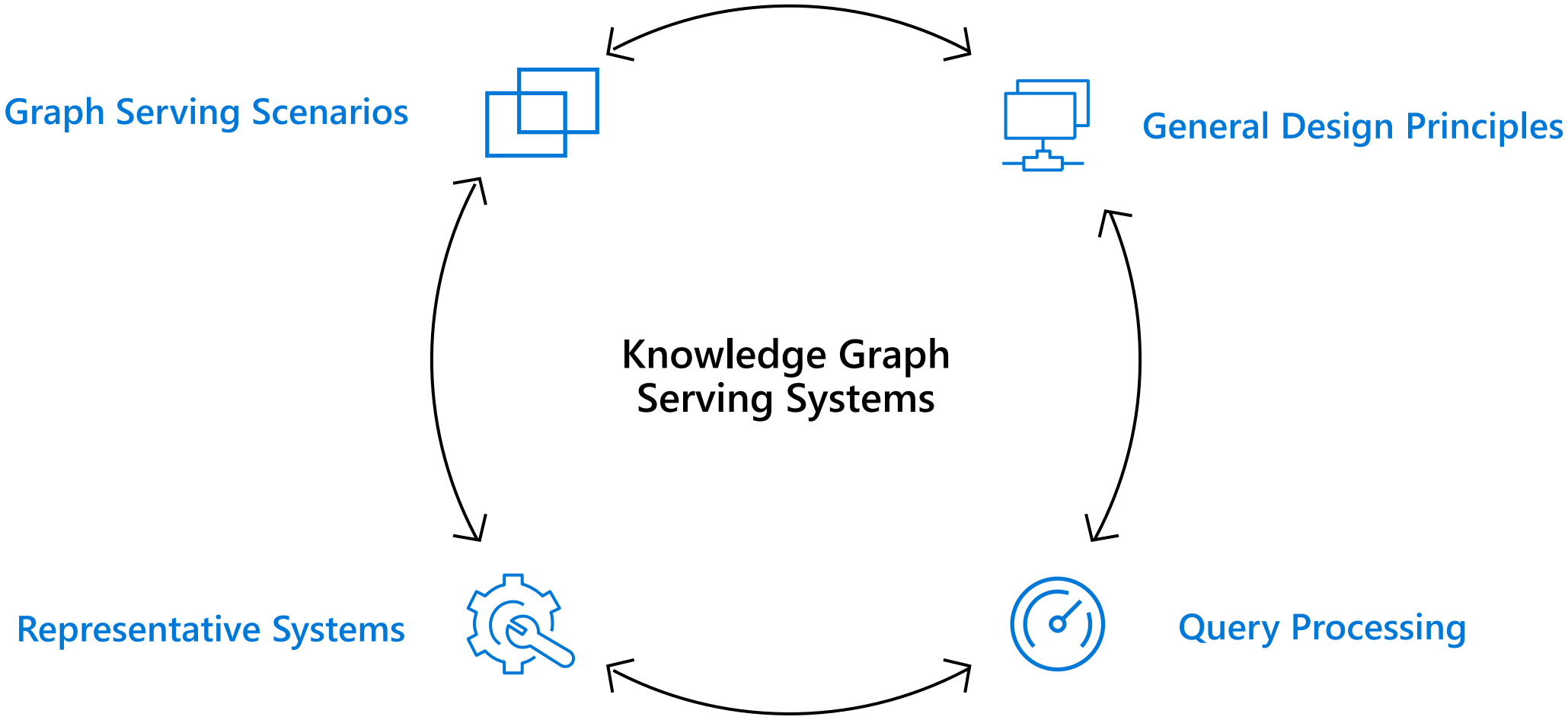
《知识图谱: 概念与技术》

第 8 讲
知识图谱管理系统

邵斌

微软亚洲研究院

binshao@microsoft.com


Graph Serving Scenarios
General Design Principles
Knowledge Graph
Serving Systems
Representative Systems
Query Processing

Outline

- Knowledge graph serving scenarios
- General design principles of knowledge graph serving systems
- Real-time query processing
- Representative graph systems
- Demo

Knowledge Serving Scenarios

A real-life relation search scenario

A News Headline

Tom Cruise Admits **Katie Holmes** Divorced Him To Protect **Suri** From Scientology

1. **Tom Cruise** – people.person.marriage – (marriage) – time.event.person – **Katie Holmes**
2. **Tom Cruise** – people.person.children – (**Suri Cruise**) – people.person.parent – **Katie Holmes**
3. **Tom Cruise** – film.actor.film – (Bambi Verleihung 2007) – film.filmactor – **Katie Holmes**
4. ...

Relation search in knowledge graph

Entity A $\cdots\rightsquigarrow$ **Entity B**

Multi-hop Relation Search

- Discover the **hidden relations** between entities
- Enable more than what entity indexes can support

Search results of Google

Google Tom Cruise, Katie Holmes

Web News Images Videos Shopping More Search tools

About 19,600,000 results (0.40 seconds)

Tom Cruise Admits **Katie Holmes** Divorced Him To Protect ...

www.huffingtonpost.com/.../**tom-cruise-katie-holmes**-protect-su...

by Stephanie Marcus

Nov 8, 2013 - **Tom Cruise** has admitted that **Katie Holmes** filed for divorce in part because of his involvement in the controversial Church of Scientology.

Images for **Tom Cruise, Katie Holmes** Report images

More images for **Tom Cruise, Katie Holmes**

Tom Cruise Comes Clean on Role of Scientology in Divorce ...

abcnews.go.com › Entertainment ABC News

Nov 9, 2013 - Amidst his court battle against tabloid headlines, **Tom Cruise** admitted that ex-wife **Katie Holmes** filed for divorce "to protect Suri from ...

Tom Cruise admits **Katie Holmes** left to protect Suri from ...

www.nydailynews.com/.../**tom-cruise**-ad... New York Daily News

by Bill Hutchinson - in 29 Google+ circles

Nov 7, 2013 - **Tom Cruise** has admitted in an explosive court deposition that actress **Katie Holmes** fled their marriage to protect their daughter from ...

Search results of Bing

bing | Tom Cruise, Katie Holmes

MS Beta 4,340,000 RESULTS Any time

News about Tom Cruise, Katie Holmes

bing.com/news

KATIE HOLMES DATING JAMIE FOXX RUMORS CONTINUE AS THE ACTRESS' EX-HUSBAND, **TOM CRUISE** WAS REPORTED TO HAVE FINALLY MOVED ON
Travelers Today · 3 days ago
Katie Holmes dating rumors again sparked as her ex-husband **Tom Cruise** was reportedly dating other woman and that…

Is **Tom Cruise** Dating Laura Prepon - **Katie Holmes** Ex Lands Scientologist Girlfriend?
The National Ledger · 10 days ago

Katie Holmes Celebrates Suri **Cruise**'s 8th Birthday
WebProNews · 3 days ago

Images of Tom Cruise, Katie Holmes

bing.com/images

Katie Holmes Celebrates Suri Cruise's 8th Birthday ...

www.webpronews.com/**katie**-**holmes**-celebrates-suri-**cruises**-8th...
Katie Holmes helped daughter Suri **Cruise** celebrate her 8th birthday in style. She treated her daughter, along with a few guests, to dinner at Nobu Next …

Tom Cruise: Katie Holmes Divorce Was A Surprise (UPDATE)

www.huffingtonpost.com/2013/04/09/**tom-cruise-katie-holmes**-divorce...
Apr 09, 2013 · **Tom Cruise** says **Katie Holmes** divorce was a surprise. Here, the former couple is pictured at the "Mission Impossible: Ghost Protocol" premiere in Dec. 2011.

See results for

Katie Holmes
American Actress
Kate Noelle "Katie" Holmes is an American actress and model who first achieved fame for her role as Joey Pot…

Tom Cruise
Film Actor
Tom Cruise, is an American film actor and producer. He has been nominated for three Academy Awards and h…

Related searches

Tom Cruise Katie Holmes **Married**
Tom Cruise Katie Holmes **Gossip**
Tom Cruise Katie Holmes **Photos**
Tom Cruise Katie Holmes **Baby**
Tom Cruise Katie Holmes **Unusual Marriage**
Katie Holmes Tom Cruise **Split**
Tom Cruise Katie Holmes **Suri Custody Settlement**
Leah Remini Problems Started Tom Cruise **Wedding**

Relation search in knowledge graph

Satori [ ] Add Search

Tom Cruise, Mimi Rogers, Nicole Kidman, Katie Holmes

Results View

94 Results (103 ms)

Results
o--film.actor.film-->(Eyes Wide Shut)--film.film.actor-->(Nicole Kidman)
o--film.actor.film-->(National Movie Awards)--film.film.actor-->(Katie Holmes)
o--film.actor.film-->(InStyle: Celebrity Weddings)--film.film.actor-->(Katie Holmes)
o--people.person.marriage-->(marriage)--time.event.person-->(Katie Holmes)
o--people.person.marriage-->(marriage)--time.event.person-->(Nicole Kidman)
o--film.actor.film-->(War of the Worlds: UK Premiere Special)--film.film.actor-->(Katie Holmes)
o--film.producer.film-->(The Others)--award.nominated_work.nomination-->(nomination)--award.nomination.nominee--(Nicole Kidman)
o--people.person.children-->(Connor Cruise)--people.person.siblings-->(Isabella Jane Cruise)--people.person.parent--(Nicole Kidman)
o--film.producer.film-->(The Others)--award.nominated_work.nomination-->(nomination)--award.nomination.nominee--(Nicole Kidman)
o--film.actor.performance-->(performance)--film.performance.film-->(Eyes Wide Shut)--film.film.actor--(Nicole Kidman)

Prev Page Next Page

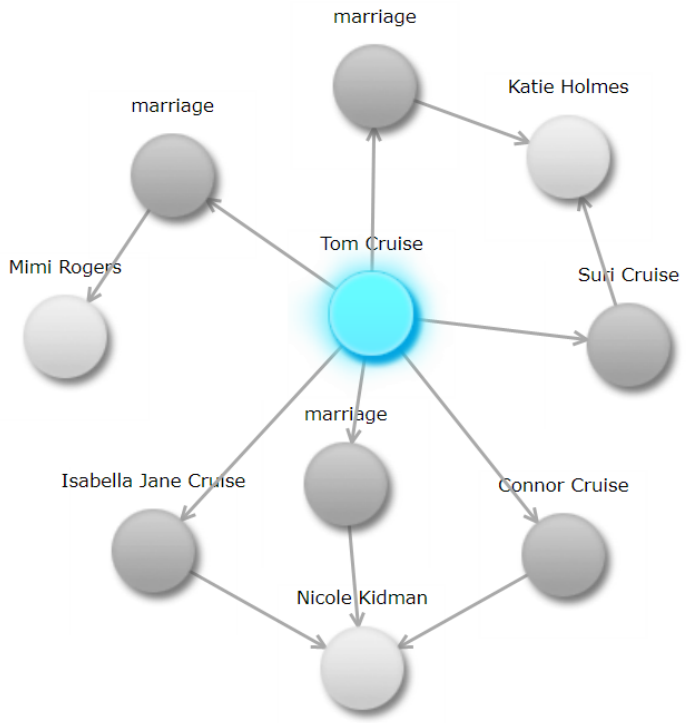
Relation search in knowledge graph

Satori

Add Search

Tom Cruise, Mimi Rogers, Nicole Kidman, Katie Holmes

Results View



Tom Cruise

Tom Cruise (born Thomas Cruise Mapother IV; July 3, 1962), is an American film actor and producer. He has been nominated for three Academy Awards and has won three Golden Globe Awards. He started his career at age 19 in the 1981 film Endless Love. After portraying supporting roles in Taps (1981) and The Outsiders (1983), his first leading role was in Risky Business, released in August 1983. Cruise became a full-fledged movie...

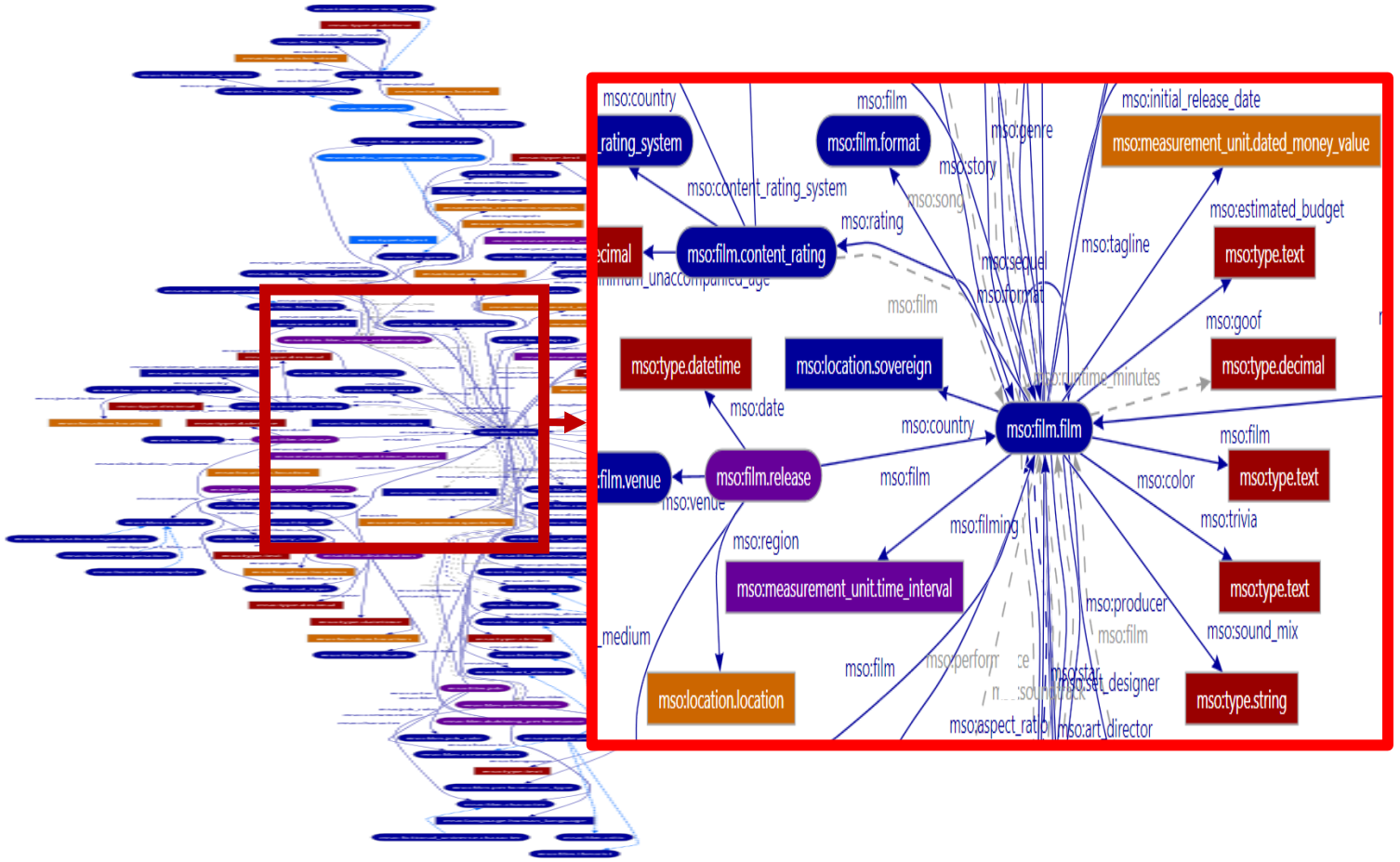
Types

award.nominee, award.winner, film.actor, film.director, film.producer, film.story_contributor ...

General Design Principles

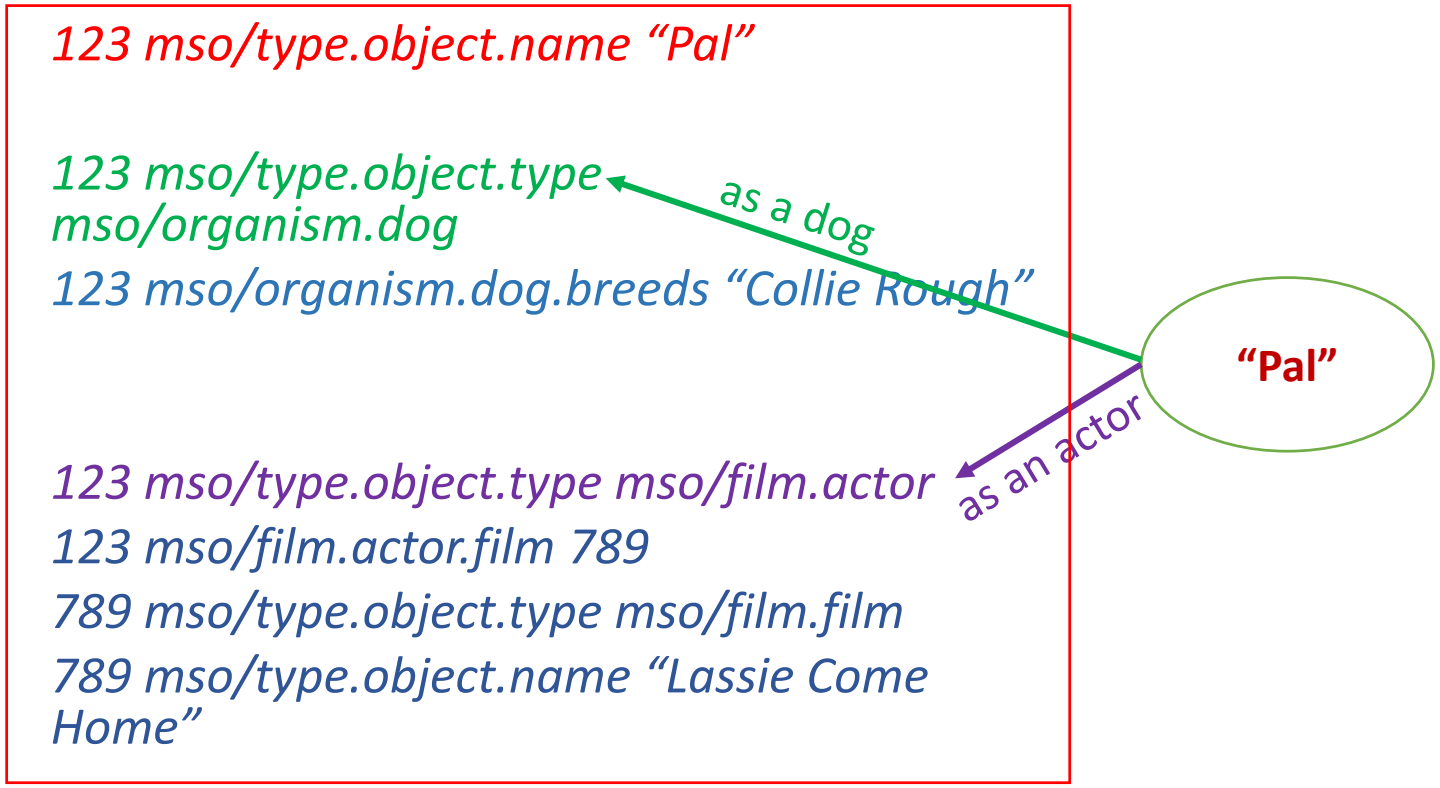
Challenges of serving knowledge graphs

- Data size
  - in the scale of terabytes
- Complex data schema
  - Rich relations

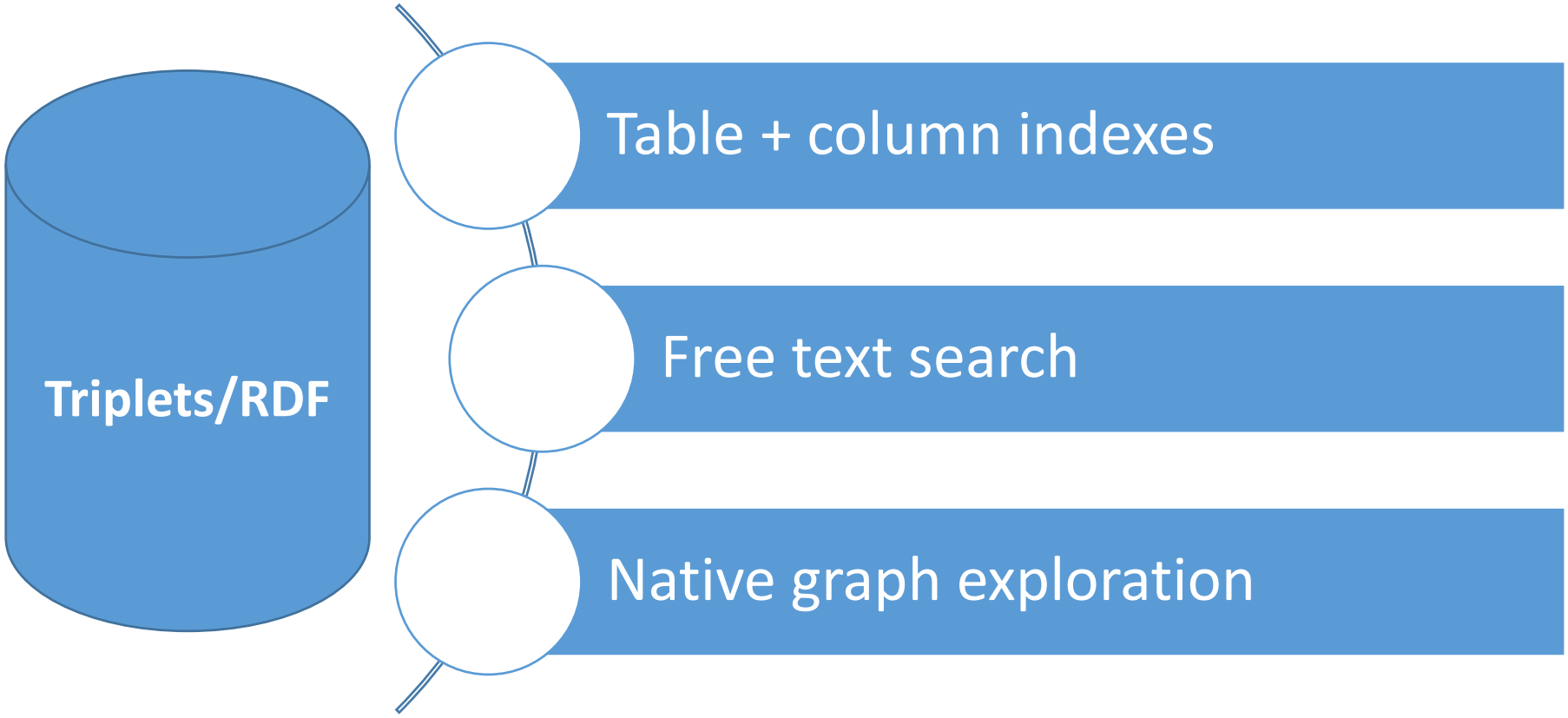


Challenges of serving knowledge graphs

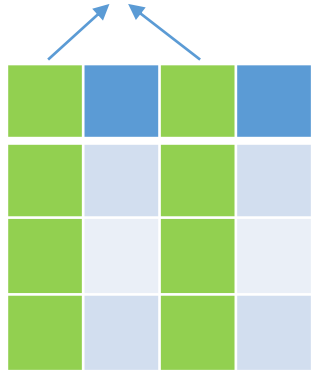
- Data size
  - In the scale of terabytes
- Complex data schema
  - Rich relations
  - Multi-typed entities

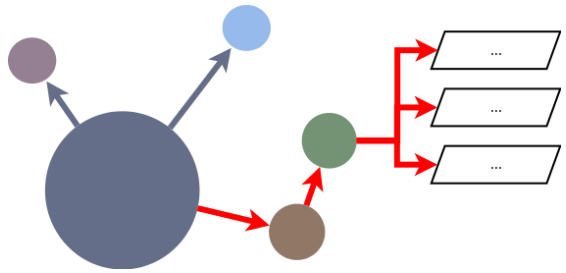


How to serve knowledge?



Column Index

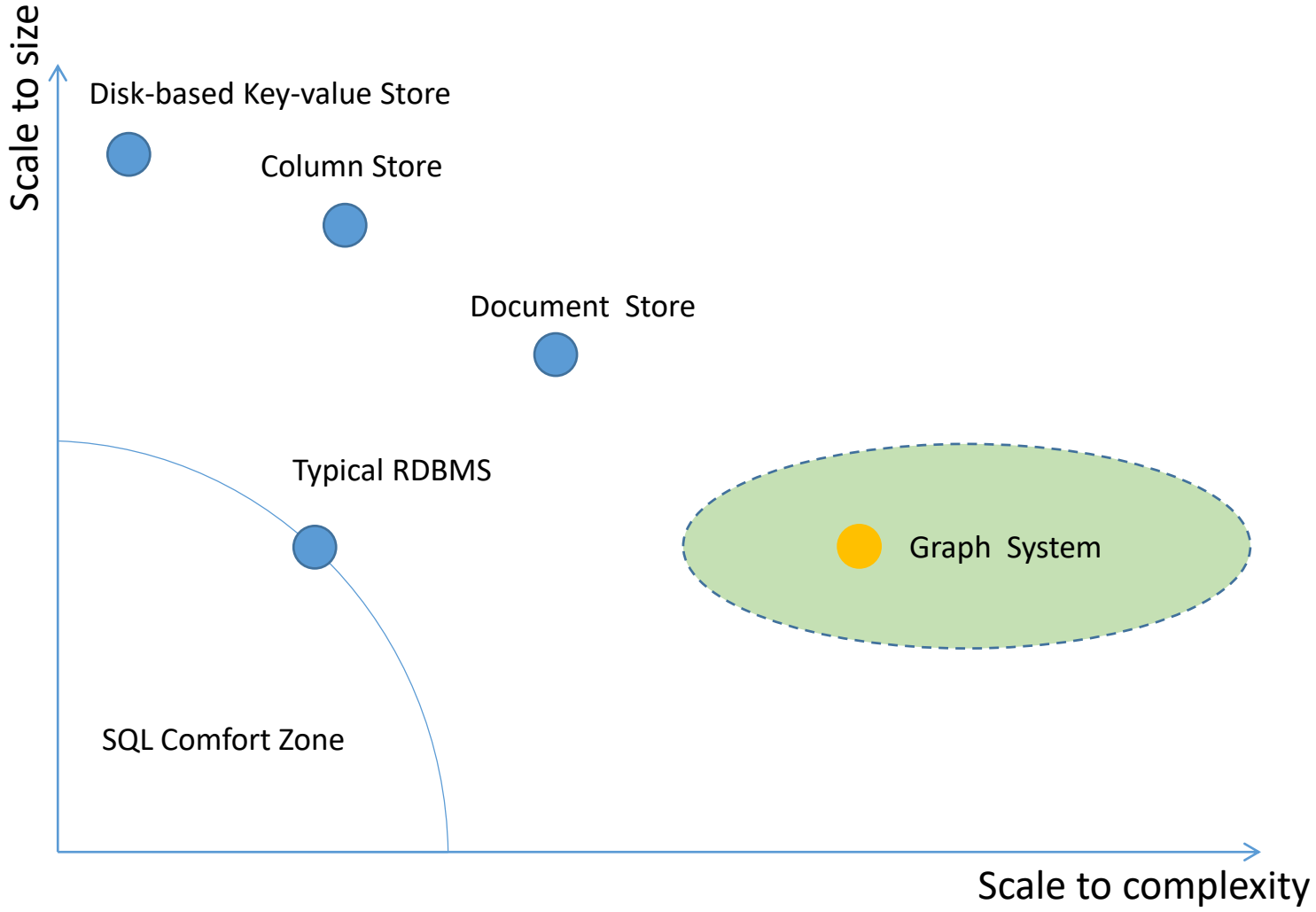




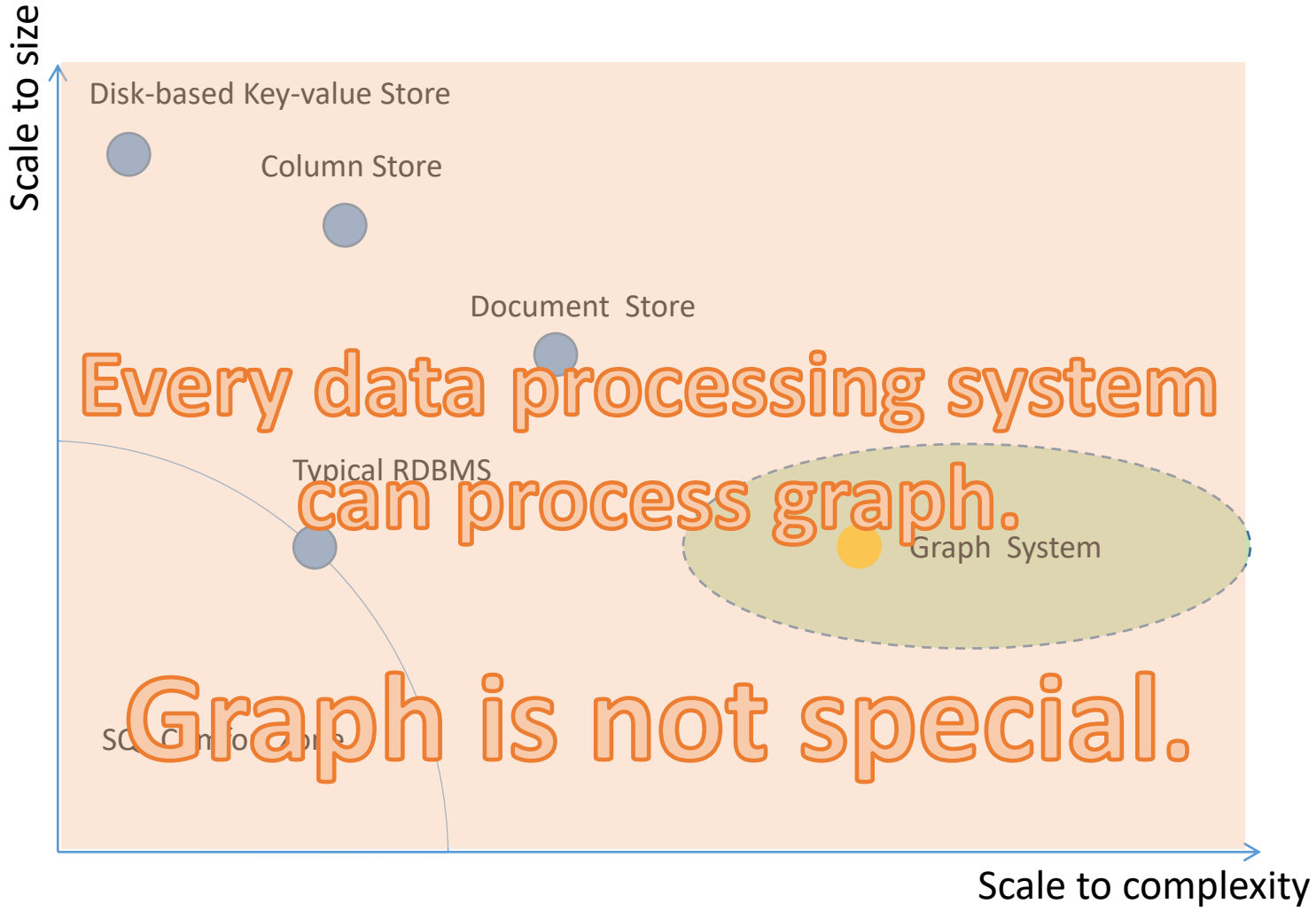
The needs ultimately determine the design

The first important rule: there is no one-size-fits-all system!

First rule: no one-size-fits-all system

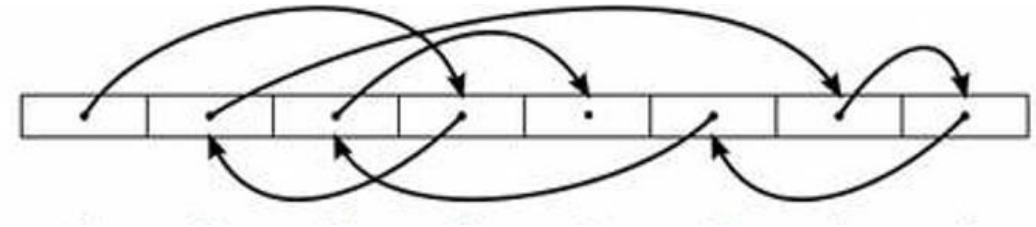


First rule: no one-size-fits-all system



Characteristics of parallel graph processing

- Random access (Poor Locality)
  - For a node, its adjacent nodes cannot be accessed without "jumping" no matter how you represent a graph
  - Not cache-friendly, data reuse is hard



- It is hard to partition data
  - In this sense, graph is "special".

- Data driven
  - the structure of computations is not known a priori

- High data access to computation ratio

Reference: Challenges in parallel graph processing

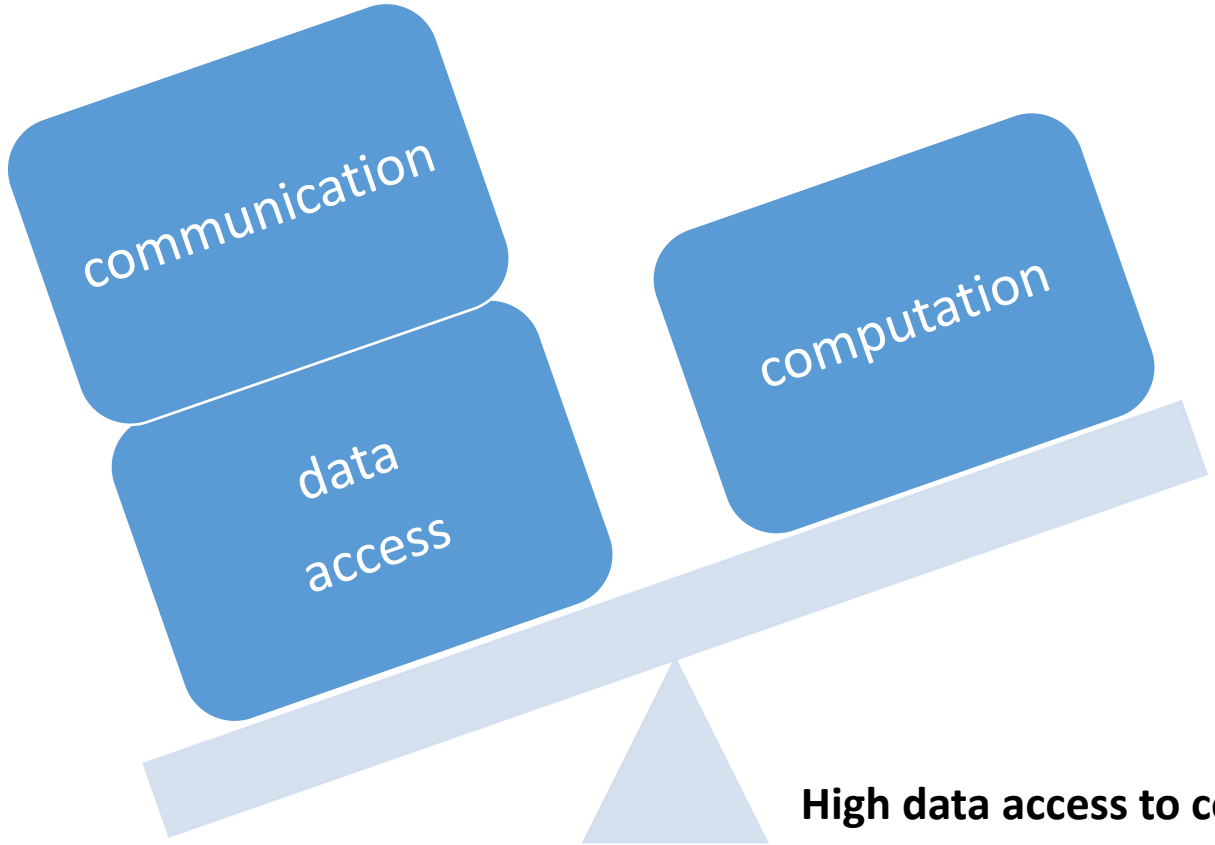
Design choices

- **First important rule: there is no one-size-fits-all system**
- Does this system support online queries, offline analytics, or both?
- Is the system optimized for response time, throughput, or both?
- Does the system scale, “out” or “up”?
- Does the system need transaction support?

Online queries vs. offline analytics

- Online query processing is usually optimized for response time
- Offline analytics is usually optimized for throughput
- Compared to offline analytics, it is harder to optimize online queries
  - Online queries are sensitive to latency
  - It is difficult to predict the data access patterns of a graph query

Query response time:
data access + communication + computation



High data access to computation ratio

System design choice

- Main storage (storage backend)
- Index
- Communication paradigm: two-sided vs. one-sided
- Scale out or scale up
- ACID Transactions or not

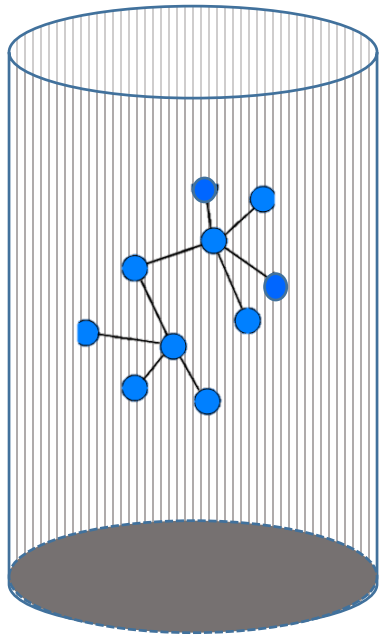
System design choice

- **Main storage (storage backend)**
- Index
- Communication paradigm: two-sided vs. one-sided
- Scale out or scale up
- ACID Transactions or not

Graph may be in the jail of storage

- Many existing data management systems can be used to process graphs

- Many existing systems are mature, but not for graphs
  - RDBMS, MapReduce
  - The commonest graph operation “traversal” incurs excessive amount of joins



Graph in the Jail of the storage

Traverse graph using joins in RDBMS

ID	name	….
1	N1	…
2	N2	…
3	N3	…
4	N4	…
5	N5	…
6	N6	…
…	…	…

Node Table: N

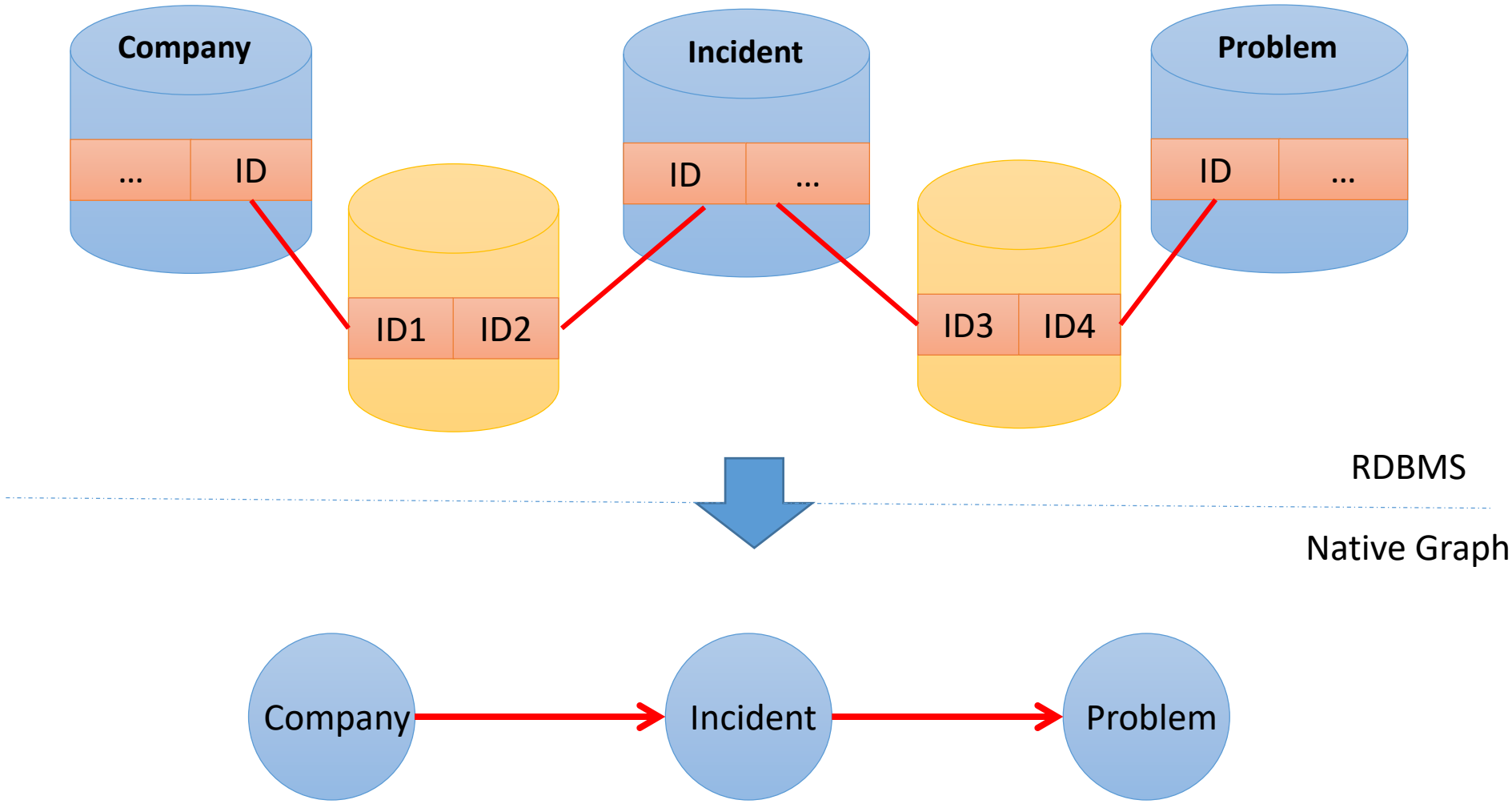
src	dst
1	3
2	4
2	1
4	3
1	5
1	6
…	…

Edge Table: E

Get neighbors of N1

```
SELECT*
FROM N
LEFT JOIN  E ON N.ID = E.dst
WHERE E.src = 1;
```

Multi-way Join vs. graph traversal



System design choice

- Main storage (storage backend)
- **Index**
- Communication paradigm: two-sided vs. one-sided
- Scale out or scale up
- ACID Transactions or not

Index

It is costly to index graph structures, use it wisely.

Query Index Examples

Algorithms	Index Size	Index Time	Update Cost
Ullmann [Ullmann76], VF2 [CordellaFSV04]	-	-	-
RDF-3X [NeumannW10]	$O(m)$	$O(m)$	$O(d)$
BitMat [AtreCZH10]	$O(m)$	$O(m)$	$O(m)$
Subdue [HolderCD94]	-	Exponential	$O(m)$
SpiderMine [ZhuQLYHY11]	-	Exponential	$O(m)$
R-Join [ChengYDYW08]	$O(nm^{1/2})$	$O(n^4)$	$O(n)$
Distance-Join [ZouCO09]	$O(nm^{1/2})$	$O(n^4)$	$O(n)$
GraphQL [HeS08]	$O(m + nd^r)$	$O(m + nd^r)$	$O(d^r)$
Zhao [ZhaoH10]	$O(nd^r)$	$O(nd^r)$	$O(d^L)$
GADDI [ZhangLY09]	$O(nd^L)$	$O(nd^L)$	$O(d^L)$

Index-based subgraph matching [Sun VLDB 2012]

Query Index Examples

Algorithms	Index Size for Facebook	Index Time for Facebook	Query Time on Facebook (s)
Ullmann [Ullmann76], VF2 [CordellaFSV04]	-	-	>1000
RDF-3X [NeumannW10]	1T	>20 days	>48
BitMat [AtreCZH10]	2.4T	>20 days	>269
Subdue [HolderCD94]	-	> 67 years	-
SpiderMine [ZhuQLYHY11]	-	> 3 years	-
R-Join [ChengYDYW08]	>175T	$> 10^{15}$ years	>200
Distance-Join [ZouCO09]	>175T	$> 10^{15}$ years	>4000
GraphQL [HeS08]	>13T(r=2)	> 600 years	>2000
Zhao [ZhaoH10]	>12T(r=2)	> 600 years	>600
GADDI [ZhangLY09]	$> 2 \times 10^5$T (L=4)	$> 4 \times 10^5$ years	>400

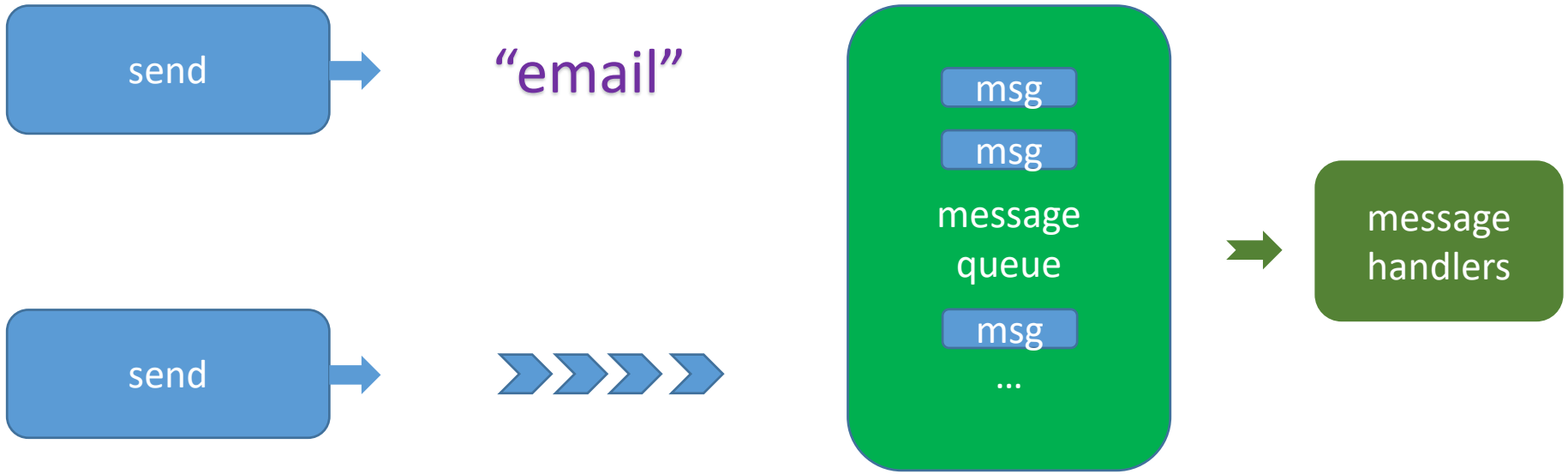
Index-based subgraph matching [Sun VLDB 2012]

System design choice

- Main storage (storage backend)
- Index
- **Communication paradigm: two-sided vs. one-sided**
- Scale out or scale up
- ACID Transactions or not

Two-sided communication

One-sided communication



System design choice

- Main storage (storage backend)
- Index
- Communication paradigm: two-sided vs. one-sided
- **Scale out or scale up**
- ACID Transactions or not

Design choice: scale-up vs. scale-out

- Supercomputer model
  - Programming model simple and efficient
    - shared memory address space
  - Expensive
  - Hardware is your ultimate limit

- Distributed cluster model
  - Programming model is complex
  - Relatively cheaper and can make use of commodity pc
  - Flexible to meet various needs

Scale “OUT”, not “UP”

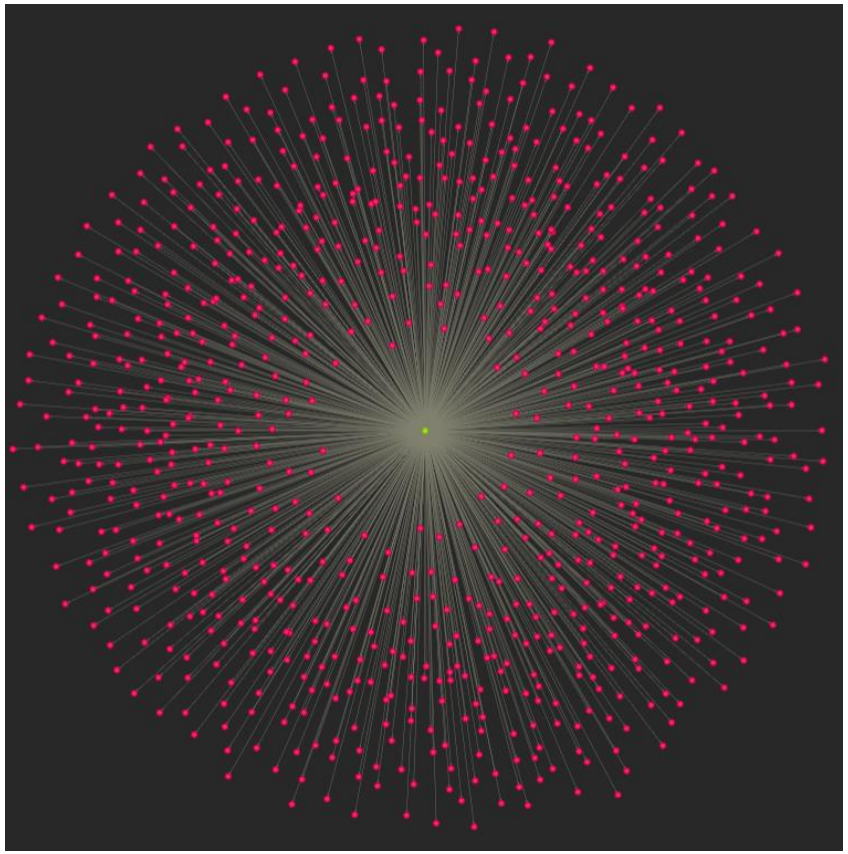
System design choice

- Main storage (storage backend)
- Index
- Communication paradigm: two-sided vs. one-sided
- Scale out or scale up
- **ACID Transactions or not**

Think twice before diving into transactions

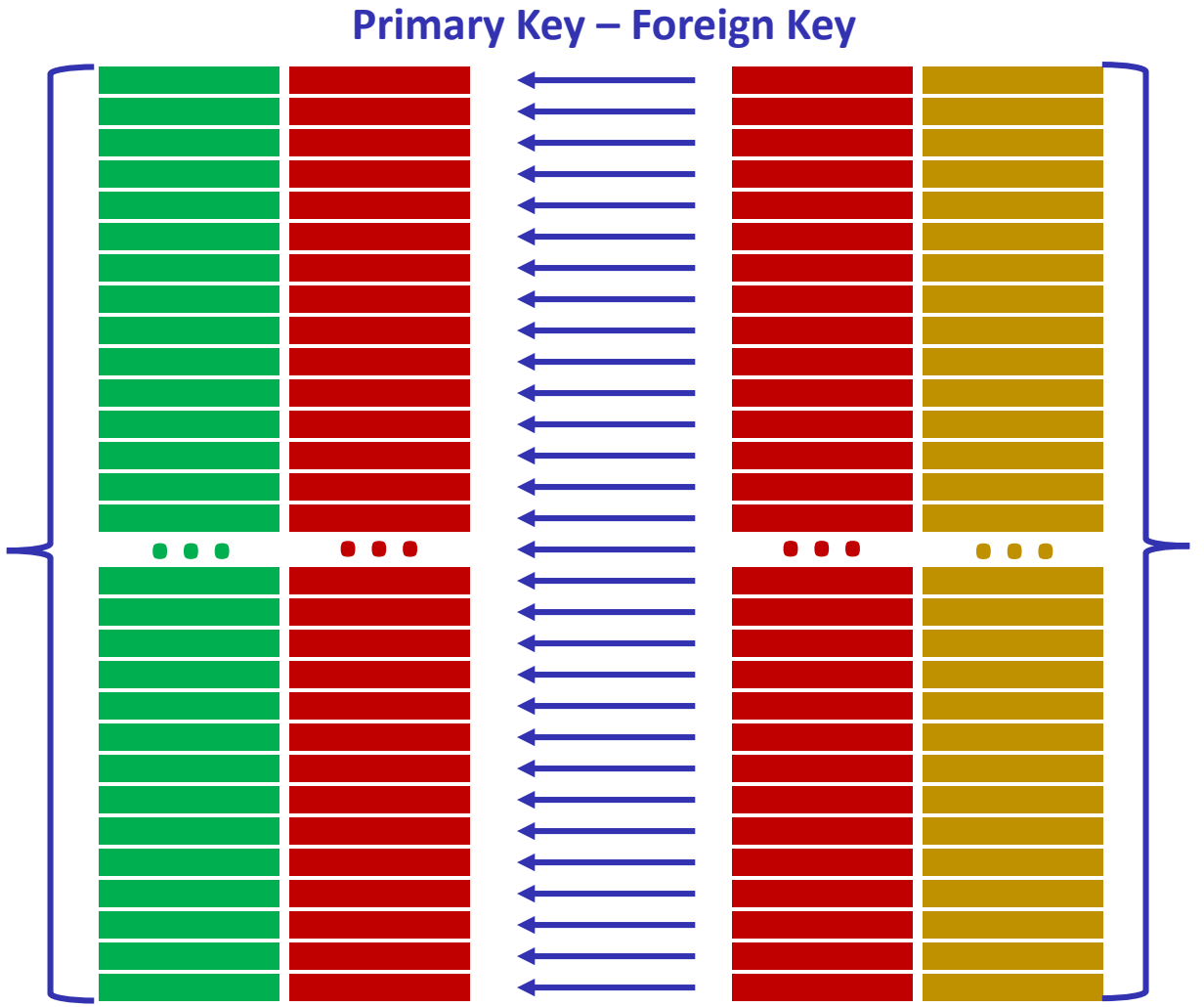
- Pros
  - Strong data consistency guarantee
- Cons
  - The hell of referential integrity
  - The disaster of cascading rollback
  - Multi-round network communications per commit for distributed transactions

The hell of referential integrity

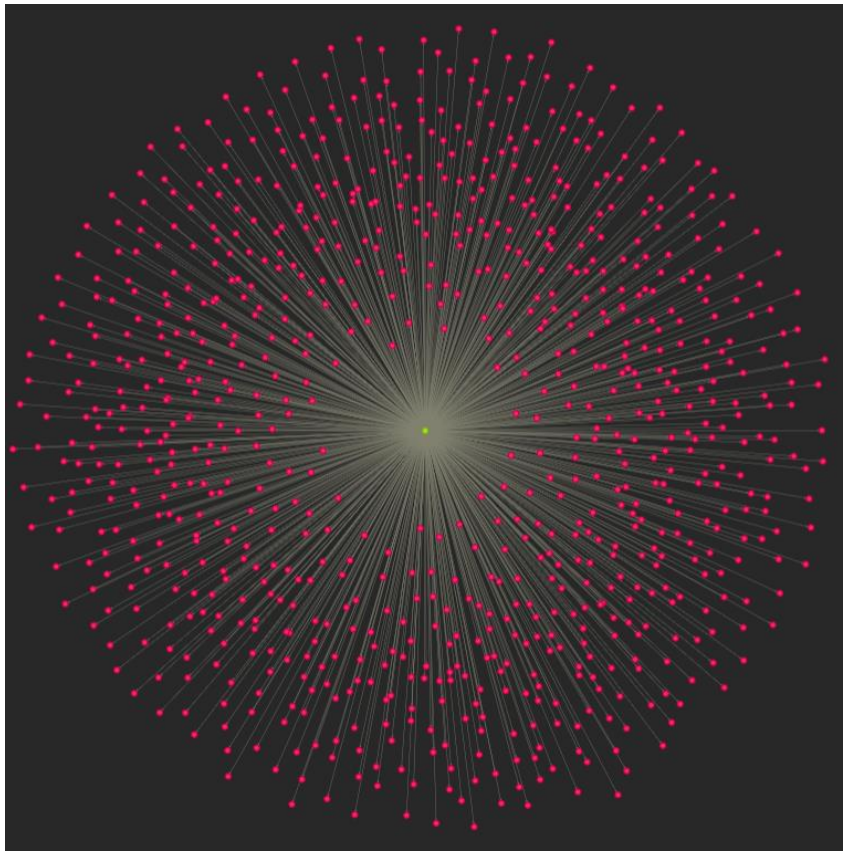


Lady Gaga in Freebase

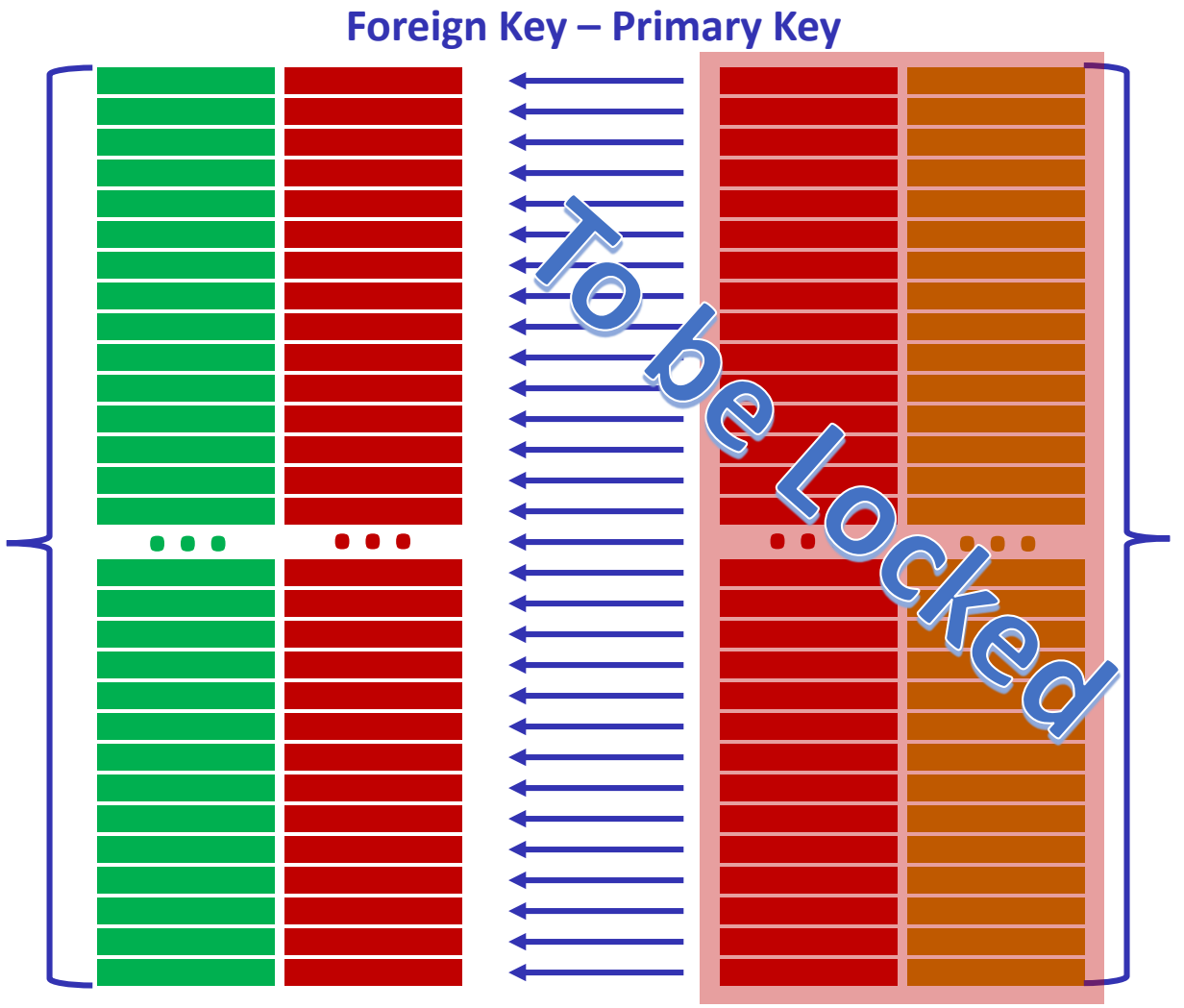
Lady Gaga



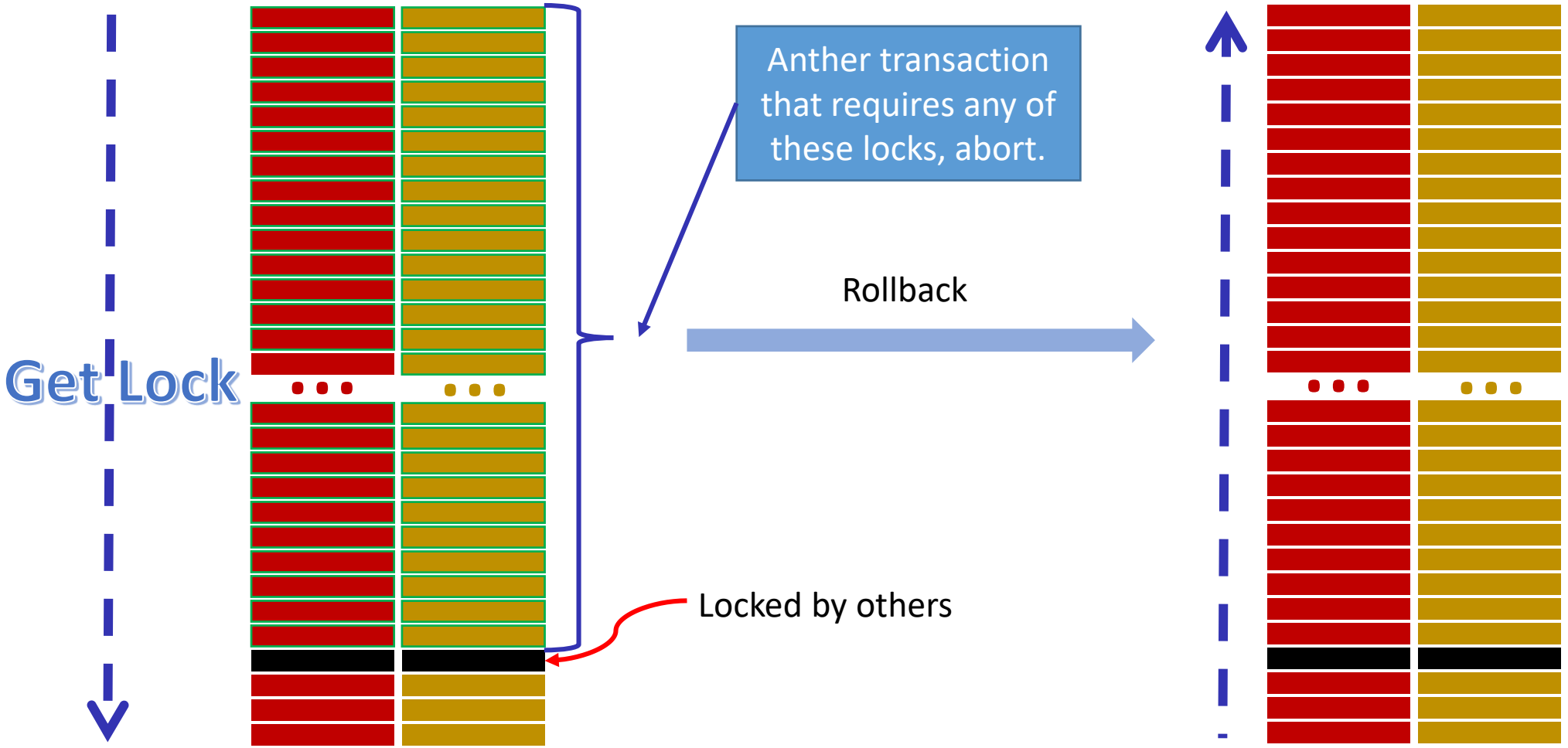
The hell of referential integrity



Lady Gaga



The disaster of cascading rollback



Real-time Query Processing

Query processing

- Where latencies come from and asynchronous fan-out search

- Index-free query processing

Query processing

- Where latencies come from and asynchronous fan-out search

- Index-free query processing

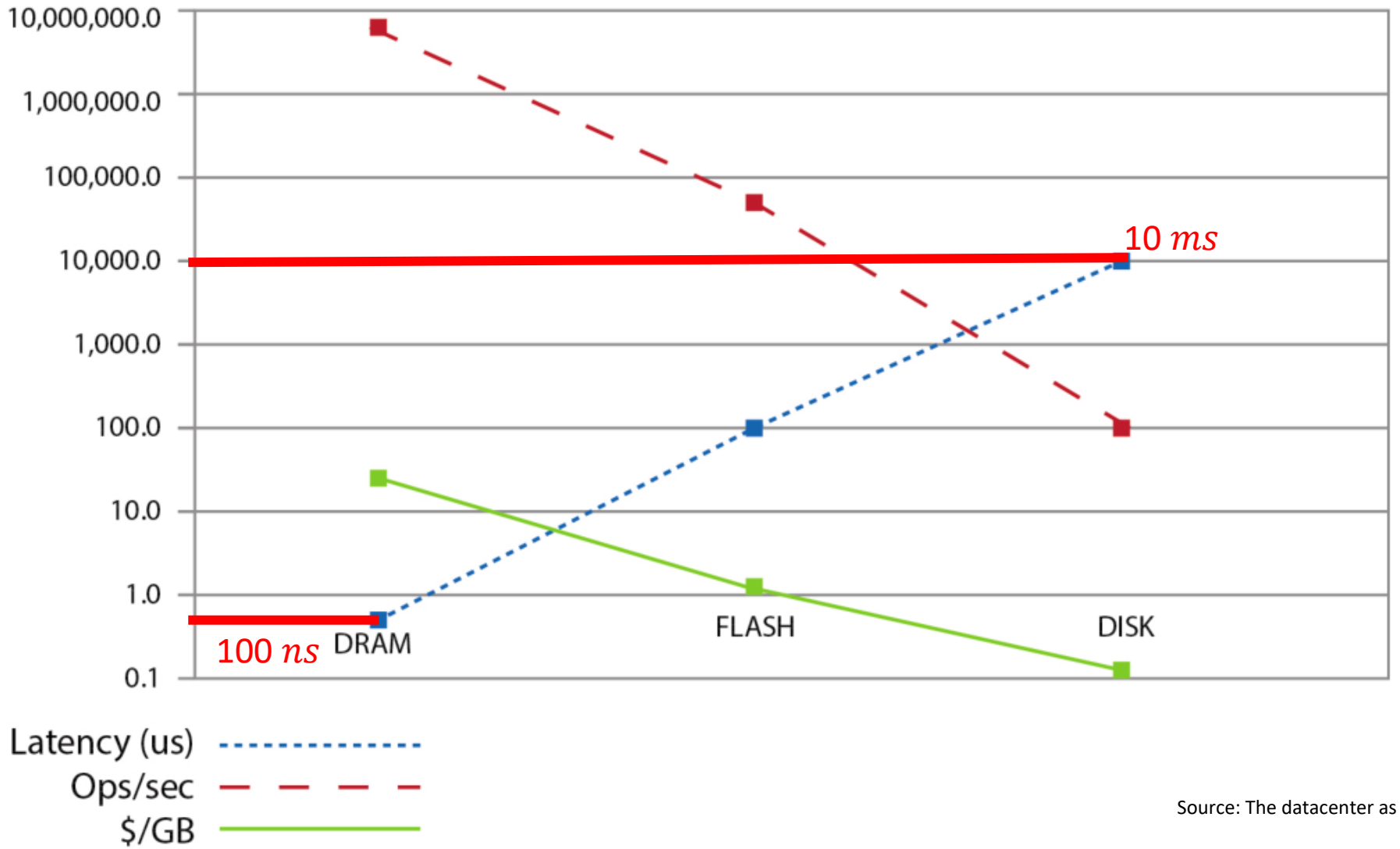
People search challenge in Facebook graph

- Among adult Facebook users, the average number of friends is 338.

338
+338 x 338
+338 x 338 x 338
=38,729,054

Can we search a person in one's 3-hop neighborhood within 500 ms?

Latency, Bandwidth, and Capacity



Source: The datacenter as a computer (book)

Disk-based approach

$$338 + 338 \times 338 + 338 \times 338 \times 338 = 38{,}729{,}054$$

$\longrightarrow$ $387{,}290{,}540$ ms $= 4.5$ days

each disk seek + read: > 10 ms

RAM-based approach

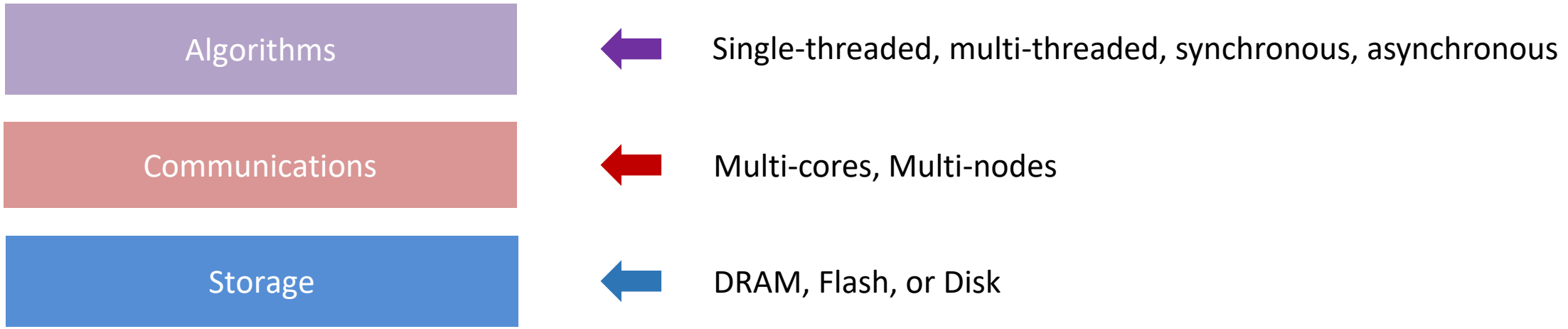
- DRAM latency: 100 ns

10 million reads/writes per second

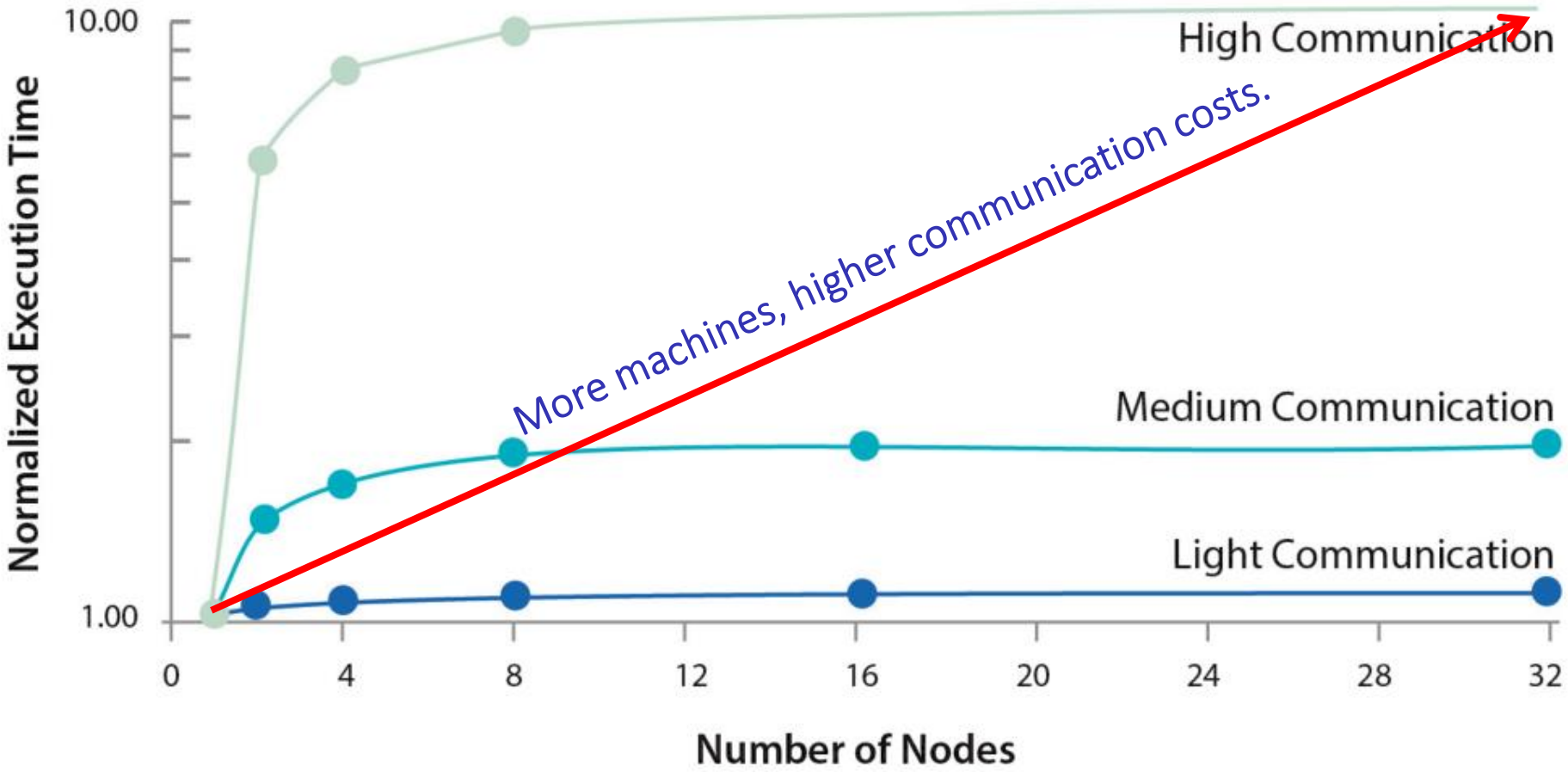
1 million node-level read/write per second

38,729,054 nodes to access, it takes at least 38 seconds.

Where do latencies come from?



Move computation, instead of data!



Source: The datacenter as a computer (book)

If you care about latency, do not use the shared-memory model in a distributed setting.

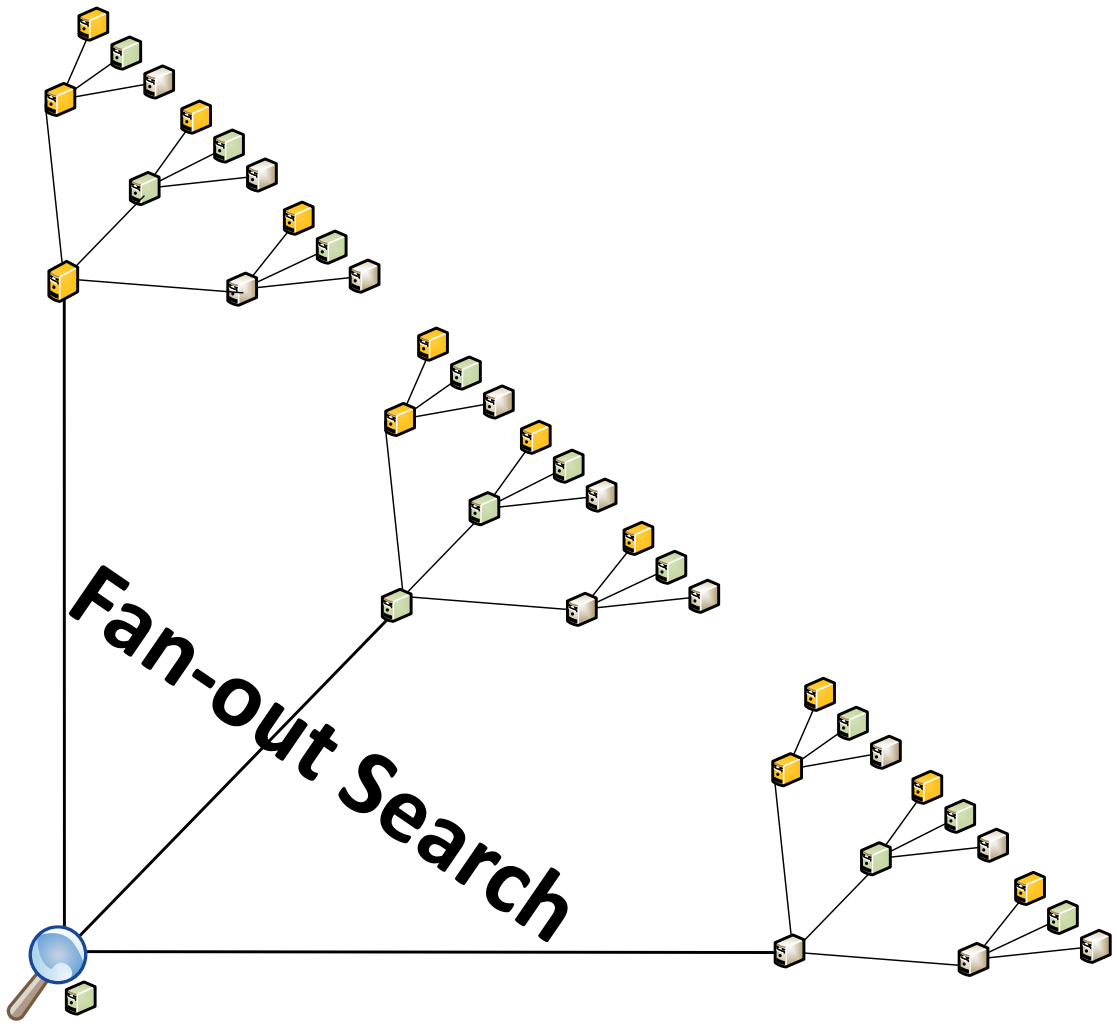
Lessons learned so far (how to reduce latencies)

- RAM (Hardware sometimes does matter a lot)
  - The stupid buy faster computers, smart ones write better programs?
- Avoid moving data
- ……

Lessons learned so far (how to reduce latencies)

- RAM (Hardware sometimes does matter a lot)
  - The stupid buy faster computers, smart ones write better programs?
- Avoid moving data
- Avoid unnecessary synchronizations

Asynchronous fan-out search



Hop	Msg #	Node # per machine
1	n	$\frac{d}{n}$
2	n^2	$\frac{d^2}{n}$
3	n^3	$\frac{d^3}{n}$

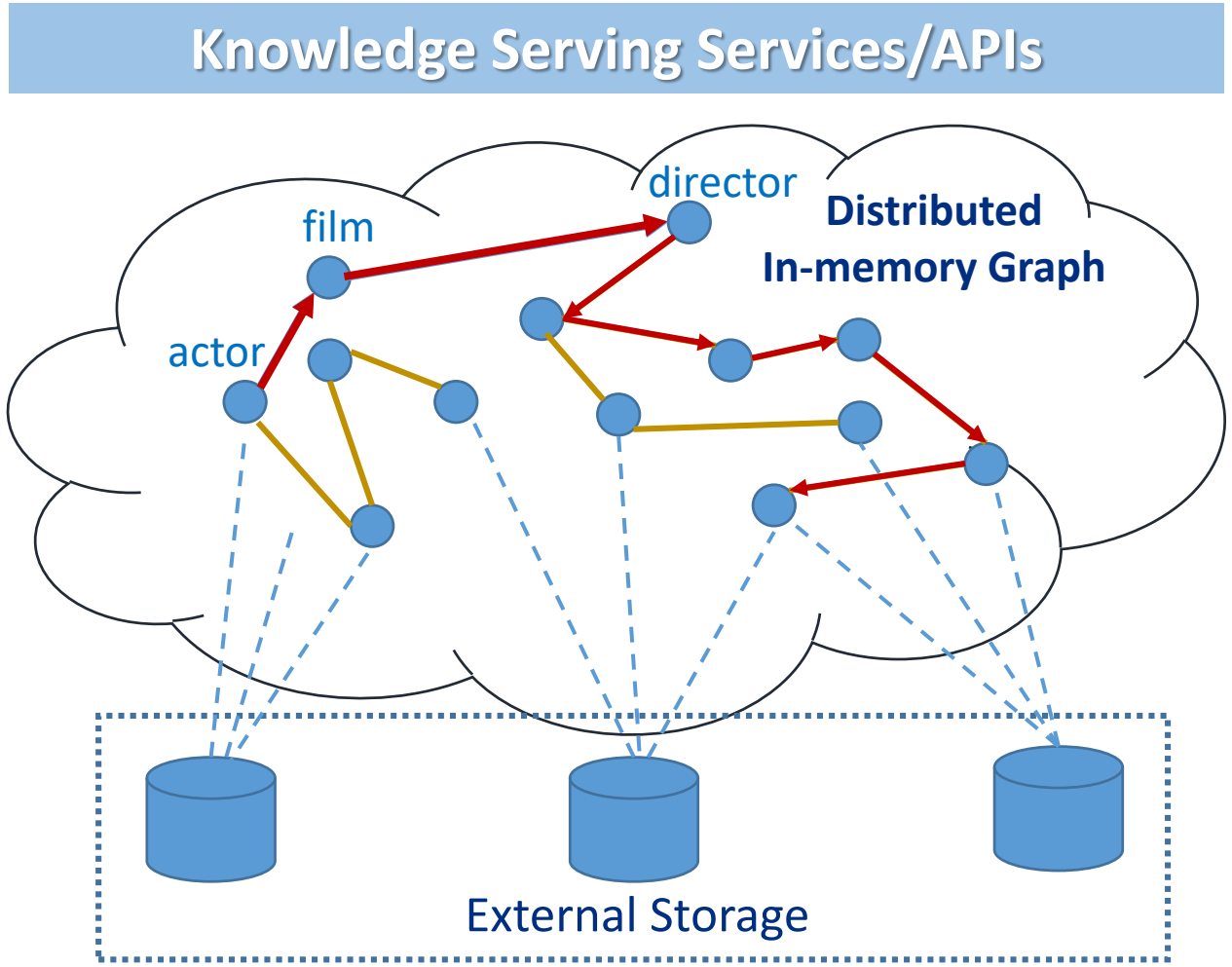
n is the server count
d is the average degree

Online query processing

- Where latencies come from and fan-out search

- Index-free query processing

Query KG via Graph Exploration



Online query example: subgraph matching

Procedure:

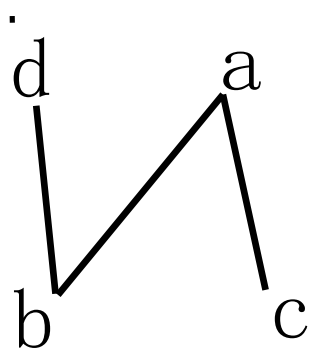
1. Break a graph into basic units (edges, paths, frequent subgraphs, …)
2. Build index for every possible basic unit
3. Decompose a query into multiple basic unit queries, and join the results

Case study: distributed subgraph matching

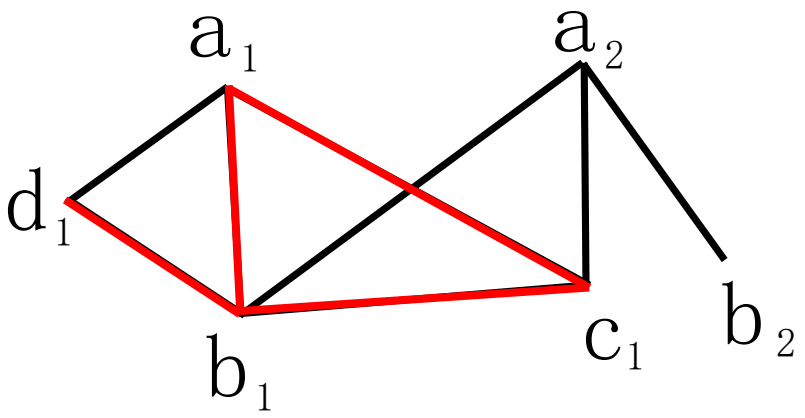
Procedure:

1. Break a query into basic units
2. **Match the basic units in parallel on the fly**
3. Join the results

Subgraph matching

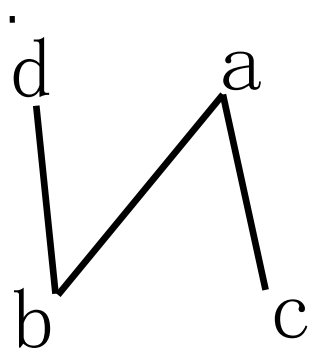


G_1

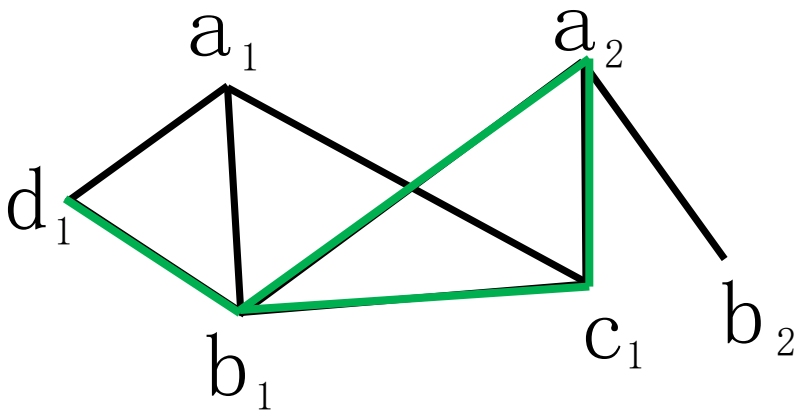


G_2

Subgraph matching

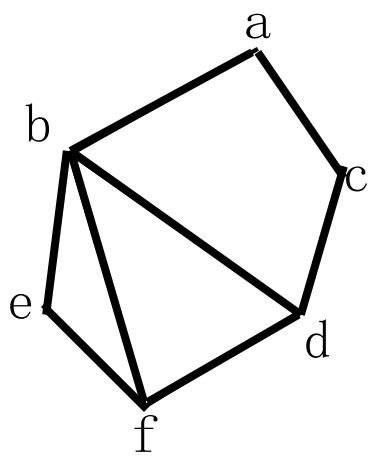


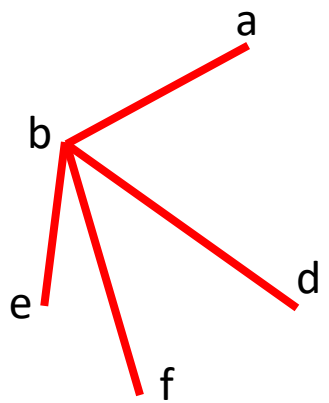
G_1

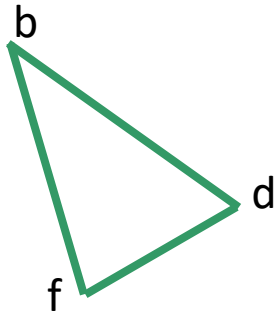


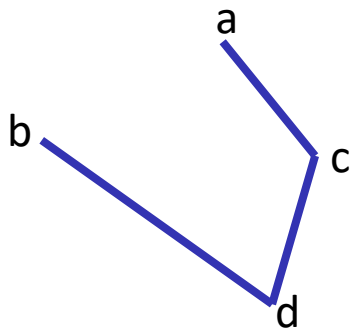
G_2

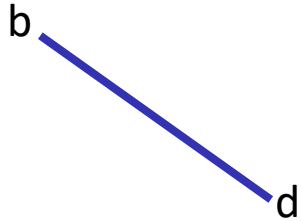
Basic unit for distributed subgraph matching





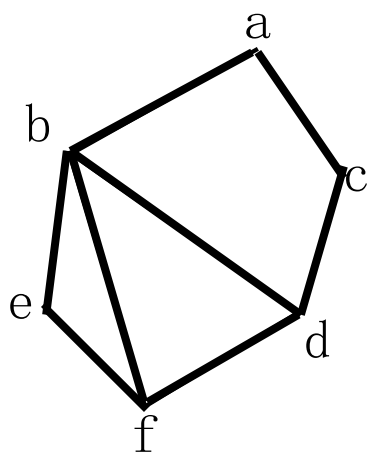


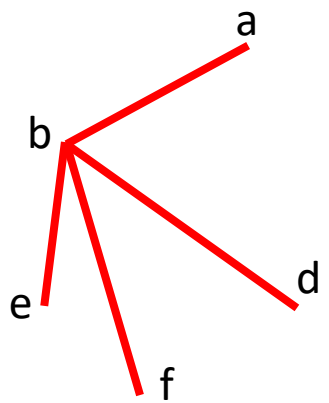


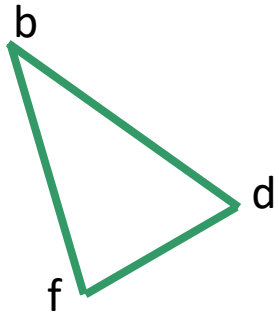


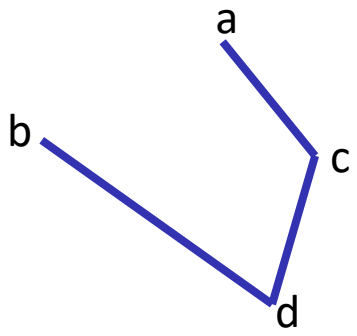
As a basic unit, which one is the best?

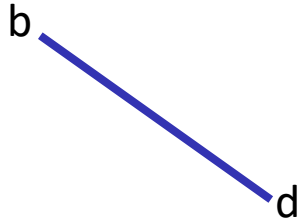
Basic unit for distributed subgraph matching





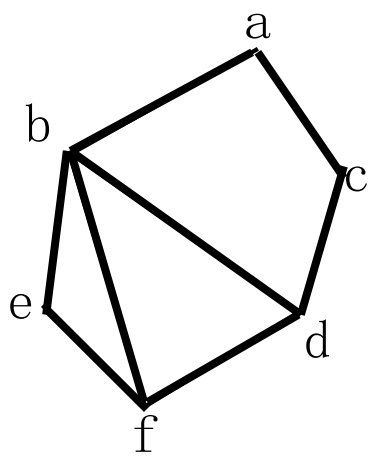


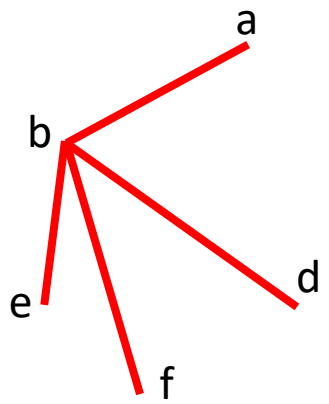




As a basic unit, which one is the best?

Basic unit for distributed subgraph matching



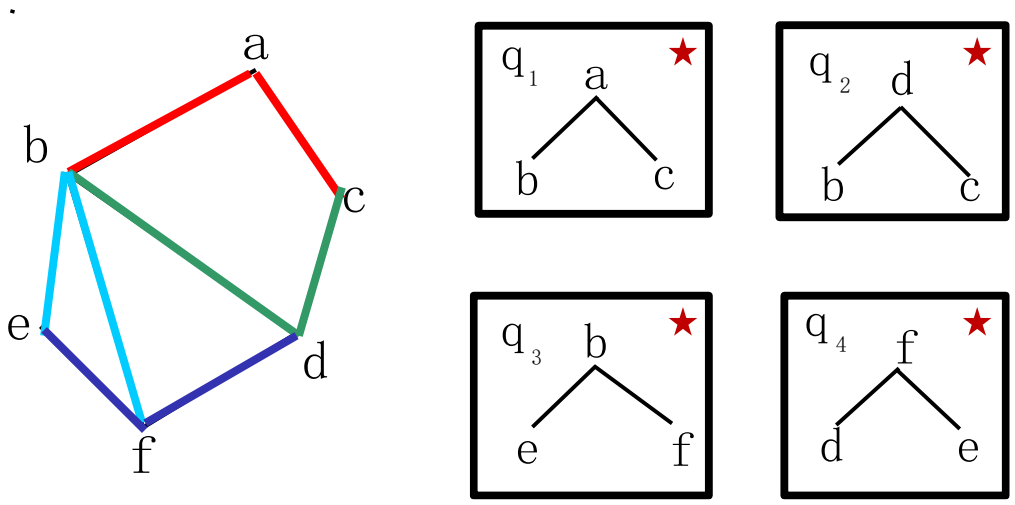


Twig

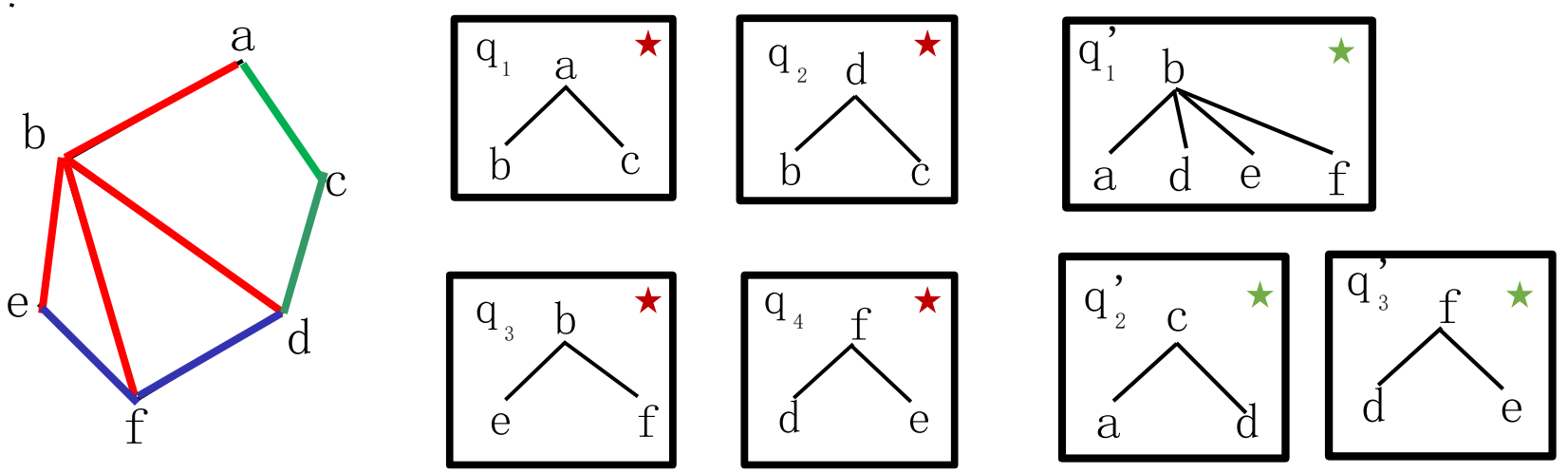
- Easy to decompose
- Height is always one
  - It at most needs to cross the network once

As a basic unit, which one is the best?

Query decomposition



Query decomposition



Query optimization problems

- How to choose a good query decomposition
- How to choose a good execution order
- How to choose a good join order

Representative Graph Systems

Existing systems

- Mature data processing systems
  - RDBMS
  - Map Reduce Systems
- Systems specialized for certain graph operations
  - PageRank, FlockDB
- General-purpose graph processing systems
  - Neo4j, Trinity, Horton, HyperGraphDB, TinkerGraph, InfiniteGraph, Cayley, Titan, PEGASUS, Pregel, Giraph, GraphLab, GraphChi, GraphX …

Representative graph processing systems

	Property graphs	Online query	Data sharding	In-memory storage	Atomicity & Transaction
Neo4j	Yes	Yes	No	No	Yes
Trinity	Yes	Yes	Yes	Yes	Atomicity
Horton	Yes	Yes	Yes	Yes	No
HyperGraphDB	No	Yes	No	No	Yes
FlockDB	No	Yes	Yes	No	Yes
TinkerGraph	Yes	Yes	No	Yes	No
InfiniteGraph	Yes	Yes	Yes	No	Yes
Cayley	Yes	Yes	SB	SB	Yes
Titan	Yes	Yes	SB	SB	Yes
MapReduce	No	No	Yes	No	No
PEGASUS	No	No	Yes	No	No
Pregel	No	No	Yes	No	No
Giraph	No	No	Yes	No	No
GraphLab	No	No	Yes	No	No
GraphChi	No	No	No	No	No
GraphX	No	No	Yes	No	No

Demo

References

- Andrew Lumsdaine, Douglas Gregor, Bruce Hendrickson, Jonathan Berry. Challenges in parallel graph processing. Parallel Processing Letters 2007.
- Bin Shao, Haixun Wang, Yatao Li. Trinity: a distributed graph engine on a memory cloud. SIGMOD 2013.
- Luiz André Barroso, Urs Hoelzle. The datacenter as a computer: An introduction to the design of warehouse-scale machines. 2009.
- Grzegorz Malewicz, Matthew H. Austern, Aart J. C. Bik, James C. Dehnert, Ilan Horn, Naty Leiser, Grzegorz Czajkowski. Pregel: a system for large-scale graph processing. SIGMOD 2010.
- Aapo Kyrola, Guy Blelloch, Carlos Guestrin. GraphChi: Large-scale graph computation on just a PC. OSDI 2012.
- Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, and Jianzhong Li. Efficient subgraph matching on billion node graphs. PVLDB 2012.
- U Kang, Charalampos E. Tsourakakis, Christos Faloutsos. PEGASUS: A peta-scale graph mining system implementation and observations. ICDM 2009.

Thanks!

https://www.graphengine.io/
https://www.binshao.info/