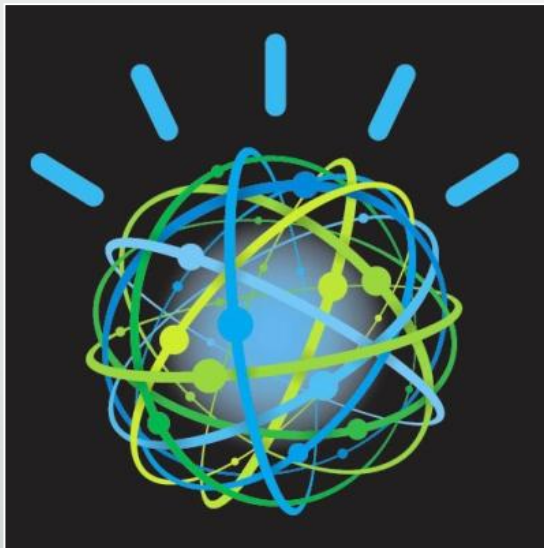




Technical Forum & Executive Briefing

17 al 21
Octubre
2011

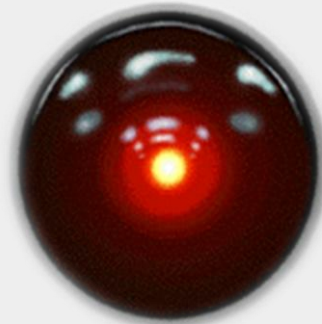
Imagine **PODER** Imagine **CAPACIDAD**



Watson

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

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: Two most successful Jeopardy contestants of all time**

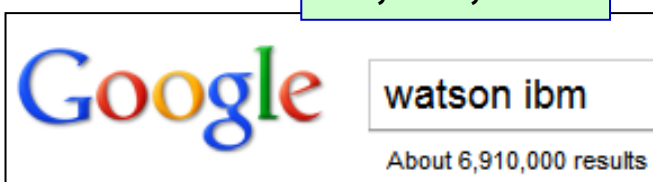
Watson's Popularity...

2,300,000+



The New York Times

78,500,000+



The Washington Post



2011 Webby Person of the Year!



Jeopardy! Basic Game Play

6 Topic Areas
5 Levels of Difficulty



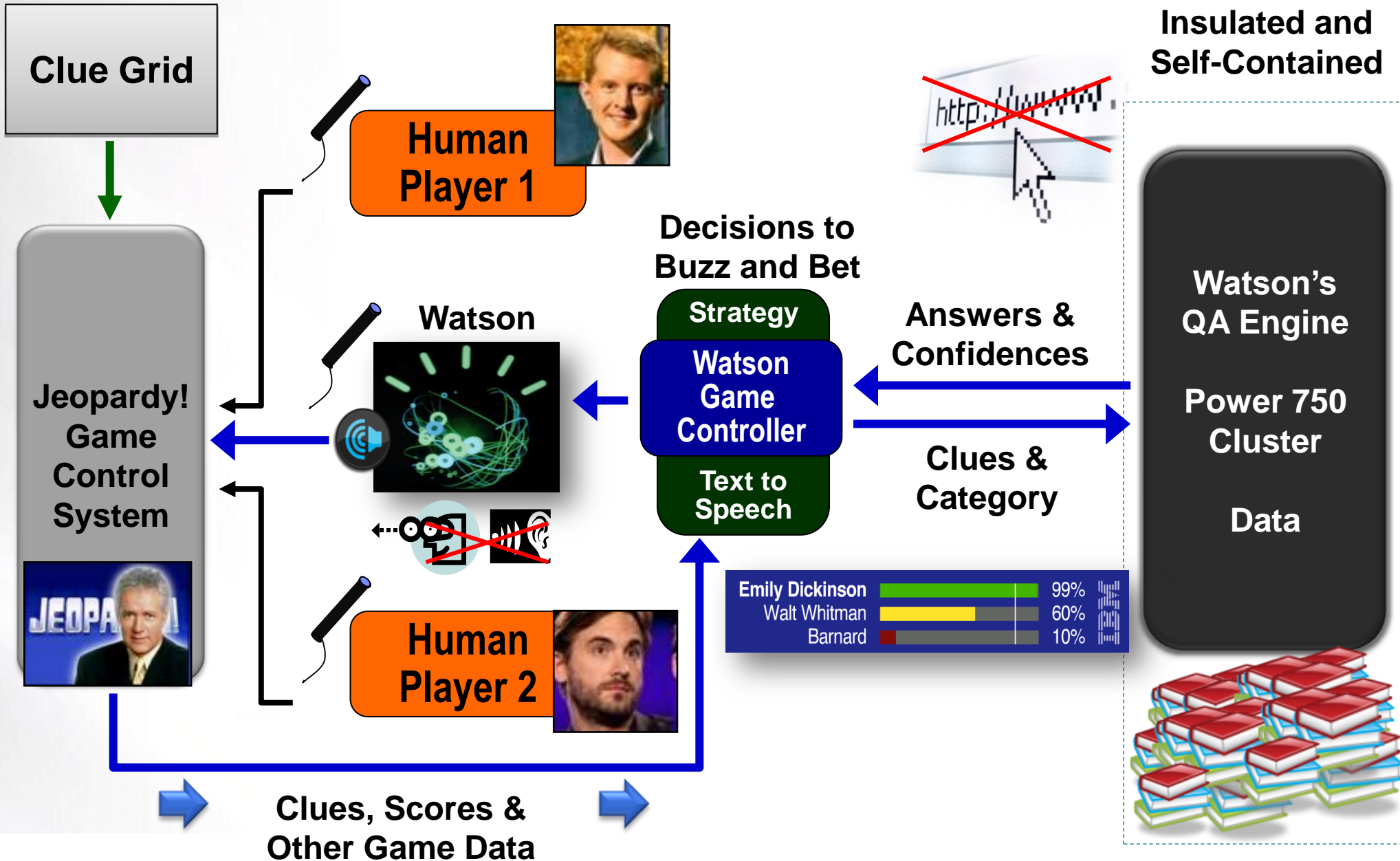
Cells	Classics	The Great Outdoors	Speak of the Dickens	Mind Your Manners	Before and After
\$200	\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000	\$1000

- 1 of 3 Players Selects a Clue
- Host reads the “Clue” out loud
- All Players compete to answer
- 1st to buzz-in gets to answer

- If Correct: Earns \$ value**
 - **Selects Next Clue**
- If Wrong: Loses \$ value**
 - **Other players buzz again**

Two Rounds Per Game + Final Question

Real-Time Game Configuration



The Jeopardy! Challenge: *A compelling and notable way to drive and measure the technology of automatic Question Answering along 5 Key Dimensions*



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

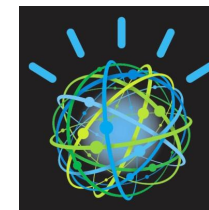
\$1000
 The first person mentioned by name in 'The Man in the Iron Mask' is this hero of a previous book by the same author.

\$600
 In cell division, mitosis splits the nucleus & cytokinesis splits this liquid cushioning the nucleus

\$2000
 Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north

Word Plays Puns Rhymes
Lyrics Dual meanings

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

1997

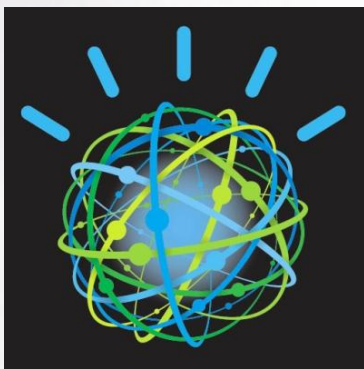


Deep Blue



Game	Chess
Architecture	Specialized Hardware and Power2 SC
# Cores	30 + 480 ASICs
Hardware	Highly Specialized
Processing	Mathematical
Data Analysis	<ul style="list-style-type: none"> ▪ Finite number of possible moves and countermoves ▪ Structured data ▪ Mathematical probability

Watson Info.....



Hardware: Cluster of 90 Power 750 (2880 Cores) @ 3.55 GHz

- **88 Compute nodes, 2 I/O nodes, 4 SAS Storage drwrs & 2 xCAT Servers**

Early implementations of Watson ran on a single processor

- **Required 2 hours to answer a single question**

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**

Data Sources

Sources of information for Watson:

- Encyclopedias
- Dictionaries
- Thesaurus
- Newswire articles
- Literary works
- Bible
- Databases, taxonomies, and ontologies.
 - IMDb
 - DBpedia
 - Wordnet
 - YAGO
- Wikipedia (Full text)

200 million pages of structured and unstructured content

- 1 Million books
- 4 TB of disk storage
- Watson: 15 TB of memory

Watson Challenge.....

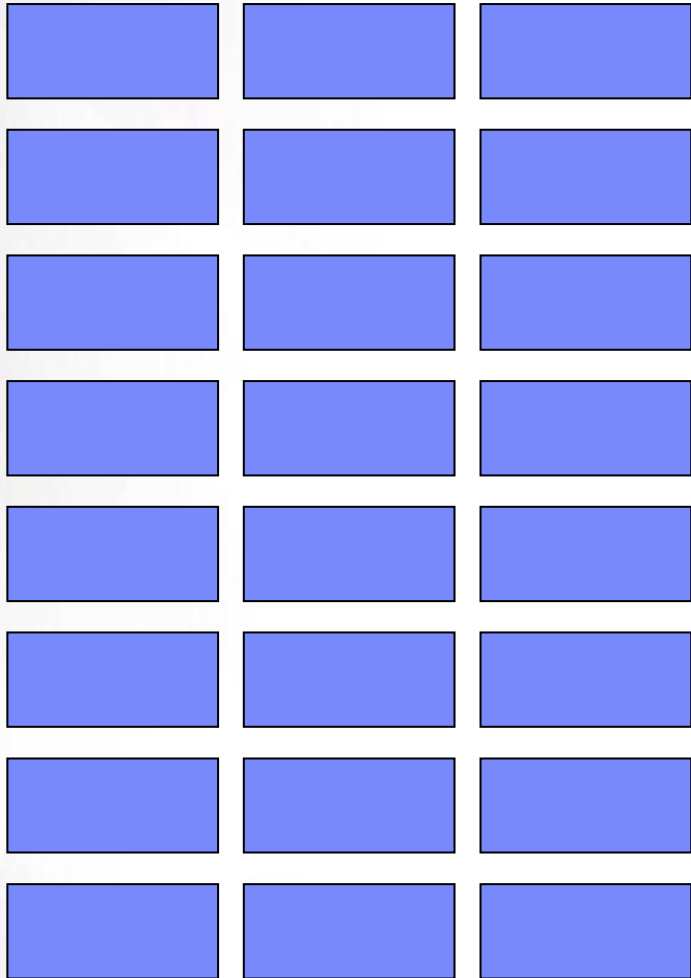


**How to Process / Analyze / Evaluate /
Prioritize all of this **Unstructured Data****

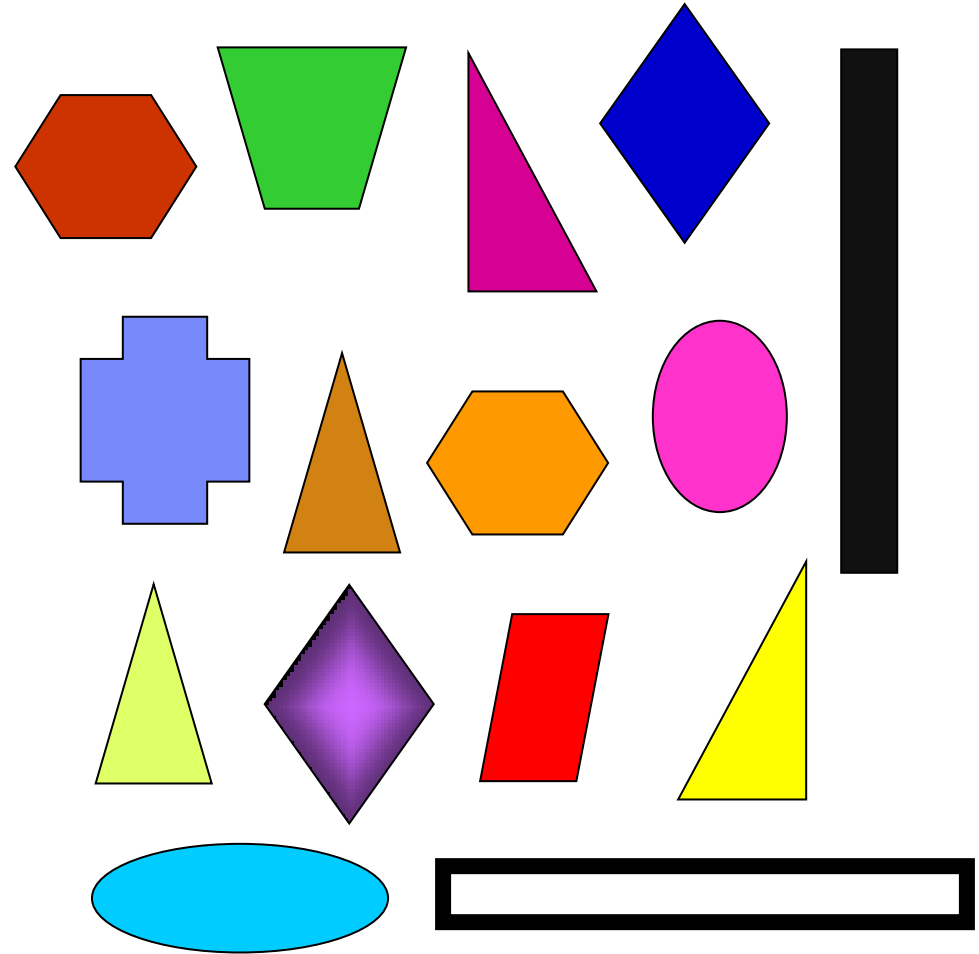
Live Audience of 10M

Data Problem.....

Structured Data



Unstructured Data



Easy Questions for Computers.....

$$25 + 25 = 50$$

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

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

Owner	Serial Number
David Jones	45322190-AK

Serial Number	Type	Invoice #
45322190-AK	LapTop	INV10895

Invoice #	Vendor	Payment
INV10895	MyBuy	\$104.56

↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
 David Jones
 David Jones

=

↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
 Dave Jones
 David Jones

≠

Structured / UnStructured Data & Questions....

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

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

Structured / Easy

City	Altitude
Mexico City	7,217 feet)

City	GDP (Pecos)
Mexico City	1.527 trillion

UnStructured / Hard

Paris, Stuttgart, and Kaliningrad

What are sister cities of Mexico City?

There are more of these in Mexico City than anyplace in the world..

What are museums?

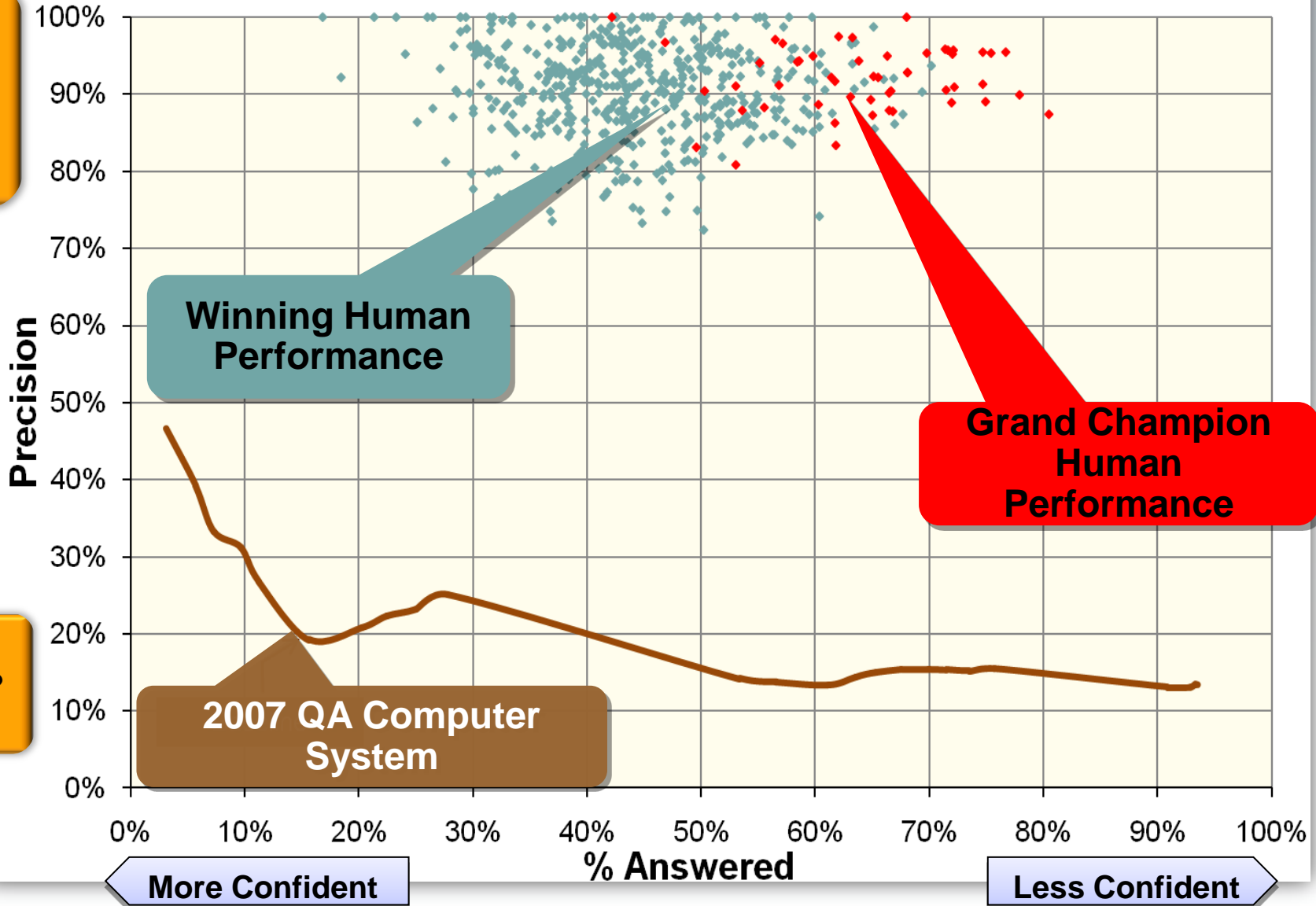
1012

How many people work at Grand Velas work during 24 Hr shift?

Best Human Performance.....

Each dot represents an actual historical human Jeopardy! game

Top human players are remarkably good.



Computers?

2007 QA Computer System

UIMA / Apache History....

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

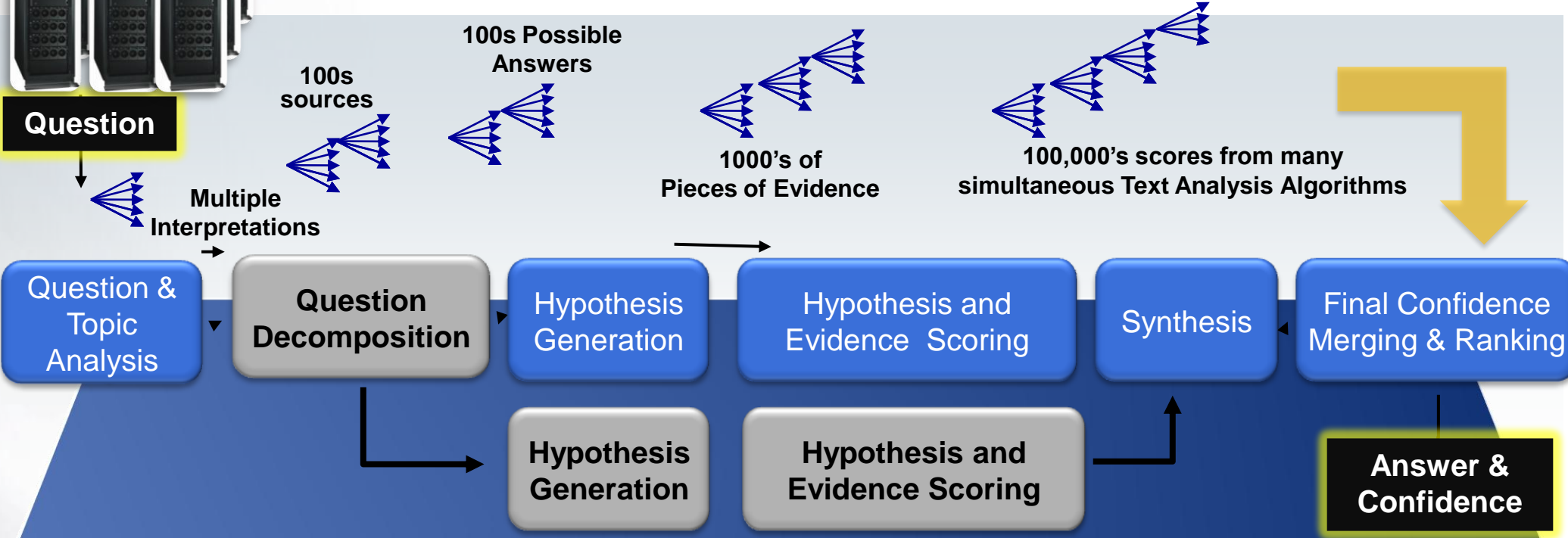
- Can be divided into independent parts, each executed by a separate processor / Computation is embarrassing parallel
- Gathers, evaluates, weighs and balances different types of evidence, delivering the answer with the best support it can find.

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.***



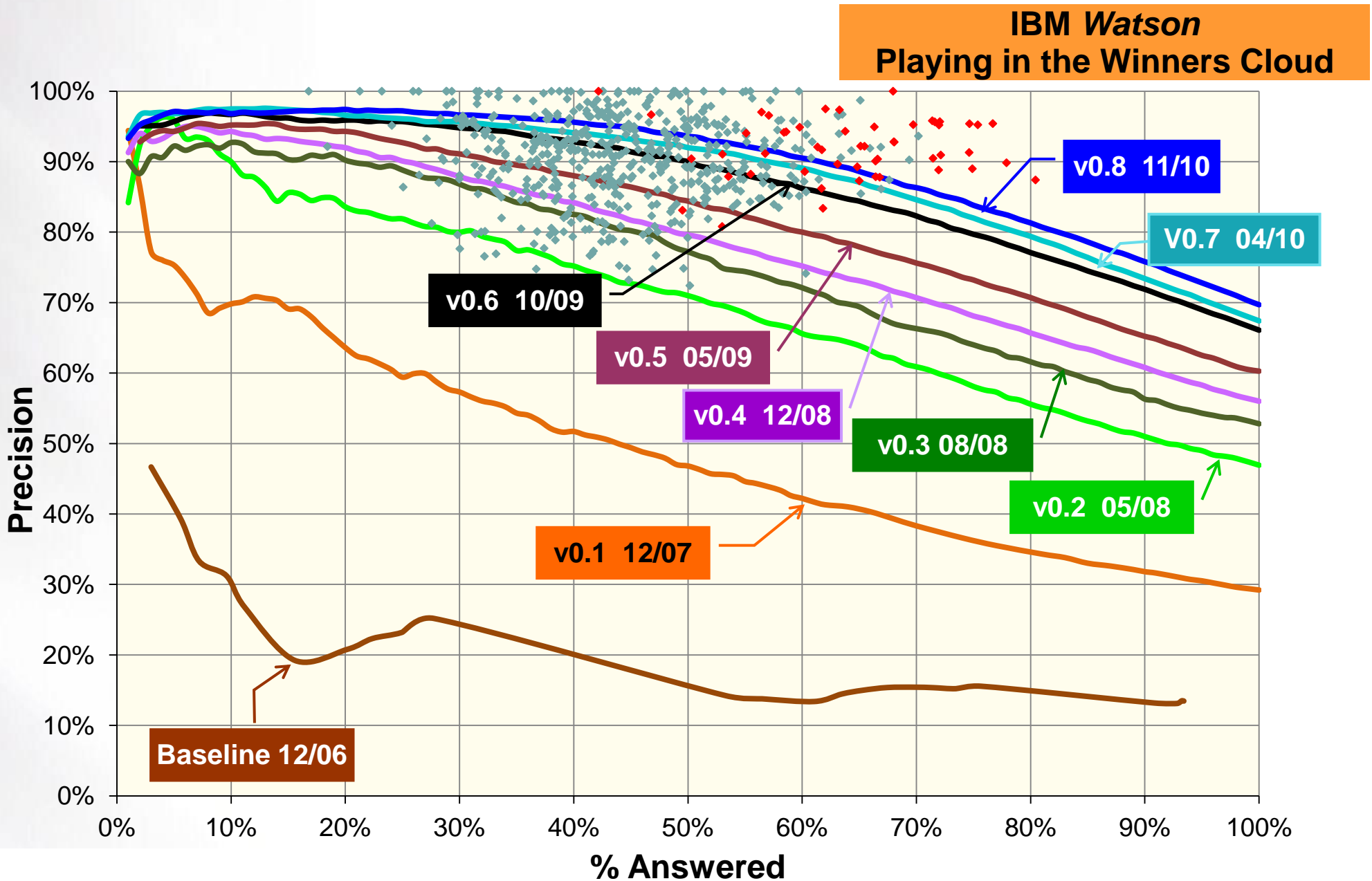
Question



built on UIMA for interoperability

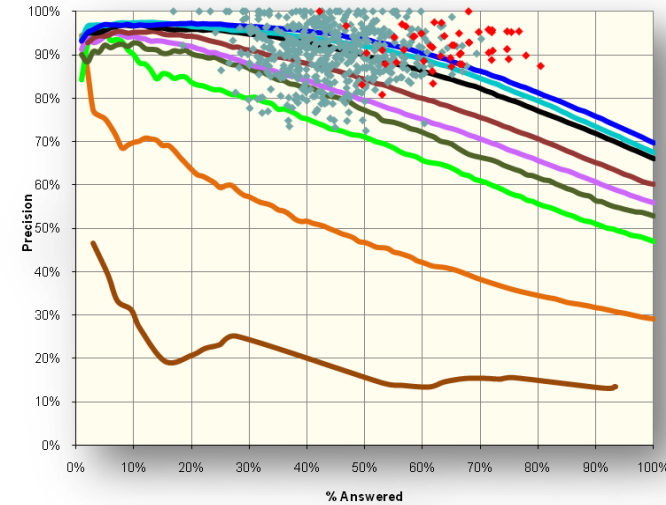
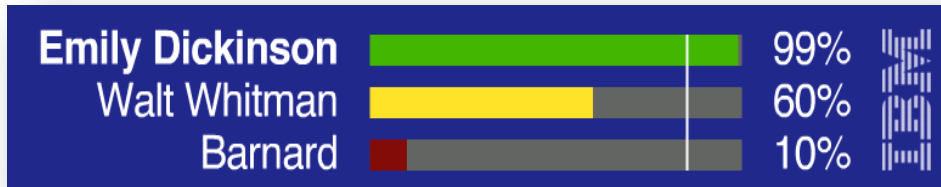
built on UIMA-AS for scale-out and speed

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



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.



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 – **Watson won Jeopardy!**

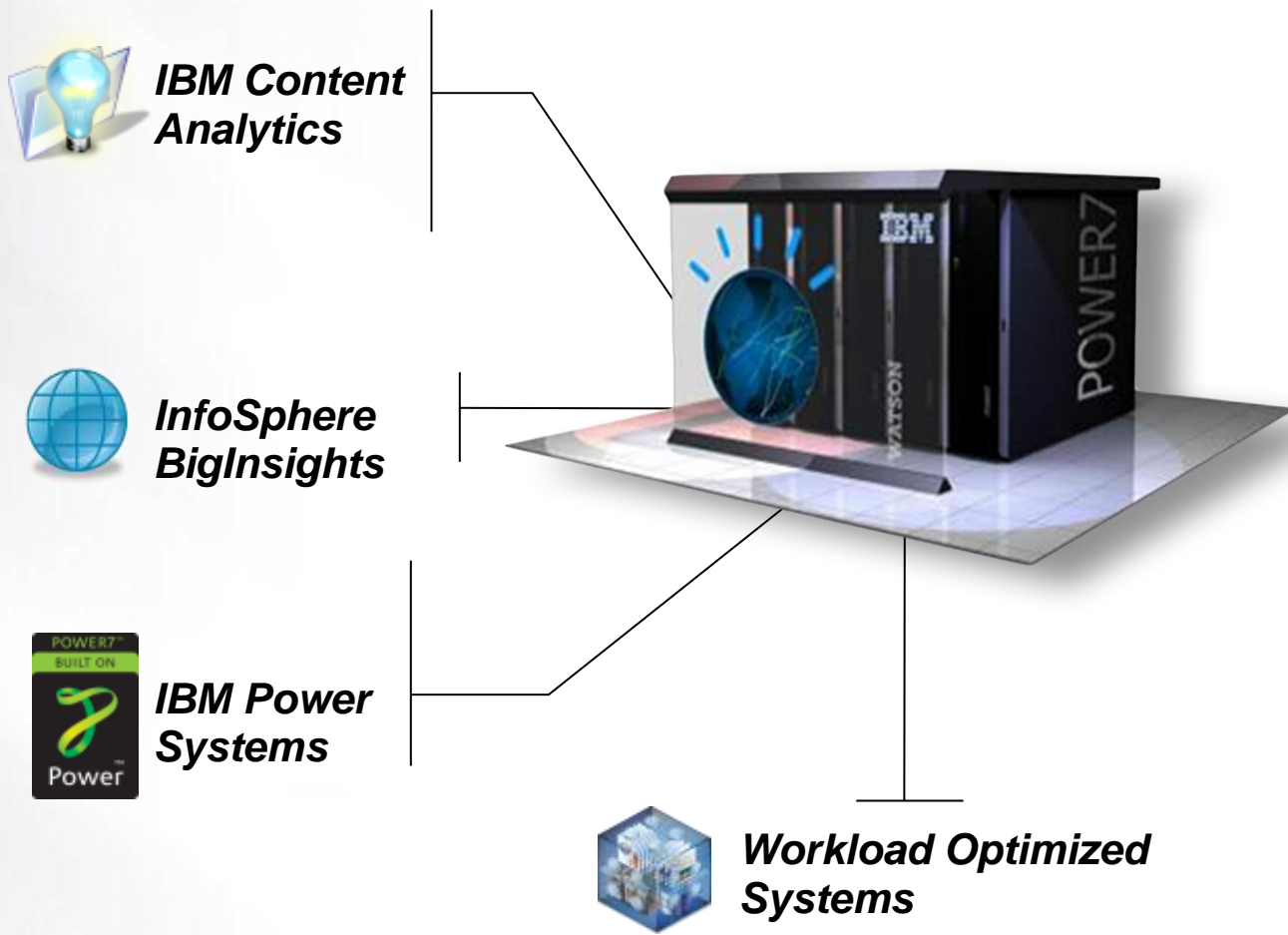


Watson Reflections



How Can IBM Help Today?

Used by Watson



Related Innovations

- InfoSphere Warehouse**
DB2, Informix, Netezza
- InfoSphere Streams**
- Business Analytics**
- ECM Solutions**
- IBM Global Business Services**

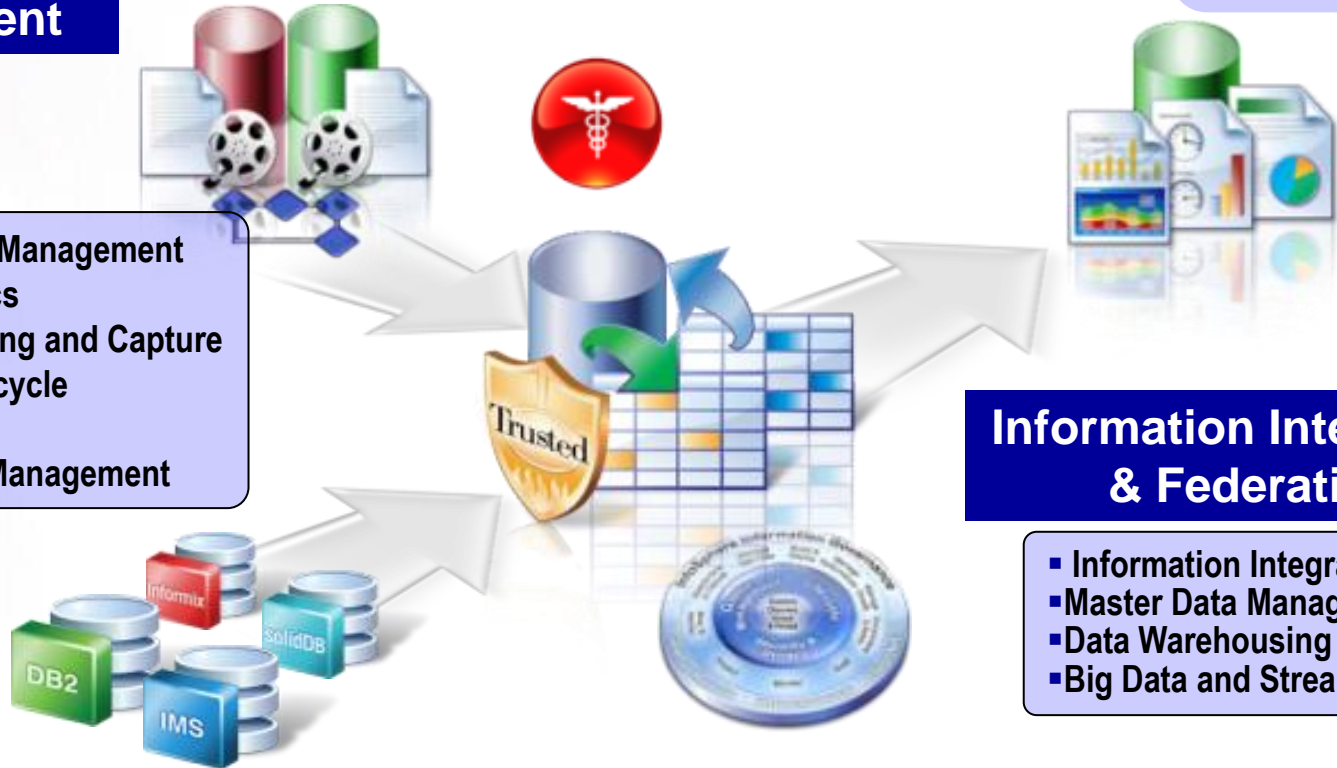
Turning information into insights

Enterprise Content Management

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

Business Analytics

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



Data Management

- Database Software
- Database Management Tools

Information Governance

- Data Lifecycle Management
- Data Security and Privacy

Information Integration & Federation

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

IBM Smart Analytics System Family

1050/5710



- **System x based technology**
- *Cost-effective solutions for analytics and reporting*
- *Compact, appliance-like single analytics solutions*
- *Available for mid-market*

9600/9610



- **System z based offering**
- *Advanced query /workload management*
- *Database designed and optimized for System z*
- *Smart Disk controller, optimized to reduce data latency*

2050/5710



- **System x based technology**
- *Cost-effective solutions for analytics and BI*
- *Balanced of power and simplicity out of the box*
- *Available for mid-market*

7700/7710



- **Based on Power technology**
- *Scaling to petabytes of data*
- *New 7700 based on POWER7 Processor and includes solid state technologies!*

5600



- **System x based technology**
- *Designed for business analytics workloads*
- *Optional Solid State Device— reducing data latency*

Netezza



- **System x based technology**
- *Designed for business analytics workloads*
- *Optional Solid State Device—*

Potential Business Applications



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

Tech Support: Help-desk, Contact Centers



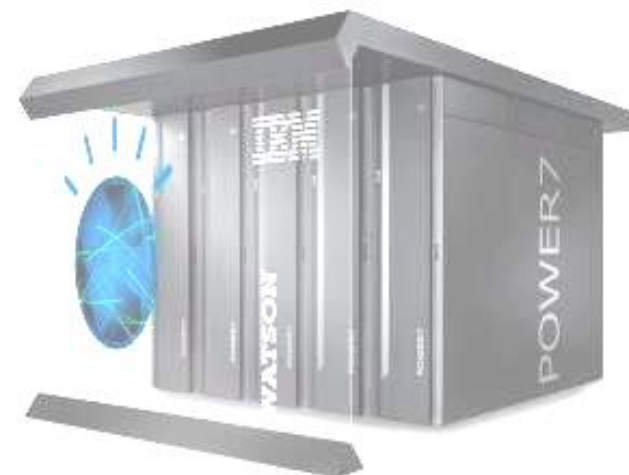
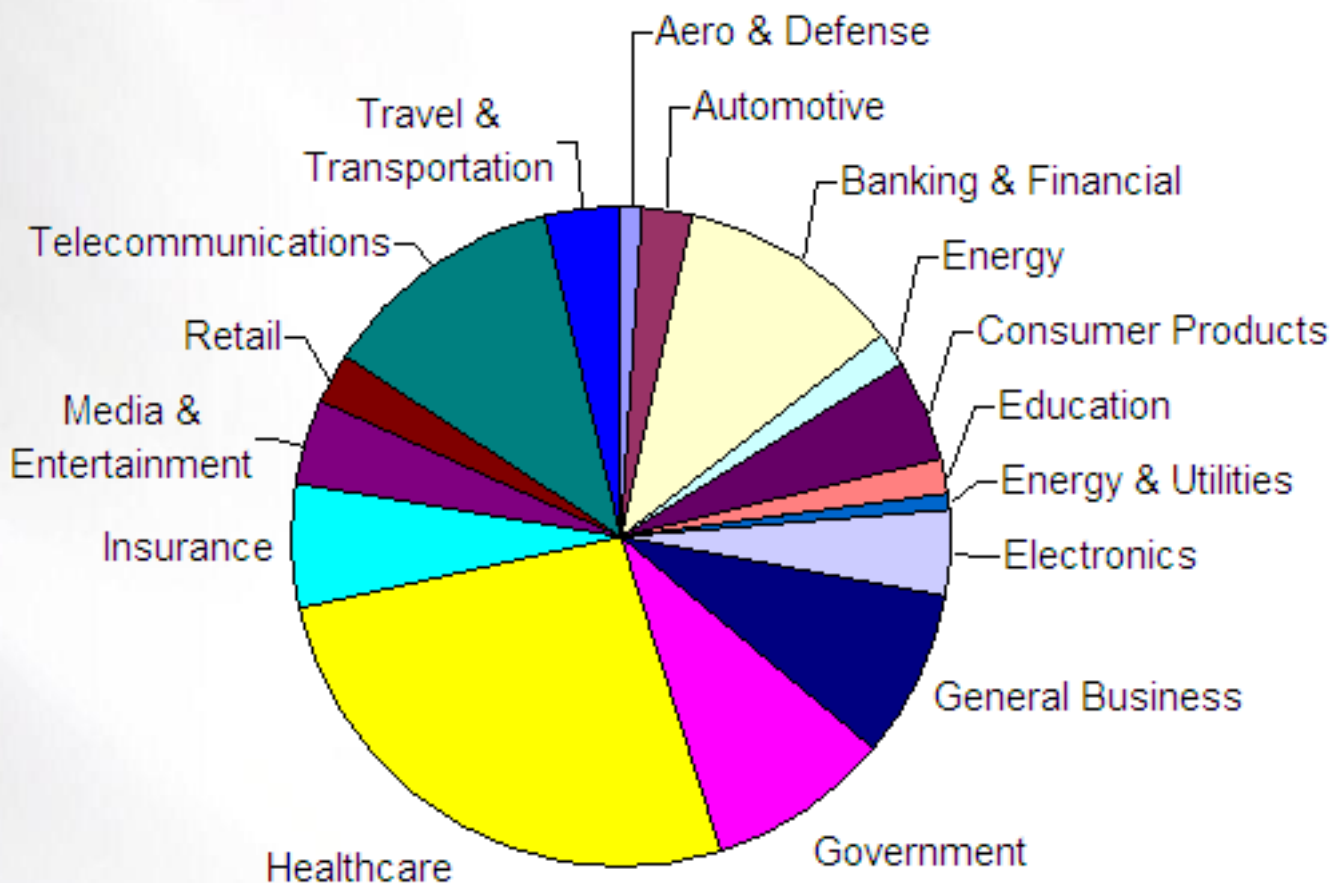
Enterprise Knowledge Management and Business Intelligence

Government: Improved Information Sharing and Security



Active Watson / DeepQA Engagements

Client inquiries arriving with a broad range of Watson use cases, from a broad range of industries



Watson Use Case – Healthcare

Challenge

Hospital providers face mounting challenges in how to **improve the quality of patient care** in an environment where patient and clinical data is proliferating and resources are limited

Problem

- **Medical information is doubling** every five years
- **1.5M errors occur** in the way medications are prescribed, delivered and taken in the U.S. each year
- **20% of all errors are diagnostic errors**
- 8.4% 2010-2011 rate of employer-based healthcare cost increase

Solution

- Watson can **analyze millions of pages per minute** to assist with diagnosis, leveraging the latest available research
- Watson can uncover patterns and **identify which patients are at risk** of contracting specific diseases
- Watson can help **identify optimal treatment** recommendations reducing waste and re-admissions

“Management of disease is phenomenally complicated. Effective storage, retrieval, analysis and use of information is critical to improve healthcare.”

Dr. Herb Chase
Professor of Clinical Medicine, Columbia University

Watson @ WellPoint



WellPoint & IBM agreement,

WellPoint will develop and launch Watson-based solution

- Help improve patient care
- Provide delivery of up-to-date, evidence-based healthcare

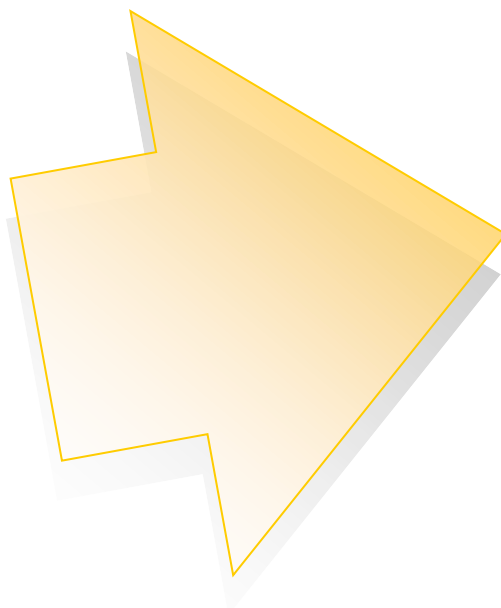
IBM will develop the base Watson healthcare technology

Targeted implementation: Early 2012,

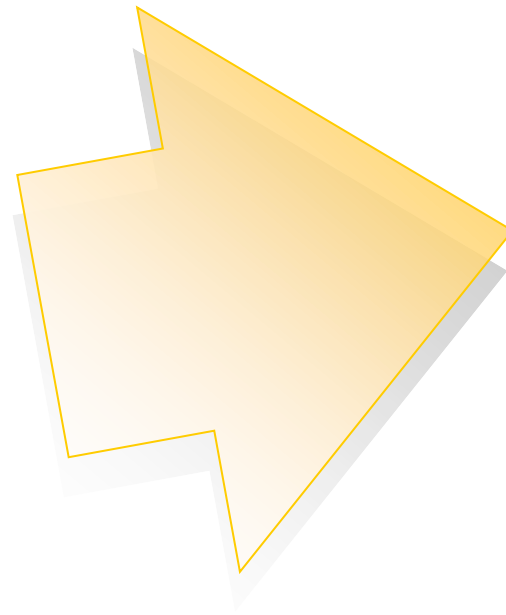
- Physician Helper: Will be a consultant with the ability to analyze various inputs to help humans make a decision
- Load Watson with medical data to analyze and identify conditions
 - ❖ Ranging from heart disease to cancer to diabetes.
 - ❖ Medical histories, test results, possible drug interactions and treatment options will be evaluated too.
- Working with select physician groups in clinical spots



Past History....

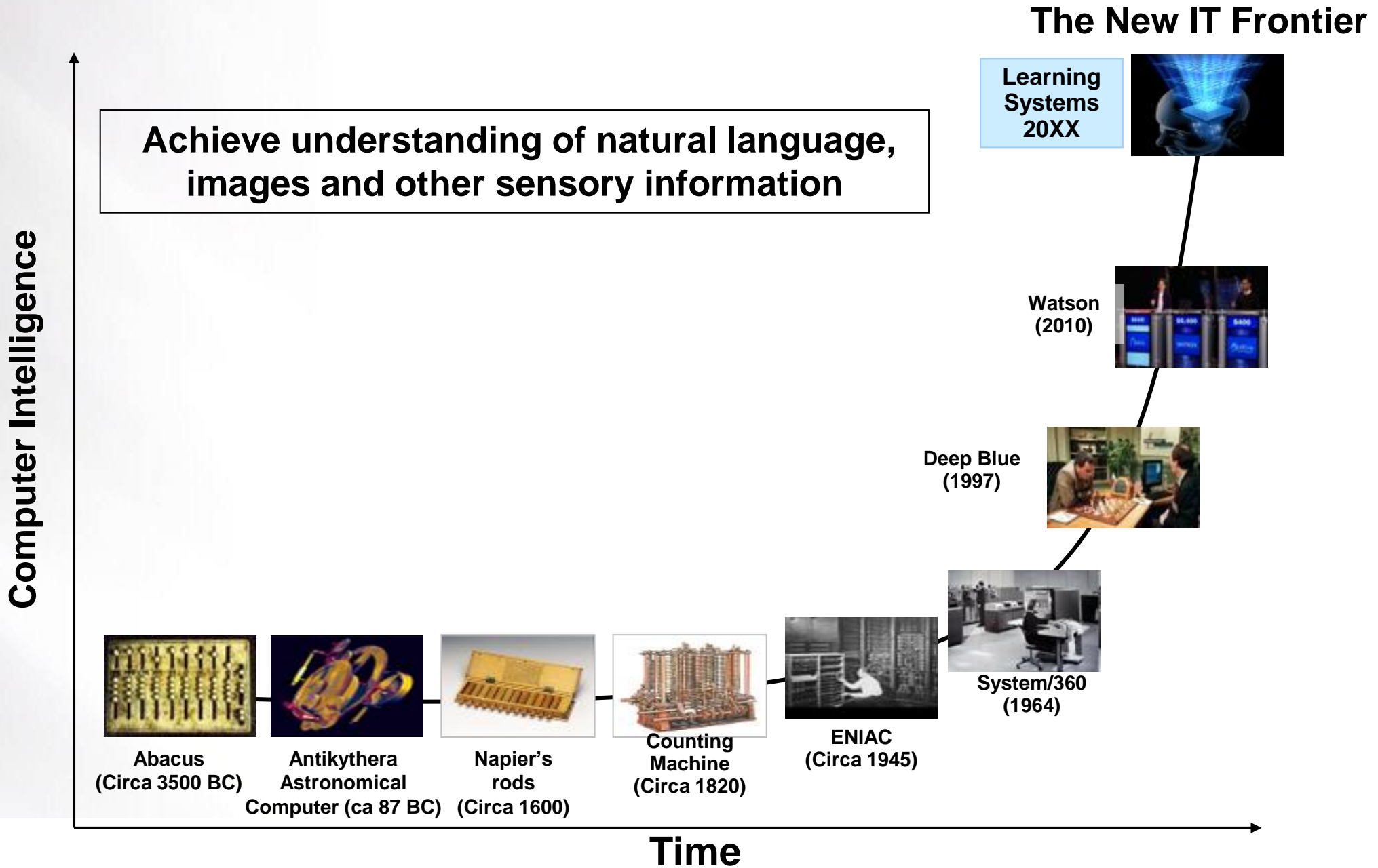


Technology that could compete against man



Technology that can think/reason like man

The Evolution of "Thinking" Machines



What do we know for sure?????

Watson

It is just the beginning...

Technology for Today

Innovation for the Future

