

# The Big Data Juggling Act



Paul Boal

2016 Amitech Solutions

[paul.boal@amitechsolutions.com](mailto:paul.boal@amitechsolutions.com)

# Remember Why Data Warehousing Started?

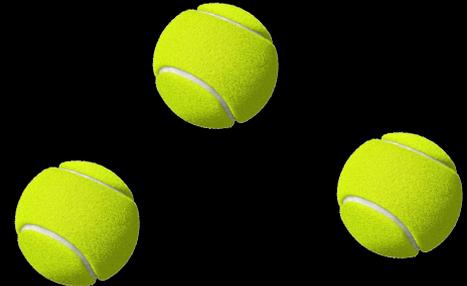
*To make complex analysis and decision making processes more efficient, consistent, and reliable.*

Guide and automate decision making

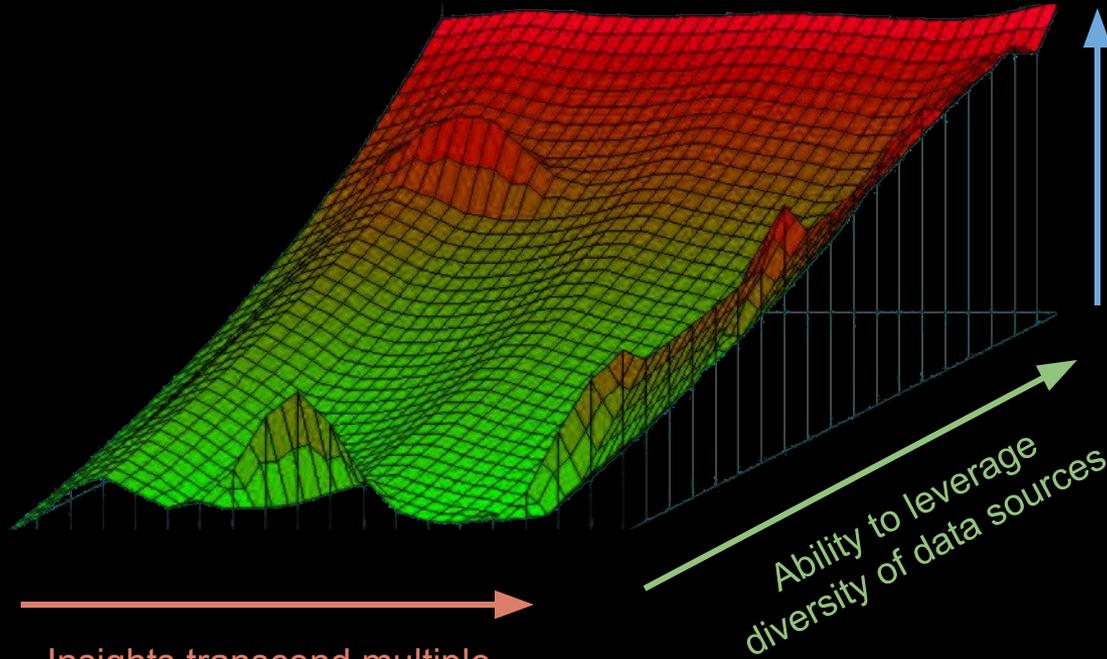
Use technology and methods to make this easier.

Look at data in consistent ways.

**Achieved Maturity!**



# Is There Anything Special About Big Data?



Insights transcend multiple  
business areas  
simultaneously

Ability to leverage  
diversity of data sources

Traditional  
Process  
Related  
Maturity



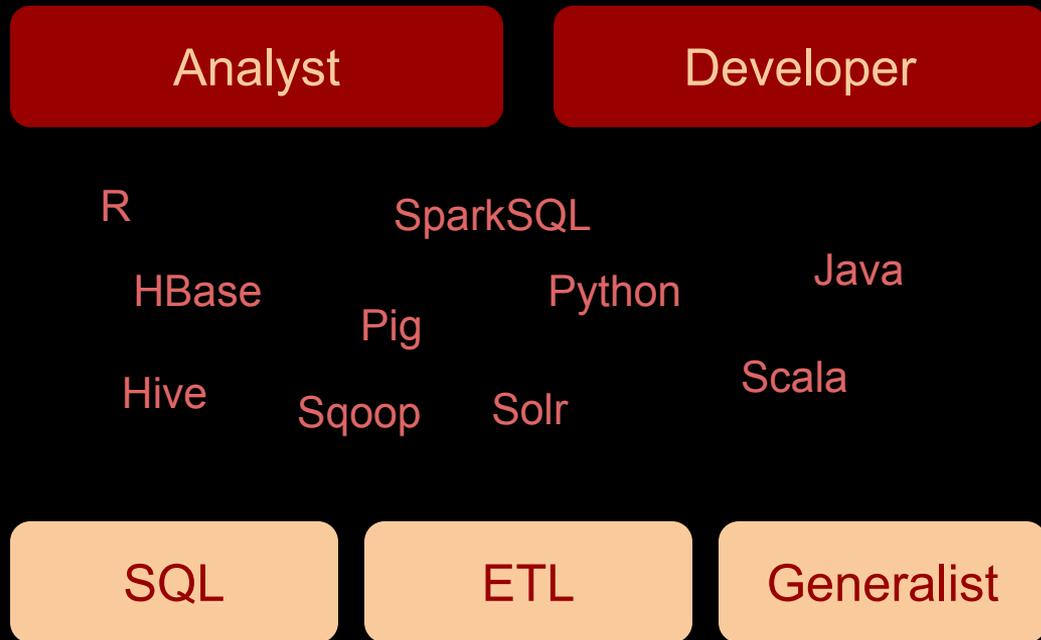
# Learn the Technology



# Learning Big Data



- Why do some learning first?
  - Set realistic expectations
  - Climb the learning curve
  - Accelerate productivity
- How to learn big data
  - Tutorials
  - Community
  - Get in and do it
- Who can learn big data
  - SQL developers
  - ETL developers
  - General programmers





# Learning Big Data - Homework

1. Download a sandbox or set up a small cloud instance
    - Hortonworks Sandbox, Microsoft Azure HDInsight, Qubole
  2. Find a problem you're interested in and load the data
    - Just load the files as is
    - Do any preprocessing you need to on the platform
  3. Run some simple analysis
    - Try Hive SQL and Pig
    - Try Spark via Scala or Pyspark
    - Try out Apache Zeppelin
- **Even if you already know how to do these things in another tool!**
    - The point isn't to prove that big data tools are best at solving this particular problem.
    - The point is to learn the technology and tools.
  - **St. Louis Meetups: Hadoop, Machine Learning, Python, R**

# Learning Big Data - Homework Examples



## 1. Public Healthcare Data

- Is there any relationship between cost of care and the quality of care?
- If there appears to be one, see if you can explain why?
- How about a relationship between the ratio of the cost care at a given hospital to median household income in that area and the quality of care?
- Pull together a whole slew of regional demographics and see if there are any correlations.
- Can you build a simple predictive model that will suggest cost of care in an area?

## 2. Public Crime Data

- Is there any relationship between amount or type of crime and the forecasted weather?
- Do you have any hypotheses on why that might be? How could you test those?
- Are there any correlations between crime and the type of businesses in a particular region?

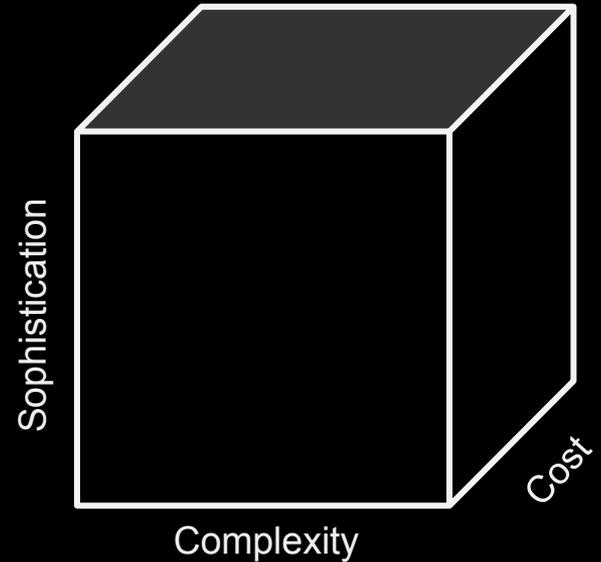
# Know the Business Value



# Where can value in big data be found?



- **Healthcare**
  - Right Living, Right Care, Right Provider, Right Value, Right Innovation
- **Logistics & Machined Industries**
  - Predictive Maintenance
- **Consumer Centric Businesses**
  - Customer Behavior, Precision Marketing, Purchasing Patterns
- **Intelligence and Security**
  - *I can't tell you*
- **Information Technology**
  - Platform Consolidation, Scalability, Cost

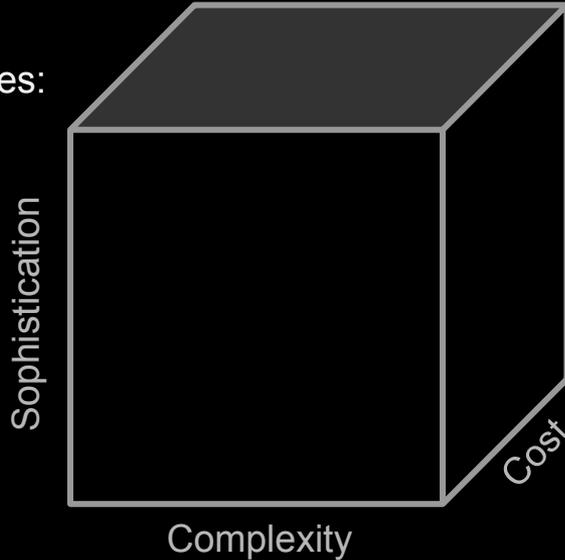


# How is value created?



That's a complex question, it requires:

- Advanced statistics
- Machine learning
- Predictive analytics
- Iterative modeling
- Hundreds of dimensions



If it didn't cost so much, we could:

- Keep that data online
- Process that faster
- Support additional users
- Consolidate platforms
- Give analysts more space

The data we have for that analysis is:

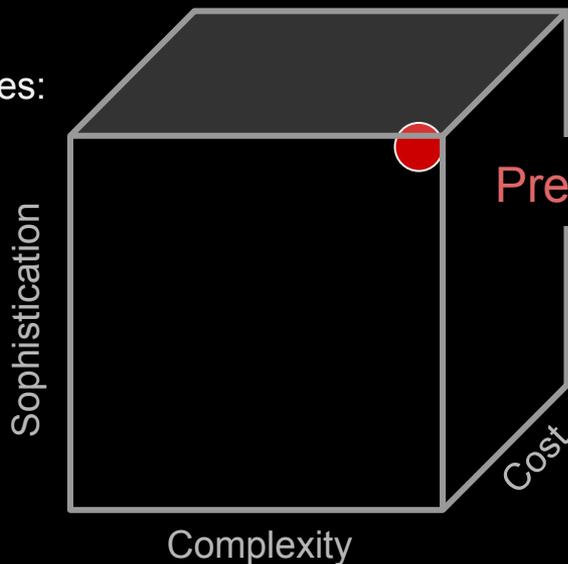
- Unstructured text
- Image based
- Streaming by quickly
- Huge in volume

# e.g. Precision Medicine



That's a complex question, it requires:

- **Advanced statistics**
- **Machine learning**
- **Predictive analytics**
- Iterative modeling
- **Hundreds of dimensions**



Precision Medicine

If it didn't cost so much, we could:

- **Keep that data online**
- **Process that faster**
- Support additional users
- Consolidate platforms
- Give analysts more space

The data we have for that analysis is:

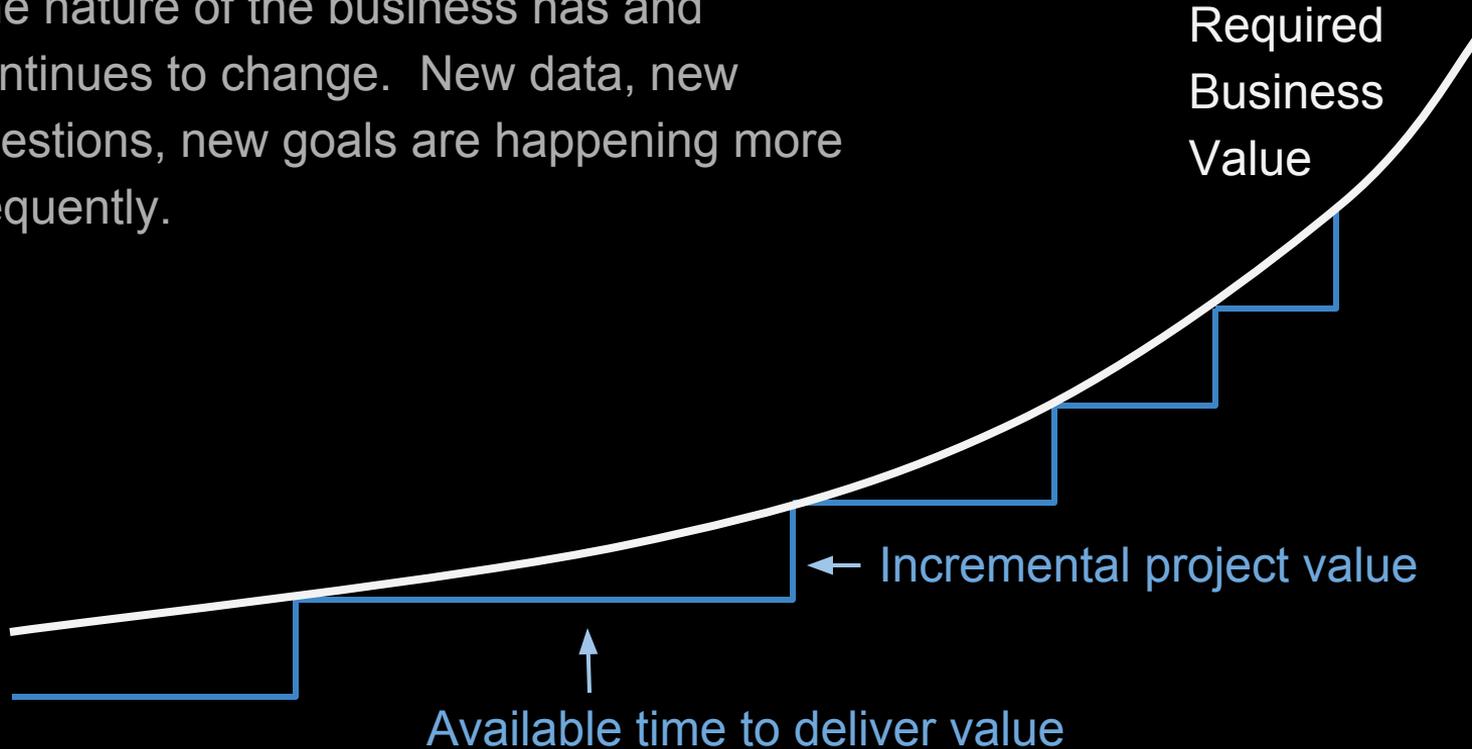
- **Unstructured text**
- **Image based**
- Streaming by quickly
- **Huge in volume**



Develop a  
New Way to  
Develop

# Have to deliver faster

The nature of the business has and continues to change. New data, new questions, new goals are happening more frequently.



# Techniques to accelerate value

- Methodology, Tools, and Architecture
  - Agile principles and methods
  - Rapid development and devops automation tools
  - Generalized architectures: data lake/library
- Organizational Responsibilities and Capabilities
  - Lab and data exploration environments
  - Flexibility and diversity of skills
  - Shared responsibilities / distributed teams
- Knowledge Sharing
  - Collaboration tools: slack
  - Lunch and learns
  - Easy knowledge bases: wiki





# Improving Development - Homework

- What are you doing that causes the most delays?
  - Waiting for others
  - Tasks that can only be done by one person
  - Mandatory process delays
- What are you doing that could be automated?
  - Show the business value of improving and automating your own processes
- What are you doing that others could learn to do?
  - Enable them with tools
  - Don't underestimate their abilities
- Do the “what if so-and-so got hit by a bus” test



# Change the Organization

# Everyone Loves Change!



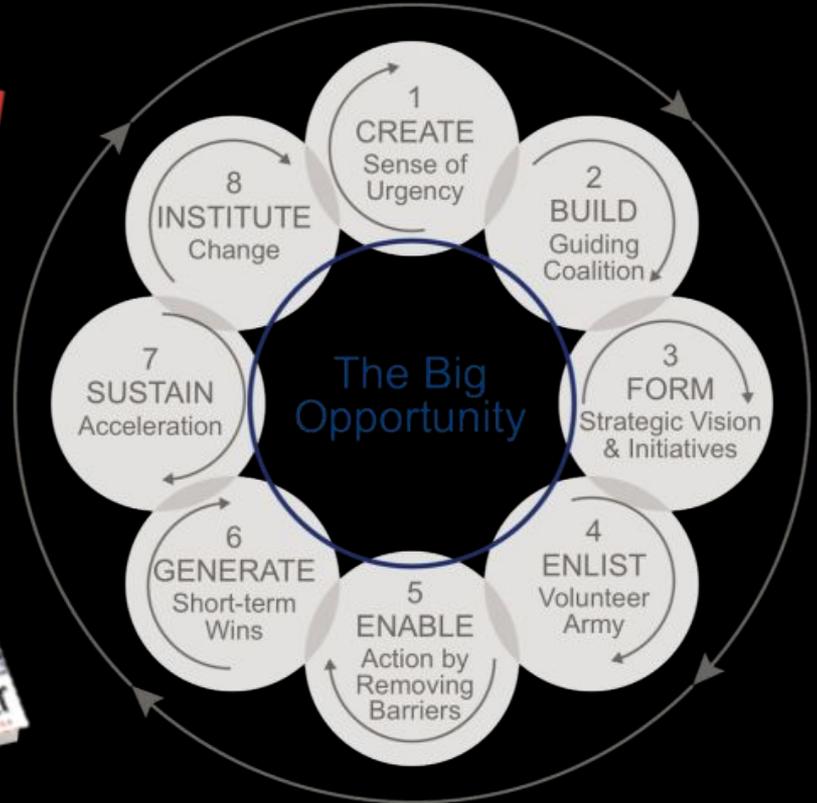
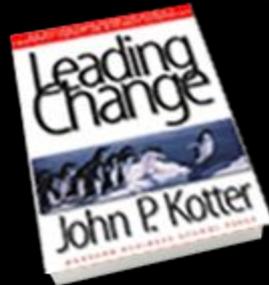
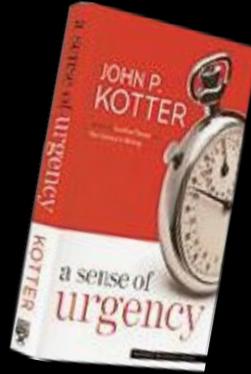
“It’s not that people don’t like change. It’s just that they only like change when it’s their idea.”



# Organizational Change Takes Work

## John P. Kotter

- Sense of Urgency
- Guiding Coalition
- Strategic Vision
- Enlist Volunteers
- Enable Action
- Short Term Wins
- Sustain Acceleration
- Institutionalize Change



# Make it Fun!

Invest time in building a real demo that's specific to your organization and uses your real data.

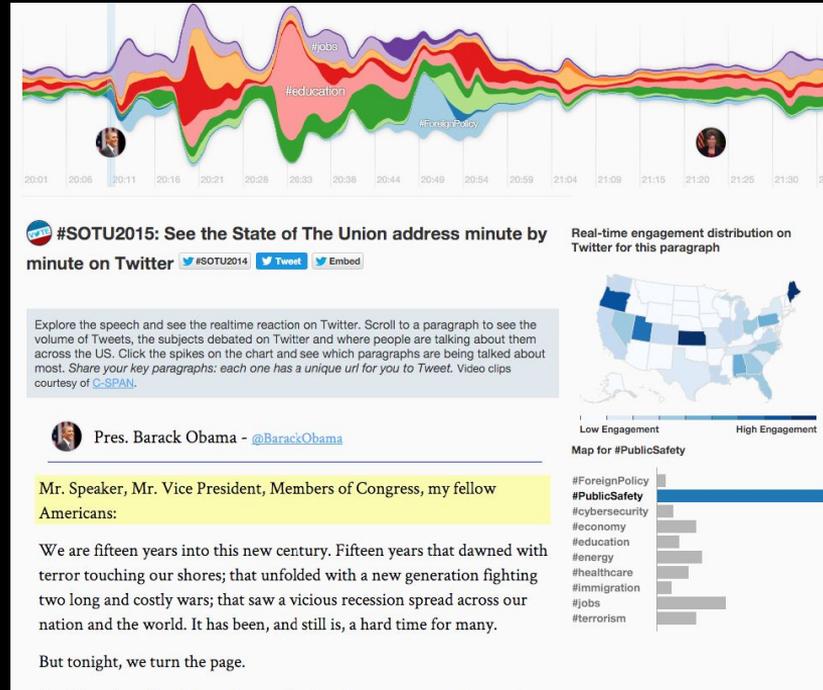
Make it fun and interesting even if it isn't entirely practical.

**MySearch**

Search



# Real-Time Analysis of State of the Union





# Organizational Change - Homework

- Read Kotter
  - Invest the time in building a demo that you can show off
  - Ask others “what would your ideal day look like?”
  - Develop a vision of how things could be
  - Collaborate with others to build energy and support
  - Don’t be afraid to fail, but do it quickly and learn
- Be Patient, Persevere

Why we're doing this hasn't really changed.

Learn new technologies.

Figure out where business value comes from.

Develop new ways to develop.

Change the organization.

