



**In-Memory  
Computing**  
S U M M I T

NORTH  
AMERICA  
2018

# Scaling up the Contacts Insights with Activity Graph

---

**Praveen Innamuri, Zhidong Ke**  
Salesforce

# Agenda

- **Introduction**
- Activity Insights Context
- Why using a Graph to model context
- Key problems solved and lessons learned
- Wrap up and QAs



# Forward-Looking Statement

## Statement under the Private Securities Litigation Reform Act of 1995

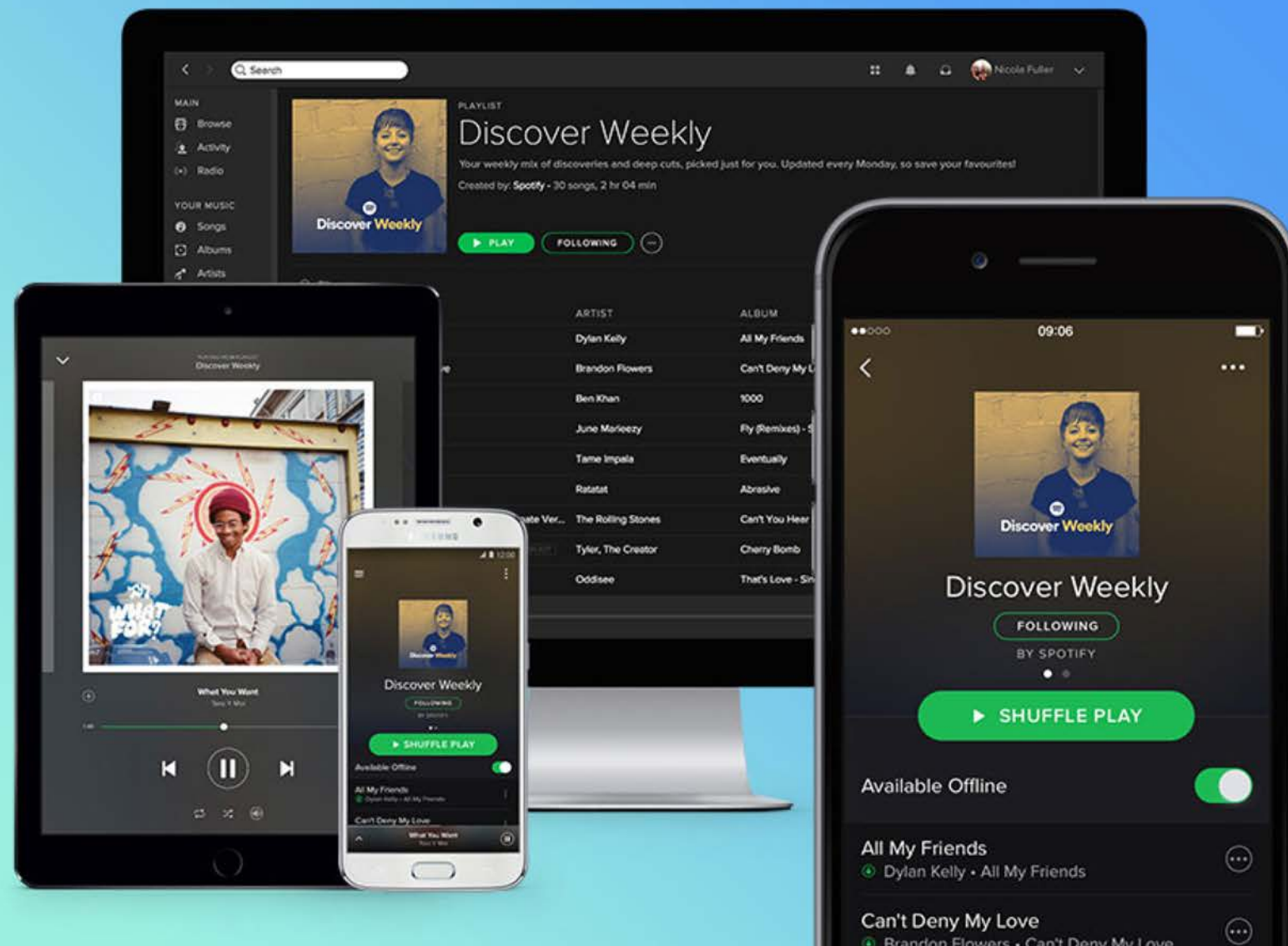
This presentation may contain forward-looking statements that involve risks, uncertainties, and assumptions. If any such uncertainties materialize or if any of the assumptions proves incorrect, the results of salesforce.com, inc. could differ materially from the results expressed or implied by the forward-looking statements we make. All statements other than statements of historical fact could be deemed forward-looking, including any projections of product or service availability, subscriber growth, earnings, revenues, or other financial items and any statements regarding strategies or plans of management for future operations, statements of belief, any statements concerning new, planned, or upgraded services or technology developments and customer contracts or use of our services.

The risks and uncertainties referred to above include – but are not limited to – risks associated with developing and delivering new functionality for our service, new products and services, our new business model, our past operating losses, possible fluctuations in our operating results and rate of growth, interruptions or delays in our Web hosting, breach of our security measures, the outcome of any litigation, risks associated with completed and any possible mergers and acquisitions, the immature market in which we operate, our relatively limited operating history, our ability to expand, retain, and motivate our employees and manage our growth, new releases of our service and successful customer deployment, our limited history reselling non-salesforce.com products, and utilization and selling to larger enterprise customers. Further information on potential factors that could affect the financial results of salesforce.com, inc. is included in our annual report on Form 10-K for the most recent fiscal year and in our quarterly report on Form 10-Q for the most recent fiscal quarter. These documents and others containing important disclosures are available on the SEC Filings section of the Investor Information section of our Web site.

Any unreleased services or features referenced in this or other presentations, press releases or public statements are not currently available and may not be delivered on time or at all. Customers who purchase our services should make the purchase decisions based upon features that are currently available.

Salesforce.com, inc. assumes no obligation and does not intend to update these forward-looking statements.





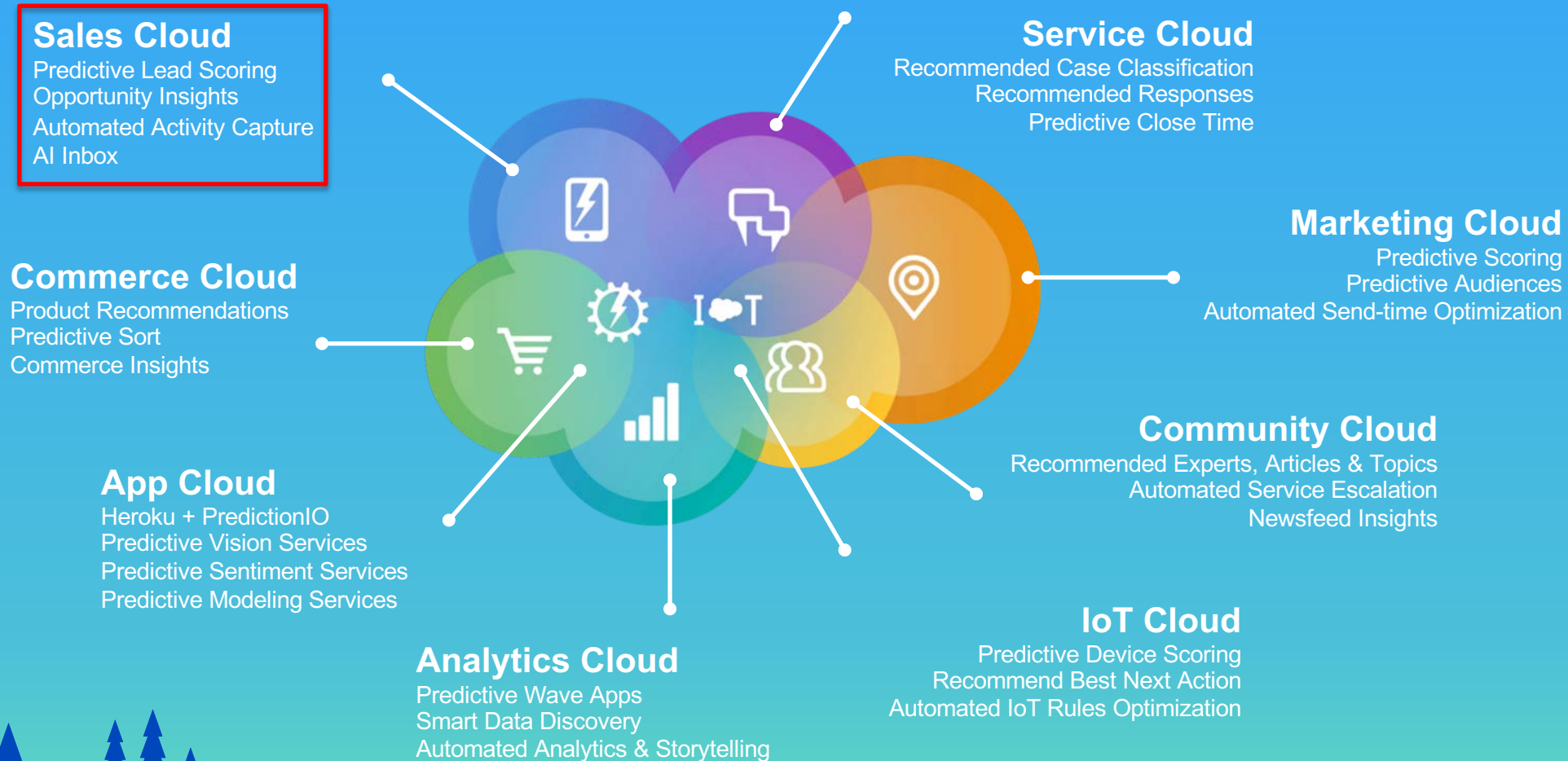
# Why I'm talking about Spotify..

- No, I'm not promoting to use Spotify.
- I should rather promote to use Salesforce products.



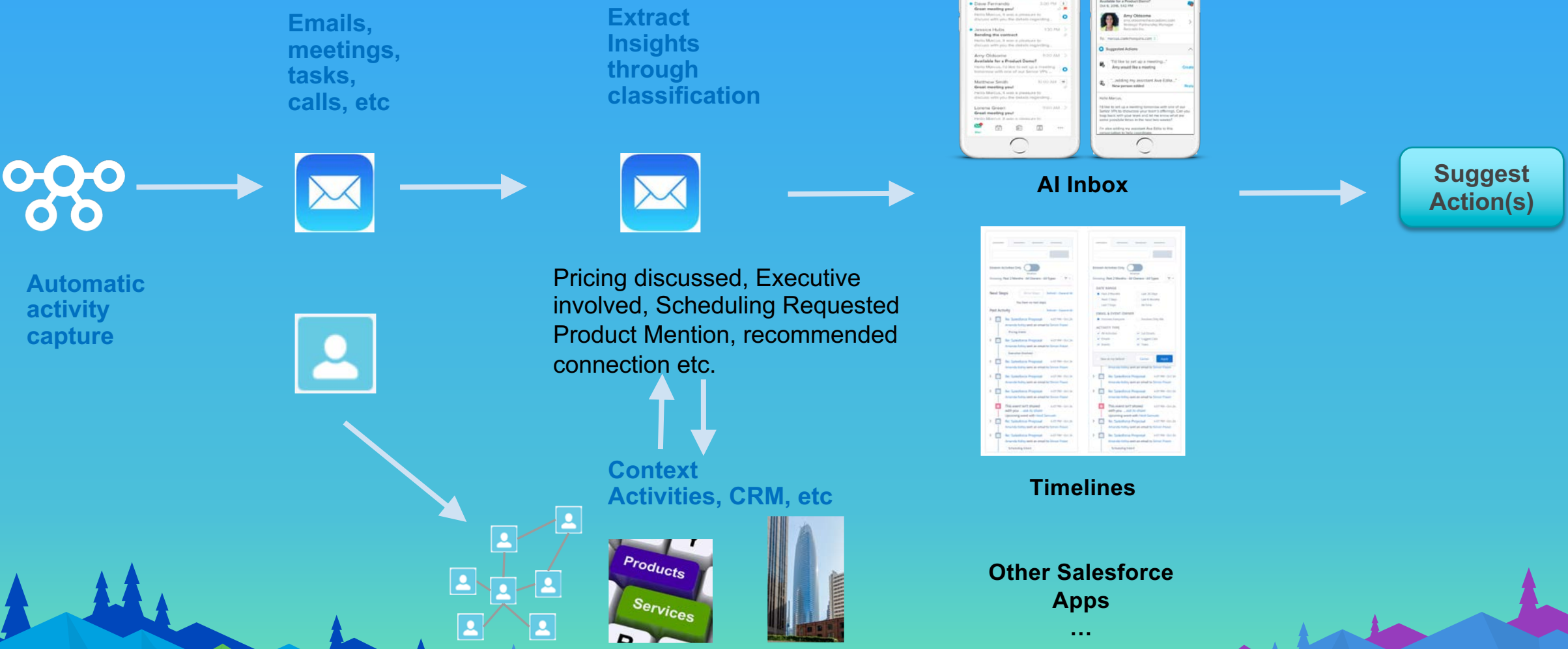
# The Age of the Customer

Salesforce Apps + AI = Whole New Customer Experience





# Augment CRM using AI and activity





# Salesforce Apps - Closest Connections

Carrier 2:57 PM

← Evan Ku



Profile Related (0)



 Acme Health Inc



 San Francisco

[Show All Contact Sources](#)

PEOPLE CONNECTED TO EVAN

 **Michael Machado**  [View](#)  
michael.machado@acmehealthinc.com  
Because of emails and meetings they h...

 **Richa Bansal**  [View](#)  
richa.bansal@acmehealthinc.com  
Because of emails and meetings t...



 **Sriram Iyer**  [View](#)  
sriram.iyer@acmehealthinc.com  
Because of emails and meetings t...



[View More](#)



Call SMS Email



Carrier 2:57 PM



← Evan Ku

 **Michael Machado**  [View](#)  
michael.machado@acmehealthinc.com  
Because of emails and meetings they h...

 **Richa Bansal**  [View](#)  
richa.bansal@acmehealthinc.com  
Because of emails and meetings they h...



 **Sriram Iyer**  [View](#)  
sriram.iyer@acmehealthinc.com  
Because of emails and meetings they h...



 **Eric Jacobs**  [View](#)  
eric.jacobs@acmehealthinc.com  
Because of emails they had together.



 **Amy Miller**  [View](#)  
amy.miller@acmehealthinc.com  
Because of emails and meetings they h...



Carrier 2:58 PM



← Evan Ku

 **Michael Machado**  [View](#)  
michael.machado@acmehealthinc.com  
Because of emails and meetings they h...


 **Richa Bansal**  [View](#)  
richa.bansal@acmehealthinc.com  
Because of emails and meetings they h...

 **Sriram Iyer**  [View](#)  
sriram.iyer@acmehealthinc.com  
Because of emails and meetings they h...

 **Eric Jacobs**  [View](#)  
eric.jacobs@acmehealthinc.com

Michael   Evan

**Strong Connection**



Michael Machado sent many emails in the last six months. They were in many meetings in the last six months.

[Email](#) [View Profile](#)



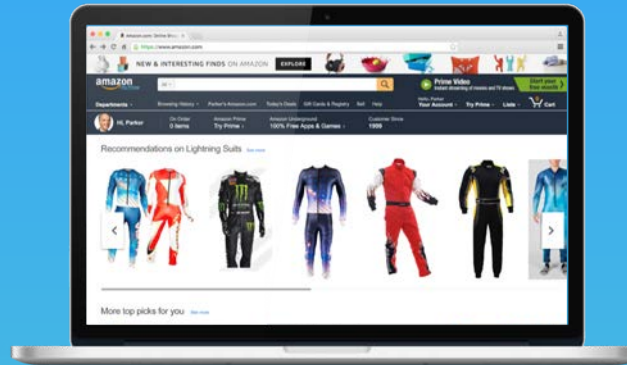
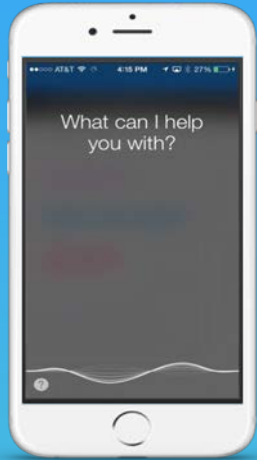
# Agenda

- Introduction
- **Contact Insights Context**
- Why using a Graph to model context
- Key problems solved and lessons learned
- Wrap up and QAs

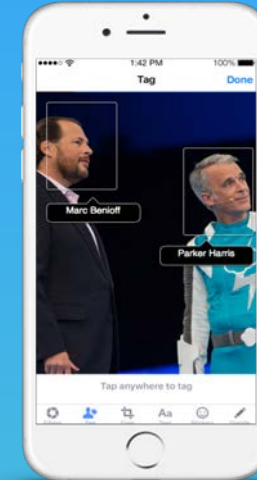


# AI & Context

Data + Algorithms + Compute = Killer Apps



amazon



facebook

What does all those apps have in common? User context

# Consumer vs Enterprise Context

## User isn't the product but the customer

- Retention, privacy, GDPR, security, auditing, etc

## Context has to be scoped

- Cannot be used globally: organization, team, user levels

## Very rich

- Goes way beyond user context: organizations/groups/teams, products and services, companies, different types of activities across many different products, etc

## Very dynamic

- Fast coming data with lots of interaction points



# Context enables us to deliver deeper insights.

## Go beyond using a single email to make classification and action recommendation

**This sender looks familiar, how well should I know him / her?**

- Are we strongly connected? Is he or she important to my accounts or opportunities? etc

**Is this email discussing products or services that my company sell?**

**Is this email discussing competitors?**

**Who, in my org, can help me sell to an individual or company?**

- Supply relevant background information on a particular individual or company
- Identify who is the key decision maker
- Give me historical information for that individual or company
- Make an introduction for me

etc



# Agenda

- Introduction
- Activity Insights Context
- **Why using a Graph to model context**
- Key problems solved and lessons learned
- Wrap up and QAs





# A graph is an efficient means for encoding relationships.

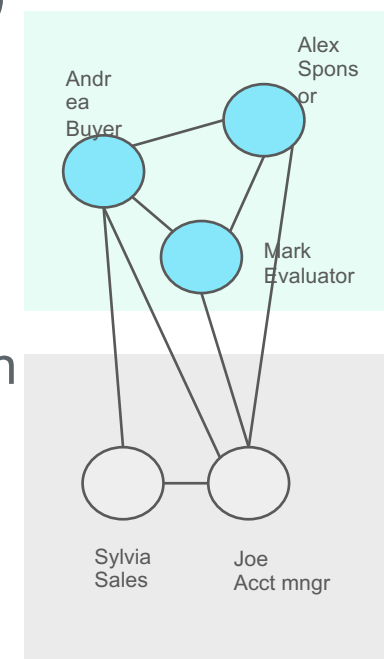
## An org can have thousands of contacts

- These contacts exist within the org itself (e.g., sales rep, account exec)
- Perhaps more importantly, contacts extend beyond the org (e.g., buyers)

15 Jan Email - Sylvia to Andrea: introduction  
 20 Jan Meeting - Created by Andrea with Sylvia  
 31 Jan Email - Andrea to Sylvia & Mark: info request  
 01 Feb Email - Sylvia to Andrea & Mark: product info  
 04 Feb Email - Andrea to Sylvia & Joe  
 17 Feb Meeting created by Andrea with Alex and Joe  
 ...

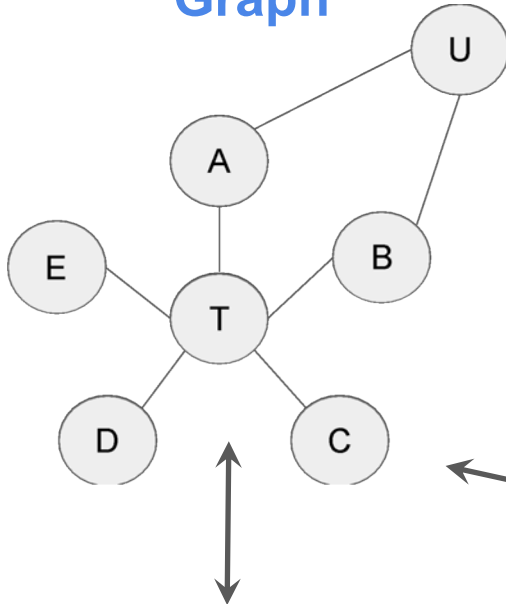
## That same org can have millions of events per week

- Events (e.g., meetings, emails, phone calls) connect contacts and indicate a relationship
- The number and nature of events between contacts can indicate strength of connection / relationship



# Coupled with AI models, our graph delivers Contextual Services

## Graph



- Identify hot leads
- Best time to email
- Recommend connections
- Updated contact info notification
- Suggest recipients, or rooms, for meetings
- Identify contact's role: economic buyer, evaluator, influencer, etc.
- Relationship with contact: e.g., strength of connection, communication topics

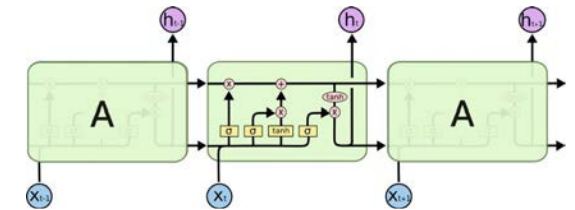
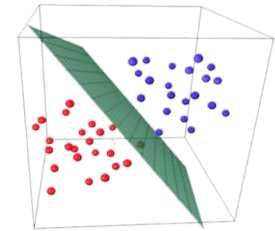
## Context

Who is a particular email from and why should I care?  
Role, latest communication, meeting history, mutual friends, contact info, etc.



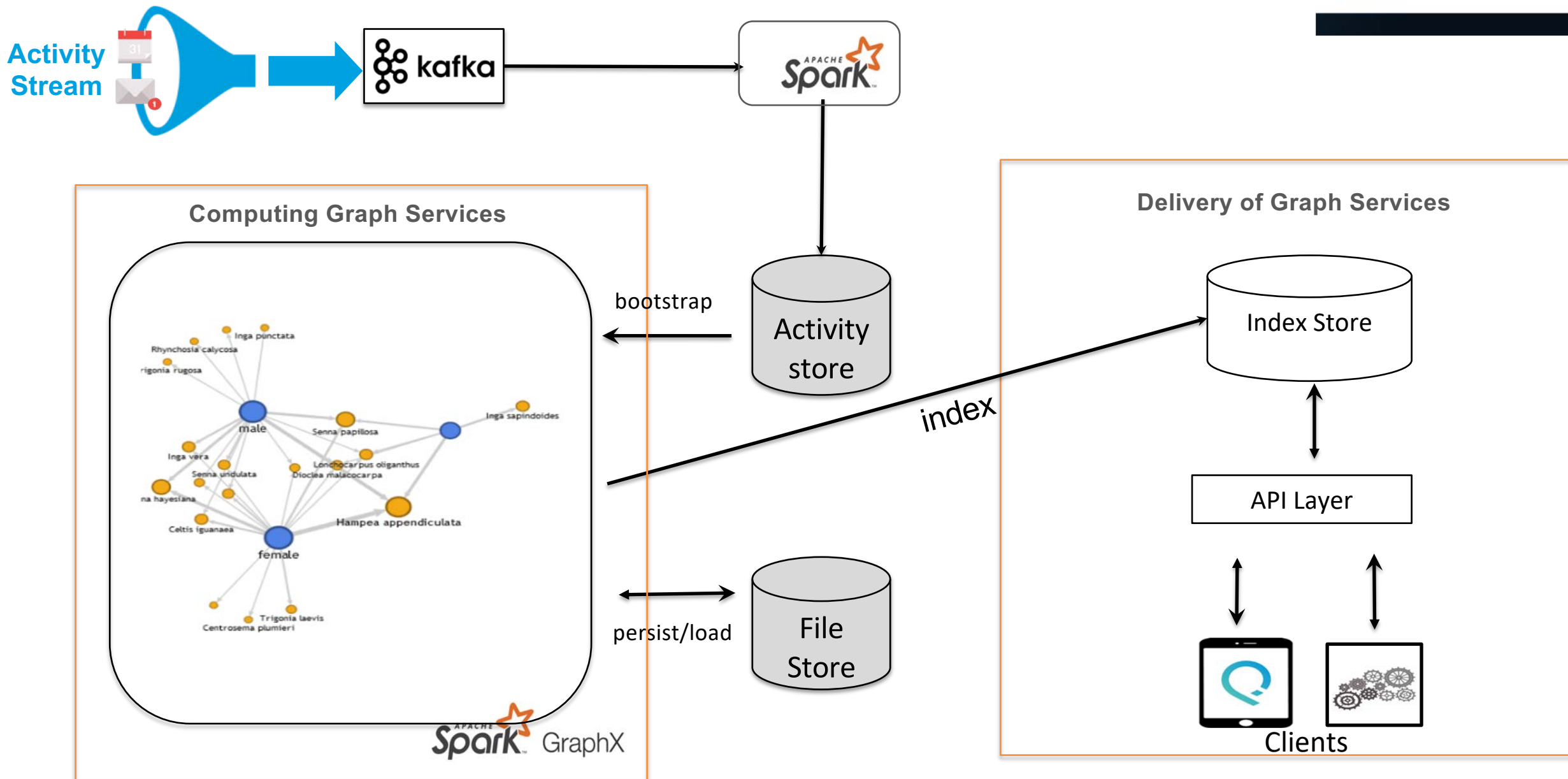
## Insights

## Models



- Pricing discussed
- Scheduling requested
- Exec involved
- etc.

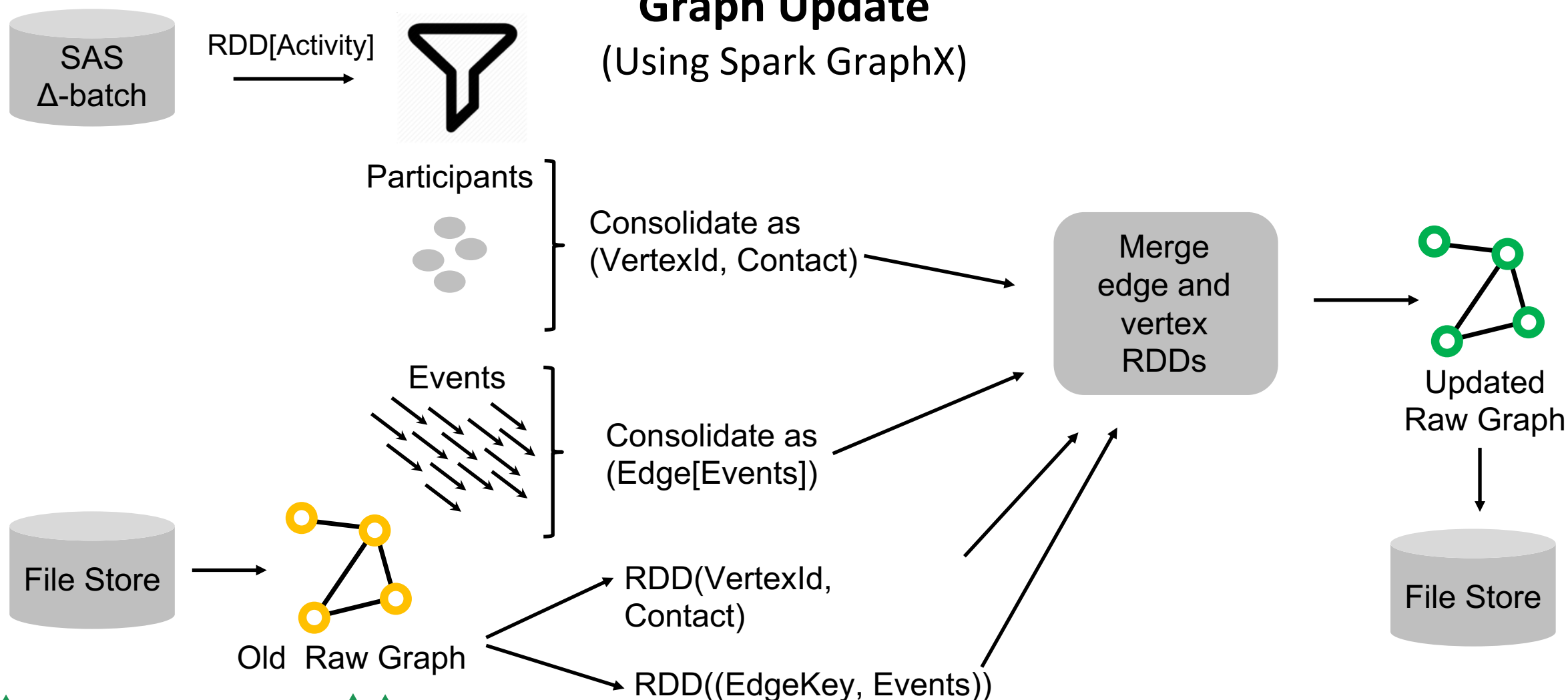
# High Level graph generation architecture



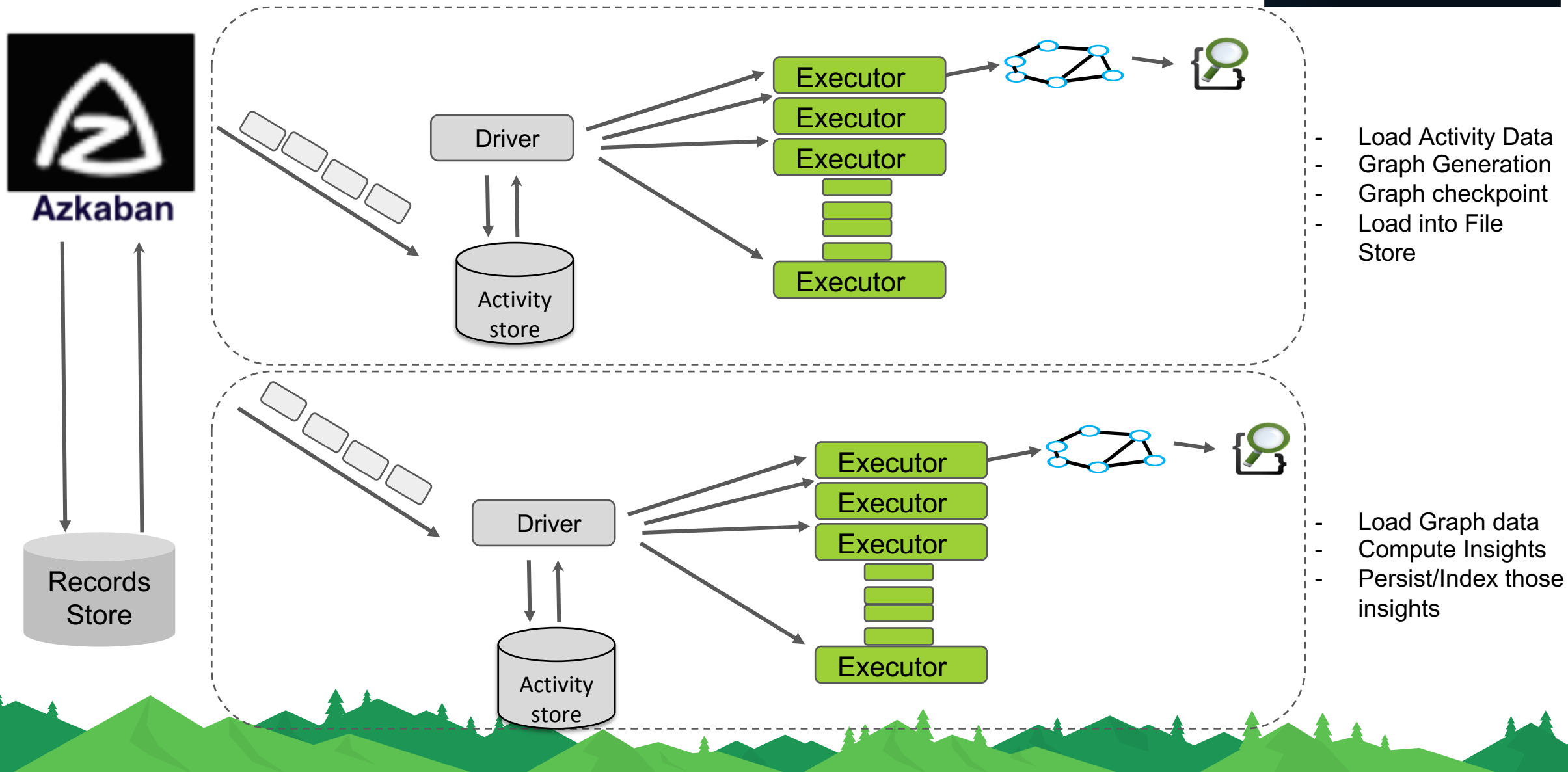
# Activity events to create / update the raw graph.

## Graph Update

(Using Spark GraphX)



# Architecture Diagram - Onboarding for all Orgs





# Agenda

- Introduction
- AI & context
- Why using a Graph to model context
- **Problems & Lessons Learned**
- Wrap up and QAs



# Memory Issues

**java.lang.OutOfMemoryError: Java heap space**

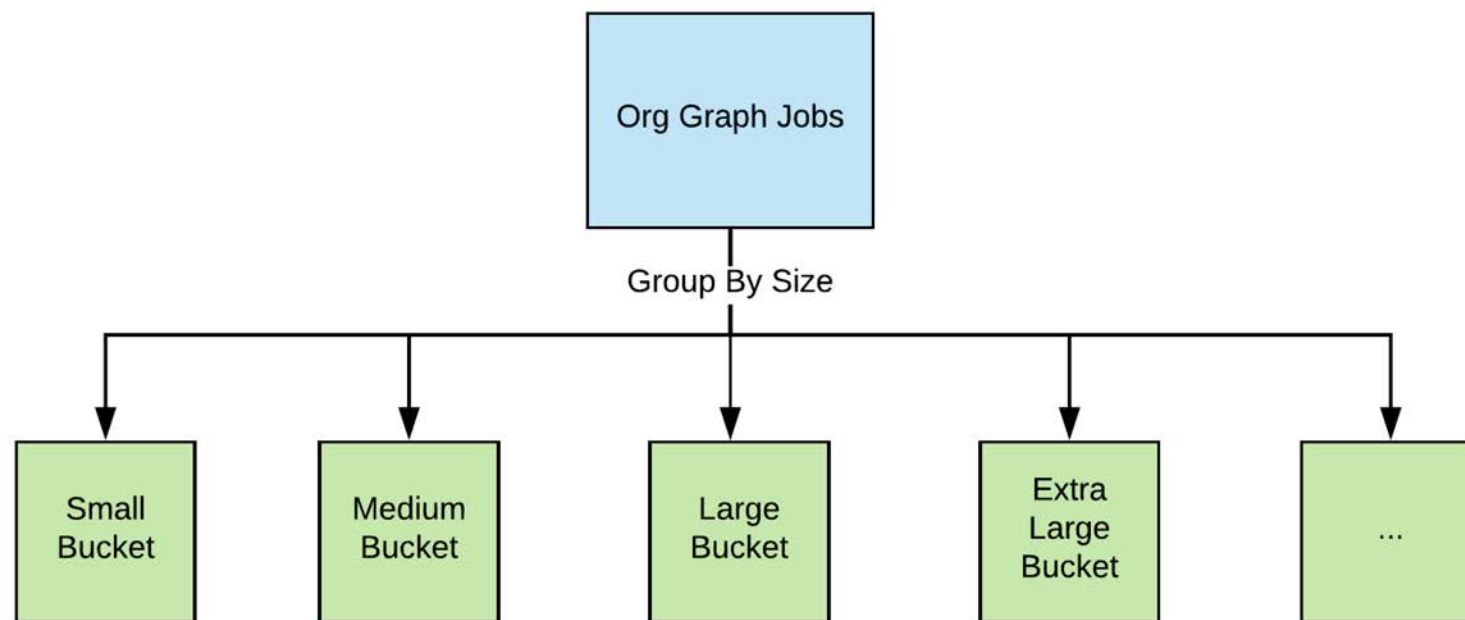
```
at java.util.Arrays.copyOf(Arrays.java:3236) ~[na:1.8.0_121]  
at java.io.ByteArrayOutputStream.grow(ByteArrayOutputStream.java:118)  
at java.io.ByteArrayOutputStream.ensureCapacity(ByteArrayOutputStream.java:93)  
at java.io.ByteArrayOutputStream.write(ByteArrayOutputStream.java:153)
```

2X memory →

Over Provision => \$\$\$



# Bucketing Strategy



## Find Right Memory, #Executors and #Partitions Per Bucket



Created by [blog.3back.com](http://blog.3back.com)

# Spark Job Got Stuck Before Reading Data

```
val df = spark.read.avro("input/*.avro")
```

Too many small files =>





# Solution

Bypass the metadata fetch

```
val df = spark.read.avro("input/*.avro")
```

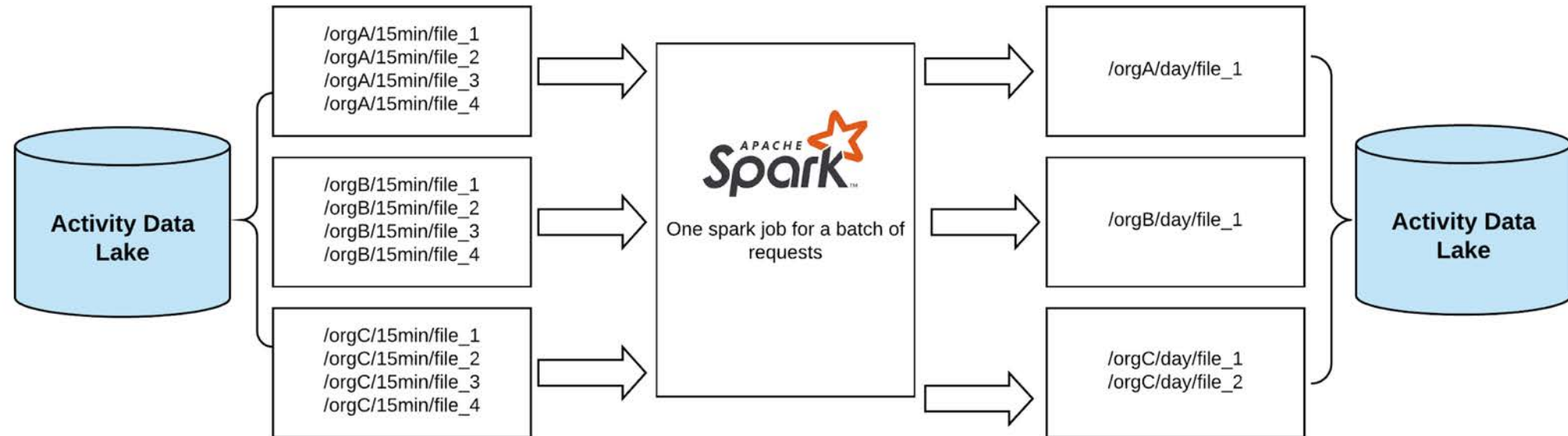


```
val input = sc.parallelize(List("input/"), 10).map(_.readData)
```



# Compaction Framework

## Compact small files



# Scaling

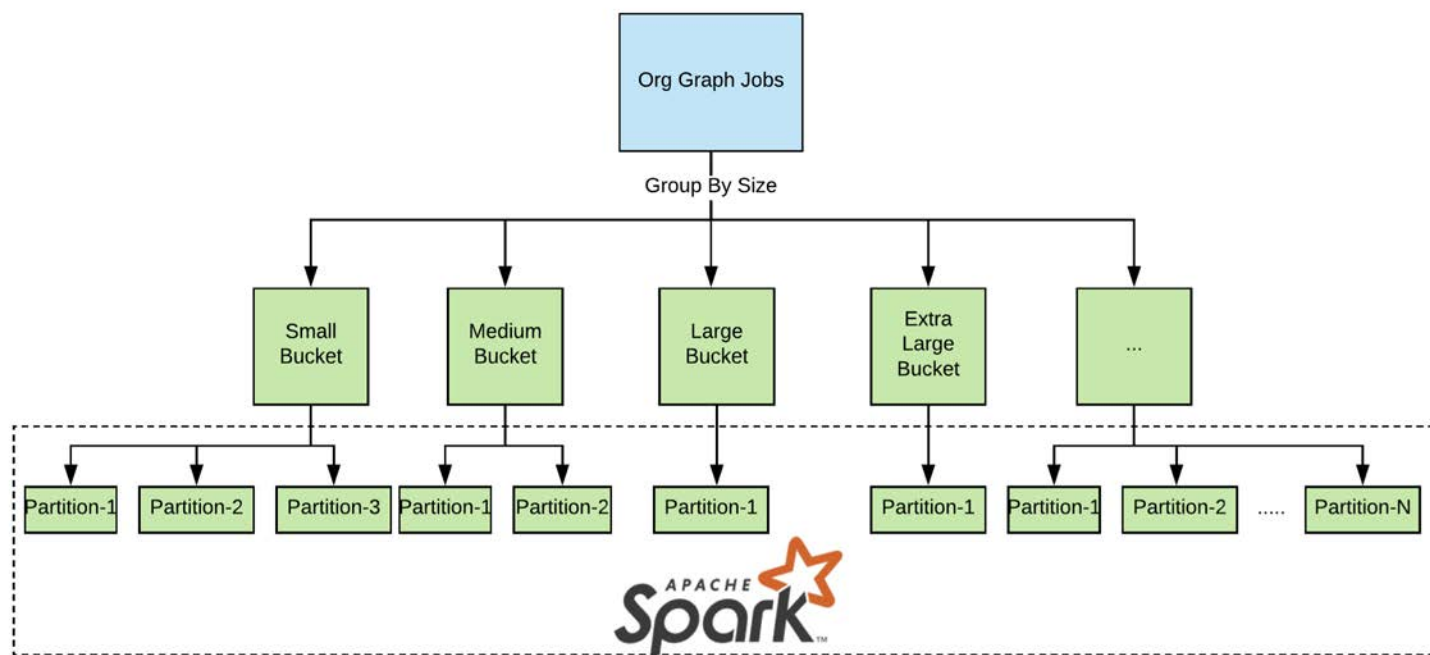
How to scale up from 0 -> Thousands of orgs?



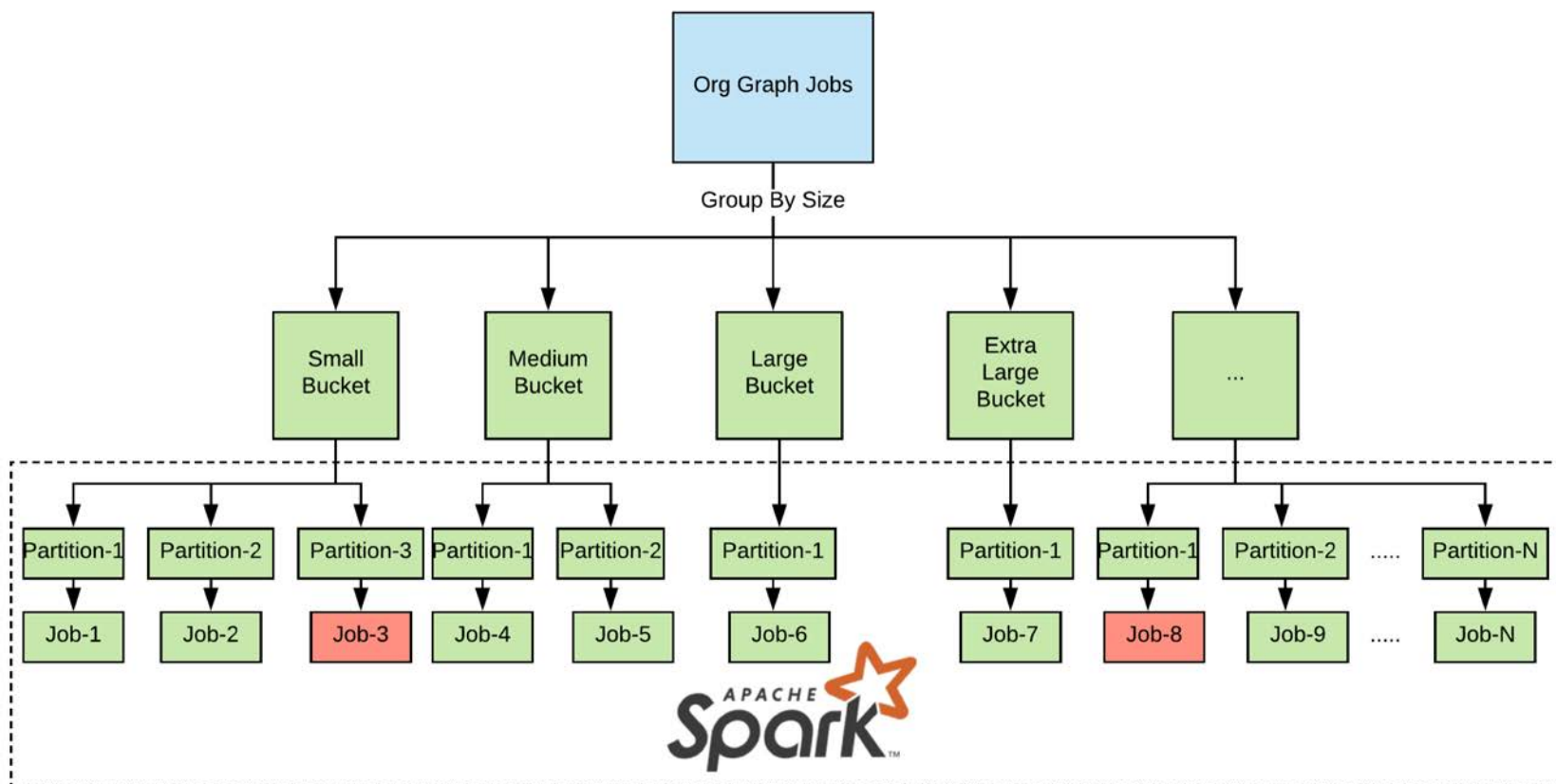
# Scale Up #Clusters

Hash partition org within each bucket

Spin up multiple spark clusters



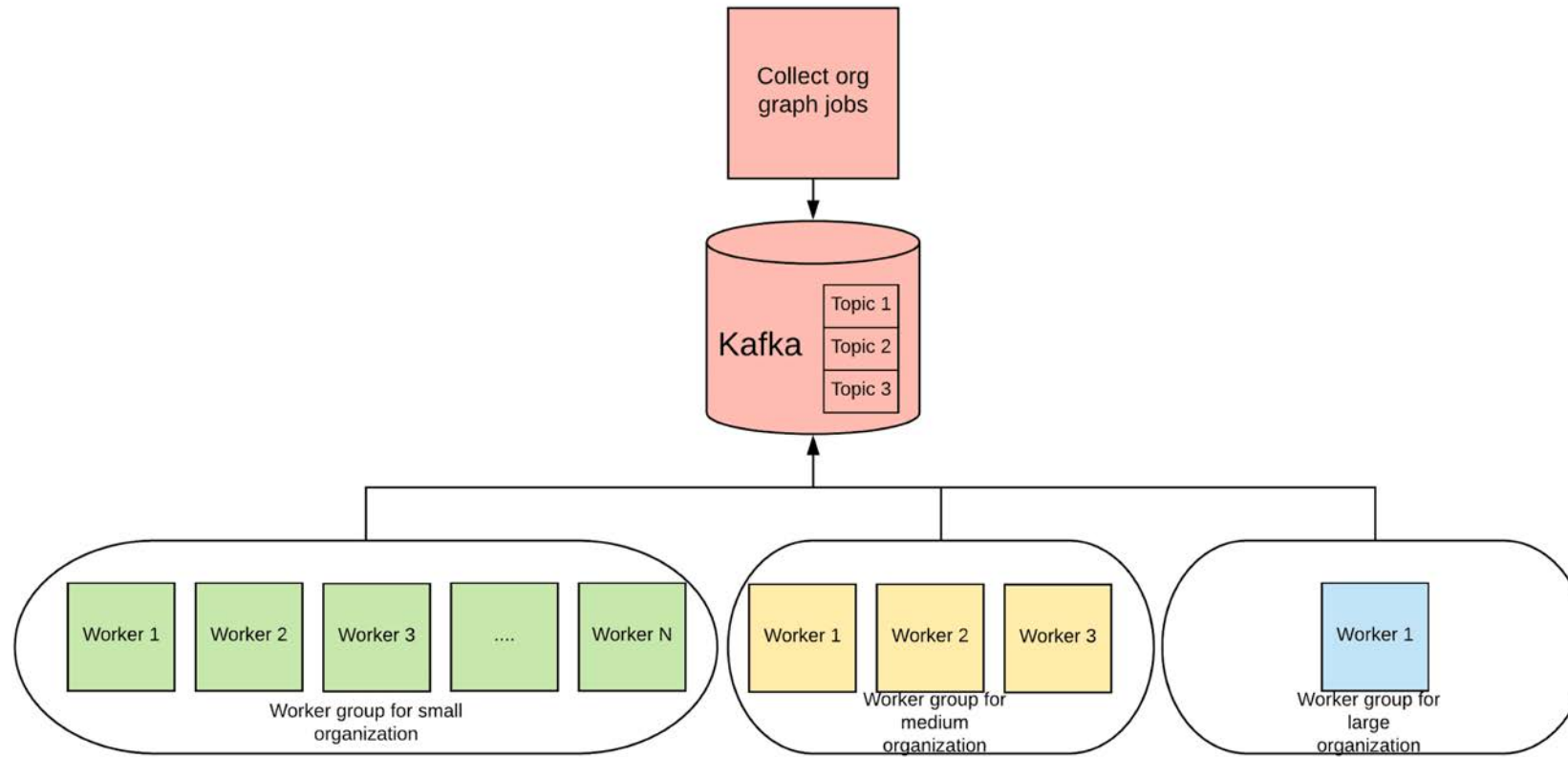
# “Hotspot” issue





# Solution

Create a request queue for each bucket

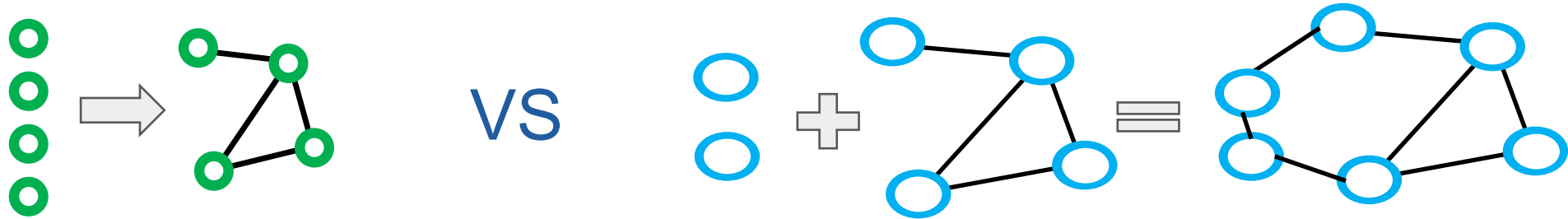


# Some Useful Tips

- Bucketing Strategy for variant data input
- Partition the orgs into small bins within each bucket
- Try scale up with multiple spark clusters
- Say no to tiny files and compact them to large chunk
- Use a simple queue with pulling module can balance the load

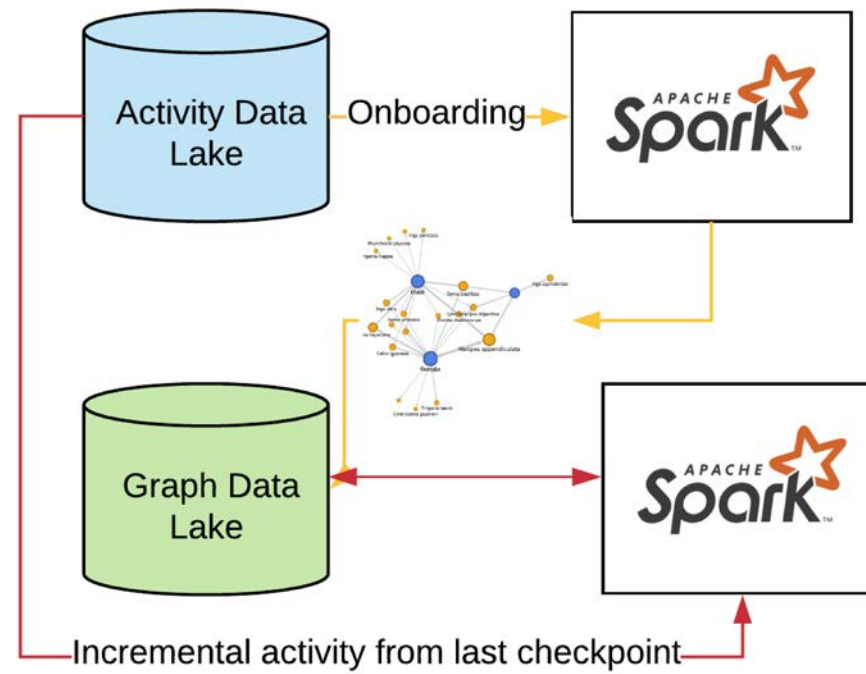


# Re-compute full graph or Incremental updates







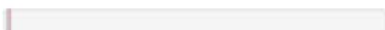
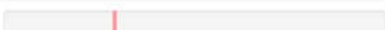


# Incremental update






- Save the intermediate graph data and checkpoint
- Incremental updating the contacts



# Failure happens

sparkJob		2018-08-20 10:36 55s	2018-08-22 19:00 40s	2d 8h 23m	Success	<a href="#">Details</a>
sparkJob		2018-08-20 10:36 55s	2018-08-22 17:52 59s	2d 7h 16m	Success	<a href="#">Details</a>
sparkJob		2018-08-20 10:36 55s	2018-08-23 08:06 36s	2d 21h 29m	Success	<a href="#">Details</a>
sparkJob		2018-08-20 10:36 55s	2018-08-22 21:07 41s	2d 10h 30m	Success	<a href="#">Details</a>
sparkJob		2018-08-20 10:36 55s	2018-08-23 02:28 17s	2d 15h 51m	Success	<a href="#">Details</a>
sparkJob		2018-08-20 10:36 55s	2018-08-22 15:36 59s	2d 5h 0m	Success	<a href="#">Details</a>
sparkJob		2018-08-20 10:42 32s	2018-08-20 10:43 04s	32 sec	Failed	<a href="#">Details</a>
sparkJob		2018-08-21 05:58 26s	2018-08-21 05:58 26s	0 sec	Cancelled	<a href="#">Details</a>

one failed job with many succeed jobs

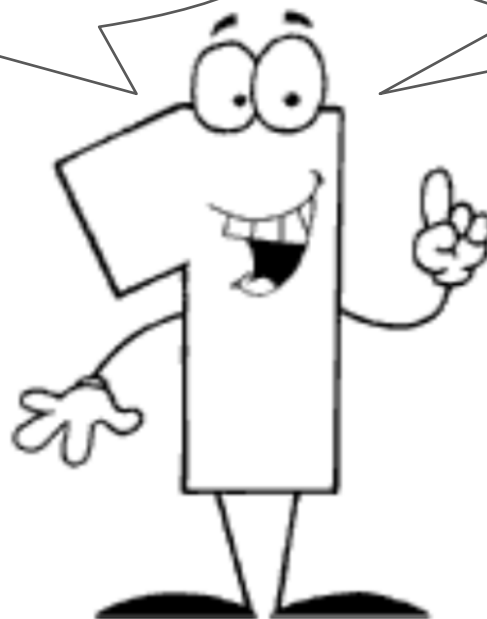
264		2018/10/02 17:27:35	85 ms	2/2 (5 skipped)	8/8 (20 skipped)
263		2018/10/02 17:27:33	2 s	3/3 (3 skipped)	12/12 (12 skipped)
262		2018/10/02 17:27:33	95 ms	2/2 (2 skipped)	8/8 (8 skipped)
261		2018/10/02 17:27:33	24 ms	1/1 (2 skipped)	4/4 (8 skipped)
260		2018/10/02 17:27:33	24 ms	1/1 (6 skipped)	3/3 (32 skipped)
259		2018/10/02 17:27:33	25 ms	1/1 (6 skipped)	4/4 (32 skipped)
258		2018/10/02 17:27:33	0.2 s	5/5 (2 skipped)	25/25 (8 skipped)
257		2018/10/02 17:27:33	18 ms	1/1	3/3
256		2018/10/02 17:27:33	14 ms	1/1	1/1

a lot of stages and jobs for graph generations

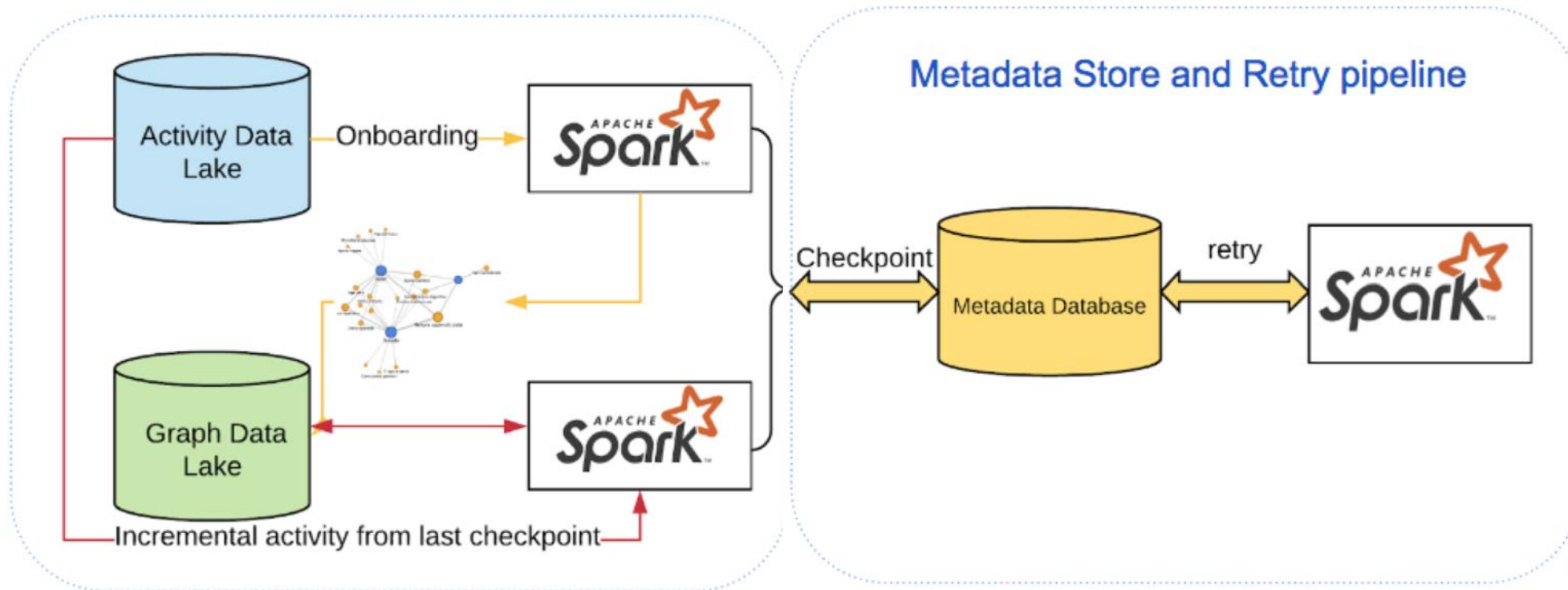


Failure  
Recover ?

Check State ?



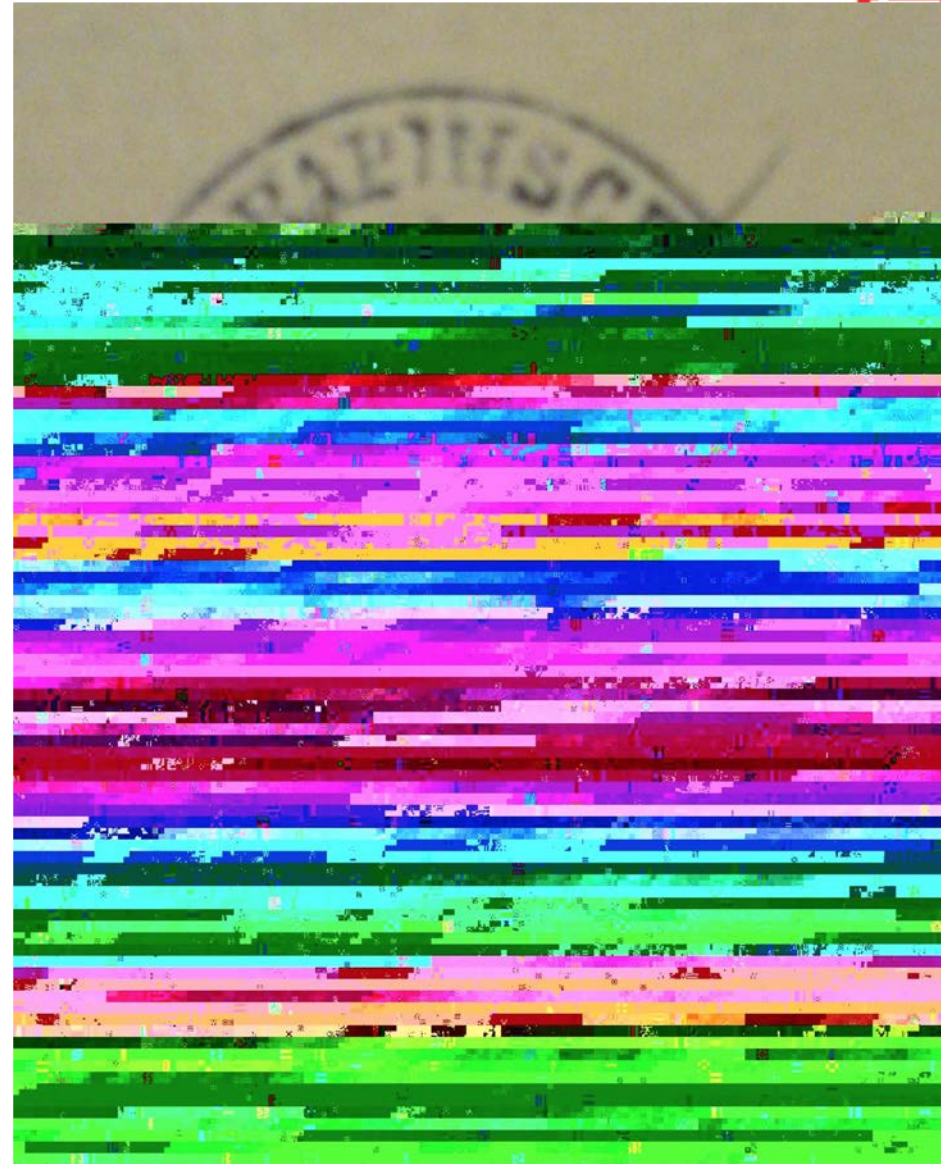
# Metadata Store



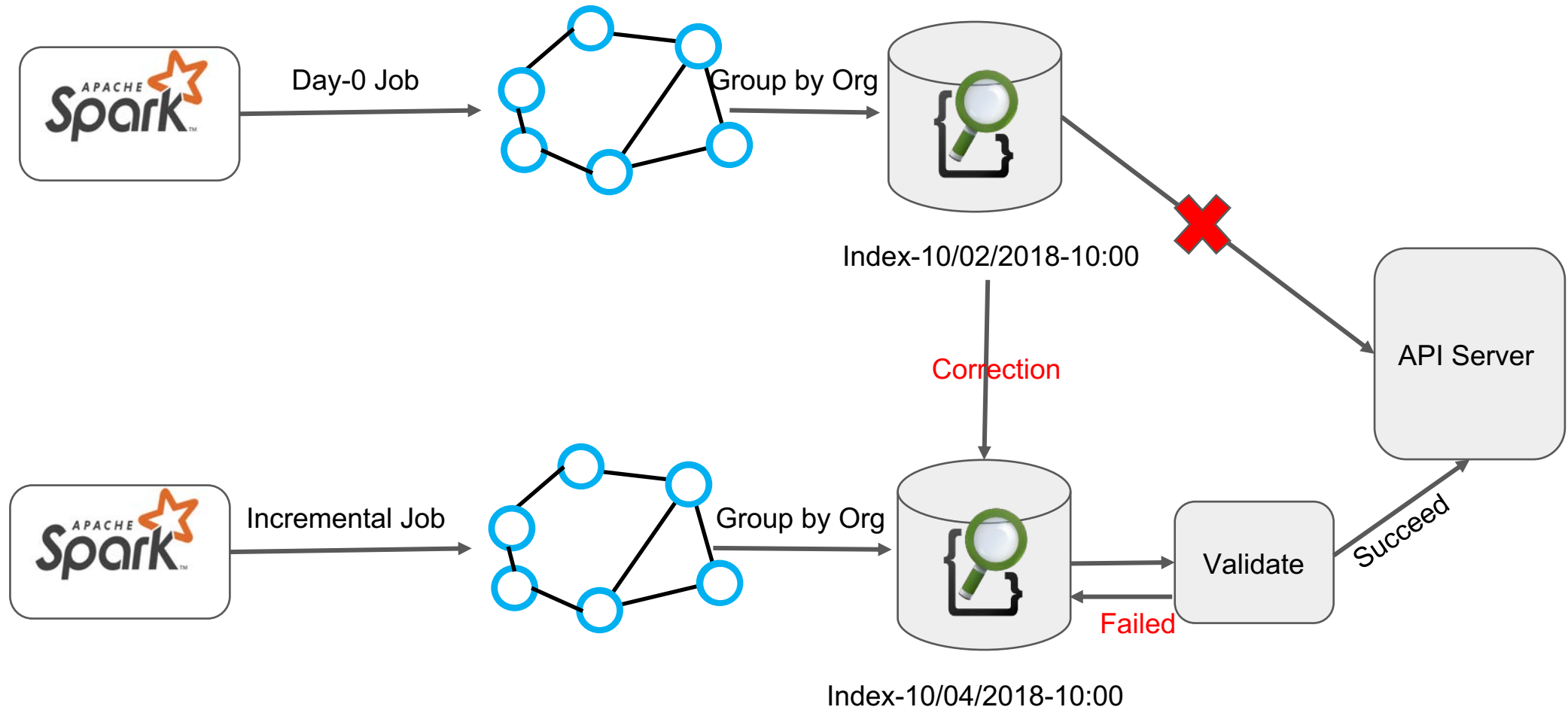


Indexing  
failures

Corruption  
Problems



# Index updates



# Some Lessons

- Use Incremental updates
- Create a metadata table for checkpoint and state store
- Create Indexes for each iteration of contact insights



# Agenda

- Introduction
- Activity Insights Context
- Why using a Graph to model context
- Key problems solved and lessons learned
- **Wrap up and QAs**



# Future Work

- Explore the graph database
- Explore the in-memory database Apache Ignite





# THANK YOU

[salesforce.com/careers](https://salesforce.com/careers)