



# Apache Kafka at LinkedIn

Guozhang Wang  
BDTC 2014

# About Me



## Guozhang Wang

Senior Software Engineer at LinkedIn

San Francisco Bay Area | Computer Software

Current The Apache Software Foundation, LinkedIn

Previous Cornell University, LinkedIn, Microsoft

Education Cornell University

View profile as ▼

**445**  
connections

 [www.linkedin.com/in/guozhangwang/](https://www.linkedin.com/in/guozhangwang/)

 Contact Info

# Agenda

- *Overview of Kafka*
- *Kafka Design*
- *Kafka Usage at LinkedIn*
- *Roadmap*
- *Q & A*



# *Why We Build Kafka?*

# We Have a lot of Data


- *User activity tracking*
  - *Page views, ad impressions, etc*
- *Server logs and metrics*
  - *Syslogs, request-rates, etc*
- *Messaging*
  - *Emails, news feeds, etc*
- *Computation derived*
  - *Results of Hadoop / data warehousing, etc*





# .. and We Build Products on Data



# Newsfeed




Share an update... 


Share with: **Public** 

Share

All Updates ▾  
Hide




**Will Khoo**  
Very cool... i can use this to work.. can i get a HOV sticker?





**Electric Car Folds in Half for Crowded City Streets**  
mashable.com • Made in South Korea, the Armadillo-T is a two-seater electric car that folds up after the driver and passenger leave the vehicle.

Like • Comment • Share • 42s ago



**Shikha Mathur's** skills and expertise were endorsed by **Raymond Reeves**.  
Shikha was endorsed for Recruiting.  
Endorse your connections • 1m ago



**Will Khoo**  


**Amid Uproar Over NSA, Malaysia Wants to Expand Its Spying Powers**  
mashable.com • The spying scandal roiling the National Security Agency in the U.S. may turn out to provide great air





# Recommendation

## People You May Know <sup>beta</sup>

See people from different parts of your professional life



All Suggestions / LinkedIn (12)

[Connect All](#)



**Brad Mauney** <sup>2nd</sup>  
Senior Product Manager, Search & Social Graph at LinkedIn  
Mountain View, California

[Connect](#)  5 shared connections



**Albert Wang** <sup>2nd</sup>  
Senior User Experience Designer at LinkedIn  
Mountain View, California

[Connect](#)  127 shared connections



**Sam Shah** <sup>2nd</sup>  
Principal Engineer at LinkedIn  
Mountain View, California

[Connect](#)  22 shared connections



**Tan Nhu** <sup>2nd</sup>  
Senior Web Developer at LinkedIn  
Mountain View, California

[Connect](#)  16 shared connections



**Vinodh Jayaram** <sup>2nd</sup>  
Software Engineering Manager at LinkedIn  
Mountain View, California

[Connect](#)  10 shared connections




**Andy Chen** <sup>2nd</sup>  
Software Engineer at LinkedIn  
Mountain View, California

[Connect](#)  78 shared connections




# Recommendation

Now it's your turn. Endorse your connections:




Does **Zhaoyin Jia** know about **Algorithms**?

Endorse




Does **Yi Shen** know about **Computer Science**?

Endorse



Does **Zhihua Geng** know about **Analytics**?

Endorse



Does **Bailu Ding** know about **Distributed Systems**?

Endorse

View more

Close

# Search

4,002 results for **software engineer linkedin**

Sort by **Relevance** ▾

- 

**Senior Software Quality Assurance Engineer**  
Epicor Software  
Newburgh, NY · Dec 9, 2013  
▶ 2 people in your network · Similar

Save Job ▾
- 

**Software Engineer in Test**  
LinkedIn  
New York · Dec 7, 2013  
▶ 1,995 people in your network · Similar

Save Job ▾
- 

**Principal Software Engineer**  
Symantec  
USA - California, Mountain View - HQ · Nov 25, 2013  
▶ 2 connections to the poster · Similar

Save Job ▾
- 

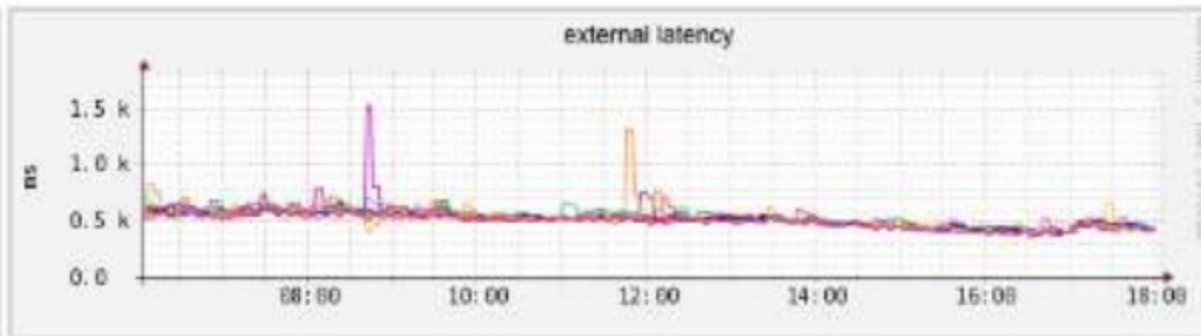
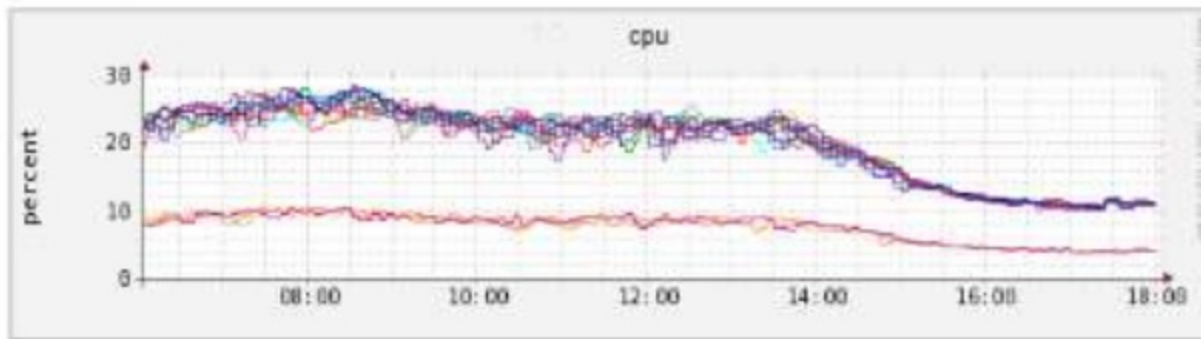
**Test Engineer OR Software Development Engineer in Test (SDET)**  
Andiamo Partners  
Newark, NJ · Dec 10, 2013  
▶ 3 people in your network · Similar

Save Job ▾
- 

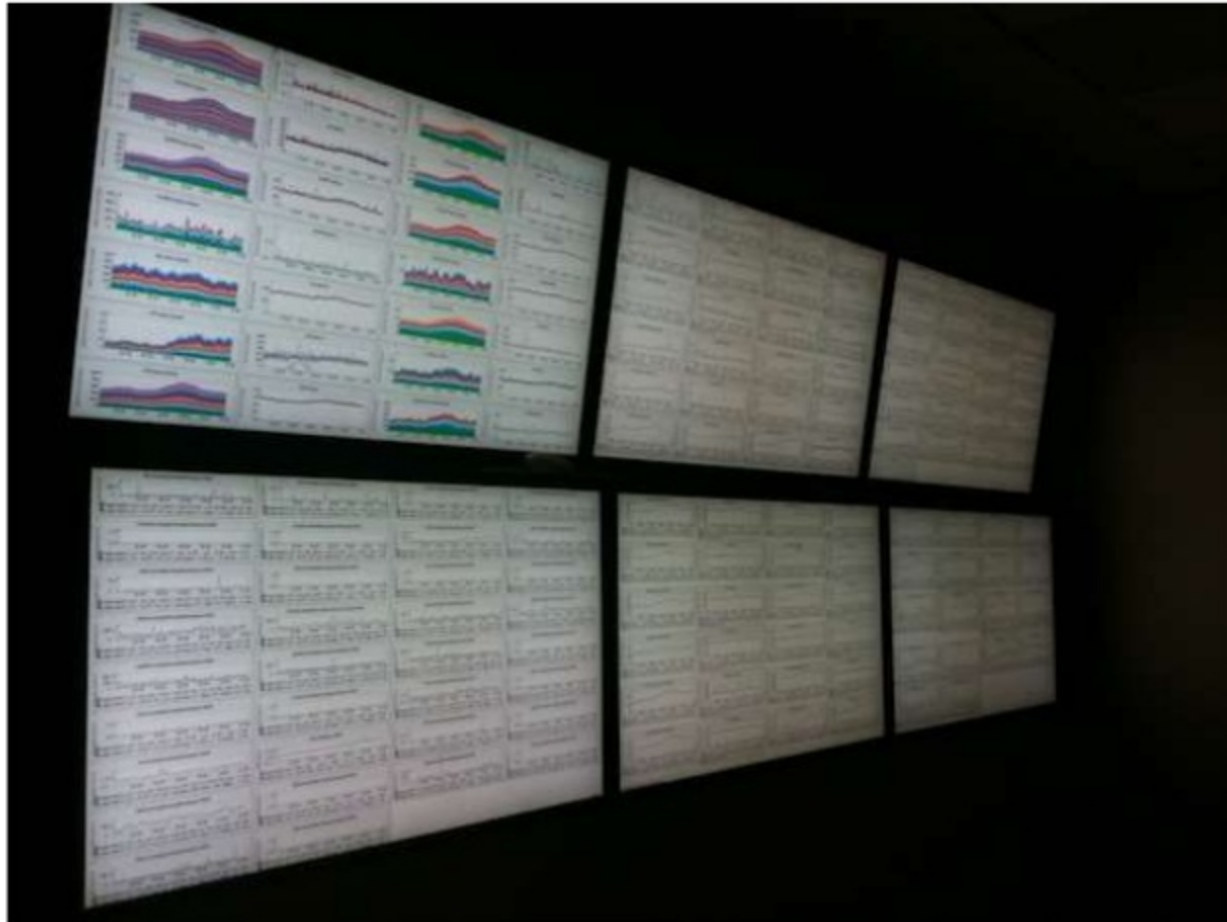
**Senior Software Developer - C#**  
Epicor Software  
Newburgh, NY · Dec 9, 2013  
▶ 2 people in your network · Similar

Save Job ▾

# Metrics and Monitoring



# .. and a LOT of Monitoring



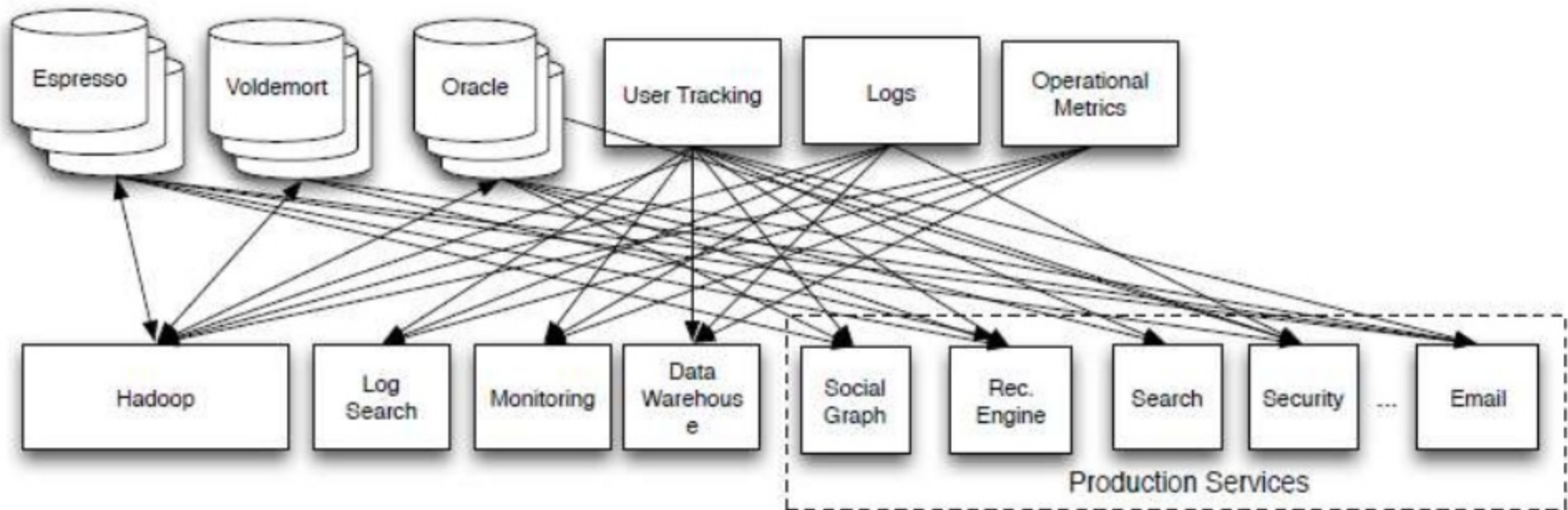
# *The Problem:*

*How to integrate this variety of data  
and make it available to all products?*

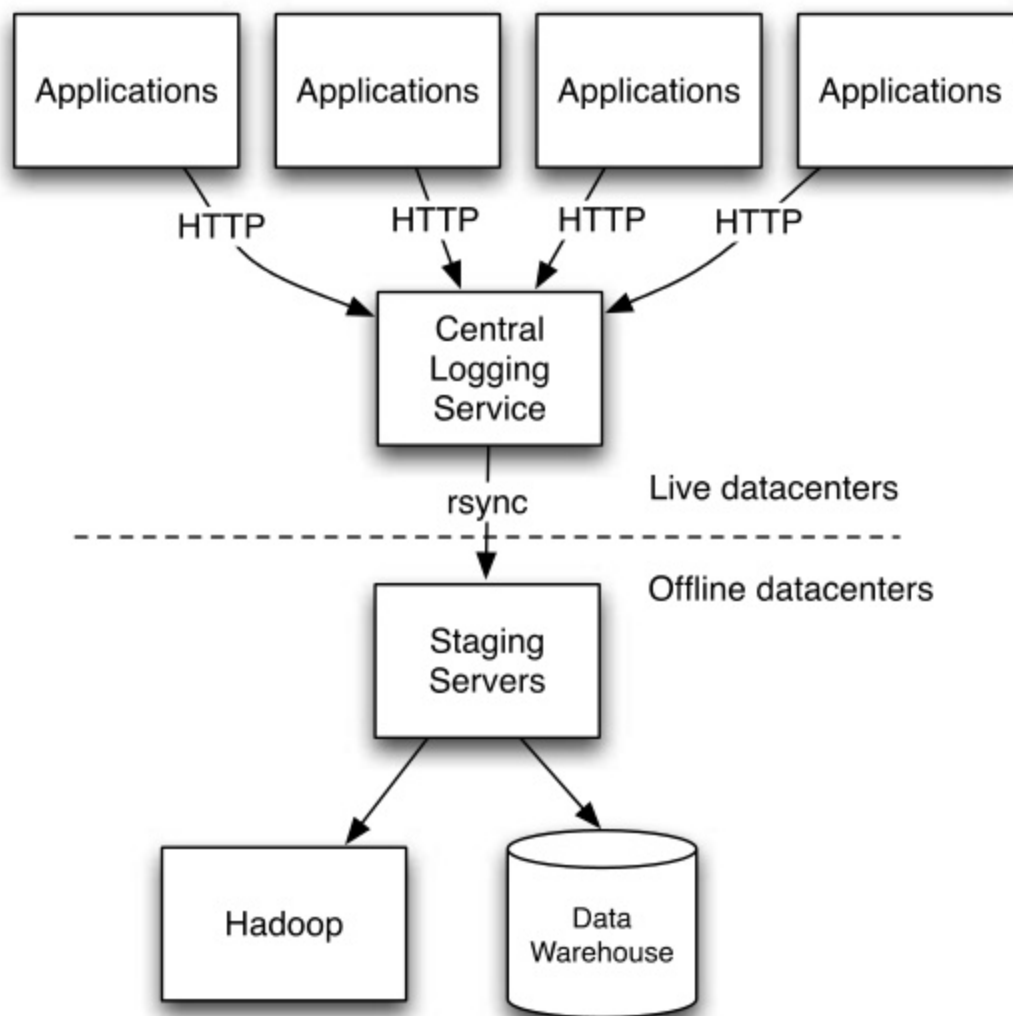


# Life back in 2010:

## Point-to-Point Pipelines



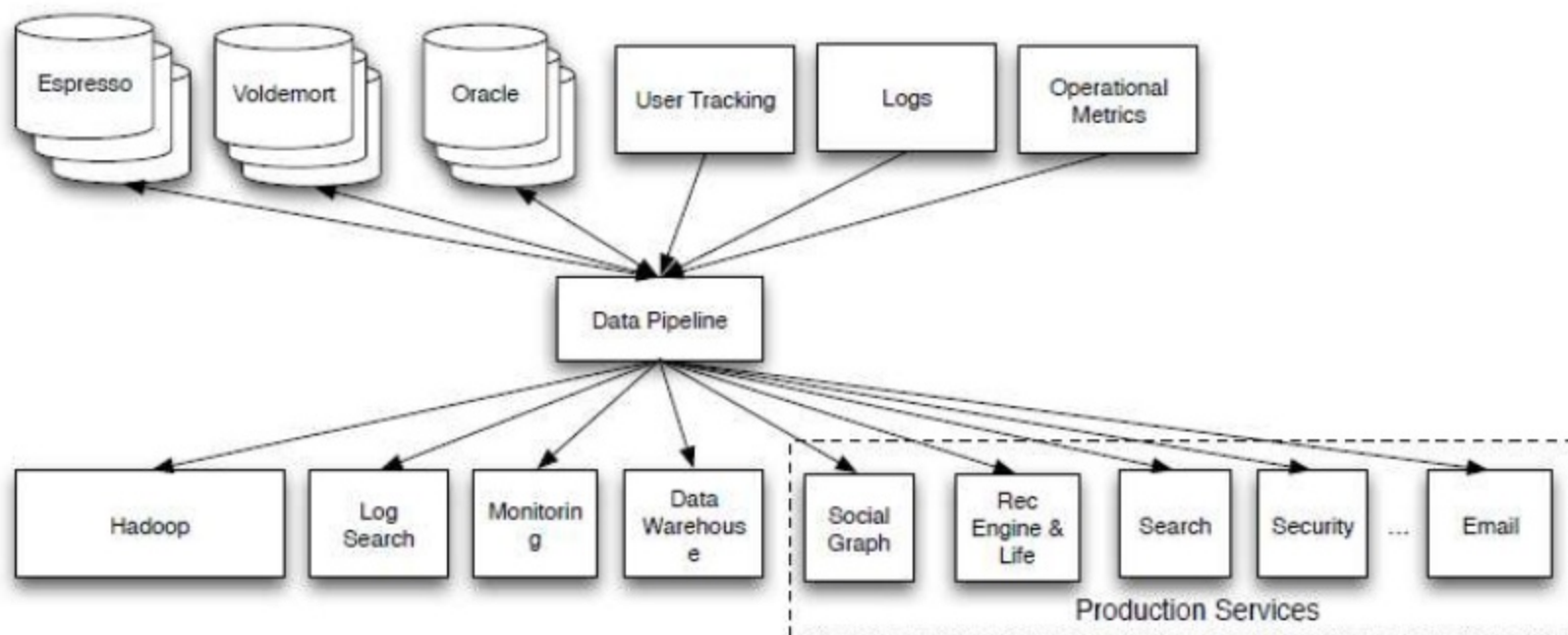
# Example: User Activity Data Flow

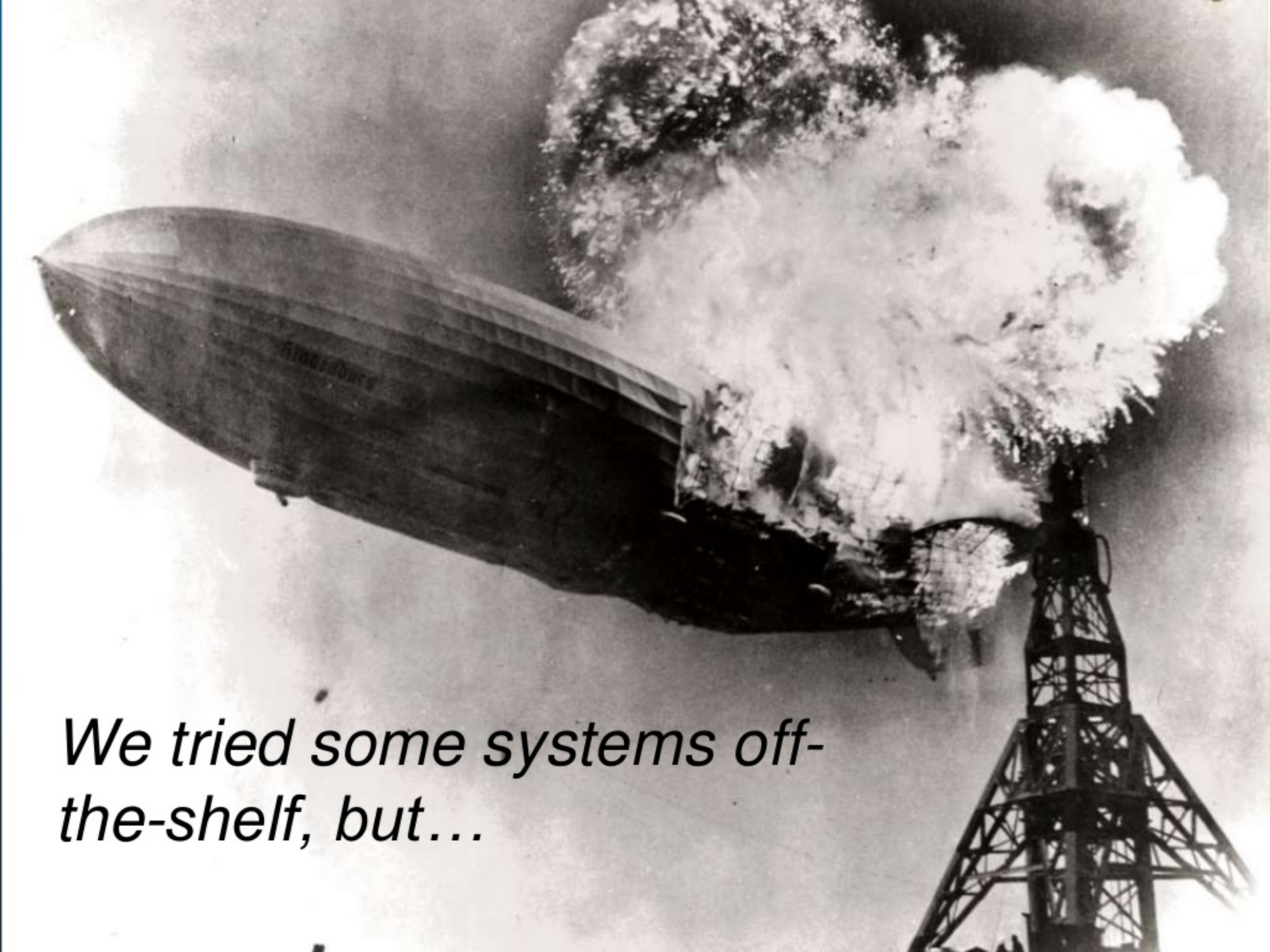




# What We Want

- A *centralized* data pipeline





*We tried some systems off-the-shelf, but...*

# What We *REALLY* Want

- A ***centralized*** data pipeline that is
  - Elastically scalable
  - Durable
  - High-throughput
  - Easy to use

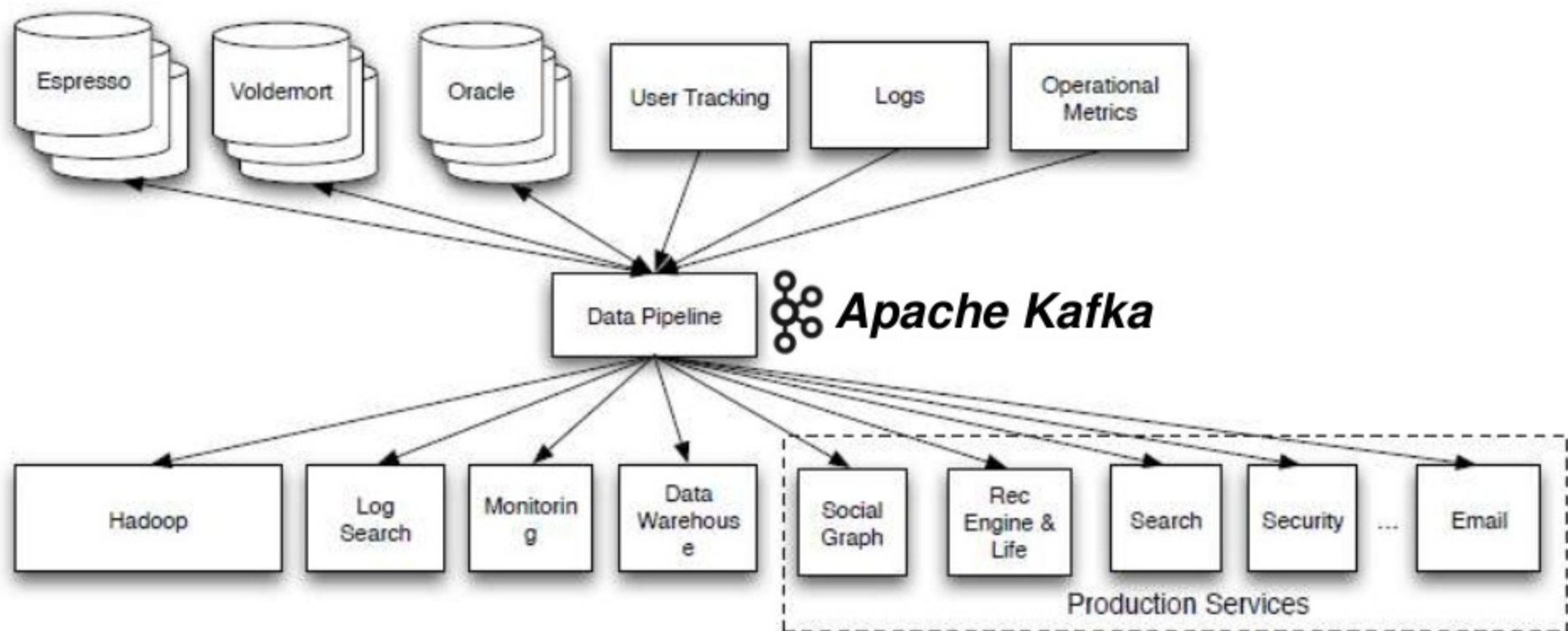




- A distributed pub-sub messaging system
  - Scale-out from groundup
  - Persistent to disks
  - High-Throughput (10s MB/sec per server)



# Life Since Kafka in Production



- Developed and maintained by 5 Devs + 2 SRE