

From Information to Foresight:

Getting Beyond the Bits



Data sources are growing, and new sources are emerging

2 Billion Internet users by 2011



Twitter processes **7 terabytes** of data every day



World Data Centre for Climate

- **220 Terabytes** of Web data
- **9 Petabytes** of additional data

New Sources of Information



Facebook processes **10 terabytes** of data every day



4.6 Billion Mobile Phones World Wide



UPS supports largest known peak database workload of **1.1 Billion** SQL statements per hour



74 Terabytes UK Land Registry - Largest known transaction processing DB



Traditional Sources of Information

The Challenge: Bring together large volumes of diverse data to find new insights



- 

Multi-channel customer sentiment and experience analysis
- 

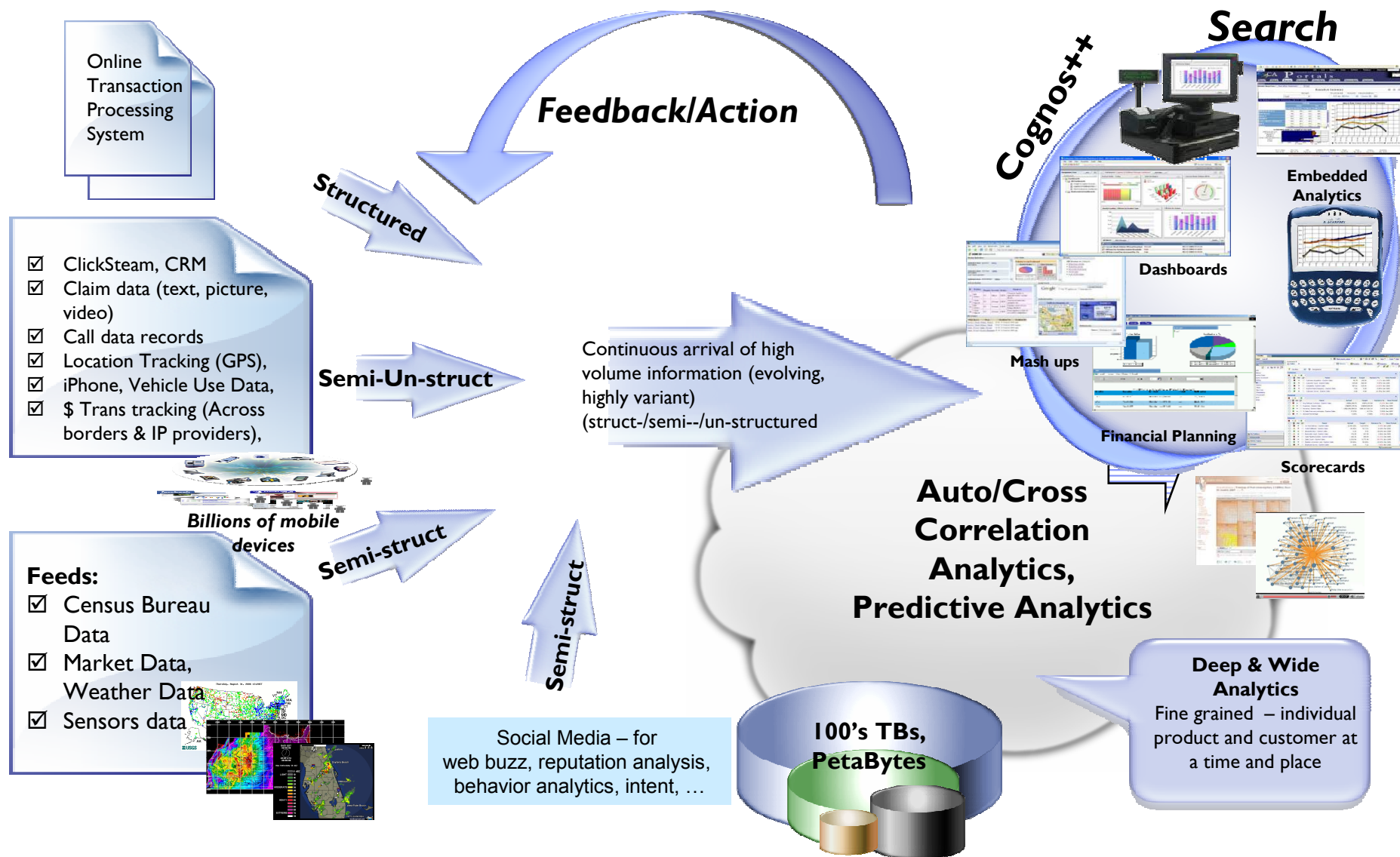
Detect life-threatening conditions at hospitals in time to intervene
- 

Predict weather patterns to plan optimal wind turbine usage, and optimize capital expenditure on asset placement
- 

Make risk decisions based on real-time transactional data
- 

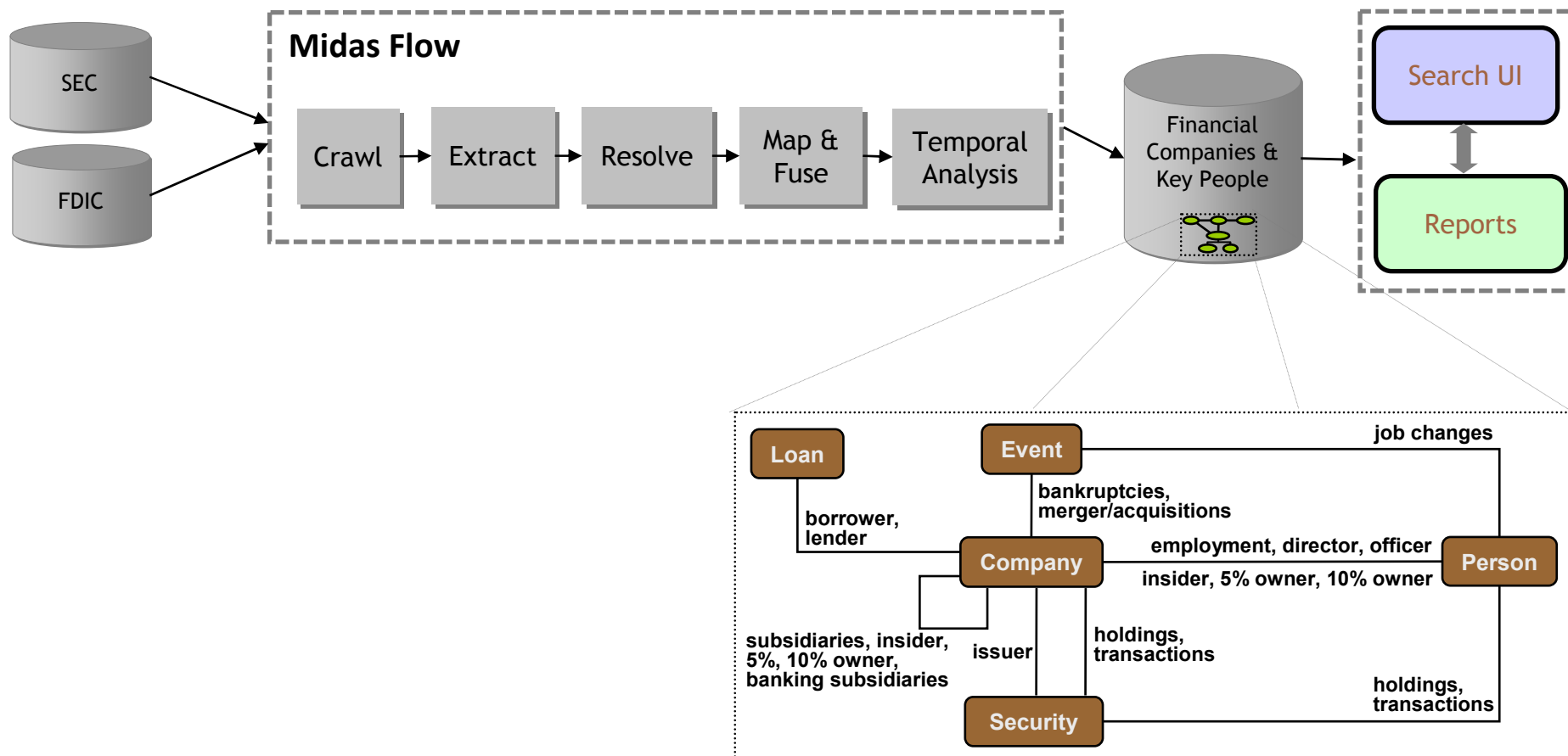
Identify criminals and threats from disparate video, audio, and data feeds

Data analytics in the enterprise means grappling with increasingly complex data and sophisticated analytics – faster than ever



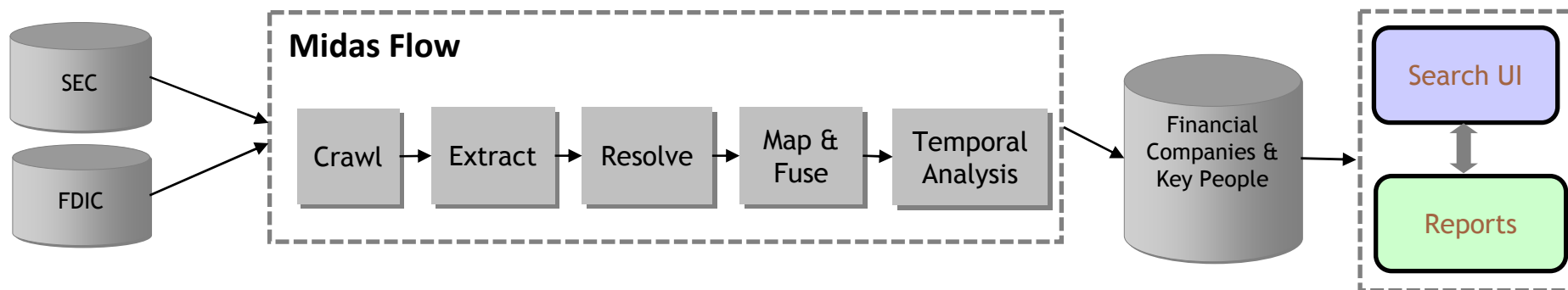
Preparing data for analysis itself requires analytics

Q: "How exposed am I to my borrowers?"



Preparing data for analysis itself requires analytics

Q: "How exposed am I to my borrowers?"



Annual Report

Loan Agreement

Proxy Statement

Insider Transaction

Unstructured, Data spread over many documents!

Counterparty Relationships

Loan Exposure

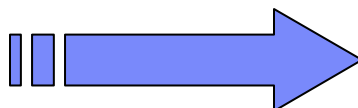
Integrated Entities and Relationships

Loan Title	Total Amount of Loan (\$ in billions)	Agreement Date
Credit Agreement (364-Day Commitment)	800,000,000	2009-06-12

Programmable analytics are needed for extraction and resolution

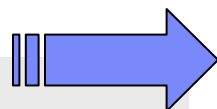
Metadata :

Tracy Guida
@tracyguida Tampa
just a Nor-Cal gal trying to fall in love with Florida



Screen name : @tracyguida
Location: Tampa, Fl
Name: Tracy Guida
IsParent: true
Gender: female
Relationships:
 <@thelorenski, acquaintance >
 <@SparklingLala, acquaintance>
 <@smileycookie, network>
Interests: sports: (Yoga, running, Steelers), food:(Burger21)

Messages :



so blessed to have a son that loves school! [#backtoschool](#)

[@tracyguida](#) Thanks for being a loyal fan, Tracy! :) [#Burger21](#)

thanks again for coming in last night, looking forward to Yoga Sunday

[@SparklingLala](#) great! I will be there Nov 10-15 for a conference, will keep in touch

[@thelorenski](#) [@HecMan44](#) I am coming to SJ in Jan, be warned :

Sore after my run this morning

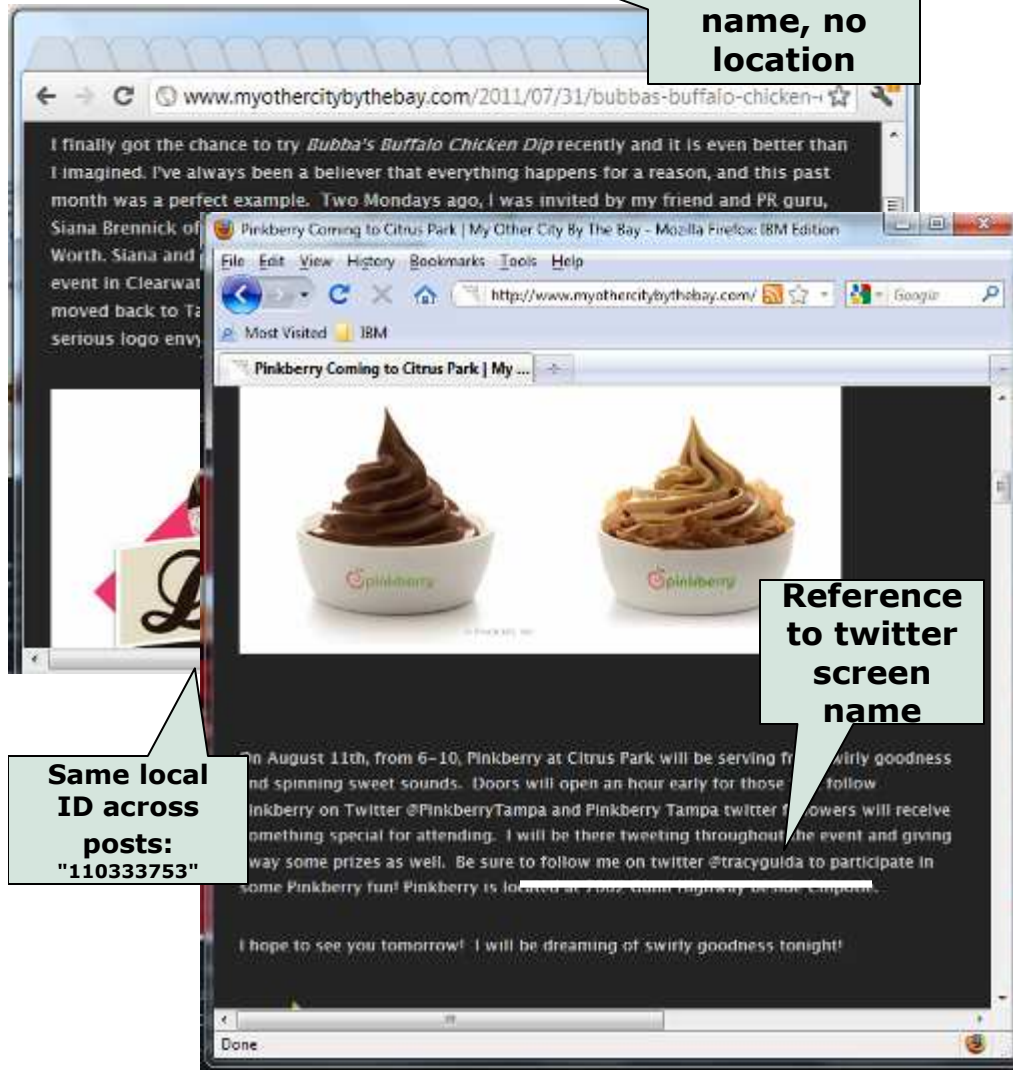
[@smileycookie](#) absolutely! I am the biggest Steelers fan living in Tampa. Going to share these cookies with fellow Steelers fans. :)

Programmable analytics are needed for extraction and resolution

boardreader™

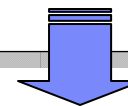
Profile Name: "Champagne Girl"

No valid name, no location



twitter

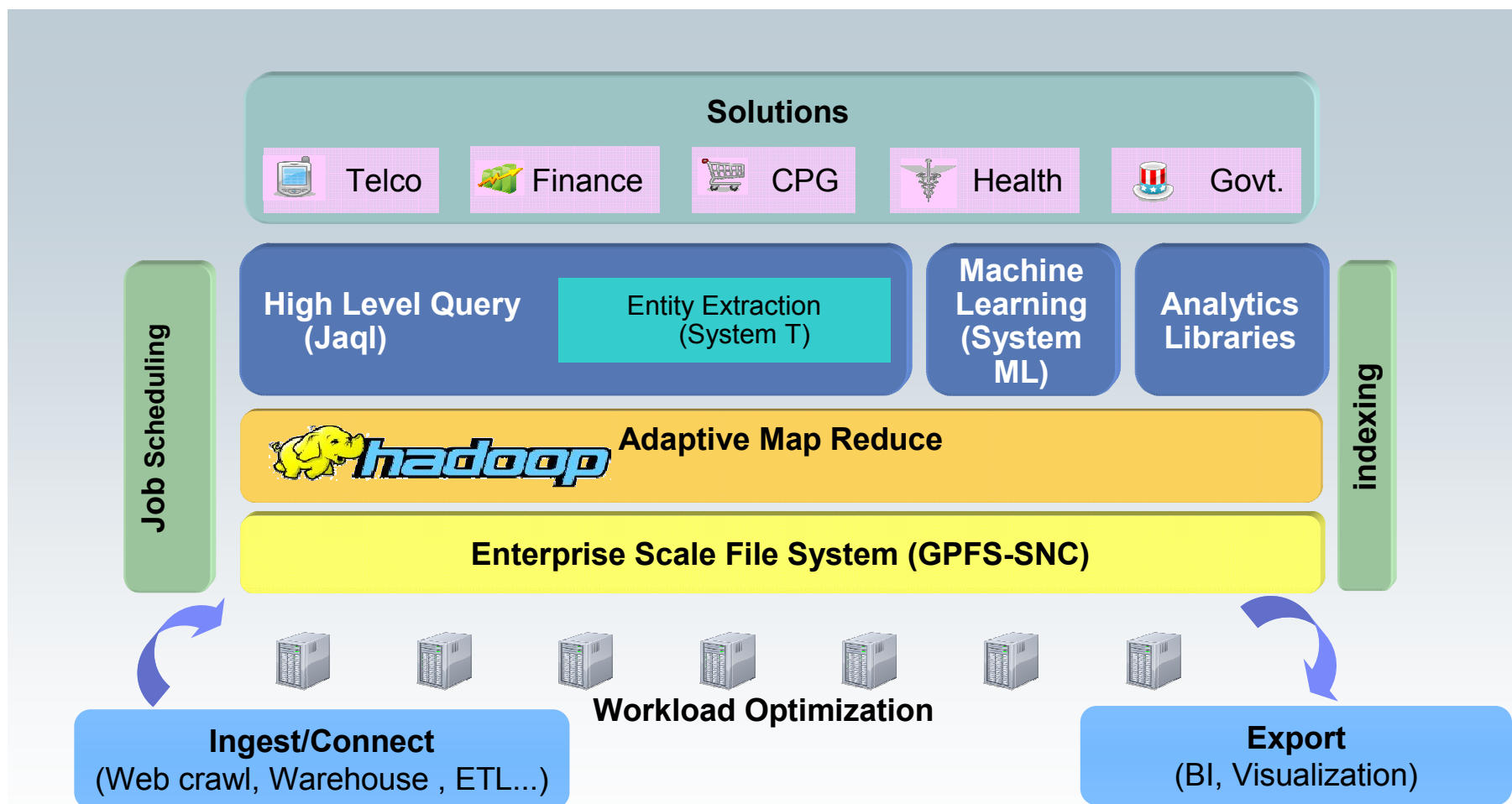
Screen name : @tracyguida
 Location: Tampa, Fl
 Name: Tracy Guida
 IsParent: true
 Gender: female
 Relationships:
 <@thelorenski, acquaintance >
 <@SparklingLala, acquaintance>
 <@smileycookie, network>
 Interests: sports: (Yoga, running, Steelers), food:(Burger21)



Global Profile:

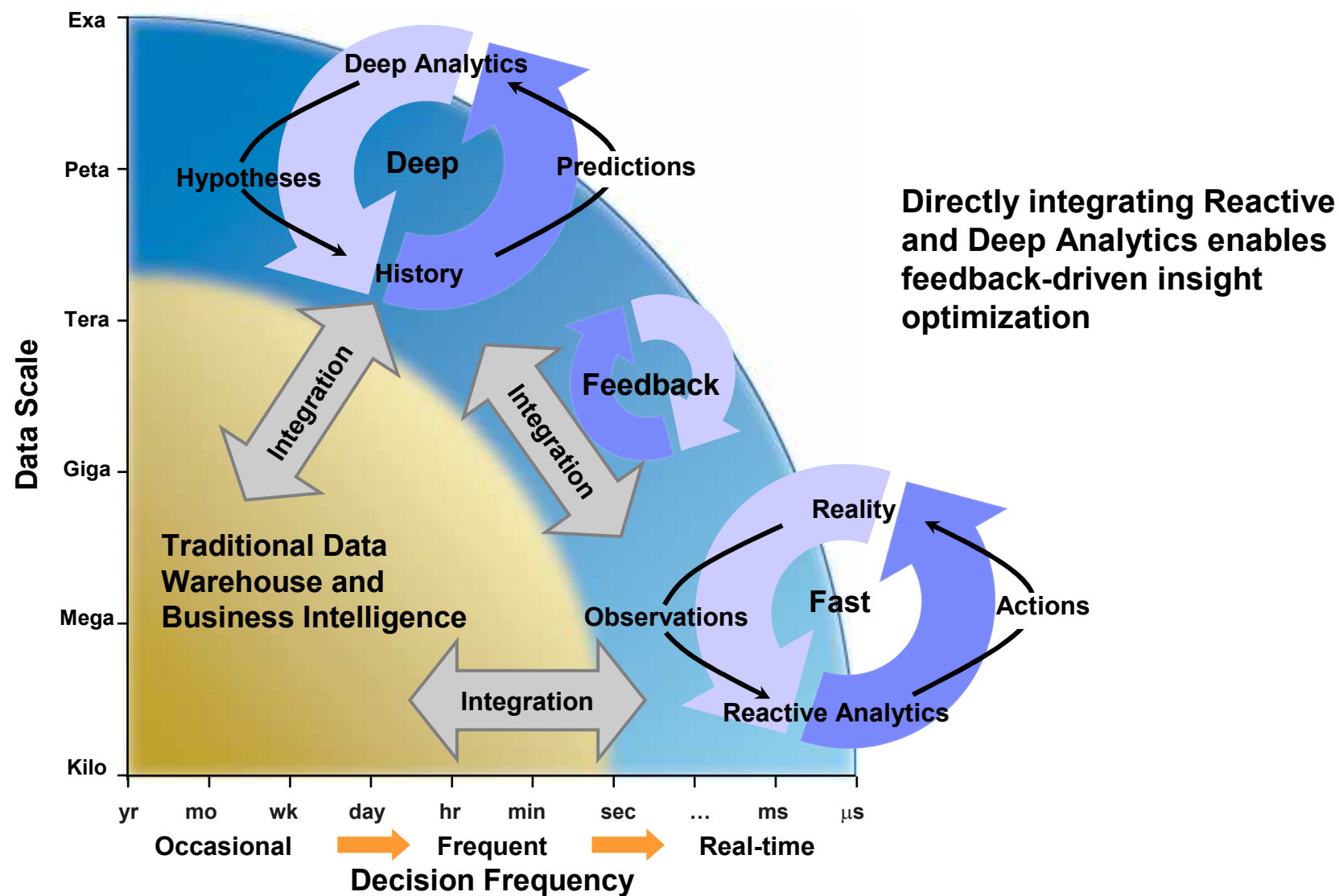
Screen name : @tracyguida
 Location: Tampa, Fl
 Name: Tracy Guida
 IsParent: true
 Gender: female
 Relationship:
 <@thelorenski, acquaintance >
 <@SparklingLala, acquaintance>
 <@smileycookie, network>
 <Siena Brennick, friend>
 Interests: sports: (Yoga, running, Steelers), food:(Burger21, Pinkberry, wine, Bubba's Buffalo Chicken Dip)
 Blogs: Source: myothercitybythebay.com, Text ="...."

Such analytics require a rich platform with design and runtime tools

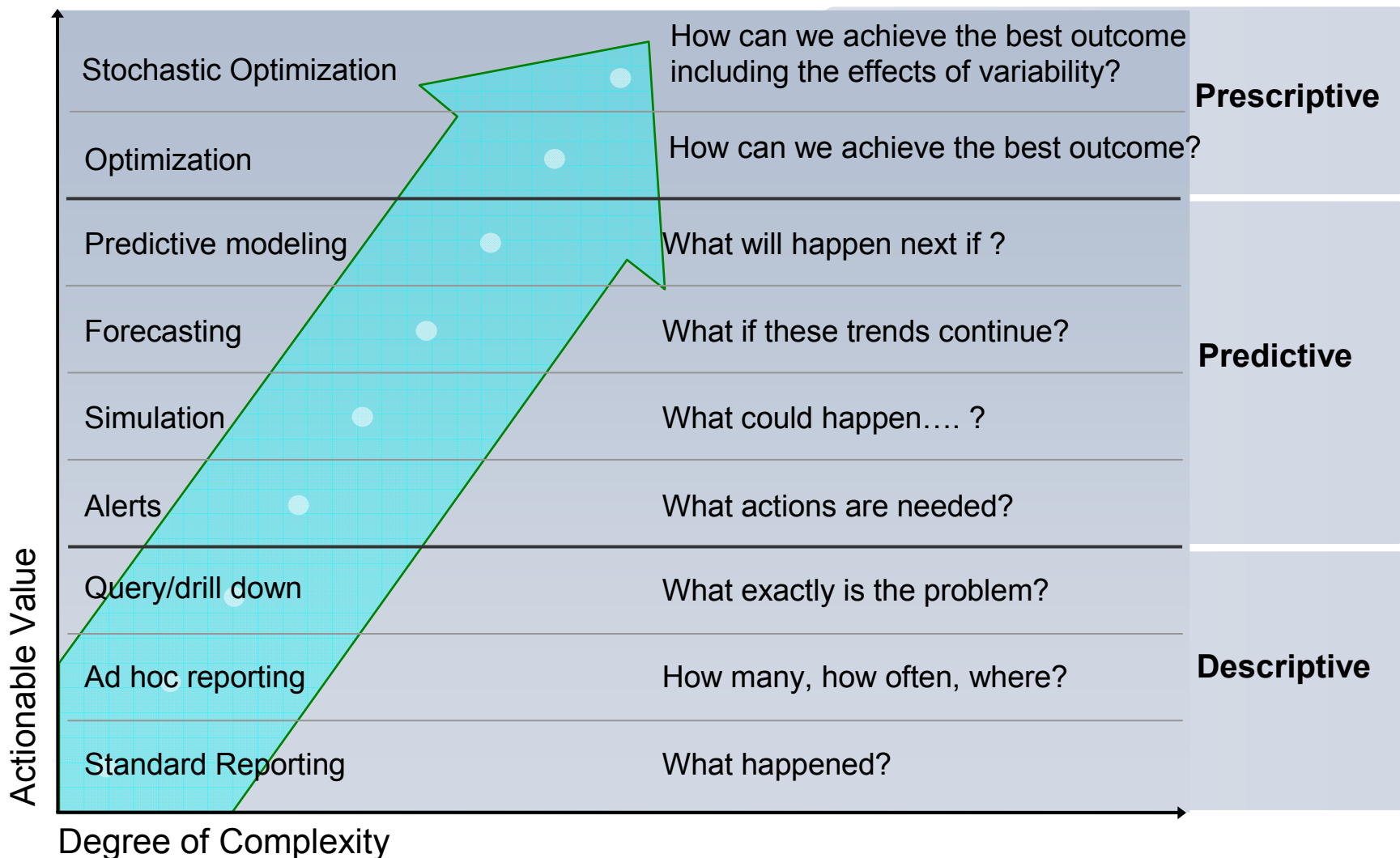


Innovation above, underneath, in and around Hadoop

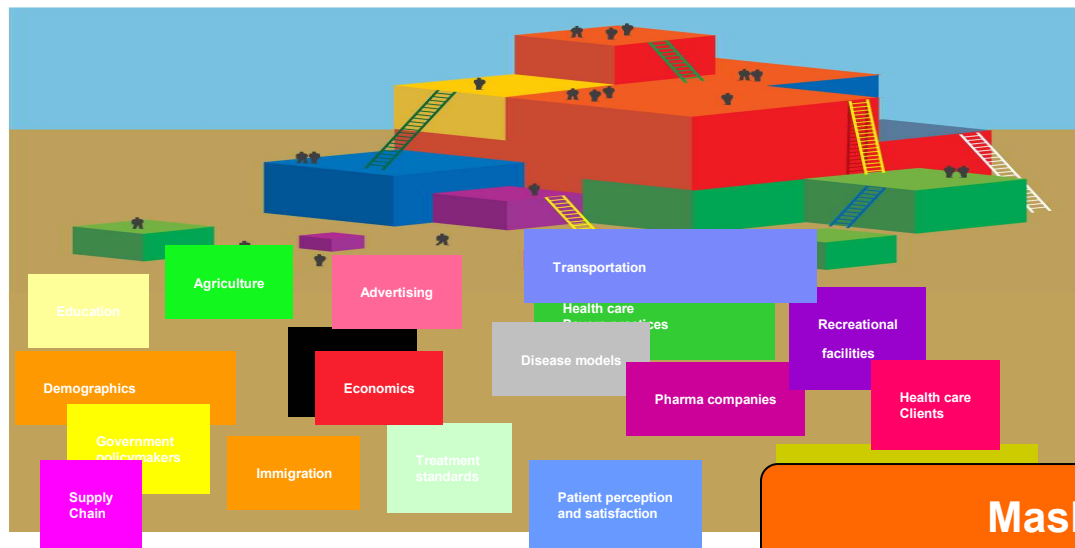
Maximum insight requires combining deep and reactive analytics



There are many kinds of analytics, often domain-specific



Today's challenges require integration of models and analytics



SPLASH: Smarter Planet Platform for Analysis and Simulation of Health

Mashup BMI estimation

Access to Physical activities
Queuing model

Store Location Model
Optimization

Transportation model
Time series

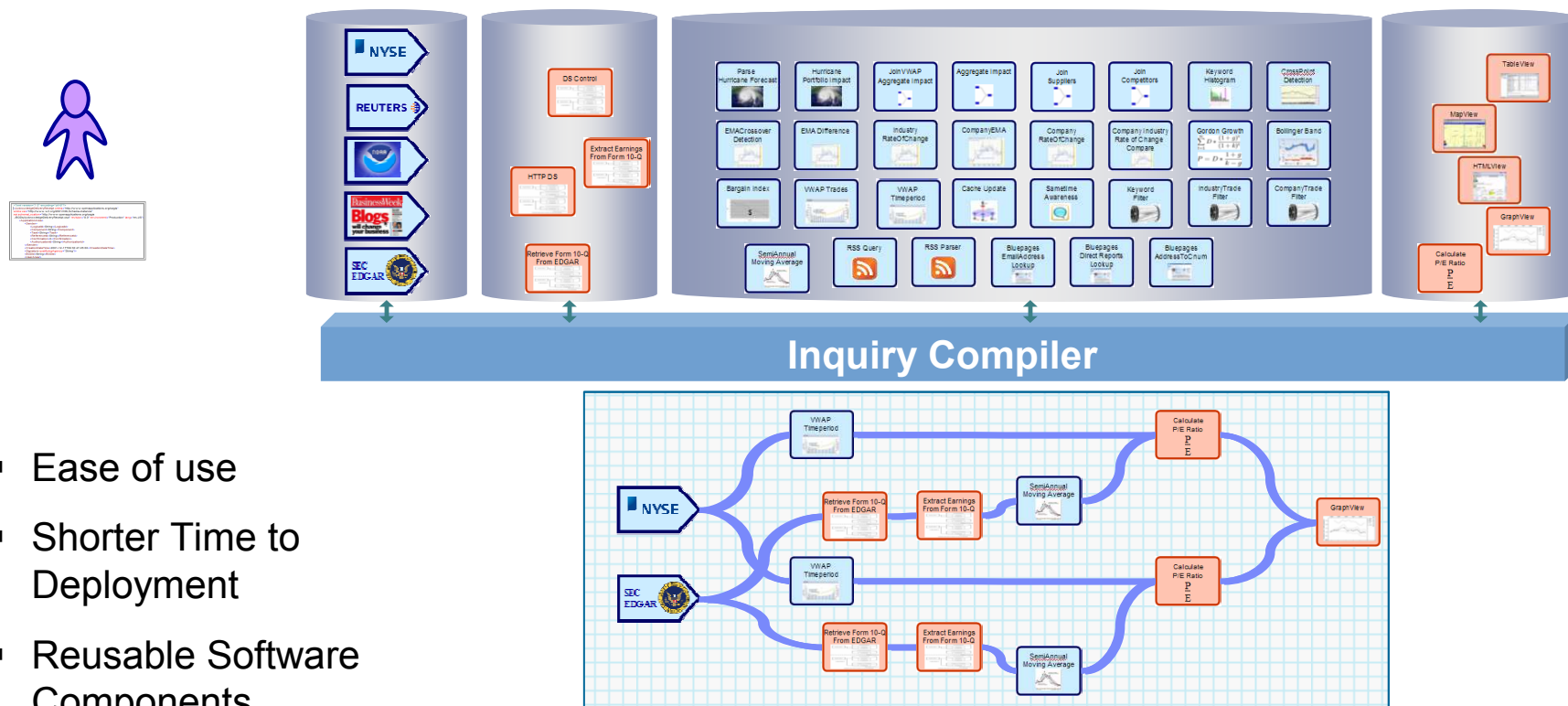
Eating & Exercise Model
Agent based

City Incentive model
Formula

Demographic, behavioral, GIS data sources for initialization of models

Identify previously unclear connections between parts of complex healthcare systems

Combining analytics is easier with the help of a compiler



- Ease of use
- Shorter Time to Deployment
- Reusable Software Components

Visualization and collaboration are key to understanding data

Logged in as **Matthew M. McKeon** | Logout

explore
 visualizations
 datasets
 comments
 my stuff

participate
 create a visualization
 upload a data set

learn more
 about Many Eyes
 about us
 FAQ

contact
 contact us

Try Our Featured Visualizations!

<p>US government expenses 1962-2004</p> <p>Where are your tax dollars going? by Frank van Ham</p>	<p>Highway vs. City Mileage</p> <p>Which cars get better mileage in the city? by Martin Wattenberg</p>	<p>Causes of Death</p> <p>by Martin Wattenberg</p>	<p>US Meat Consumption per Capita, 1909-2004</p> <p>Meat eating habits are changing dramatically in the US. by Fernanda B. Viegas</p>
---	--	--	---

Featured Data Sets:
 Motor Vehicle Fatalities per Million People
 Selected Cereal Nutrition Data

The best way to see your data, ask questions, and share your discoveries. Harness the collective intelligence of the net for insight and analysis.

many eyes
 for shared visualization and discovery

<http://www.many-eyes.com>

IBM Many Bills
 A Visual Bill Explorer

Browse Collections | My Bill Tray | Sign In | Help | Blog | Search Search Bills... Go

Many Bills
 is a Visualization of Congressional Legislation

Each bill is represented as a column of color coded sections. Colors correspond to topics we assign to sections, offering insight into the content of bills.

Search Bills
 Search our database of bills for a set of words or phrases.
 Search Bills... Go
 Advanced...

Recent Searches
 learning communities
 disaster response
 federal bureau of investigation
 tobacco programming
 111b parity
 marriage
 mass transit
 grant research computer science

Your Congressperson
 Choose a state: **Massachusetts**

Sen. Scott Brown (Rep)
 See more of Scott's Legislative Activity

Sen. John Kerry (Dem)
 See more of John's Legislative Activity

Enter your zip code to find your district representative. ##### Go

Recently Added Bills
 in the News | Hot Bills | Recently Active

H.R. 627
 Credit Card Accountability Responsibility and Disclosure Act of 2009
 Sponsored by President

H.R. 760
 Advanced Broadband Infrastructure Bond Initiative of 2009
 Introduced by House

S. 425B IS
 Growing Safe Food Act of 2009
 Introduced by Senate

H. R. 3624
 Poison-Free Poultry Act of 2009
 Introduced by House

Featured Collections:
 A collection is a set of one or more bills, that is saved as a unit with a title and description. Make your own collection and share it with others!
 More Featured Collections...

American Housing Bills (42)
 on Law FL on Apr 10 2010
 All bills currently impacting the residential real estate industry.

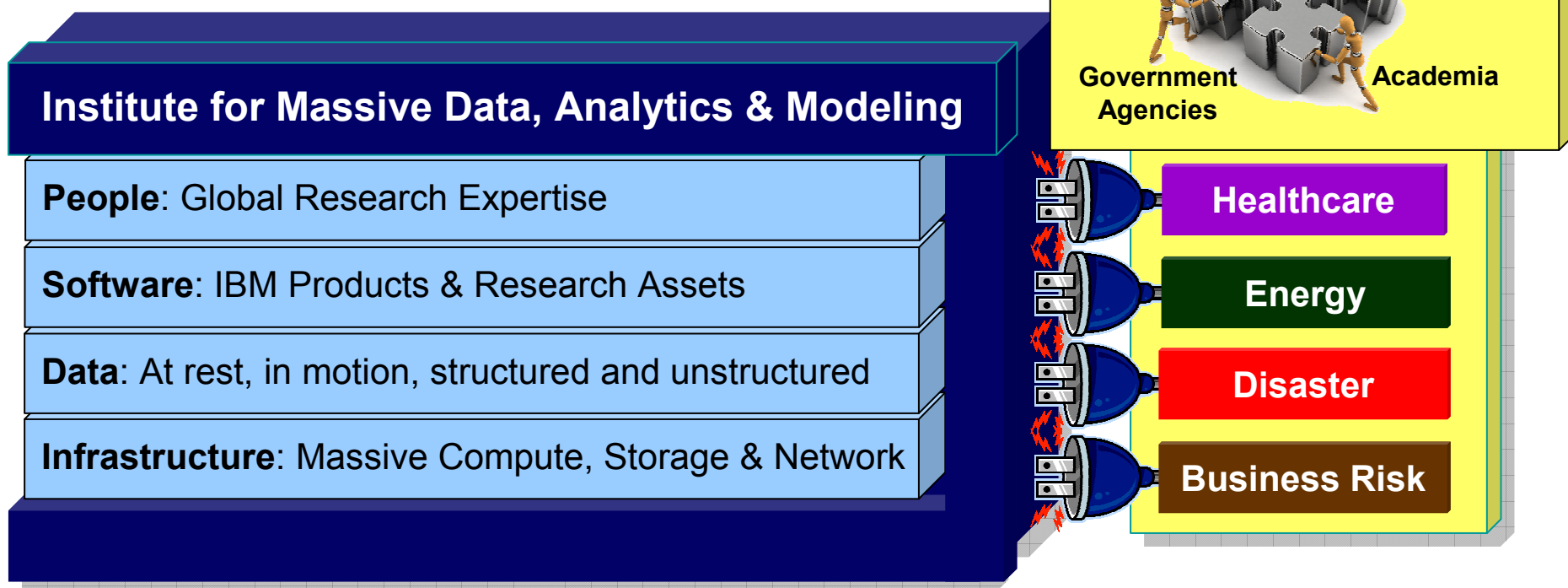
Feedback

<http://manybills.us>

The Institute for Massive Data, Analytics and Modeling

Taking Analytics to the Limit

- Enable solution of real-world problems
- Drive new breakthroughs across stack
- Leverage unique IBM (Research) capabilities
- World-class platform for data-intensive research



Summary

- Better collection of data affords many opportunities
 - To do better science
 - To improve our world
 - To make more money 😊
- To benefit, we need to do more than collect data!
 - Data preparation is a challenging problem
 - Tooling and runtimes are both important
 - Model integration is a key emerging challenge
 - Supporting collaboration and understanding both data and analysis is essential
- One of the key challenges for gaining insight is “integrating” and enabling people
- The Institute for Massive Data, Analytics and Modeling offers exciting possibilities
 - For computer scientists: drive new innovation in information management & mining
 - For domain researchers: leverage data for new discoveries in that field
 - For the world: new techniques to solve big problems
 - Through bringing together data, technology and people with diverse skills