



# Deal Personalization Systems @ Groupon

**GROUPON®**

Ameya Kanitkar  
[ameya@groupon.com](mailto:ameya@groupon.com)

ГРУПОН

Pilot / MVP

MVP

# Relevance & Personalization Systems @ Groupon





# Our Relevance Scenario

Users



**GROUPON** Featured Deal All Deals Getaways Goods Reserve Home-Garden Gifts Amaya K. ▾

All Deals New Deals Food & Drink Events & Activities Beauty & Spas Health & Fitness Home & Garden Automotive Shopping

Refer Friends. Get \$10\*

Q Pizza, massage, etc. Location: San Francisco, CA GO

### AstaYoga – Mission Dolores

\$25 for Five Yoga Classes (\$70 Value)

Value	Discount	You Save
\$70	64%	\$45

Give as a Gift Learn more

Limited time remaining!

Over 770 bought Limited quantity available

The deal is on!

Like 62 Send Mail Tweet Share

**In a Nutshell**  
The spacious, serene studio practices Ashtanga yoga and helps students achieve physical and spiritual balance.

**The Fine Print**  
Expires 180 days after purchase. Limit 1 per person, may buy 1 additional as a gift. Limit 1 per visit. Must activate by expiration date. Reservation required. Classes must be used by same person. First class must be redeemed in person. See the rules that apply to all deals.

**"Great Outdoors" Deals For You**

- \$14 for \$124 at Mount Shasta Hotel and Lodge, Mount Shasta, CA
- \$124 for \$258 at Big Bear Lake Mallard Bay Resort, Big Bear Lake, CA
- \$125 for \$236 at Hotel Madeline Telluride, Telluride, CO
- \$12 for \$24 at Outside Magazine Subscription Online Deal
- \$24.99 for \$45 at Portable Outdoor Lights Online Deal

**GROUPON** Featured Deal All Deals Getaways Goods Reserve Home-Garden Gifts Amaya K. ▾

All Deals New Deals Food & Drink Events & Activities Beauty & Spas Health & Fitness Home & Garden Automotive Shopping

Refer Friends. Get \$10\*

Q Pizza, massage, etc. Location: San Francisco, CA GO

### Citipets – Bayview

One or Three Days of Dog Daycare or One or Three Nights of Dog Boarding (Up to 51% Off)

Value	Discount	You Save
\$20	50%	\$10

Give as a Gift Learn more

Limited time remaining!

Over 50 bought Limited quantity available

The deal is on!

Like 2 Send Mail Tweet Share

**In a Nutshell**  
CPH and First Aid-certified staff care for pets, who have access to communal sleeping rooms and play in a huge indoor-outdoor facility.

**The Fine Print**  
Expires 180 days after purchase. Limit 1 per person, may buy 1 additional as a gift. Limit 1 per visit. Valid only for option purchased. Appointment required. 24hr cancellation notice required. Must sign waiver. Not valid on holiday weekends. May redeem across visits. Dogs must meet all playground requirements listed here. See the rules that apply to all deals.

**"Great Outdoors" Deals For You**

- \$14 for \$124 at Mount Shasta Hotel and Lodge, Mount Shasta, CA
- \$124 for \$258 at Big Bear Lake Mallard Bay Resort, Big Bear Lake, CA
- \$125 for \$236 at Hotel Madeline Telluride, Telluride, CO
- \$12 for \$24 at Outside Magazine Subscription Online Deal
- \$24.99 for \$45 at Portable Outdoor Lights Online Deal

Grow Your Business

**GROUPON** Featured Deal All Deals Getaways Goods Reserve Home-Garden Gifts Amaya K. ▾

All Deals New Deals Food & Drink Events & Activities Beauty & Spas Health & Fitness Home & Garden Automotive Shopping

Refer Friends. Get \$10\*

Q Pizza, massage, etc. Location: San Francisco, CA GO

### Regalito Rosticceria – Mission Dolores

Mexican Dinner for Two or Four with Appetizer, Entrees, and Dessert (Up to 47% Off)

Value	Discount	You Save
\$55.50	46%	\$25.50

Give as a Gift Learn more

Limited time remaining!

Over 800 bought Limited quantity available

The deal is on!

Like 23 Send Mail Tweet Share

**In a Nutshell**  
Chef in open kitchen roast free-range chicken, slow-roast citrus-marinated pork shoulder, and sauté North Carolina blue pears in vina.

**The Fine Print**  
Expires 180 days after purchase. Limit 1 per person, may buy 1 additional as a gift. Limit 1 per table. Valid only for option purchased. Reservation required. Valid only Monday-Thursday. Must use promotional value in 1 visit. See the rules that apply to all deals.

**"Great Outdoors" Deals For You**

- \$14 for \$124 at Mount Shasta Hotel and Lodge, Mount Shasta, CA
- \$124 for \$258 at Big Bear Lake Mallard Bay Resort, Big Bear Lake, CA
- \$125 for \$236 at Hotel Madeline Telluride, Telluride, CO
- \$12 for \$24 at Outside Magazine Subscription Online Deal
- \$24.99 for \$45 at Portable Outdoor Lights Online Deal

**GROUPON** Featured Deal All Deals Getaways Goods Reserve Home-Garden Gifts Amaya K. ▾

All Deals New Deals Food & Drink Events & Activities Beauty & Spas Health & Fitness Home & Garden Automotive Shopping

Refer Friends. Get \$10\*

Q Pizza, massage, etc. Location: San Francisco, CA GO

### K1 Speed – Multiple Locations

\$44 for a Racing Package with Four Races and Two Yearly Licenses (Up to \$91.96 Value)

Value	Discount	You Save
\$91.96	52%	\$47.96

Give as a Gift Learn more

Time left to buy 6 days 4:44:21

Over 610 bought Limited quantity available

The deal is on!

Like 52 Send Mail Tweet Share

**In a Nutshell**  
Eco-friendly electric karts, speeds up to 45 mph, suits for kids and adults, collection of racing memorabilia on display.

**The Fine Print**  
Expires Dec 4, 2013. Limit 2 per person, may buy multiple as gifts. Must activate licenses by expiration date, licenses expire 1 year from activation date. All junior racers must be at least 48" and adult racers must be at least 4'10" tall to race. All racers should be under 7' tall and 300 lbs. Subject to availability. Not valid for group events. May split Groupon.

**"Great Outdoors" Deals For You**

- \$14 for \$124 at Mount Shasta Hotel and Lodge, Mount Shasta, CA
- \$124 for \$258 at Big Bear Lake Mallard Bay Resort, Big Bear Lake, CA
- \$125 for \$236 at Hotel Madeline Telluride, Telluride, CO
- \$12 for \$24 at Outside Magazine Subscription Online Deal
- \$24.99 for \$45 at Portable Outdoor Lights Online Deal



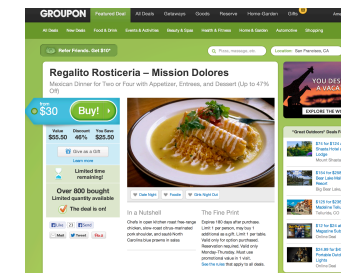
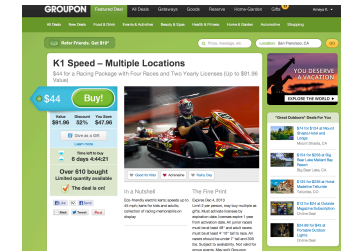
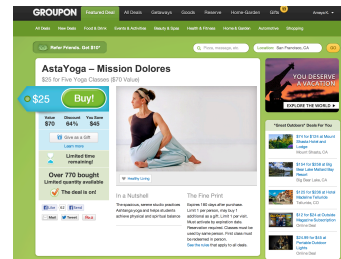
# Our Relevance Scenario

Users

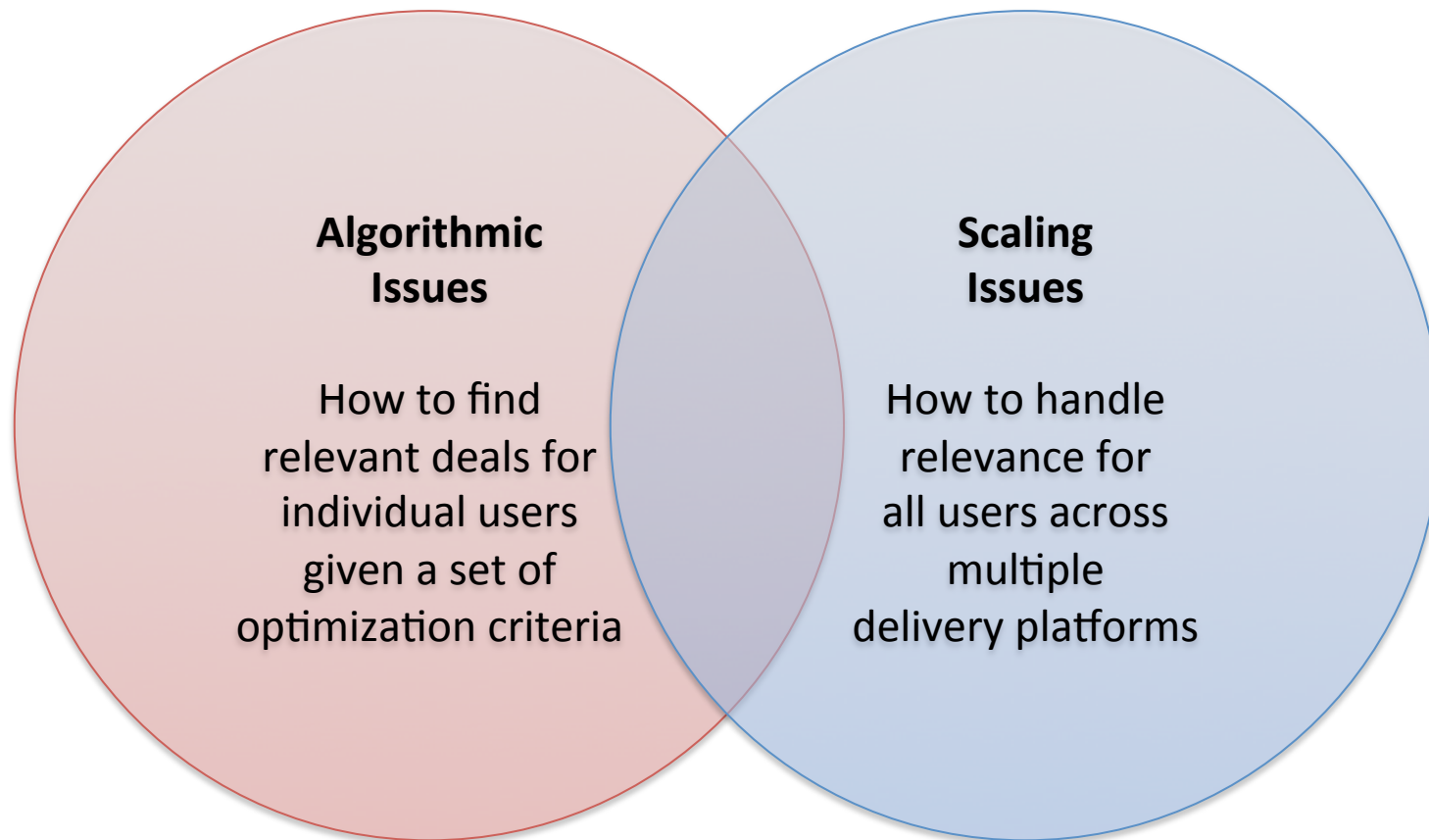


How do we surface relevant deals ?

- Deals are perishable (Deals expire or are sold out)
- No direct user intent (As in traditional search advertising)
- Relatively Limited User Information
- Deals are highly local



# Two Sides to the Relevance Problem



# Developing Deal Ranking Algorithms

- **Exploring Data**

- ➡ Understanding signals, finding patterns

- **Building Models/Heuristics**

- ➡ Employ both classical machine learning techniques and heuristic adjustments to estimate user purchasing behavior

- **Conduct Experiments**

- ➡ Try out ideas on real users and evaluate their effect

**T-Bones Marketplace – Marshall**  
\$15 for \$30 Worth of Meats, Produce, and Prepared Foods

**\$15** Sold Out

Value	Discount	You Save
\$30	50%	\$15

Buy it for a friend!

This deal sold out at:  
3:15PM  
05/15/2011

**SOLD OUT**  
Over 5,000 bought

Share:

**Highlights**

- Independent boutique grocery store
- Signature prepared foods
- Daily lunch & dinner specials

**The Fine Print**

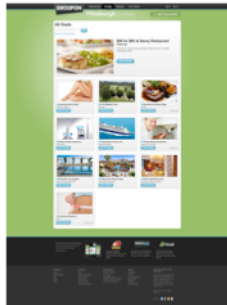
Expires Nov 16, 2011  
Limit 1 per person, may buy multiple as gifts. Limit 1 per visit. Not valid for catering.  
[See the rules](#) that apply to all deals.



# Data Infrastructure

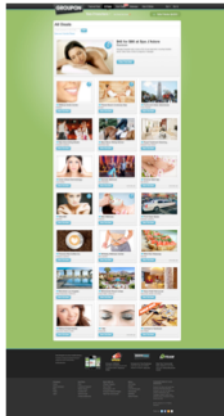
## Growing Deals

2011



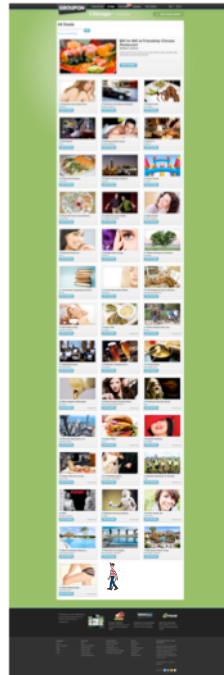
20+

2012



400+

2013



2000+

## Growing Users

- 100 Million+ subscribers
- We need to store data like, user click history, email records, service logs etc. This tunes to billions of data points and TB's of data

# Deal Personalization Infrastructure Use Cases

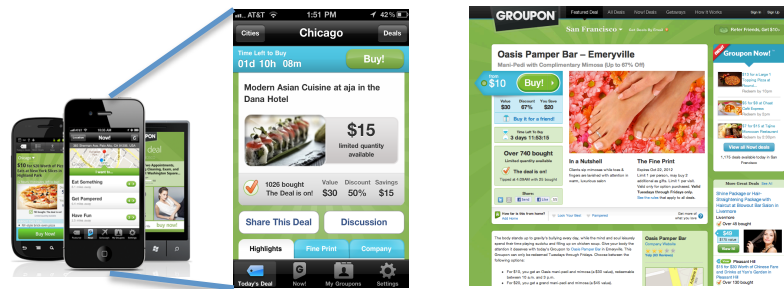
## Deliver Personalized Emails



Personalize billions of emails for hundreds of millions of users

**Offline System**

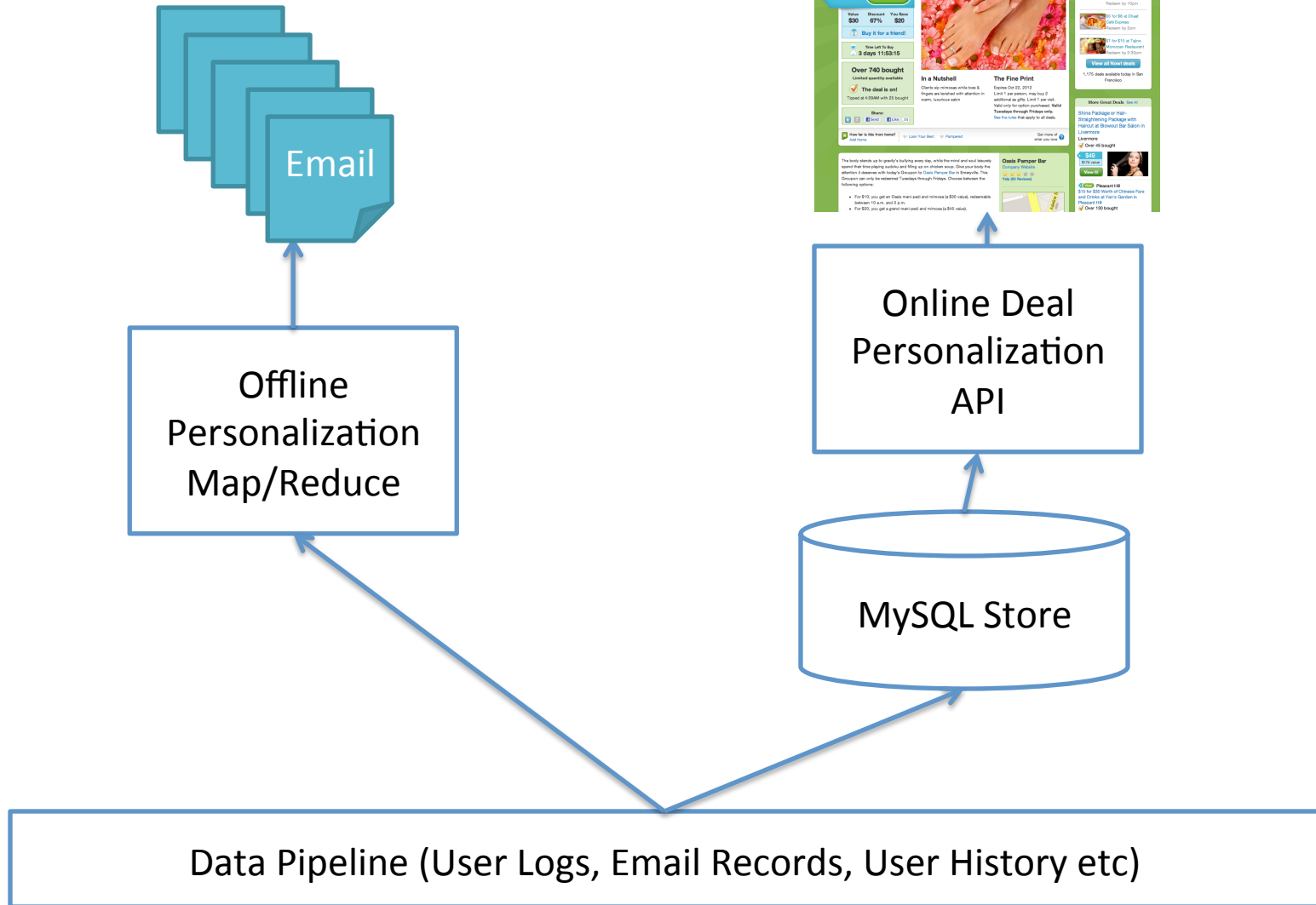
## Deliver Personalized Website & Mobile Experience



Personalize one of the most popular e-commerce mobile & web app for hundreds of millions of users & page views

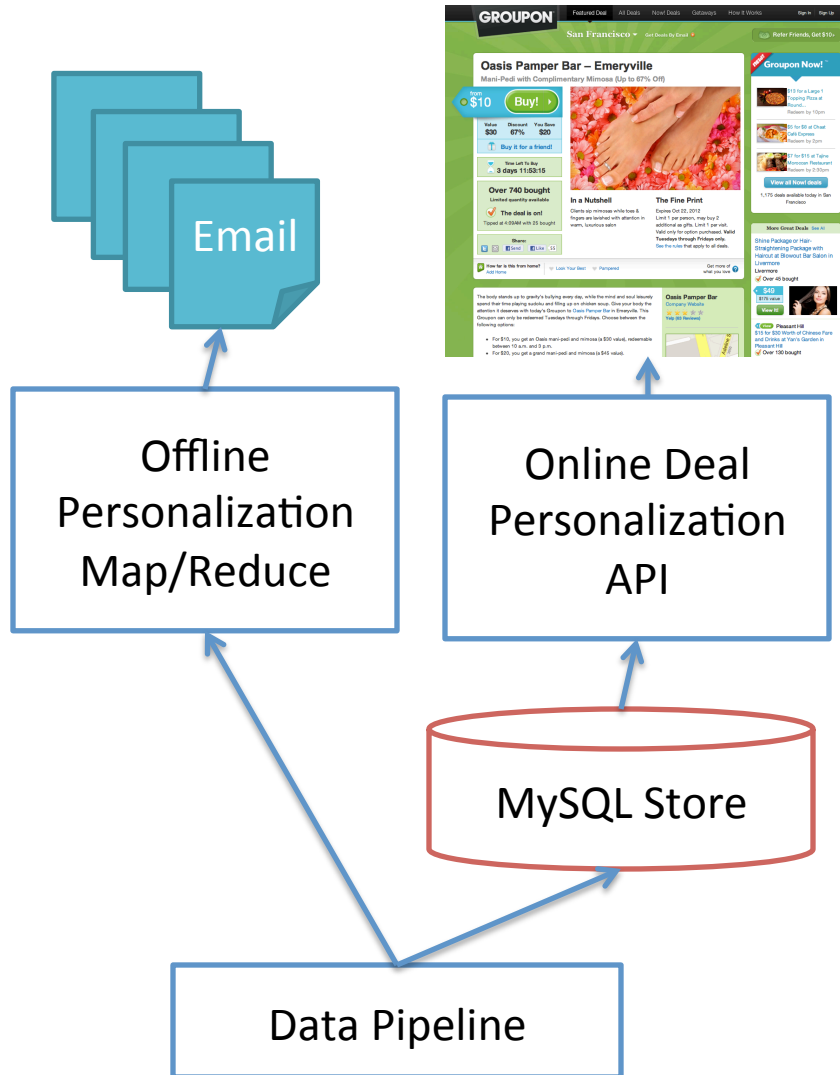
**Online System**

# Earlier System



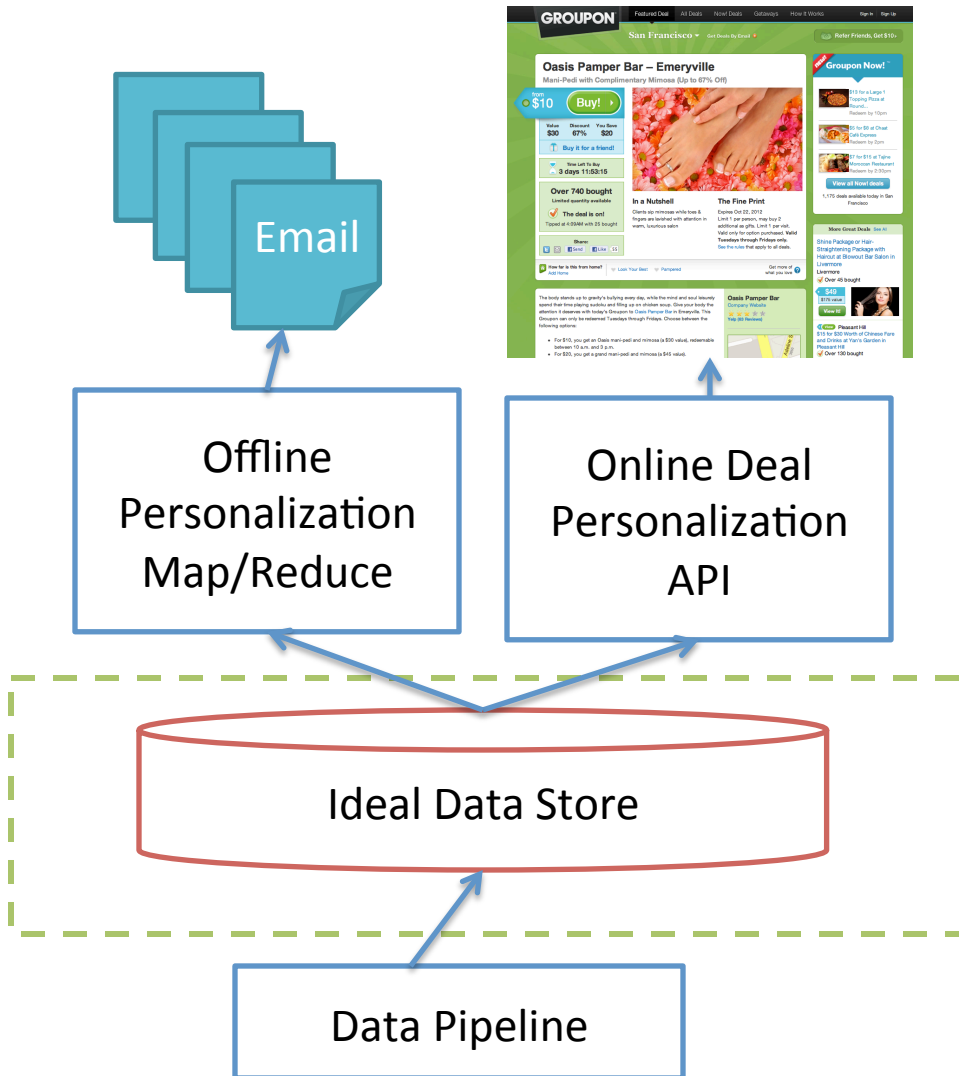


# Earlier System



- Scaling MySQL for data such as user click history, email records was painful unless we shard data
- Need to maintain two separate data pipelines for essentially the same data.

# Ideal System



- Common data store that serves data to both online and offline systems
- Data store that scales to hundreds of millions of records
- Data store that plays well with our existing hadoop based systems
- Data store that supports get() put() access patterns based on a key (User ID).

# Why HBase?

- Open Source distributed map data store modeled after Google's Big Table
- Distributed Data Store: Store data on 1-700 node cluster. Linear scaling. Add capacity by adding more machines.
- Very light schema. Each row may have any number of columns. Columns need not be defined upfront.  
(Something like: Row1-> Map<byte[], byte[])



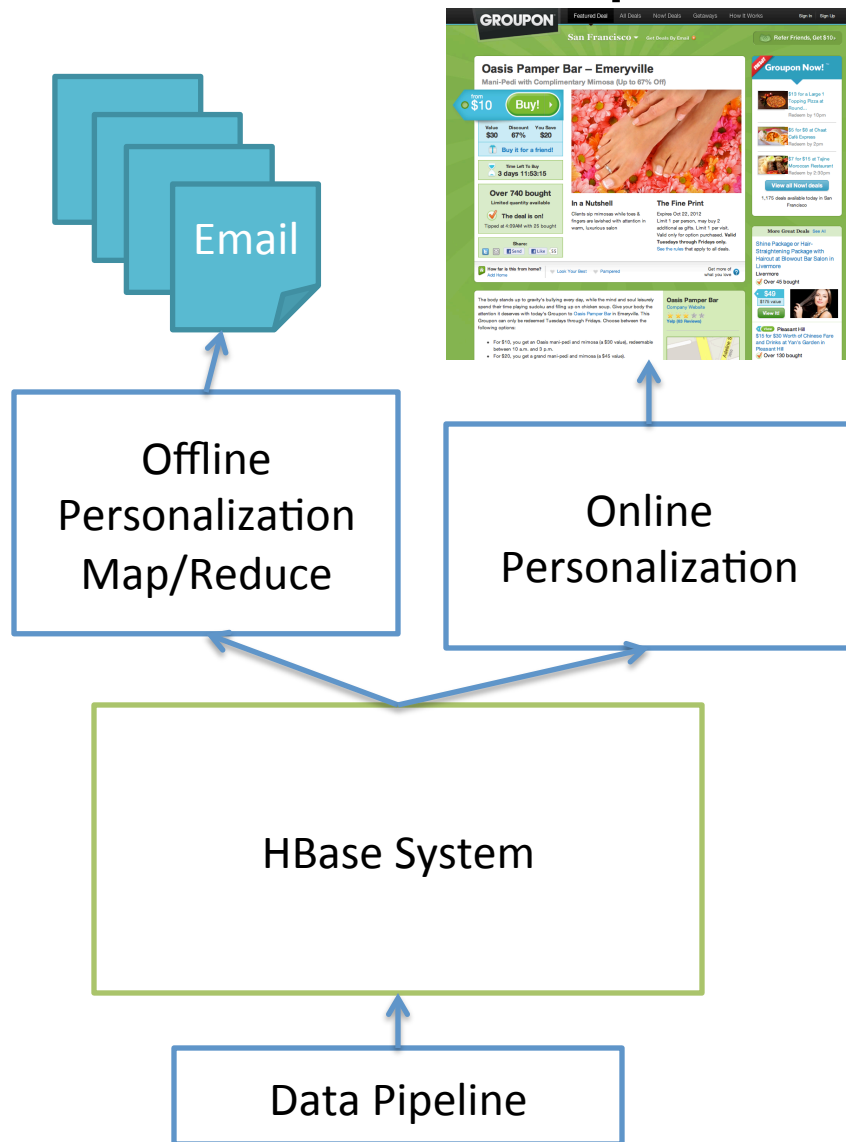
# Why HBase?

- Consistent Database. Highly available. Automatically shards/ scales. Can scale to billions of rows and multi terabyte data sizes
- Writes : 1-10 ms, Reads 20-50 ms
- Tight out of the box integration with Hadoop and Map Reduce

# HBase Table

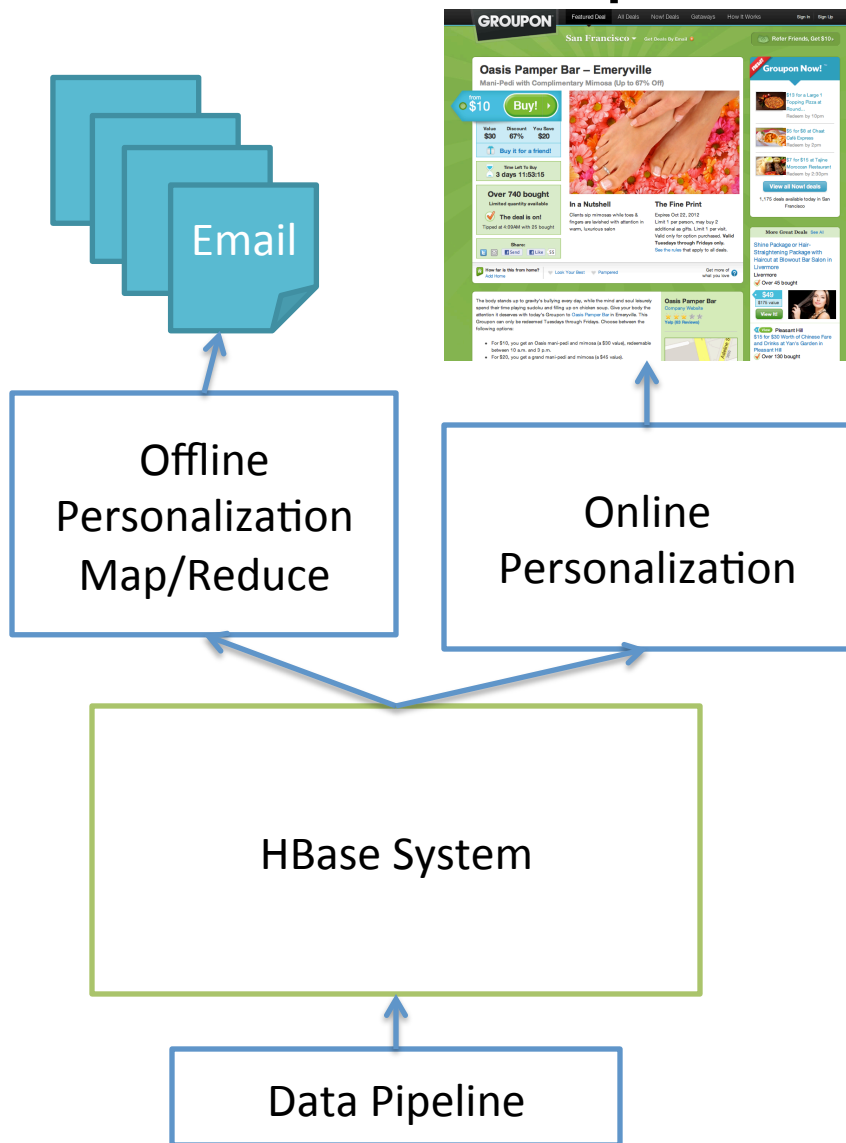
Row	Cf:<qual>	Cf:<qual>	....	Cf:<qual>
row1	Cf1:qual1	Cf1:qual2		
row11	Cf1:qual2	Cf1:qual22	Cf1:qual3	
row2	Cf2:qual1			
rowN				

# Architecture Options





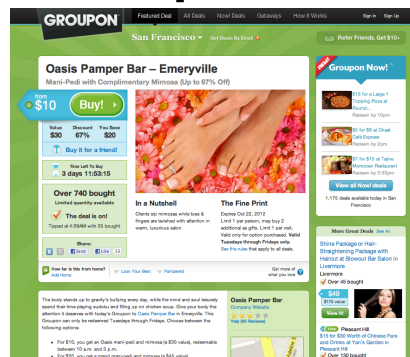
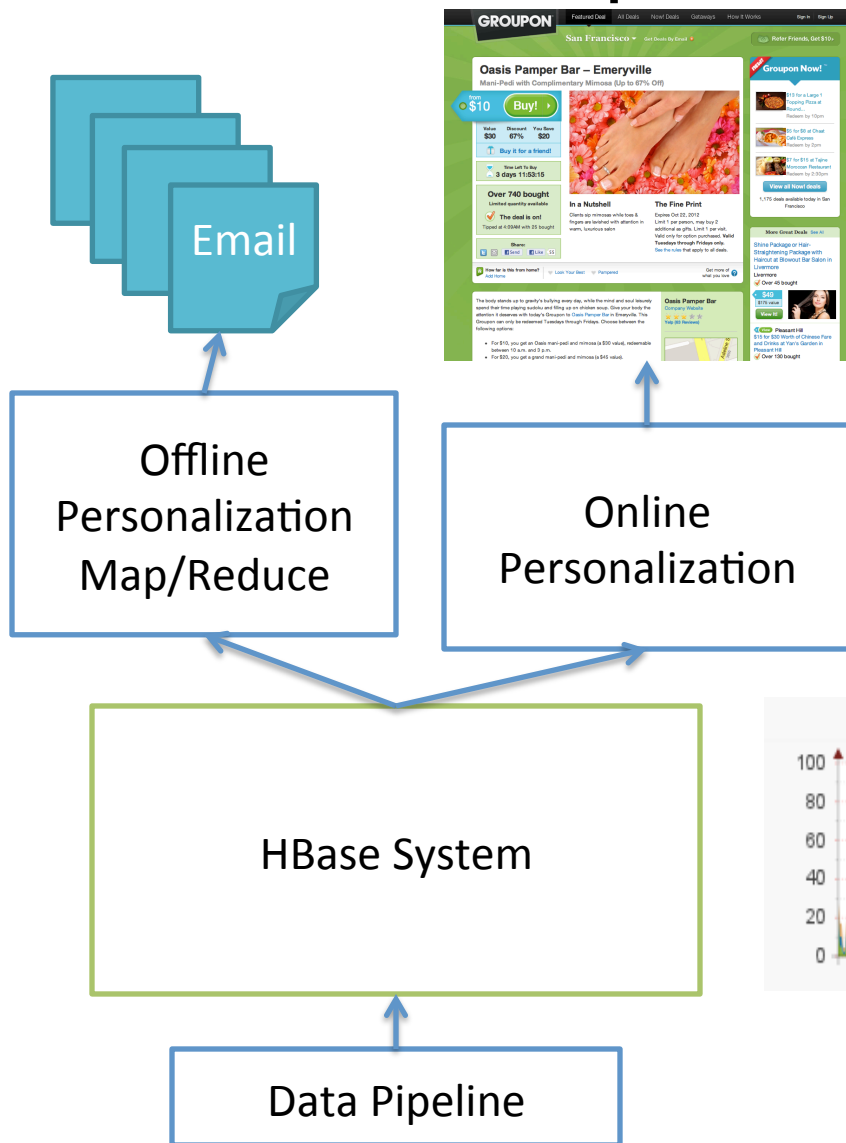
# Architecture Options



## Pros

- Simple design
- Consolidated system that serves both online and offline personalization

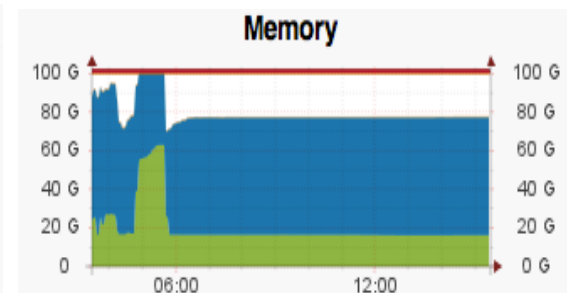
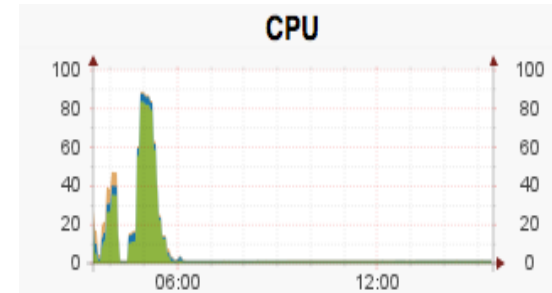
# Architecture Options



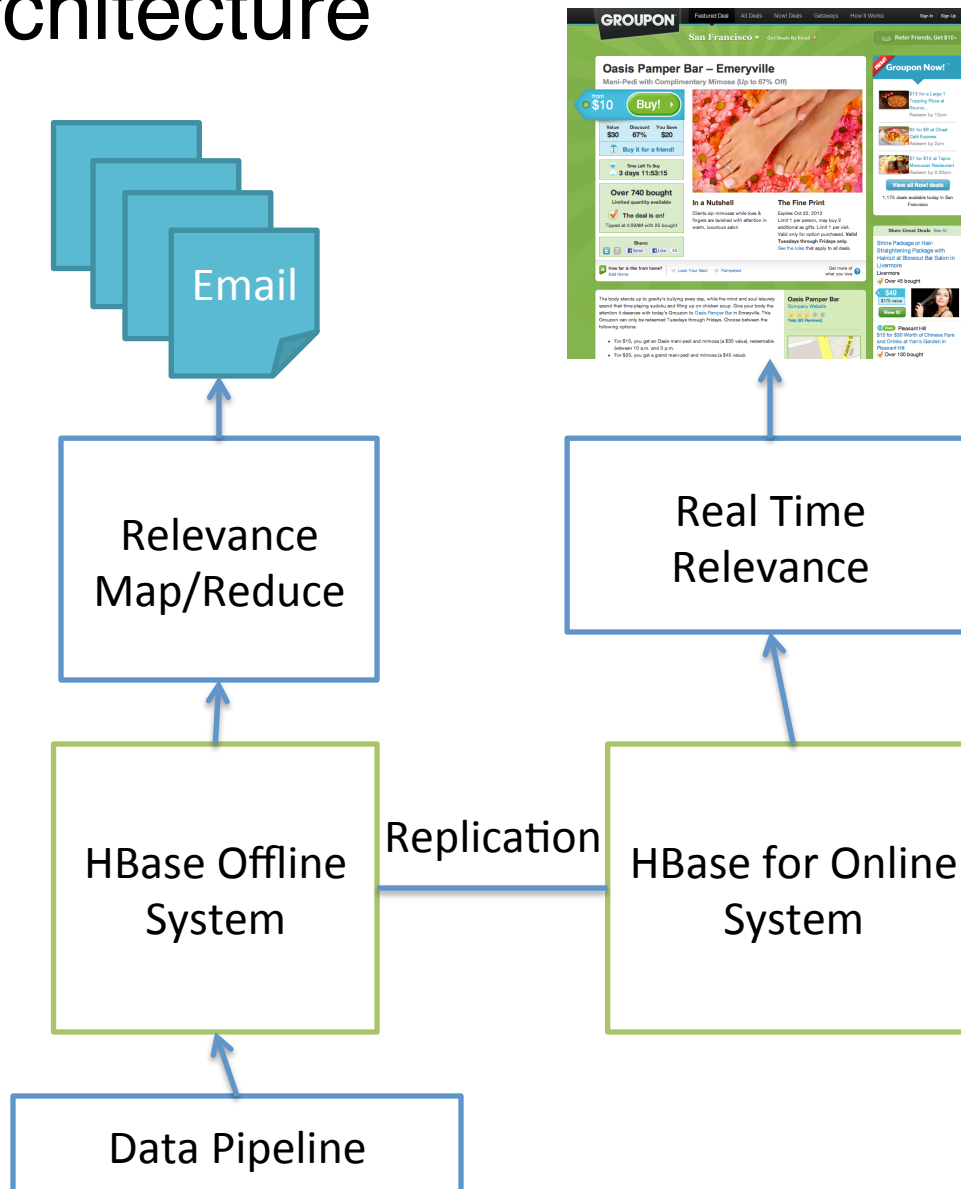
## Cons

- We now have same uptime SLA on both offline and online system
- Maintaining online latency SLA for bulk writes and bulk reads is hard.

And here is why...

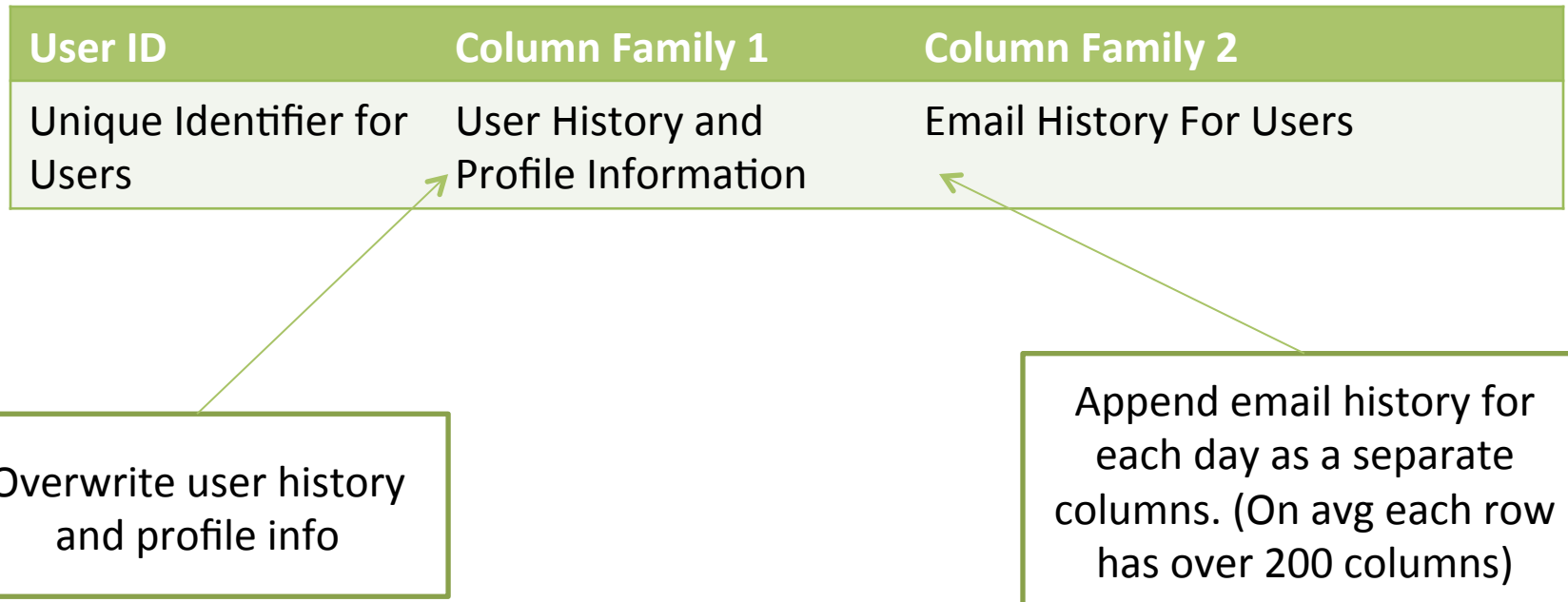


# Architecture



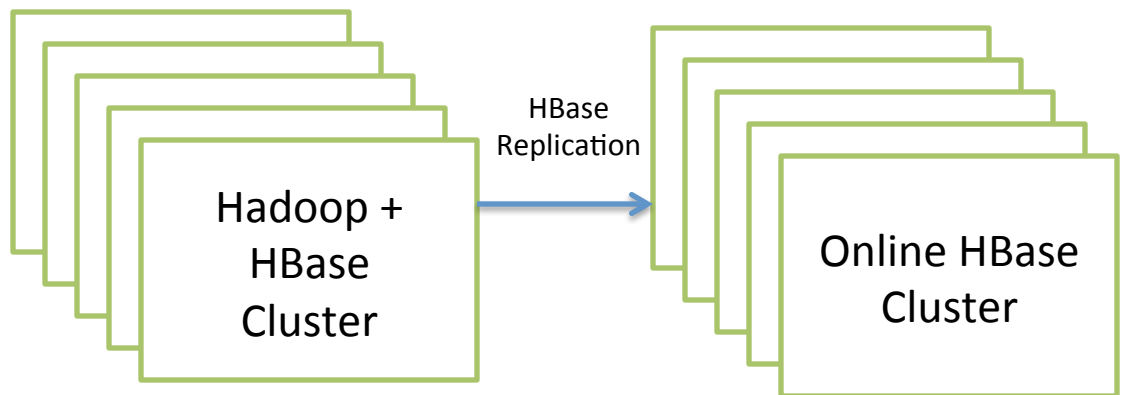
- We can now maintain different SLA on online and offline systems
- We can tune HBase cluster differently for online and offline systems

# HBase Schema Design



- Most of our data access patterns are via “User Key”
- This makes it easy to design HBase schema
- The actual data is kept in JSON

# Cluster Sizing



100+ machine Hadoop cluster, this runs heavy map reduce jobs  
The same cluster also hosts 15 node HBase cluster

10 Machine dedicated HBase cluster to serve real time SLA

- **Machine Profile**
- **96 GB RAM (HBase 25 GB)**
- **24 Virtual Cores CPU**
- **8 2TB Disks**
- **Data Profile**
- **100 Million+ Records**
- **2TB+ Data**
- **Over 4.2 Billion Data Points**

## Other Takeaways

- Choose data storage format carefully. (We are using JSON, but one can consider Avro, Protobufs etc)
- Always store compressed data. We use LZO, its easy to map reduce
- Always store processed data in HBase.
- HBase needs some tuning before it scales. Tuning garbage collection is important. So is various timeouts and caching parameters, cluster can be unstable without these tuning parameters.



# Questions?



Questions?

**GROUPON®**

Thanks!

[ameya@groupon.com](mailto:ameya@groupon.com)

[www.groupon.com/techjobs](http://www.groupon.com/techjobs)