



Social Network Analysis at LinkedIn

Mitul Tiwari

Search, Network, and Analytics (SNA)

LinkedIn

Who am I?



Mitul Tiwari

Senior Research Engineer at LinkedIn

San Francisco Bay Area | Computer Software


Current

Senior Research Engineer at LinkedIn 

Past

Lead Member of Technical Staff at Kosmix 

Member of Technical Staff at Kosmix 

Summer Intern at Google Inc. 

[see all](#) ▾

Education

The University of Texas at Austin

The University of Texas at Austin

Indian Institute of Technology, Bombay

Recommendations

2 people have recommended Mitul

Connections

500+ connections

Websites

[Personal Website](#)

[Blog](#)

Twitter

[mitultiwari](#)

Public Profile

<http://www.linkedin.com/in/mitultiwari>

Outline

- About Me
- About LinkedIn
- Analytics products at LinkedIn
- Fun Facts: Sequencing the Startup DNA

LinkedIn by the numbers

- 120,000,000+ users (August 4, 2011)
- 2+ new user registrations per second
- 81+ Million monthly unique users* (Comscore)
- 2 Billion People Searches in 2010
- 7.1 Billion page views in Q2 2010
- 2+ Million companies with LinkedIn Company Pages
- 2+ M LinkedIn Groups

Broad Range of Products

Neal Sample 1st

CTO - Open Commerce at eBay
San Francisco Bay Area | Internet

Current CTO - X.commerce at eBay

Past VP of Architecture, Technology Product Management, Developers Program at eBay
VP Social, Open and Participation Platforms at Yahoo!
Chief Architect (VPE/Fellow) at Yahoo!

Education Stanford University
University of Wyoming
Stanford University

Recommendations 33 people have recommended Neal

Connections 500+ connections

Public Profile <http://www.linkedin.com/in/nealsample>

Share PDF Print vCard Flag

Jobs Home | Saved Jobs | Saved Searches | Advanced Search | Refer Connections

Sign up for the free webinar "Job Seeking on LinkedIn" for tips and tricks

Search for Jobs

Job Title, Keywords, or Company Name Advanced

Jobs you may be interested in Beta [Email Alerts](#) | [See More](#)

- Software Architect**
Intuit - Menlo Park, CA
- Software Architect**
Nomis Solutions - San Bruno, CA
- Software Architect w/ Reverse Proxy Implementation experience**
AT&T Interactive - San Francisco
- Network Software Architect**
AGCM Research - San Francisco Bay Area
- Software Architect - MVC - Cloud Computing - Mobile Commerce**
CyberCoders - San Francisco ,CA

Inbox 92 | Companies | News | More | People | Search...

Invitations (92) [Add Connections](#)

- Wendy Pfeiffer
Sr. Director, Enterprise Shared Services
- Alex Schwartz
Site Reliability Engineer at LinkedIn
- Tony Ng
Director of Architecture at eBay
Tony has indicated you are a Colleague at Yahoo!

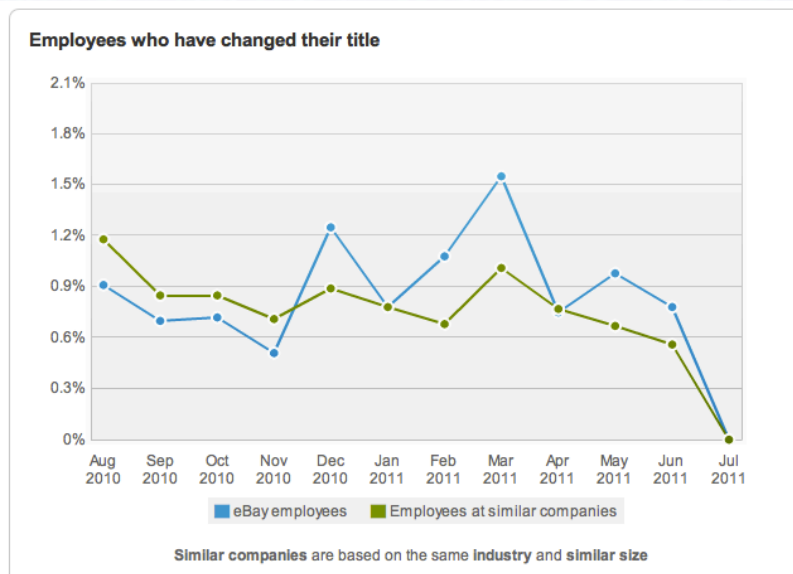
Messages (0) [Compose Message](#)

- Christopher Brereton (Responded)
Hey Tom! Jun 14
- Kevin Krawez
Your call Apr 19
- Aaron Tran
VMware seeking references for Staff/Principal SW-Systems Management... Mar 30

Kara Swisher
archives | ethics statement | bio | e-mail | RSS | Follow @karaswisher

Viral Video: Like Palm's Creepy Naked Lady, TouchPad's Floating Celeb Heads Get the HP Boot

AUGUST 19, 2011 AT 12:13 AM PT [Tweet](#) [Like](#) [Share](#) [Share](#) [Print](#)



INTERNET TODAY · AUG 22

The most shared news by people in the Internet Industry. Take the tour »

FRONT PAGE | INTERNET | MARKETING &... | COMPUTER... | ONLINE MEDIA | IT | TELECOMMUNIC...

Skype To Acquire Year-old Group Messaging Service GroupMe

techcrunch.com [Share](#) [Save](#)

Skype will acquire group messaging service GroupMe, a service that was born at a hackathon at TechCrunch Disrupt in New York in 2010. GroupMe was founded by Jared Hecht and Steve Martocci. The terms of the deal, including...

Trending in Information Technology and Services, Internet, Computer Software and Venture Capital & Private Equity

2nd 2nd 2nd

Khan Academy Integrates With Digital Textbooks

mashable.com

The 12-minute lectures that Bill Gates has called "the start of a revolution" will now be linked with the material in some digital textbooks.

Trending in Internet, Computer Software, Marketing and Advertising and Online Media

390

DISCOVER MORE IN INTERNET

- 13 Sites to Download Free eBooks
59K+ views · itmanagersinbox.com
- How to track a laptop Video - 5min.com
19K+ views · 5min.com
- Random Key Generator
148K+ views · randomkeygen.com
- MAKE | \$25 USB Stick PC
96K+ views · makezine.com
- 15 Keyboard Shortcuts To Enhance Your PC Productivity : Technology :...
18K+ views · openforum.com

Tips For Spying On Your PPC Competitors


A primer on PPC Competitive...


User Profile

LinkedIn Account Type: Pro Tom Quiggle [Add Connections](#)

Home Profile **Contacts** Groups Jobs **Inbox 92** Companies News More People Search... [Advanced](#)


[Earn Your M.S. in Finance - 15 month program designed for busy working professionals. Click for info!](#) From: Saint Mary's College of California



Jeff Weiner 1st 



CEO at LinkedIn
Mountain View, California | Internet




Jeff Weiner @dtunkelang adds great examples to recent talk re: the importance of understanding mutual fit when recruiting.



Dream. Fit. Passion. thenoisychannel.com · via Daniel Tunkelang
A few days ago, our CEO Jeff Weiner led a session at LinkedIn on how to "close" candidates — that is, how to persuade candidates to join your team once you have found and interviewed them. Since not everyone has the opportunity...

5 hours ago · Like (1) · Comment · Send a message · Share · See all activity

Current CEO at LinkedIn 
Member, Board of Directors at DonorsChoose 
Member, Board of Directors at Malaria No More

Past Executive in Residence at Accel Partners 
Executive in Residence at Greylock 
Executive Vice President Network Division at Yahoo! 
[see all](#) >

Education University of Pennsylvania - The Wharton School

Recommendations 8 people have recommended Jeff

Connections 500+ connections


Websites Company Website


Twitter [Follow](#) @jeffweiner

Public Profile <http://www.linkedin.com/in/jeffweiner08>

[Share](#) [PDF](#) [Print](#) [vCard](#) [Flag](#)

Suggest Connections

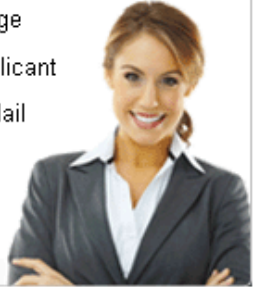
[Send Jeff a message](#)
[Suggest a profile update for Jeff](#)
[Save Jeff's Profile](#) 

LinkedIn Premium 

Get Hired Faster
with Job Seeker Premium

- Get noticed with a Job Seeker Badge
- Move to the top as a Featured Applicant
- Contact recruiters directly with InMail


[Learn More](#)



Jeff's Activity

Jeff Weiner likes this update:


Mike Gamson Another step in the right direction for the Khan Academy. . .



Khan Academy Integrates With Digital Textbooks mashable.com
The 12-minute lectures that Bill Gates has called "the start of a revolution" will now be linked with the material in some digital textbooks.

4 hours ago · Like (4) · Comment · Share

Jeff Weiner @dtunkelang adds great examples to recent talk re: the importance of understanding mutual fit when recruiting.



Dream. Fit. Passion. thenoisychannel.com · via Daniel Tunkelang
A few days ago, our CEO Jeff Weiner led a

Summary

Internet executive with over 16 years of experience, including general management of mid to large size organizations, corporate development, product development, business operations, and strategy.

Connections

LinkedIn Account Type: Pro Tom Quiggle Add Connections

Home Profile Contacts Groups Jobs Inbox 91 Companies News More People Search... Advanced

My Connections Imported Contacts Profile Organizer Network Statistics Add Connections Remove Connections

Share your phone, IM and more with your connections. Update it now »

Filter Connections Select: All, None abc

Filter	Name	Company	Connections
<input type="checkbox"/>	Henke, David	Senior Vice President Engineering/Operations - LinkedIn	500+
<input type="checkbox"/>	Herendeen, Julie	VP, Marketing - Lookout	436
<input type="checkbox"/>	Hoang, Phu	Executive Chairman - Buysight, Inc.	500+
<input type="checkbox"/>	Hoffman, Havi	Content wrangler at Mozilla Labs - Mozilla Corporation	500+
<input type="checkbox"/>	Hom, Allen	Campus Recruiter - LinkedIn	500+
<input type="checkbox"/>	Hsiung, Ernie	Co-Founder - MoodThingy	496
<input type="checkbox"/>	Hull, Mark	Co-Founder - Riley Games, Inc.	500+
<input type="checkbox"/>	Johnson, Mark	Chief Data Officer - Groupon	500+

Quickly view and organize your connections?
Select a category or individual to see contact info, send a message and more.

1 outstanding sent invitation | [Export connections](#)

Ads by LinkedIn Members

[Engineers: Get PDH Free](#)

Free PDH: Ethical Decision Making for Engineers. Learn more.
From: Victoria Zambito

[Management Leadership](#)

Learn Ivy League Management At eCornell. Best in Class Training.
From: LTG LinkedIn

[What's this?](#)

Communications

Communications

The screenshot displays the LinkedIn inbox interface for a user named Tom Quiggle. The top navigation bar includes links for Home, Profile, Contacts, Groups, Jobs, Inbox (with a notification badge of 91), Companies, News, and More. A search bar is located on the right side of the navigation bar. Below the navigation bar, there are tabs for Messages and Invitations (with a notification badge of 91). The Messages tab is active, showing a list of messages. On the left side, there is a sidebar with a search box for the inbox, a 'Compose Message' button, and a list of folders: Inbox, Sent, Archived, and Trash. The message list includes the following entries:

Select	From	Subject	Date	Actions
<input type="checkbox"/>	Christopher Brereton (Responded)	Hey Tom!	Jun 14	Reply · Forward · Archive · Delete
<input type="checkbox"/>	Kevin Krawez	Your call	Apr 19	Reply · Forward · Archive · Delete
<input type="checkbox"/>	Aaron Tran	VMware seeking references for Staff/Principal SW-Systems Management developers	Mar 30	Reply · Forward · Archive · Delete
<input type="checkbox"/>	Lynne Peters (Declined)	NoSQL Architect Opportunity	Mar 29	Reply · Forward · Archive · Delete
<input type="checkbox"/>	Courtney Gartin - We're Hiring!!! (Declined)	Your experience is AMAZING! Great Opportunities & Your Expertise Requested - 7907	Mar 24	Reply · Forward · Archive · Delete
<input type="checkbox"/>	Catherine Porter	Catherine Porter invites you to join LI for Education (LIFE) on LinkedIn	Mar 18	Reply · Forward · Archive · Delete
<input type="checkbox"/>	Mark Gambirasi (Declined)	Unique, Client-Facing Systems Engineer Position with Rising Star Cloud Computing Company	Mar 16	Reply · Forward · Archive · Delete

On the right side of the inbox, there are advertisements. The first is for 'Autodesk Courses' with the text 'Learn Autodesk through our classroom and virtual training options.' and 'From: EMA Design Automation'. The second is for 'Cash for your Business' with the text 'We're open for your Yahoo business. Get up to \$40K to grow your business.' and 'From: Kabbage, Inc'. A 'What's this?' link is located below the second advertisement.

Hiring Solutions

This screenshot shows the LinkedIn Recruiter interface. At the top, there's a navigation bar with 'Post a Job', 'Create a Project', and 'Advanced Search'. Below this is a sidebar with navigation options: Home, Clipboard (14), Searches, My Projects, Team Projects, Communication, and Jobs. The main content area is divided into three sections: 'Inbox (3 new)' with a feedback score of five stars, 'Alerts' showing saved searches for 'Product Manager', 'Senior Product Manager', and 'Director, Engineering', and 'Team Activity' showing recent profile additions and job postings.

This screenshot displays the search results for 'product manager' in the LinkedIn Recruiter interface. The search results are sorted by 'Keyword relevance' and show 77,087 results. The top result is for 'Jeff Robertson', Sr. Product Manager at WebEx, located in the San Francisco Bay Area. Below this, the profile of 'Deborah Jones', Sr. Product Manager at Sunquest Information Systems, is shown. The right sidebar contains a search filter section with options to refine by Location, Relationship, and Industry.

Location	Count
All Locations	
United States	77,087
San Francisco Bay Area	18,115
Greater Boston Area	7,435
Greater Seattle Area	6,933
United Kingdom	6,487

Relationship	Count
All LinkedIn Members	
1st Connections	10
2nd Connections	3,612
Group Members	876
3rd + Everyone Else	72,836

Industry	Count
All Industries	
Information Technology and Services	82,535
Computer Software	77,087
Marketing and Advertising	57,210
Telecommunications	35,106
Financial Services	34,563

Job Search

The screenshot displays the LinkedIn job search interface. At the top, the user's name 'Tom Quiggle' and 'Add Connections' are visible. The navigation bar includes 'Home', 'Profile', 'Contacts', 'Groups', 'Jobs', 'Inbox 92', 'Companies', 'News', and 'More'. A search bar is present with a 'Jobs' dropdown and a search button labeled 'Advanced'. Below the navigation, the 'Jobs Home' section is active, showing 'Saved Jobs' and 'Saved'. A 'Job Search' sidebar on the left contains filters for Keywords (distributed database), Job Title (architect), Company, Location (Located in or near), Country (United States), Postal Code, and Within (10 mi (15km)). A 'Refine By' section lists company filters like Microsoft, CyberCoders, and IBM. A dropdown menu is open over the 'Jobs' link, listing options: Find Jobs, Job Seeker Premium, Career Explorer beta, Post a Job, Manage Jobs, and Hiring Solutions. The main content area shows a list of job listings, including 'Java Architect with data warehousing for a Cloud technology co.' by CyberCoders, 'Architect- Hadoop Engineering' by Greenplum, 'TEGRA INFRASTRUCTURE ARCHITECT' by NVIDIA, 'Senior Software and Systems Architect' by Sportvision, 'Software Solution Architect' by FutureDial, 'Platform Architect - LAMP - Subversion - Build - Release - Ops' by CyberCoders, and 'Database Architect (Oracle)' by Comcast. On the right, there are promotional banners for LinkedIn Premium, tips to stand out from the crowd, and features like Profile Organizer and InMail.


Company Pages

LinkedIn Account Type: Pro ▼ Tom Quiggle [Add Connections](#)

Home Profile Contacts Groups Jobs **Inbox 92** Companies News More

Companies > Dell

Overview Careers **Products & Services** Analytics



Listen. Learn. Deliver. That's what we're about.

For over 26 years, Dell has empowered communities and people everywhere to use technology to realize their dreams. Customers trust us to deliver technology solutions that help them do and achieve more at home, work, school or anywhere in their world.

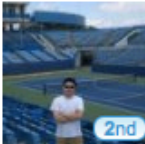

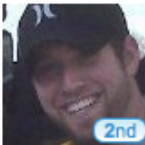


Dell has 144,246 followers

[Follow Company](#) [Share](#)

38 Products & Services

221 Recommendations

33 people in your network recommend Dell


-  **Chen Li** Software QA Engineer at LinkedIn
Recommends Dell Efficient Workforce »
-  **Will Koch** Category Head, Tech & Telco, EMEA at LinkedIn
Recommends Dell EqualLogic Solutions »
-  **Steven Brooks** Sr. UX Designer at LinkedIn
Recommends Dell Latitude E6520 Laptop »
-  **Bhairavi Vyas** Lead Engineer - QA at LinkedIn
Recommends Dell Streak »
-  **Madhu Gupta** Head of API Platform Product and Strategy

Dell Product and Service Spotlight


PowerEdge First Server Solutions

Designed for your small business needs.

[Shop Now >](#)



Featured Products and Services





Dell Efficient Workforce

The Efficient Workforce focuses on empowerment. Your workforce is changing, and you need to change with it to stay competitive. Your employees want to be able to work from anywhere, at anytime and use the most efficient technologies available. They bring powerful new opportunities into the...

[Recommend](#) · [Share](#)

10 recommendations






Dell Workstation Solutions

The Power and Performance your Work Demands Dell Precision™ workstations are our highest-performing and most scalable systems, specifically designed for graphics-intensive professionals in engineering, product design, animation and...

21 recommendations



11




Outline

- About Me
- About LinkedIn
- Analytics products at LinkedIn
- Fun Facts: Sequencing the Startup DNA

What do I mean by *Analytics Products*?

People You May Know

People you may know

	Jay Kreps, Principle Engineering Manger at LinkedIn + Connect	×
	Jeremy Gillick, Senior Web Developer at LinkedIn + Connect	×
	Albert Wang, User Experience Design at LinkedIn + Connect	×

[See more »](#)

Profile Stats: WVMP

Profile Stats Pro

Last 90 Days May 10, 2011 – August 8, 2011 [Settings](#)

Who's Viewed Your Profile

LAST THREE MONTHS



Graduate Student at Stanford University



Someone in the Executive Leadership function in the Staffing and Recruiting industry from San Francisco Bay Area



Dj Das  
Business Operations / Product Management / Big Data Evangelist
San Francisco Bay Area | Information Technology and Services
In Common: [▶ 8 shared connections](#) [▶ 1 shared group](#)



Software Developer at SAS



Ravikant D V S
Research Assistant at Cornell University
Austin, Texas Area | Computer Software
In Common: [▶ 5 shared connections](#)



Doug Judd
CEO at Hypertable, Inc.
San Francisco Bay Area | Internet
In Common: [▶ 34 shared connections](#) [▶ 2 shared groups](#)



Bruce Ferguson  
High Frequency Trader / Quant Strategist
Toronto, Canada Area | Investment Management
In Common: [▶ 1 shared connection](#)

Trends

Views [Appearances in Search](#)

Total Views **489**



Top Search Keywords

1. hadoop	6%
2. kosmix	6%
3. mitul tiwari	3%
4. data mining	3%
5. linkedin	2%
6. mapreduce	2%
7. computer science	2%
8. hadoop mapreduce	2%
9. hadoop microsoft	1%
10. tweets	1%
11. ruby	1%
12. machine learning	1%

Viewers of this profile also ...

Viewers of this profile also viewed...



Subramanyam (Manyam) Mallela
Director of Engineering, Online...



Digvijay Lamba
Chief Architect Categorization at Kosmix



Vijay Chittoor
Founder at Mertado



Anand Rajaraman
Founder, Kosmix; Founding Partner,...



Indrani Chakravarty
Member of Technical Staff at...



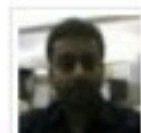
Ankur Jain
Director, Product Management at...



Sri Subramaniam
Engineering & Product Executive



Venky Harinarayan
SVP, Global eCommerce at Walmart



Gaurav Bhalotia
Director, at Retrevo

Skills

Search Skills & Expertise

Related Skills

- HBase
- MapReduce
- Nutch
- Solr
- Lucene
- AWS
- Collaborative Filtering
- EC2
- Guice
- Weka
- S3
- Recommender Systems
- MarkLogic
- Amazon Web Services
- Memcached
- RDFS
- HAProxy
- Web Crawling
- Berkeley DB
- CouchDB

Hadoop ▲ 46% y/y

Primary Industry: Internet

Apache Hadoop is a Java software framework that supports data-intensive distributed applications under a free license. It enables applications to work with thousands of nodes and petabytes of data. Hadoop was inspired by Google's MapReduce and Google File System (GFS) papers. Hadoop is a top-level Apache project, being built and used by a community of ... [More on 'Hadoop' at Wikipedia >](#)

✓ Listed on your profile
Edit Your Skills

[Share](#) [Tweet](#)

89th Fastest growing skill

	Relative Growth	Size
HBase		
Solr		
Hadoop		
MapReduce		

Related Companies

- The Apache Software Foundation**
Computer Software, United States
[Follow](#)
- Cloudera**
Computer Software, San Francisco Bay Area
[Follow](#)
- Yahoo!**
Internet, San Francisco Bay Area
[Follow](#)
- LinkedIn**
Internet, San Francisco Bay Area
[Stop following](#)
- eBay**
Internet, San Francisco Bay Area
[Follow](#)

[More companies...](#)

Hadoop Jobs

- Senior Hadoop Systems Developer**
Cloudera, Inc. - Palo Alto or San Francisco CA
- Software Engineer – Hadoop Engineering**
Greenplum - San Francisco Bay Area
- Sr. Software Engineer - Hadoop**
Apple Inc. - Cupertino, CA
- Software Architect (Java, Hadoop)**
NetApp - Sunnyvale CA

Hadoop Professionals

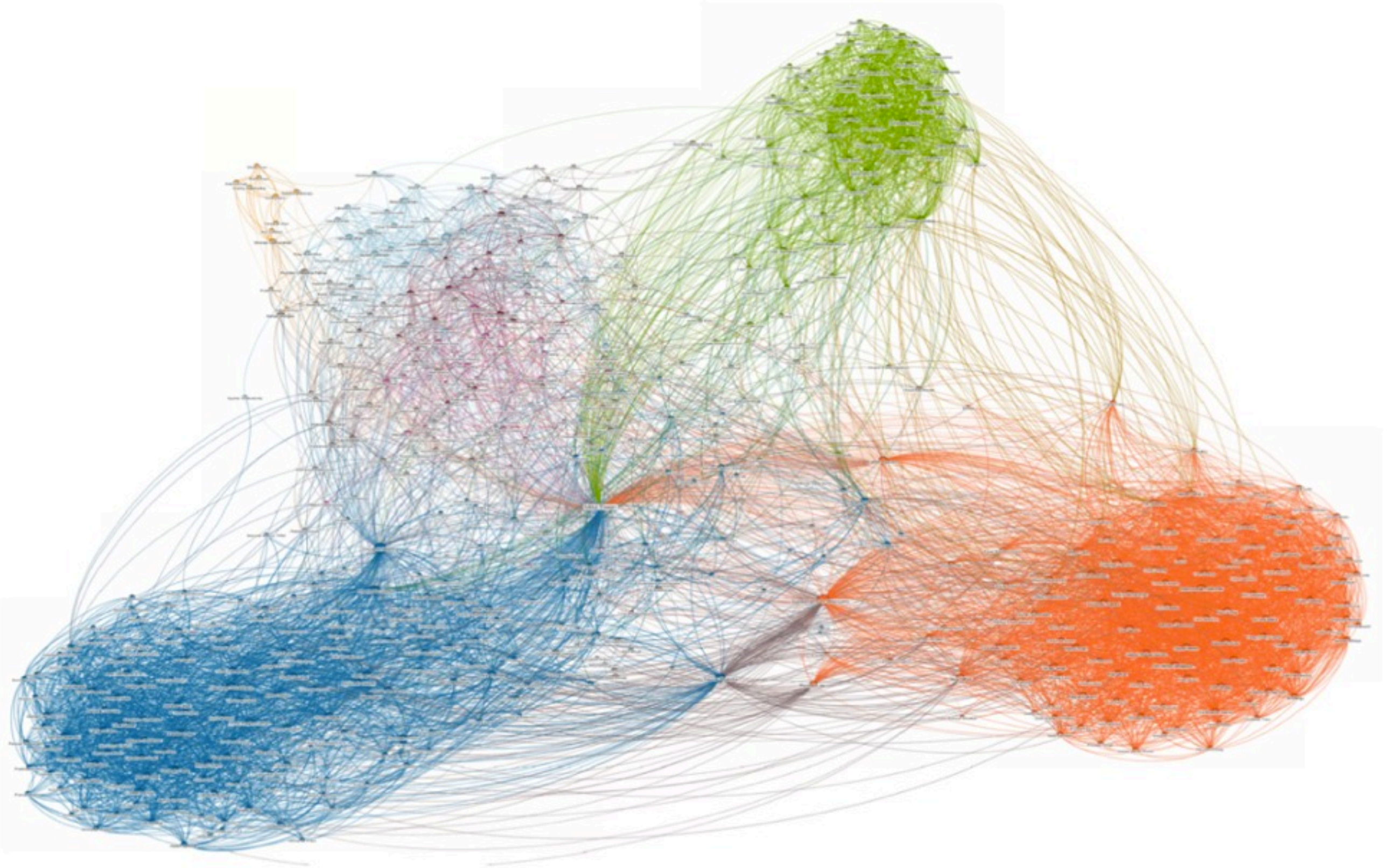
- Tom White** (2nd)
Engineer at Cloudera
Tom White has been an Apache Hadoop committer since February 2007, and is a member of the Apache Software...
- Doug Cutting** (2nd)
Me
I am a founder of the Lucene, Nutch, Hadoop and Avro open source projects.
- Patrick Hunt** (2nd)
Member at The Apache Software Foundation
- Arun C Murthy** (2nd)
Founder and Architect at Hortonworks Inc.
I am a Founder and Architect at Hortonworks Inc. Hortonworks was formed by the key architects and core Hadoop...
- Devaraj Das** (2nd)
Founder at HortonWorks
Distributed/Grid Computing, Security technologies (at all layers of the stack - application, tcp, IP, physical...
- Jakob Homan** (1st)
Senior Software Engineer at LinkedIn
Apache Hadoop Committer and PMC member. Trying to be everywhere at once on LinkedIn's SNA team - Hadoop, Hive,...
- Koji Noguchi** (2nd)
Hadoop Debugger
- Chris Douglas** (2nd)
Technical Yahoo!

[More Hadoop Professionals...](#)

Hadoop Groups

- Hadoop Users** You are a member
A group for Hadoop users.

InMaps



Related Searches

- Millions of Searches everyday
- Help users to explore and refine their queries

Related searches for **Web Developer**

web designer

php

web development

ruby on rails

software engineer

programmer

developer

html

Analytics Products: Key Ideas

- Recommendations
 - People You May Know, Related Searches, Viewers of this profile ...
- Insight
 - Profile Stats: Who Viewed My Profile, Skills
- Visualization
 - InMaps

Analytics Products: Challenges

- LinkedIn: largest professional network
- 120+ million members on LinkedIn
- Billions of pageviews
- Terabytes of data to process

Outline

- Analytics products at LinkedIn
- Deep Dive - Connection Strength
- Deep Dive - Related Searches
- Fun Facts: Sequencing the Startup DNA

Connection Strength

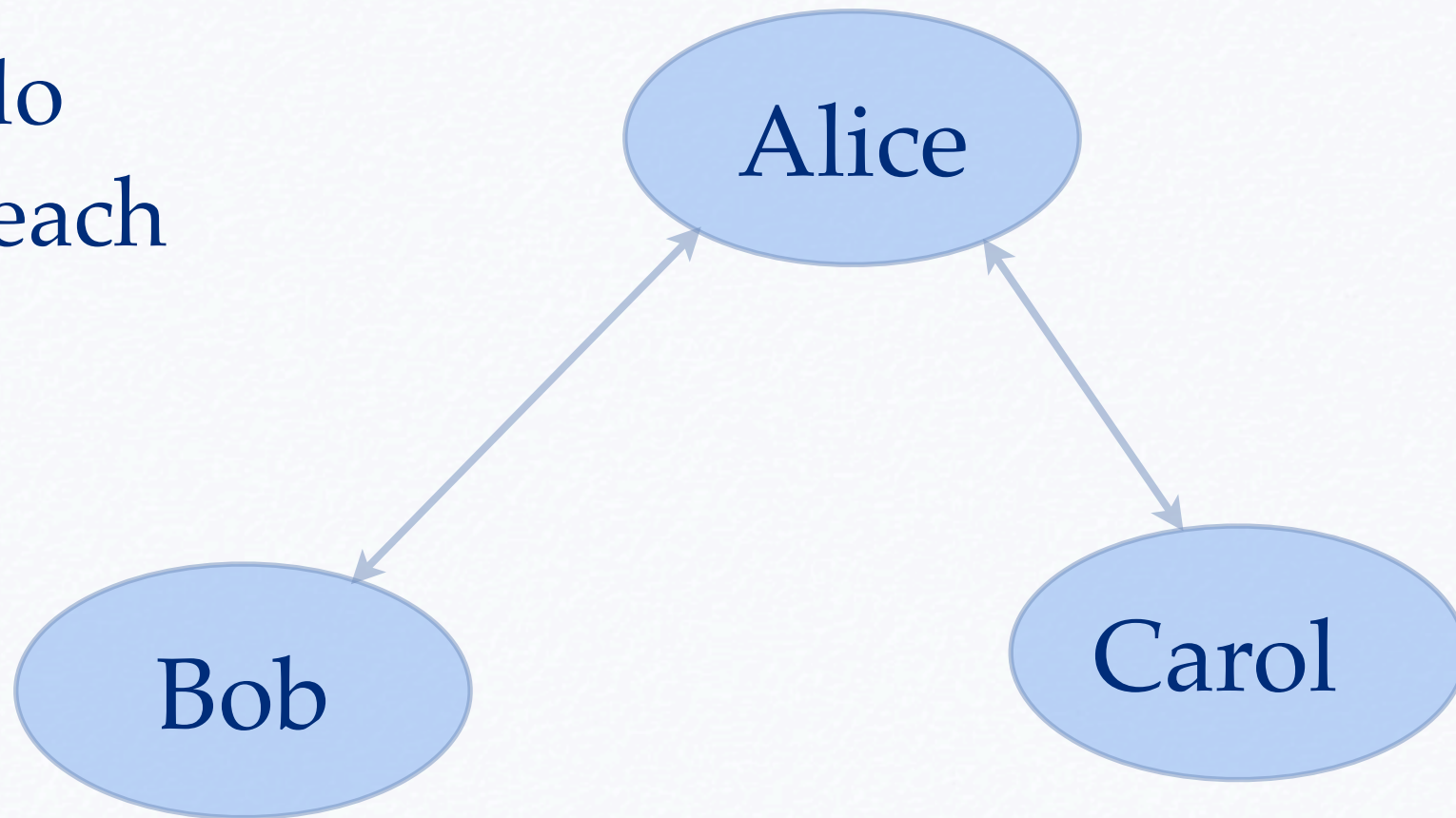
- Let's build "Connection Strength"
- Systems and Tools we use
- Hadoop MapReduce
- Managing workflow
- Serving data in production

Connection Strength

- How well do people know each other?
- Connection Strength
- Application: reorder updates to show updates from close connections

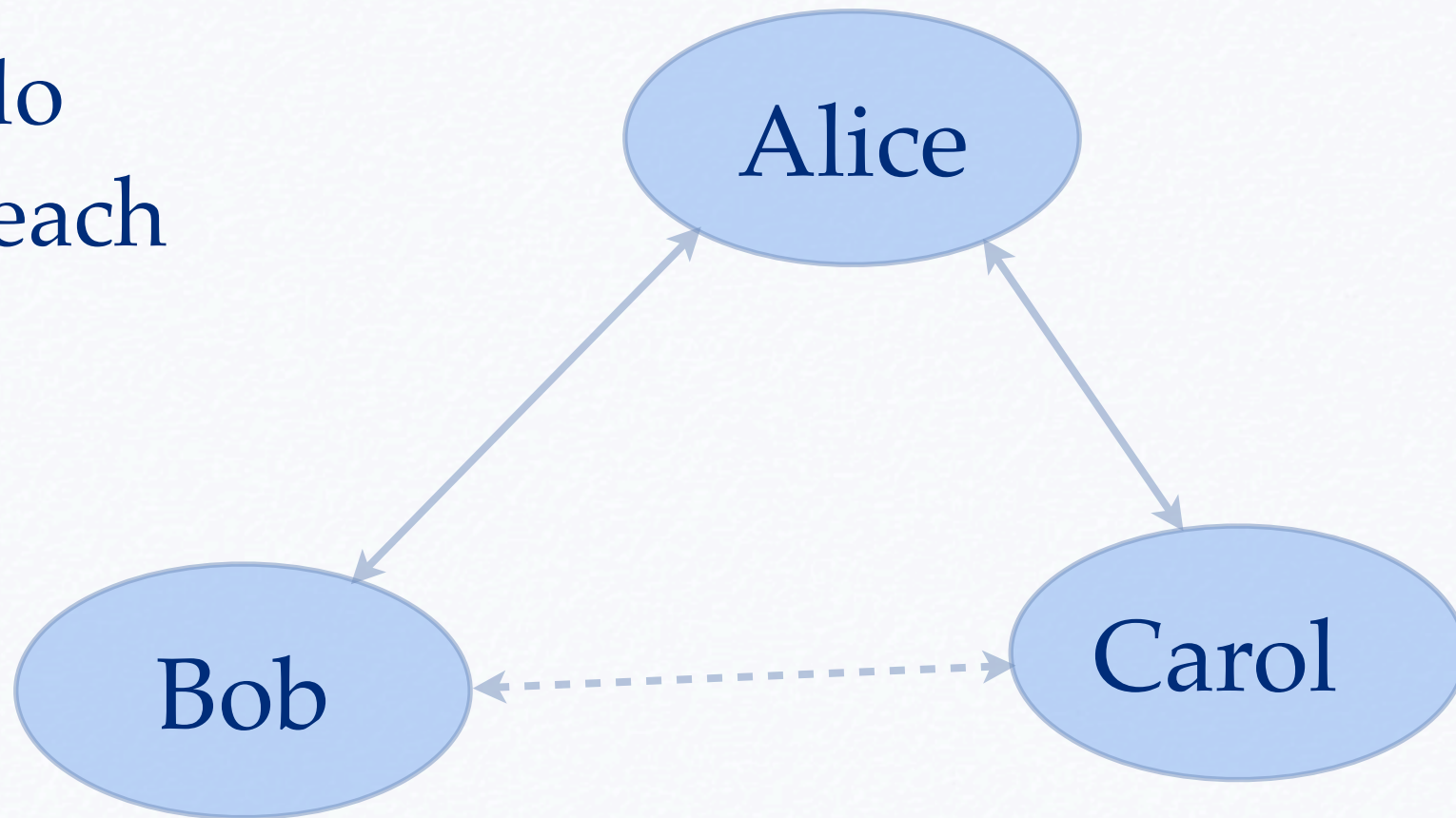
Connection Strength

How well do
people know each
other?



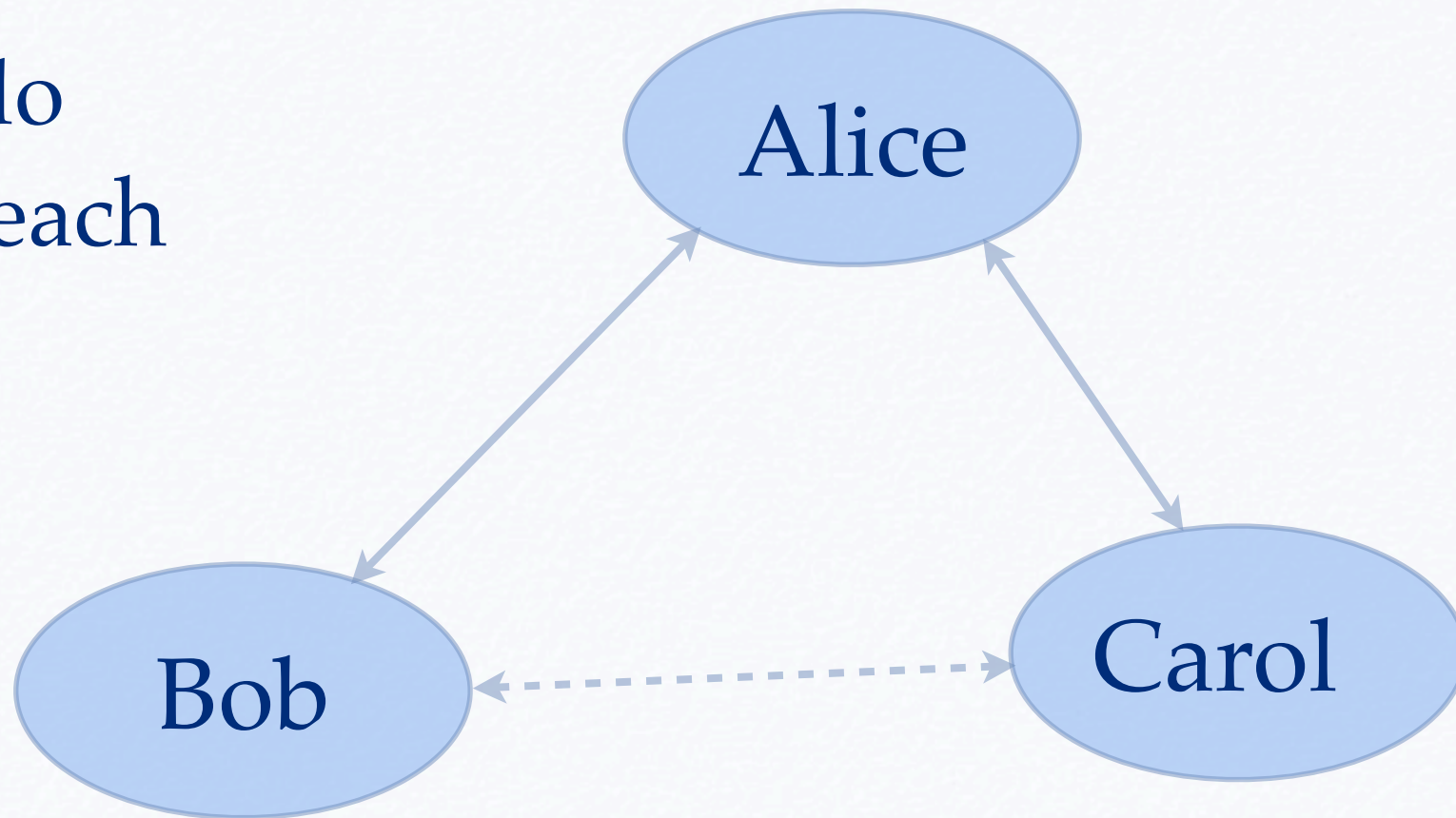
Connection Strength

How well do
people know each
other?



Connection Strength

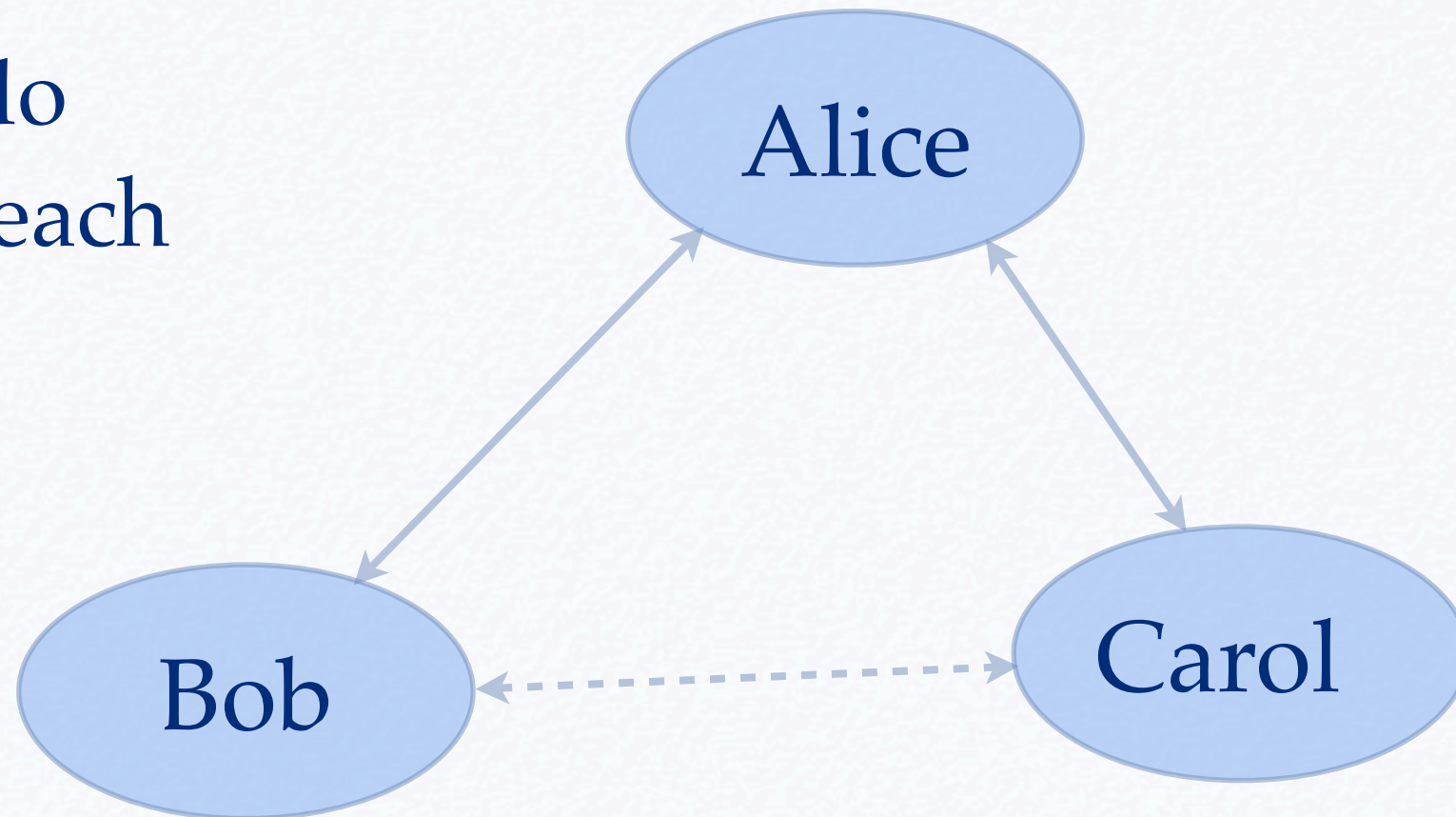
How well do
people know each
other?



Triangle closing

Connection Strength

How well do
people know each
other?



Triangle closing

$\text{Prob}(\text{Bob knows Carol}) \sim \text{the \# of common connections}$

Triangle Closing in Pig

```
-- connections in (source_id, dest_id) format in both directions
connections = LOAD `connections` USING PigStorage();
group_conn = GROUP connections BY source_id;
pairs = FOREACH group_conn GENERATE
    generatePair(connections.dest_id) as (id1, id2);

common_conn = GROUP pairs BY (id1, id2);
common_conn = FOREACH common_conn GENERATE
    flatten(group) as (source_id, dest_id),
    COUNT(pairs) as common_connections;
STORE common_conn INTO `common_conn` USING PigStorage();
```

Pig Overview

- Load: load data, specify format
- Store: store data, specify format
- Foreach, Generate: Projections, similar to select
- Group by: group by column(s)
- Join, Filter, Limit, Order, ...
- User Defined Functions (UDFs)

Triangle Closing in Pig

```
-- connections in (source_id, dest_id) format in both directions
connections = LOAD `connections` USING PigStorage();
group_conn = GROUP connections BY source_id;
pairs = FOREACH group_conn GENERATE
    generatePair(connections.dest_id) as (id1, id2);

common_conn = GROUP pairs BY (id1, id2);
common_conn = FOREACH common_conn GENERATE
    flatten(group) as (source_id, dest_id),
    COUNT(pairs) as common_connections;
STORE common_conn INTO `common_conn` USING PigStorage();
```

Triangle Closing in Pig

-- connections in (source_id, dest_id) format in both directions

```
connections = LOAD `connections` USING PigStorage();
```

```
group_conn = GROUP connections BY source_id;
```

```
pairs = FOREACH group_conn GENERATE
```

```
    generatePair(connections.dest_id) as (id1, id2);
```

```
common_conn = GROUP pairs BY (id1, id2);
```

```
common_conn = FOREACH common_conn GENERATE
```

```
    flatten(group) as (source_id, dest_id),
```

```
    COUNT(pairs) as common_connections;
```

```
STORE common_conn INTO `common_conn` USING PigStorage();
```


Triangle Closing in Pig

-- connections in (source_id, dest_id) format in both directions

```
connections = LOAD `connections` USING PigStorage();
```

```
group_conn = GROUP connections BY source_id;
```

```
pairs = FOREACH group_conn GENERATE
```

```
    generatePair(connections.dest_id) as (id1, id2);
```

```
common_conn = GROUP pairs BY (id1, id2);
```

```
common_conn = FOREACH common_conn GENERATE
```

```
    flatten(group) as (source_id, dest_id),
```

```
    COUNT(pairs) as common_connections;
```

```
STORE common_conn INTO `common_conn` USING PigStorage();
```

Triangle Closing in Pig

```
-- connections in (source_id, dest_id) format in both directions
connections = LOAD `connections` USING PigStorage();
group_conn = GROUP connections BY source_id;
pairs = FOREACH group_conn GENERATE
    generatePair(connections.dest_id) as (id1, id2);

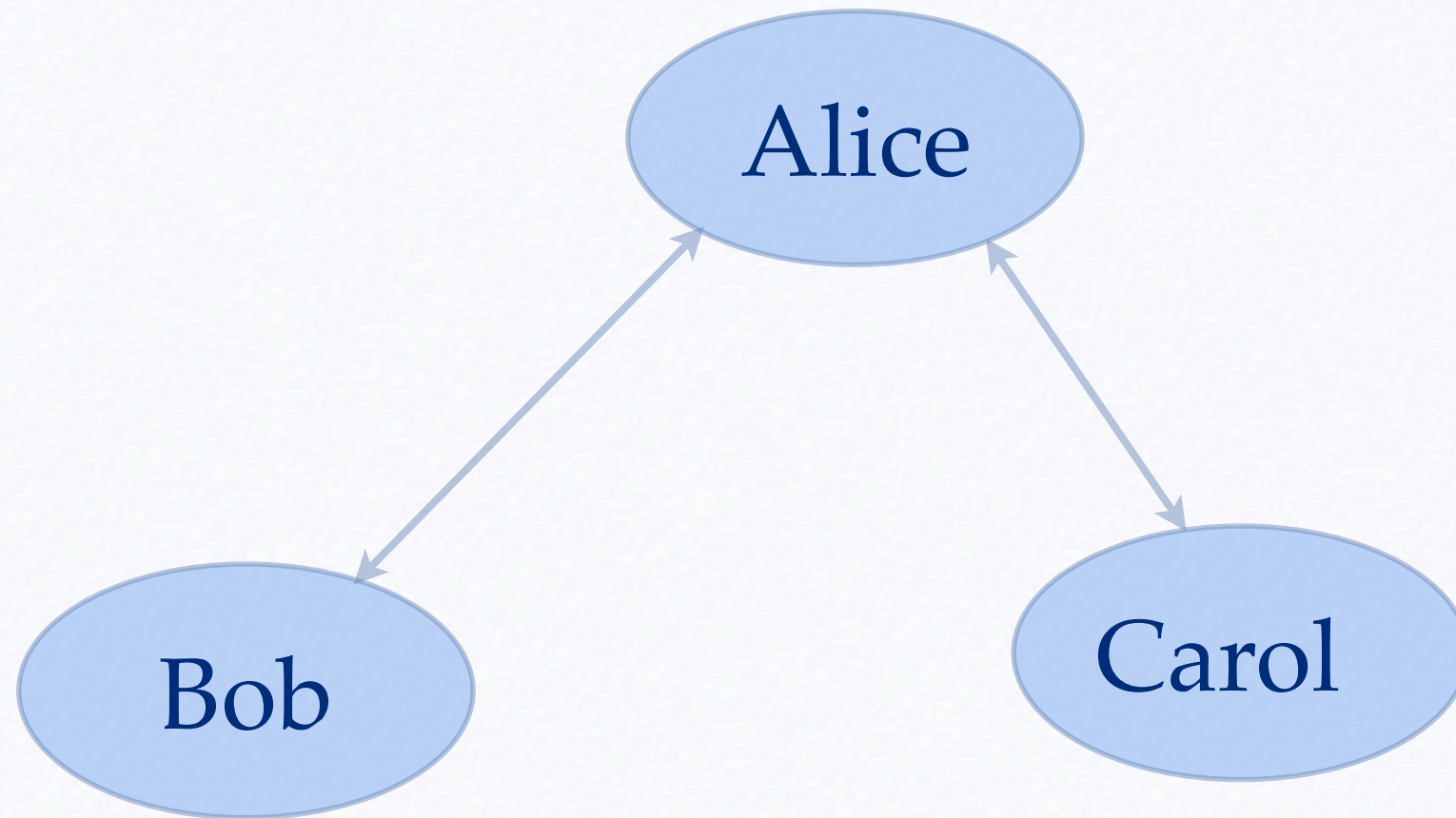
common_conn = GROUP pairs BY (id1, id2);
common_conn = FOREACH common_conn GENERATE
    flatten(group) as (source_id, dest_id),
    COUNT(pairs) as common_connections;
STORE common_conn INTO `common_conn` USING PigStorage();
```


Triangle Closing in Pig

```
-- connections in (source_id, dest_id) format in both directions
connections = LOAD `connections` USING PigStorage();
group_conn = GROUP connections BY source_id;
pairs = FOREACH group_conn GENERATE
    generatePair(connections.dest_id) as (id1, id2);

common_conn = GROUP pairs BY (id1, id2);
common_conn = FOREACH common_conn GENERATE
    flatten(group) as (source_id, dest_id),
    COUNT(pairs) as common_connections;
STORE common_conn INTO `common_conn` USING PigStorage();
```

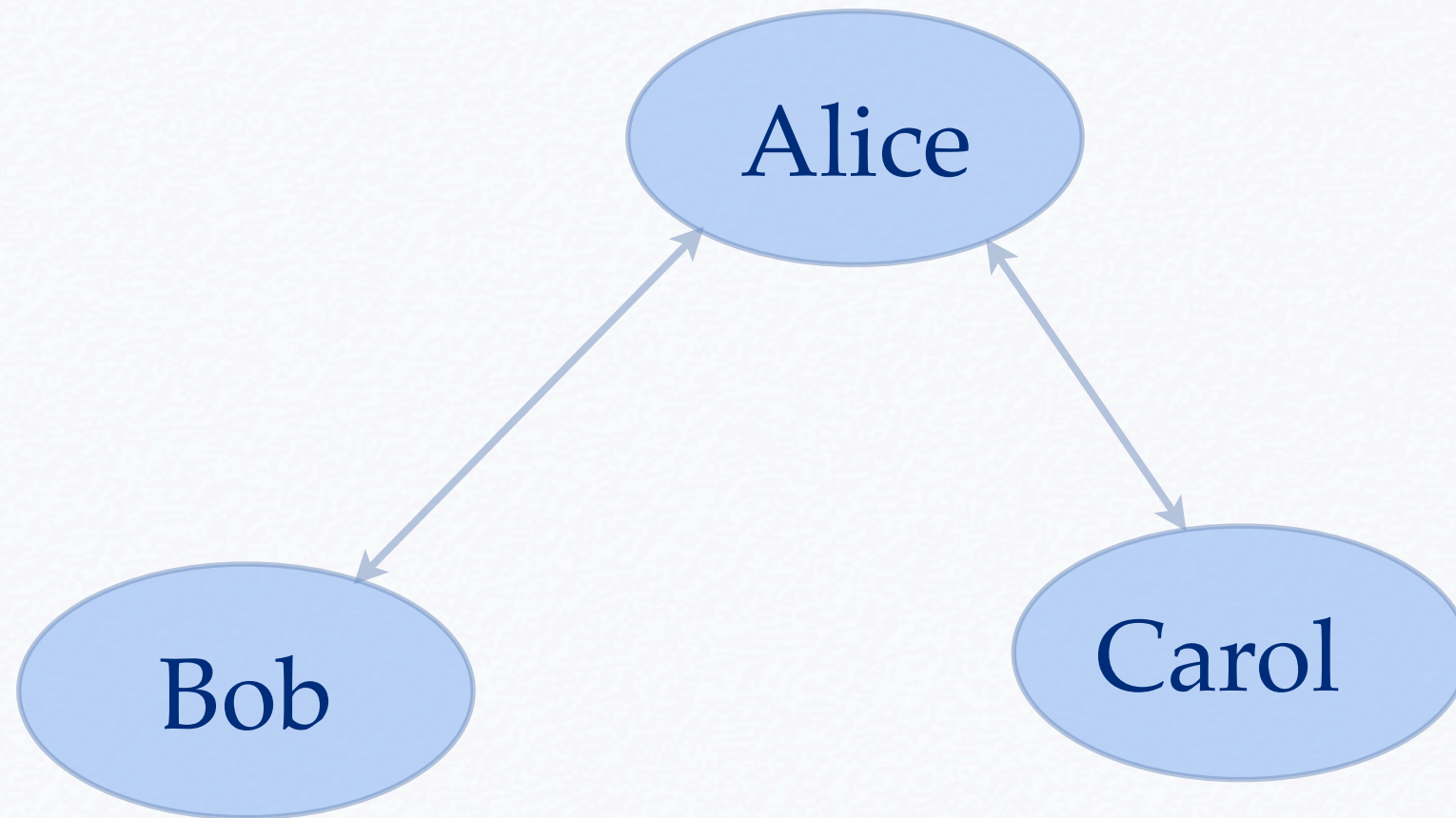
Triangle Closing Example



1. $(A,B), (B,A), (A,C), (C,A)$
2. $(A, \{B,C\}), (B, \{A\}), (C, \{A\})$
3. $(A, \{B,C\}), (A, \{C,B\})$
4. $(B,C,1), (C,B,1)$

```
connections = LOAD `connections` USING  
PigStorage();
```

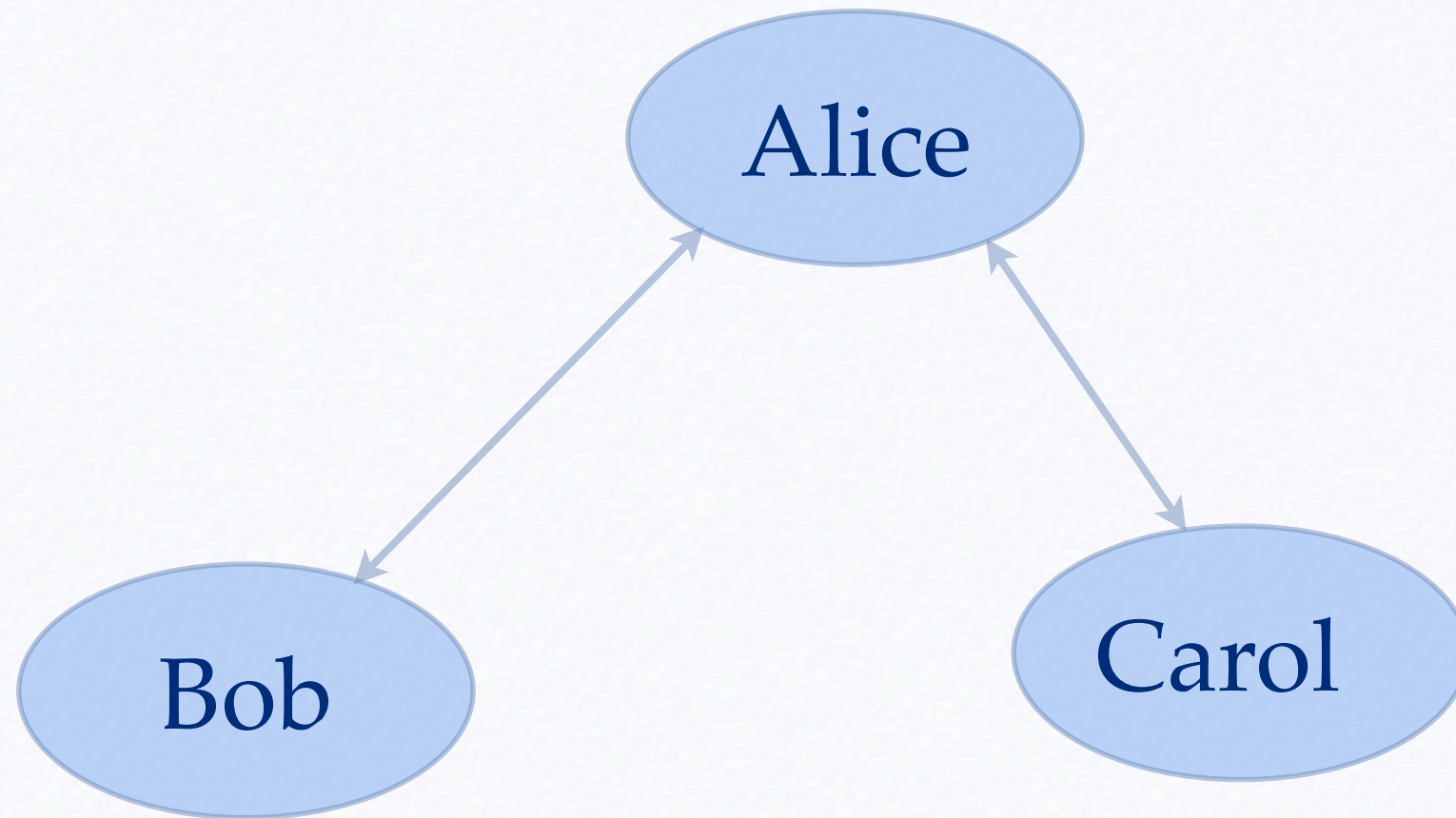

Triangle Closing Example



1. $(A,B), (B,A), (A,C), (C,A)$
2. $(A, \{B,C\}), (B, \{A\}), (C, \{A\})$
3. $(A, \{B,C\}), (A, \{C,B\})$
4. $(B,C,1), (C,B,1)$

group_conn = GROUP connections BY source_id;

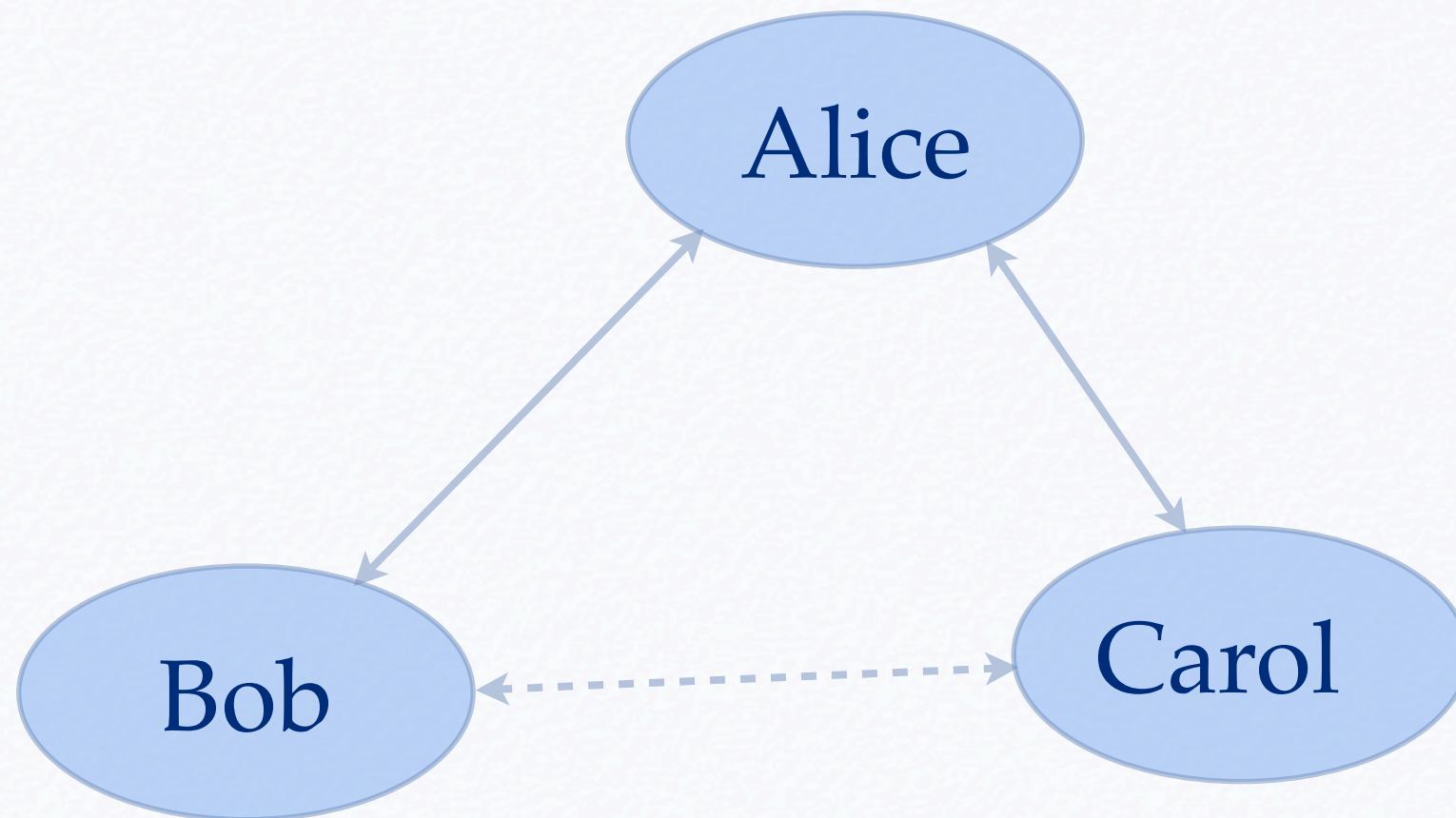
Triangle Closing Example



1. (A,B), (B,A), (A,C), (C,A)
2. (A, {B,C}), (B, {A}), (C, {A})
3. (A, {B,C}), (A, {C,B})
4. (B,C,1), (C,B,1)

pairs = FOREACH group_conn GENERATE
generatePair(connections.dest_id) as (id1, id2);

Triangle Closing Example



1. (A,B),(B,A),(A,C),(C,A)
2. (A,{B,C}),(B,{A}),(C,{A})
3. (A,{B,C}),(A,{C,B})
4. (B,C,1), (C,B,1)

```
common_conn = GROUP pairs BY (id1, id2);  
common_conn = FOREACH common_conn  
GENERATE flatten(group) as (source_id, dest_id),  
COUNT(pairs) as common_connections;
```

Connection Strength

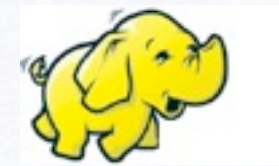
- Age
- Company Overlap
- School Overlap

Related Searches

- Let's build "Related Searches"
- Systems and Tools we use
- Hadoop MapReduce
- Managing workflow
- Serving data in production

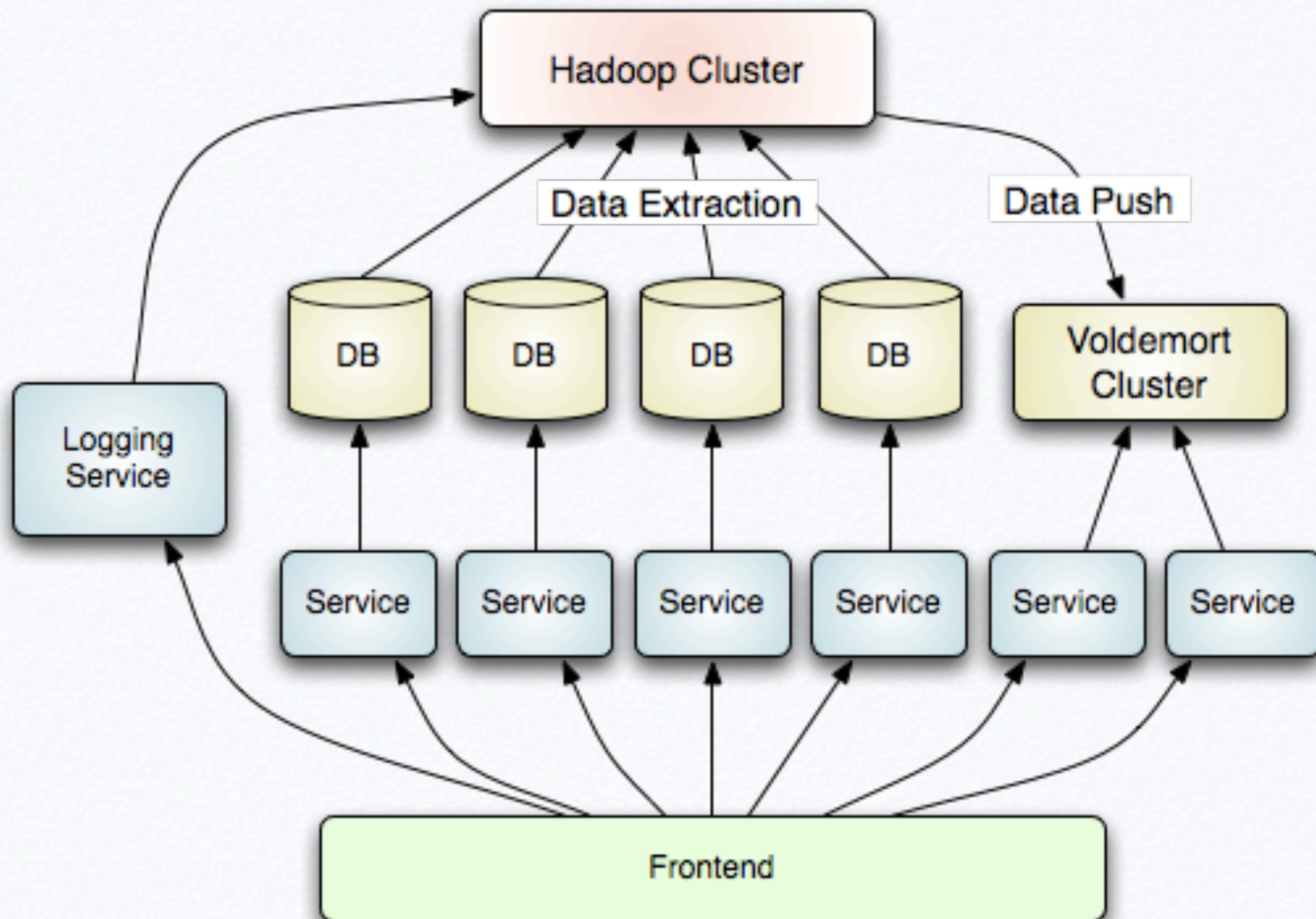
Systems and Tools

- Kafka (LinkedIn)
- Hadoop (Apache)
- Azkaban (LinkedIn)
- Voldemort (LinkedIn)



Data Cycle

20,000 Foot View Of The Data Cycle



Systems and Tools

- Kafka
 - publish-subscribe messaging system
 - transfer data from production to HDFS
- Hadoop
- Azkaban
- Voldemort

Systems and Tools

- Kafka
- Hadoop
 - Java MapReduce and Pig
 - process data
- Azkaban
- Voldemort

Systems and Tools

- Kafka
- Hadoop
- Azkaban
 - Hadoop workflow management tool
 - to manage hundreds of Hadoop jobs
- Voldemort

Systems and Tools

- Kafka
- Hadoop
- Azkaban
- Voldemort
 - Key-value store
 - store output of Hadoop jobs and serve in production

Related Searches

- Let's build "Related Searches"
- Systems and Tools we use
- Hadoop MapReduce
- Managing workflow
- Serving data in production

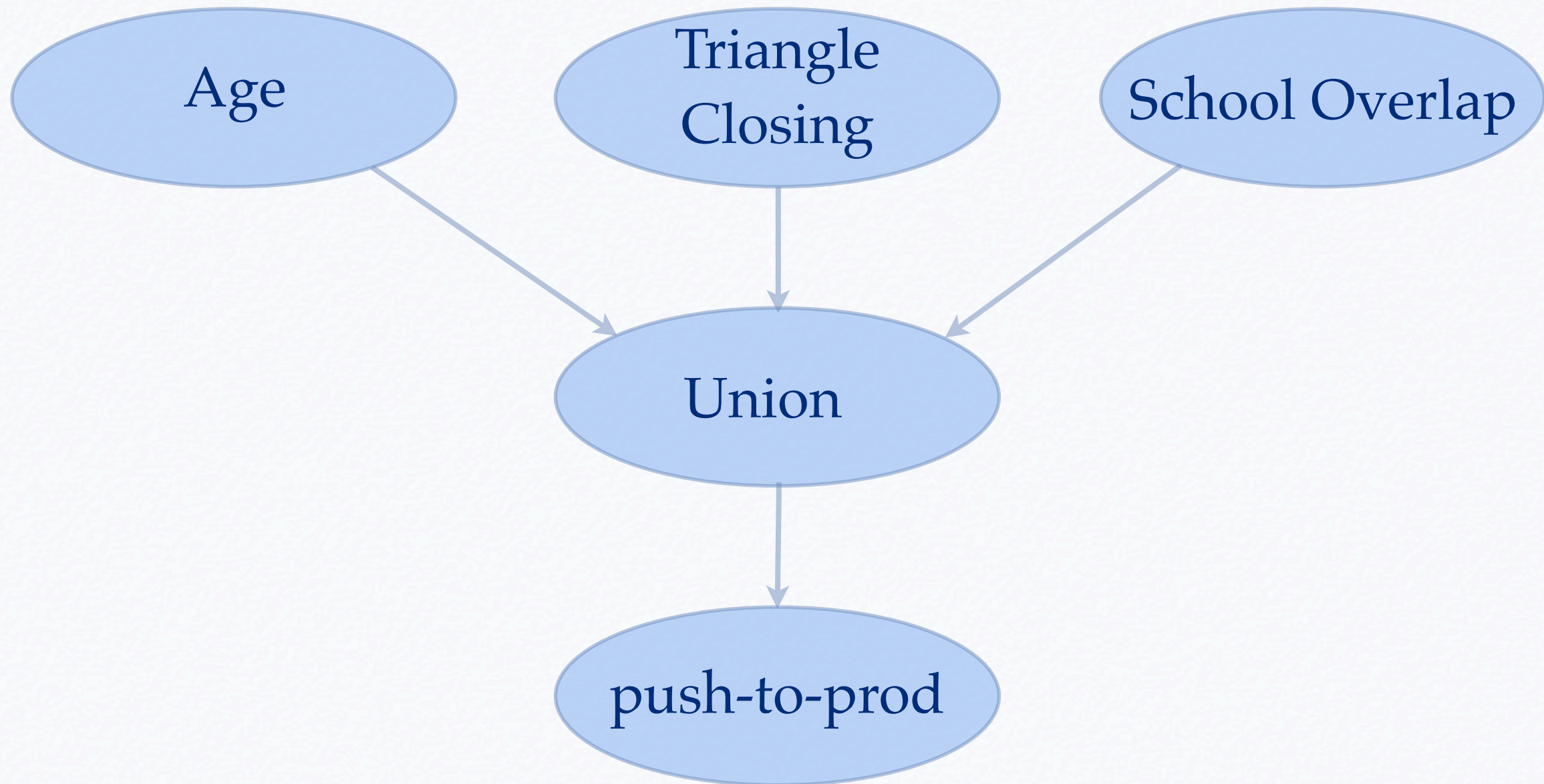
Hadoop MapReduce

- Large-scale distributed data processing
- Map phase: partition the problem in smaller steps
- Reduce phase: aggregate the output of smaller steps
- Allows parallelism and fault-tolerance

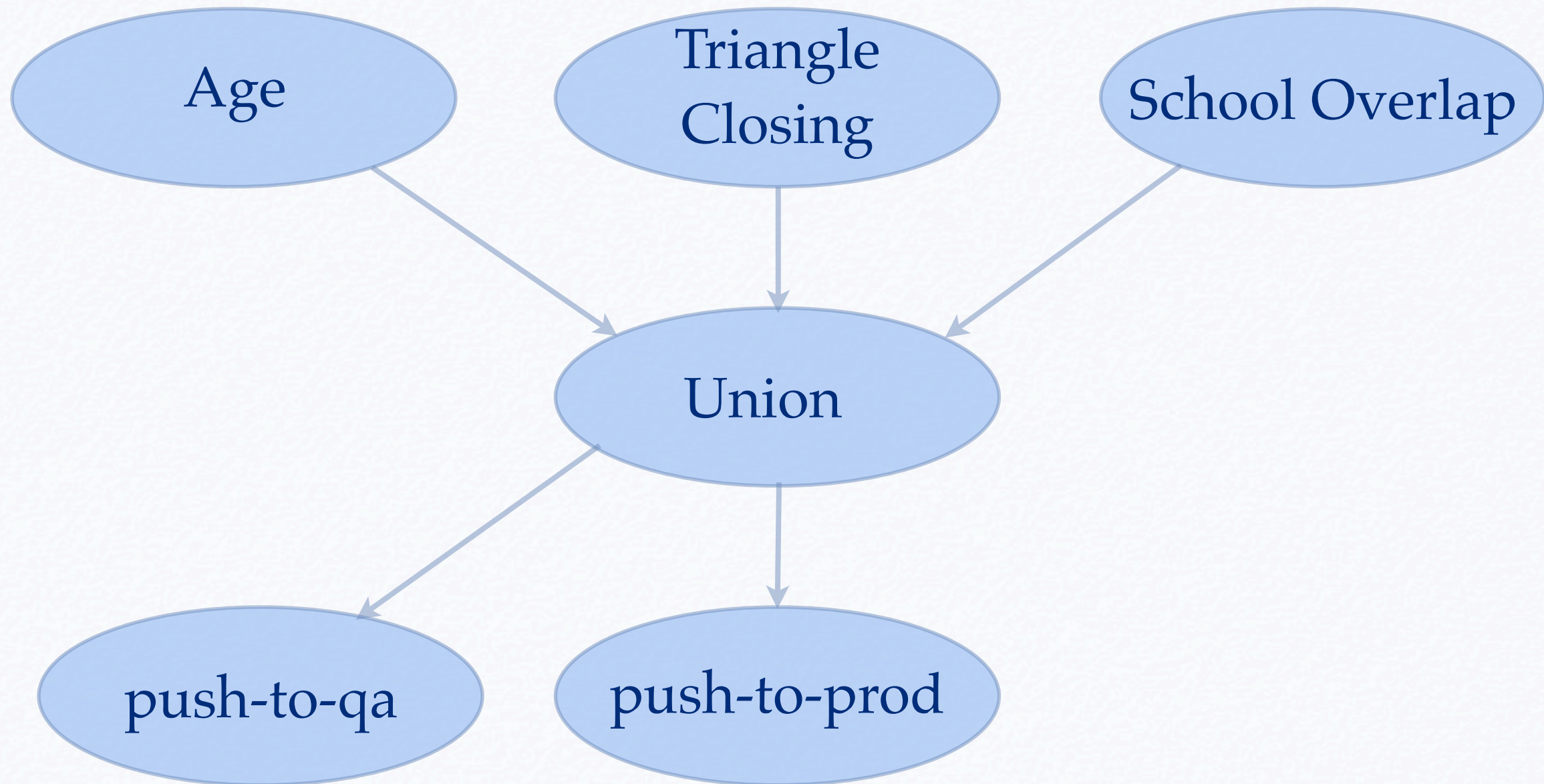
Related Searches

- Let's build "Related Searches"
- Systems and Tools we use
- Hadoop MapReduce
- Managing workflow
- Serving data in production

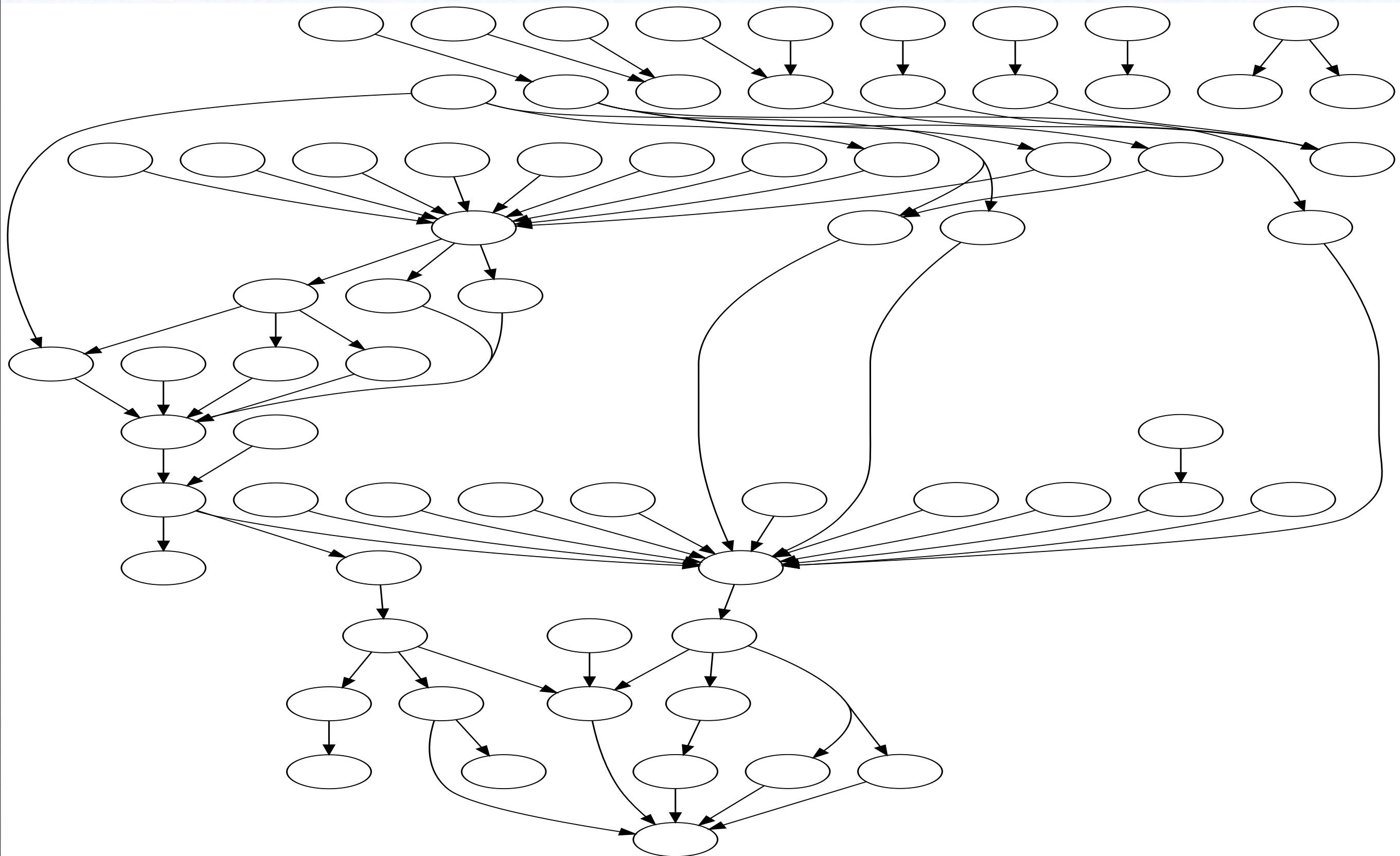
Our Workflow



Our Workflow



Sample Workflow

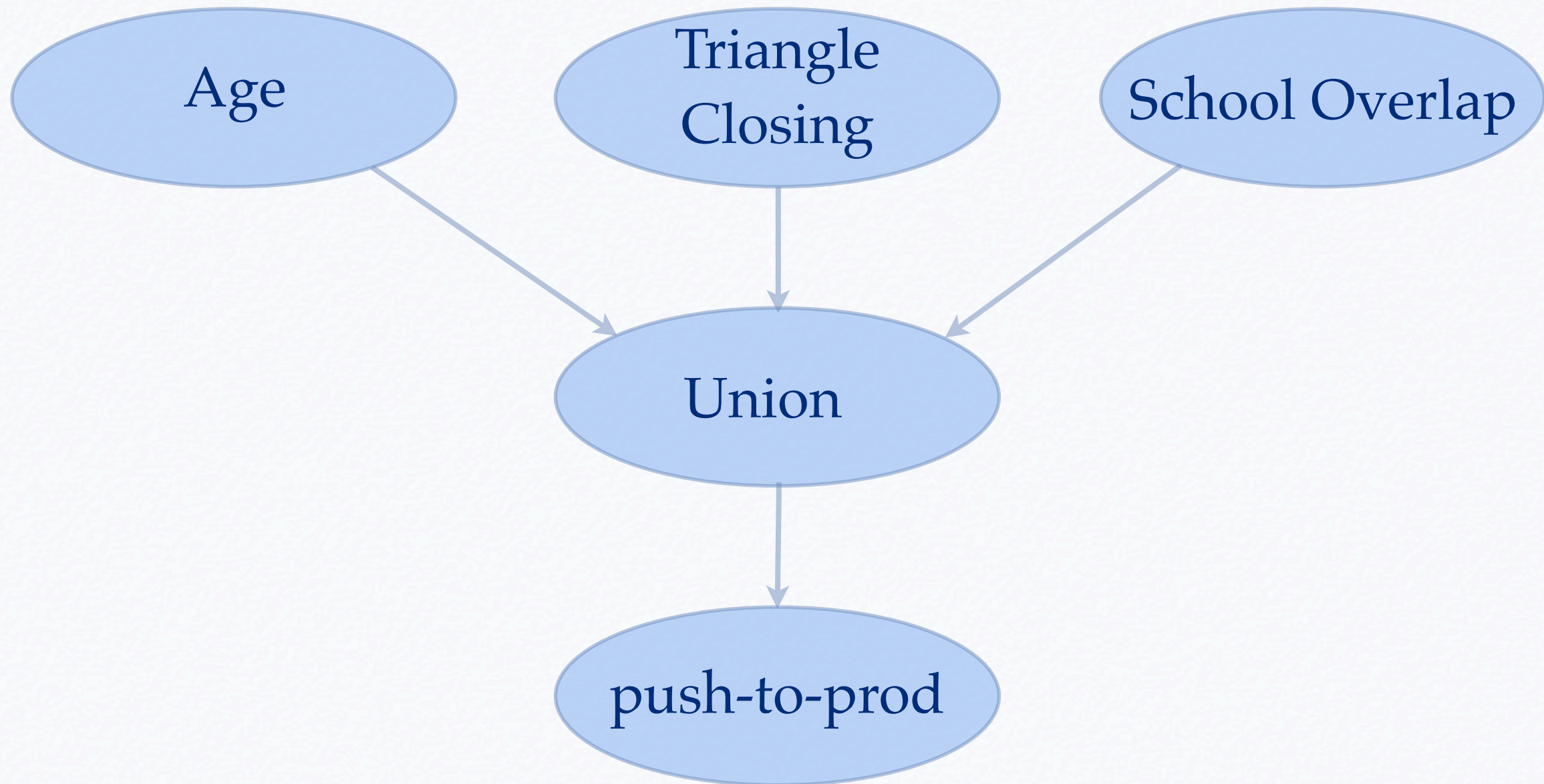


Azkaban Workflow Management

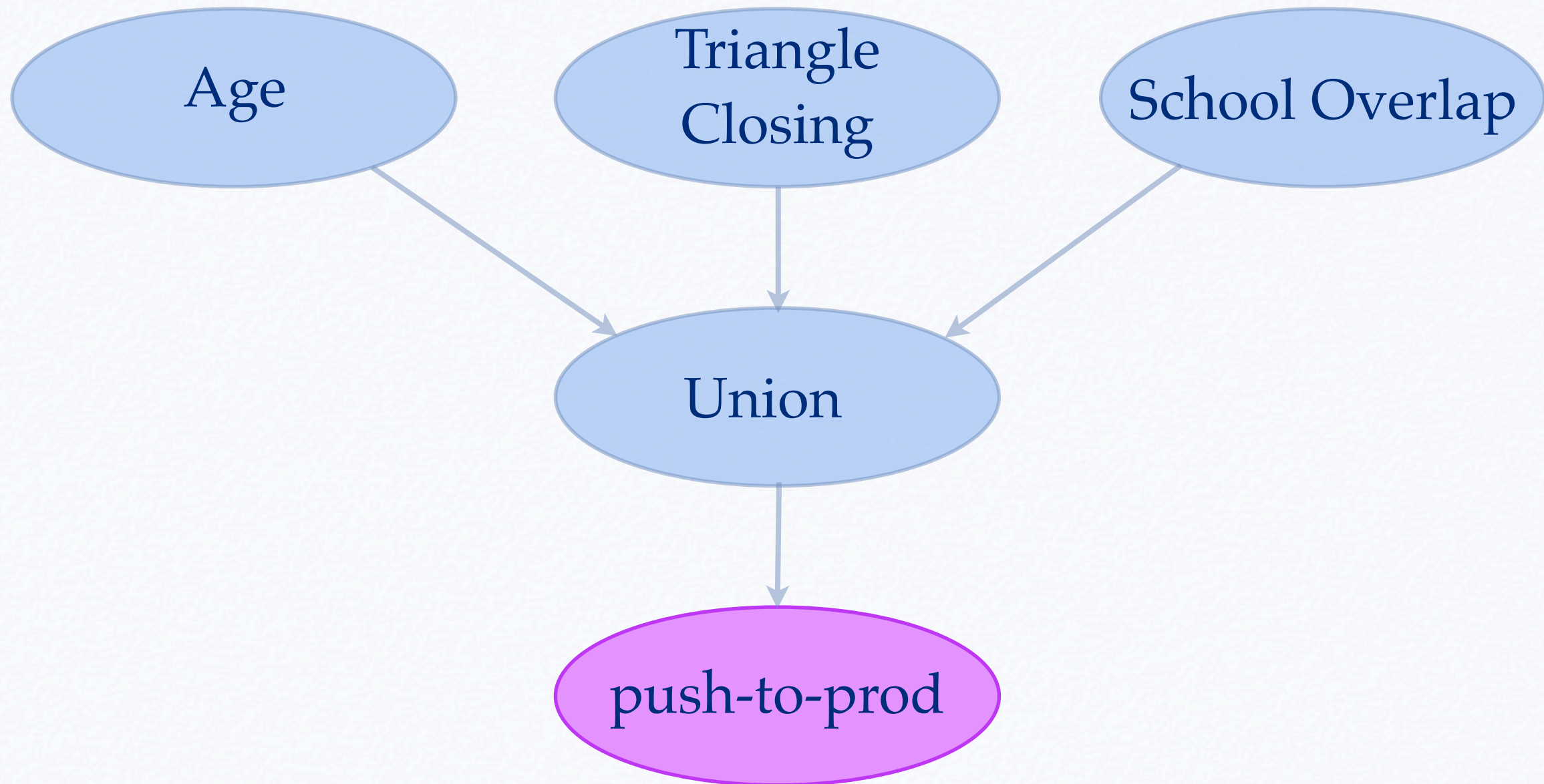


- Dependency management
- Regular Scheduling
- Monitoring
- Diverse jobs: Java, Pig, Clojure
- Configuration / Parameters
- Resource control / locking
- Restart / Stop / Retry
- Visualization
- History
- Logs

Our Workflow



Our Workflow



Related Searches

- Let's build "Related Searches"
- Systems and Tools we use
- Hadoop MapReduce
- Managing workflow
- Serving data in production

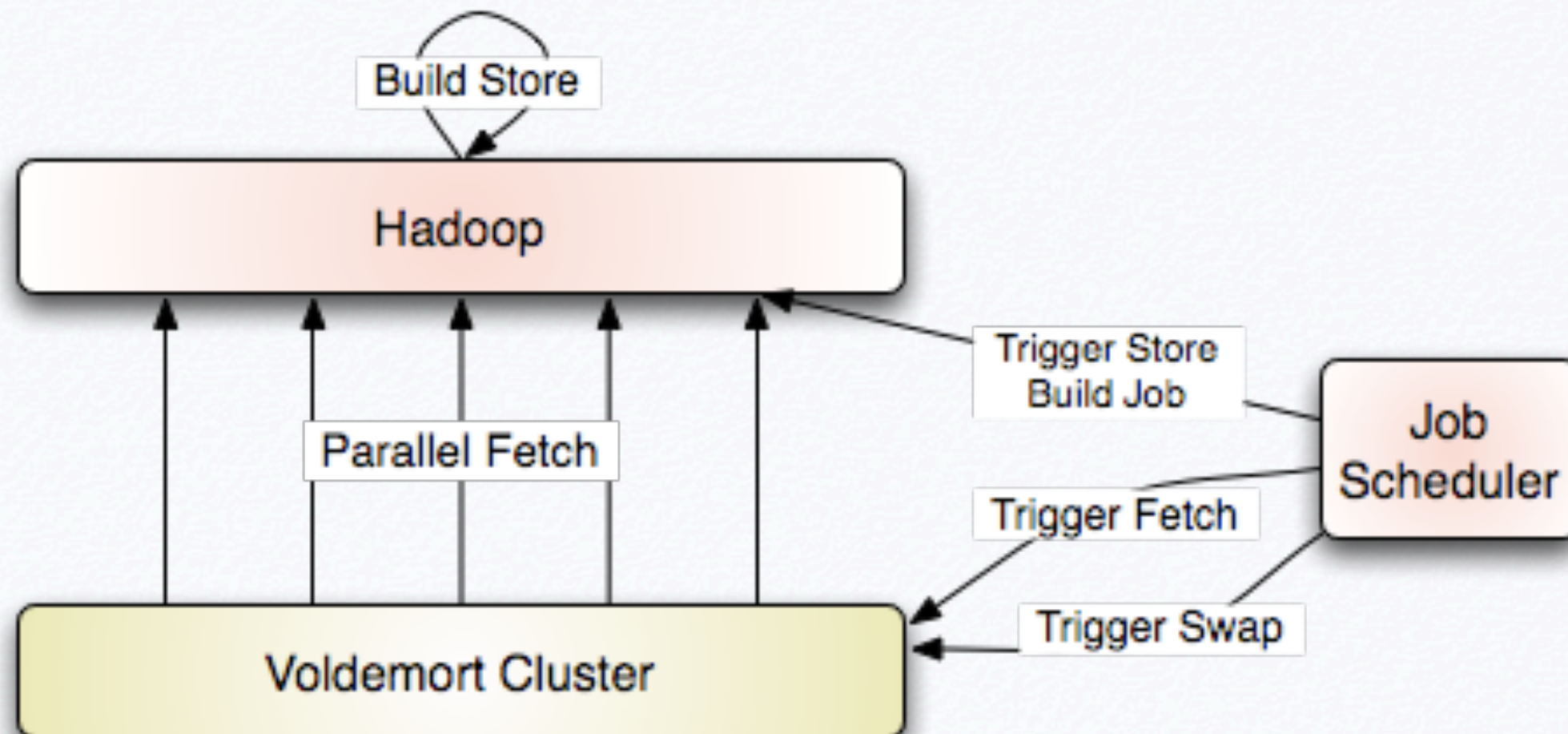
Voldemort

- Large amount of data / Scalable
- Quick lookup / low latency
- Versioning and Rollback
- Fault tolerance through replication
- Read only
- Offline index building



Voldemort RO Store

Read-Only Store Build and Swap Process



Outline

- Analytics products at LinkedIn
- Deep Dive - Connection Strength
- Deep Dive - Related Searches
- Fun Facts: Sequencing the Startup DNA

Related Searches

- Millions of Searches everyday
- Help users to explore and refine their queries

Related searches for **Web Developer**

web designer

php

web development

ruby on rails

software engineer

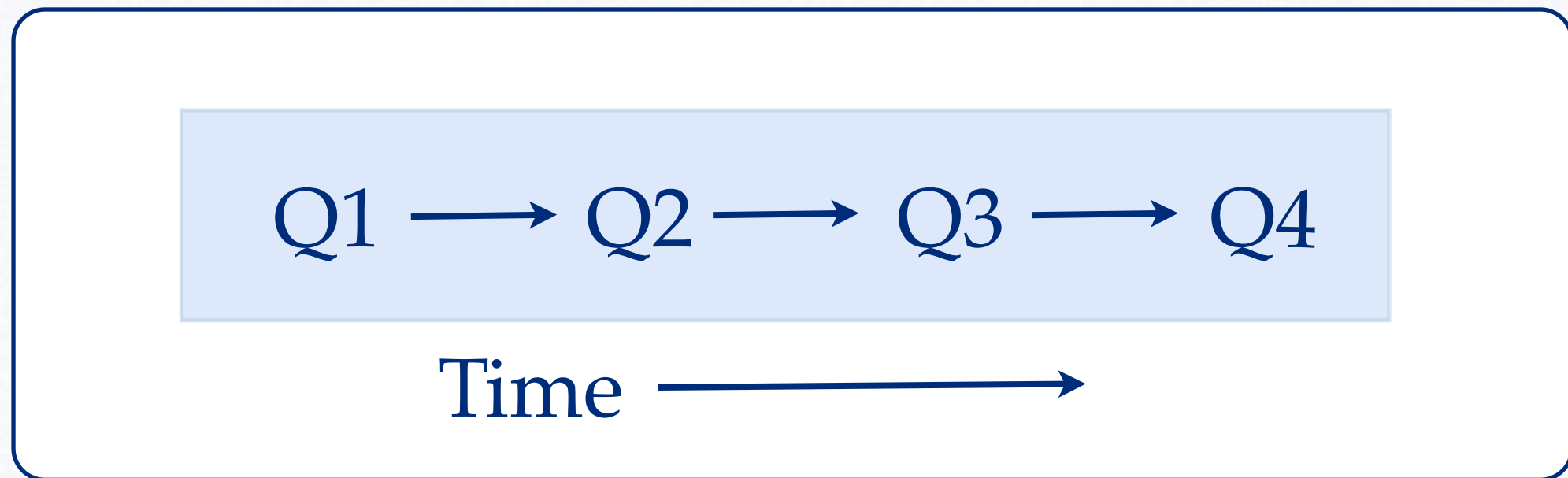
programmer

developer

html

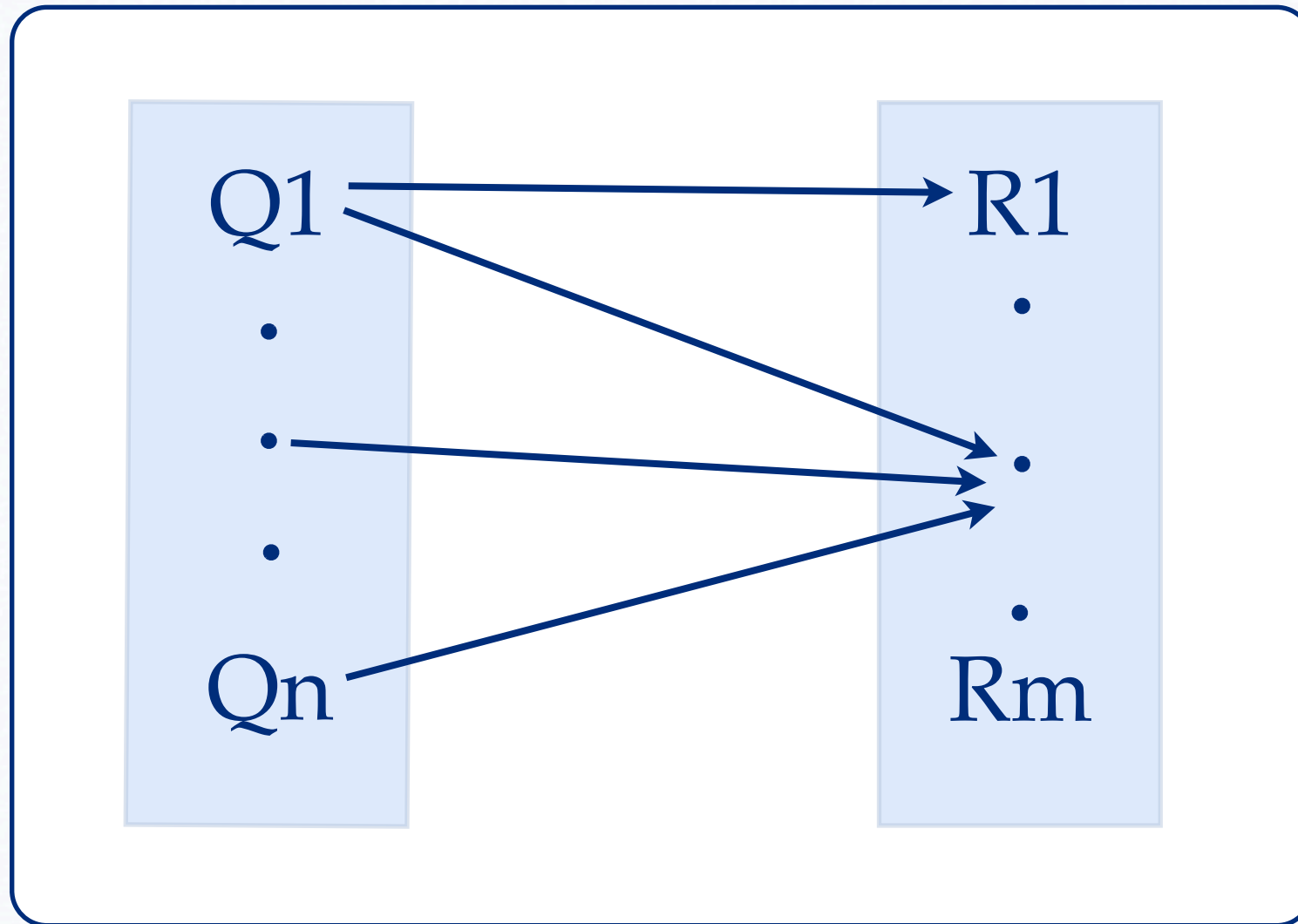
How to build Related Searches?

- Collaborative Filtering: searches done in the same session



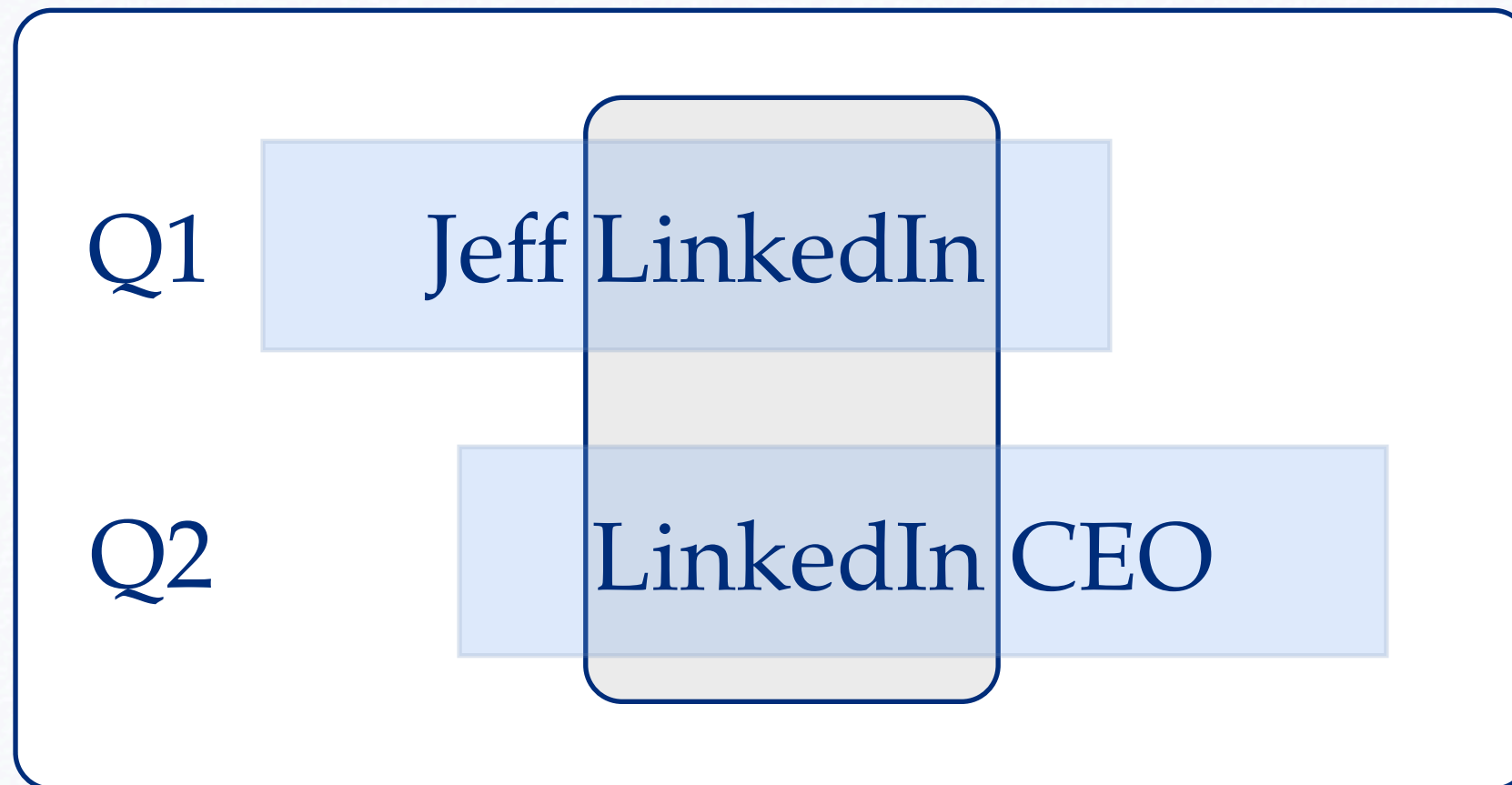
How to build Related Searches?

- Searches correlated by results clicks



How to build Related Searches?

- Searches with overlapping terms



Outline

- Analytics products at LinkedIn
- Deep Dive - Connection Strength
- Deep Dive - Related Searches
- Fun Facts: Sequencing the Startup DNA

Sequencing the Startup DNA

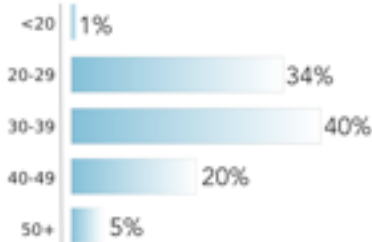
Sequencing the Startup DNA



Top* 10 most entrepreneurial business schools

1. STANFORD GRADUATE SCHOOL OF BUSINESS
2. MIT Sloan School of Management
3. Berkeley Haas School of Business
4. MIT Sloan School of Management
5. Wharton University of Pennsylvania
6. MIT Sloan School of Management
7. MIT Sloan School of Management
8. MIT Sloan School of Management
9. MIT Sloan School of Management
10. MIT Sloan School of Management

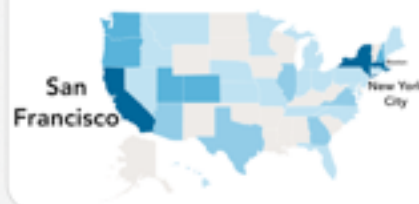
Founders' age at first startup



Where do founders come from?



Top* startup regions



Average position length (before first startup):

2.5 years

Founders and their fields of study



Top* industries connected to founders

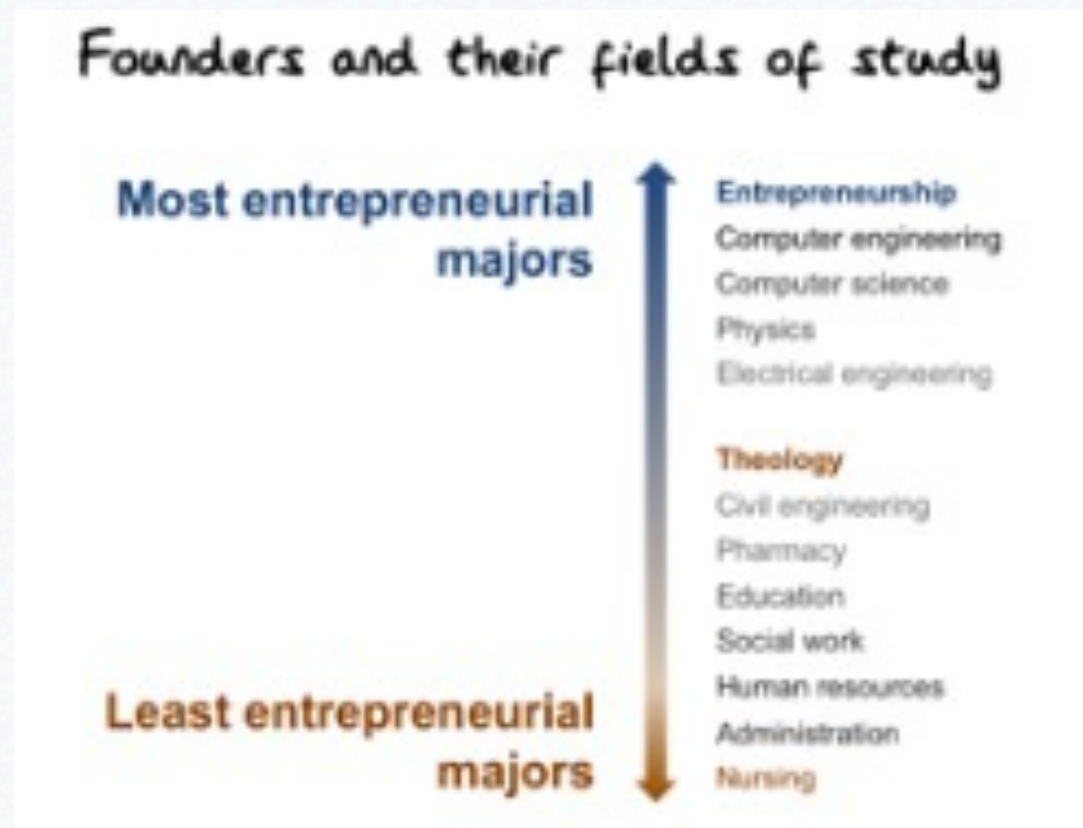
1. Venture Capital
2. Online Publishing
3. Internet
4. Recruiting

* Top, most entrepreneurial = over-represented among entrepreneurs

Done by Monica Rogati at LinkedIn

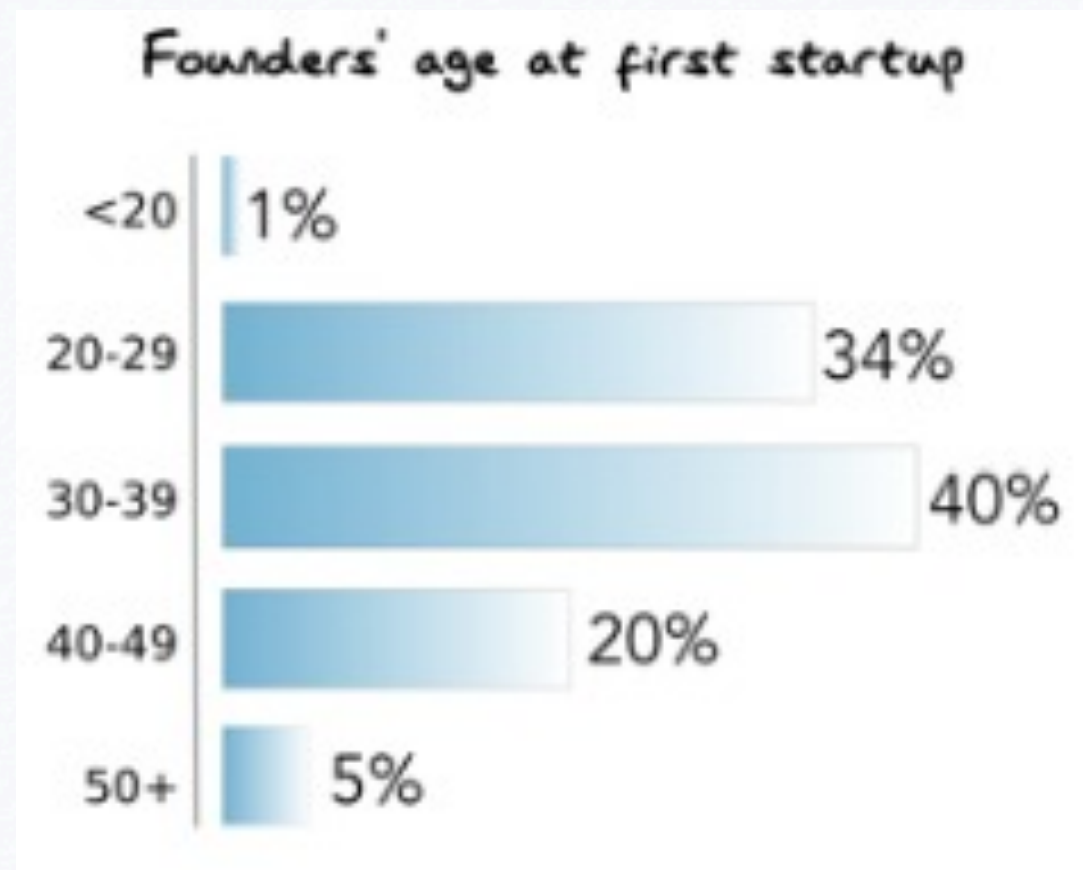
Sequencing the Startup DNA

- Top majors: Entrepreneurship, Computer Engineering
- Bottom: Nursing, Administration



Sequencing the Startup DNA

- Most founders age at first startup between 20 and 40



Sequencing the Startup DNA

- Top regions: San Francisco, NYC



Sequencing the Startup DNA

- Top Business Schools:
Stanford, Harvard, MIT Sloan

Top* 10 most entrepreneurial business schools

1.  **STANFORD**
GRADUATE SCHOOL OF BUSINESS
2.  **HARVARD**
BUSINESS
SCHOOL
3.  **MIT Sloan**
MANAGEMENT
4.  **Berkeley**
Haas School of Business
5.  **Tuck School of Business**
at Dartmouth
6.  **Wharton**
UNIVERSITY OF PENNSYLVANIA
7.  **Columbia**
Business
School
8.  **BABSON**
9.  **UNIVERSITY**
VIRGINIA
Darden School of Business
10.  **JOHNSON**
Cornell University

Things Covered

- Analytics products at LinkedIn
- Deep Dive - Connection Strength
- Deep Dive - Related Searches
- Fun Facts: Sequencing the Startup DNA

SNA Team

- Thanks to SNA Team at LinkedIn
- <http://sna-projects.com>
- We are hiring!

Questions?