

MINORITY REPORT

Meets Big Data

THE TECHNOLOGY
THE INTERFACE
THE DESIGN

Eva Andreasson, Cloudera @EvaAndreasson
Justin Langseth, Zoomdata @justinlangseth



We already have...

Iris Scanning



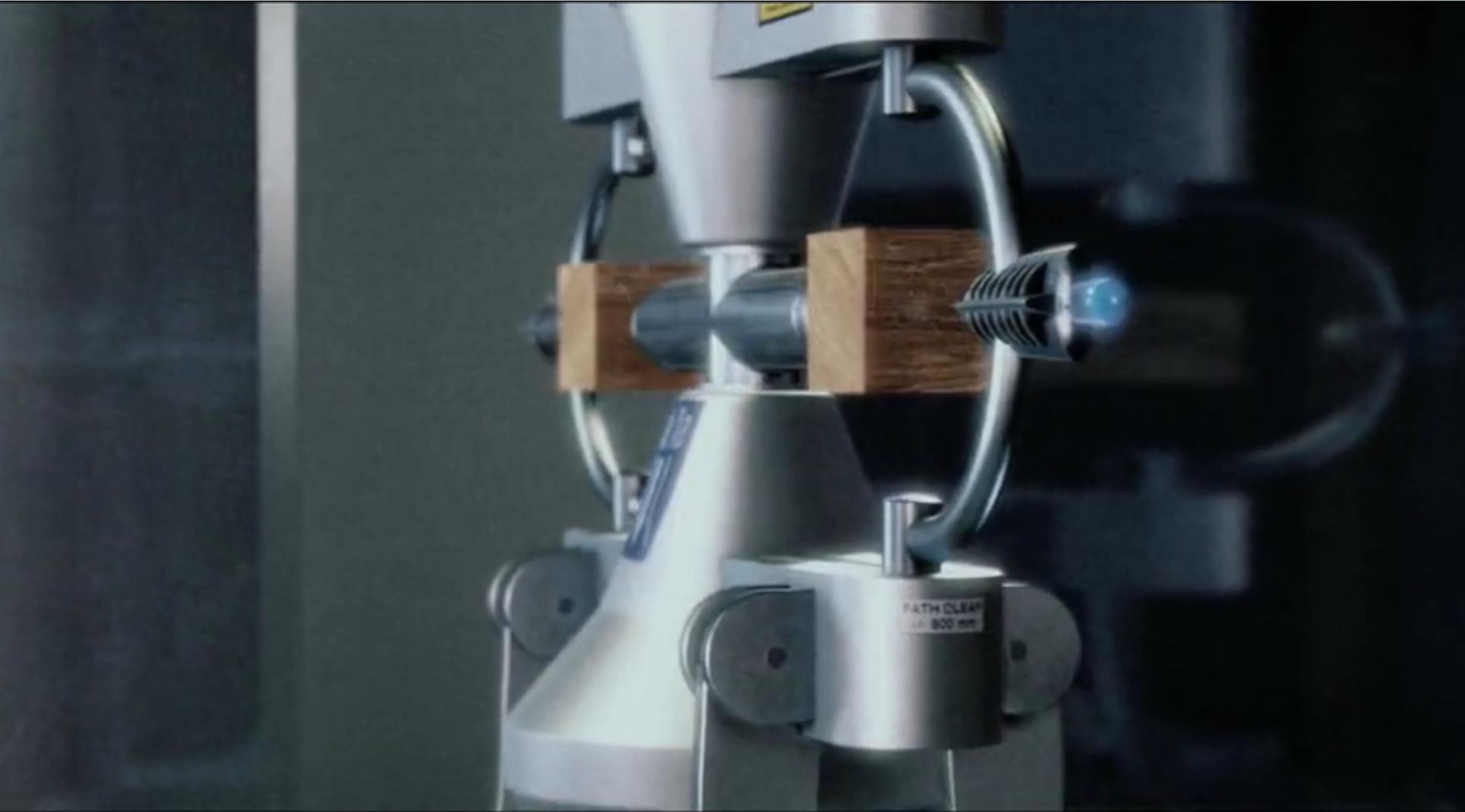
Targeted Advertising



Google Glass

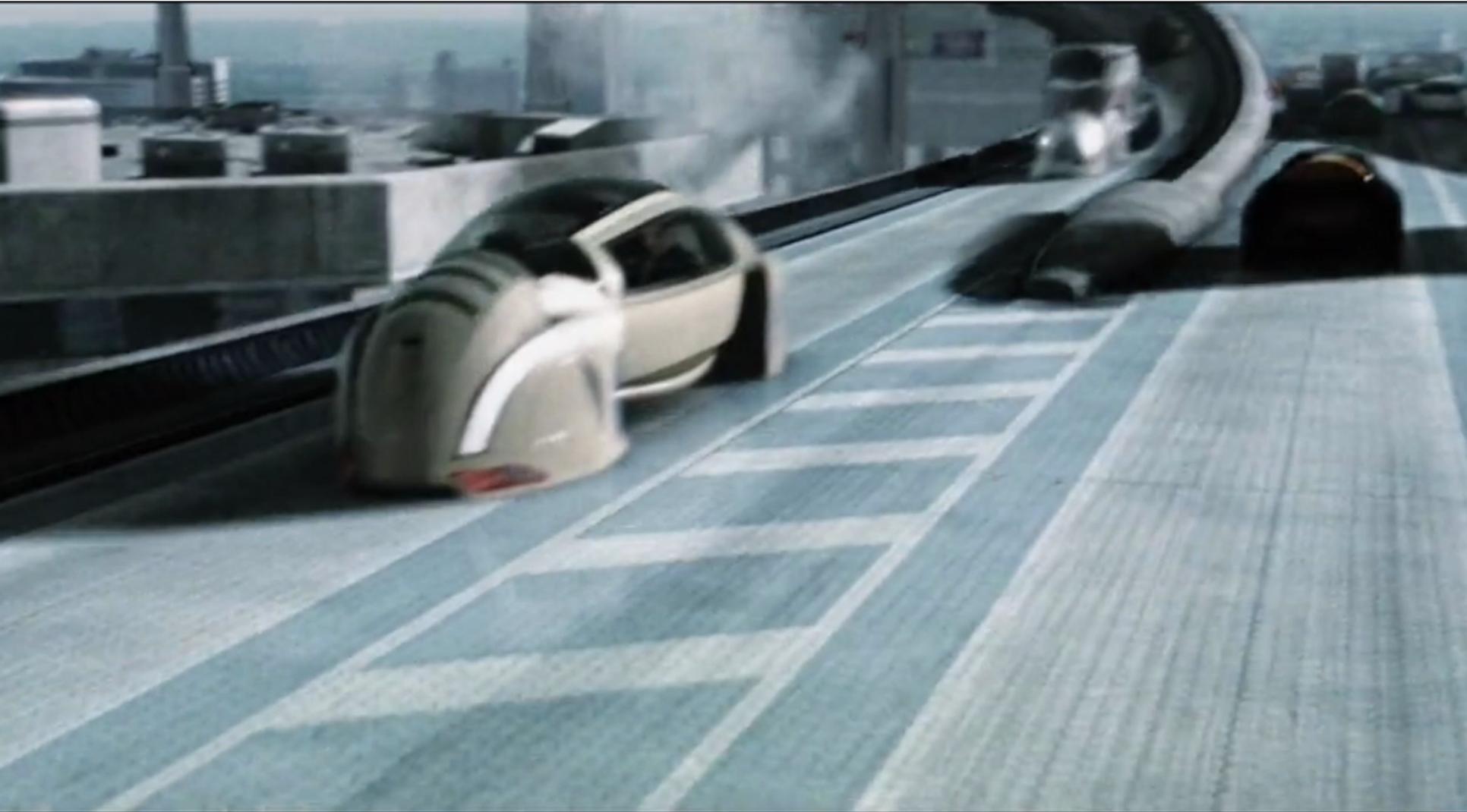


3D Printing



We can look forward to...

Self-driving Cars



Smart Shopping



Robotic Spiders



The Daily Prophet



High-stakes Frogger



And, to the return of...

Jetpacks



Floppy Disks



But what we really,
really want is...

The Interface



Our Focus

**Designing the Minority Report
GUI, Leveraging Big Data,
for Normal Humans**



Supply Chain Viz



Drug Research



Ad Audience Analysis



Info/Corp Security

Curiosity Fabric



How do you actually *build* a
curiosity fabric?

Are You a Data Geek?

**DATA
GEEKS**



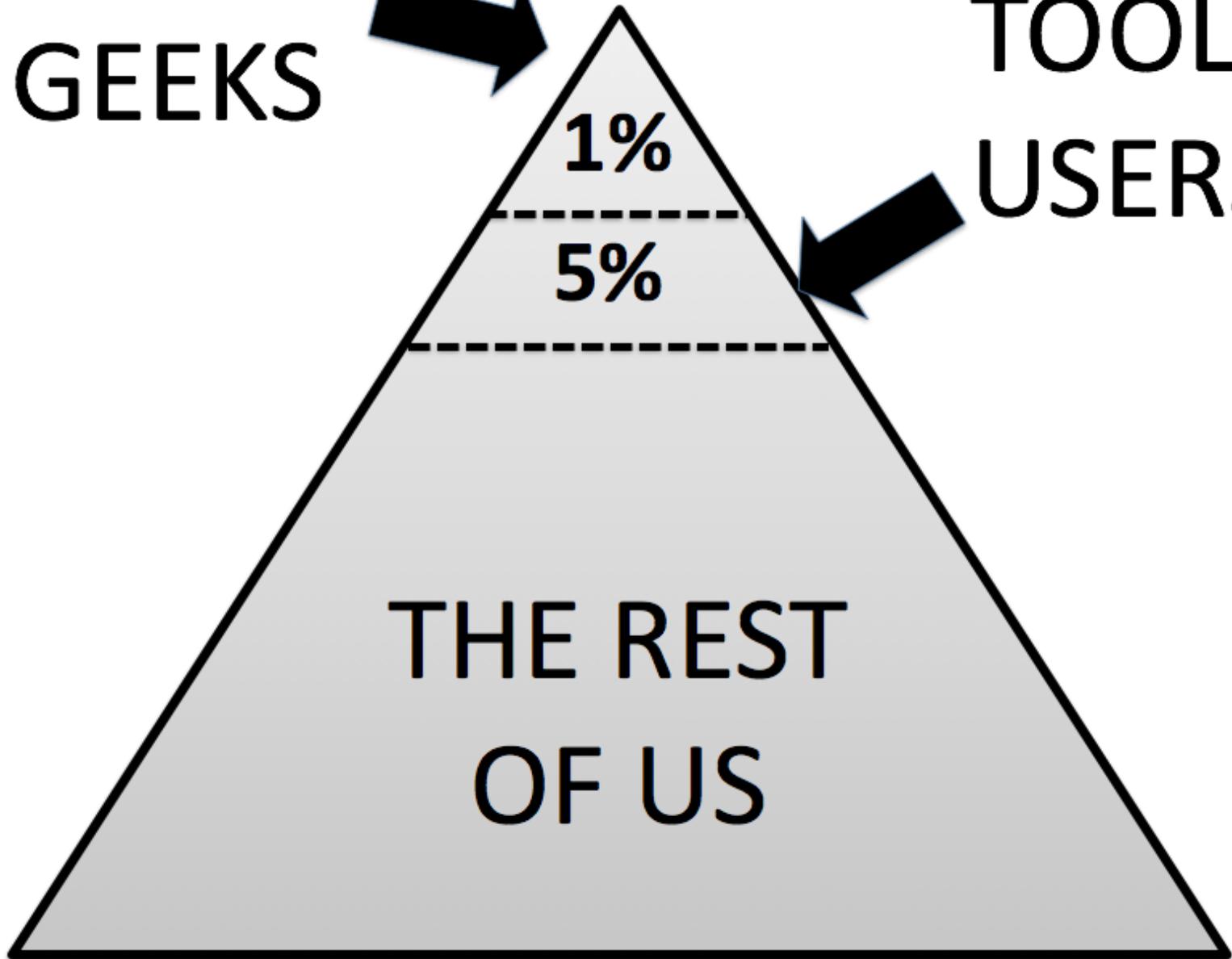
**TOOL
USERS**

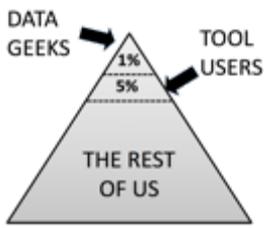


1%

5%

**THE REST
OF US**





DATA GEEK

GENOME PROJECT

Geek #1

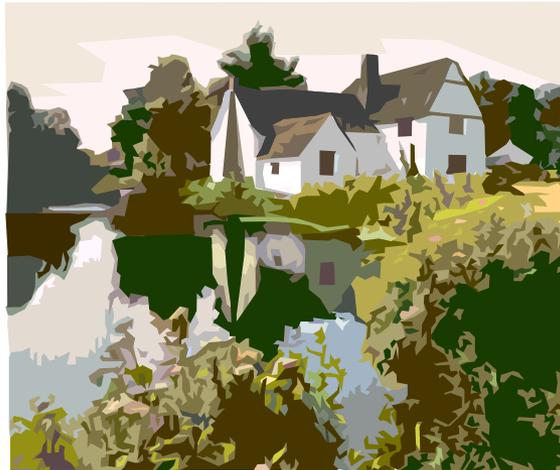


Gilbert

Architect Data Geek

This geek is responsible for loading, storing, and sometimes structuring data.

The Rooms of Your House



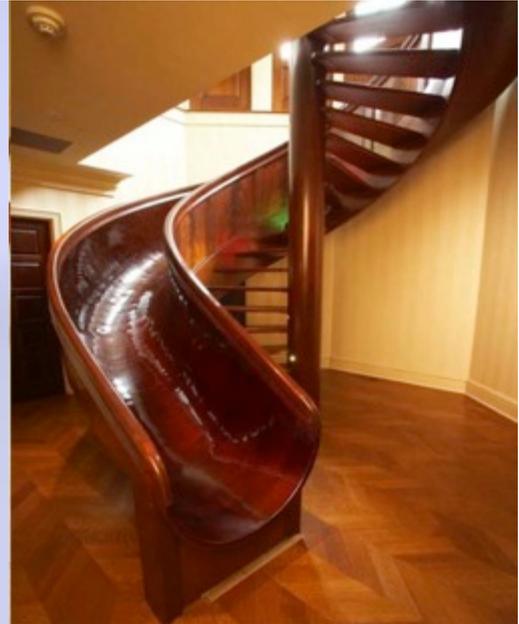
Real Time Ingest and Ingest



Cost Efficient, Multi-Type Storage



Right Tool for the Right Task



Streamline Multi-Workflows

- Silos?



Streamline Multi-Workflows

- Integration!



Visibility, Audit, and Access Control



What Lego Pieces to Use

- EDW

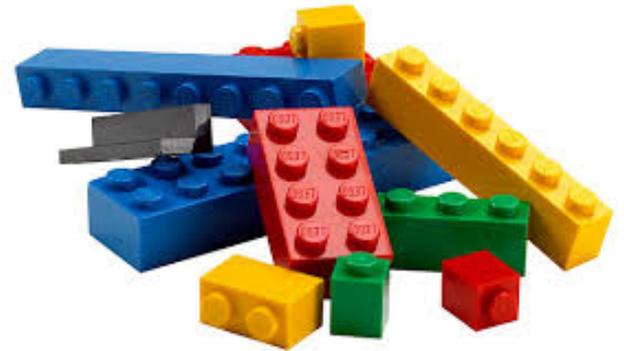


What Lego Pieces to Use

- EDW



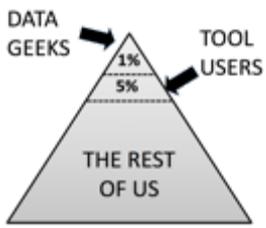
- Apache Hadoop



Introducing: The Enterprise Data Hub

- Hadoop at the core
 - Open source
 - Scalable
 - Flexible
- Add to that:
 - Integrated with ecosystem
 - Multi-workload
 - Impala, Search, Spark, ...
 - Easily managed
 - Cloudera Manager
 - Secure
 - Sentry
 - Cloudera Navigator





DATA GEEK GENOME PROJECT

Geek #2



Analytics Geek

The analytics geek installs, configures, trains, and uses data analytics and machine learning systems.

Lewis

Let's take a step back...

288 megapixels

X 5 bytes/pixel

X 500 fps

X 1 sec = 720 GB

X 2 eyes

The allegory of the bear = 144 TB

Today's bears
represent
opportunity...

Curiosity Fabric

MVP Requirements

Realtime and historical,
handle structured, unstructured, semistructured, log, graph,
be blazingly fast,
fuse data within and across sources,
leverage advanced analytics, statistics, and ML,
be easy, beautiful, collaborative,
on desktop, web, mobile, and touch devices.

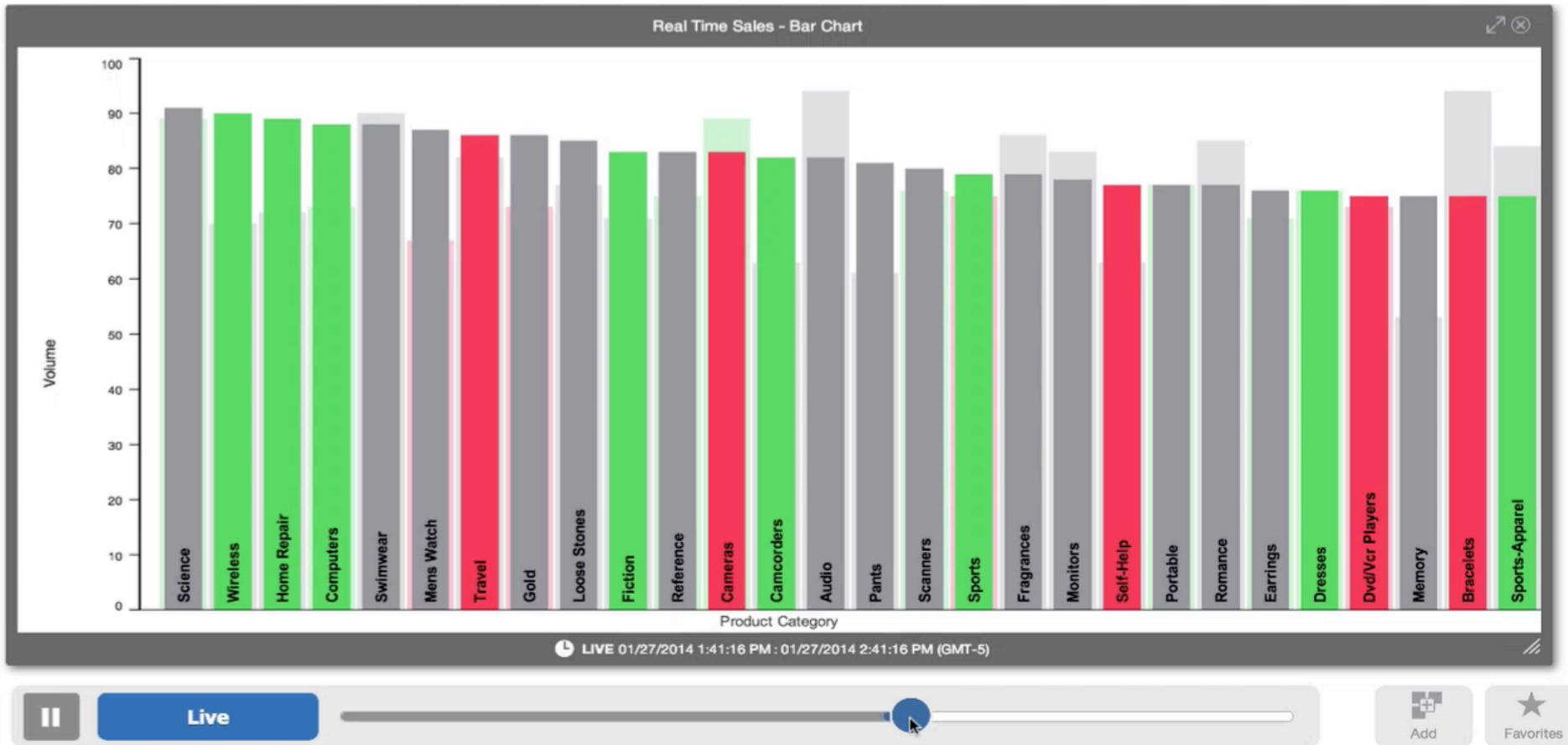
~~Copy?~~

Design.

Innovate.

Requirement: Realtime and historical

Design: Data DVR, stream processing



Requirement: handle structured, unstructured, semistructured, log, graph,

Design: Pluggable big data connectors



Requirement: Blazingly Fast

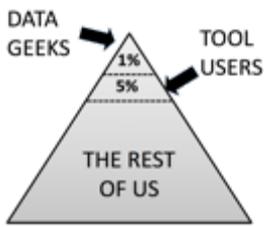
Design: Micro-query stream aggregation



Curiosity Fabric

MVP Requirements

Realtime and historical,
handle structured, unstructured, semistructured, log, graph,
be blazingly fast,
fuse data within and across sources,
leverage advanced analytics, statistics, and ML,
be easy, beautiful, collaborative,
on desktop, web, mobile, and touch devices.



DATA GEEK GENOME PROJECT

Geek #3

Data Artist Geek

The data artist geek codes customized javascript / d3 data visualizations to allow for intuitive understanding of complex systems.



Poindexter



© 2013 Google
© 2009 GeoBasis-DE/BKG
US Dept of State Geographer

Requirement: Be easy, beautiful, and collaborative

Design: Reimagined, touch-first BI interface

