

Data without Limits

Dr. Werner Vogels

CTO, Amazon.com

[Shop All Departments](#) ▼Search

GO

Cart

[Your Lists](#) ▼[Books](#)[Advanced Search](#)[Browse Subjects](#)[New Releases](#)[Bestsellers](#)[The New York Times® Bestsellers](#)[Libros En Español](#)[Bargain Books](#)[Textbooks](#)

Prime

Get it tomorrow, **Friday**, June 11Get it **Saturday**, June 12 **One-Day 1-Click®—\$3.99**

Order within 7 hr 9 min

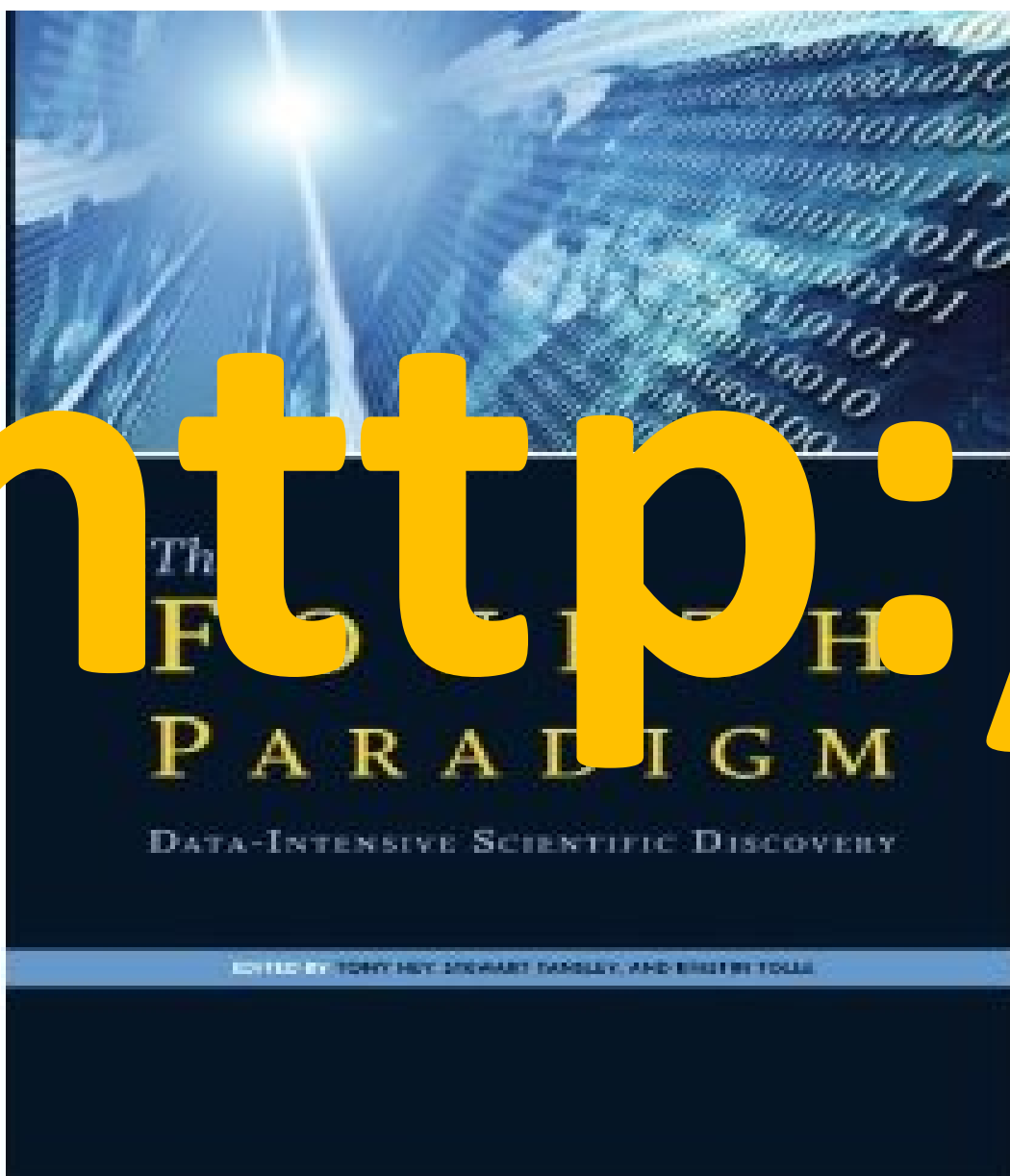
Two-Day 1-Click®—FREE

Order within 3 hr 9 min

Ship to:



[The Fourth Paradigm: Data-Intensive Scientific Discovery](#) and over 540,000 other books are available for **Amazon Kindle** – Amazon's new wireless reading device. [Learn more](#)

[See larger image](#)[Share your own customer images](#)

[Publisher: learn how customers can search inside this book.](#)

The Fourth Paradigm: Data-Intensive Scientific Discovery (Paperback)

~ [Tony Hey](#) (Editor), [Stewart Tansley](#) ▼ (Editor), [Kristin Tolle](#) (Editor)

★ ★ ★ ★ ☆ ▼ (4 customer reviews)

List Price: ~~\$46.07~~Price: **\$39.09** Eligible for free shipping with Amazon Prime.
You Save: ~~\$6.98~~ (15%)**In Stock.**Ships from and sold by **Amazon.com**. Gift-wrap available.**4 used** from **\$28.99**

Formats	Amazon Price	New from	Used from
Kindle Edition	\$0.99	—	—
Paperback	\$39.09 ✓Prime	\$39.09	\$28.99

Buy New

Quantity: **Add to Cart**

or Buy now with

Two-Day 1-Click®—FREEShip to: ☐ Add gift-wrap/note[Add to Wish List](#) ▼[Add to Shopping List](#)

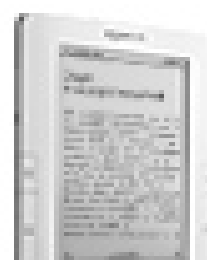
Buy Used

Used - Very Good [See details](#)**\$28.99** Eligible for free shipping with Amazon Prime.

Fulfilled by Amazon

Add to Cart

or Buy now with

Two-Day 1-Click®—FREE

Start reading [The Fourth Paradigm: Data-Intensive Scientific Discovery](#) on your Kindle **in under a minute.**

Don't have a Kindle? Get your

BIG-DATA

DATA Intensive
Centric

BIG-DATA

BIG-DATA

When your data sets become so large that you have to start innovating how to collect, store, organize, analyze and share it

3Vs **Volume**
Velocity
Variety

BIG-DATA

The collection and analysis of
large amounts of data to create
a competitive advantage

BIGGER IS BETTER

Frequently Bought Together

Customers buy this item with [K&N HP-7014 Cartridge Oil Filter](#) by K&N **\$10.39**



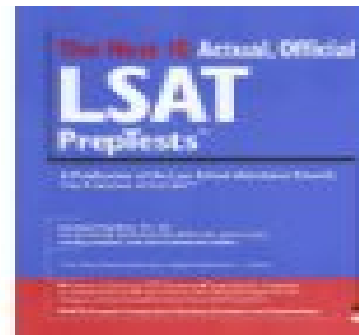
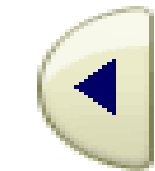
Price For Both: \$60.65

Add both to Cart

Add both to Wish List

[Show availability and shipping details](#)

Customers Who Bought This Item Also Bought



[The Next 10 Actual, Official LSAT PrepTests \(Ls...](#) by Wendy Margolis
★★★★☆ (25)
\$19.80



[Ultimate Collection ~ Jackson 5](#)
★★★★☆ (75)
\$10.68



[Acqua Di Giò By Giorgio Armani For Men. Eau De Toilette Spray](#)
★★★★☆ (263)
\$2.49 - \$158.00



[Motorola Signal Booster 484095-001-00 Bi-Directional RF Amplifier](#)
★★★★☆ (528)
\$31.98



[Microsoft Windows 7 Professional, Full Version \(PC DVD\)](#)

by Microsoft (Oct 22, 2009)

Platform: Windows 7

Average Customer Review: ★★★★★ (52)

In stock

RRP: £219.99

Price: **£157.53**

[17 used & new](#) from **£145.99**

Add to Basket

Add to Wish List

☐ I own it ☒ Not interested | ★★★★★ Saved

Recommended because you added [Japanese Kitchen Knives: Essential Techniques and Recipes](#) to your Wish List ([Fix this](#))

UNCERTAINTY

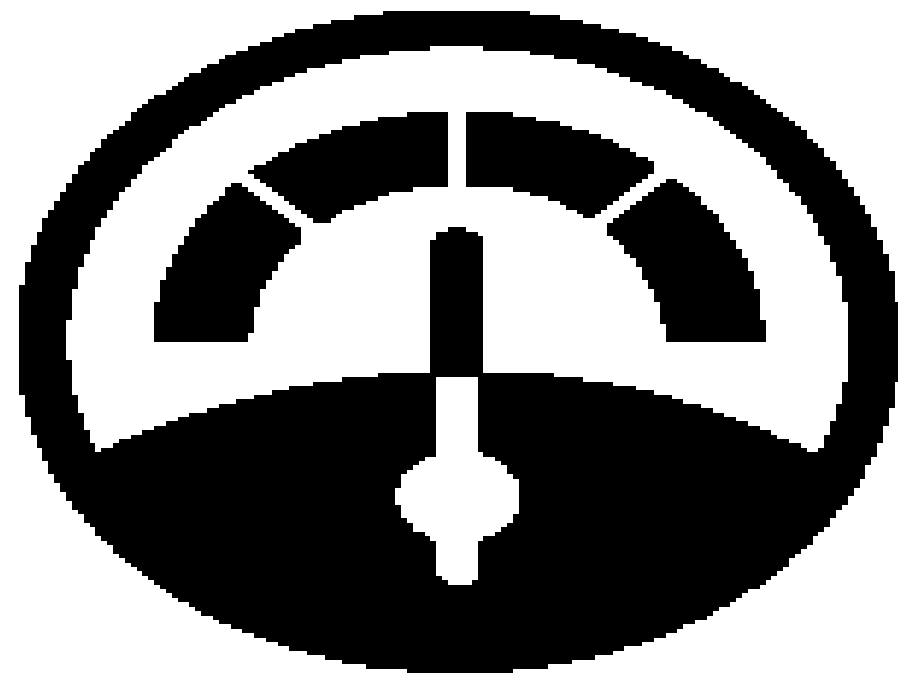
BIG-DATA REQUIRES

NO LIMITS

COLLECT | STORE | ORGANIZE | ANALYZE | SHARE

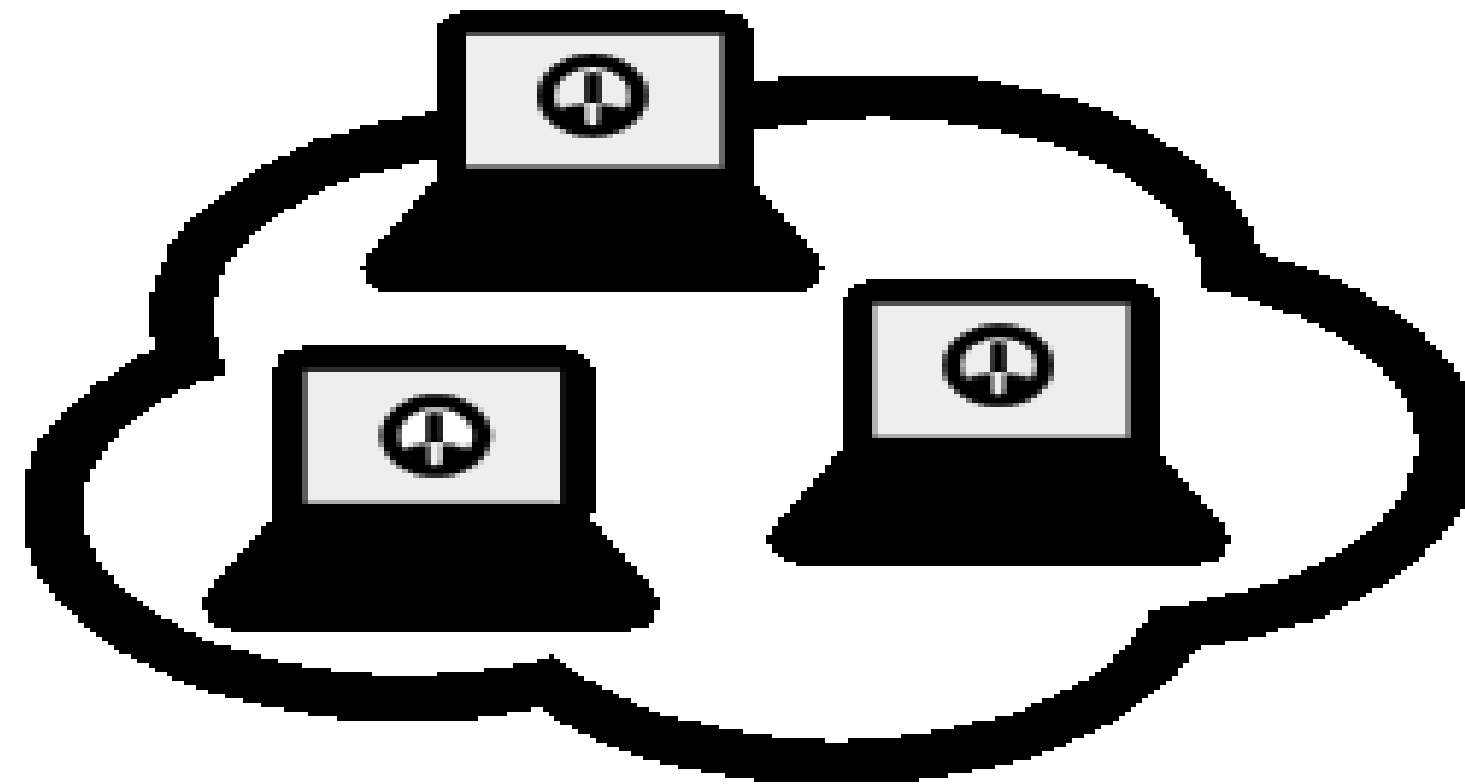
COLLECT | STORE | ORGANIZE | ANALYZE | SHARE

wakooopa



Step 1: Tracking

We've created a unique tracking application. It keeps track of all website visited, software used, and/or ads seen.



Step 2: Panel

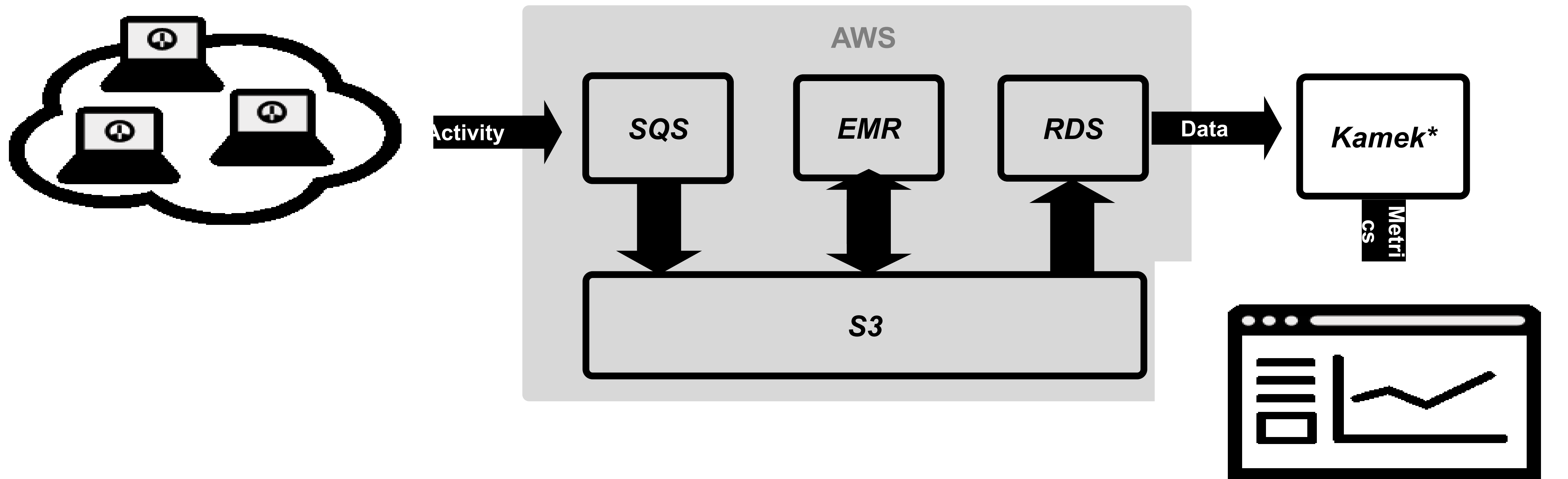
We invite members of a research panel to install it. We know not only their digital habits, but also their offline demographics and behavior.



Step 3: Dashboard

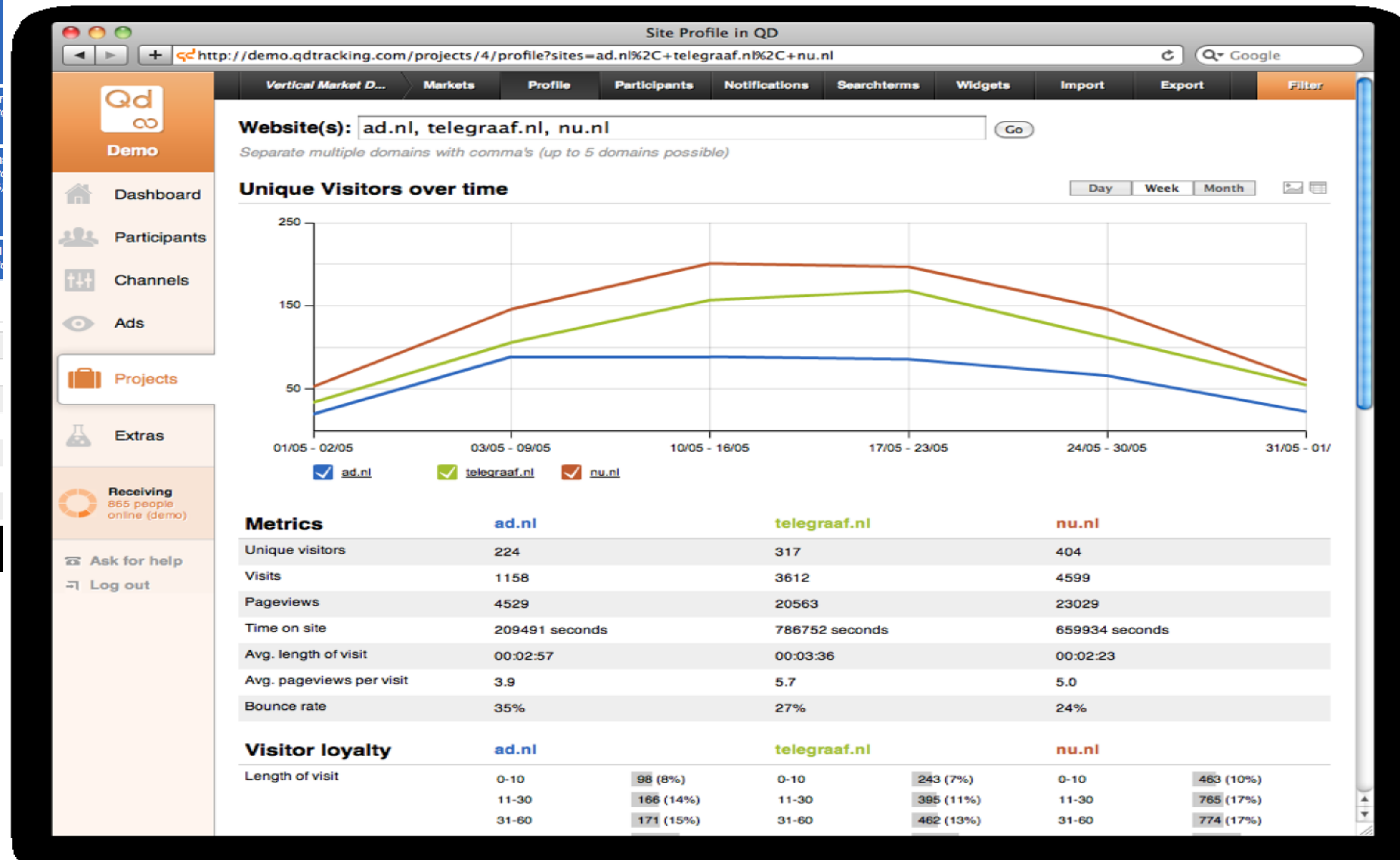
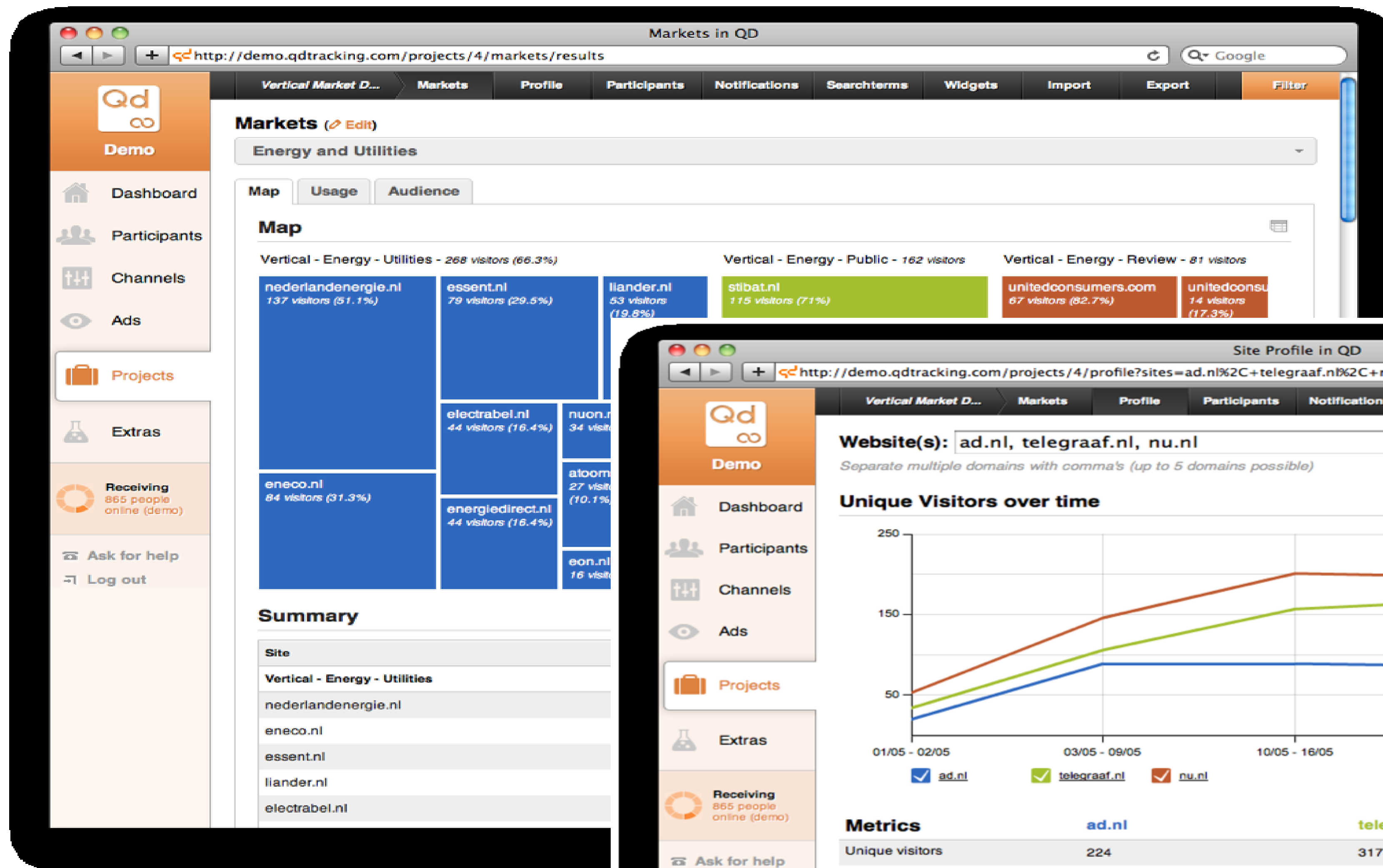
Usage data now begins to pour into the Wakoopa dashboard in real-time. Log in, and create beautiful visualizations and useful reports.

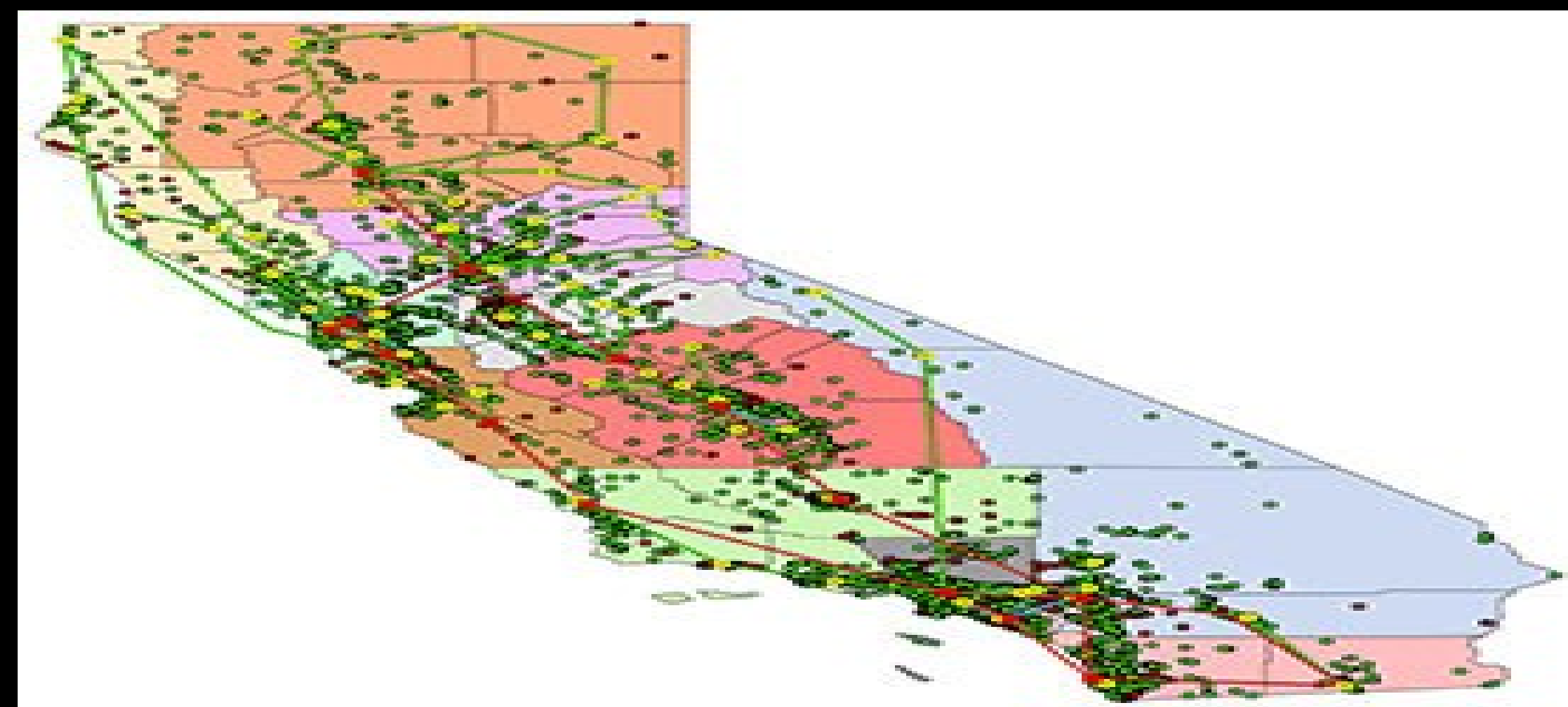
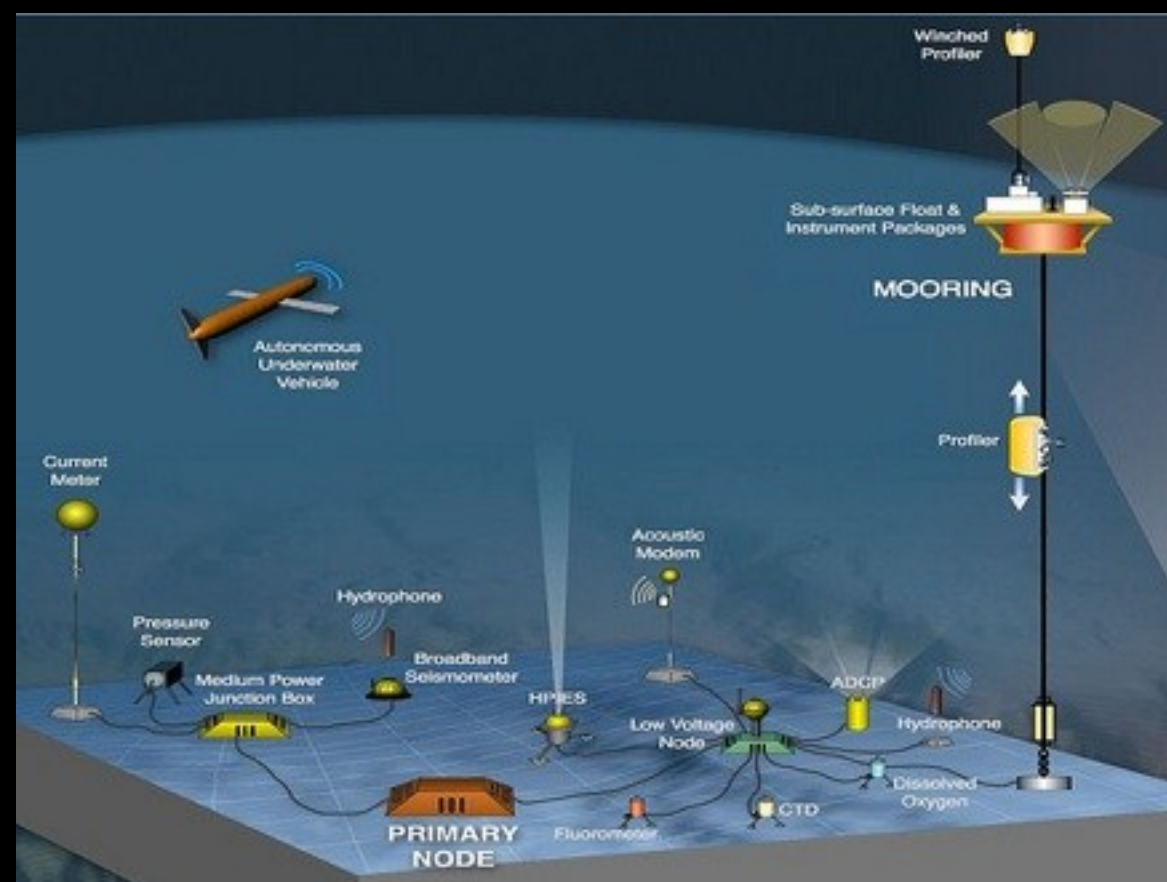
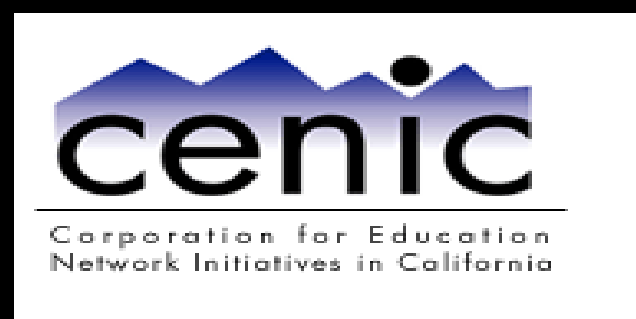
wakooopa



Wakoopa dashboard







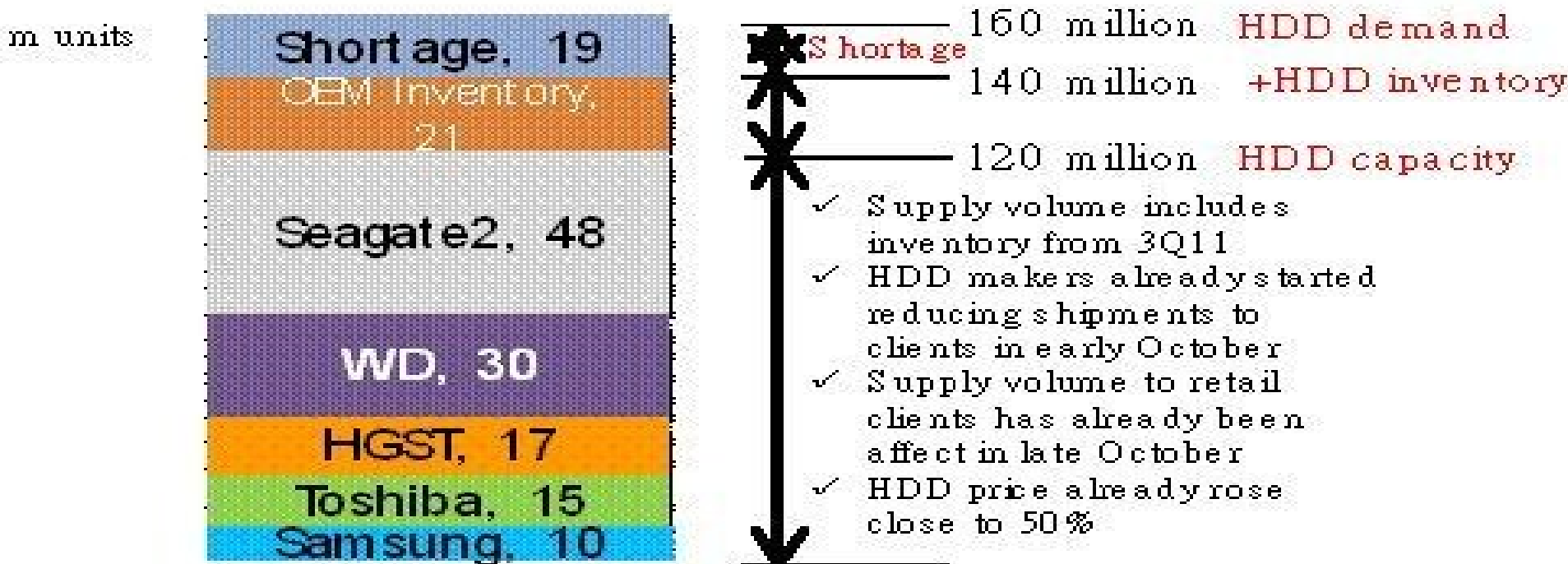
COLLECT | **STORE** | ORGANIZE | ANALYZE | SHARE

razorfish™

Storage Muck

4Q11 HDD shortage to reach 19 million units

4Q11 HDD shipments and supply gap estimate



3 Source: DIGITIMES Research, October 2011



Database Muck

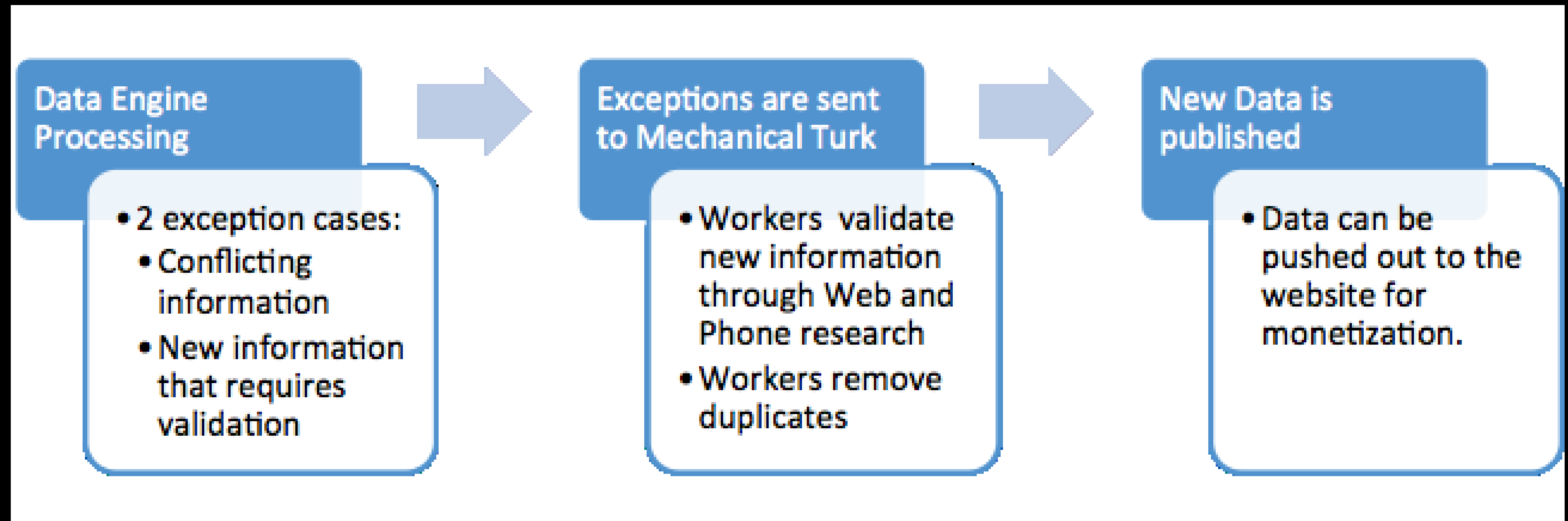
Amazon DynamoDB

COLLECT | STORE | **ORGANIZE** | ANALYZE | SHARE

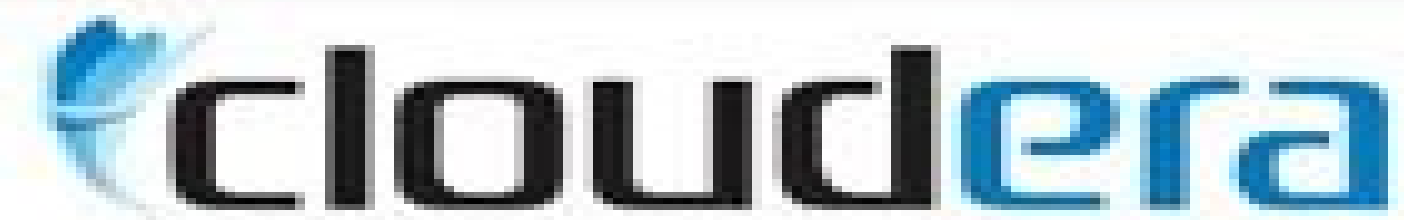
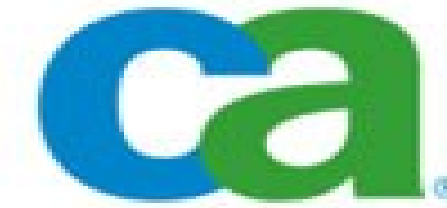
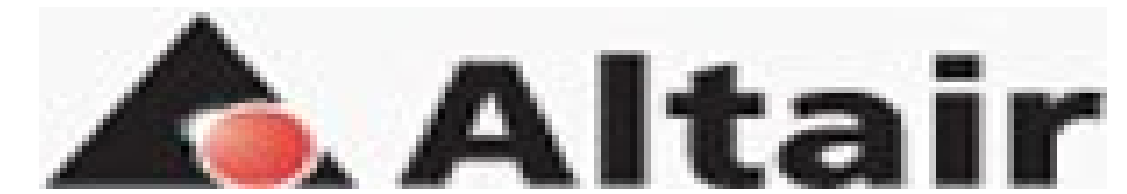
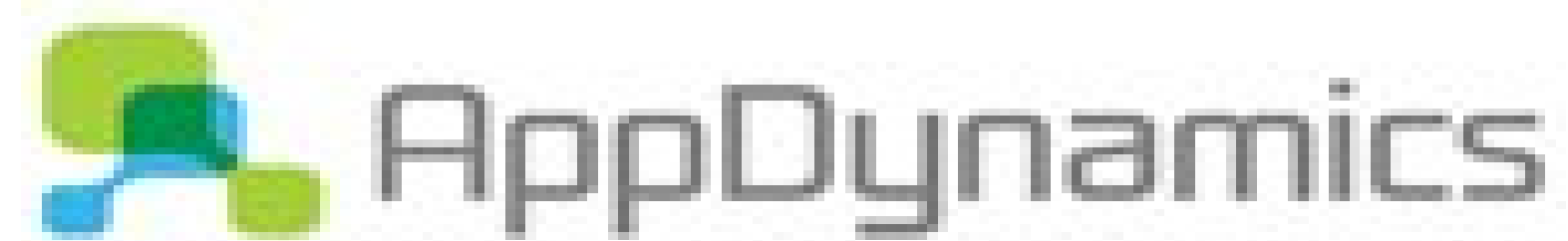
DATA QUALITY CONTROL

- Control Data
- Correct Data
- Validate Data
- Enrich Data

- A large provider of business listings (over 20MM in the US) needs to determine where each data element belongs and if it is valid.
- 1 MM new pieces of data are reviewed a day.



COLLECT | STORE | ORGANIZE | **ANALYZE** | SHARE



Computational Muck

MAPREDUCE

AMAZON ELASTIC MAPREDUCE

HADOOP

Hadoop - Elastic Map-Reduce (2009)

Logs

S3

EMR
Hadoop

Hive

Business
Intelligence

- Web Access Logs
- Streaming Service Logs
- Terabyte per day scale
- Easy Hadoop via Amazon EMR
- Hive SQL “Data Mart”
- Gateway to Datacenter BI

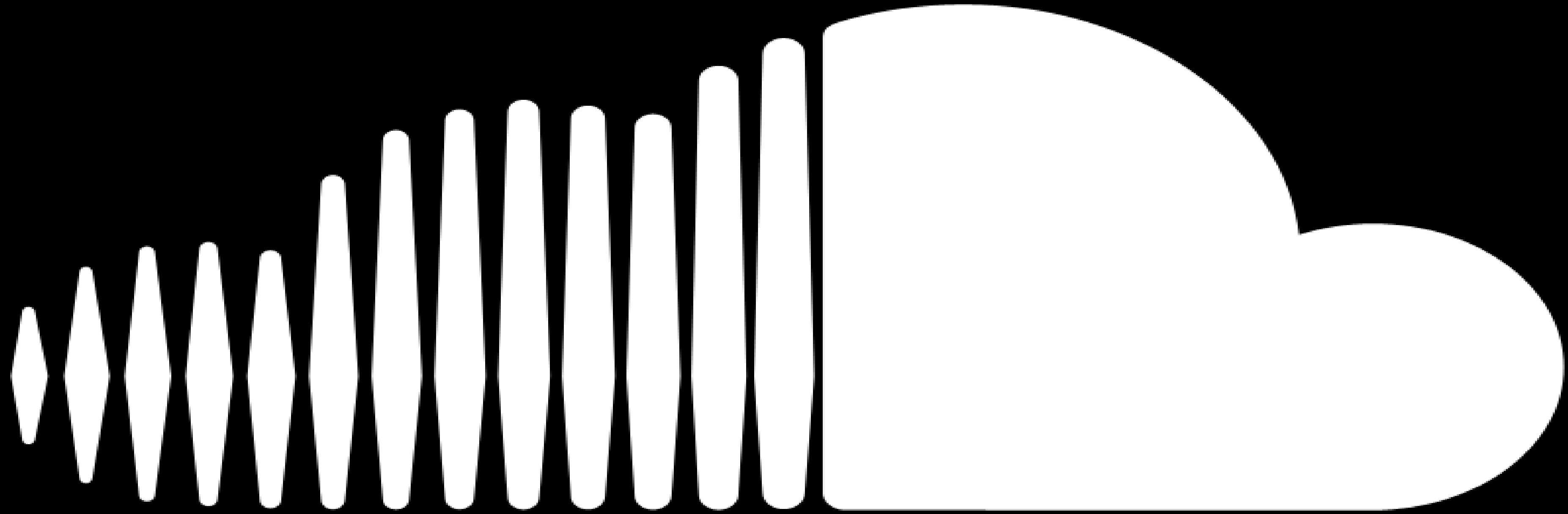
Slideshare.net talks

evamtse “Netflix: Hive User Group” <http://slidesha.re/aqJLAC>

adrianco “Crunch Your Data In The Cloud” <http://slidesha.re/dx4oCK>

NETFLIX





SOUNDCLOUD

[Post Project](#)[Find Freelancers](#)[Browse Projects ▼](#)[Make Money ▼](#)[Help ▼](#)

Hire online for a fraction of the cost!

Freelancer.com is the world's largest outsourcing marketplace, empowering entrepreneurs & small businesses worldwide.

- ✓ Outsource anything you can think of!
- ✓ Projects start at \$30 and the average job is under \$200
- ✓ Programmers, designers, content writers are ready now!
- ✓ Only pay freelancers once you are happy with their work

What type of work
do you need done?

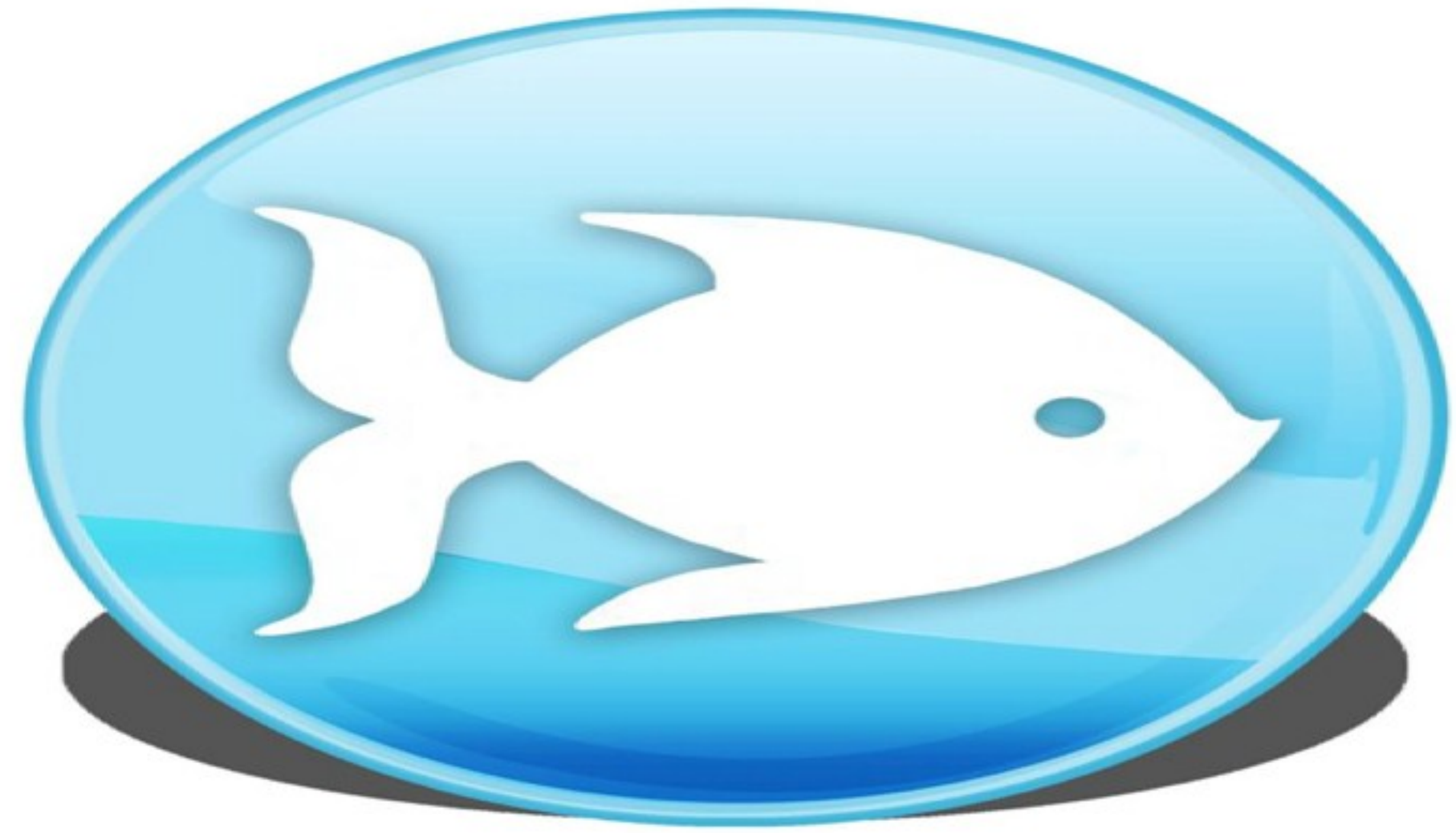
Select a Category ▼

Post a Project »

Post your first project FREE!

NEW!
HOURLY
PROJECTS

Etsy



yelp*

► Lingit lager programvare **utviklet for lese- og skrivesvake.**



Hjelpemidler



Få hjelp nå!

Vår programvare hjelper deg når du skal lese eller skrive bokmål, nynorsk eller engelsk.

[► Les mer](#)

Lingit **Skolepakke**



Skolepakke

En enklere og bedre hverdag for elever med lese- og skrivevansker

[► Les mer](#)

Nettbutikk

Bestill nå!

Du kan bestille vår programvare for lese- og skrivestøtte direkte i nettbutikken.

[► Gå til nettbutikken](#)

Nyheter

» [Foreldre betaler for hjelp til dyslektiske barn](#)

VG Nett skriver om foreldre som betaler mer enn 50.000 kr i året...

» [Barnestemme](#)

Lingit skal utvikle en barnestemme basert på HTS-teknologi i et prosjekt finansiert av...

[Last ned DEMO](#)[Last ned oppdateringer](#)



RECOVERY.GOV

TRACK THE MONEY

Recovery.gov is the U.S. government's official website providing easy access to data related to Recovery Act spending and allows for the **reporting of potential fraud, waste, and abuse.**

Connect With Us


 Search

Go

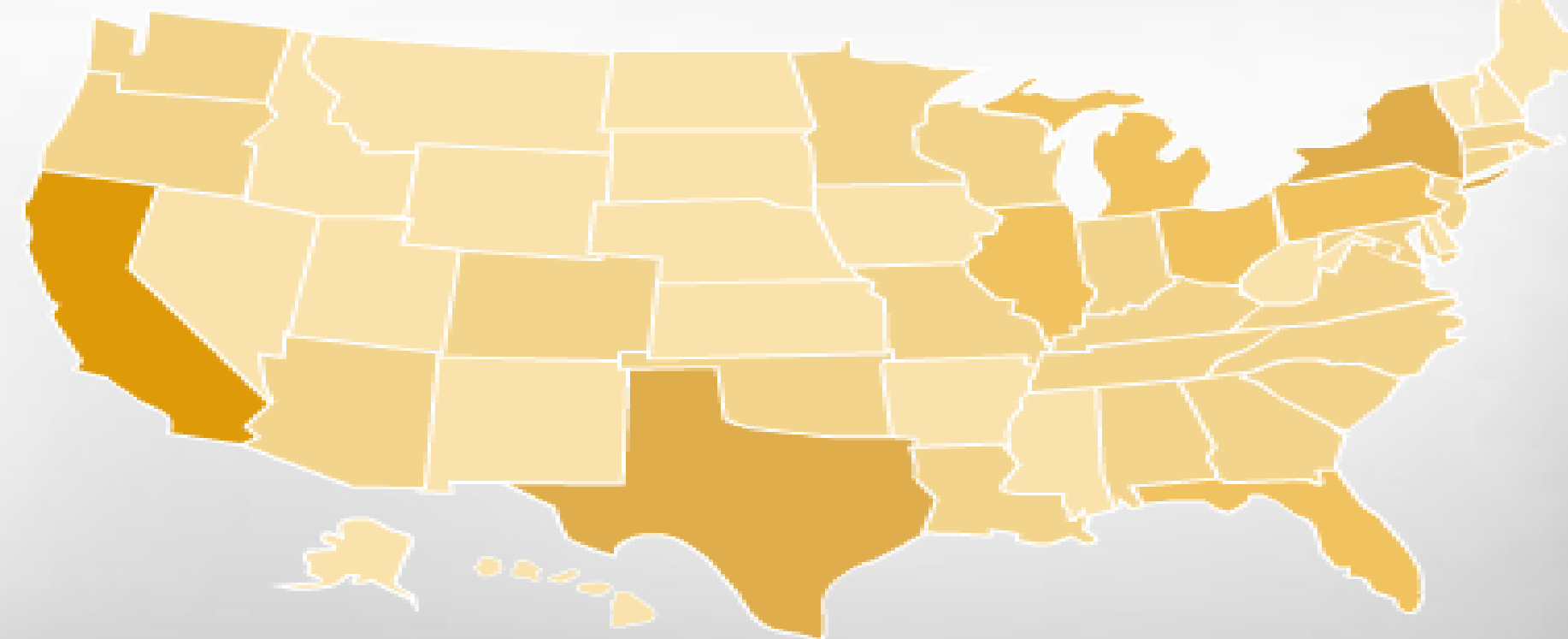
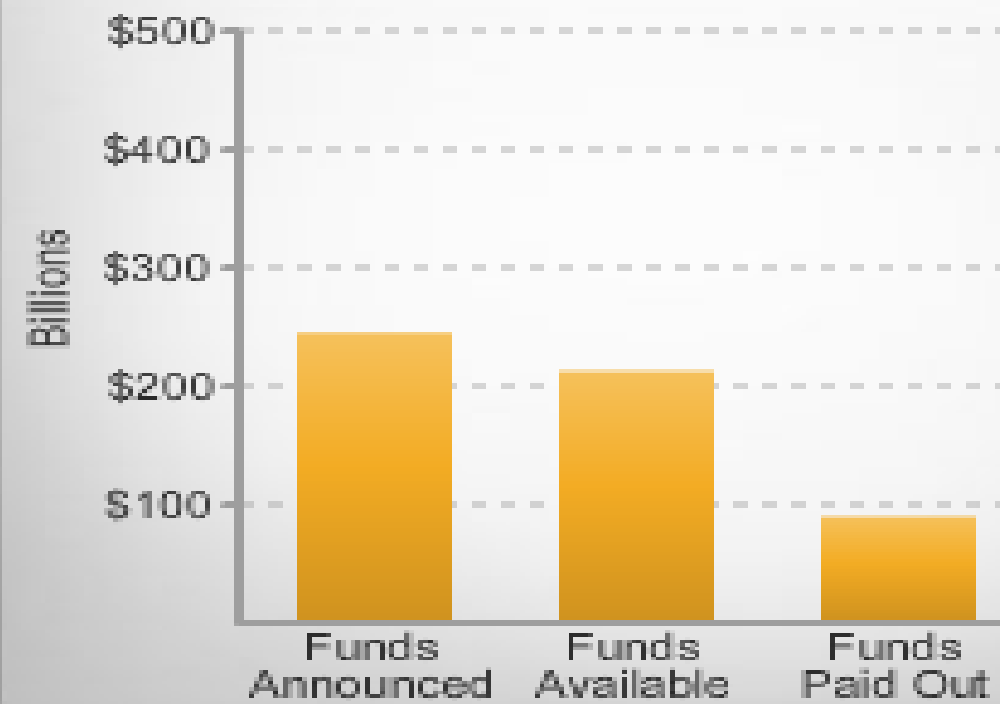
[HOME](#)[ABOUT](#)[ACCOUNTABILITY](#)[WHERE IS THE MONEY GOING?](#)[OPPORTUNITIES](#)[NEWS](#)[FAQS & RESOURCES](#)[CONTACT US](#)

AGENCY REPORTED

RECIPIENT REPORTED

U.S. TOTAL

for contracts, grants, loans and entitlements



Roll over graph and map to see Recovery data.
Click on any state for more data.

Updated: 09/21/2009

[Territories](#)[Text View of Data](#) [Source](#)

View Recovery Information in Your Neighborhood

Go

PA: WATER SYSTEM UPGRADES

September 25, 2009

Hazleton City Authority, PA has started an upgrade of its drinking water system as the result of a \$12.5M Recovery Grant. Source: Hazleton City Authority

Find out more about PA: Water System Upgrades >>



1 2 3 4 5 || ▶

ESTIMATED JOBS CREATED/SAVED

Recipient Data Will Be Released on October 15th, 2009

- » Find more information about Recovery related jobs.
- » How Job Data is Estimated - CEA (PDF 58kb)

A company receiving Recovery Act funding through a contract or grant and projects yet to be contracted out (Contracts Up for Bid) may offer job opportunities.

CONTRACTS

Source: [FedBizOpps.gov](#)

DATA, DATA & MORE DATA

To see where the money is going, make your selections below and click go.

Contracts ☐ Grants ☐ Loans ☐

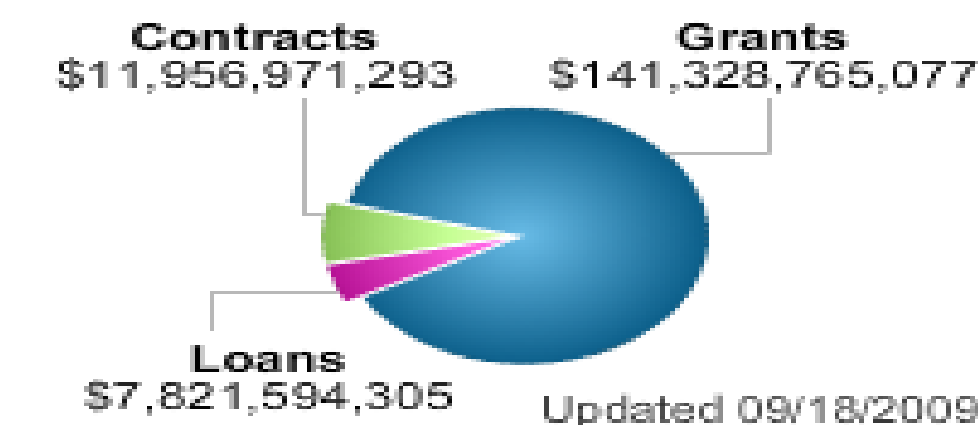
-- Agency --

-- State --

-- Amount --

Go

Funds Available by Award Type

[Source](#)

Updated 09/18/2009

» To access data to create your own reports visit the [Download Center](#).

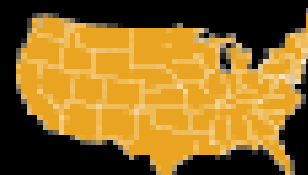
CHAIRMAN'S CORNER

Earl E. Devaney is chairman of the Recovery Accountability and Transparency Board, which manages this website and oversees spending under the American Recovery and Reinvestment Act of 2009.

September 29, 2009

After overcoming many challenges, the Recovery Board is launching its new state-of-the-art website. We plan to hold federal agencies and recipients of Recovery money accountable to you, our citizens. Please watch my video to



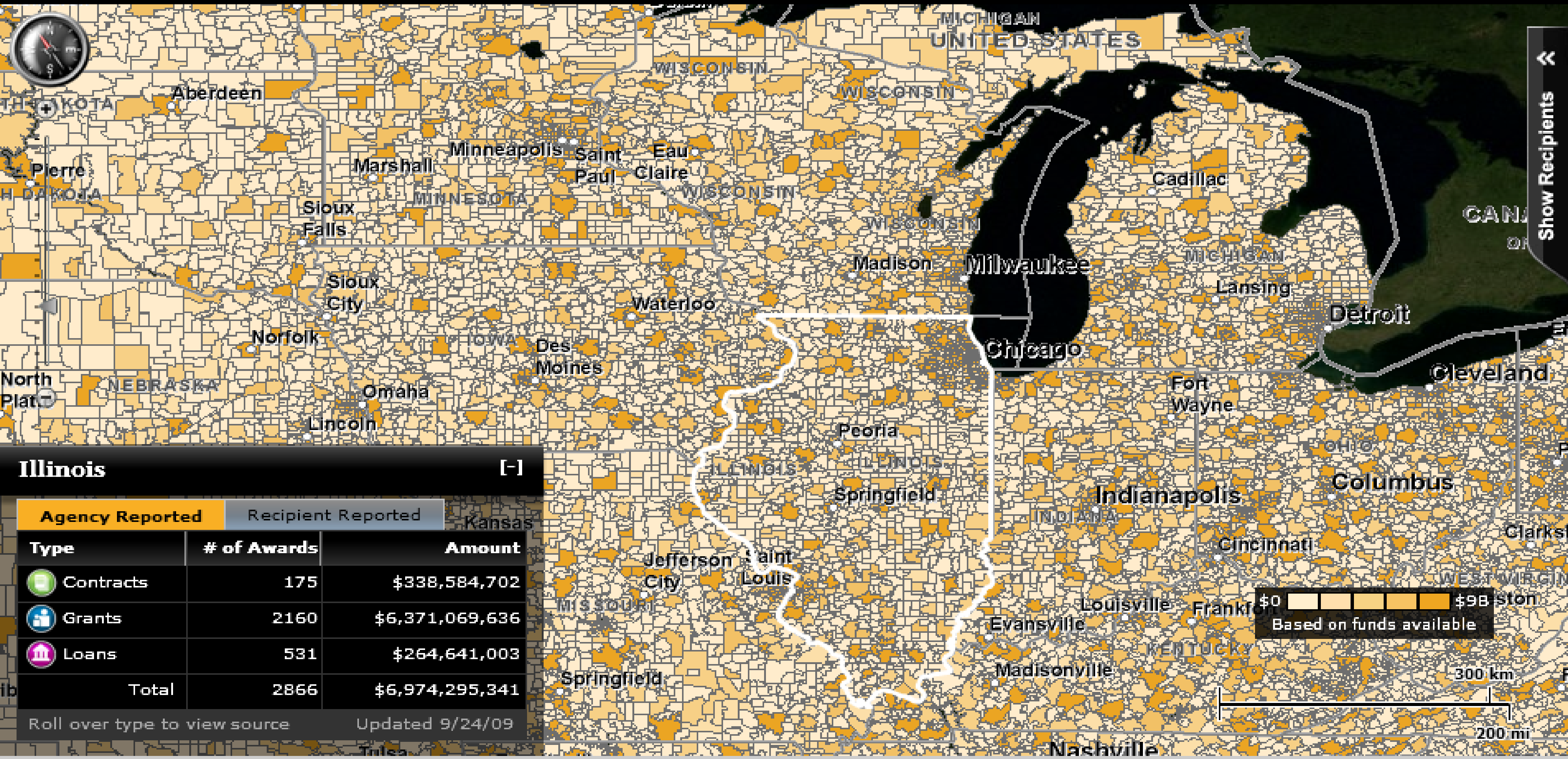


Map View:

Funding Map

Street Map

Aerial Map



Illinois

[1]

Agency Reported		Recipient Reported		Kansas	
Type	# of Awards		Amount		
Contracts	175		\$338,584,702		
Grants	2160		\$6,371,069,636		
Loans	531		\$264,641,003		
Total	2866		\$6,974,295,341		

Roll over type to view source Updated 9/24/09



ESRI



Microsoft®





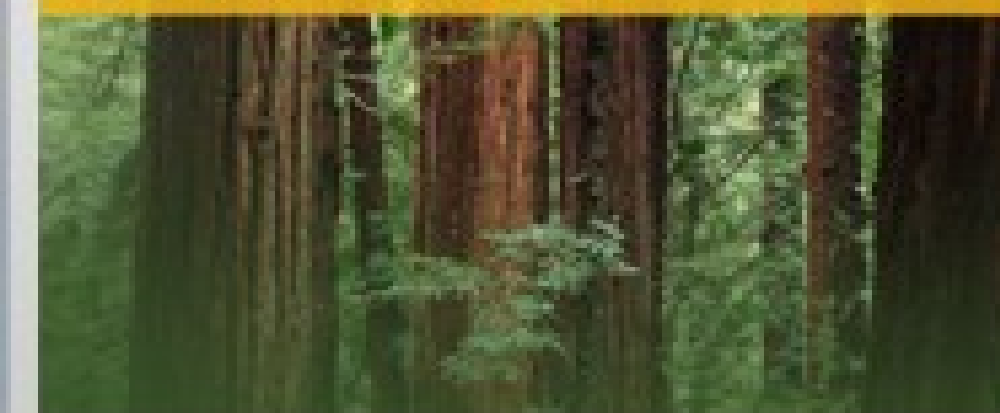
SAP CI LOGIN



SAP CARBON IMPACT OnDemand

BUSINESS SOFTWARE TO MANAGE ENERGY IN FACILITIES,
CARBON AND PRODUCT LIFECYCLE ASSESSMENTS

WHY SAP
CARBON IMPACT

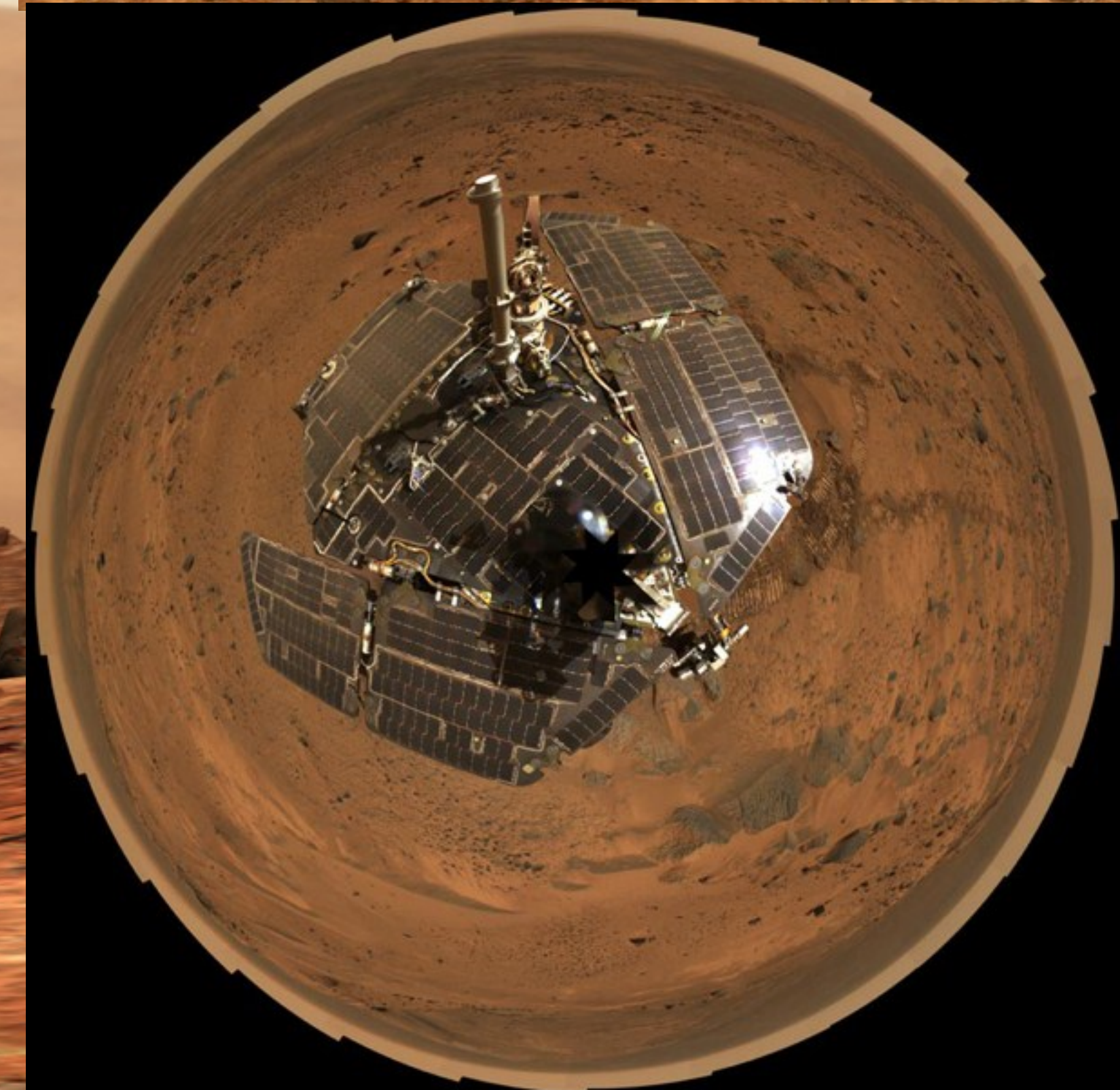
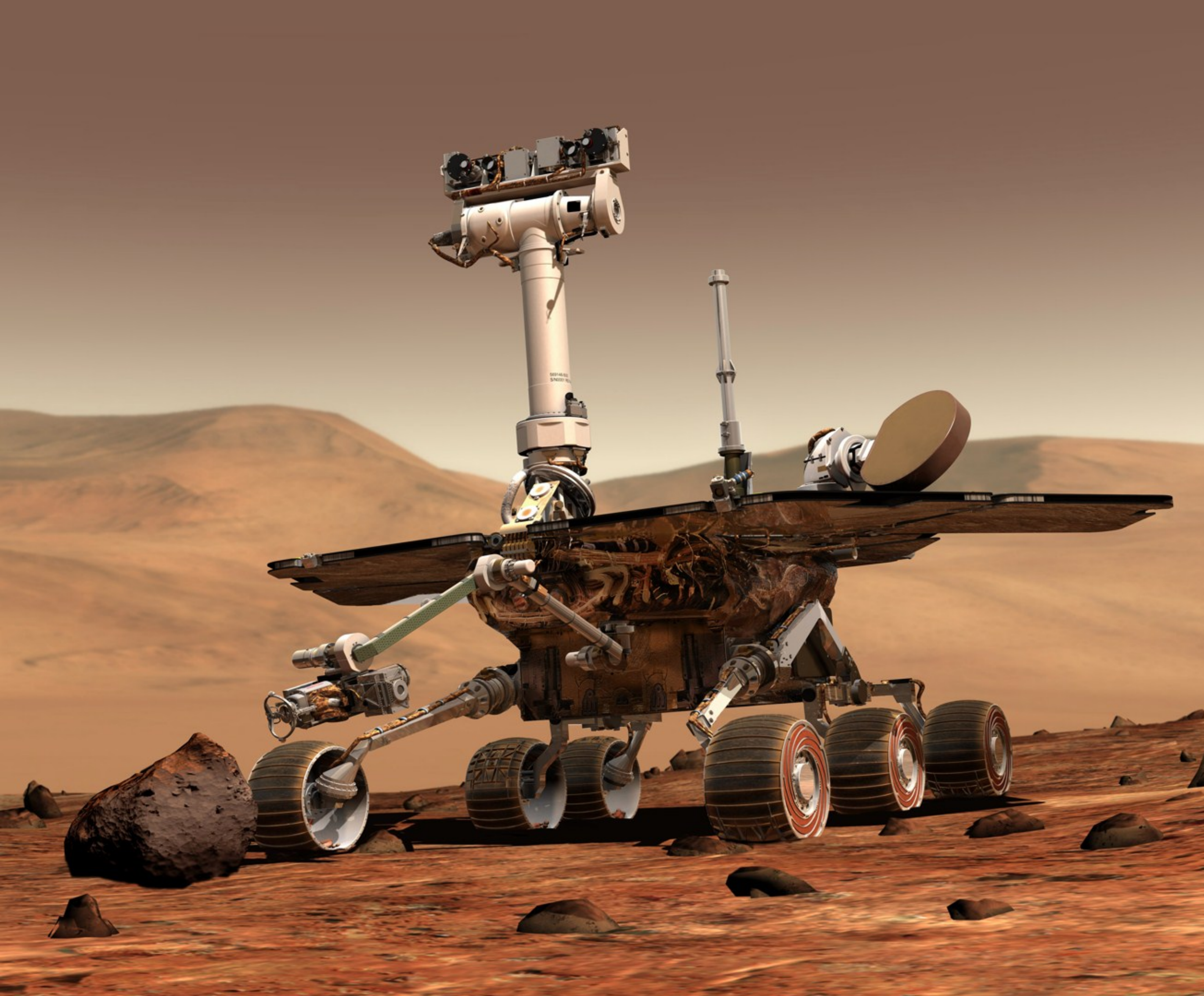


THE CARBON AND
ENERGY IMPERATIVE



SAP'S COMMITMENT
TO SUSTAINABILITY





COLLECT | STORE | ORGANIZE | ANALYZE | **SHARE**

A Request New Replay

Symbol:

Date:

Time:

Timespan:

☐ Quotes ☐ Trades ☒ All

- B** Replays
- GOOG 08/15/2008 10:15:00.000
 - SIRI 08/15/2008 12:00:00.000
 - GOOG 08/15/2008 12:00:00.000
 - NDAQ 08/18/2008 12:00:00.000



E Last Sale: 506.5700 Quantity: 200 Exchange: ADFN Time: 10:10:30.997

		Best Bid	Best Offer		
NBBO	200	506.18	506.58	400	NBBO
NSDQ	200	506.18	506.58	300	NSDQ
CINN	300	506.17	506.58	100	ARCX
ARCX	100	506.15	506.59	100	CINN
ADFN	500	496.10	510.00	10000	ADFN
CBOE	0	0.00	510.22	100	ISEG
CHXE	0	0.00	999,999.00	0	CBOE
ISEG	0	0.00	999,999.00	0	CHXE

Timespan 10:05:00.000 10:24:59.999

F Trades

High: 509.2745 Trades: 1,787

Low: 505.5000 Shares: 262,375

VWAP: 506.7651

Time	Price	Quantity	MC
10:10:27.400	506.6300	100	D
10:10:27.730	506.4000	100	D
10:10:27.730	506.3900	100	Q
10:10:27.730	506.3900	200	Q
10:10:27.730	506.3900	300	Q
10:10:27.777	506.4000	100	D
10:10:28.103	506.4000	495	D
10:10:28.980	506.2300	100	Q
10:10:29.527	506.4500	100	Q
10:10:29.527	506.4900	500	Q
10:10:29.543	506.5000	100	Q
10:10:29.543	506.5000	100	Q
10:10:30.933	506.5700	100	C
10:10:30.997	506.5700	200	D
10:10:31.933	506.5700	100	Q
10:10:31.933	506.5800	100	Q
10:10:31.933	506.5800	100	Q
10:10:31.933	506.5800	100	Q

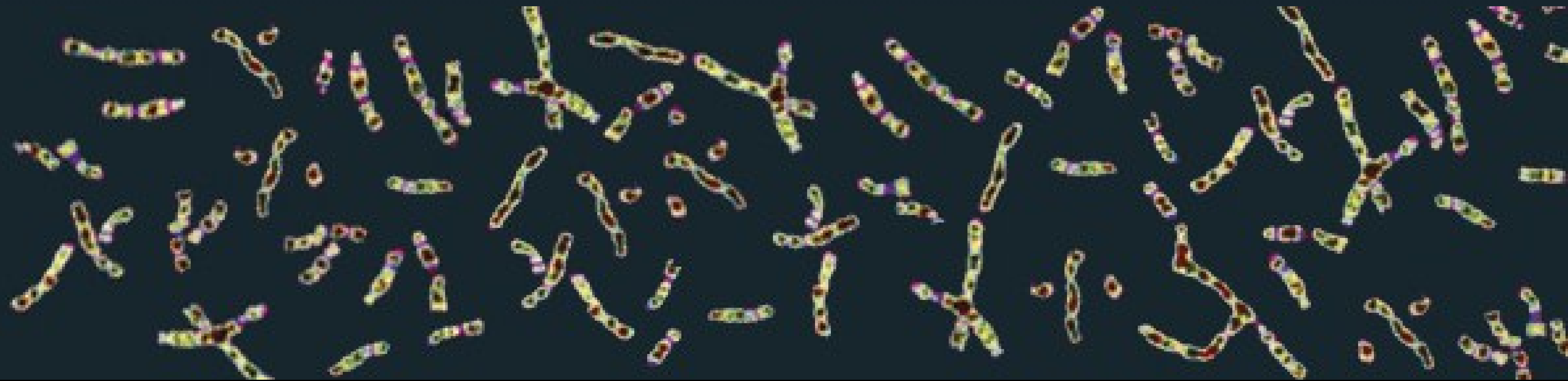
G Quotes

	Low Bid	High Bid	Low Ask	High Ask
NBBO	505.50	508.28	505.72	508.29
ARCX	505.50	508.28	505.72	508.42
NSDQ	505.42	508.20	505.73	508.50
ADFN	496.10	506.19	507.01	510.00
CINN	490.01	508.05	505.72	508.49
CBOE	0.00	506.54	506.82	999,999.
ISEG	0.00	506.38	510.22	510.22
CHXE	0.00	0.00	999,999.	999,999.

<http://aws.amazon.com/publicdatasets>

1000 Genomes

A Deep Catalog of Human Genetic Variation



Big Data Verticals

Media/ Advertising

Targeted
Advertising

Image and
Video
Processing

Oil & Gas

Seismic
Analysis

Retail

Recommendations

Transaction
Analysis

Life Sciences

Genome
Analysis

Financial Services

Monte Carlo
Simulations

Risk
Analysis

Security

Anti-virus

Fraud
Detection

Image
Recognition

Social Network/ Gaming

User
Demographics

Usage
analysis

In-game
metrics

COLLECT | STORE | ORGANIZE | ANALYZE | SHARE