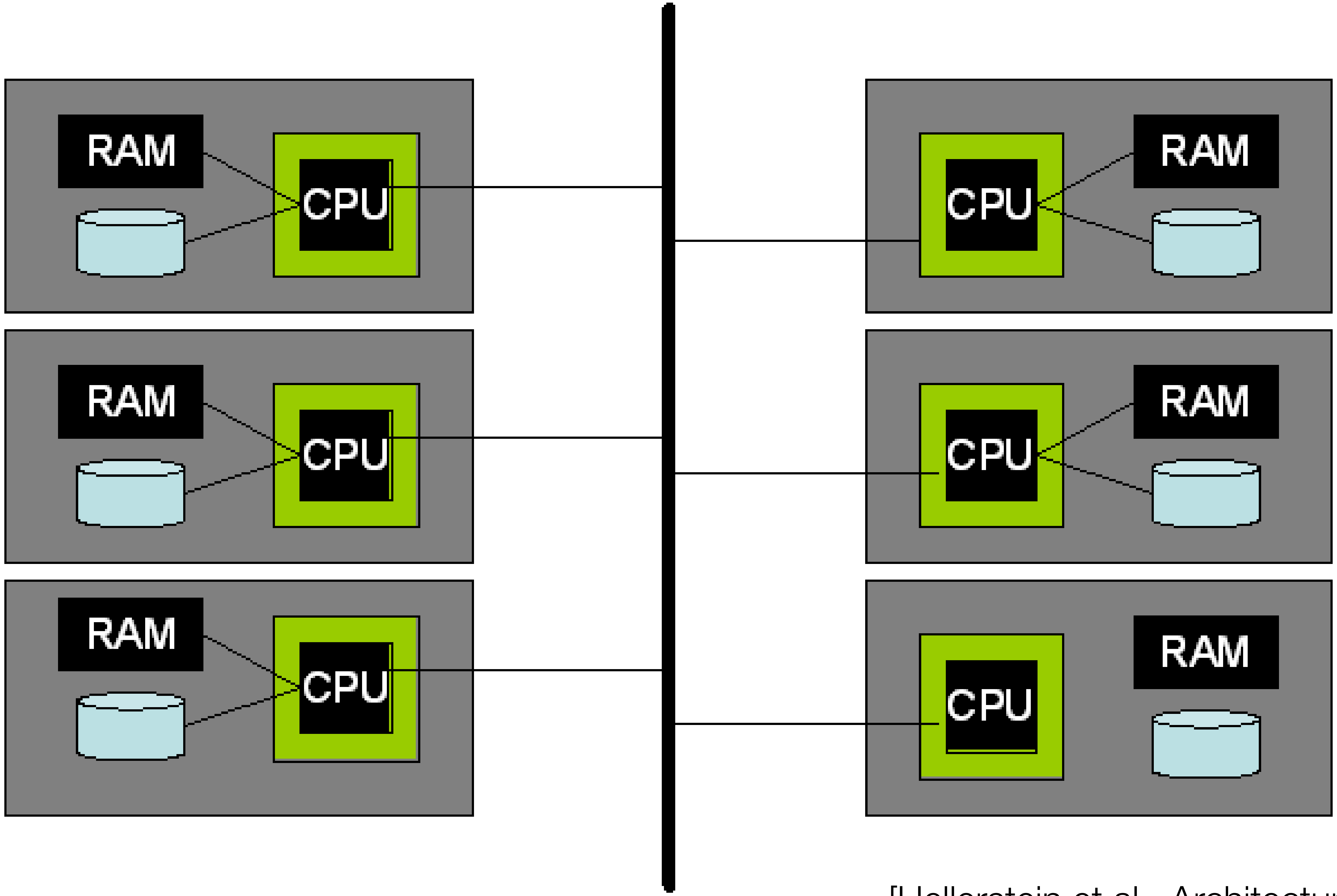
TrafficDB: Shared-Memory Data Store

- Traffic-aware route planning
- Want up-to-date data for all
- Thousands of requests per second
  - High-Frequency Reads
  - Low-Frequency Writes
- "Data must be stored in a region of RAM that can be shared and efficiently accessed by *several* different application processes"



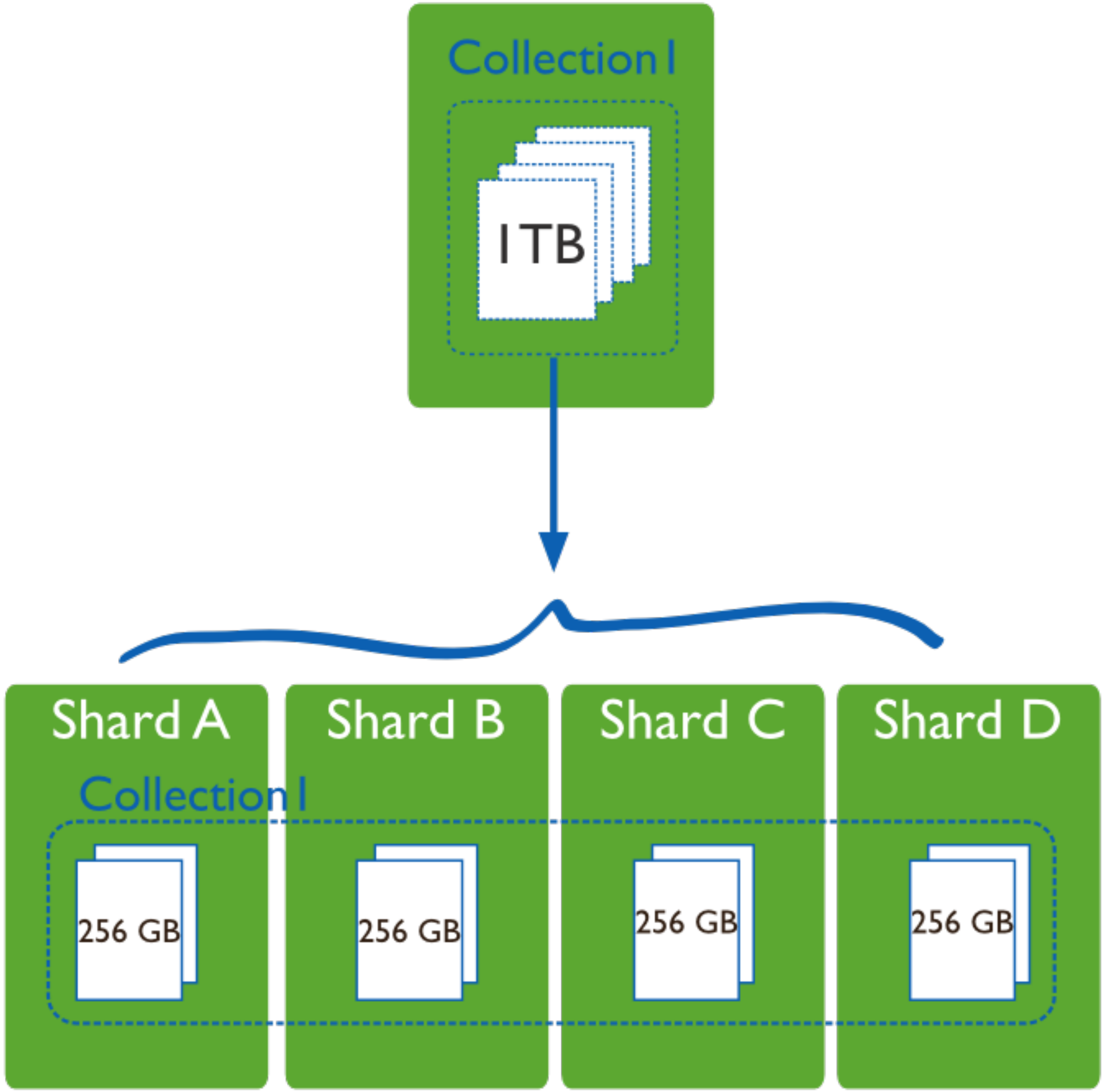
[R. Fernandes et al., 2016]

Parallel DB Architecture: Shared Nothing



[Hellerstein et al., Architecture of a Database System]

Sharding



[MongoDB]

Relational Databases: One size fits all?

- Lots of work goes into relational database development:
  - B-trees
  - Cost-based query optimizers
  - ACID (Atomicity, Consistency, Isolation, Durability)
- Vendors have stuck with this model since the 1980s
- Having different systems leads to business problems:
  - cost problem
  - compatibility problem
  - sales problem
  - marketing problem

[Stonebraker and Çetinetmel, 2005]

ACID Transactions

- Make sure that transactions are processed reliably
- **A**tomicity: leave the database as is if some part of the transaction fails (e.g. don't add/remove only part of the data) using rollbacks
- **C**onsistency: database moves from one valid state to another
- **I**solation: concurrent execution matches serial execution
- **D**urability: endure hardware failures, make sure changes hit disk

Stonebraker: The End of an Architectural Era

- "RDBMSs were designed for the business data processing market, which is their sweet spot"
- "They can be beaten handily in most any other market of significant enough size to warrant the investment in a specialized engine"
- Changes in markets (science), necessary features (scalability), and technology (amount of memory)
- RDBMS Overhead: Logging, Latching, and Locking
- Relational model is not necessarily the answer
- SQL is not necessarily the answer

Row Stores

Primary Key

Row

id	scientist	death_by	movie_name
1	Reinhardt	Crew	The Black Hole
2	Tyrell	Roy Batty	Blade Runner
3	Hammond	Dinosaur	Jurassic Park
4	Soong	Lore	Star Trek: TNG
5	Morbius	The machine	Forbidden Planet
6	Dyson	SWAT	Terminator 2: Judgment Day

[J. Swanhart, Introduction to Column Stores]

OLTP vs. OLAP

- Online Transactional Processing (OLTP) often used in business applications, data entry and retrieval transactions
- OLTP Examples:
  - Add customer's shopping cart to the database of orders
  - Find me all information about John Hammond's death
- OLTP is focused on the day-to-day operations while Online Analytical Processing (OLAP) is focused on analyzing that data for trends, etc.
- OLAP Examples:
  - Find the average amount spent by each customer
  - Find which year had the most movies with scientists dying

Inefficiency in Row Stores for OLAP

select sum(metric) as the_sum from fact

1. Storage engine gets *a whole row* from the table

6	15	on_hold	247	122	9	72	76	5	66

↓

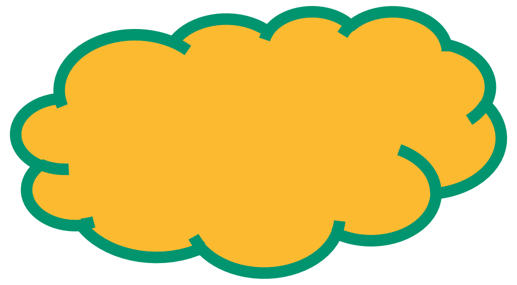
2. SQL interface extracts only requested portion, adds it to “the_sum”

247

↓

3. IF all rows scanned, send results to client, else GOTO 1

↓



[J. Swanhart, Introduction to Column Stores]

Column Stores

row id = 1

row id = 6

id	Title	Person	Genre
1	Mrs. Doubtfire	Robin Williams	Comedy
2	Jaws	Roy Scheider	Horror
3	The Fly	Jeff Goldblum	Horror
4	Steel Magnolias	Dolly Parton	Drama
5	The Birdcage	Nathan Lane	Comedy
6	Erin Brokovitch	Julia Roberts	Drama

Each column has a file or segment on disk

[J. Swanhart, Introduction to Column Stores]

Horizontal Partitioning vs. Vertical Partitioning

Original Table

CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
1	TAEKO	OHNUKI	BLUE
2	O.V.	WRIGHT	GREEN
3	SELDA	BAĞCAN	PURPLE
4	JIM	PEPPER	AUBERGINE

[M. Drake]

Horizontal Partitioning vs. Vertical Partitioning

Vertical Partitions

VP1

CUSTOMER ID	FIRST NAME	LAST NAME
1	TAEKO	OHNUKI
2	O.V.	WRIGHT
3	SELDA	BAĞCAN
4	JIM	PEPPER

VP2

CUSTOMER ID	FAVORITE COLOR
1	BLUE
2	GREEN
3	PURPLE
4	AUBERGINE

Original Table

CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
1	TAEKO	OHNUKI	BLUE
2	O.V.	WRIGHT	GREEN
3	SELDA	BAĞCAN	PURPLE
4	JIM	PEPPER	AUBERGINE

Horizontal Partitions

HP1

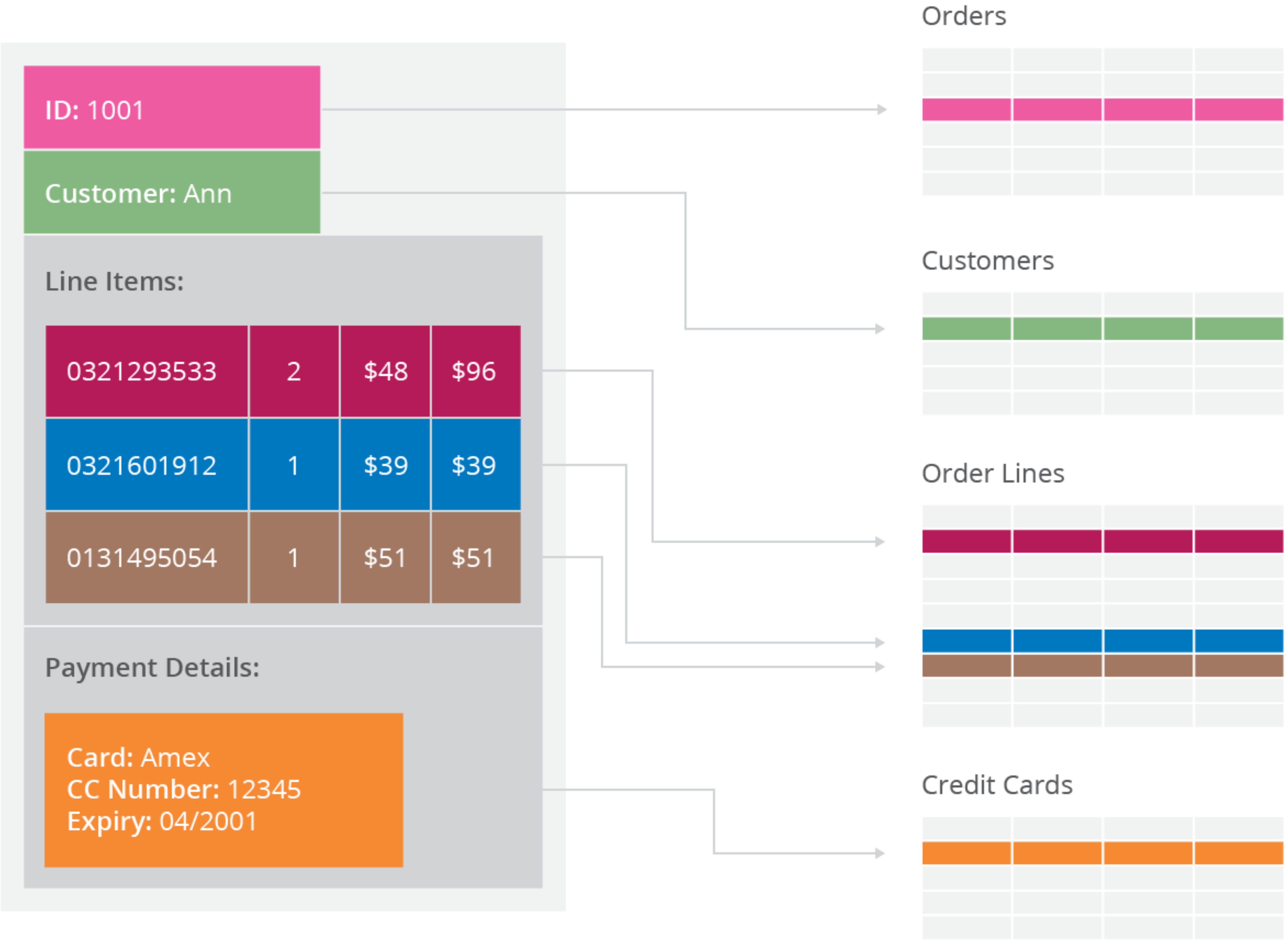
CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
1	TAEKO	OHNUKI	BLUE
2	O.V.	WRIGHT	GREEN

HP2

CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
3	SELDA	BAĞCAN	PURPLE
4	JIM	PEPPER	AUBERGINE

[M. Drake]

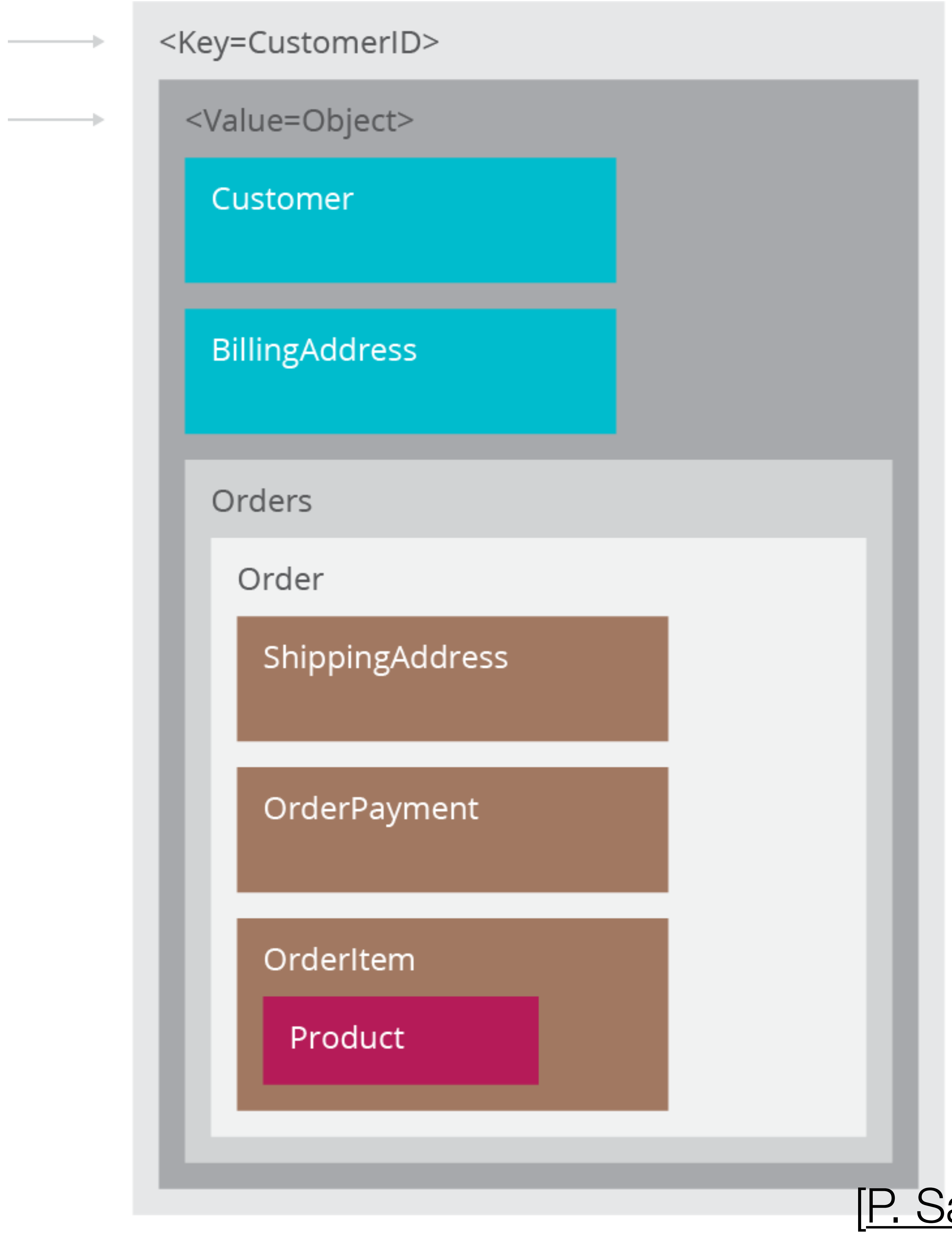
Problems with Relational Databases



[P. Sadalage]

NoSQL: Key-Value Databases

- Always use primary-key access
- Operations:
  - Get/put value for key
  - Delete key
- Examples
  - Memcached
  - Amazon DynamoDB
  - Project Voldemort
  - Couchbase



[P. Sadalage]

NoSQL: Document Databases

- Documents are the main entity
  - Self-describing
  - Hierarchical
  - Do not have to be the same
- Could be XML, JSON, etc.
- Key-value stores where values are "examinable"
- Can have query language and indices overlaid
- Examples: MongoDB, CouchDB, Terrastore

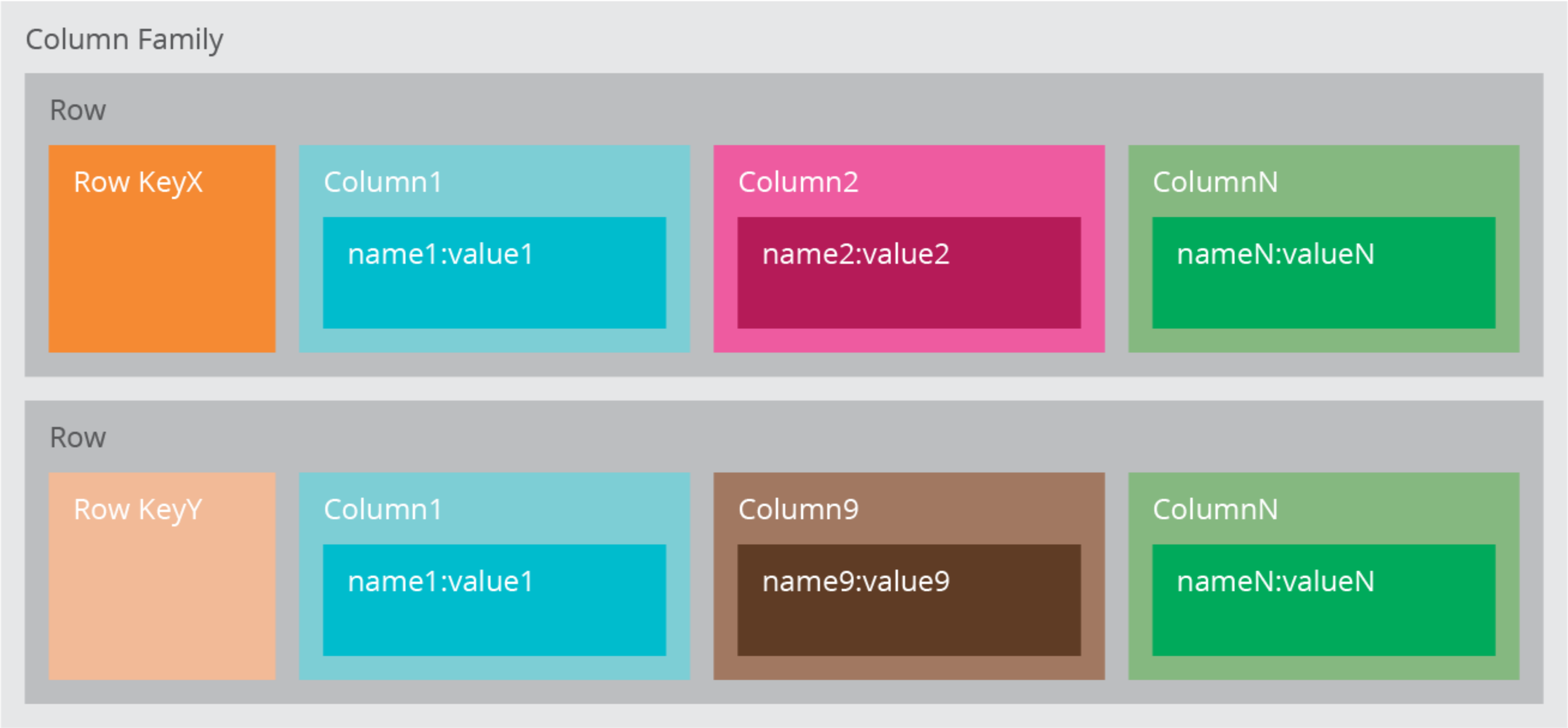
<Key=CustomerID>

```
{
    "customerid": "fc986e48ca6"
    "customer":
    {
    "firstname": "Pramod",
    "lastname": "Sadalage",
    "company": "ThoughtWorks",
    "likes": [ "Biking","Photography" ]
    }
    "billingaddress":
    { "state": "AK",
        "city": "DILLINGHAM",
        "type": "R"
      }
}
```

[P. Sadalage]

NoSQL: Column Stores

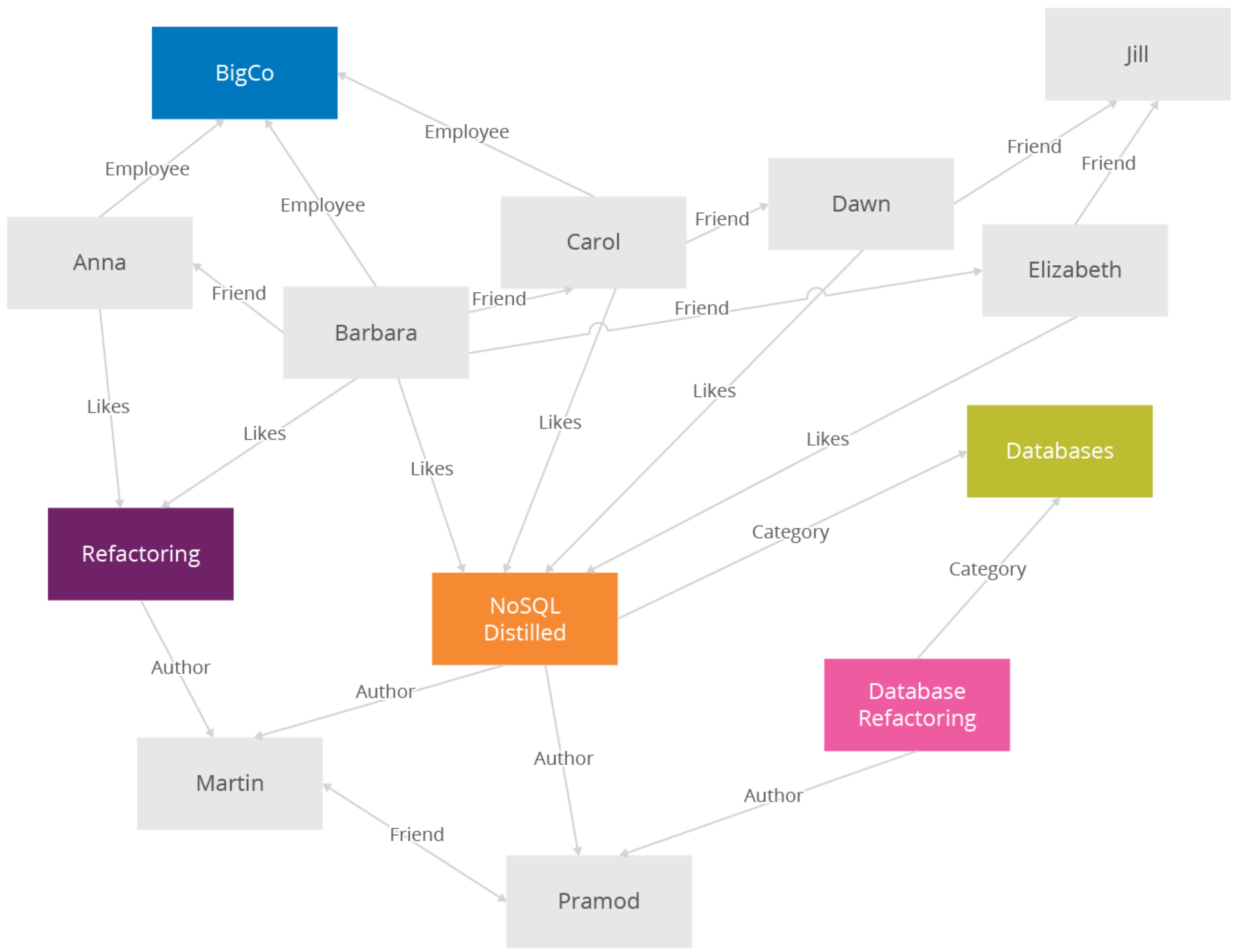
- Instead of having rows grouped/sharded, we group columns
- …or families of columns
- Put similar columns together
- Examples: Cassandra, HBase



[P. Sadalage]

NoSQL: Graph Databases

- Focus on entities and relationships
- Edges may have properties
- Relational databases required a set traversal
- Traversals in Graph DBs are faster
- Examples:
  - Neo4j
  - Pregel



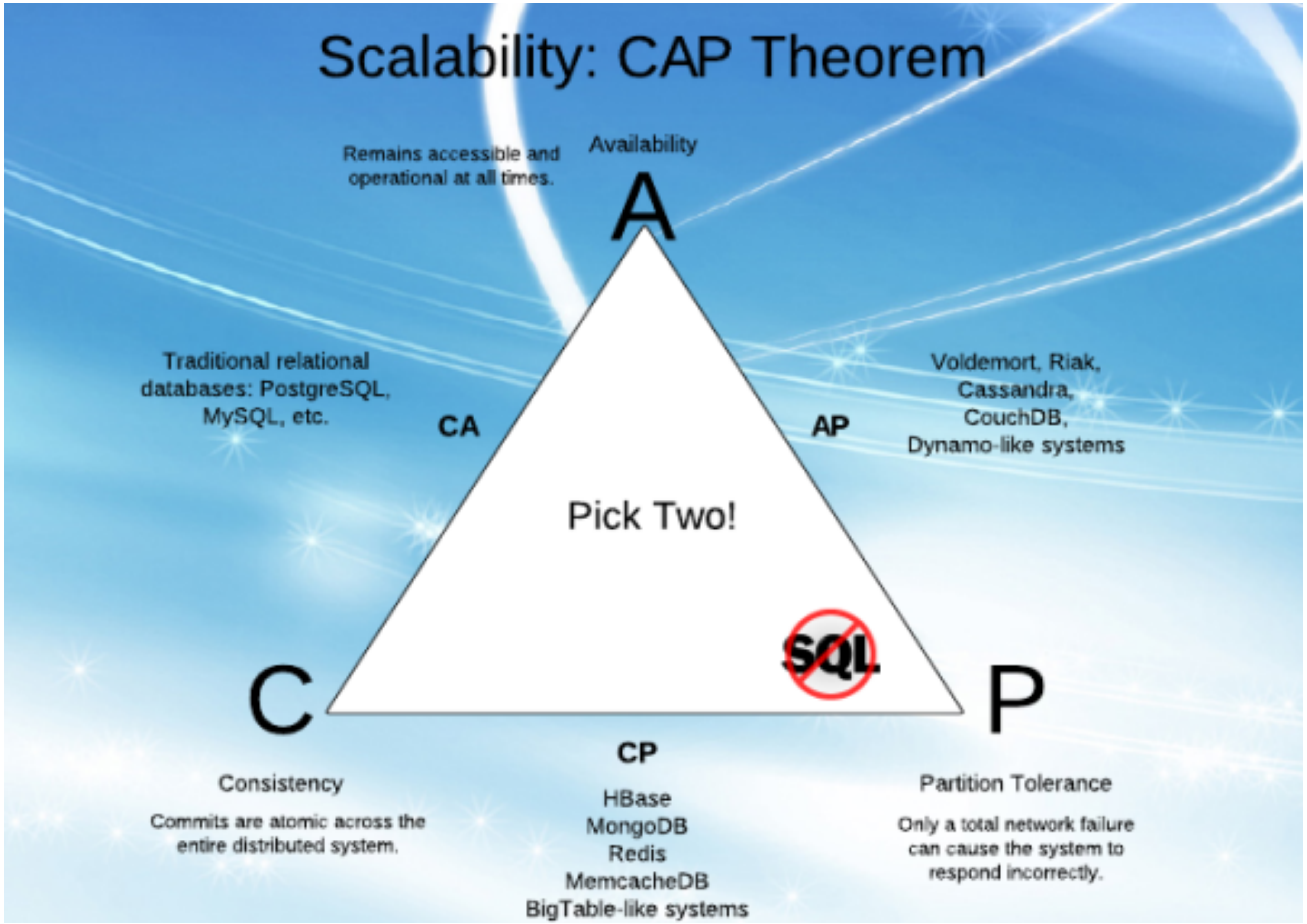
[P. Sadalage]

Distributing Data

- Aggregate-oriented databases
- Sharding (horizontal partitioning): Sharding distributes different data across multiple servers, so each server acts as the single source for a subset of data
- Replication: Replication copies data across multiple servers, so each bit of data can be found in multiple places. Replication comes in two forms,
  - Source-replica replication makes one node the authoritative copy that handles writes, replica synchronizes with the source and may handle reads.
  - Peer-to-peer replication allows writes to any node; the nodes coordinate to synchronize their copies of the data.

[P. Sadalage]

CAP Theorem



[E. Brewer]

CAP Theorem

- Consistency: every read would get you the most recent write
- Availability: every node (if not failed) always executes queries
- Partition tolerance: system continues to work even if nodes are down
- Theorem (Brewer): It is impossible for a distributed data store to simultaneously provide more than two of Consistency, Availability, and Partition Tolerance

Think about RDBMS Transactions…

Cassandra:
A Decentralized Structured Storage System

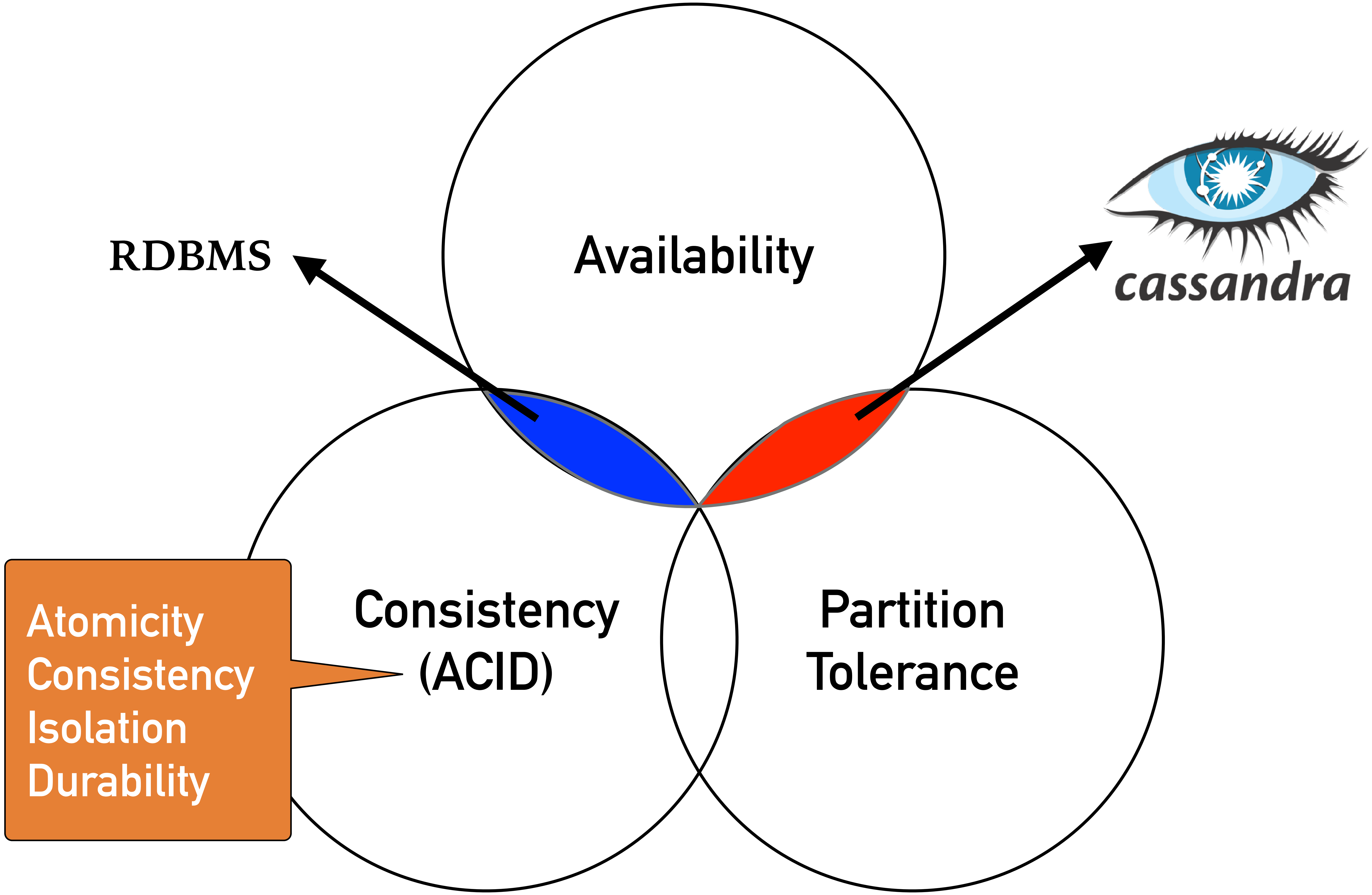
A. Lakshman and P. Malik

What is Cassandra?

- Fast Distributed (Column Family NoSQL) Database
  - High availability
  - Linear Scalability
  - High Performance
- Fault tolerant on Commodity Hardware
- Multi-Data Center Support
- Easy to operate
- Proven: CERN, Netflix, eBay, GitHub, Instagram, Reddit

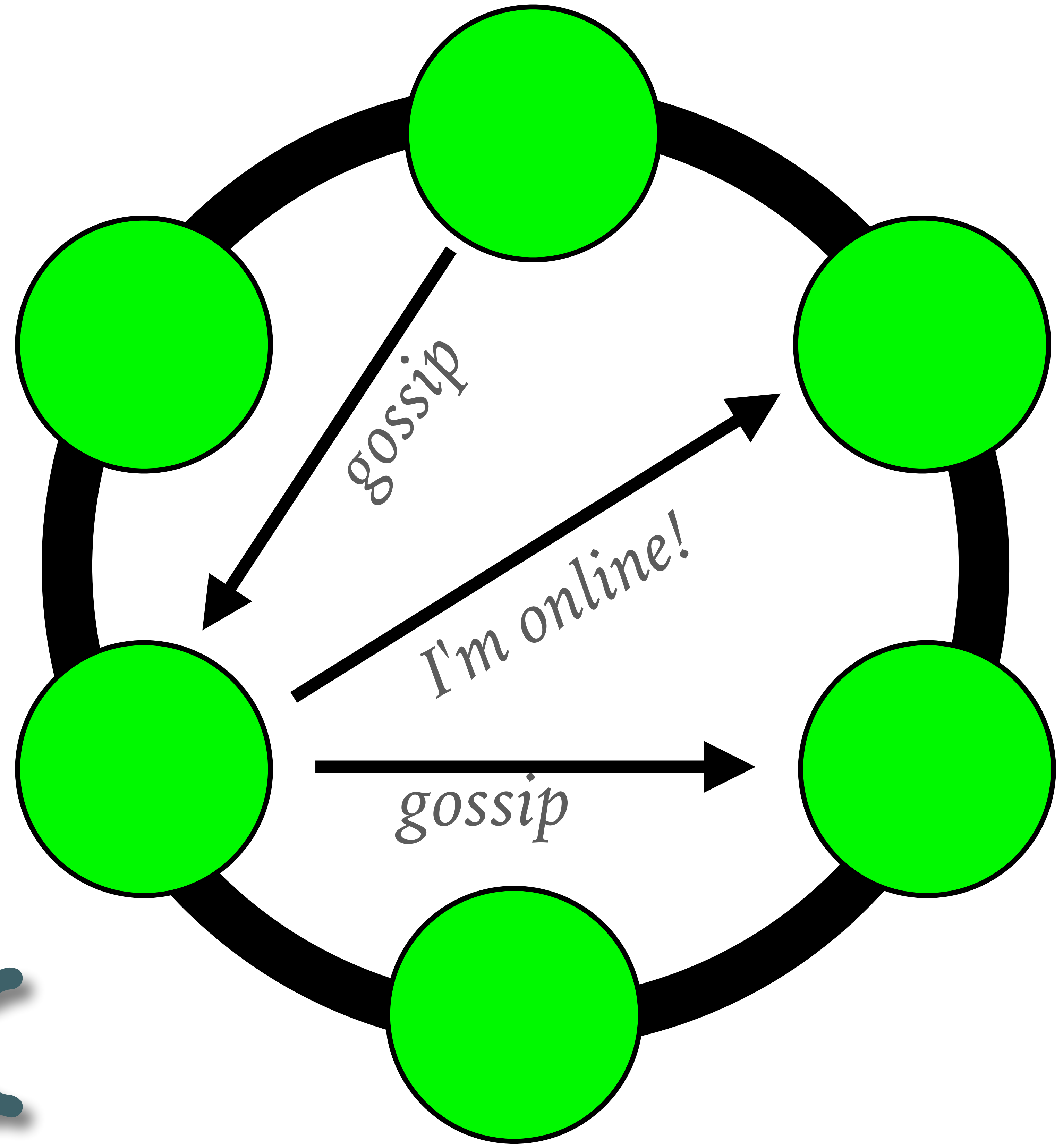
[G. Atil]

Cassandra and CAP



[G. Atil]

Cassandra: Ring for High Availability



[G. Atil]

Slides: Introduction to Cassandra

Robert Stupp

Next Class's Reading Response

- Spanner: Google's Globally-Distributed Database
- Reading Response for Monday:
  - Focus on main concepts in the paper
  - Submit to Blackboard