

# Daniel Molnar, 6Wunderkinder GmbH

# Who we are?

6Wunderkinder makes Wunderlist in Berlin

- ▶ Productivity app on iPhone, iPad, Mac, Android, Windows, Kindle Fire and the Web
- ▶ 13+ million users, 5 years, headcount of 67
- ▶ From monolithic Rails to polyglot microservices (Scala, Clojure, Go) heavy on AWS

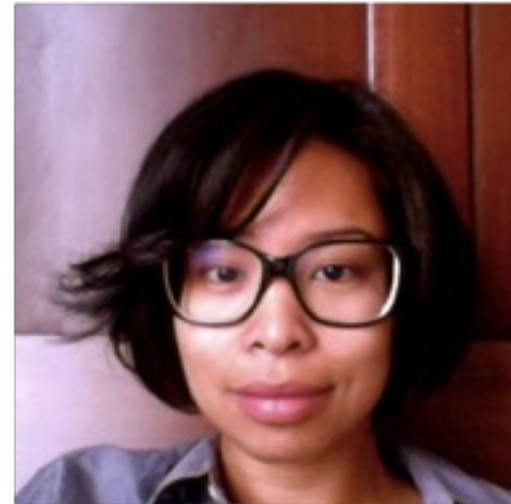


# Data Team

Tightly-knit team of all-rounders



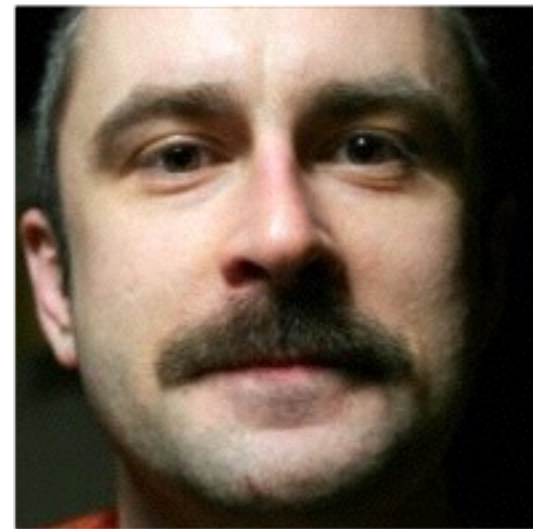
**Torsten Becker**  
Infrastructure  
BI + ML



**Jenny Herald**  
BI + UX  
Finance



**Faludi Bence**  
Infrastructure  
BI + ML



**Molnár Dániel**  
Infrastructure  
BI + ML



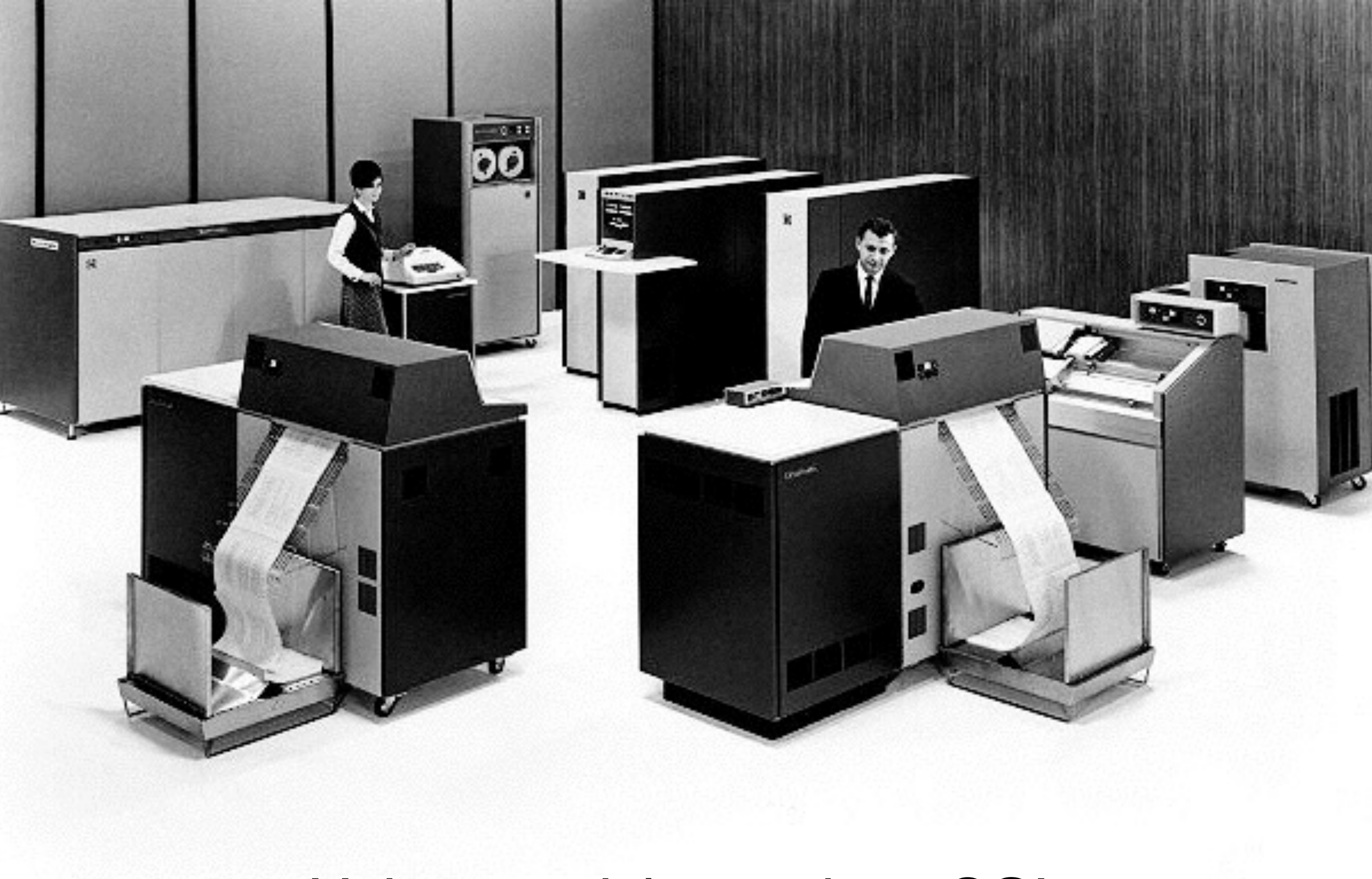
# Data Stack Philosophy

I Was Made For Lovin' You

- ▶ Closet clean
- ▶ Borges' Labyrinths
- ▶ Backwards straight
- ▶ Data mythology
- ▶ Self service





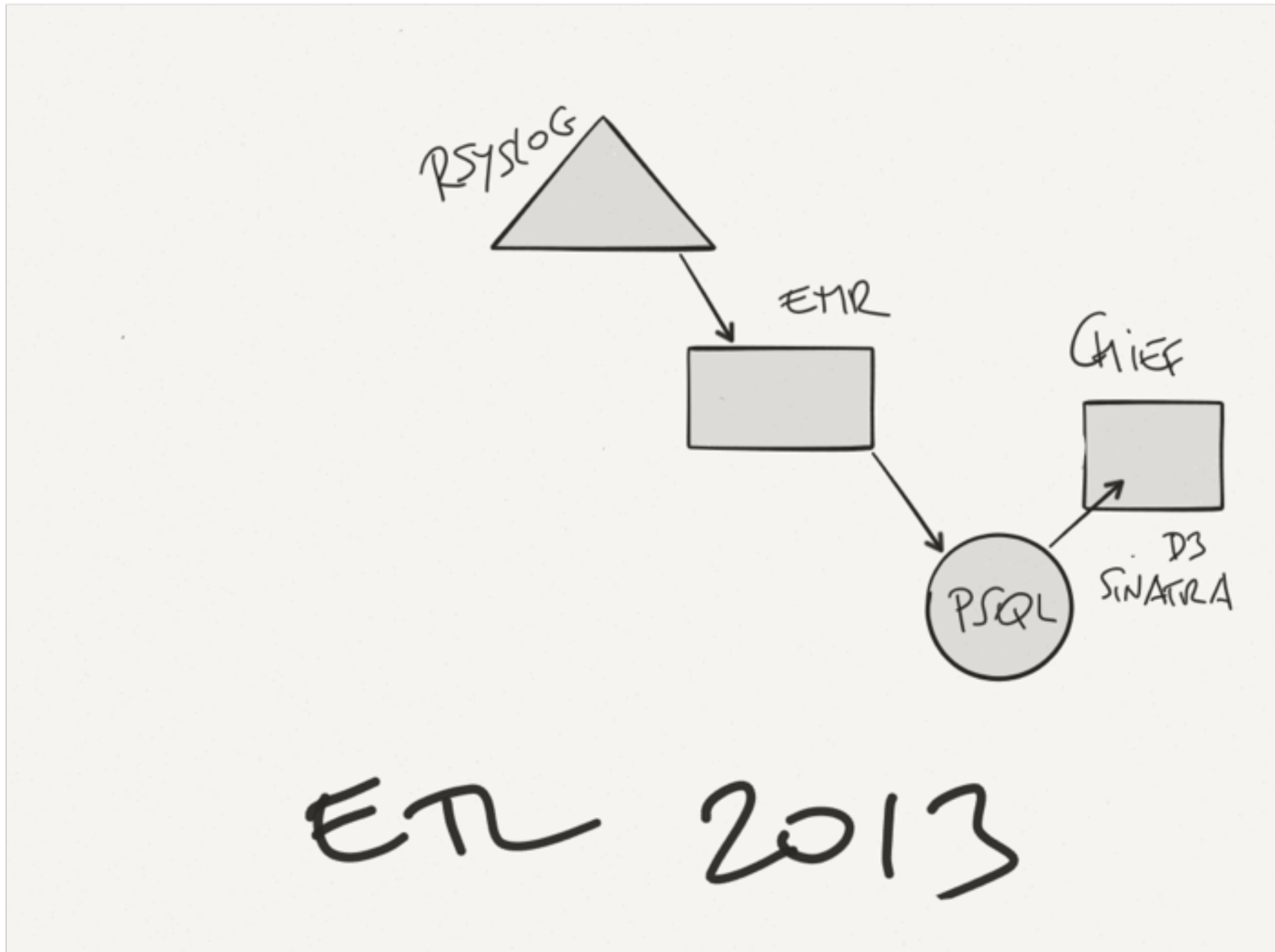


# Unix + cronjob + make + SQL

Choose Boring Technology

# How It Started?

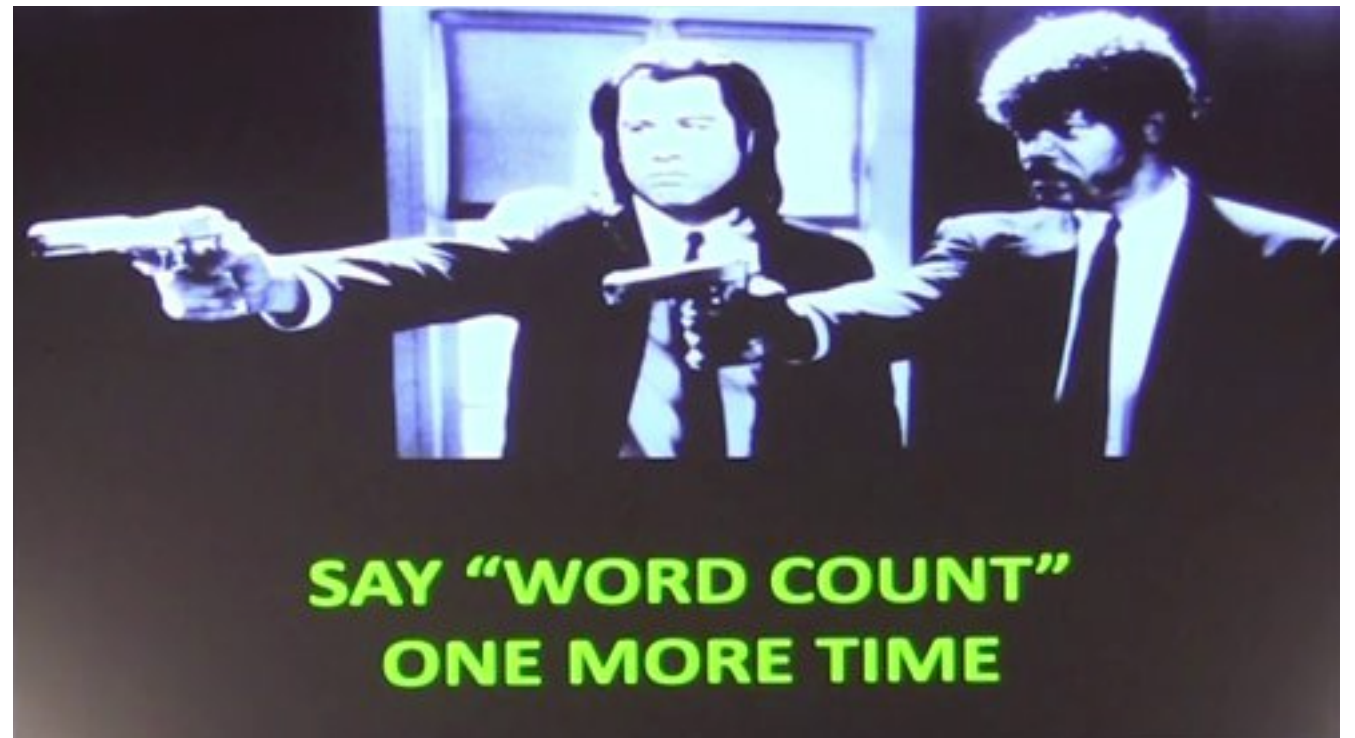
Mapreduce My Heart



# Logging

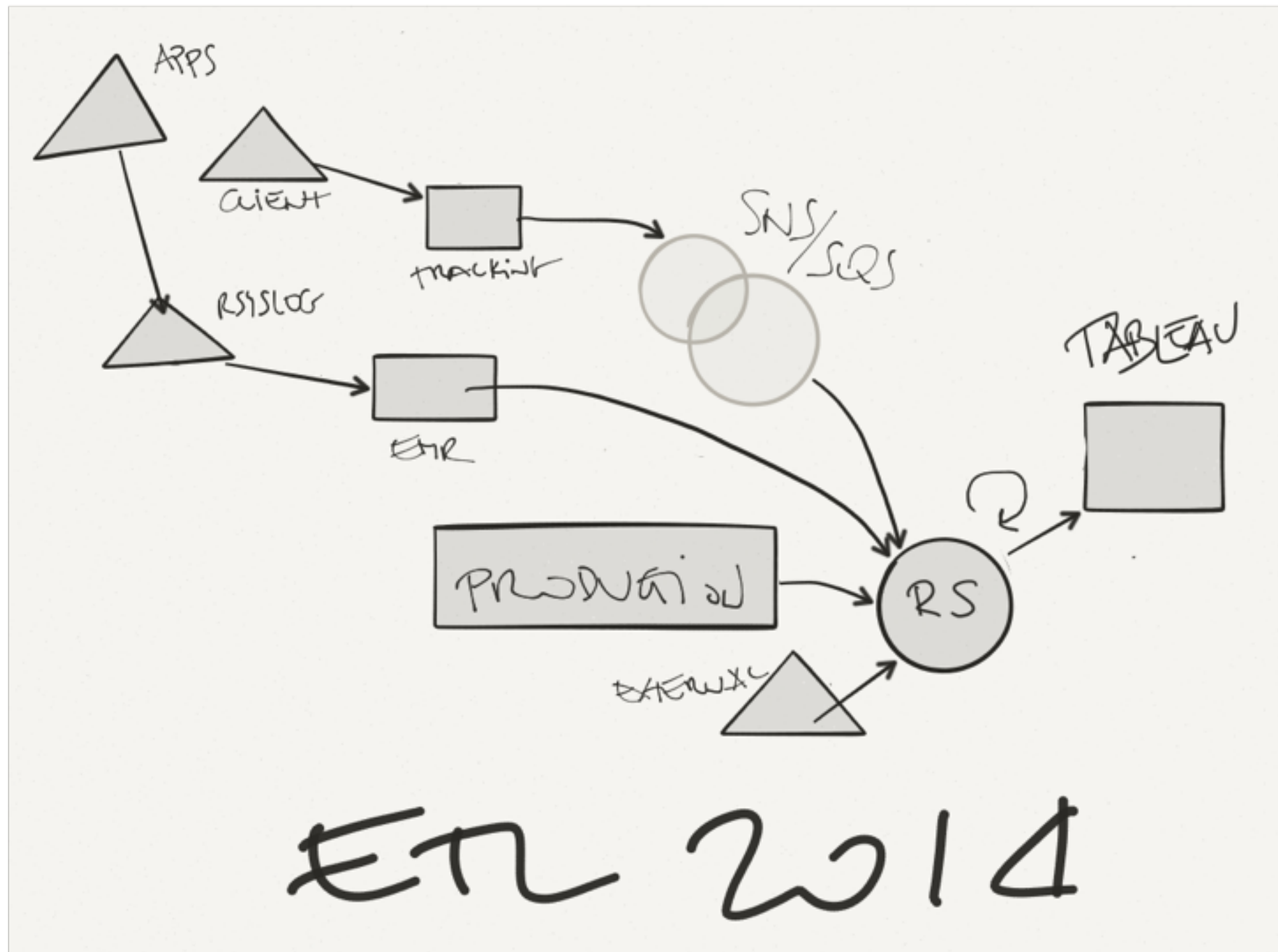
Say “Google Analytics” one more time

- ▶ No to: GA (no raw, no attribution, sampling, off by X%), Kinesis, Snowplow
- ▶ Tools: Railslog, Noxy, homebred tracker, Adjust
- ▶ Mr Beaver (EMR job in Scala)
- ▶ Tracker in node.js > SNS > SQS
- ▶ 63 TB logs + 9 TB dumps
- ▶ Logging distributed systems is MEH (Monitorama PDX 2014 James Mickens)



# Getting into Gear

Already Too Many Lines





# ETL

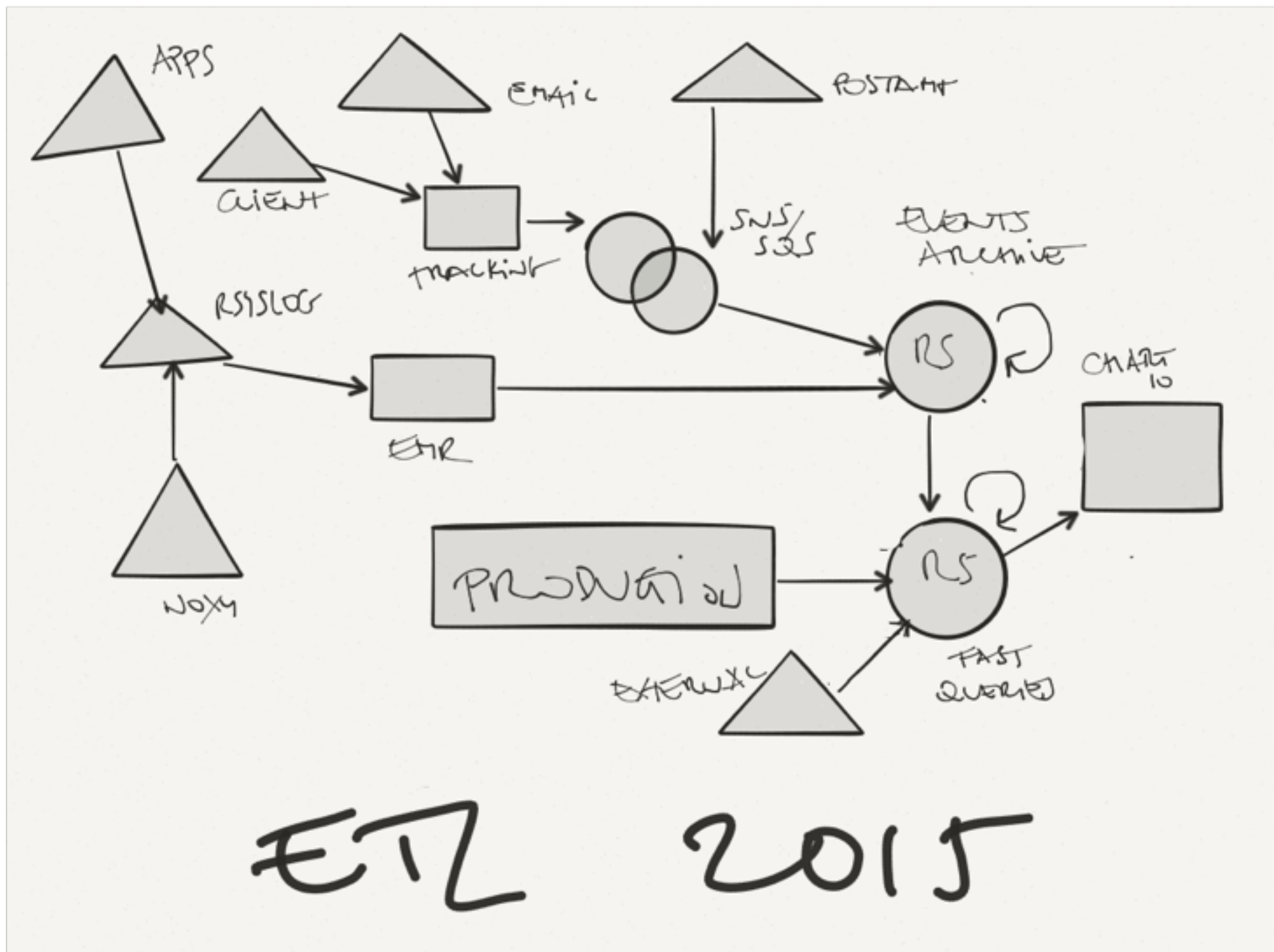
Don't Forget My Plumber

- ▶ No to: Amazon Data Flow, Oozie, Luigi
- ▶ Nightly cronjob + make + 240 ETL SQLs
- ▶ 41 sources (events, production DBs, App Annie, Mailchimp, payment providers, Maxmind)
- ▶ Inject variables and logic into SQL with ERB
- ▶ Timing with a bash wrapper



# State of the Union

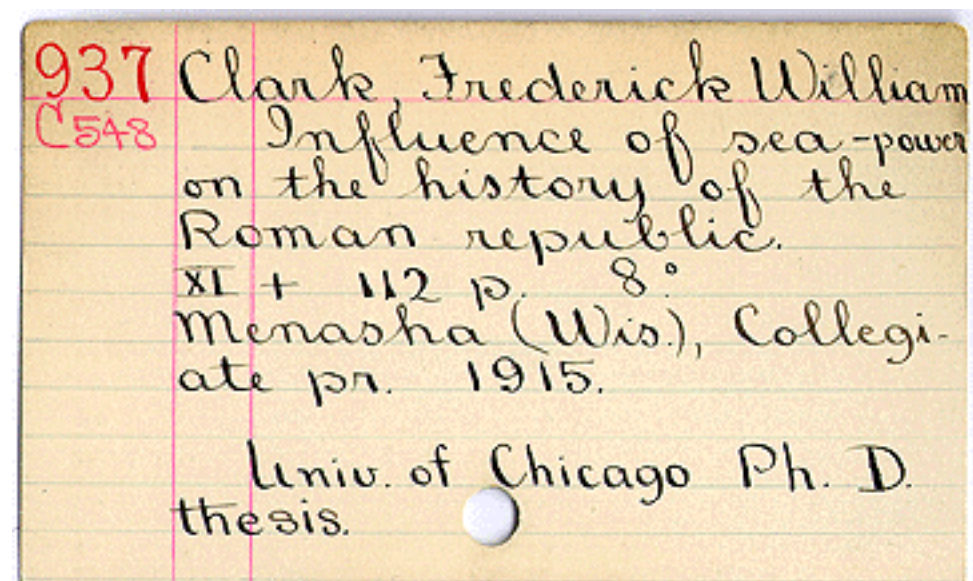
Mysterious Arrows All Around



# Datawarehouse

Drop That Table Like It's Hot

- ▶ No to: Hadoop, Hive, Impala
- ▶ Tools: (PSQL) Redshift
- ▶ Barebone DW, JSON, window functions
  - + crack (superfast, cheap)
  - GUI, support, reboots
- ▶ Don't join, filter - no starschema
- ▶ 32 small SSDs to  
5 small HDs (cold) + 16 small SSDs (hot)
- ▶ 4 TB in 280 tables
- ▶ 'real' schema





# Reporting and Datavis

"If you go micro, it's pretty hard to distinguish between bad data and crazy people."

- ▶ No to: Localytics, Looker
- ▶ Tools: (Sinatra + D3 > Tableau) > Chart.io
- ▶ Tableau
  - + value for money
  - cashcow, Windows server, Mac app, Redshift connector)
- ▶ 240 chart.io SQLs



# Business Intelligence

Friends don't let friends calculate p-values (without fully understanding them)

- ▶ Tools: Wizard (OSX), iPython notebooks
- ▶ Default KPIs, DAU (active), MAU
- ▶ Monthly and weekly cohorts
- ▶ Segments based on platform, geography and activity
- ▶ Funnels for segments



# Experiments

You are not LinkedIn

- ▶ Tools: Optimizely, homebred system
- ▶ A/B tests on app features + any messaging
- ▶ A/A — illusory A/B
- ▶ Too small > Bayesian > less certainty
- ▶ Short-Term Bias, Regression to the Mean, Random Variation
- ▶ Chris Stucchio, Evan Miller

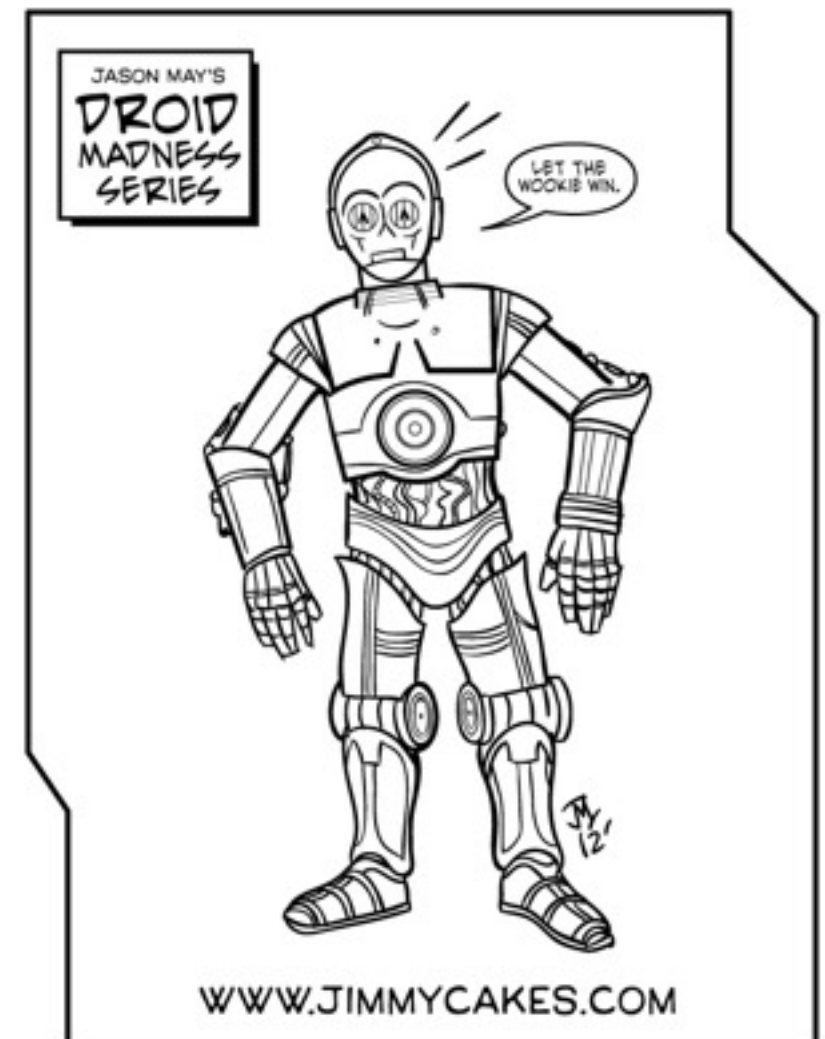




# Machine Learning

7 years on a PhD in ML to build "suggested pokes" at Facebook

- ▶ No PhDs
- ▶ The Mailchimp way
- ▶ We use LDA and NLP



- ▶ Data quality
- ▶ [Surveymonkey.com](https://www.surveymonkey.com)
- ▶ [Usertesting.com](https://www.usertesting.com) and [Lookback](https://lookback.io)
- ▶ Back of a napkin



Whatnot?