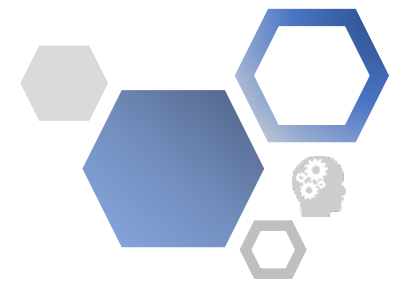




# Overview of the IBIC, Big Data, Advanced Analytics and Cognitive Computing

IBM Buffalo Innovation Center  
Jay Goodwyn, Director

# Topics



- **IBM Buffalo Innovation Center**
- **Big Data**
- **Advanced Analytics**
- **Cognitive Computing**

# What is the IBM Buffalo Innovation Center (IBIC)?



## Collaboration between IBM, SUNY Polytechnic and its Private and Public Partners

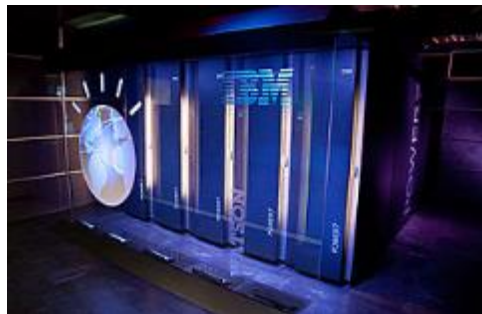
- 500 high-tech jobs committed by IBM within a 7 year period
- Located at the Key Center, temporarily at 50 Fountain Plaza in downtown Buffalo, 4<sup>th</sup> floor North Tower
- Permanent home will be the top 6 floors of the south tower; the 3<sup>rd</sup> floor for the data center (total 7 floors)

## Design point is as an analytics center

- Focus on **Advanced Analytics, Big Data, Watson (Cognitive) Computing**
- Represent the entirety of IBM for client solutions (SWG, GBS, AMS, Research, STG)
- Primary industry focus in Government and Healthcare; Secondary cross-industry within New York State

## Develop integral relationships with local colleges and universities

- Help design and implement an analytics curriculum
- Develop a recruiting process
- Implement internship and residency programs
- Focus on local colleges and universities (UB, Buf State, Canisus, ECC, etc) as well as regional schools (RIT, RPI, Syracuse, Cornell)



# The Big Data phenomenon

1 zettabyte = 1k exabytes = 1m petabytes = 1B terrabytes

## In just two days

we now generate as much data as was generated in total through 2003

## 80% of all data

is unstructured and growing 15 times the rate of structured data

## Over 1 billion tweets

are sent every 3 days

## 5 million trade events

are clocked every day

We are here

44 zettabytes

4.4 zettabytes

2010

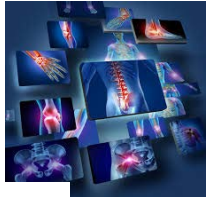
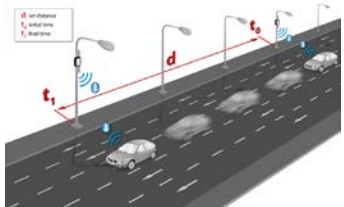
2013

2020

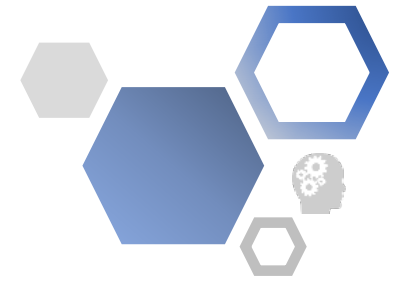
Structured data

Unstructured data

**DARK DATA**



# Another key driving factor behind Big Data is storage costs



The first hard disk drive came from IBM. It was called the IBM Model 350 Disk File and was a huge device. It had **50 24-inch disks** contained inside a cabinet that was as large as a cupboard and anything but lightweight. This hulk of a storage unit could store a whopping 5 MB of data.

1 Terabyte of Storage

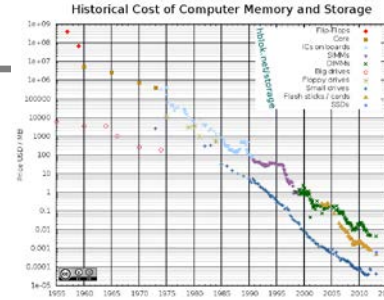
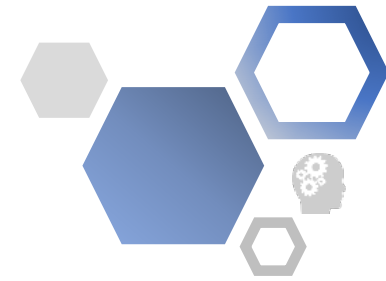
$$= 1,000,000,000 / 5 \text{ Megabytes}$$

$$= 200,000 \text{ units}$$

\$3,500 per month per unit

$$= \$3,500 \times 200,000 = \$700\text{m (per month)}$$

# A Little Perspective



## 1956

### IBM Released 350 RAMAC disk file

- It took 200,000 of these to store 1 TB
- They would take up over 3000 soccer fields
- It cost \$700M to store 1 TB (storage only!)

## 1976

### IBM Released the 3350 Direct Access Storage

- It took 1575 of these to store 1 TB
- It cost \$100M to store 1 TB

## 1997

### IBM Released the Deskstar 16GP

- It took 5752 of these to store 1 TB
- It cost \$3M to store 1 TB

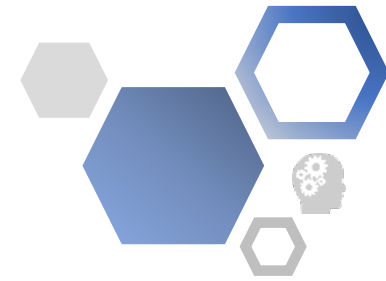
## Today

### You can buy a 512GB Memory Card for \$279

- It takes 2 memory cards to store 1 TB
- It cost \$500 to store 1 TB

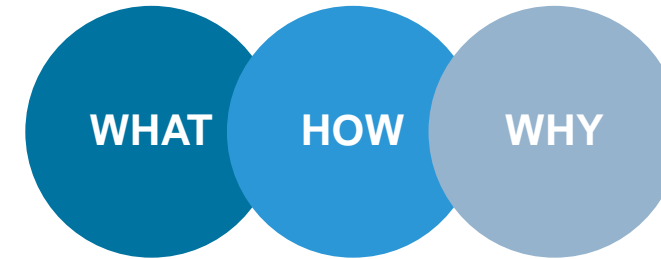
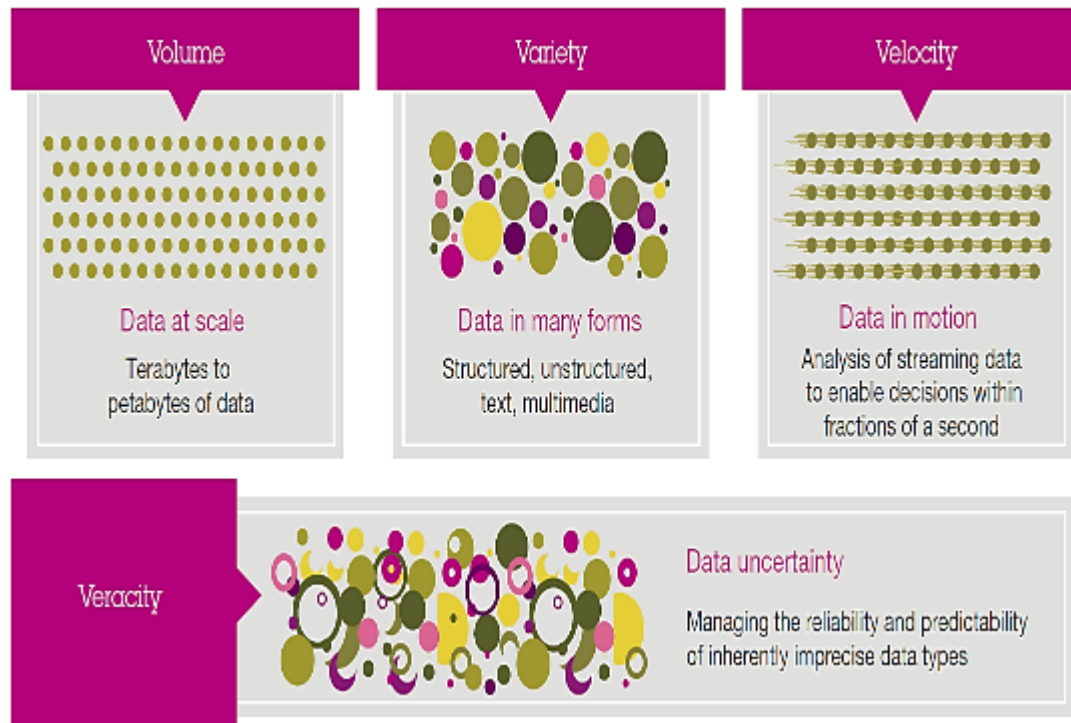


# There is more to Big Data than the 3 / 4 V's of Big Data



*Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.*

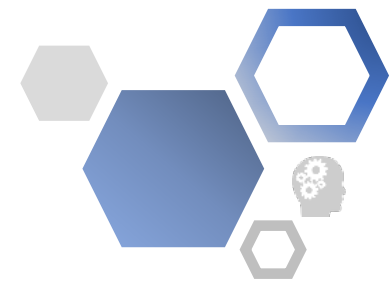
*(Gartner Definition of Big Data)*



- 1 **New dimensions of data complexity** ("What?") – the "Vs"
- 2 **New capabilities** for data processing and analytics ("How?") – practices and methods, skills, technologies
- 3 **Systematic exploitation for business applications** ("Why?") – insight, decisions, business models

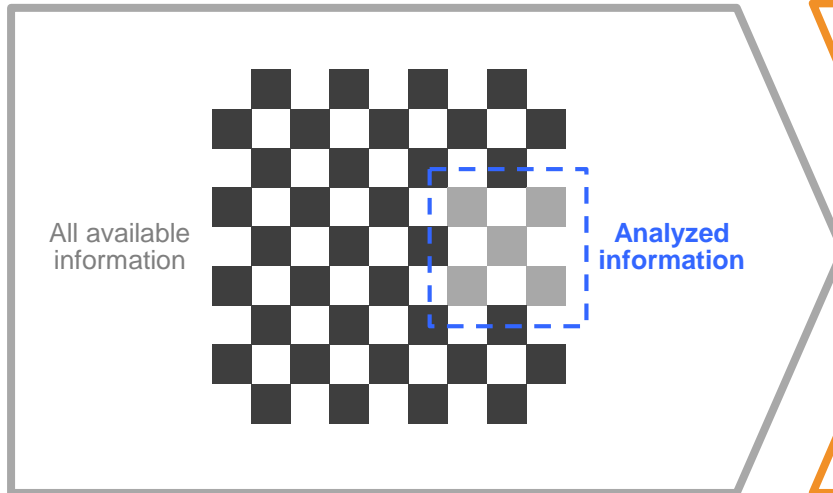
Source: [www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/](http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/)

# Paradigm shifts enabled by big data



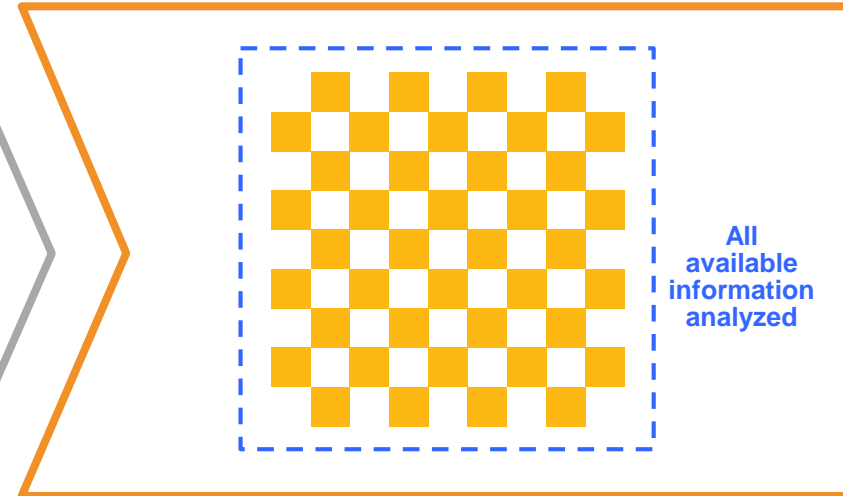
Consider ALL data as opposed to structured

TRADITIONAL APPROACH



Analyze small subsets  
of information

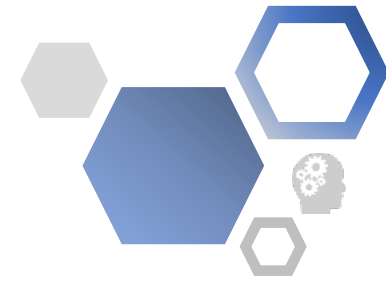
BIG DATA APPROACH



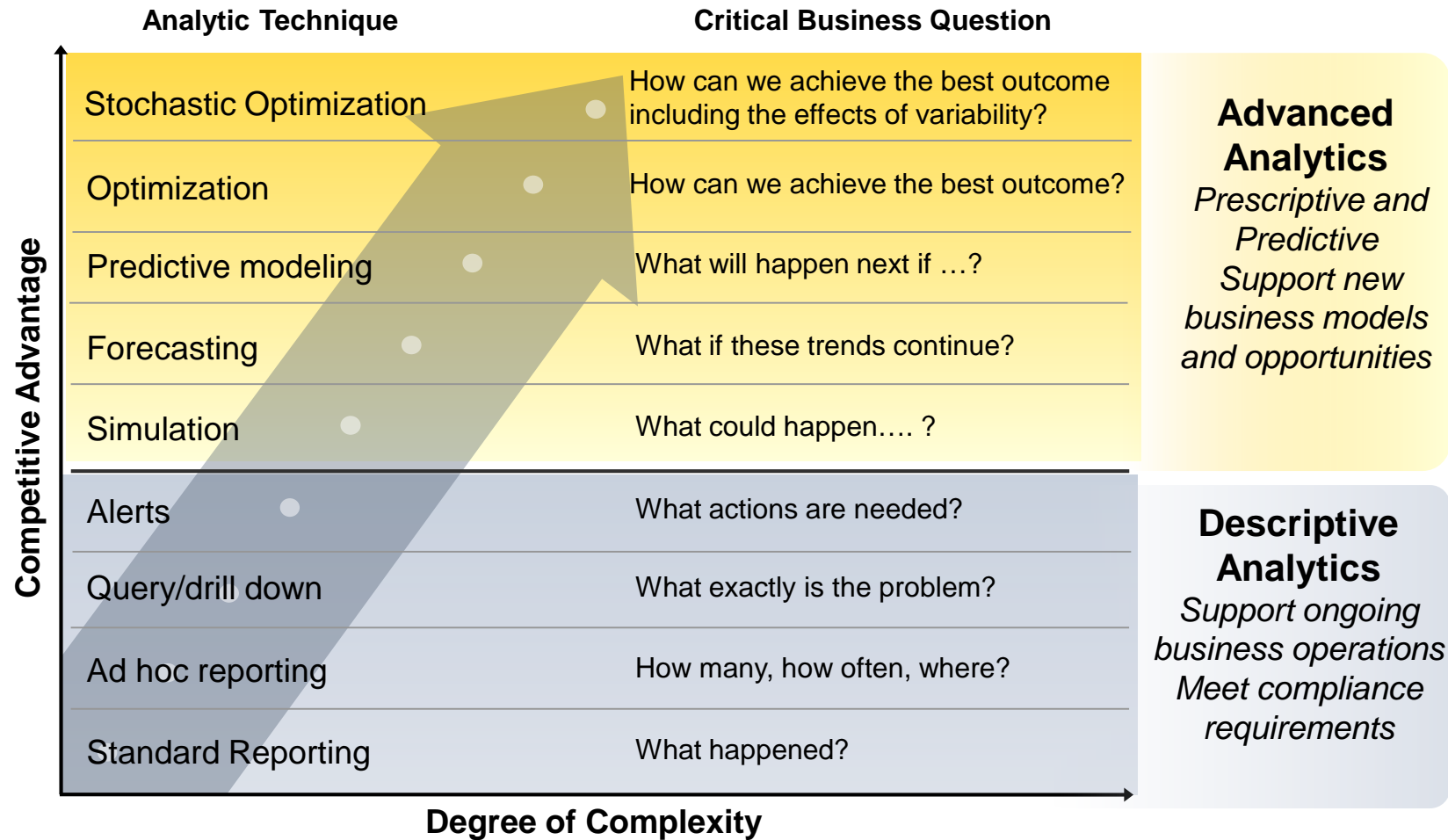
Analyze  
*all* information



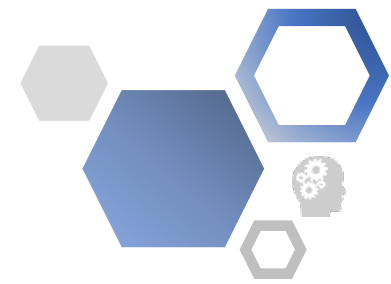
# What is Advanced Analytics?



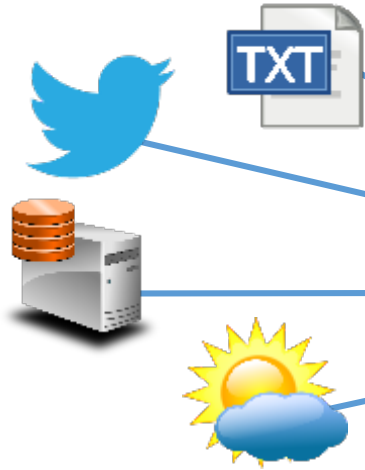
How can we distinguish between "analytics" and "advanced analytics?"



# Advanced Analytics - Example



## Historic Information



- Enterprise data
- Data warehouse
- Sensor data
- Streaming data
- Text documents
- Tweets
- Weather info
- Other external data

## Data Integration



- Data prep and load
- Connect data relationships
- Quality audit
- Annotate and interpret

## Analytic Model



- Build forecasting model that best evaluates correct outcomes based on data input

## Predictive Analytics



- Predict outcomes
- Compare against actual outcomes
- Adjust model as appropriate

# Case Study: Rapid Analytics for Hidden Defection (Banking Client)

## Client Profile

- Regional banking corporation with assets of approximately \$37 billion as of 2013 and nearly 5000 employees in the New York and Pennsylvania regions.

## Project Objective

- Implement rapid analytics to predict “hidden defection” of clients who are on the path to discontinuing business with the bank.

## Expected Benefits

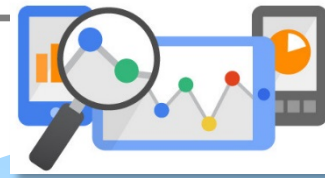
- Enhance the bank’s ability to predict attrition of clients, allowing the bank to proactively prevent clients’ defection by sending out pinpointed marketing offers and communications.



## Data Ingest & Preparation

Collaboration between the IBIC and the client to understand, gather, define, and map, and load the client’s data.

**Project Results**  
Gathered, mapped, and loaded 4.3M unique individual records and 6.05M unique household records with 1.4M transactions over a 27 month period.



## Modeling & Analytics

Build, test, and run the analytics model and reports to discover correlations and patterns in the data.

**Project Results**  
Applied multiple predictive modeling algorithms to the data and built analytics models in SPSS Modeler tool to uncover possible correlations in the data.



## Evaluation & Insights

Develop data insights, refine as required, and draw conclusions based on the analytics model’s findings.

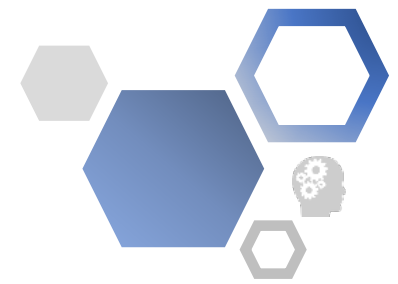
**Project Results**  
Defined two customer groups based on account activity to **predict attrition with 75% accuracy for 70% of all attrition customers.** Explored breakdowns of the data for each customer group to derive behavior insights.

## Next Steps

Final report with results, conclusions, and recommendations.

**Project Results**  
Delivered analytics models for each of the two defined customer groups. Recommended development options for the client, such as adoption for an Analytics Foundry.

# A New Era of Computing



Tabulating Systems Era

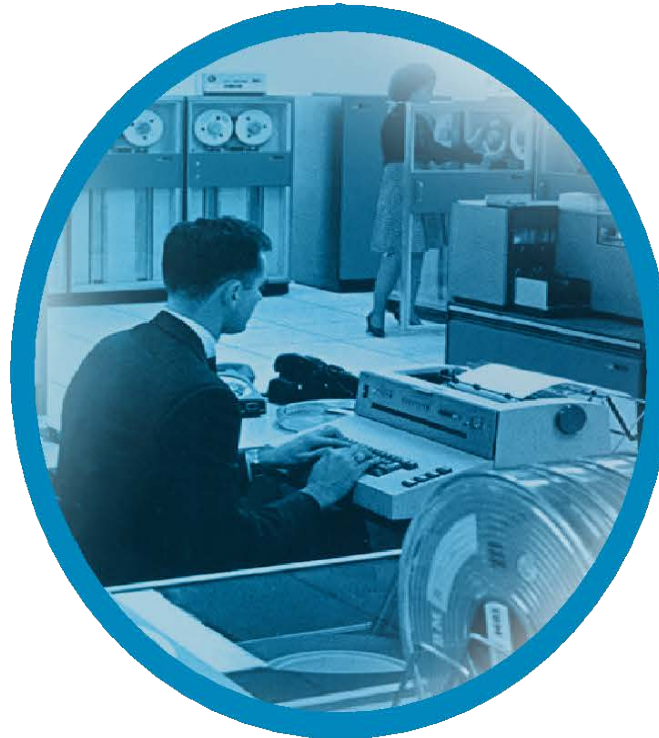
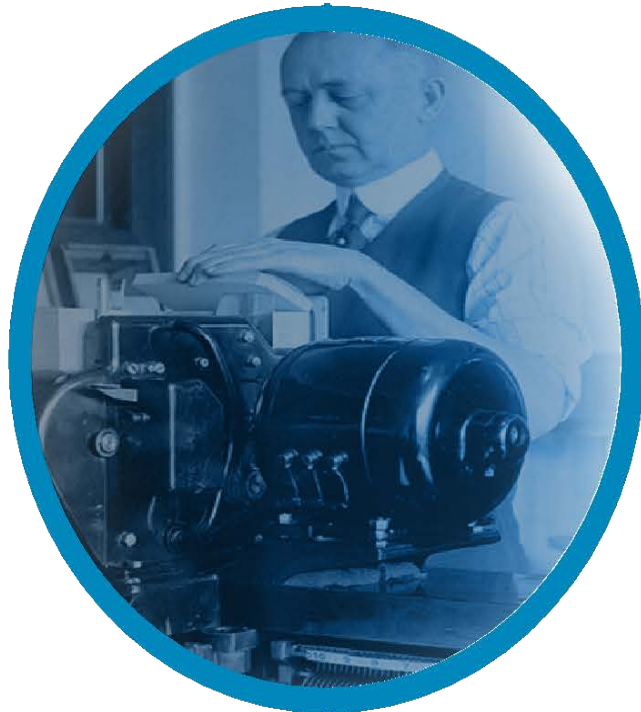
Programmable Systems Era

Cognitive Systems Era

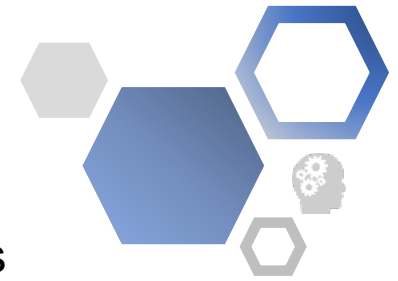
1900

1950

2011



# The Jeopardy Challenge



On February 14<sup>th</sup>, 2011, IBM's Watson took on two of the most successful Jeopardy contestants ever...and won.

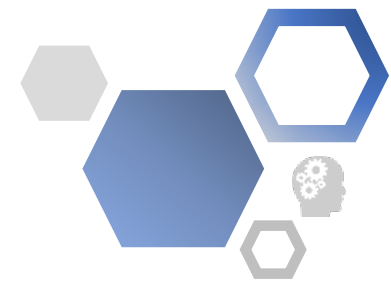


- Read over 200 million pages of information
- Had to understand the question, develop hypotheses, compute confidence levels, and decide whether to answer *faster than the best human beings*
- Machine learning took Watson from taking hours to get to low probability to answering Jeopardy questions within 2 ½ seconds correctly
- Watson is more than a Q&A system; the implications of machine learning reach far beyond what was witnessed on Jeopardy

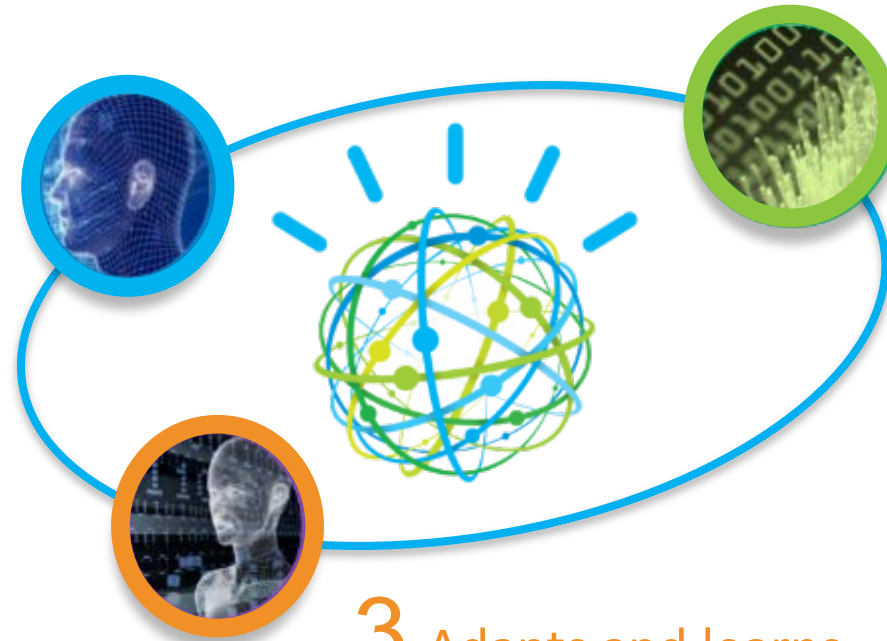
[https://www.youtube.com/watch?v=Y\\_cqBP08yuA](https://www.youtube.com/watch?v=Y_cqBP08yuA)

# The concept

---



1 Understands natural language, context and the complexities of human communication

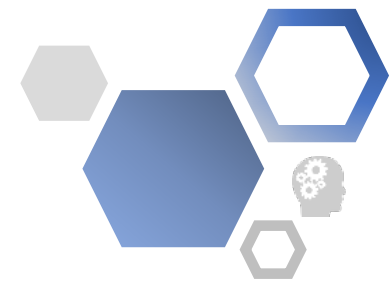


2 Generates and evaluates evidence-based hypotheses

3 Adapts and learns from training, interaction and outcomes

# Commercialized First in Healthcare

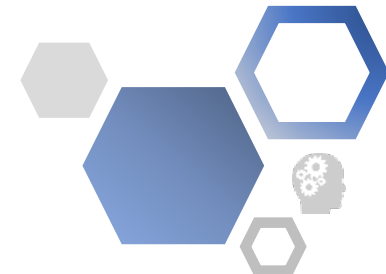
---



The most obvious and immediately effective place for Watson's ability was in the Healthcare space. With so much healthcare data available, the possibilities for Watson can only increase.

- Watson for Oncology developed in partnership with Memorial Sloan Kettering
  - Watson reads all published medical literature and new drug trials and brings that together with the experience of MSK's Oncologists to recommend treatment plans
- Other Use Cases
  - Clinical Trial Matching – identify candidates for new drug trials
  - Skin Lesion Analysis for Cancer
  - ER Training
  - Genomic Sequencing
- Watson Health Cloud
  - Will bring together clinical, research and social data from a diverse range of health sources to creating a secure, cloud-based data sharing hub





Thank  
You