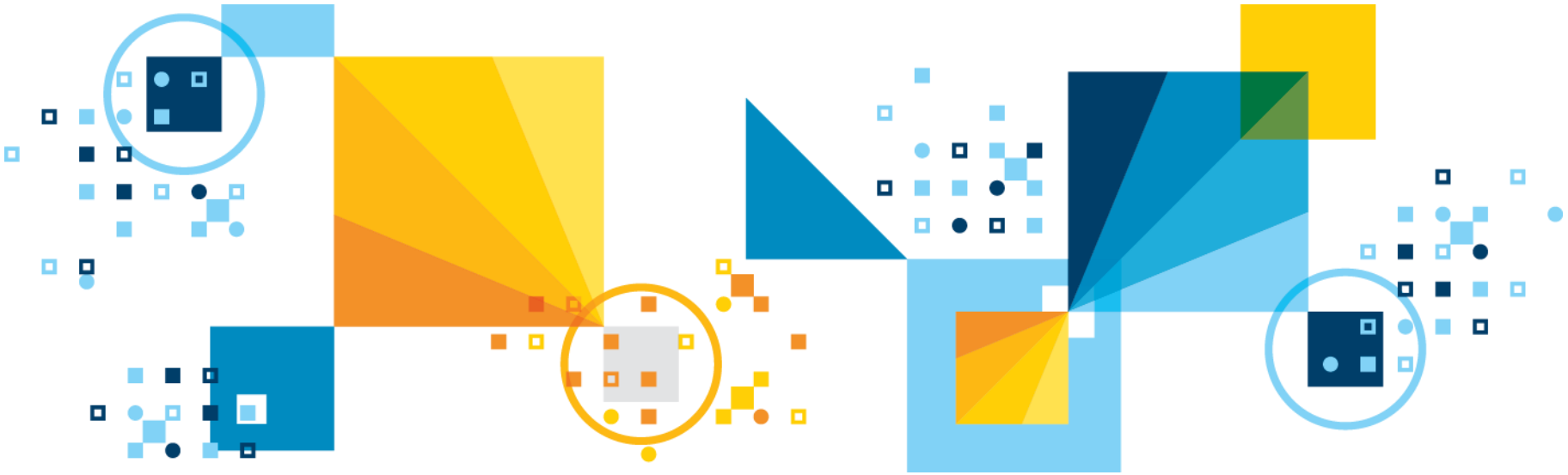
Data Science and Data Scientist

Dr. Alex Liu, Principal Data Scientist



Data Science Example

Google Flu Trend Analytics

Detecting outbreaks
two weeks ahead
of CDC

Estimating which cities are
most at risk.

Data Science Example

elections2012

Live results | President | Senate | House | Governor | Choose your

Numbers nerd Nate Silver's forecasts prove all right on election night

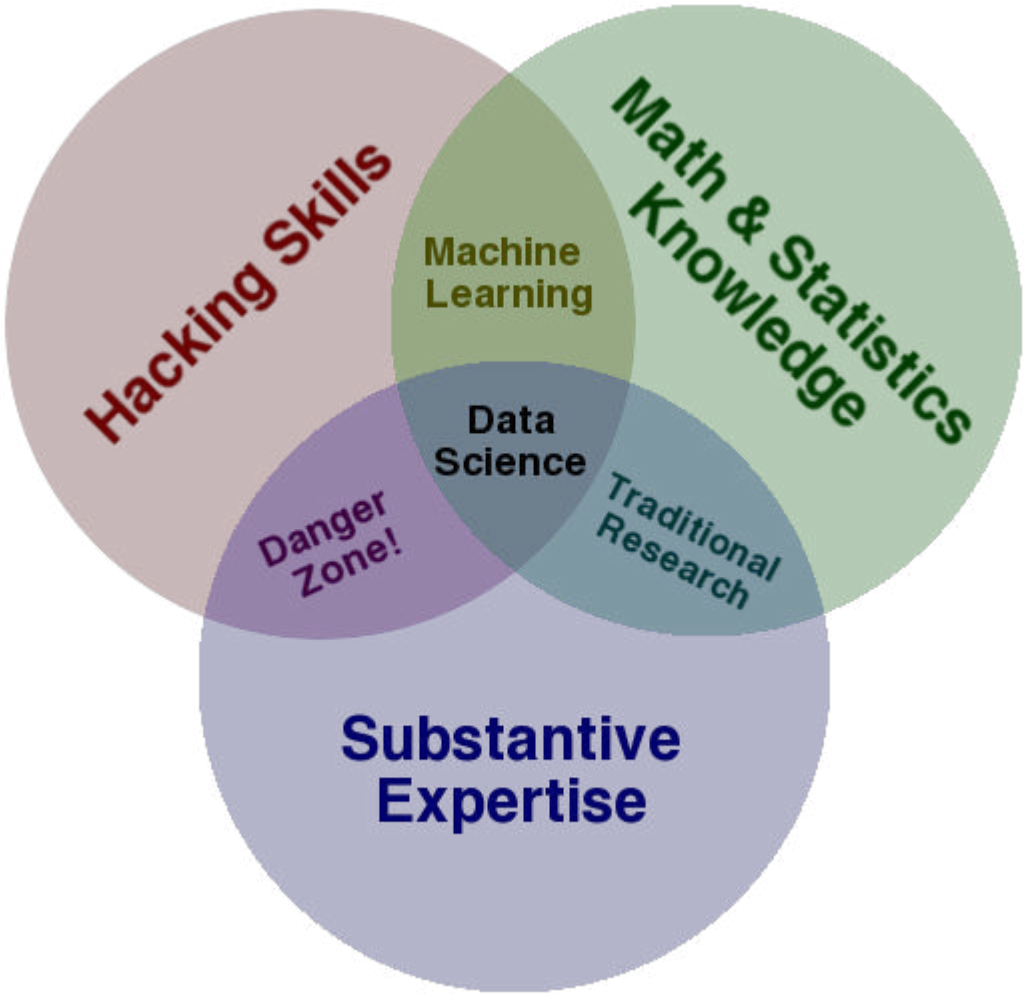
FiveThirtyEight blogger predicted the outcome in all 50 states, assuming Barack Obama's Florida victory is confirmed

Luke Harding
guardian.co.uk, Wednesday 7 November 2012 10.45 EST

More data science examples …

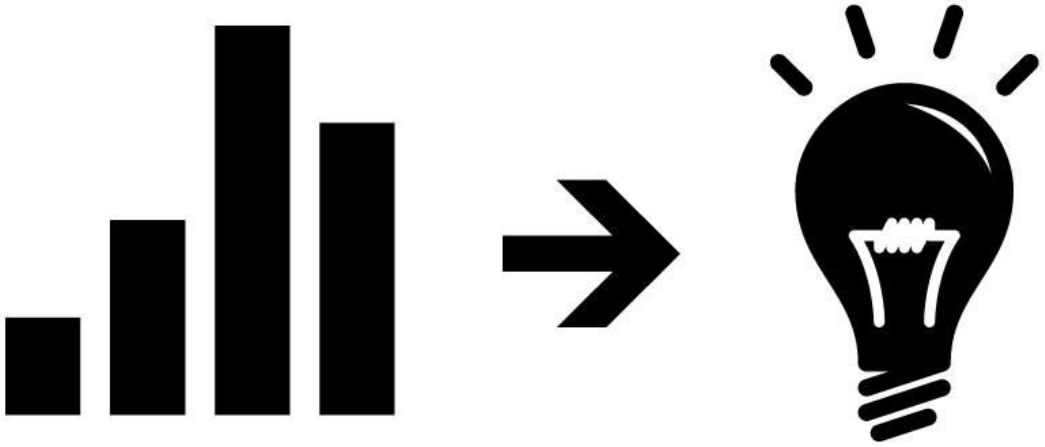
Capabilities	Outcomes
Know Everything about your Customer Analyze all sources of data to know your customers as individuals	**Creates customized offers up to 125x faster with better results**
Innovate New Products at Speed and Scale Capture all sources of feedback and analyze vast data to drive innovation	**Reduced processing time in half**
Instant Awareness of Fraud and Risk Analyze all available data, detect fraud and manage risk in real-time	**Identified fraud which previously went undetected**
Exploit Instrumented Assets Predict and prevent maintenance, develop new products & services	**Loads hurricane data in seconds and performs risk analysis in near real-time for greater reliability**

Data Science – One Definition by Drew Conway

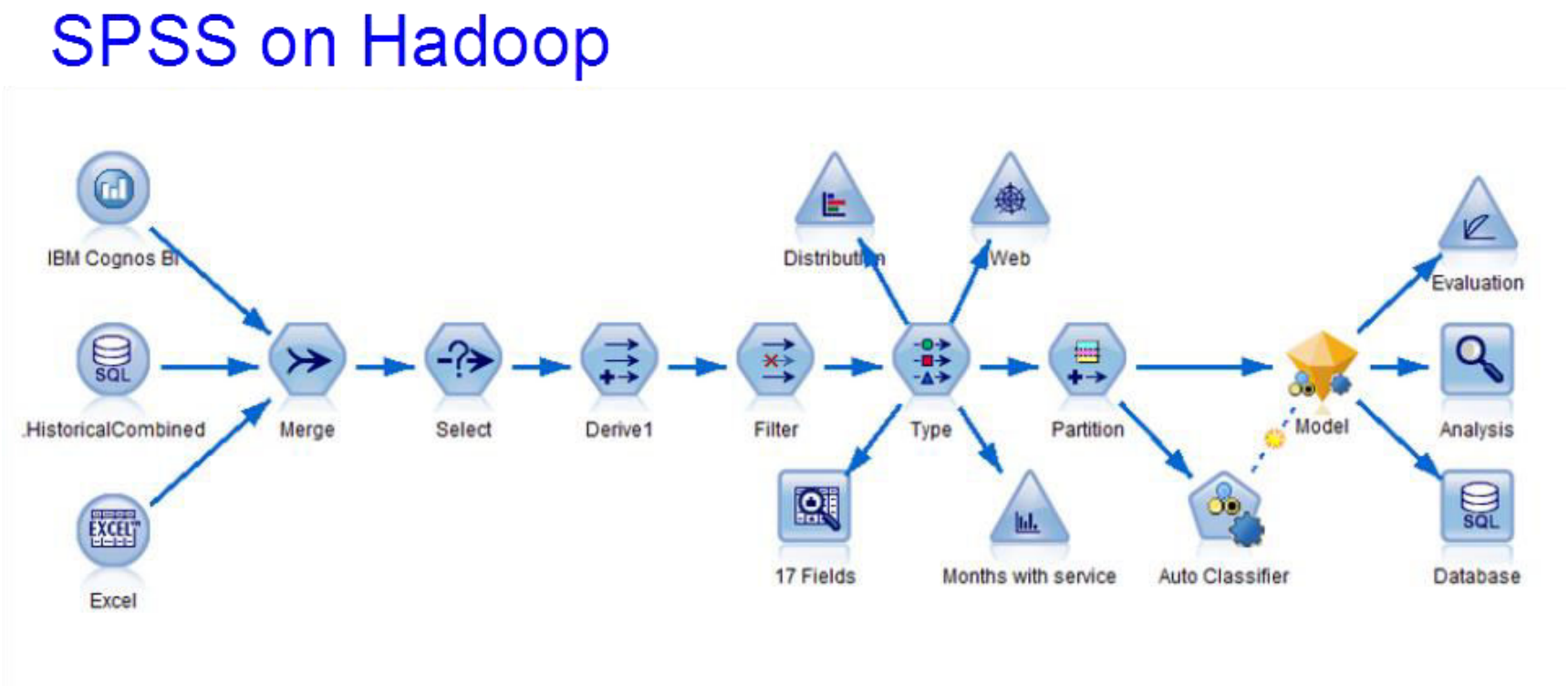


Data Science Definition

- **Data Science** is an interdisciplinary field about processes and systems to extract knowledge or insights from large volumes of **data** in various forms either structured or unstructured, which is a continuation of some of the data analysis fields such as **data** mining and predictive analytics, as well as knowledge discovery and **data** mining (KDD).
- **Data Science** is about turning data into insights.



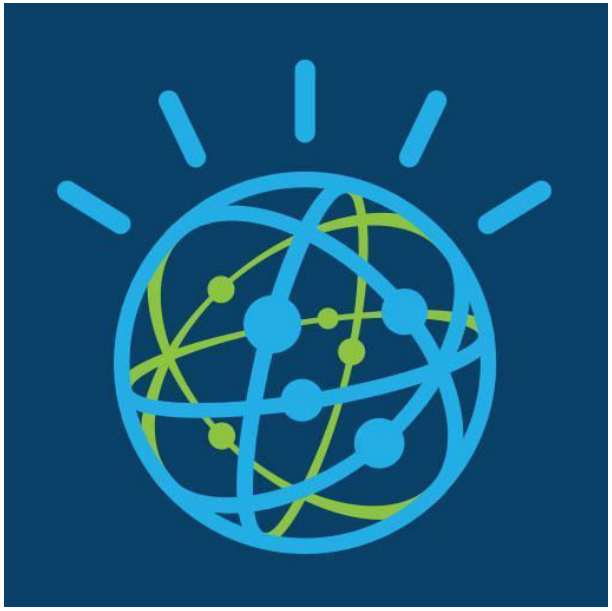
Data Science is a process

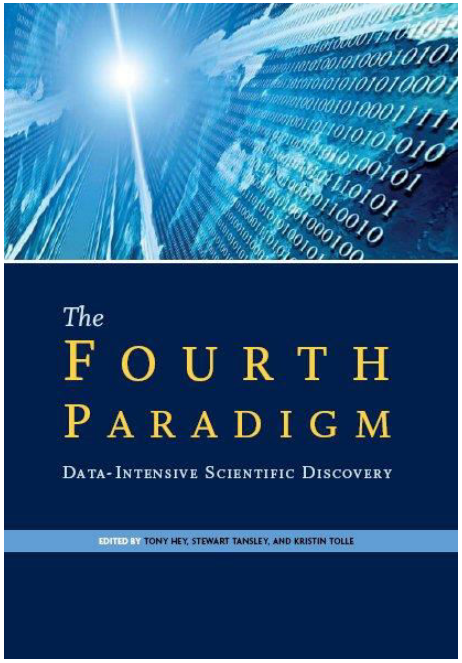


4Es – Equation – Estimation – Evaluation - Explanation

Data Science – a new science paradigm

- **Data Science** is a new science paradigm, under which the knowledge discovery processes and systems are dramatically different from that in the past, and even how scientists work and get organized is dramatically different from the past.
- **Data Science** is a new research paradigm, under which researchers must obtain intelligent assistance to deal with huge amount of data, large selection of **e**quations and models, large selection of **e**stimation algorithms, and complicated results **e**valuation and **e**xplanation.





Data Scientist

Data Scientist:
The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data.

by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

70 Harvard Business Review October 2012

Data Scientist – A Definition

- **A data scientist is a scientific professional who process large amount of data to discover insights.**
- A data scientist represents an evolution from a business or data analyst role. The formal training is similar, with a solid foundation typically in computer science and applications, modeling, statistics, analytics, math or even applied social science. What sets the data scientist apart is strong business acumen, coupled with the ability to communicate findings to both business and IT leaders in a way that can influence how an organization approaches a business challenge. Good data scientists will not just address business problems, they will pick the right problems that have the most value to the organization.
- Whereas a traditional data analyst may look only at data from a single source – a CRM system, for example – a data scientist will most likely explore and examine data from multiple disparate sources. The data scientist will sift through all incoming data with the goal of discovering a previously hidden insight, which in turn can provide a competitive advantage or address a pressing business problem. A data scientist does not simply collect and report on data, but also looks at it from many angles, determines what it means, then recommends ways to apply the data.

Source: http://www-01.ibm.com/software/data/infosphere/data-scientist/

Data Scientist Skills