IBM Power Systems™

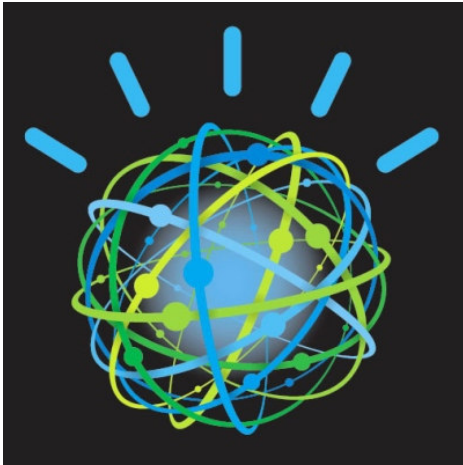## IBM Systems & Technology Group

Watson

Patrick O'Rourke
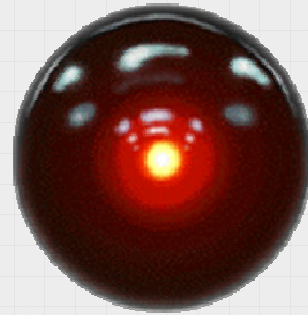**pmorour@us.ibm.com**
Executive Briefing Center

Power is performance redefined

Deliver IT services faster,
with higher quality,
and with superior economics

www.ibm.com/power

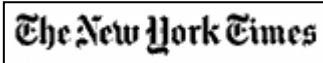# From Science Fiction to Reality

Watson takes on Jeopardy!

Advanced computing system has potential to take business intelligence to a new level

- **Date: February 14 / 15 / 16 2011**

- **IBM Research project named "Watson"**

- **Competition with humans at the game of Jeopardy:**
  - **Human vs. Machine contest.**

- **Competition:**
  - **Ken Jennings & Brad Rutter**
  - **Two most successful Jeopardy contestants of all time**

Watson's Popularity…

YouTube

NOVA

NBC NIGHTLY NEWS

CBS NEWS

WIRED

LATE SHOW with David Letterman

The Tonight Show with Jay Leno

WIKIPEDIA
The Free Encyclopedia

The New York Times

THE INDEPENDENT

Google watson ibm
About 6,910,000 results

eWEEK.com

Forbes.com

The Washington Post

TV GUIDE

facebook

2011 Webby Person of the Year!

THE WEBBY AWARDS

IBM WATSON

EMC 2011 Data Hero Awards

EMC² DATA HERO AWARD WINNER

Technology: Application

EMC WORLD 2011—LAS VEGAS—May 9, 2011 – Technology Application award to IBM Watson

- Awards honor those delivering unique and innovative solutions and techniques for using Big Data to profoundly impact individuals, organizations, industries and the world.

- Awards are selected by a prestigious independent judge panel

IBM WATSON

The Jeopardy! Challenge: *A compelling and notable way to* drive *and* measure *the technology* of automatic Question & Answering

Broad/Open Domain

Complex Language

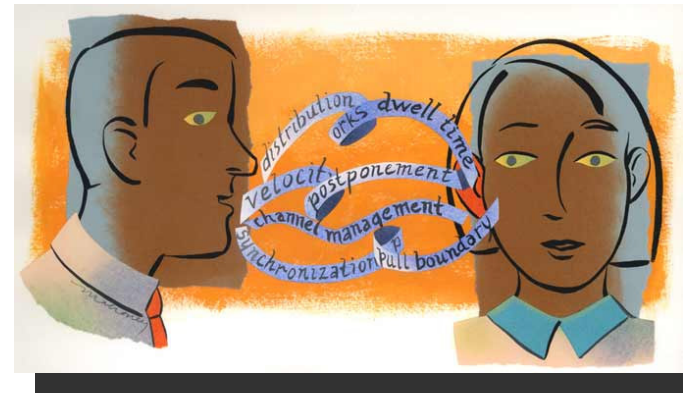**High Precision**

Accurate Confidence

High Speed

$200
If you're standing, it's the direction you should look to check out the wainscoting.

$1000
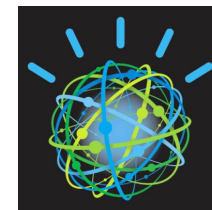**The first person mentioned by name in 'The Man in the Iron Mask' is this hero of a previous book by the same author.**

Word Plays   Puns   Rhymes
Lyrics   Dual meanings

IBM WATSON

Watson Overview

Watson History.

- 3+ years development by IBM scientists
- Software: IBM Research Software Stack
- Hardware: Power Systems

Why Jeopardy?

- Grand challenge for a computing system
- Broad range of subject matter,
- Speed at which contestants must provide both accurate responses
- Determine a confidence they are correct

Tale of the Tape

	Deep Blue
1997	
Game	Chess
Architecture	Specialized Hardware and Power2 SC
# Cores	30 + 480 ASICs
Hardware	Highly Specialized
Processing	Mathematical / Sturctured
Data Analysis	▪ **Finite number of possible moves and countermoves** ▪ **Structured data** ▪ **Mathematical probability**

IBM**WATSON**

Watson Info.....

Hardware:

- Cluster: 90 Power 750 (2880 Cores) @ 3.55 GHz
- 80 Teraflops
- 88 Compute nodes, 2 I/O nodes
  - ❖ 15TB of memory
- 4 SAS Storage drwrs & 2 xCAT Servers

Software:

- SLES 11, JAVA, CNFS, GPFS, xCat, Apache Hadoop

Middleware:

- Apache UIMA (Open source)

Applications:

- DeepQA - Main analytical engine which ran on POWER 7
- Lenovo desktop : Voice synthesis, strategies for betting, buzzing in, clue selection & exchanging info with Jeopardy Computers
- Mac notebook: Avatar

Power 750 System

Watson Environment	
POWER7 Architecture	32 Cores @ 3.55 GHz 88 compute nodes 2 I/O nodes
DDR3 Memory	60 Nodes 128GB 30 Nodes 256GB
System Unit SFF Bays	6 HDD 146GB @ 15k
System Unit IO Expansion Slots	(1) FC 1983 Dual port 1Gb Ethernet (1) FC 5769 10Gb Fiber SR In the 2 I/O nodes: (1) FC 5903 SAS RAID Controller
Integrated Virtual Ethernet	Dual 10 Gb IVE / HEA
Storage Drawers	Four SAS enclosure FC 5888 ▪ Two per I/O node ▪ Twelve 300GB
IO Drawers	None

Watson's Sources of Information

Watson reads roughly 200 million pages of content (equivalent to one million books), written in natural human language … in less than 3 seconds

Organized by topic, but additional information
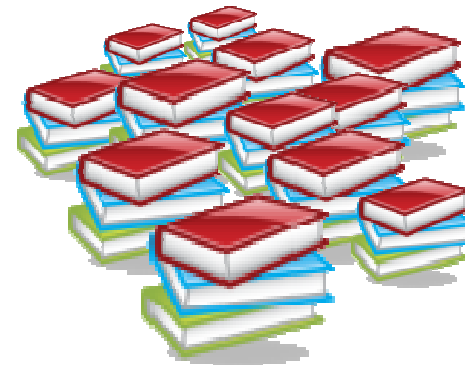   comes from actually reading the sentences

Unstructured Sources
Wikipedia (Full text)
IMDb
Encyclopedias
Dictionaries
Thesauri
Newswire Articles
Literary Works (including the Bible)

Structured Sources
DBpedia
Wordnet
YAGO

**200 million pages structured and unstructured content
= 1 Million books
= 4 TB storage+16 TB memory**

These structured data sources were typically used to obtain information about the English language, e.g. nouns, verbs, adjectives and adverbs grouped into sets of cognitive synonyms

IBM WATSON

Watson Odds and Ends…

22 Different Process Types
- Heavily Parallelized

389 Processes
- 199 C++
- 190 Java

Threads were heavily core multi-threaded
- Large memory: 256 GB System
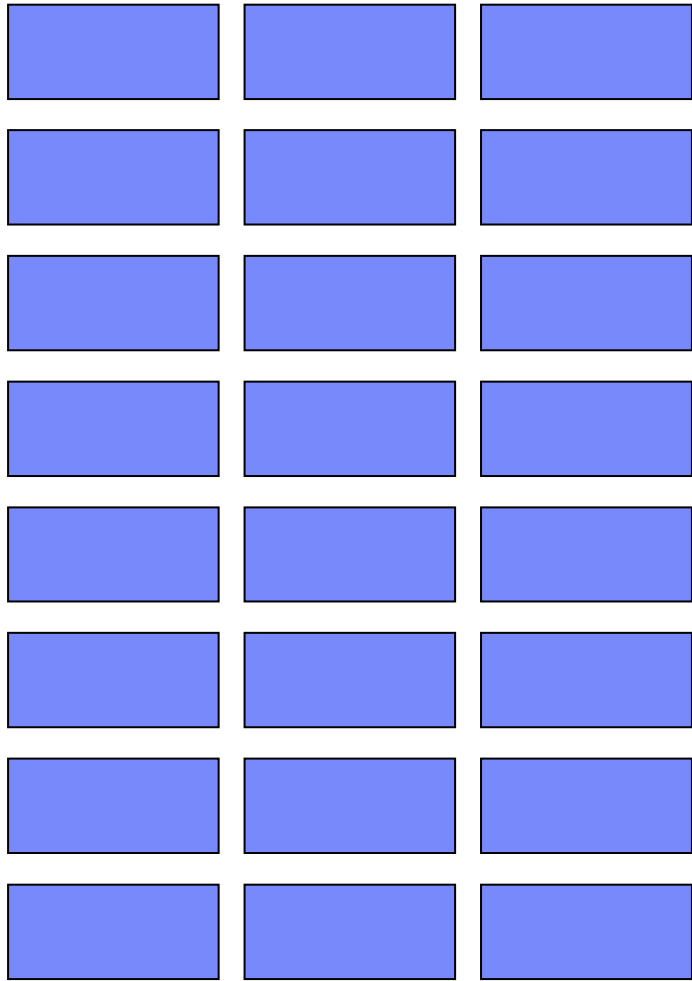- Smaller memory: 128 GB System

Watson Challenge…..

How to Process / Analyze / Evaluate / Prioritize all of this Unstructured Data

Data Problem…..

Structured Data

Unstructured Data
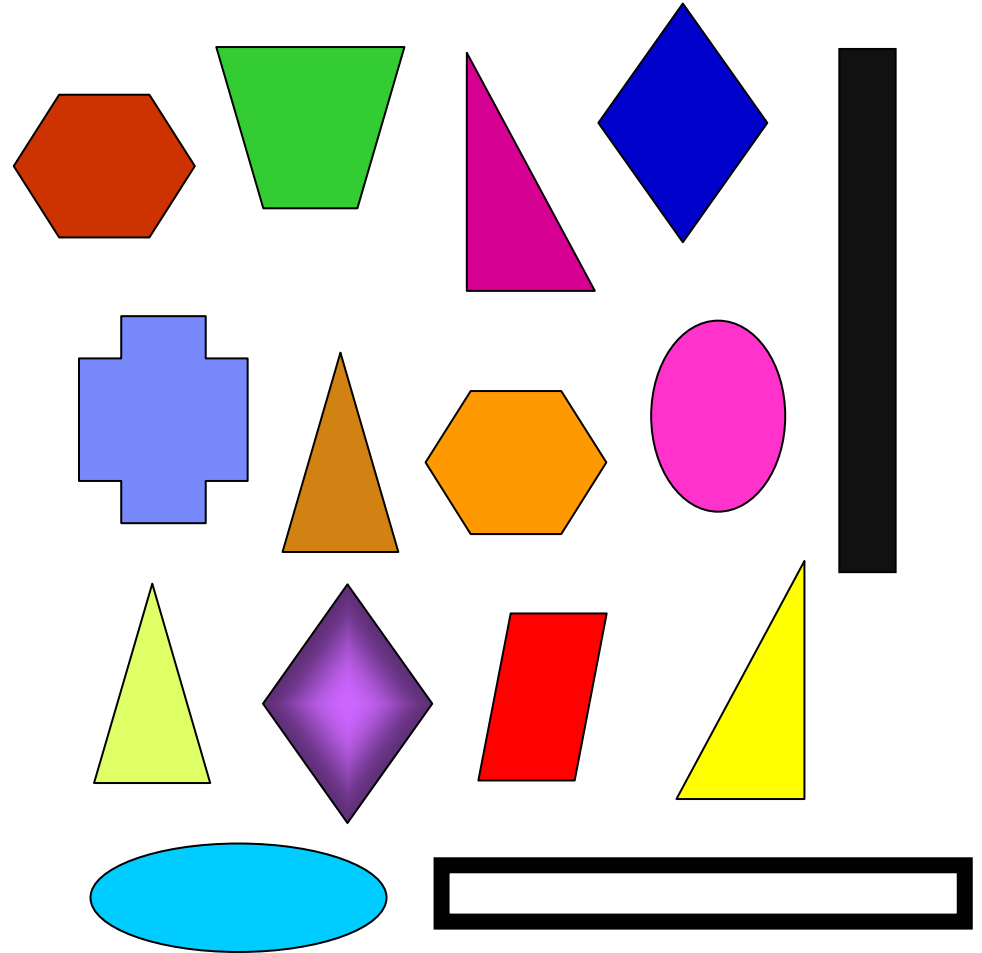
Easy Questions…..

$\ln((12{,}546{,}798 * \pi))^2 / 305.992 = \quad 1.0$

Select *Payment* where *Owner*="David Jones" and *Type(Product)*="Laptop",

Owner	Serial Number
David Jones	45322190-AK

Invoice #	Vendor	Payment
INV10895	MyBuy	$104.56

Serial Number	Type	Invoice #
45322190-AK	LapTop	INV10895

David Jones

David Jones

=

Dave Jones

David Jones

≠

Hard Questions....

Computer programs are natively explicit, fast and exacting in their calculation..

Natural Language is implicit, highly contextual, ambiguous and often imprecise.

Structured

Person	Birth Place
A. Einstein	ULM

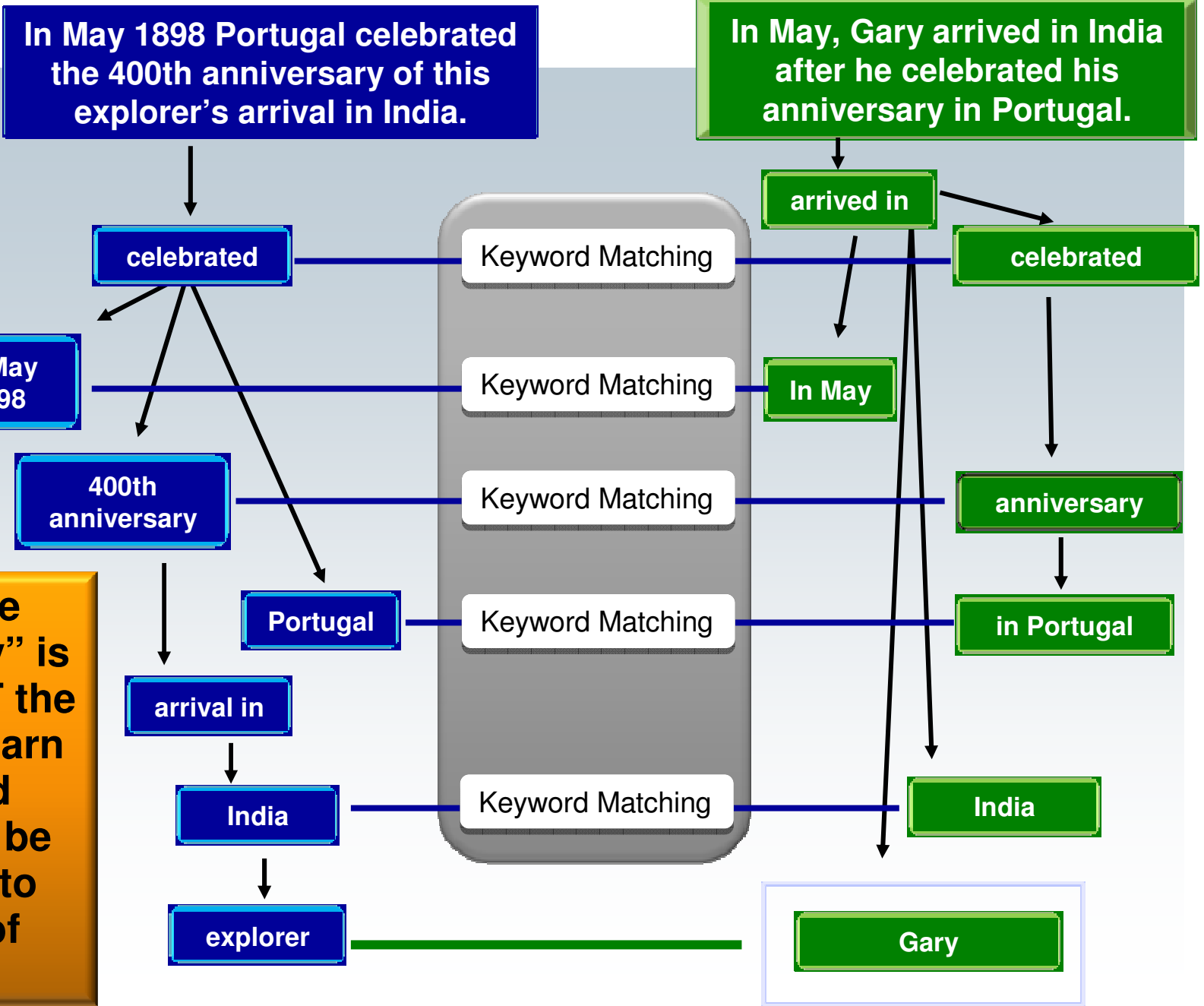| Person	Organization
J. Welch	GE

UnStructured

1. **Where was X born?**

   Otto chose a water color of this city to send to Albert Einstein as a remembrance of Einstein´s birthplace.

2. **X ran this?**

   If leadership is an art then surely, Jack Welch proved himself a master painter during his tenure at this company..

IBM WATSON

Keyword Evidence

In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India.

In May, Gary arrived in India after he celebrated his anniversary in Portugal.

celebrated — Keyword Matching — celebrated

arrived in

In May 1898 — Keyword Matching — In May

400th anniversary — Keyword Matching — anniversary

Portugal — Keyword Matching — in Portugal
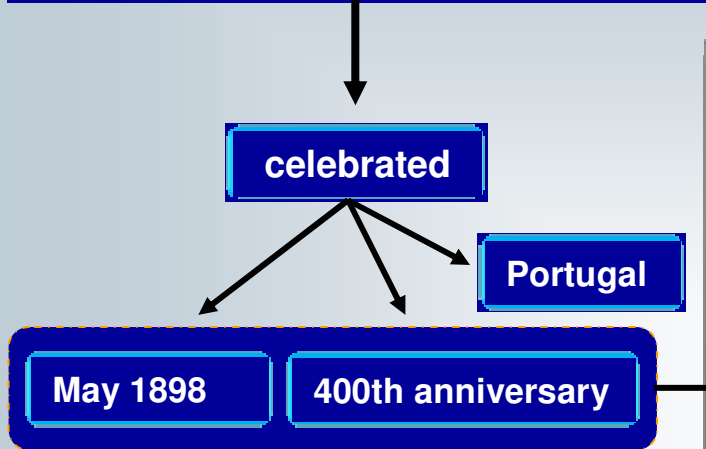
arrival in

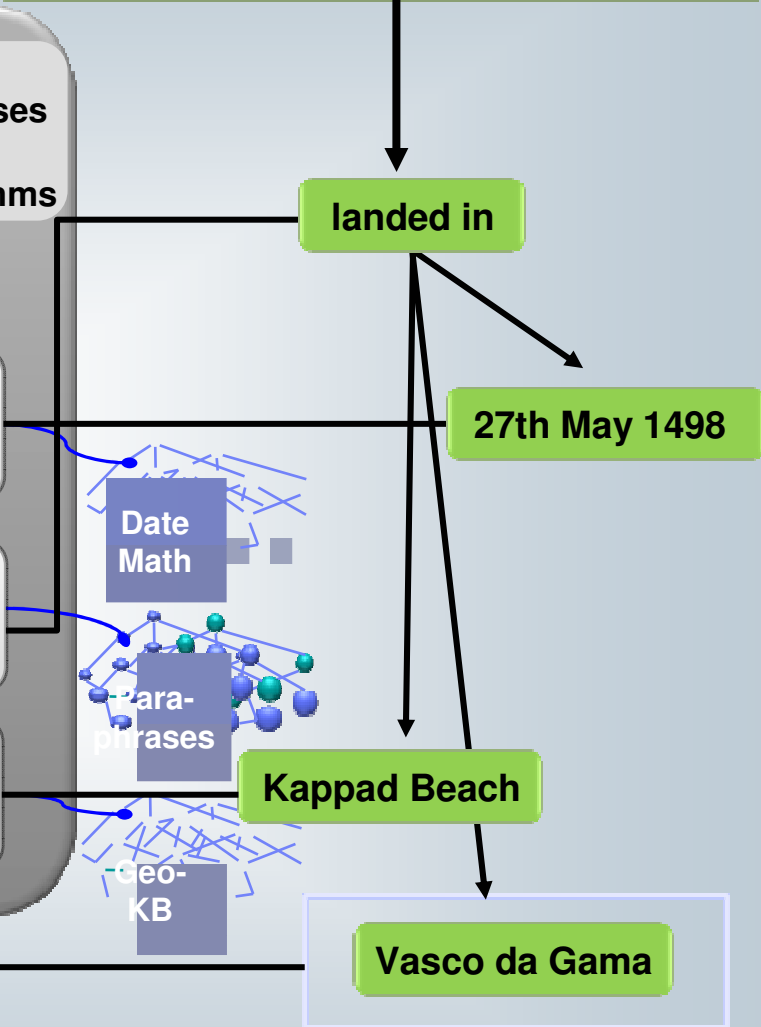India — Keyword Matching — India

explorer

Gary

This evidence suggests "Gary" is the answer BUT the system must learn that keyword matching may be weak relative to other types of evidence

Deeper Evidence

In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India.

On the 27th of May 1498, Vasco da Gama landed in Kappad Beach

- *Search* Far and Wide
- Explore many hypotheses
- Find Judge Evidence
- Many inference algorithms

celebrated

Portugal

May 1898 400th anniversary

landed in

Temporal Reasoning

27th May 1498

Date Math

arrival in

Statistical Paraphrasing

Para-phrases

India

GeoSpatial Reasoning

Kappad Beach

Geo-KB
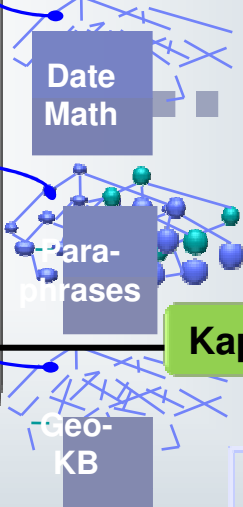
Stronger evidence can be much harder to find and score.

explorer

Vasco da Gama

The evidence is still not 100% certain.

IBM WATSON

Best Human Performance.....

Each dot represents an actual historical human Jeopardy! game

Top human players are *remarkably* good.
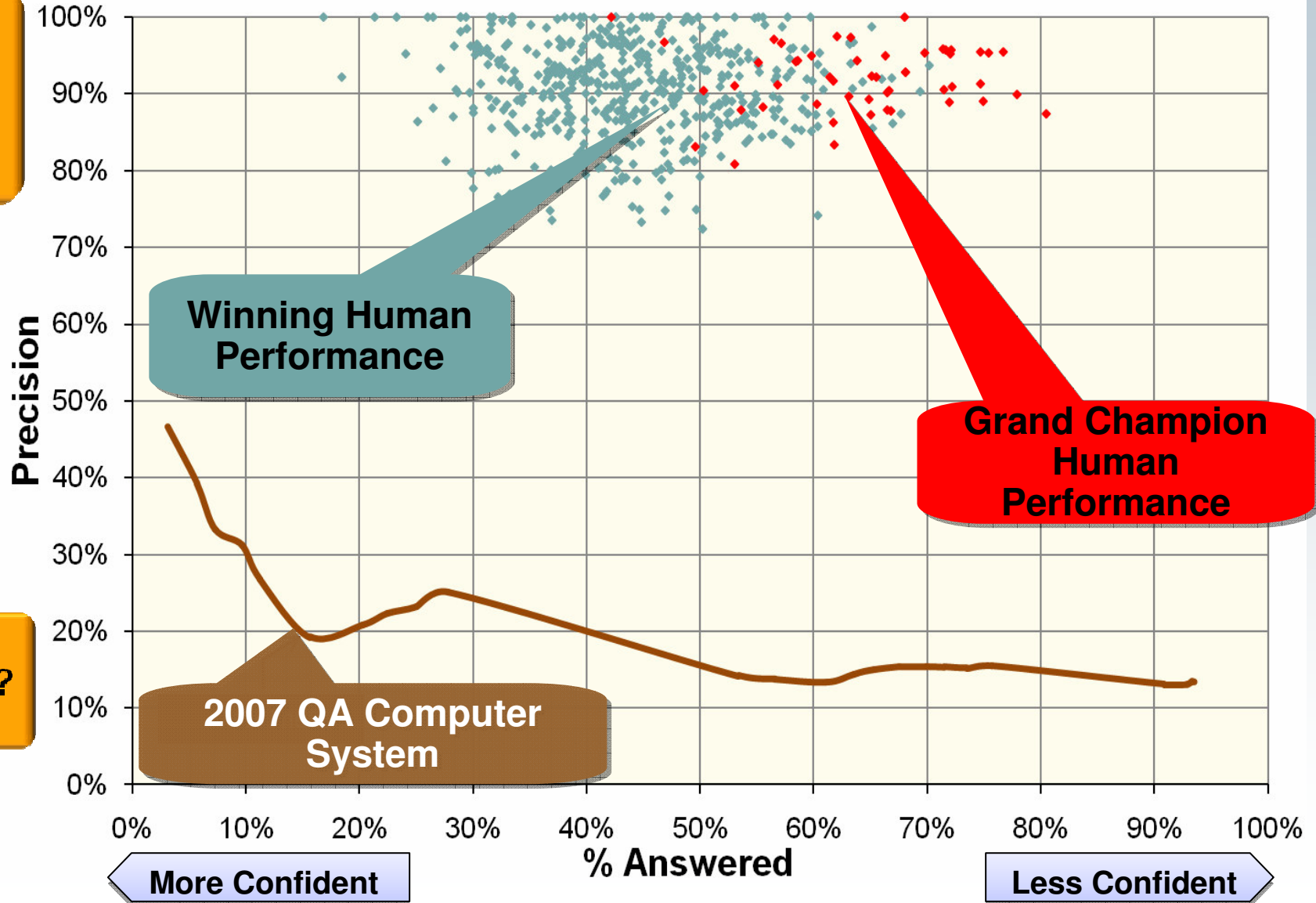
Winning Human Performance

Grand Champion Human Performance

Computers?

2007 QA Computer System

Precision

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

0%   10%   20%   30%   40%   50%   60%   70%   80%   90%   100%

% Answered

◁ **More Confident**

Less Confident ▷

IBM Apache UIMA Hadoop, and DeepQA,

UIMA is the industry standard for Content Analytics
- Unstructured Information Management Architecture.

UIMA SDK was originally developed by IBM
- 2006: SDK available at alphaWorks®.
- 2008: IBM donated UIMA SDK to **Apache**
- Ongoing: Development by the Apache UIMA community

Apache UIMA add-on: **UIMA Asynchronous Scaleout** (AS)
- Provides the ability to scale out in a clustered environment

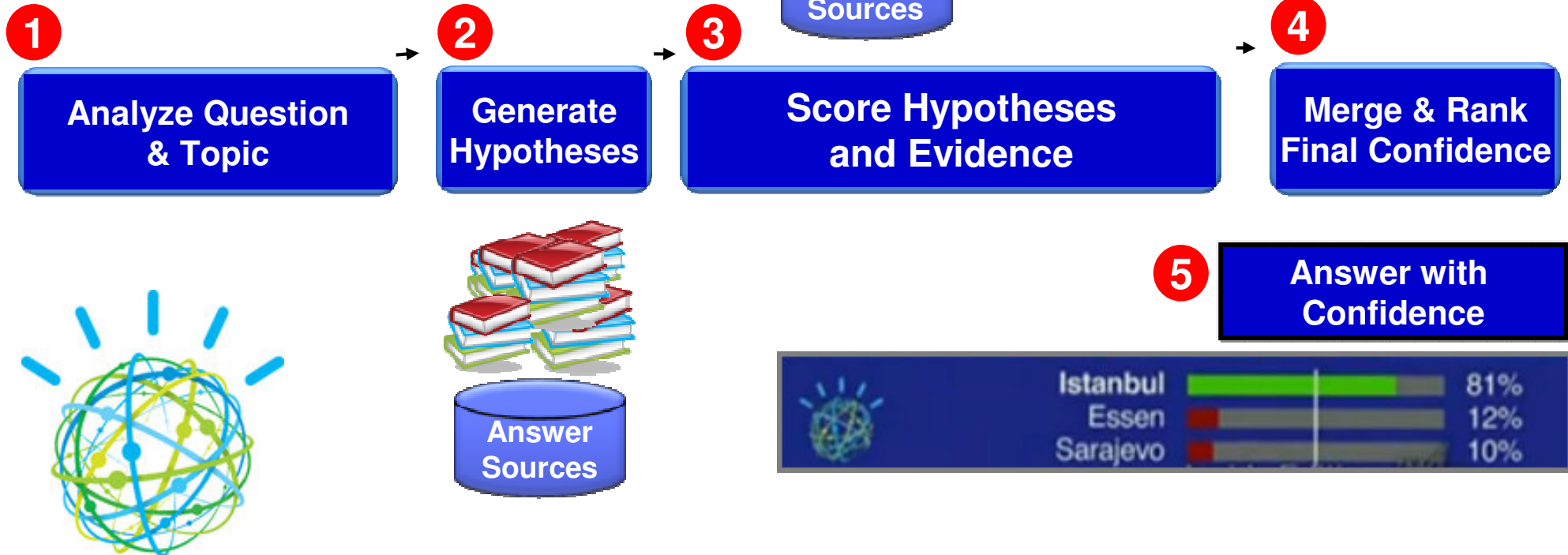**Apache Hadoop**: Framework used by Watson to facilitate preprocessing the large volume of data, created in-memory datasets used at run-time

DeepQA: Collection of Algorithms
- Can be divided into independent parts, each executed by a separate processor / Computation is embarrassing parallel
- Gathers, evaluates, weighs and balances different types of evidence, delivering the answer with the best support it can find.

IBM DeepQA: How Watson answered questions

Each year the EU selects capitals of culture; one of the 2010 cities was this Turkish "meeting place of cultures"

eviden

Models   Models

Evidence Sources

1 **Analyze Question & Topic**

2 **Generate Hypotheses**

3 **Score Hypotheses and Evidence**

4 **Merge & Rank Final Confidence**

Answer Sources

5 **Answer with Confidence**

Istanbul		81%
Essen		12%
Sarajevo		10%

IBMWATSON

DeepQA: The Technology Behind *Watson*

*Generates and scores many hypotheses using a combination of 1000's **Natural Language Processing**, **Information Retrieval**, **Machine Learning** and **Reasoning Algorithms**.*

Question

100s sources

100s Possible Answers

Multiple Interpretations

1000's of Pieces of Evidence

100,000's scores from many simultaneous Text Analysis Algorithms

Potential Answers

| Question & Topic Analysis | Question Decomposition | Hypothesis Generation | Hypothesis and Evidence Scoring | Synthesis | Final Confidence Merging & Ranking |

Hypothesis Generation

Hypothesis and Evidence Scoring

Answer & Confidence

built on UIMA *for interoperability*

built on UIMA-AS *for scale-out and speed*

DeepQA: Progress in Answering Precision 6/2007-11/2010

**IBM *Watson*
Playing in the Winners Cloud**

v0.8 11/10

V0.7 04/10

v0.6 10/09

v0.5 05/09

v0.4 12/08

v0.3 08/08

v0.2 05/08

v0.1 12/07

Baseline 12/06

Precision

% Answered
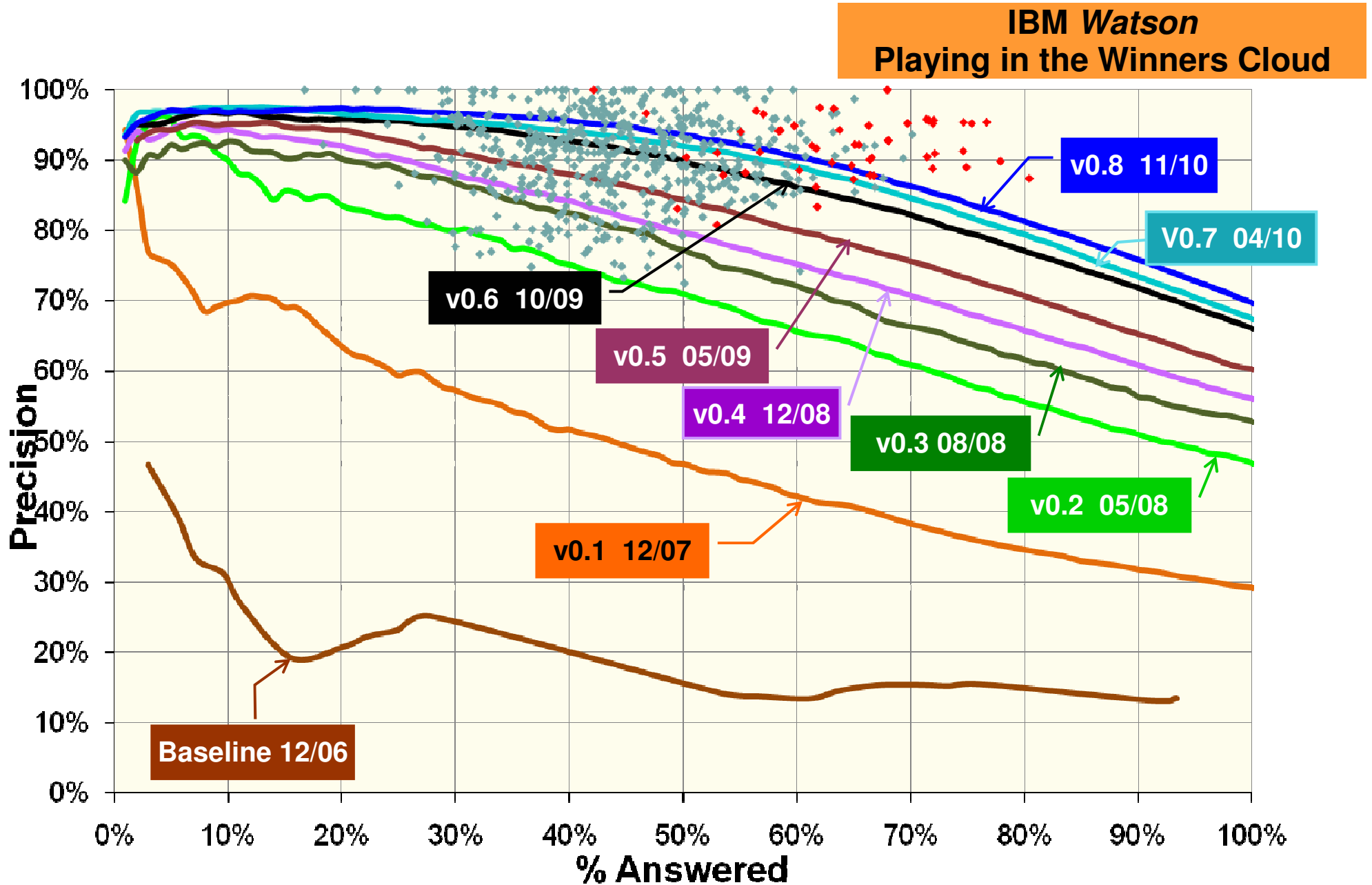
Precision / Confidence & Speed

Deep Analytics – Combining many analytics in a novel architecture, we achieved very high levels of *Precision* and *Confidence* over a huge variety of *as-is* content.

Emily Dickinson		99%
Walt Whitman		60%
Barnard		10%
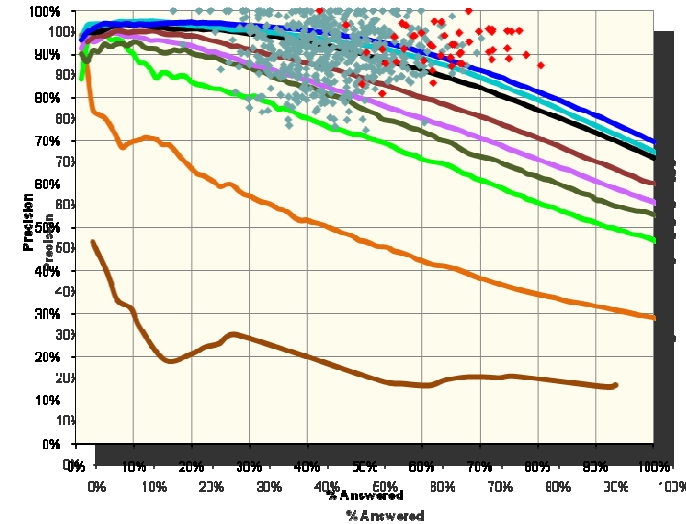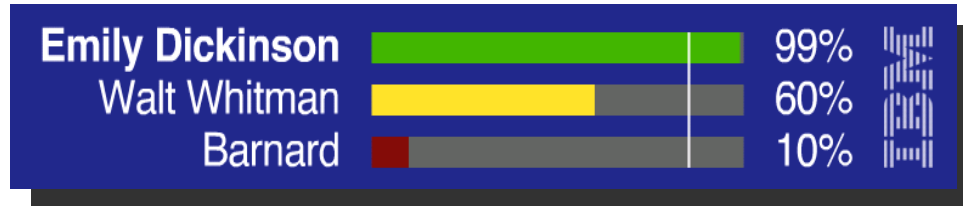
Speed – By optimizing Watson's computation for Jeopardy! on over **2,800 POWER7** processing cores we went **from 2 hours per question on a single CPU to an average of just 3 seconds.**

Results – in 55 real-time sparring games against former **Tournament of Champion Players last year**, Watson put on a very competitive performance in all games -- placing 1st in 71% of the them!

Watson has much to learn about Chicago!

Category names on Jeopardy! are tricky

"What US city" wasn't in the question

Multiple cities named Toronto in the US

Toronto, Canada, has an American League baseball team

Watson found little evidence to connect either city's airport to WWII

With Toronto at 14% confidence, would not have buzzed in

Chicago at 11% was a very close second on list of possible answers
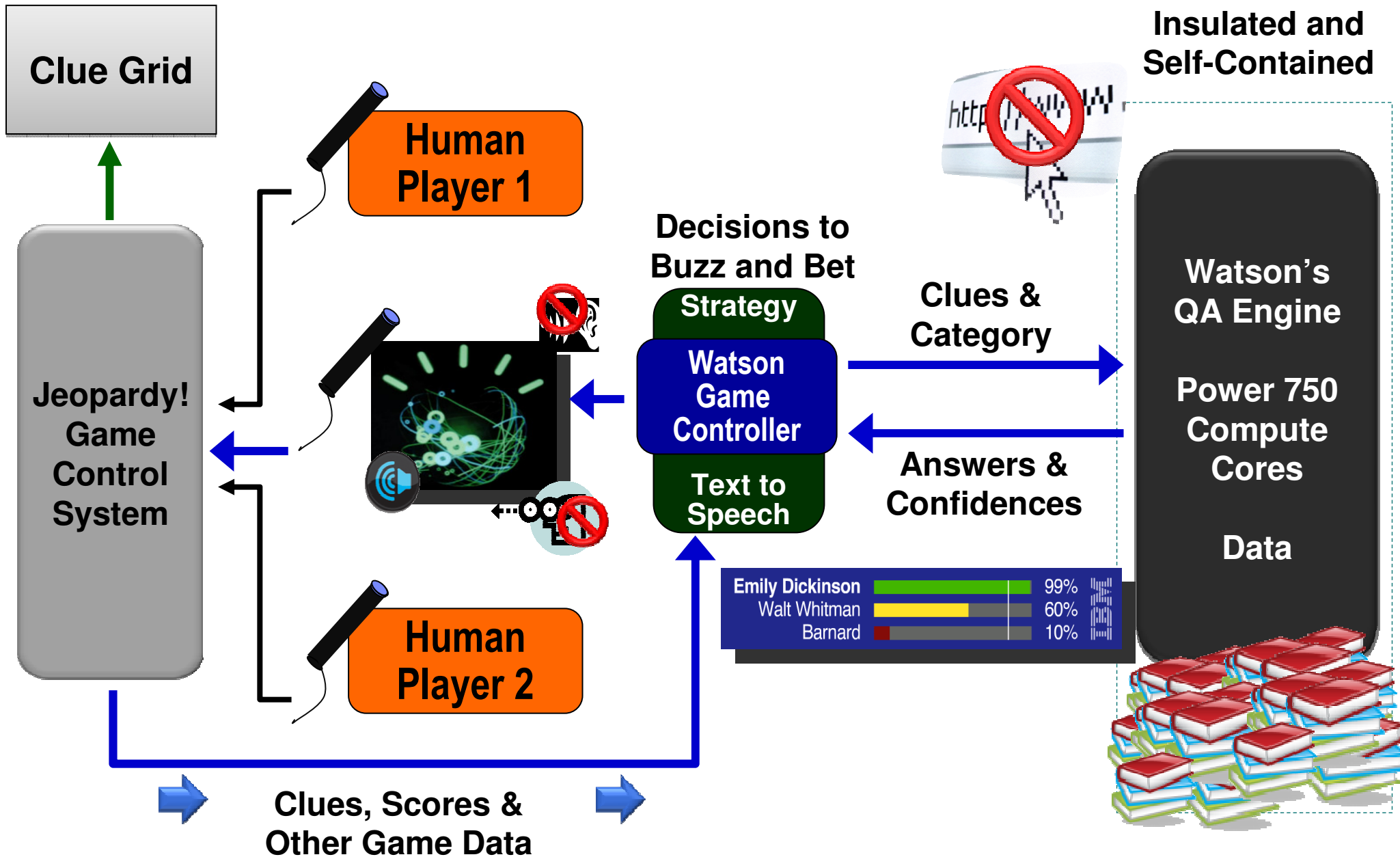
Easy for humans, difficult for Watson

US Cities

Its largest airport is named for a World War II hero, Its second largest for a World War II battle.

Watson: Toronto???
Brad: Chicago
Ken: Chicago

Real-Time Game Configuration

Clue Grid

Human Player 1

Jeopardy! Game Control System

Insulated and Self-Contained

Decisions to Buzz and Bet

Strategy

Watson Game Controller

Text to Speech

Clues & Category

Answers & Confidences

Watson's QA Engine

Power 750 Compute Cores

Data

Emily Dickinson		99%
Walt Whitman		60%
Barnard		10%

Human Player 2

Clues, Scores & Other Game Data

IBM WATSON

Trivia: Who is….

Who is Ed Toutant????

- Former Power Systems Engineer
- Former Game Show Contestant
- Watson Sparing Partner
- ????

IBM WATSON

Watson Reflections

Next Steps for Watson / IBM?

Commercialization of Watson/Power
- Power as strategic development platform in early pilots and productization
- Vectoring clients to existing Business Analytics and Information Mgmt offerings
  (e.g. pureScale, Cognos, Content Analytics …)
- "Roadmap to Watson" for longer-term Client engagements

Deep Collaboration
- Healthcare Pilot Applications (e.g. MD Anderson Cancer Center)
- Target broad scale applicability, validation of technology and business model
- Consumability of Jeopardy! accelerating analytics into daily practice
- Technology transfer agents for other research and commercialization activity

Client Engagements
- Collaborative STG/SWG/GBS/Research focus
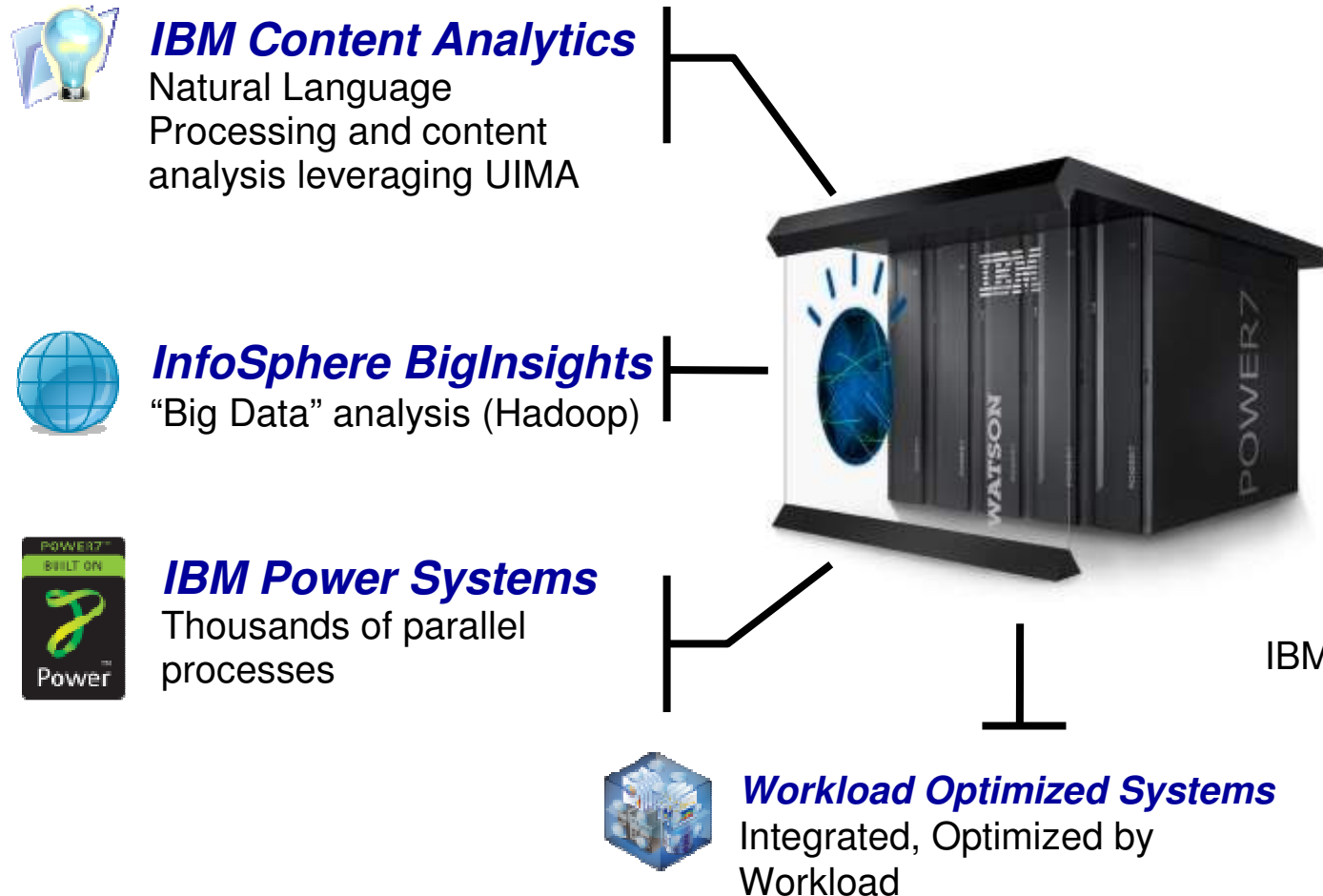- Identify collaborative opportunities, client priorities, and next steps
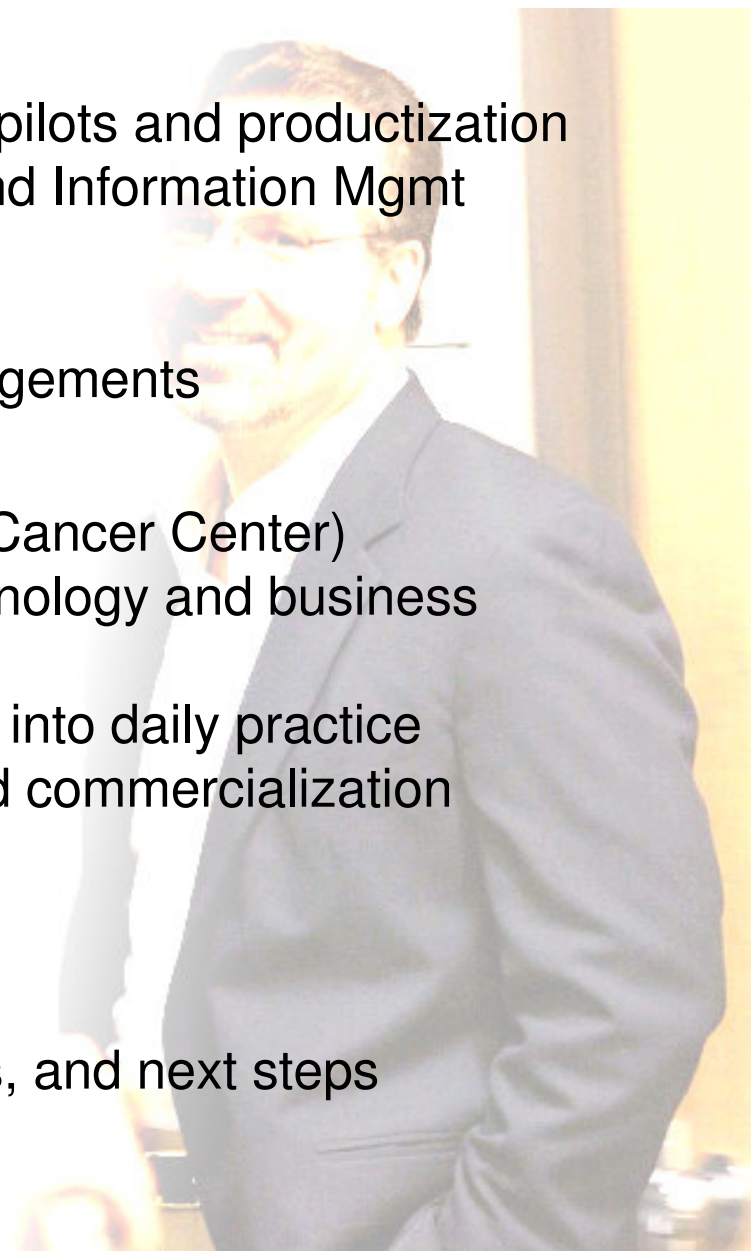
Bridging from Watson to the present …..

IBM Business Analytics and Optimization solutions

Used by Watson

IBM Content Analytics
Natural Language Processing and content analysis leveraging UIMA

InfoSphere BigInsights
"Big Data" analysis (Hadoop)

IBM Power Systems
Thousands of parallel processes

Workload Optimized Systems
Integrated, Optimized by Workload

Related Innovations

InfoSphere Warehouse DB2, Informix, Netezza
Aggregating and storing data and content

InfoSphere Streams
Massively parallel analysis

Business Analytics
BI, Predictive Analytics and more

ECM Solutions
IBM eDiscovery Analyzer
IBM Classification Module
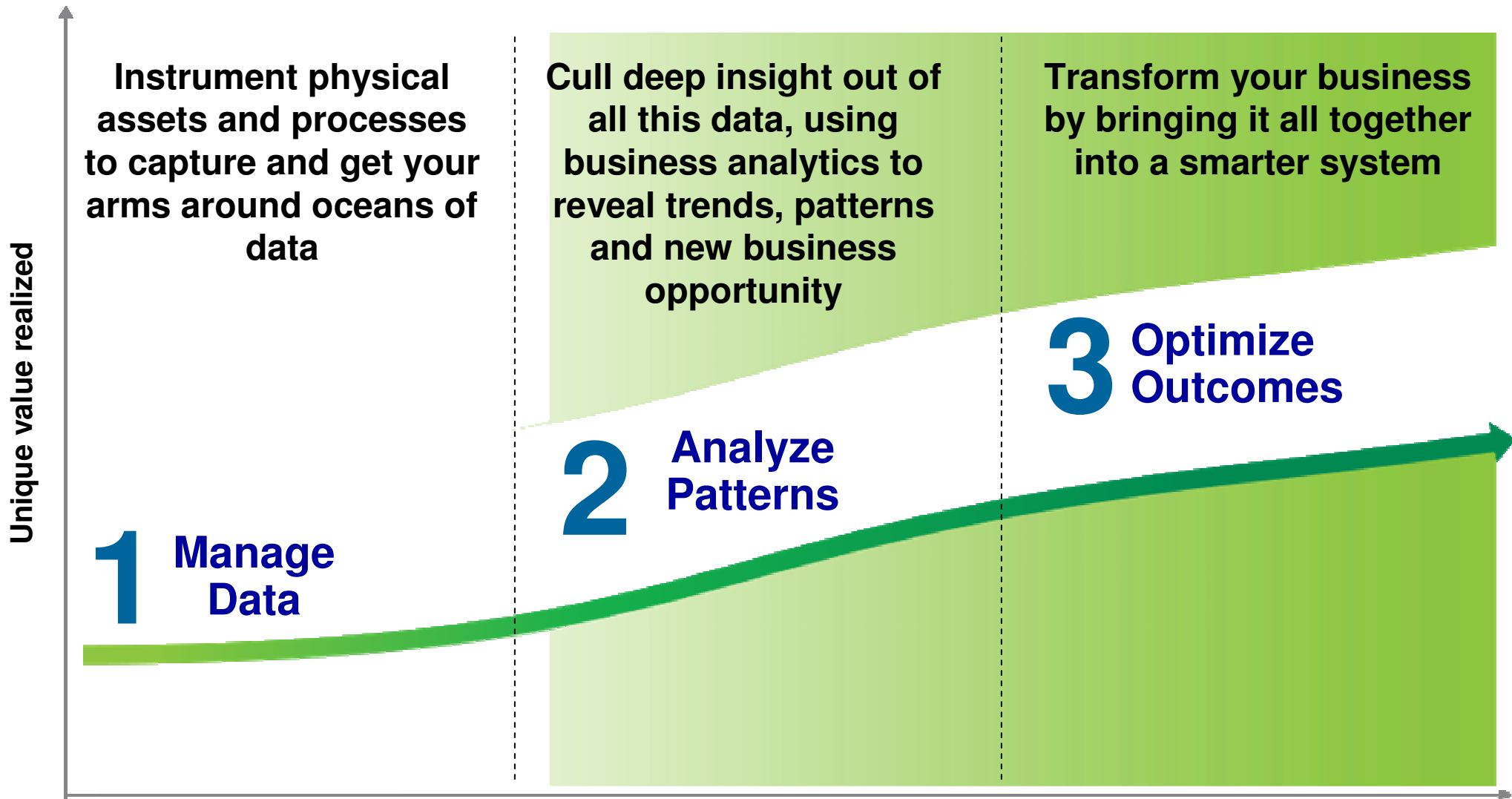IBM OmniFind Enterprise Search

IBM Global Business Services
Research, expertise and analytical assets

IBM WATSON

Applying Watson's capabilities for business

Helping clients with deployments which tend to follow three phases:

Unique value realized

Instrument physical assets and processes to capture and get your arms around oceans of data

Cull deep insight out of all this data, using business analytics to reveal trends, patterns and new business opportunity

Transform your business by bringing it all together into a smarter system

3 Optimize Outcomes

2 Analyze Patterns

1 Manage Data

IBM WATSON

Turning information into insights

Business Analytics

- Analytic Applications
- Business Intelligence
- Predictive Analytics
- Financial Performance Management
- Governance, Risk & Compliance
- Web Analytics

Enterprise Content Management

- Advanced Case Management
- Content Analytics
- Document Imaging and Capture
- Information Lifecycle Governance
- Social Content Management

Information Integration & Federation

- Information Integration
- Master Data Management
- Data Warehousing
- Big Data and Streams

Data Management

- Database Software
- Database Management Tools

Information Governance

- Data Lifecycle Management
- Data Security and Privacy
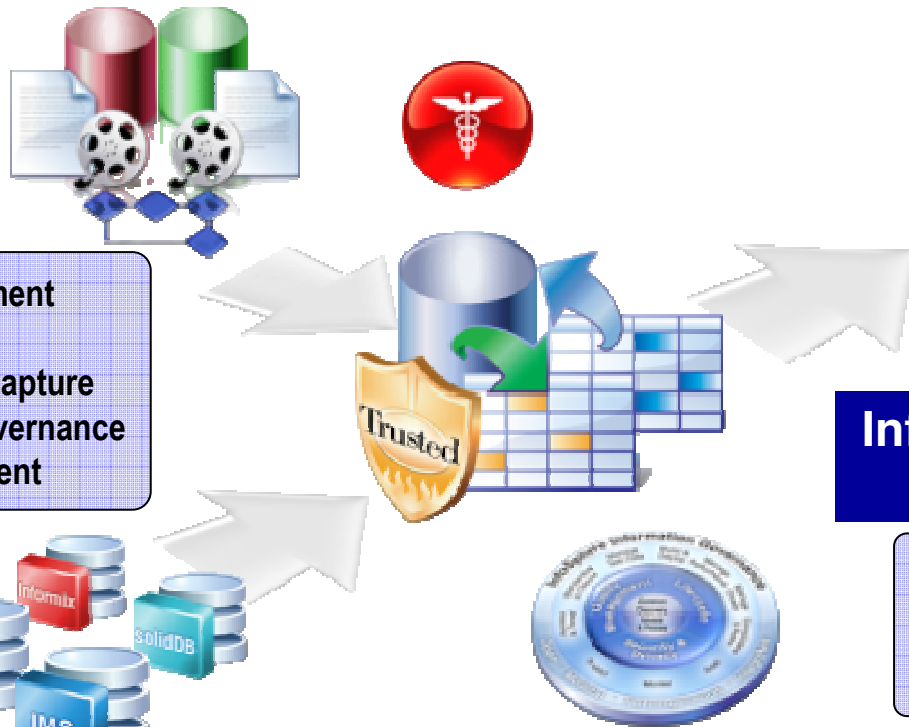
Potential Business Applications

Healthcare / **Life Sciences**: Diagnostic Assistance, Evidenced-Based, Collaborative Medicine

Tech Support: Help-desk, Contact Centers

Enterprise Knowledge Management and Business Intelligence

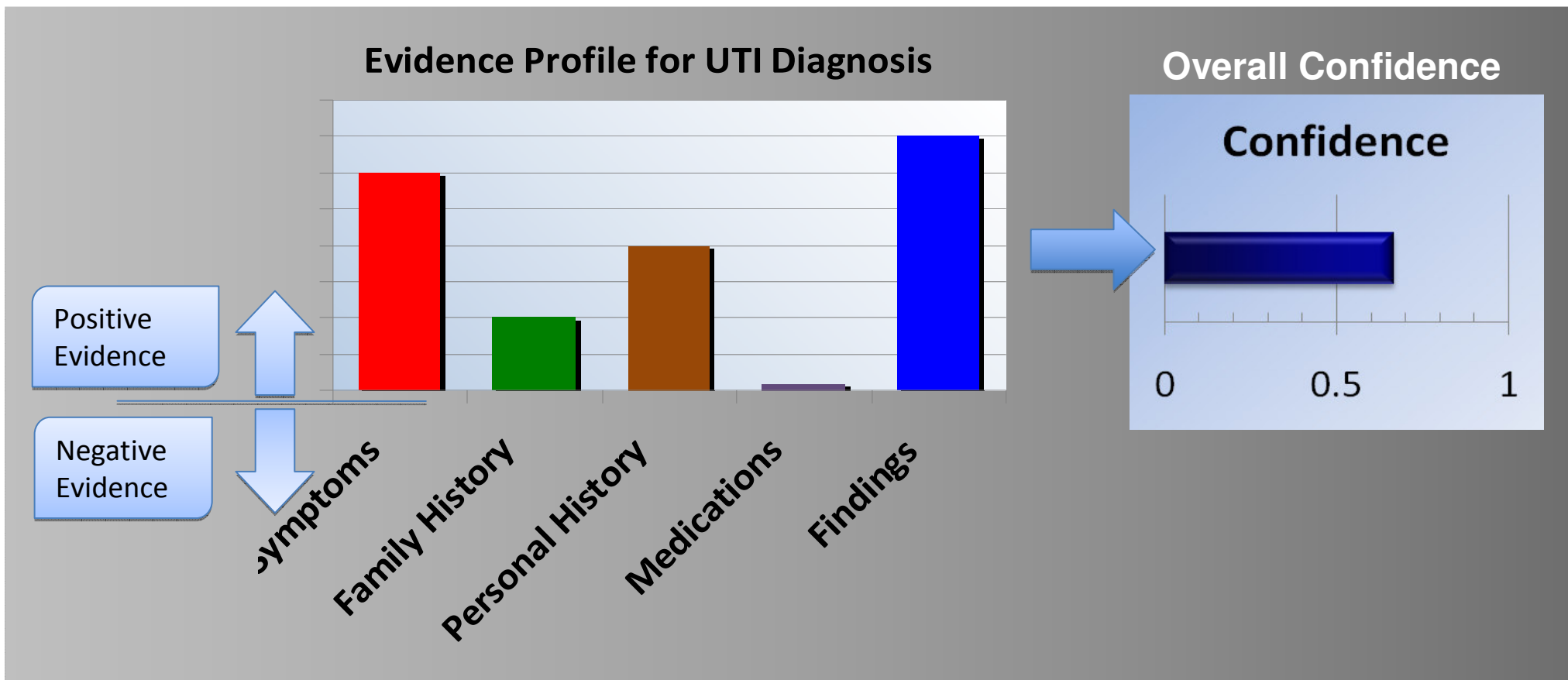**Government:** Improved Information Sharing and Security

Evidence Profiles from disparate data sources

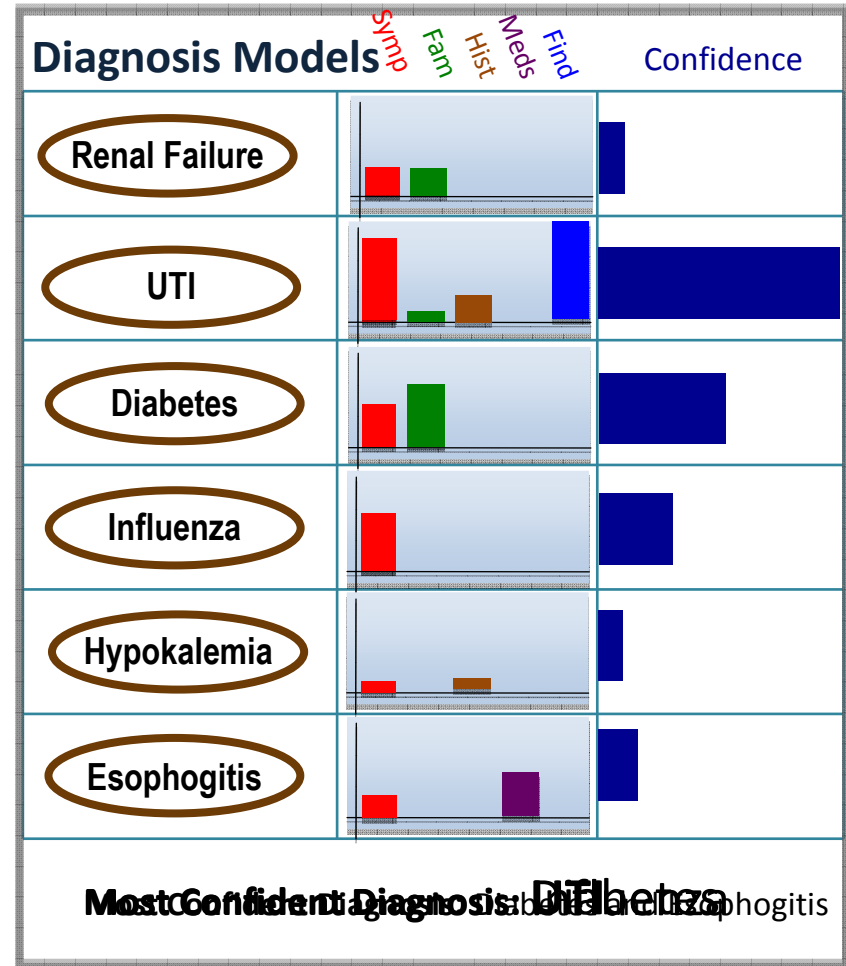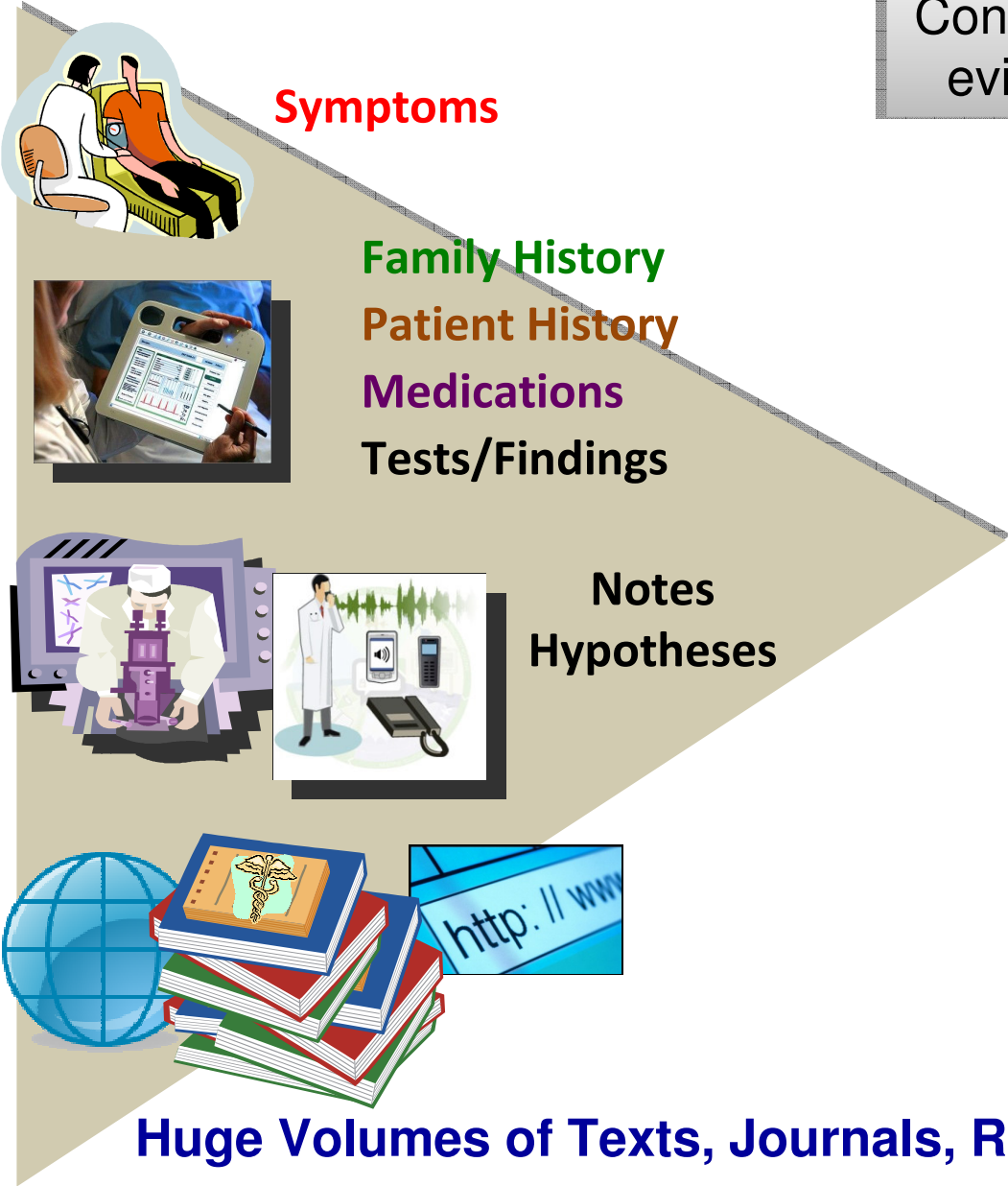Each dimension contributes to supporting or refuting hypotheses based on
- **Strength of evidence**
- **Importance of dimension for diagnosis** (learned from training data)

Evidence dimensions are combined to produce an overall confidence

Continuous Evidence-Based Diagnostic Analysis

Symptoms

Family History

Patient History

Medications

Tests/Findings

Notes

Hypotheses

Huge Volumes of Texts, Journals, References, DBs etc.

Considers and synthesizes a broad range of evidence improving quality, reducing cost

Diagnosis Models — Symp Fam Hist Meds Find — Confidence

- Renal Failure
- UTI
- Diabetes
- Influenza
- Hypokalemia
- Esophogitis

Most Confident Diagnosis: Diabetes

Active Watson / DeepQA Engagements

Client inquiries arriving with a broad range of Watson use cases, from a broad range of industries
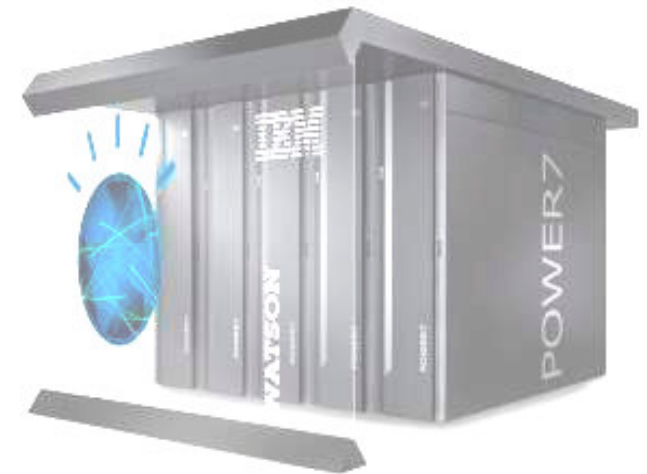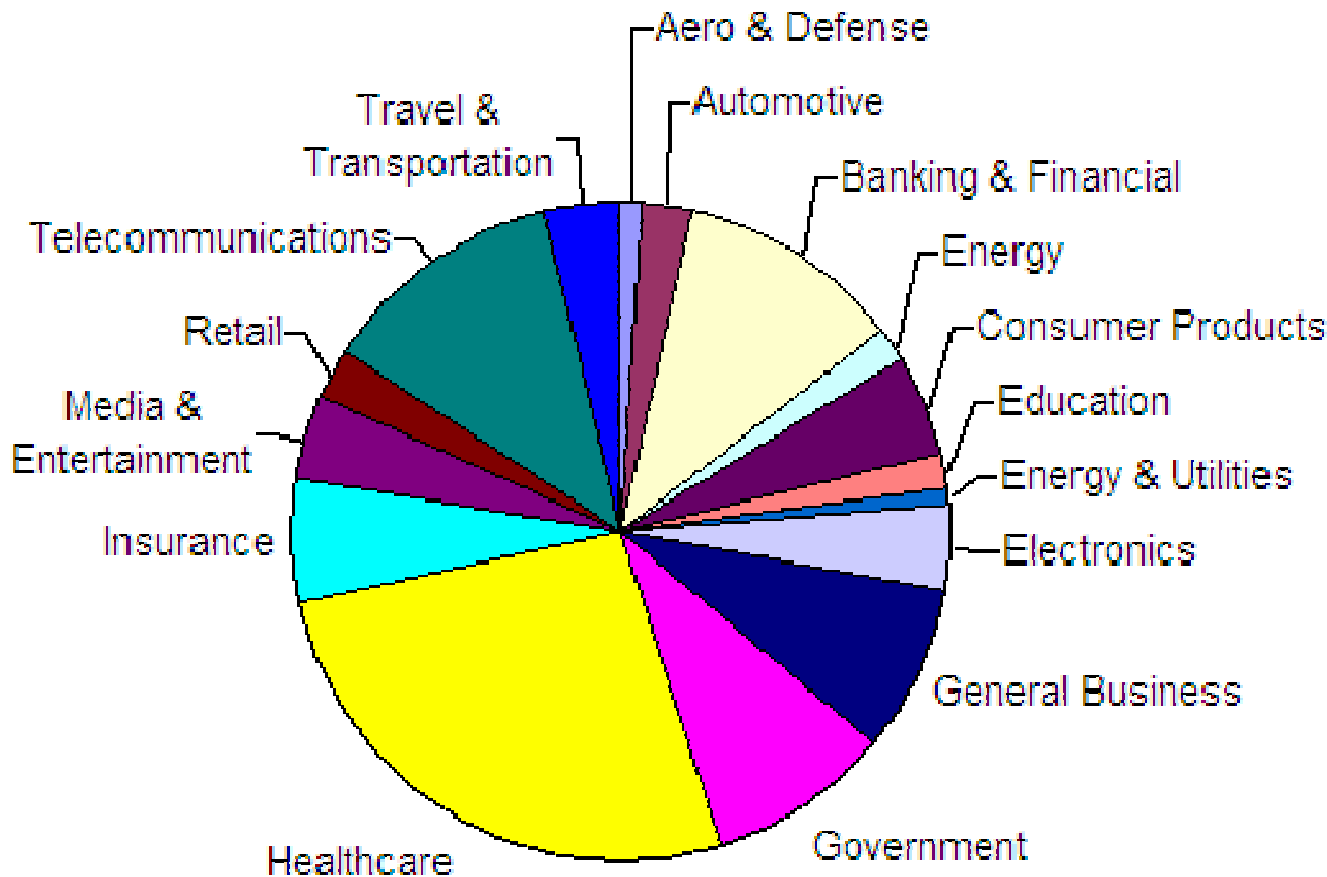
Past History….

Technology that could compete against man

IBM WATSON

The Evolution of Science

Era of Natural Philosophy

Era of Modern Science

Learning Systems (XXI Century)

DNA (1953 AD)

Computing (1946 AD)

Industrial Revolution

Evolution of Science

Platonic Academy (387 BC)

Quantum Physics (1925 AD)

Astronomy (Babylon, 1900 BC)

Mathematics (India, 499 BC)

Scientific Revolution (1543 AD)

Newton's Laws (1687 AD)

Relativity (1905 AD)

Time

IBM WATSON
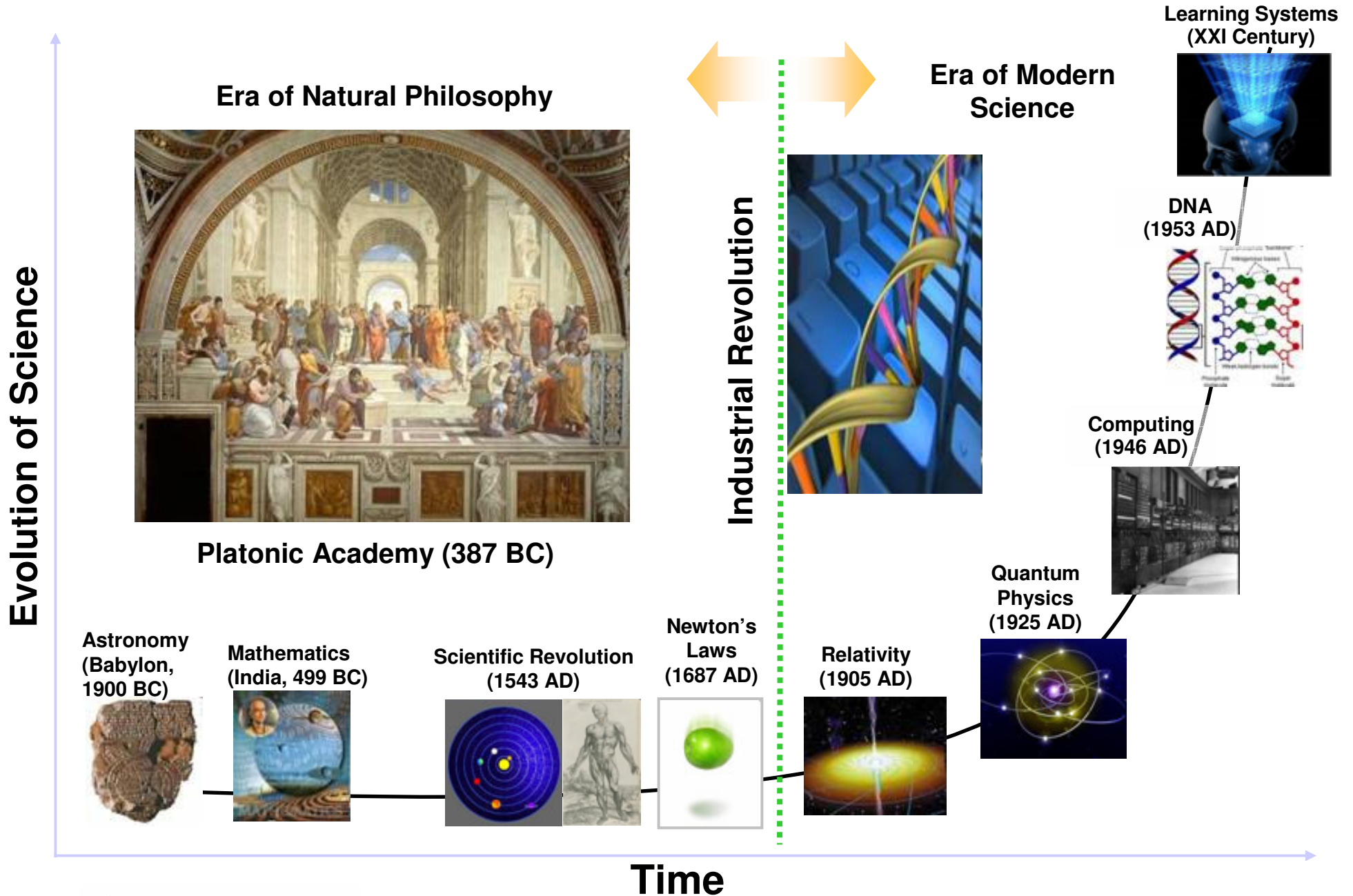
The Evolution of Technology

Achieve understanding of natural language, images and other sensory information

Computer Intelligence

Watson (2010)

Deep Blue (1997)

ENIAC (Circa 1945)

System/360 (1964)

Abacus (Circa 3500 BC)

Napier's rods (Circa 1600)

Antikythera Astronomical Computer (ca 87 BC)

Counting Machine (Circa 1820)

Time

IBM WATSON
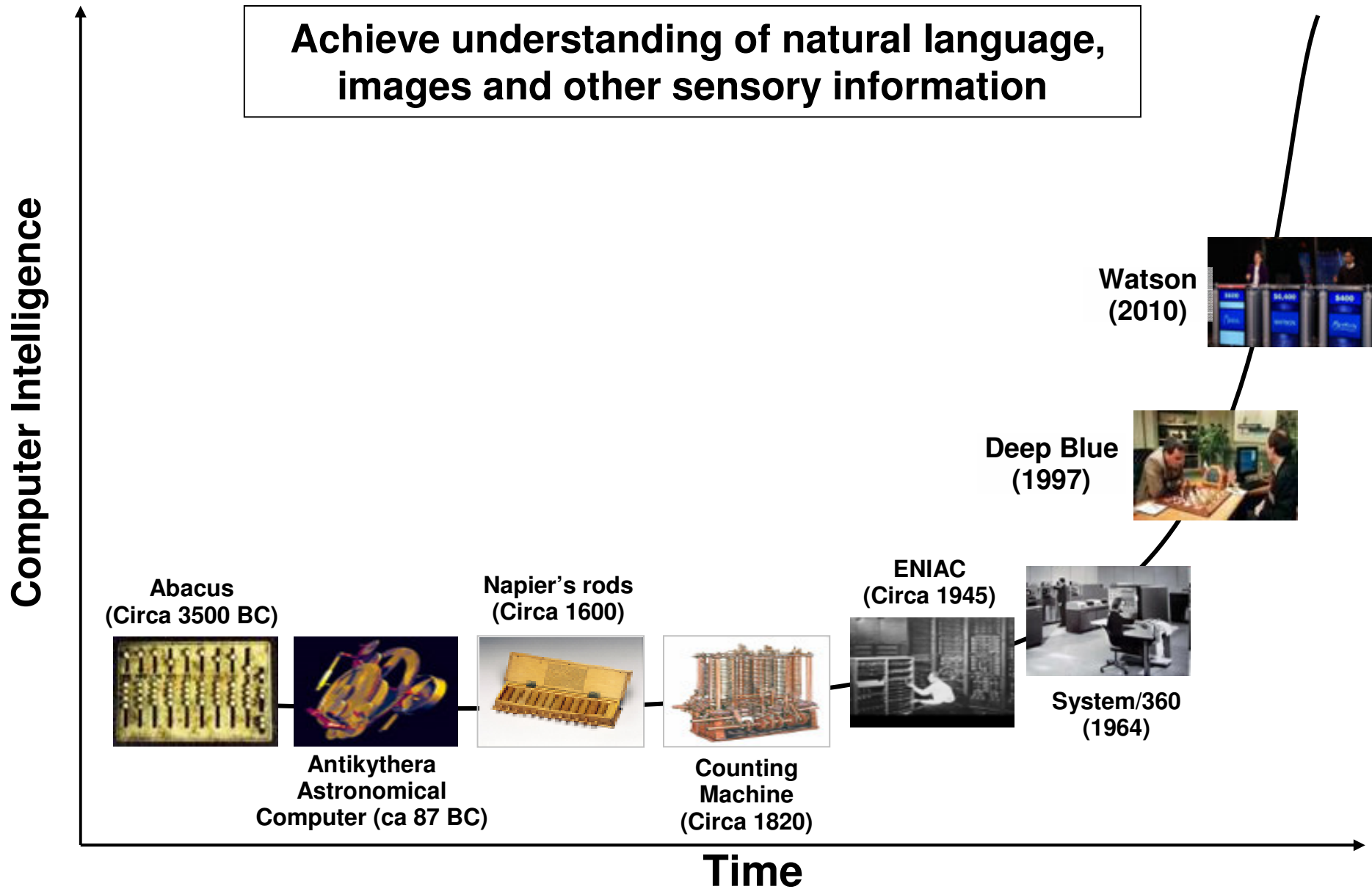
Device and Technology Roadmap
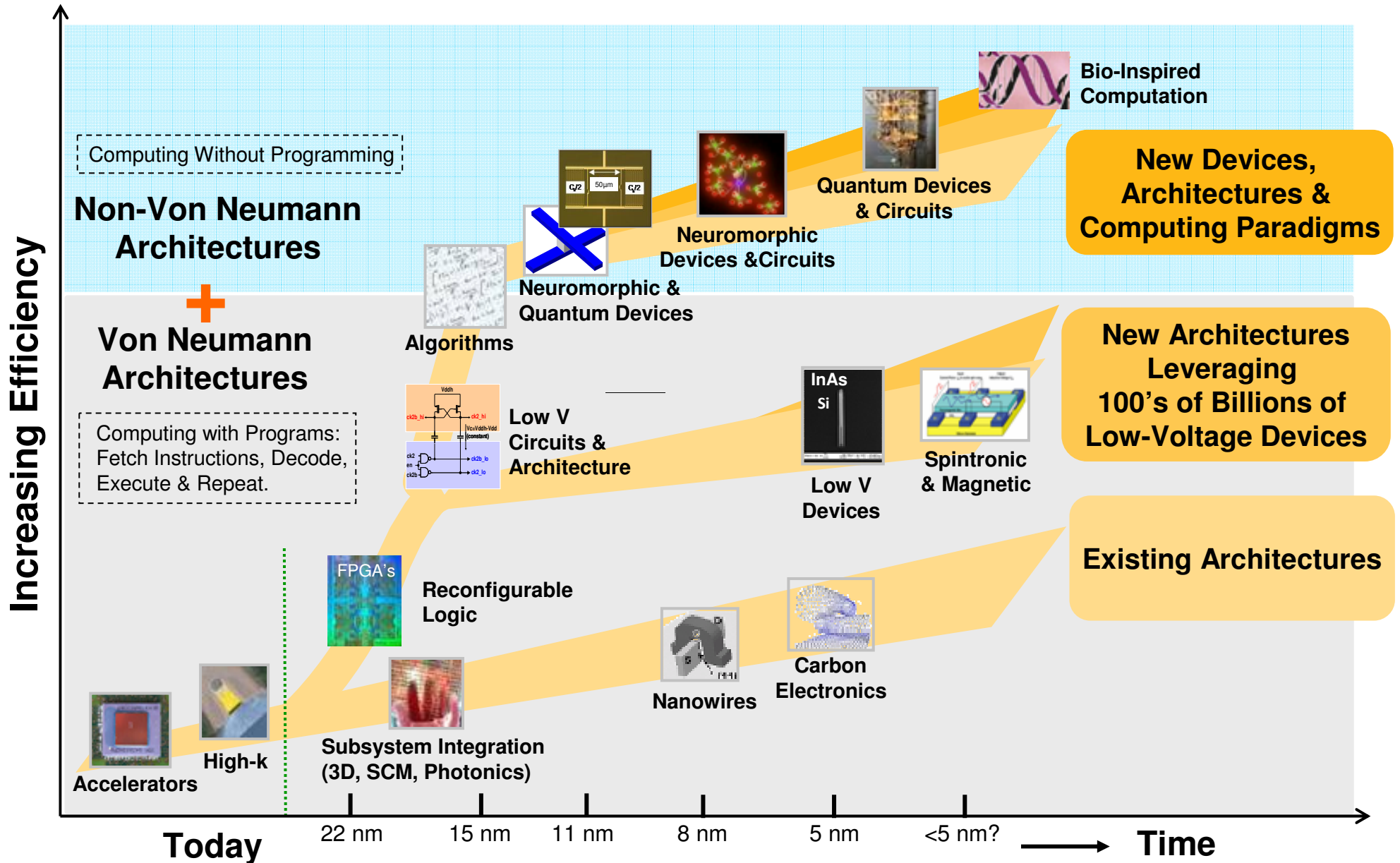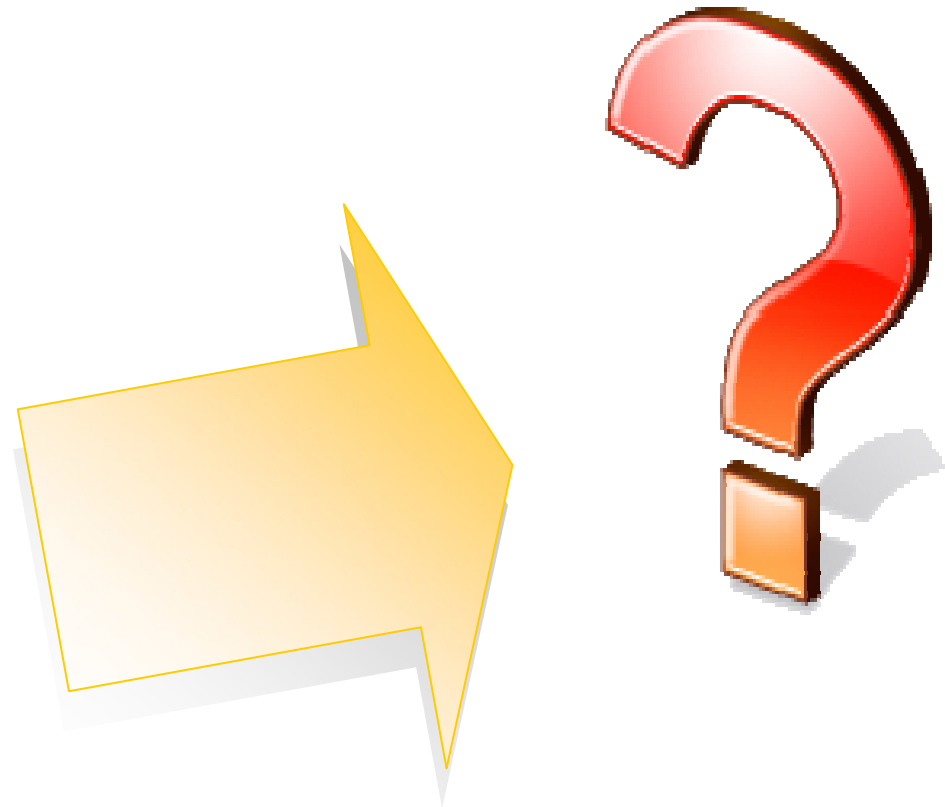
Technology that can think/reason like man

Data Problem of the future…..

Future Data Components

Audio / Voice

Images

Video

3D Content

Dynamic

Real Time

Sensory
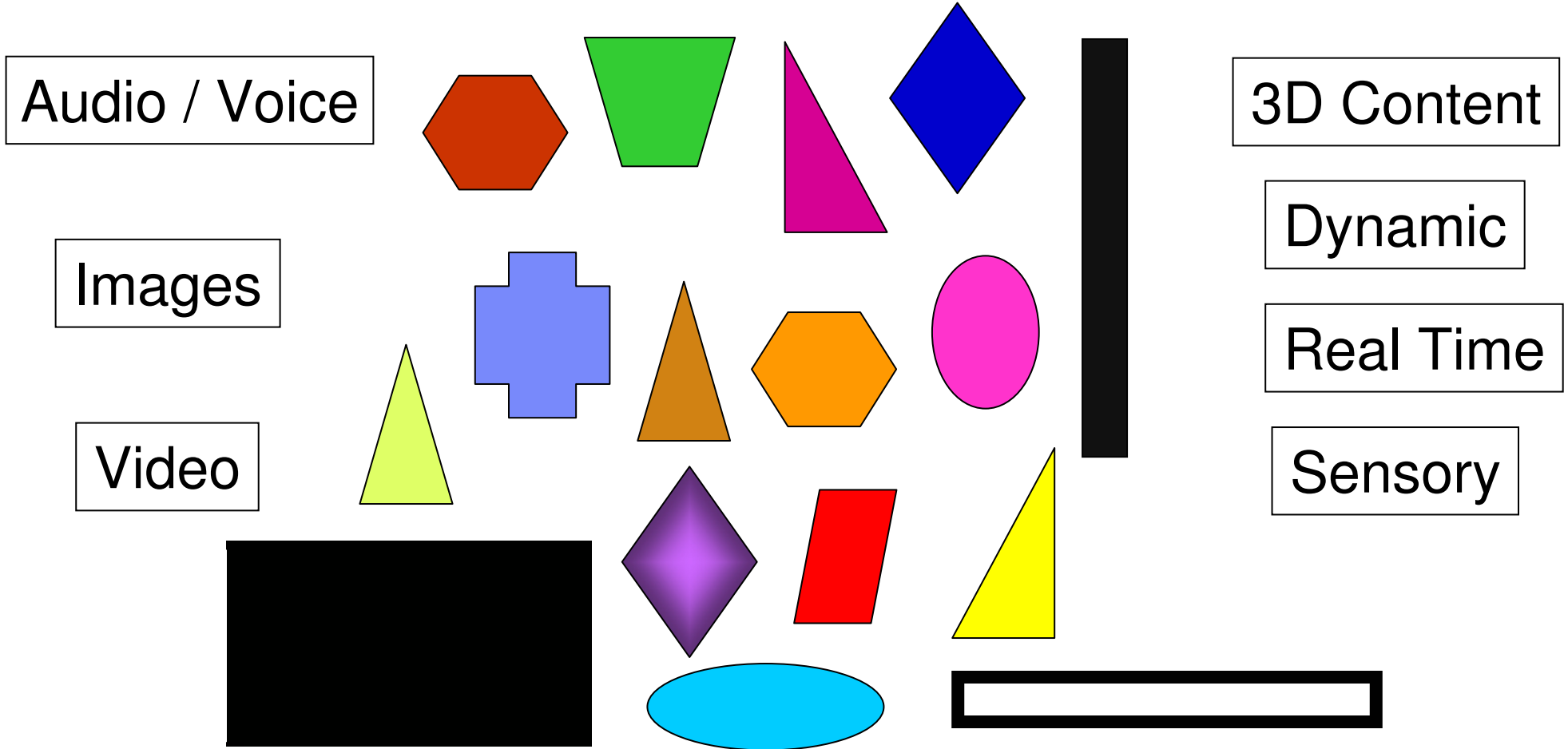
The Evolution of "Thinking" Machines

New IT Frontier

Learning Systems 20XX

Achieve understanding of natural language, images and other sensory information

Computer Intelligence

Watson (2011)

Deep Blue (1997)

System/360 (1964)

ENIAC (Circa 1945)

Counting Machine (Circa 1820)

Napier's rods (Circa 1600)

Antikythera Astronomical Computer (ca 87 BC)

Abacus (Circa 3500 BC)

Time

IBM WATSON
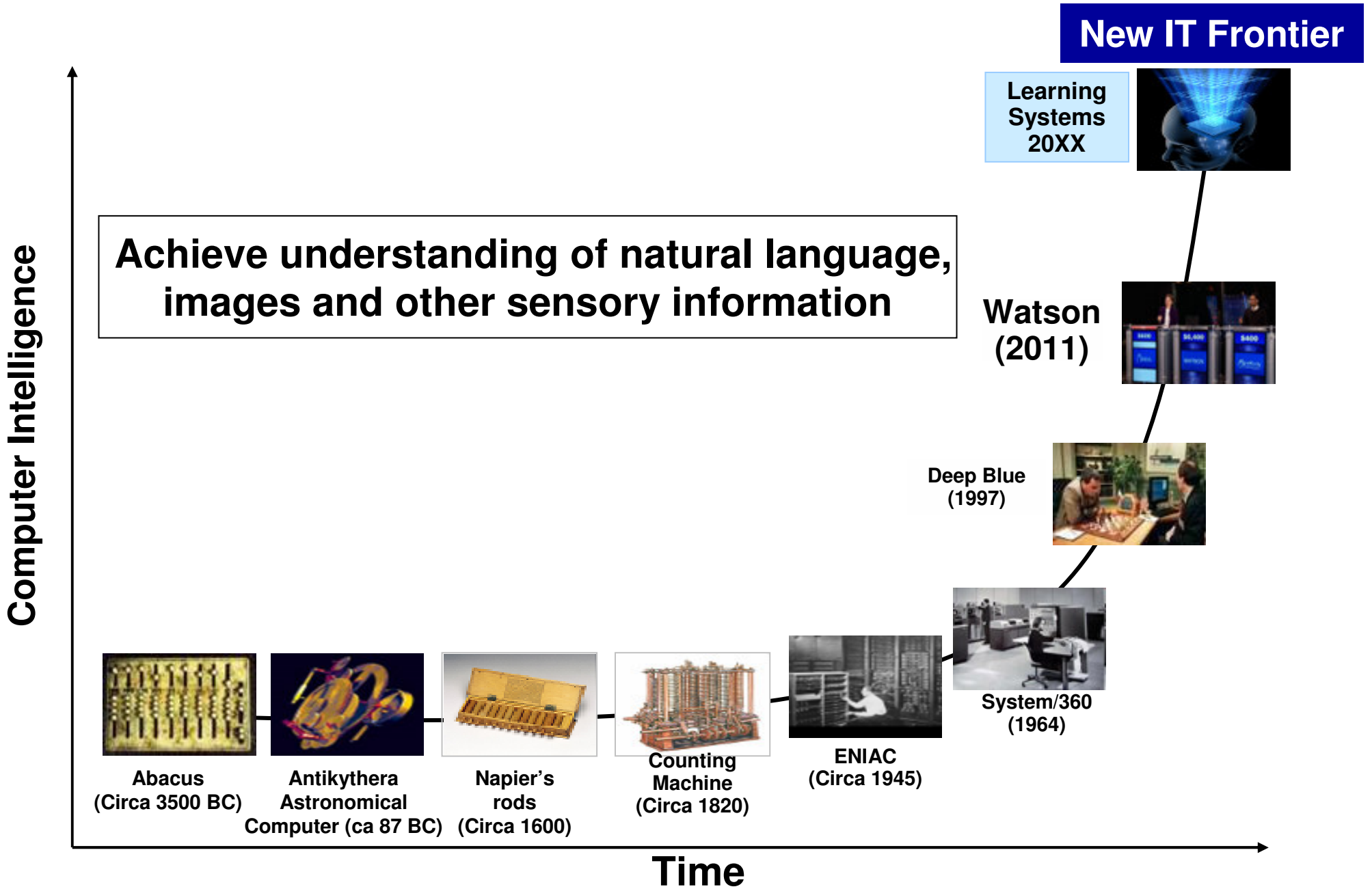
Learning Systems Roadmap to Meet the Challenge

The Research Division will ensure IBM is the world leader in Learning Systems. We will define & benchmark progress through a series of Grand Challenges.

Dynamic Learning Systems

Autonomous Learning Systems

Achieve understanding of natural language, images and other sensory information. Hypothesis and question generation across arbitrary domains; meta-heuristic to automate algorithm choices
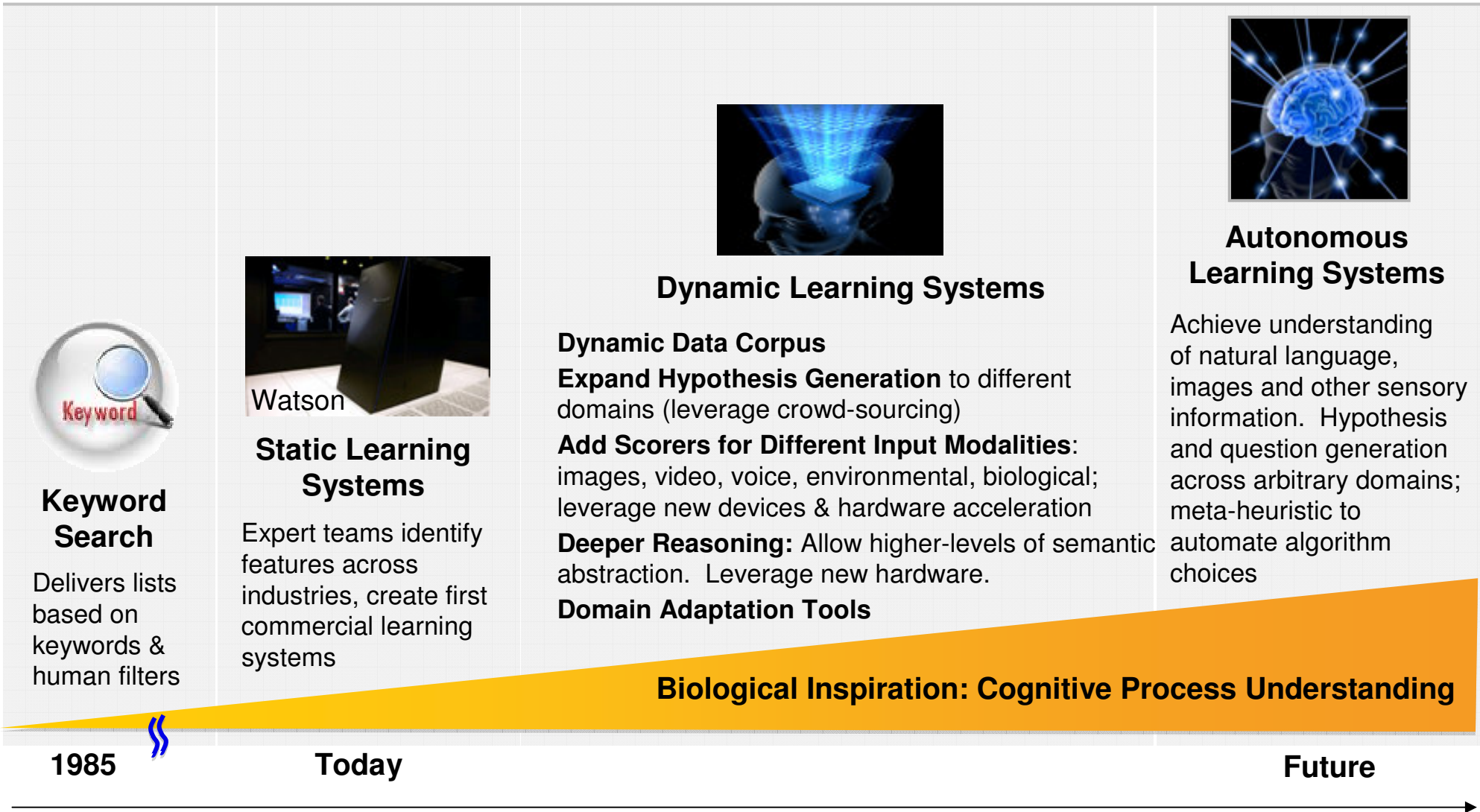
Dynamic Data Corpus

Expand Hypothesis Generation to different domains (leverage crowd-sourcing)

Add Scorers for Different Input Modalities: images, video, voice, environmental, biological; leverage new devices & hardware acceleration

Deeper Reasoning: Allow higher-levels of semantic abstraction. Leverage new hardware.

Domain Adaptation Tools

Static Learning Systems

Watson

Expert teams identify features across industries, create first commercial learning systems

Keyword Search

Delivers lists based on keywords & human filters

Biological Inspiration: Cognitive Process Understanding

| 1985 | Today | Future |

Greater Autonomy

44   IBM WATSON

The Future ….

Watson
It is just the beginning…