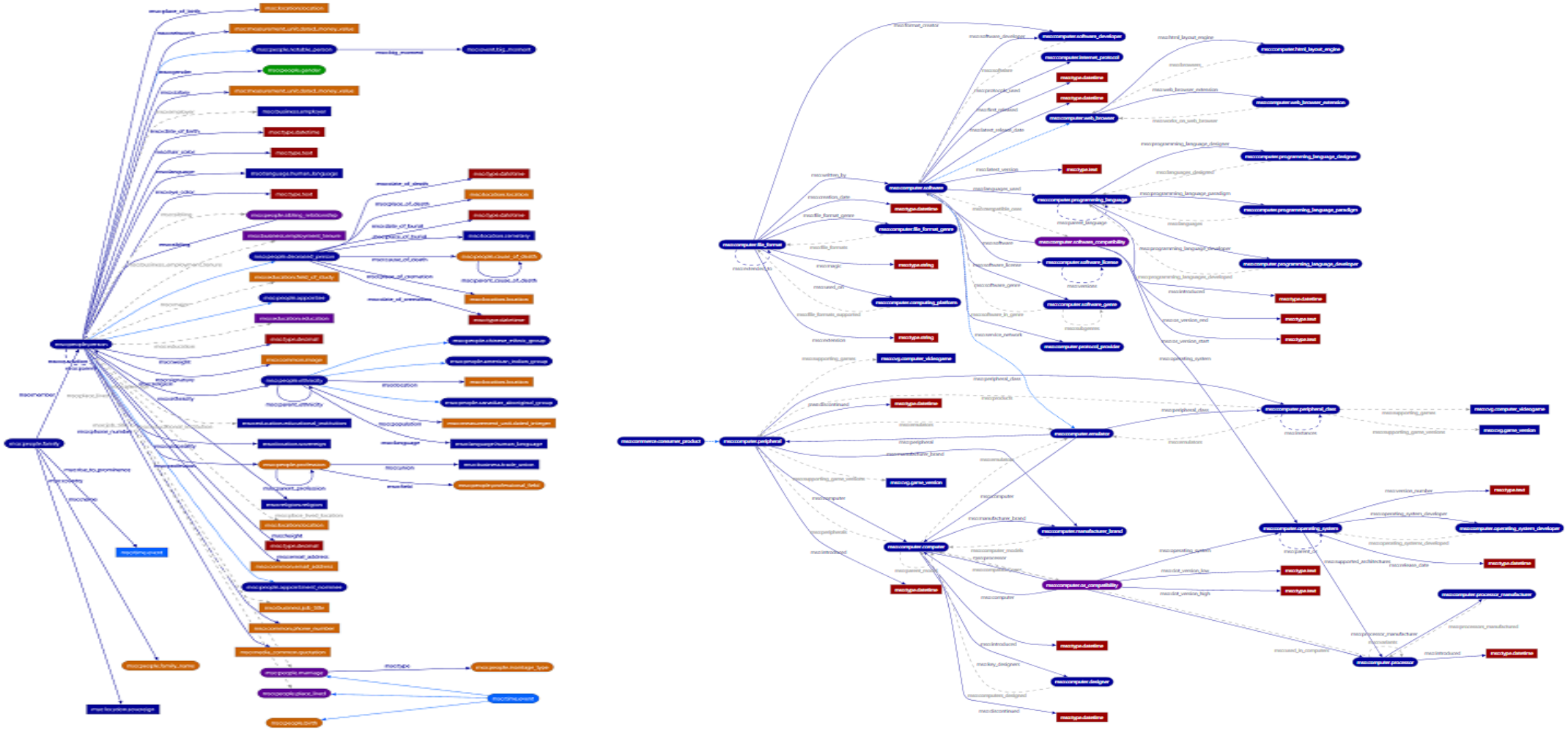
# Real-time Knowledge Graph Serving

Bin Shao

Microsoft Research Asia (Beijing, China)

# MKG: billions of entities with rich relationships



A small portion of the schema graph

# How to serve knowledge?

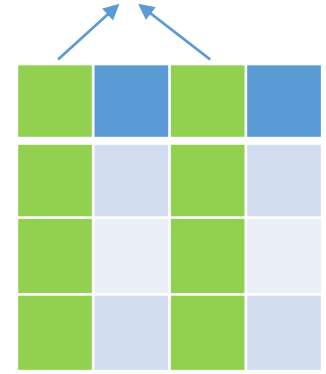
**Knowledge data, e.g. Triplets/RDF**

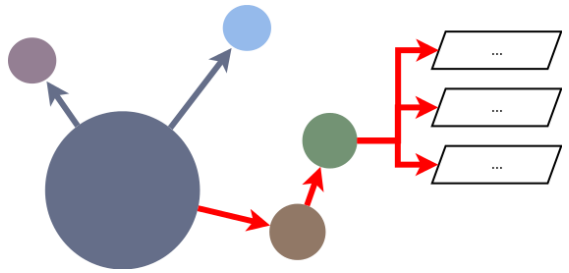
Table + column indexes

Free text search

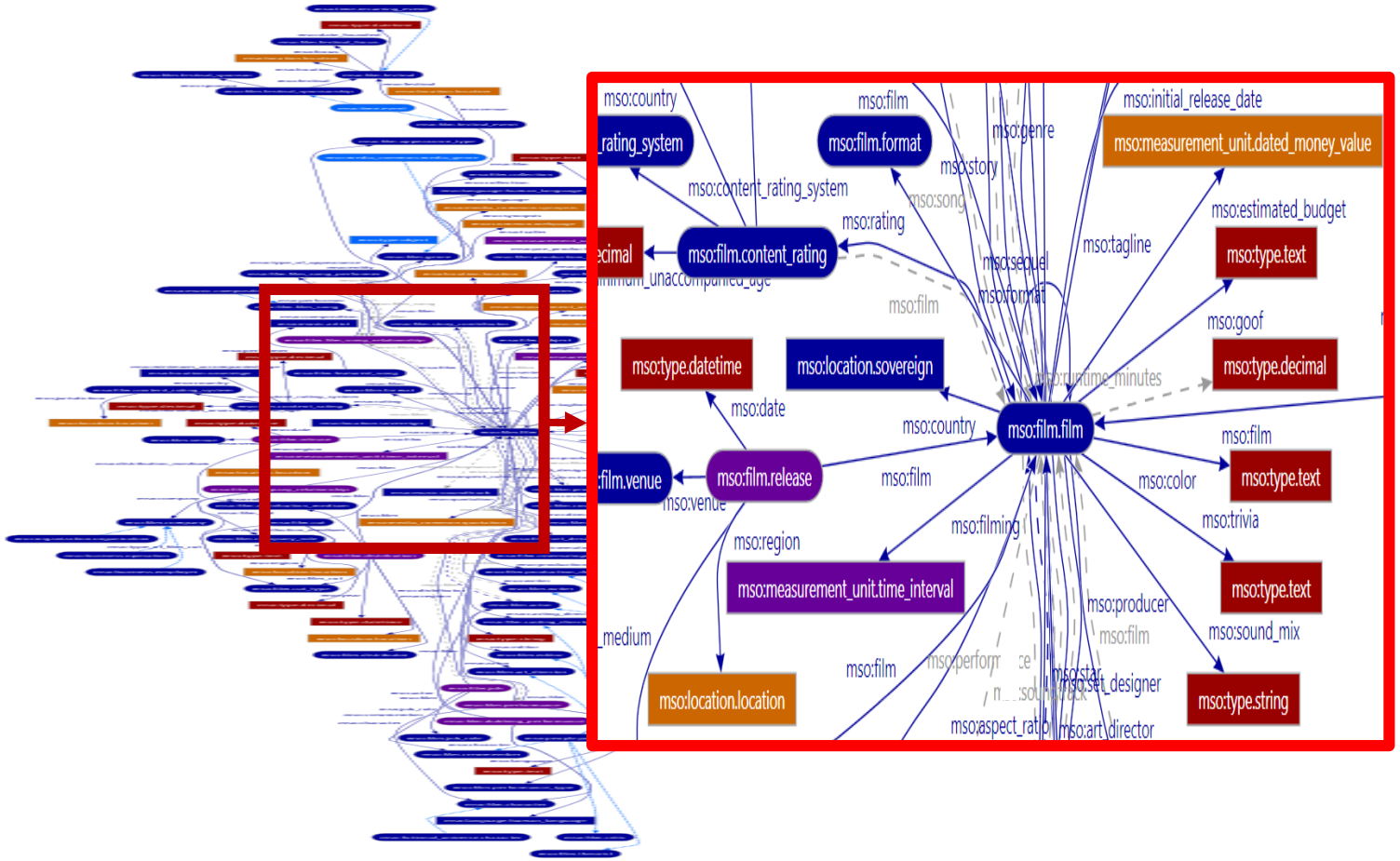
Native graph exploration

Column Index



# Complexity of the knowledge graph

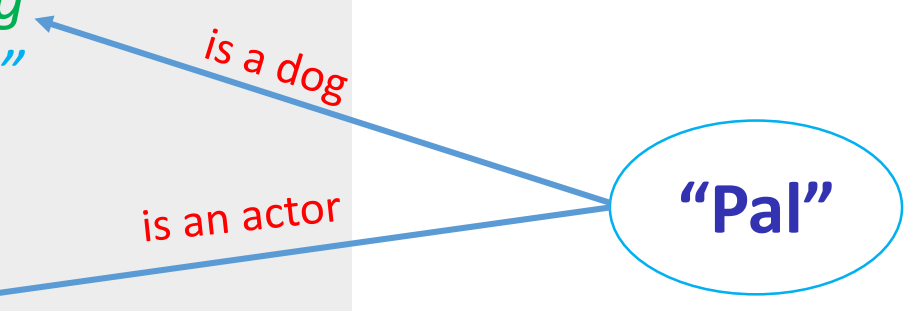


# Multi-typed entities

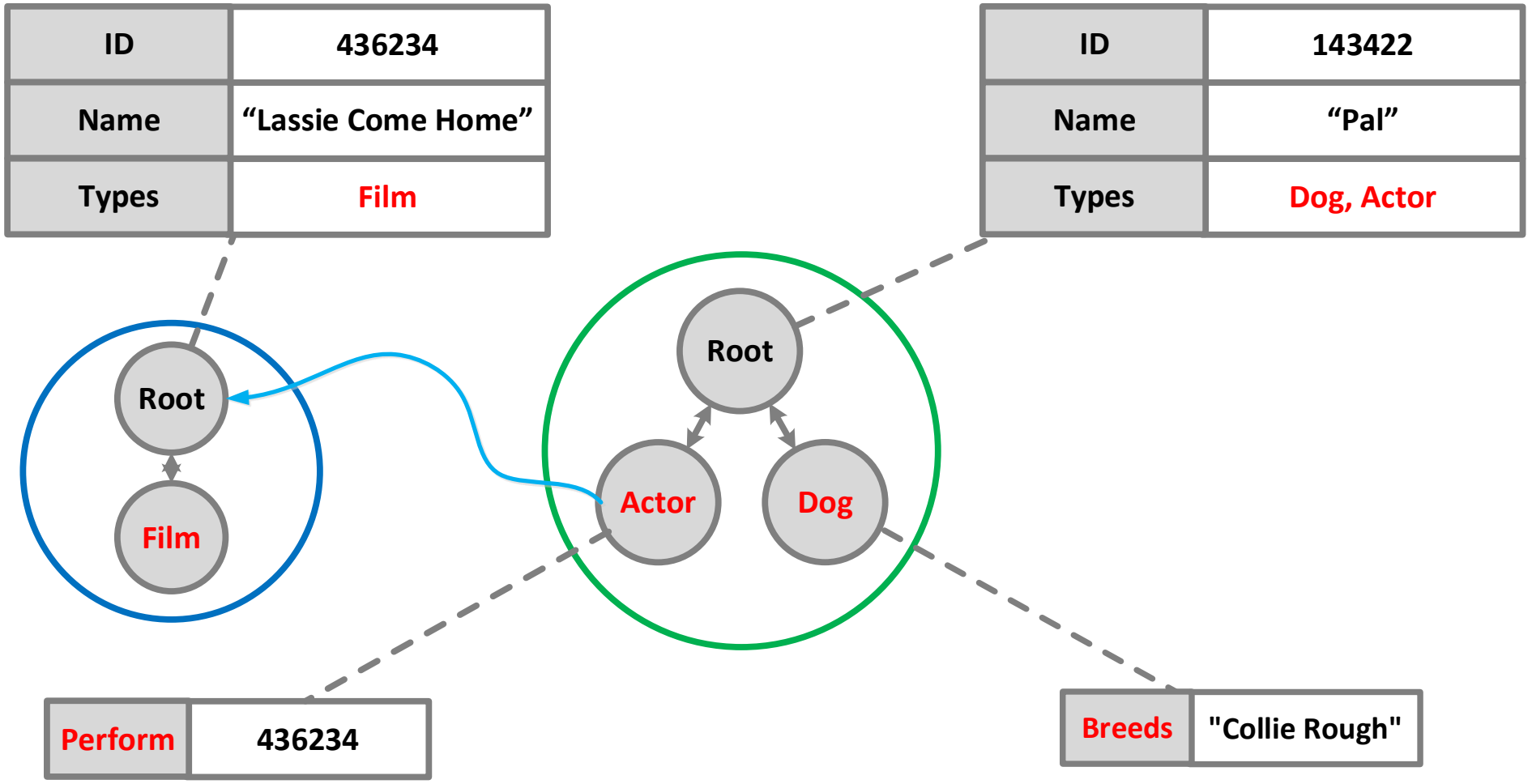
*123 mso/type.object.name “Pal”*

*123 mso/type.object.type  mso/organism.dog*
*123 mso/organism.dog.breeds “Collie Rough”*

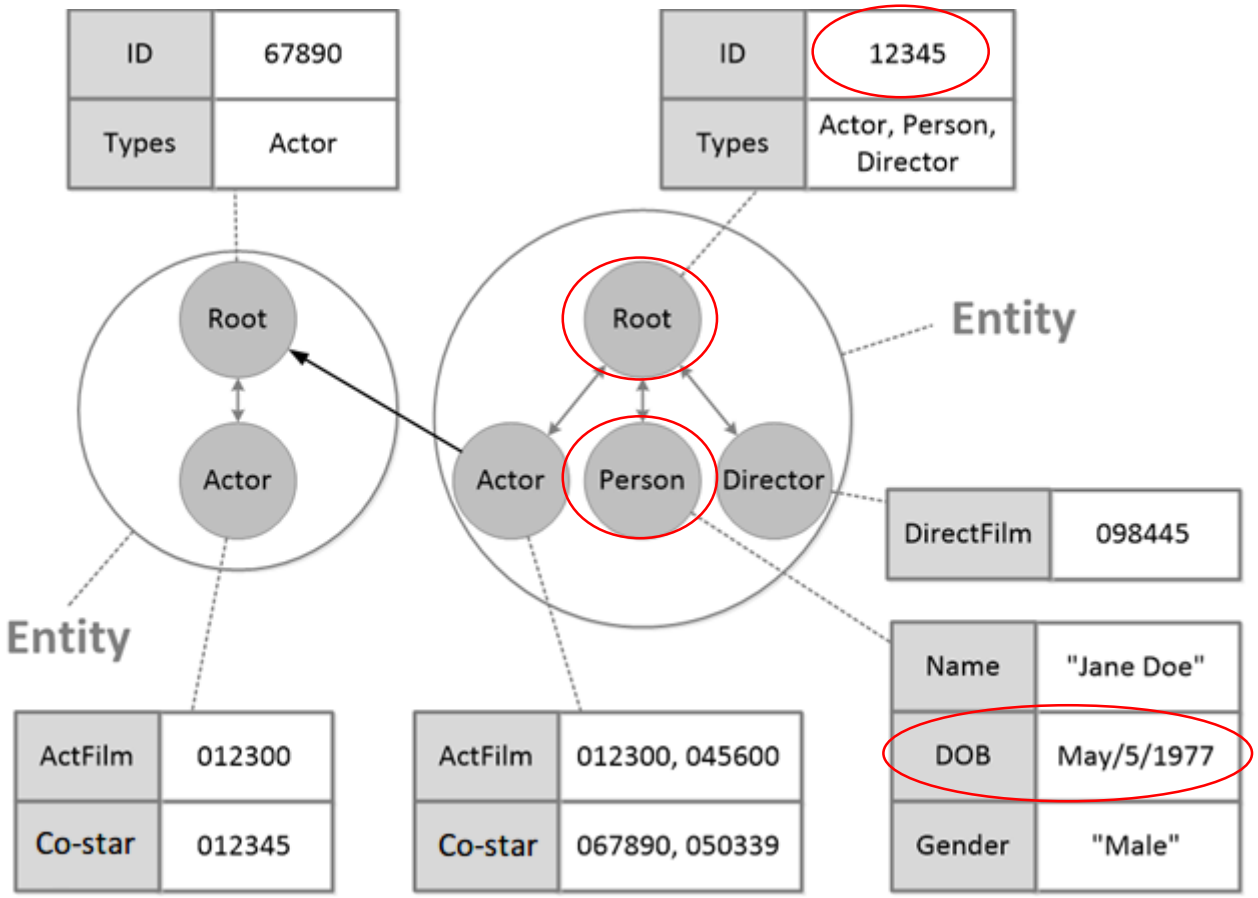
*123 mso/type.object.type mso/film.actor*
*123 mso/film.actor.film 789*
*789 mso/type.object.type mso/film.film*
*789 mso/type.object.name “Lassie Come Home”*



# Modeling multi-typed entities in a strongly-typed manner

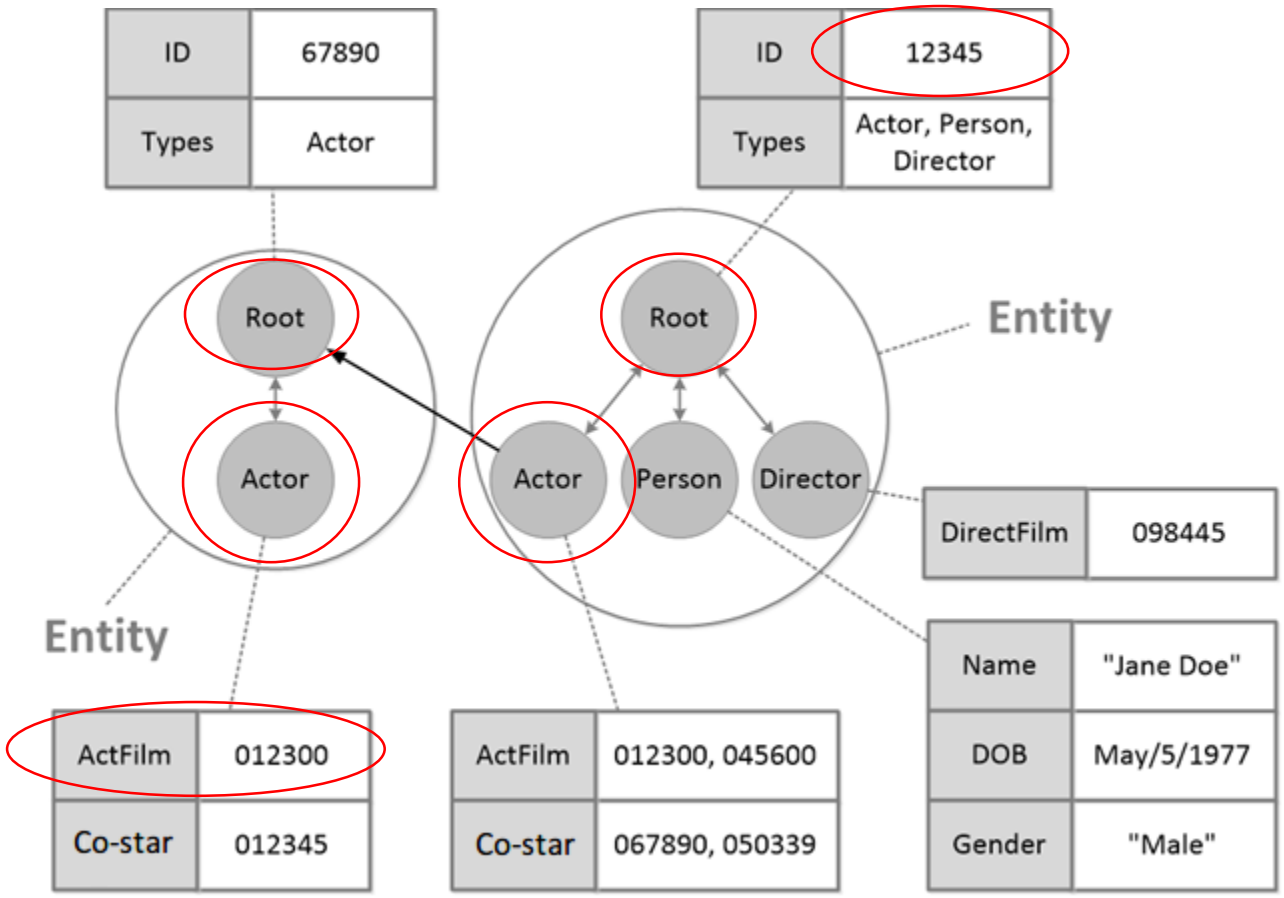


# Strongly-typed data access



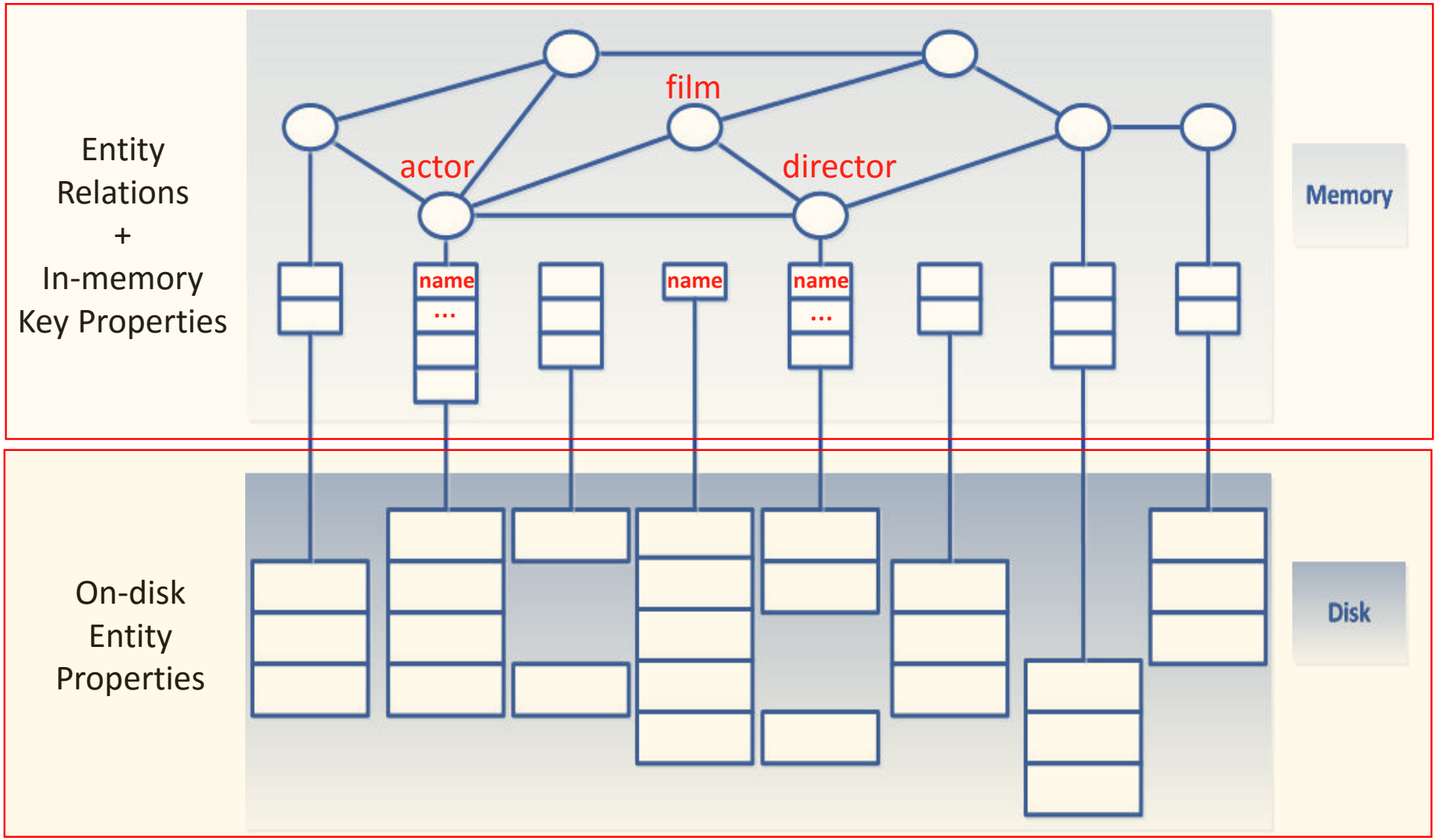

Get the DOB of entity 12345

# Strongly-typed data access

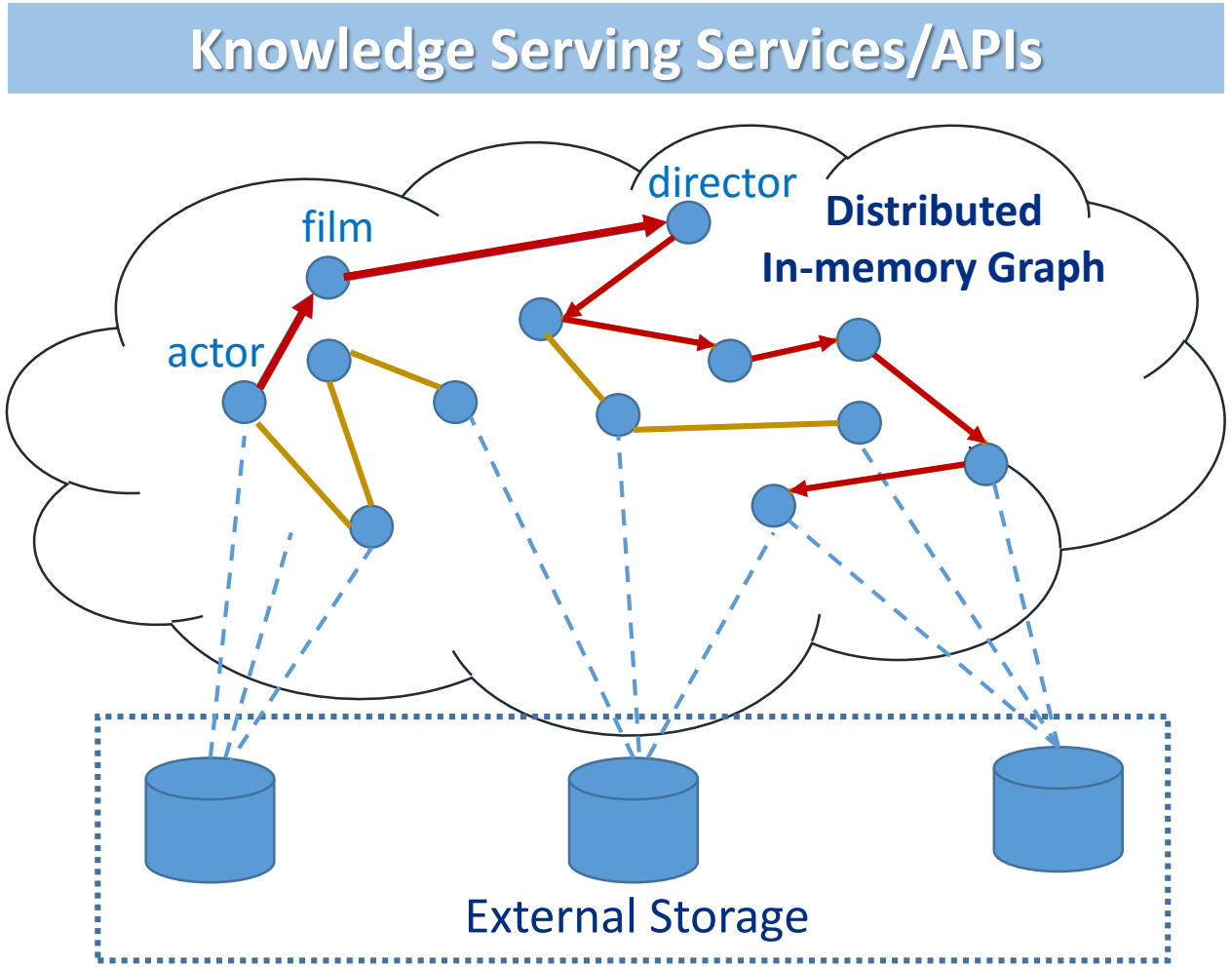



Get the films of actors co-starring with entity 12345

# Storage architecture



# Query knowledge via graph exploration



# Knowledge serving interfaces

Satori

Search

## Harvard University

award.presenting_organization, award.ranked_item, award.winner, book.author, education.academic_institution, education.educational_institution ...

## Harvard University

organization.organization, type.object

## harvard university

internet.social_network_user, people.person, type.object

## harvard university

internet.social_network_user, people.person, type.object

## Harvard University

local.entity, type.object

## Harvard University

local.entity, type.object

Prev Page Next Page

# Harvard University

Harvard University is an American private Ivy League research university located in Cambridge, Massachusetts, United States, established in 1636 by the Massachusetts legislature. Harvard is the oldest institution of higher learning in the United States and the first corporation (officially The President and Fellows of Harvard College) chartered in the country. Harvard's history, influence,...

**Types**

award.presenting_organization, award.ranked_item, award.winner, book.author, education.academic_institution, education.educational_institution ...

**Predicates**

education.educational_institution.total_enrollment
education.educational_institution.color
education.educational_institution.subsidiary_or_constituent_schools
education.educational_institution.number_of_staff
education.educational_institution.honorary_degrees_awarded
education.educational_institution.school_sports_team

Prev Page Next Page

**Values**

"Harvard Extension School"
"Harvard Medical School"
"Harvard Business School"
"Harvard College"
"Harvard Division of Continuing Education"
"John F. Kennedy School of Government"

Prev Page Next Page

Entity Explorer

# Schema Graph

Meta Graph of Satori

Schema Type:

mso/people.person Go

Schema Path:

mso/people.person ⇢ mso/book.author Go

**Fields:**

| | |
|---|---|
| .bust_measurement | mso/type.decimal |
| .date_of_birth | mso/type.datetime |
| .eye_color | mso/type.text |
| .first_name | mso/type.string |
| .hair_color | mso/type.text |
| .height | mso/type.decimal |
| .hips_measurement | mso/type.decimal |
| .last_name | mso/type.string |
| .waist_measurement | mso/type.decimal |
| .weight | mso/type.decimal |

**Links:**

| | |
|---|---|
| .business_employment_tenure | mso/business.employment_tenure |
| .children | mso/people.person |
| .city_of_birth | mso/location.location |

| | | |
|---|---|---|
| mso/people.person | .quotation | mso/media_common.quotation |
| mso/media_common.quotation | .character | mso/fictional_universe.character |
| mso/fictional_universe.character | .appears_in_these_fictional_universes | mso/fictional_universe.universe |
| mso/fictional_universe.universe | .literary_series | mso/book.literary_series |
| mso/book.literary_series | .author | mso/book.author |

# Schema Graph Services

# Satori Knowledge Graph Access API

| API Names | Availability | Description |
|---|---|---|
| GetEntityIdByName | Available | Gets a list of Trinity entity Ids by the specified entity name. |
| GetPredicatesByEntityId | Available | Gets a list of predicates for the entity with the specified Trinity entity Id. |
| GetValuesByEntityPredicate | Available | Gets the values of the specified predicates for the specified entity. |
| GetSubjectsByPredicateObject | Available | Gets the subjects for the given object and a predicate. |
| GetEntityIdBySatoriId | Available | Gets the corresponding Trinity entity Id for the specified Satori Guid. |
| GetSatoriIdByEntityId | Available | Gets the corresponding Satori Guid for the specified Trinity entity Id. |
| GetRankedEntityIdByName | Available | Gets a list of Trinity entity Ids by the specified entity name sorted by their static rank. |
| **GetScoredValuesByEntityPredicate** | Available | Gets the values of the specified predicates for the specified entity, sorted by confidence score. |
| GetSortScoredValuesByEntityPredicate | Available | Gets the values of the specified predicates for the specified entity, sorted by the column index (1 for... |
| GetEntityDescription | Available | Gets the description of the specified entityid. |

Testing: GetScoredValuesByEntityPredicate

Please input test parameters below:

EntityId 24604518639751

Predicate mso/film.actor.film

Submit

| PredicateValue | ConfidenceScore | OverallScore |
|---|---|---|
| 2987469205879 | 0.71 | 1.311128 |
| 116281907553515 | 0.71 | 1.409593 |
| 265920831012309 | 0.71 | 1.416611 |
| 58184534540412 | 0.71 | 1.710736 |
| 237856925167352 | 0.71 | 1.228339 |
| 57628320423344 | 0.71 | 1.272193 |

Page 1 of 1

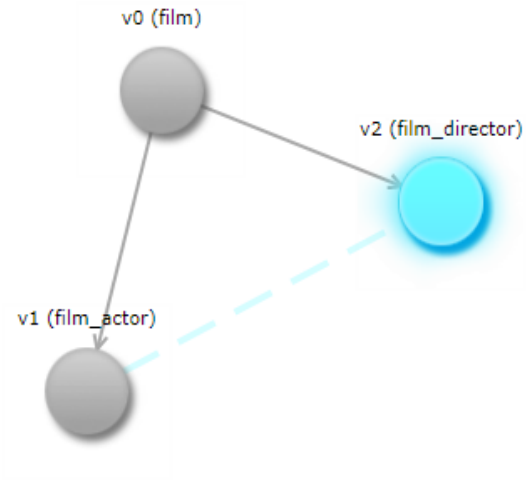
# Knowledge Serving APIs

Home TQL Schema Service API

Satori film Add

# Graphical Query Interface

View TQL Results




Node Information

Alias: v2
Type: film_director
URI: mso_film_director

Conditions:

Outputs:

Steve Saari
Steve Sacks
Steve Sale
Steve Salinaro
Steve Salisian
Steve Sanders
Steve Sanguedolce
Steve Saporito
Steve Savage
Steve Savitz
Steve Scheffler

You could add some conditions a

name = Steve S

Submit

Powered By Trinity Graph Engine

Contact Us

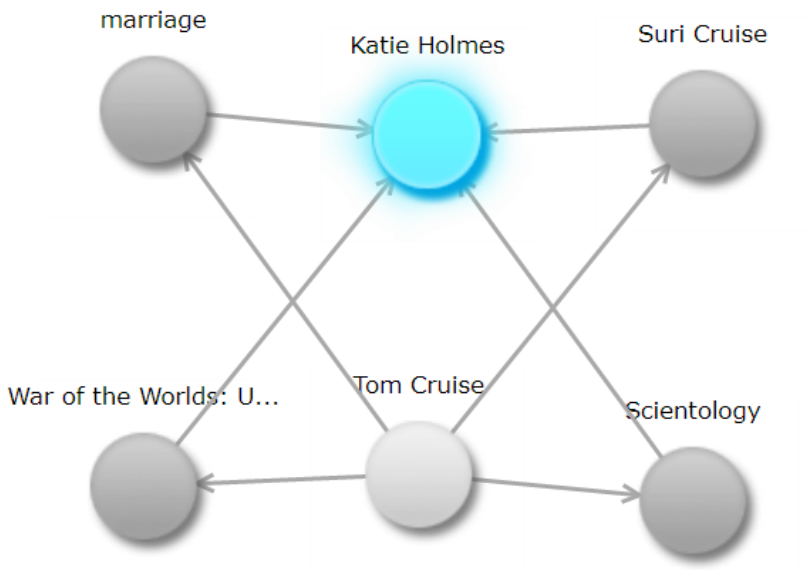
# Multi-hop relation search in knowledge graph

Home | Schema | API | Relation Search | People Relation Search

Satori [ ] Add Search

**Tom Cruise, Katie Holmes**

Results View




## Katie Holmes

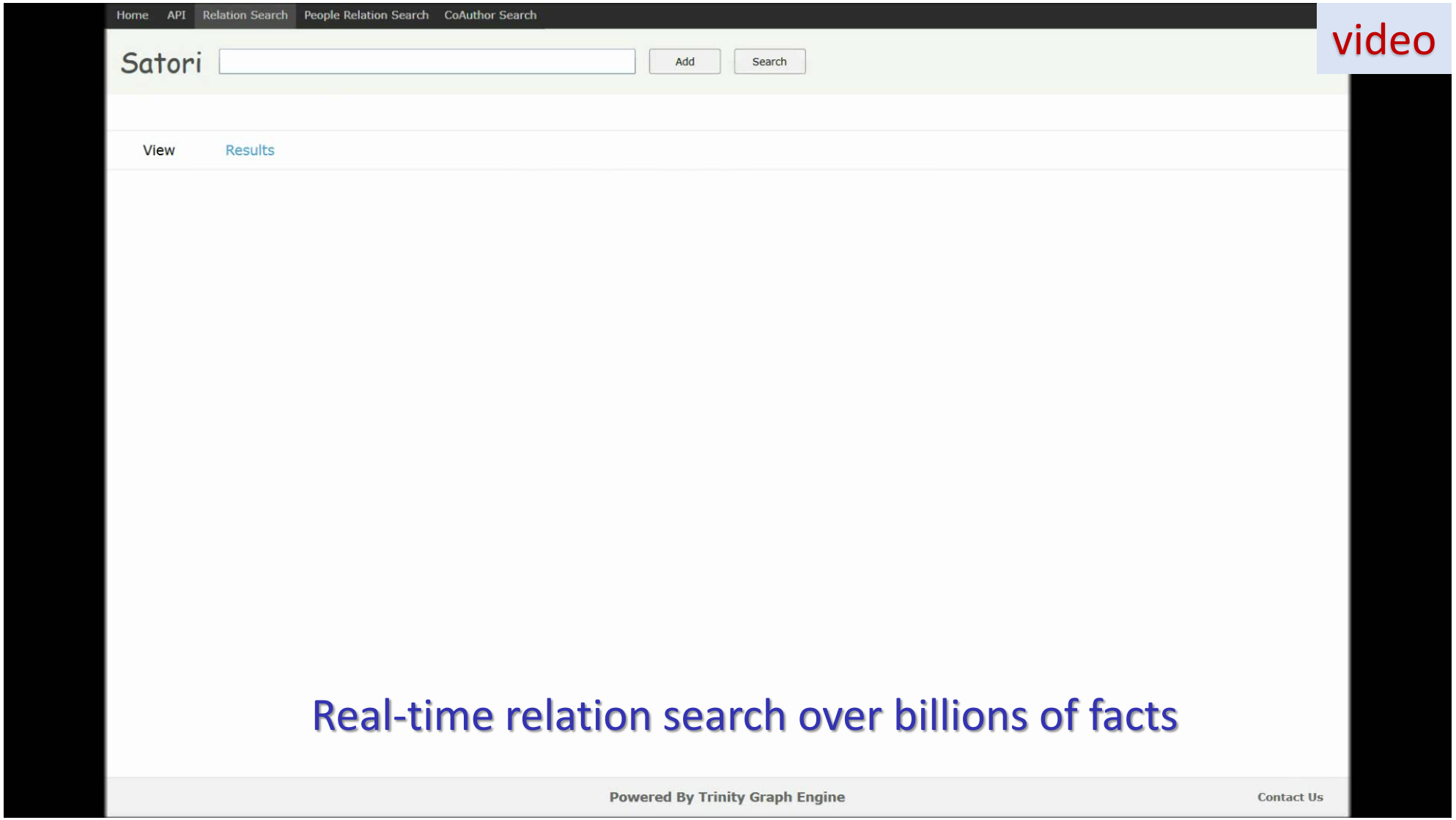
Kate Noelle "Katie" Holmes (born December 18, 1978) is an American actress and model who first achieved fame for her role as Joey Potter on The WB television teen drama Dawson's Creek from 1998 to 2003....

**Types**

award.nominee, award.ranked_item, award.winner, film.actor, film.writer, medicine.notable_person_with_medical_condition ...

**Powered By Trinity Graph Engine** **Contact Us**

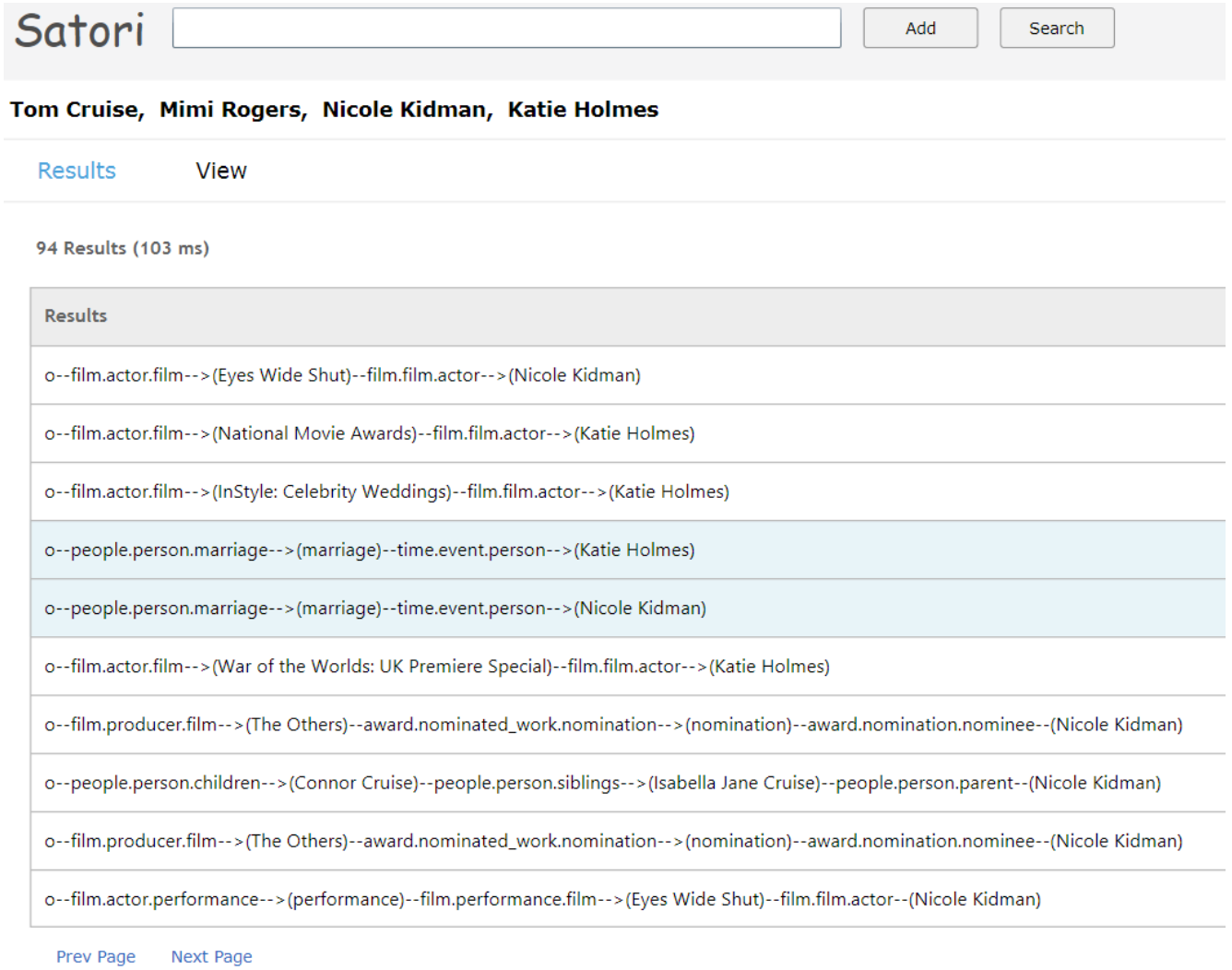
# Relation search in a large knowledge graph




video

Real-time relation search over billions of facts

# Multi-hop relation search in knowledge graph



# Applications of knowledge serving services

# 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

# 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. ...

# 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

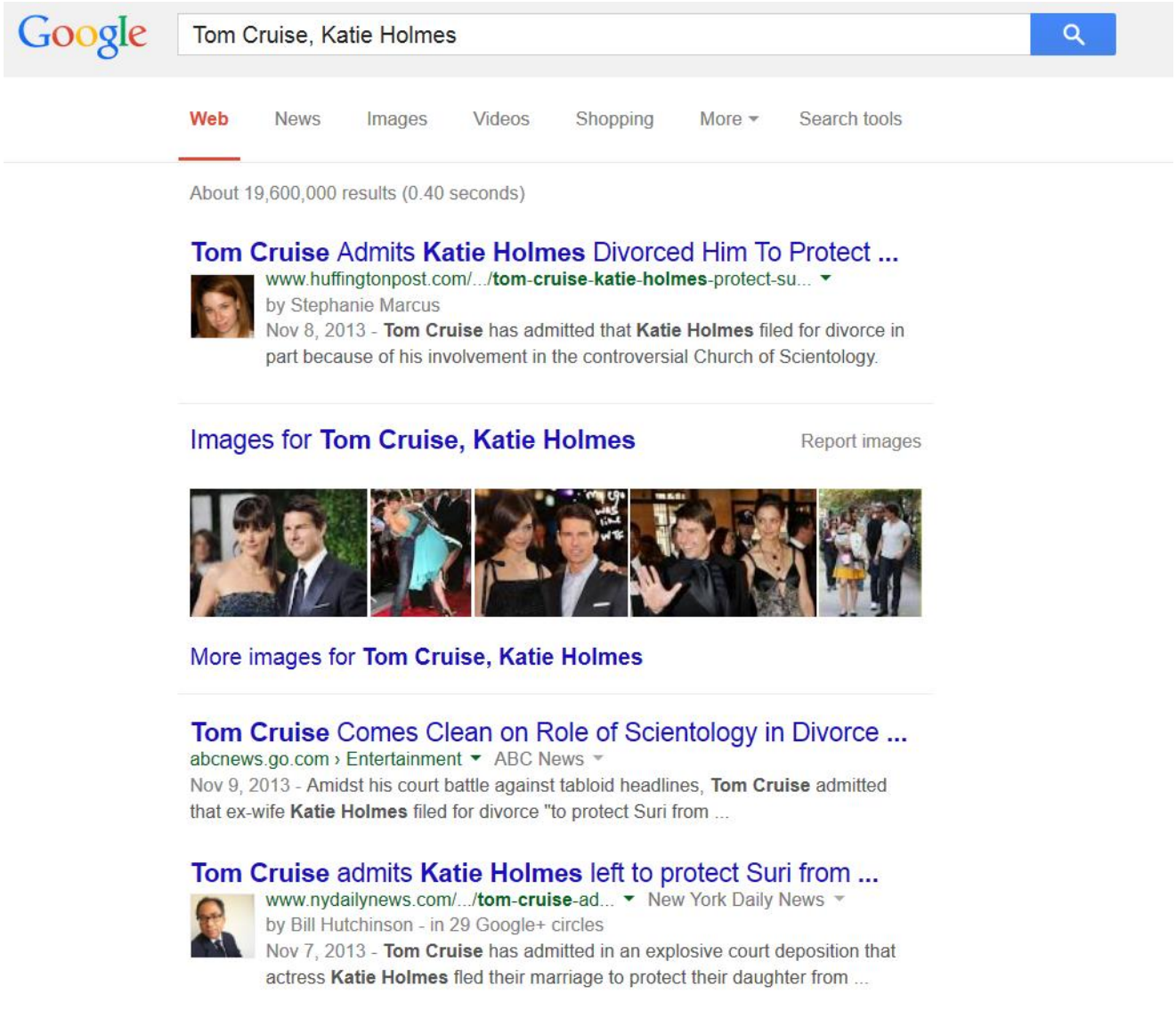
Katie Holmes
American Actress
Kate Noelle "Katie" Holmes is an American actress and model who first achieved fame for her role as Joey Pot…

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

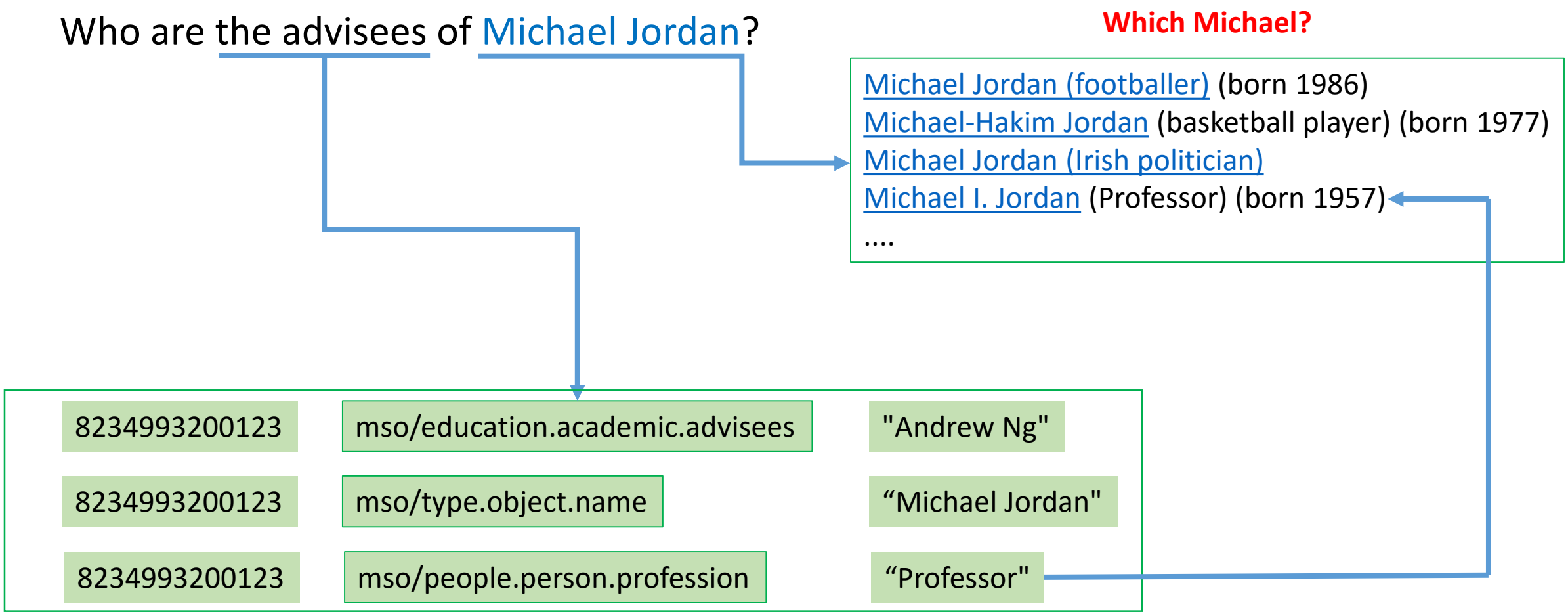
Related searches

- 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**

# Search results of Google

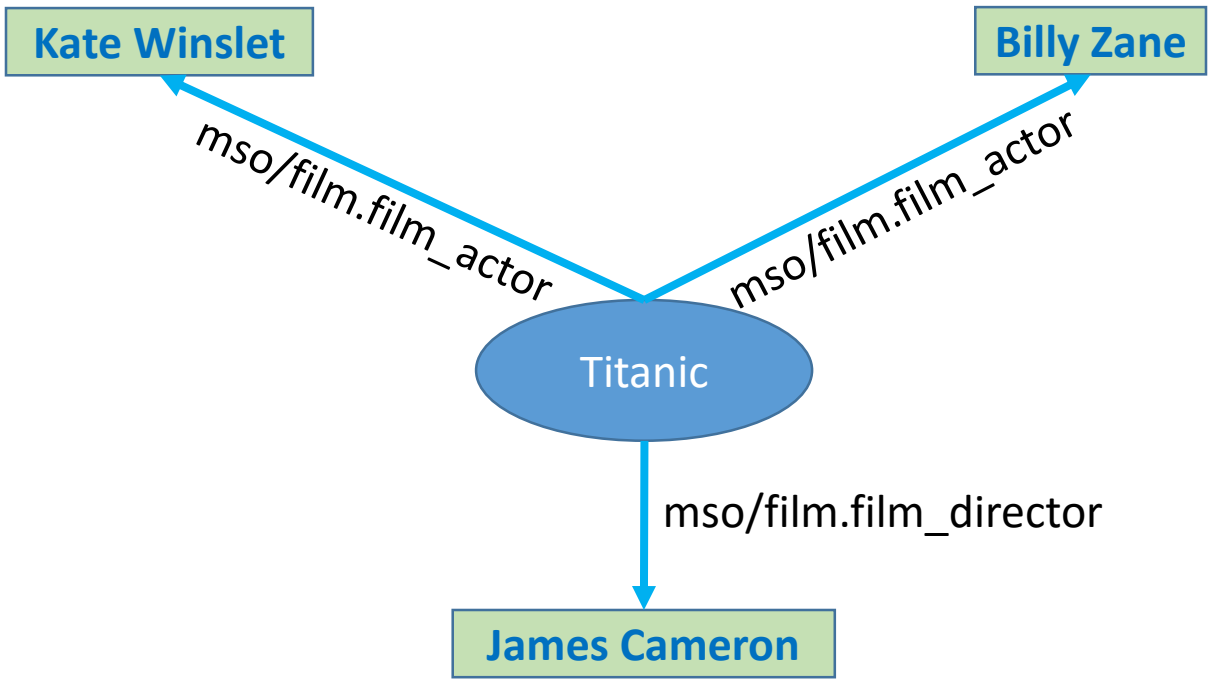


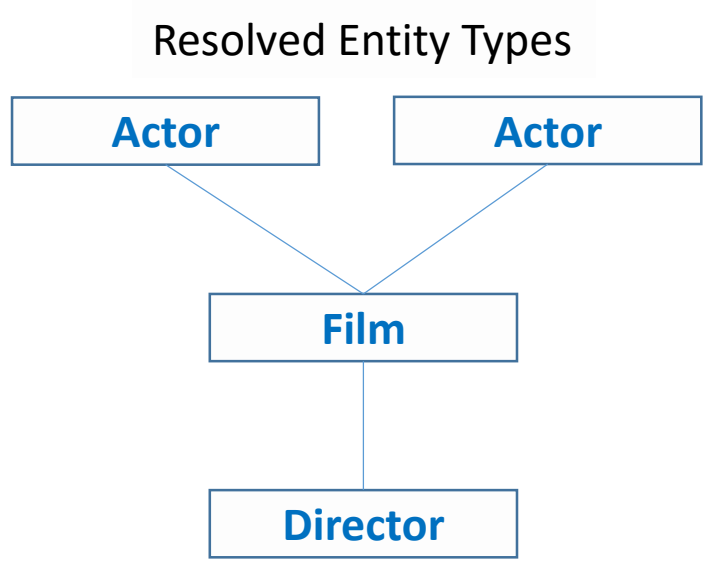
# Entity disambiguation and type resolving



# Discover linking entities

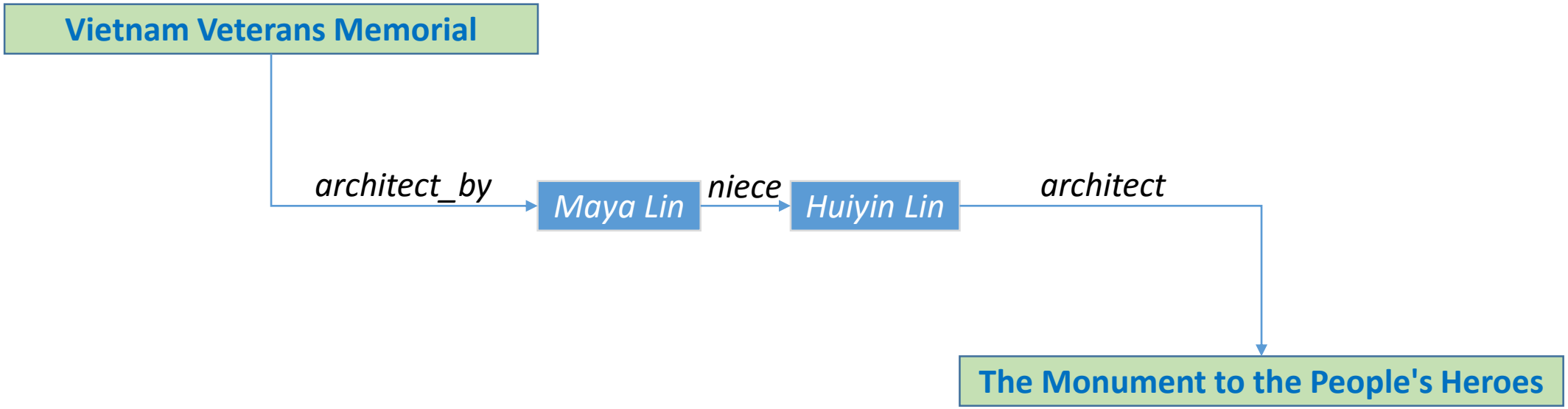
Given three entities *"Kate Winslet"*, *"Billy Zane"*, and *"James Cameron"*



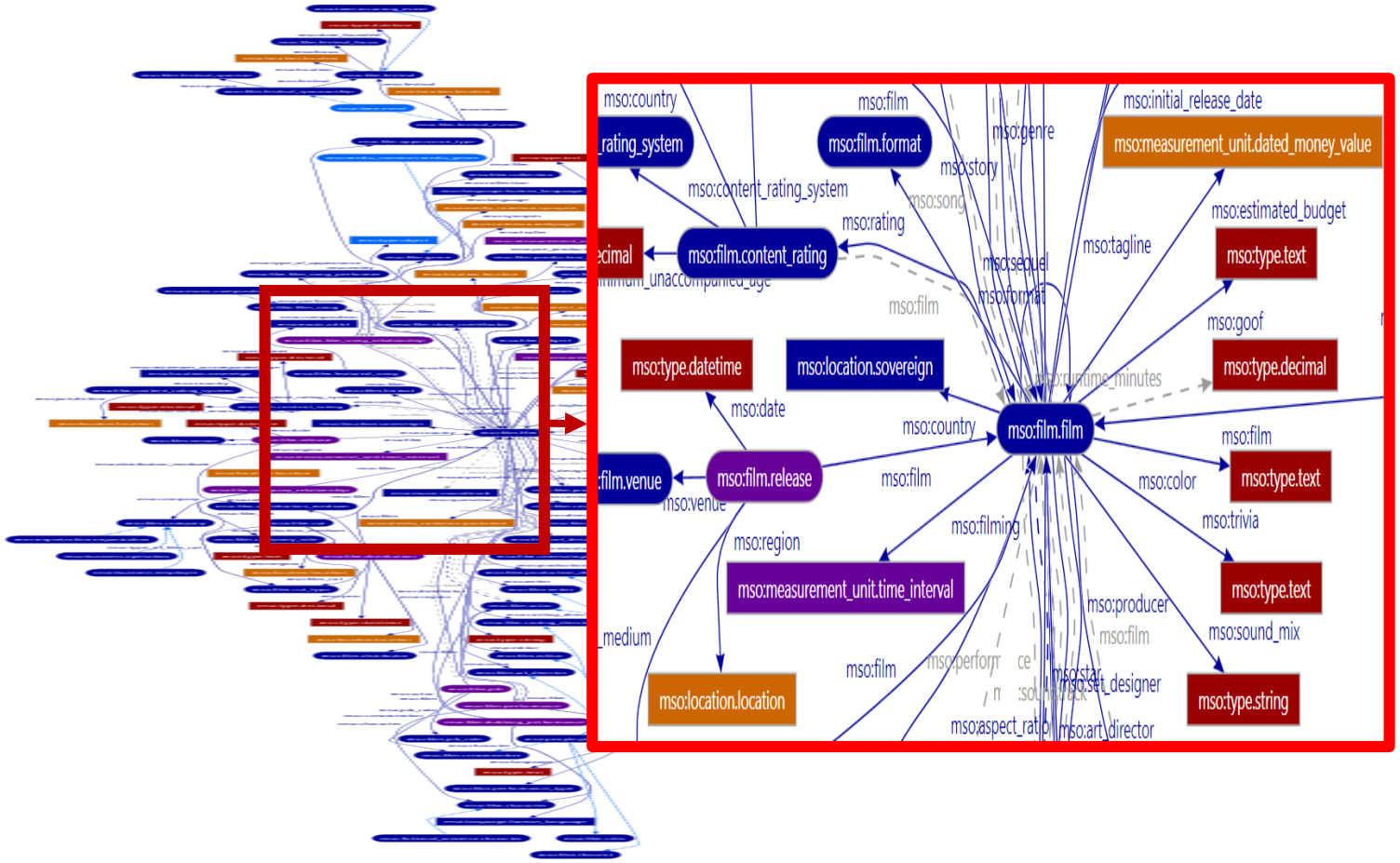
# Discover linking relations

Given two entities *"Vietnam Veterans Memorial"* and *"The Monument to the People's Heroes"*



# How can we make it fast enough

- Big data
  - emm, we have a large variety of tools available
- But, how do we handle “big schema” …


mso:country
mso:film
mso:initial_release_date
mso:film.format
mso:genre
mso:measurement_unit.dated_money_value
mso:story
mso:content_rating_system
mso:song
mso:estimated_budget
mso:rating
mso:film.content_rating
mso:tagline
mso:type.text
mso:sequel
mso:format
mso:film
mso:goof
mso:type.datetime
mso:location.sovereign
mso:type.decimal
mso:runtime_minutes
mso:date
mso:country
mso:film.film
mso:film
mso:film.release
mso:film
mso:color
mso:type.text
mso:venue
mso:trivia
mso:region
mso:filming
mso:measurement_unit.time_interval
mso:type.text
mso:producer
mso:film
mso:sound_mix
mso:film
mso:location.location
mso:set_designer
mso:type.string
mso:art_director

# Big Schema

# How can we make it fast enough

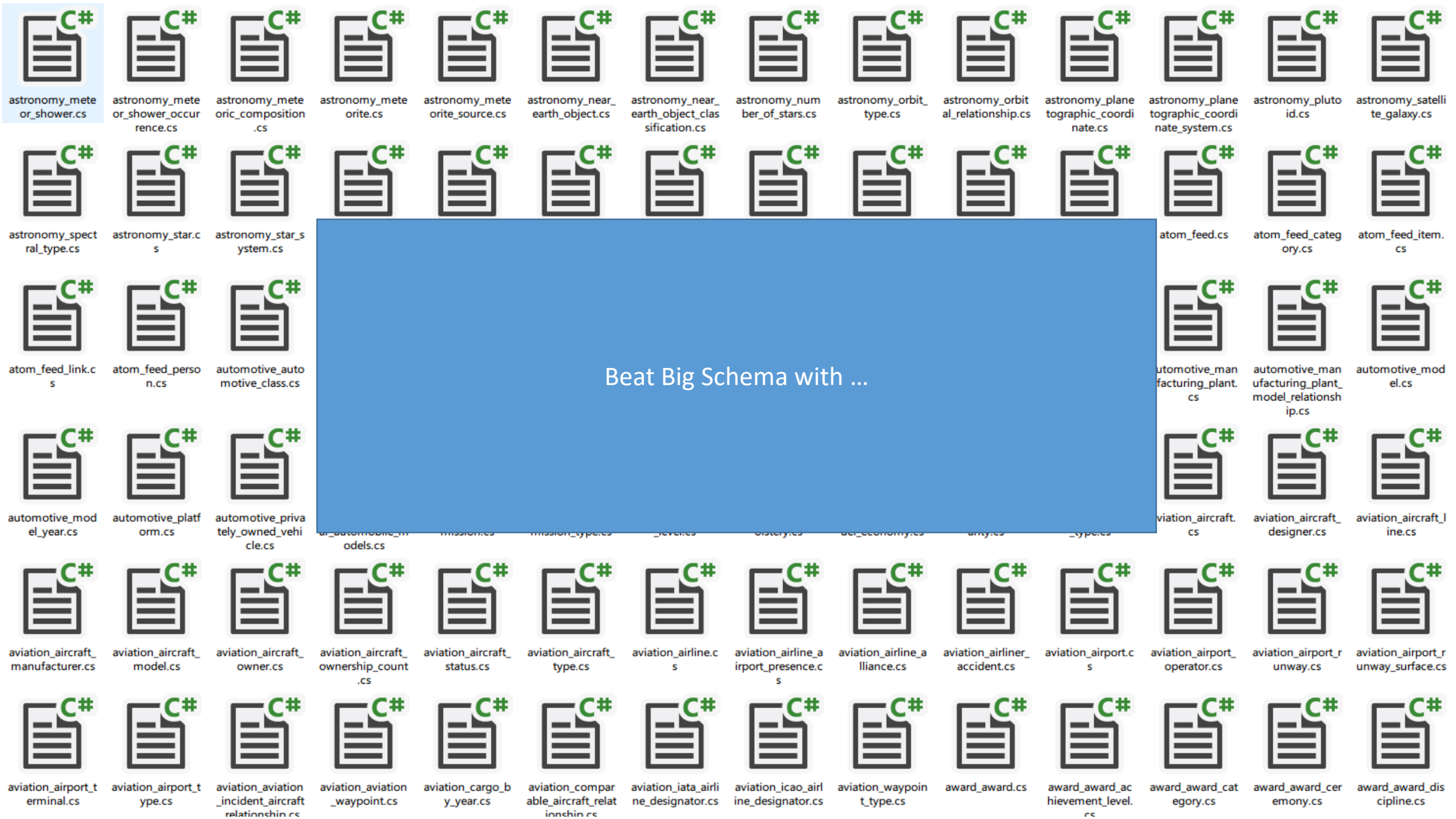
- Big data
  - emm, we have a large variety of tools available

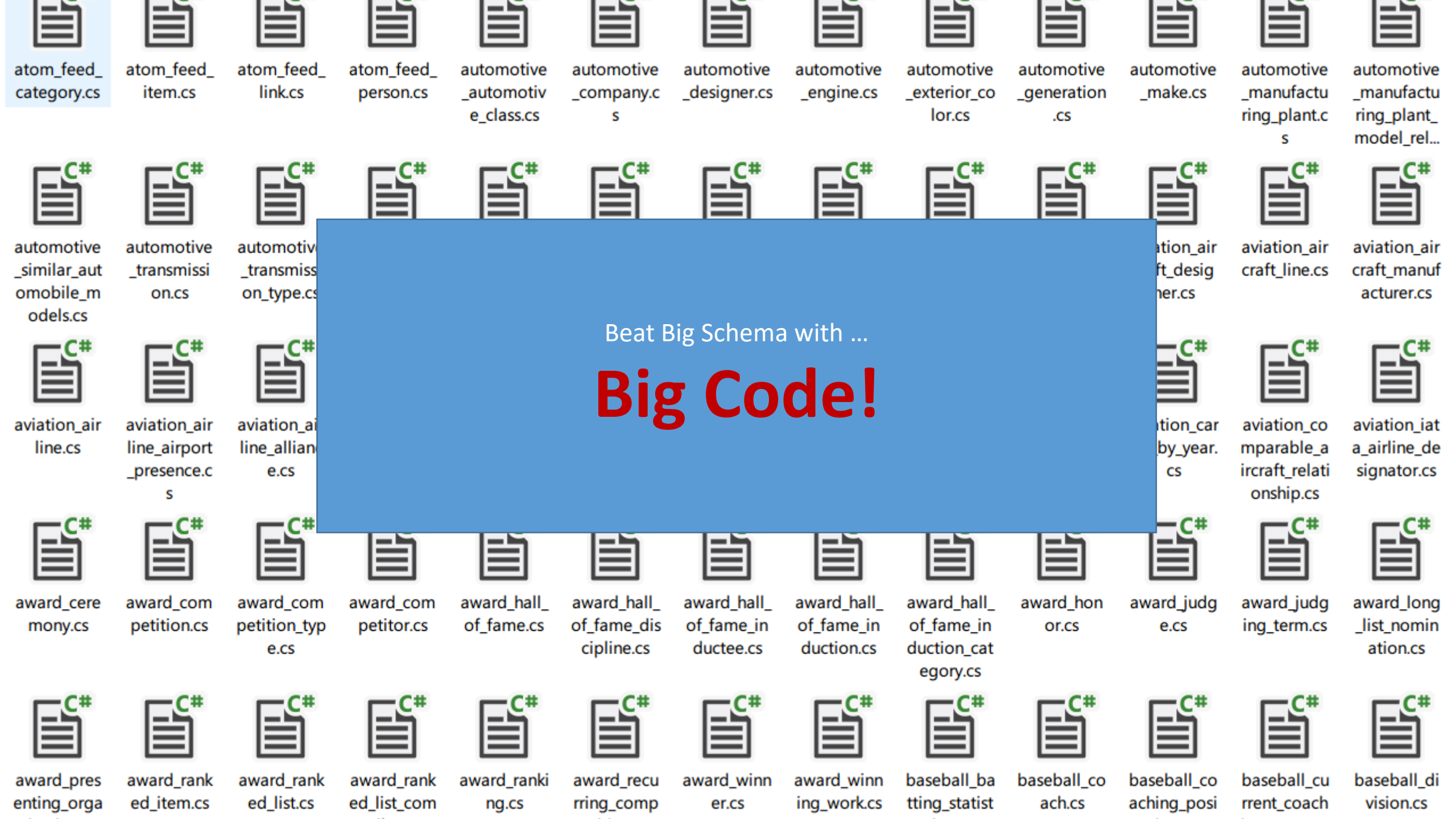
- But, how do we handle “big schema” …

  If we treat everything as texts and build indexes for these piles of words

  - Inefficient data processing (weakly-typed system)
  - Limited search functionality we can provide

Beat Big Schema with …


astronomy_meteor_shower.cs
astronomy_meteor_shower_occurrence.cs
astronomy_meteoric_composition.cs
astronomy_meteorite.cs
astronomy_meteorite_source.cs
astronomy_near_earth_object.cs
astronomy_near_earth_object_classification.cs
astronomy_number_of_stars.cs
astronomy_orbit_type.cs
astronomy_orbital_relationship.cs
astronomy_planetographic_coordinate.cs
astronomy_planetographic_coordinate_system.cs
astronomy_pluto_id.cs
astronomy_satellite_galaxy.cs
astronomy_spectral_type.cs
astronomy_star.cs
astronomy_star_system.cs
atom_feed.cs
atom_feed_category.cs
atom_feed_item.cs
atom_feed_link.cs
atom_feed_person.cs
automotive_automotive_class.cs
automotive_manufacturing_plant.cs
automotive_manufacturing_plant_model_relationship.cs
automotive_model.cs
automotive_model_year.cs
automotive_platform.cs
automotive_privately_owned_vehicle.cs
aviation_aircraft.cs
aviation_aircraft_designer.cs
aviation_aircraft_line.cs
aviation_aircraft_manufacturer.cs
aviation_aircraft_model.cs
aviation_aircraft_owner.cs
aviation_aircraft_ownership_count.cs
aviation_aircraft_status.cs
aviation_aircraft_type.cs
aviation_airline.cs
aviation_airline_airport_presence.cs
aviation_airline_alliance.cs
aviation_airliner_accident.cs
aviation_airport.cs
aviation_airport_operator.cs
aviation_airport_runway.cs
aviation_airport_runway_surface.cs
aviation_airport_terminal.cs
aviation_airport_type.cs
aviation_aviation_incident_aircraft_relationship.cs
aviation_aviation_waypoint.cs
aviation_cargo_by_year.cs
aviation_comparable_aircraft_relationship.cs
aviation_iata_airline_designator.cs
aviation_icao_airline_designator.cs
aviation_waypoint_type.cs
award_award.cs
award_award_achievement_level.cs
award_award_category.cs
award_award_ceremony.cs
award_award_discipline.cs
Beat Big Schema with …


atom_feed_category.cs
atom_feed_item.cs
atom_feed_link.cs
atom_feed_person.cs
automotive_automotive_class.cs
automotive_company.cs
automotive_designer.cs
automotive_engine.cs
automotive_exterior_color.cs
automotive_generation.cs
automotive_make.cs
automotive_manufacturing_plant.cs
automotive_manufacturing_plant_model_rel...
automotive_similar_automobile_models.cs
automotive_transmission.cs
aviation_aircraft_line.cs
aviation_aircraft_manufacturer.cs
aviation_airline.cs
aviation_airline_airport_presence.cs
aviation_comparable_aircraft_relationship.cs
aviation_iata_airline_designator.cs
award_ceremony.cs
award_competition.cs
award_competition_type.cs
award_competitor.cs
award_hall_of_fame.cs
award_hall_of_fame_discipline.cs
award_hall_of_fame_inductee.cs
award_hall_of_fame_induction.cs
award_hall_of_fame_induction_category.cs
award_honor.cs
award_judge.cs
award_judging_term.cs
award_long_list_nomination.cs
award_presenting_orga
award_ranked_item.cs
award_ranked_list.cs
award_ranked_list_com
award_ranking.cs
award_recurring_comp
award_winner.cs
award_winning_work.cs
baseball_batting_statist
baseball_coach.cs
baseball_coaching_posi
baseball_current_coach
baseball_division.cs
Beat Big Schema with …
Big Code!


# Freebase Graph:

- Generated lines of code for Freebase: **8,868,163**
- Bytes of code: **446,747,058**

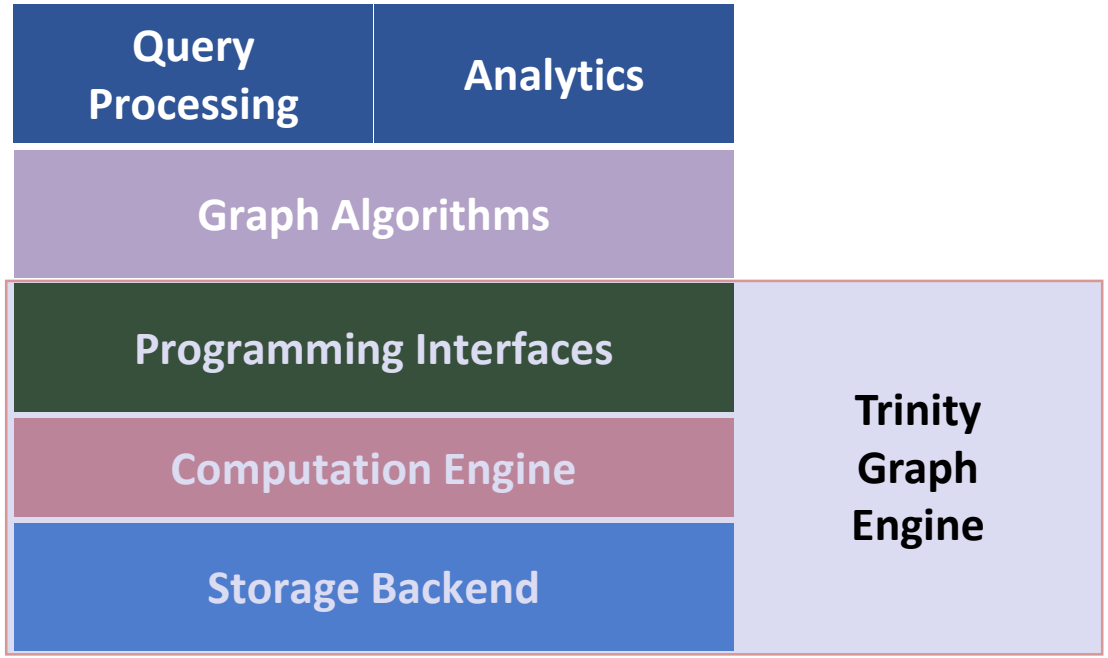
# What is the huge amount of code for?

- Provides extremely fine-grained data access methods best matching the data

= Efficiency

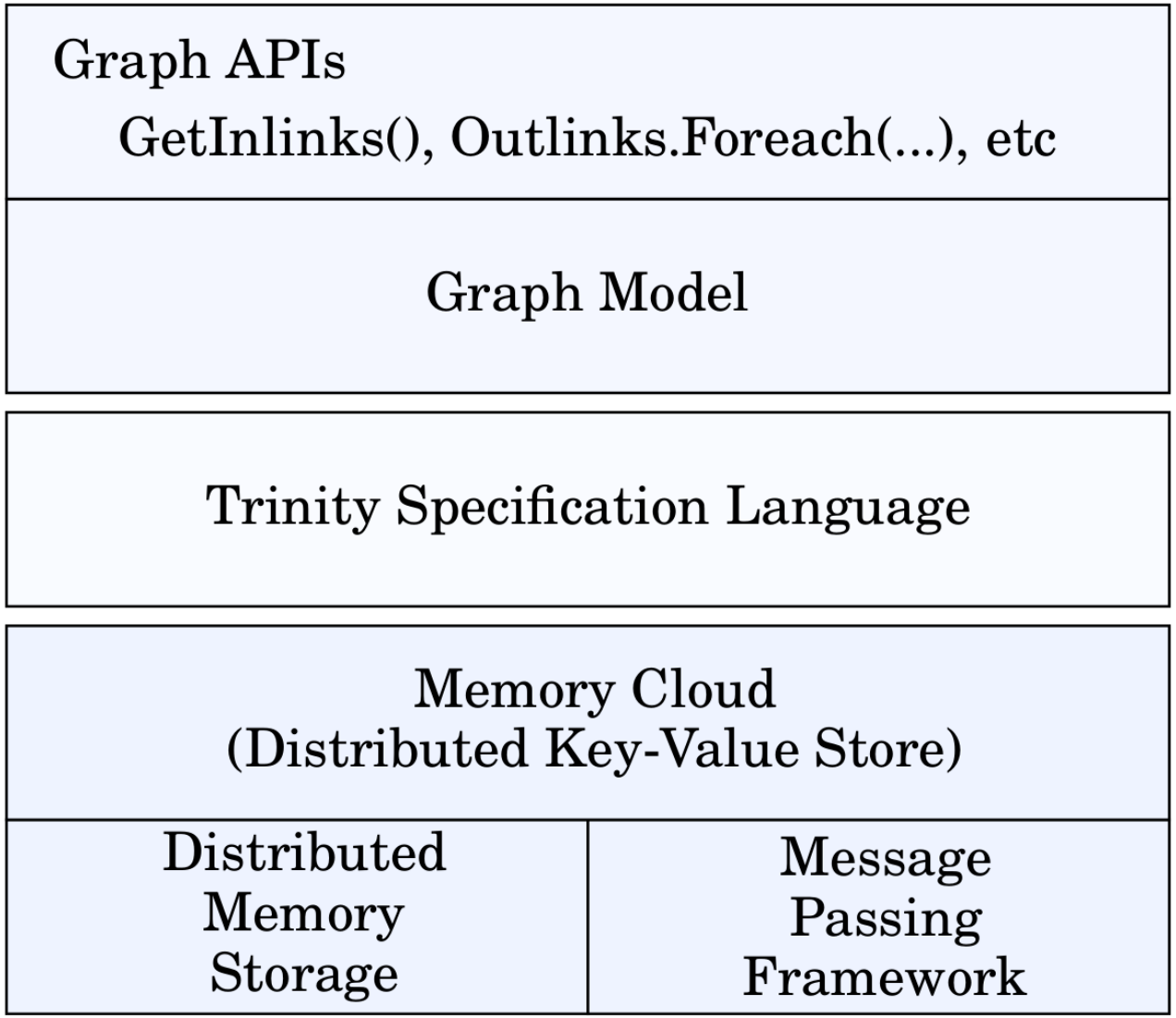
# A brief introduction to the underlying knowledge serving infrastructure
# Trinity Graph Engine

# What Trinity Graph Engine is

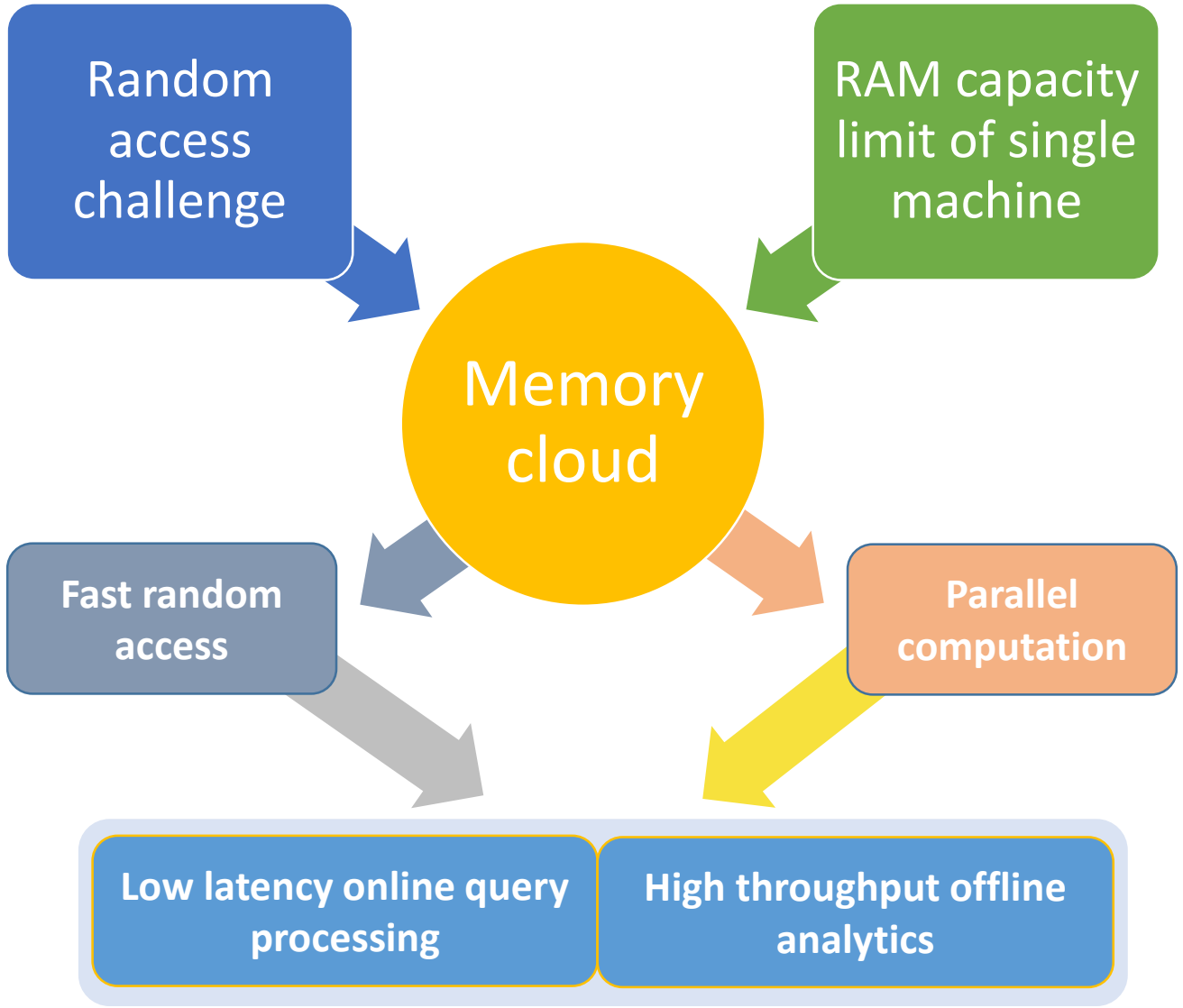



A general graph processing stack

# System stack



# Design rationale of memory cloud



# Design philosophy

**Not a one-size-fits-all graph system, but a graph engine**

Flexible data and computation modeling capability

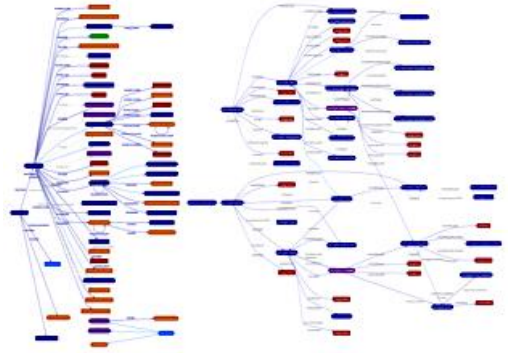
Trinity can morph into
a large variety of graph processing systems

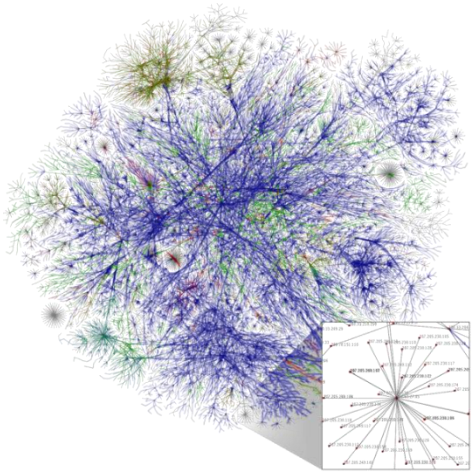
***Trinity*** **= Graph Modeling Tools +**
**Distributed In-memory Data Store +**
**Declarative Programming Model**

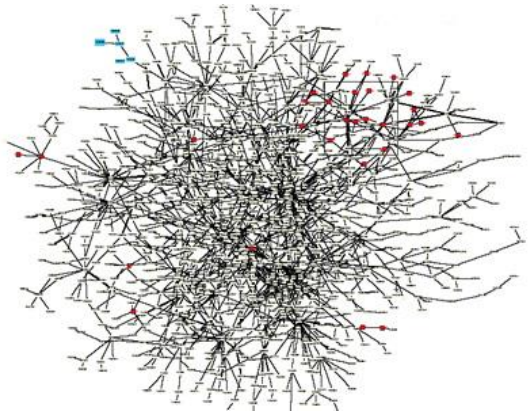
Trinity is a highly extensible graph computation engine.

# Why is extensibility important?

## Diversity challenges in graph computing







Diversity of graphs

- Online query processing
  - Shortest path query
  - Subgraph matching query
  - SPARQL query
  - …
- Offline graph analytics
  - PageRank
  - Community detection
  - …
- Other graph operations
  - Graph generation, visualization, interactive exploration, etc.

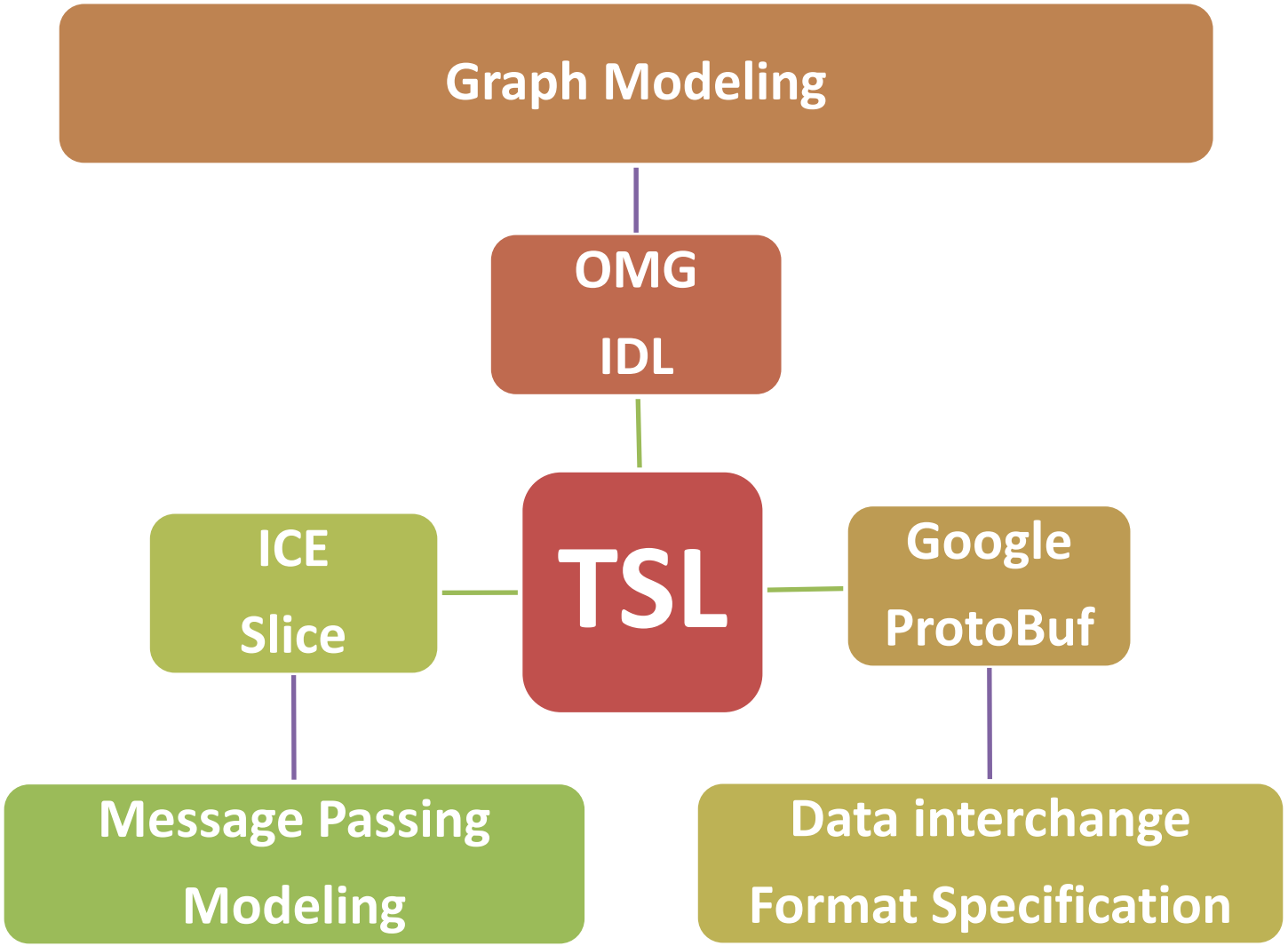
Diversity of computations

# High extensibility via Trinity Specification Language

- Due to the diversity of graphs and the diversity of graph applications, it is hard to efficiently support various graph computations using a fixed graph model and fixed computation paradigms.

- Instead of using a fixed graph model and fixed computation paradigms, Trinity allows users to specify graph model and distributed computation protocols via a declarative language called TSL (**Trinity Specification Language**).

# Trinity Specification Language



# Modeling a simple movie and actor Graph

```
[CellType: NodeCell]
cell struct Movie
{
    string Name;
    [EdgeType: SimpleEdge, ReferencedCell: Actor]
    List<long> Actors;
}
[CellType: NodeCell]
cell struct Actor
{
    string Name;
    [EdgeType: SimpleEdge, ReferencedCell: Movie]
    List<long> Movies;
}
```

# Runtime object (C#, Java, etc.)

- User friendly manipulation interfaces
  - *int id = object.Id*; or *object.inlinks[0] = 3242*;
- Large memory overhead
  - An empty c# runtime object requires 24 bits on 64-bit system and 12 bit on 32-bit system
- Object cannot be referenced across machine boundary
- Costly serialization and deserialization

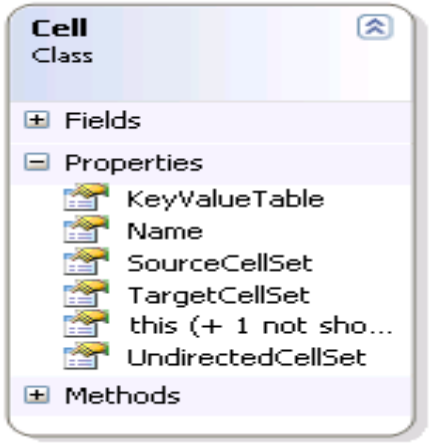
# Trinity Cell

- User-specified data schema (strongly-typed)
- Compact (blob)
- Globally addressable
- Zero serialization/deserialization overhead

# Blob vs. runtime object

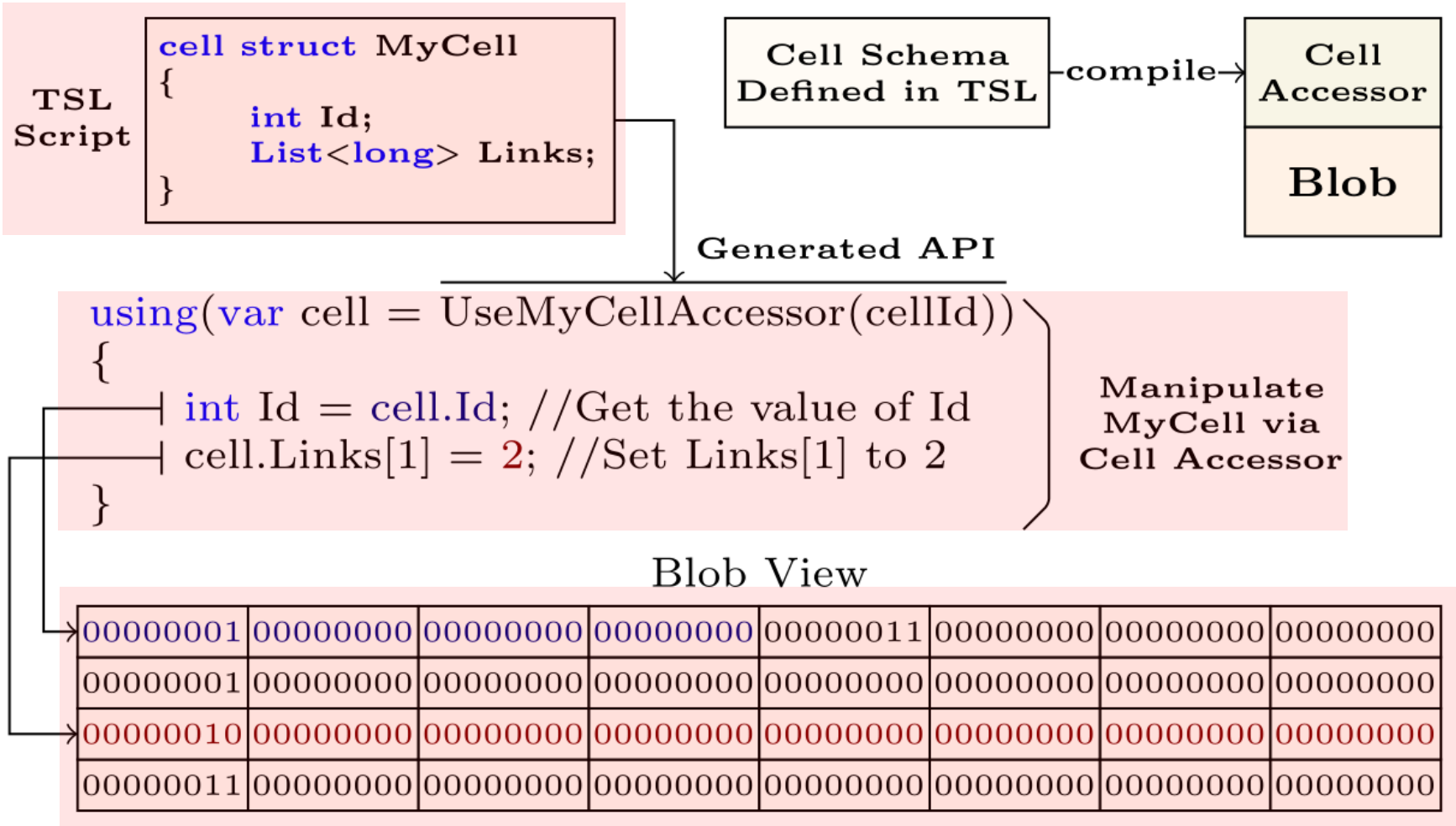
Economical and efficient

***vs.***




Easy to use

# TSL-enabled cell accessor: efficient and user-friendly



# Modeling message passing

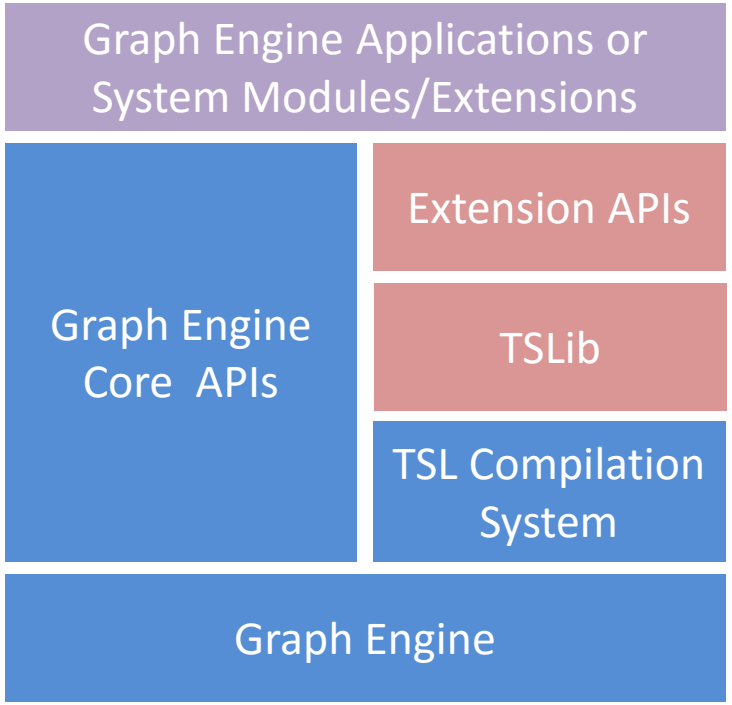
```
struct MyMessage
{
        string Text;
}
protocol Echo
{
        Type: Syn;
        Request: MyMessage;
        Response: MyMessage;
}
```

# Why TSL?

- TSL allows users to define graph schemata,  and communication protocols through declarative interfaces.

- TSL makes Trinity memory cloud beyond a key-value store
  - Users are allowed to freely define the data schema
  - TSL makes message passing programming ever so easy

# High extensibility




Graph Engine API Stack

With the **TSL** subsystem, Trinity can morph into a large variety of graph processing systems.

# Strengths of Graph Engine

- Solid technology accumulation in parallel large graph processing
  - Stabilized for a few years
- Seamless integration with Visual Studio and Azure
  - Offers developers the best experience in both development and deployment
- Unique capability with rich features
  - The first cloud-ready distributed engine for real-time graph processing

# Graph Engine VSExtension

Free

Graph Engine SDK for developing and deploying Graph Engine a

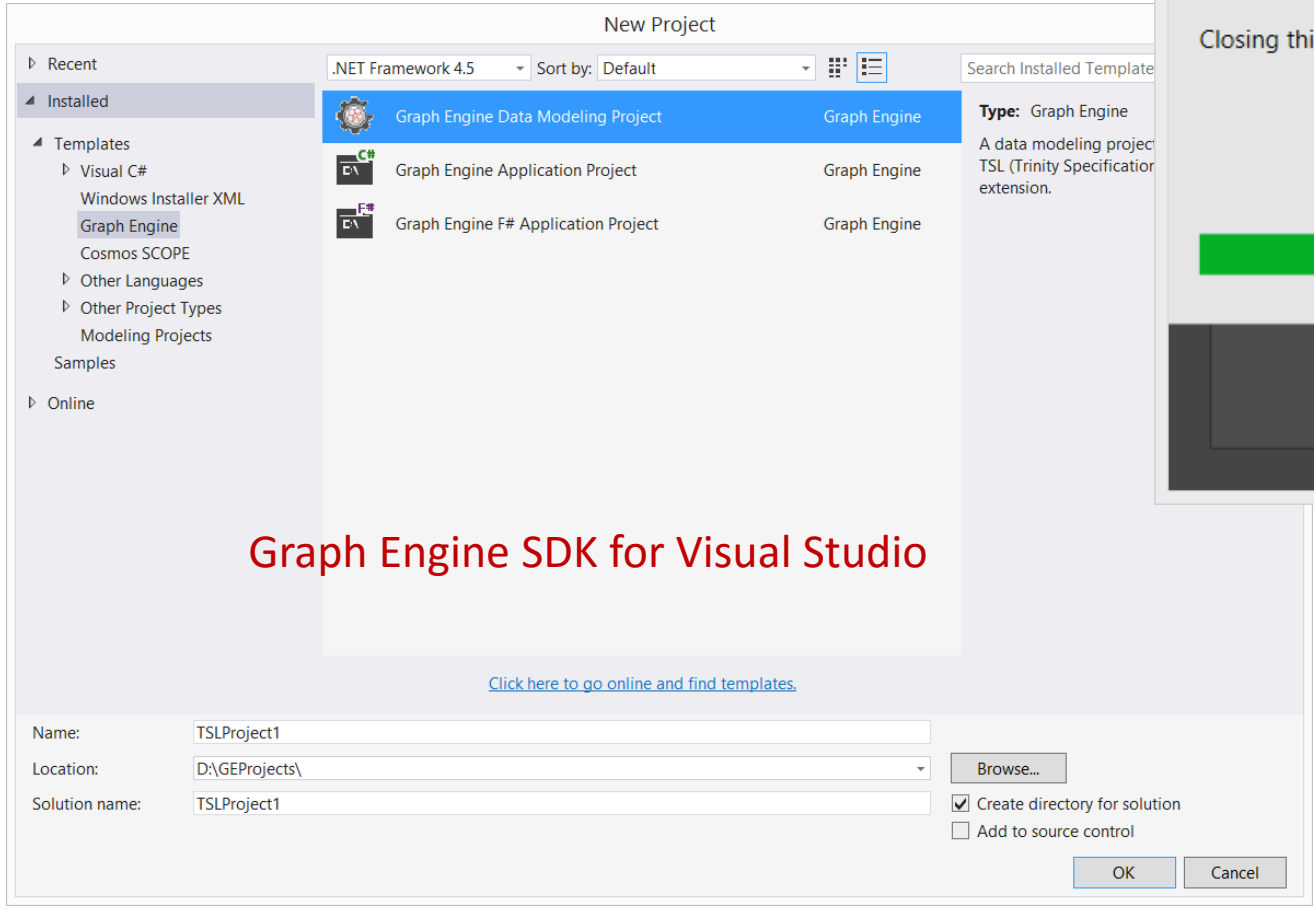
CREATED BY Graph Engine Team (Microsoft Corporation)

UPDATED

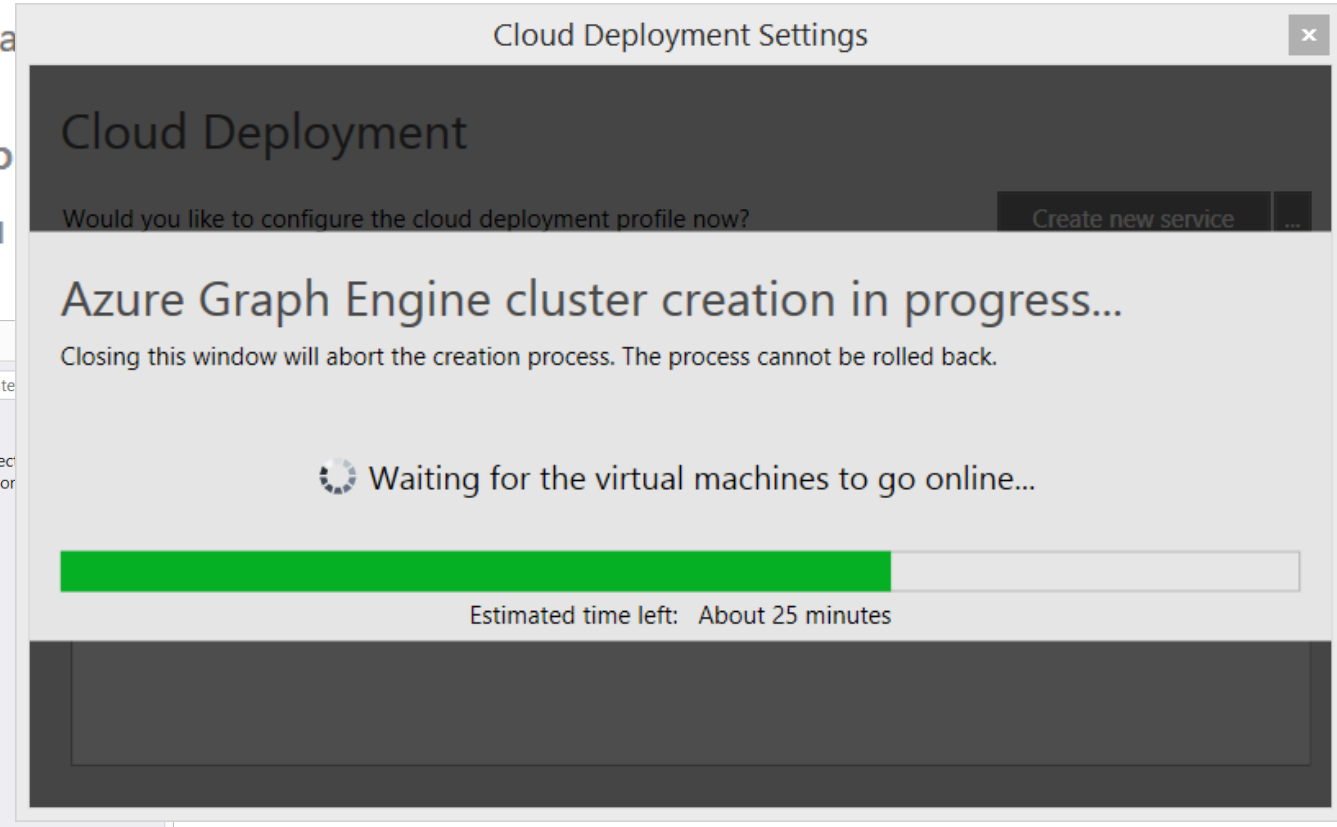
VERSION

REVIEWS ★★★★½ (3) Review

LICENSE




Graph Engine SDK for Visual Studio




Seamless integration with Visual Studio and Azure

# Graph Engine

SERVING BIG GRAPH IN REAL-TIME

http://graphengine.io

**Graph Engine Website**

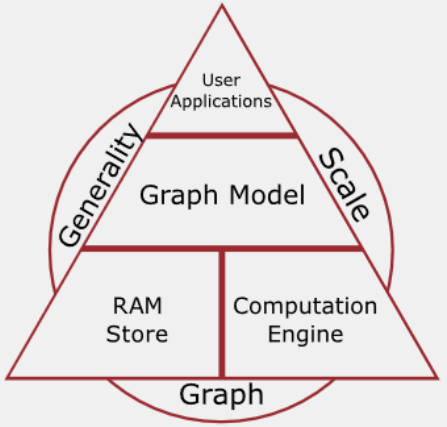
## Graph Engine

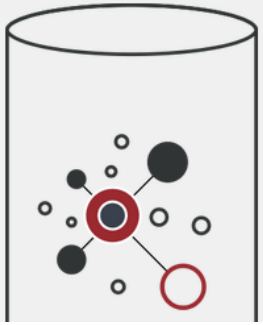
### = RAM Store + Computation Engine + Graph Model

Graph Engine (GE) is a highly modularized graph processing system, underpinned by a strongly-typed RAM store and a general computation engine.

The distributed RAM store provides a globally addressable high-performance key-value store over a cluster of machines. Through the RAM store, GE enables the fast random data access power over a large distributed data set.

The capability of fast data exploration and distributed parallel computing makes GE a natural large graph processing platform. GE supports both low-latency online query processing and high-throughput offline analytics on billion-node large graphs.



## Strongly-typed RAM Store

### Schema Matters

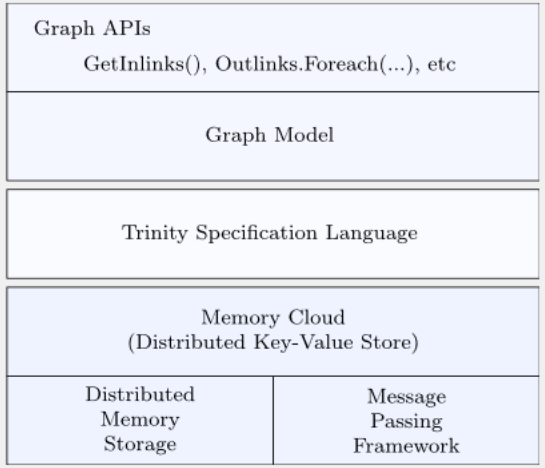
Schema does matter when we need to process data efficiently. Strongly-typed data modeling is crucial for compact data storage, fast data access, and clear data semantics.

### One Byte Counts

GE is good at managing billions of run-time objects of varied sizes. One byte counts as the number of objects goes large. GE provides fast memory allocation and efficient memory reallocation with

# Graph Engine Basics

GE is both a RAM store and a computation engine. As a RAM store, GE organizes the main memory* of a cluster of machines as a globally addressable address space (a memory cloud) to store large scale data sets; as a computation engine, GE provides user-customized APIs to implement graph processing logic.




http://graphengine.io

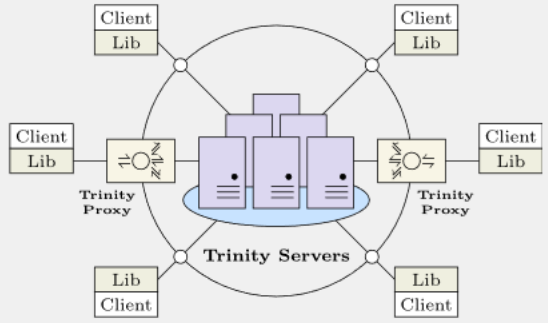
**Online Manual**

This figure shows the stack of GE system layers. The memory cloud is a distributed key-value store, which is supported by a memory storage module and a message passing framework. The memory storage module manages the main memory of a cluster of machines and provides mechanisms for concurrency control. The network communication module provides an efficient, one-sided, machine-to-machine message passing infrastructure.

GE provides a specification language called TSL (Trinity specification language) that bridges the graph model and the underlying storage and computation infrastructure. It is hard, if not entirely impossible, to support efficient, general-purpose graph computations using a fixed graph schema due to the diversity of graph data and application needs. Instead of using a fixed graph schema and fixed computation models, GE allows users to use TSL to specify graph schemata, communication protocols, and computational paradigms.

GE has two running modes, *embedded mode* and *distributed mode*. In the embedded mode, GE serves as an in-process library. In the distributed mode, GE can be deployed on one or more machines.

When deployed in the distributed mode, GE consists of a number of system components that communicate through a network. A GE component may have one or more following roles: I) storing data; II) handling messages and performing computations; III) interacting with clients. According to the roles played by the components, we classify GE components into three categories: *Server*, *Proxy*, and *Client*.




- *Server*. A server plays two roles: storing data and performing computations on the data. Computations usually involve sending messages to and receiving messages from other GE components.
- *Proxy*. A proxy handles messages but does not own a data partition. It usually serves as a middle tier between servers and clients. For example, a proxy can serve as a query aggregator: it dispatches requests received

# Thanks!

http://www.graphengine.io/