Taming the Inferno of the Online Ad Markets

- billions of microtransactions per day
- dozens of publisher, advertiser, & audience attributes
Goal: Fast Analytics
Over 100s of Terabytes
Goal: Fast Analytics
Over 100s of Terabytes

queries in seconds

data crunched in minutes

dashboard
database

ingestion
Solution 1: MPP Database

Queries in minutes

Data crunched in minutes

Dashboard

Database

Ingestion

Hadoop
Solution 2:
HBase

Queries in seconds

Data crunched in hours

Dashboard

Database

Ingestion

Hadoop
Solution 3: Do It Ourselves: Druid

queries in seconds

data crunched in minutes

dashboard

database

ingestion

Hadoop

Druid
Four Principles of Druid’s Performance at Scale

**SUMMARIZE**

- 100x smaller vs raw data

**DISTRIBUTE**

- 100x throughput vs a single node (with 100 cores)

**PARALLELIZE**

**STORE IN-MEMORY**

- 100x faster vs disk

Druid can filter and aggregate over 1 billion rows per second on a 50-core cluster, or 20m rows per core per second.

= $10^6$

factor speed-up
Consequences of Druid: Faster Queries

photo credit tonylanciabeta http://www.flickr.com/photos/tonysphotos/3305157904/sizes/o/in/photostream/
Consequences of Druid: Fresher Data

photo credit: Lars P. http://www.flickr.com/photos/lars_p/4911238308/sizes/o/in/photostream/
Consequences of Druid: Scalable in the Cloud

photo credit: MonkeyAt Large http://www.flickr.com/photos/monkeyatlarge/16645379/sizes/l/in/photostream/
One Billion Rows Per Second: Analytics for the Digital Media Markets

QUESTIONS? CONTACT ME AT MIKE@METAMARKETSGROUP.COM

MICHAEL DRISCOLL CO-FOUNDER & CTO @medriscoll