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

© 2011 IBM Corporation
From Science Fiction to Reality
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…

2011 Webby Person of the Year!
EMC 2011 Data Hero Awards

- 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.
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
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

<table>
<thead>
<tr>
<th>Game</th>
<th>Chess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture</td>
<td>Specialized Hardware and Power2 SC</td>
</tr>
<tr>
<td># Cores</td>
<td>30 + 480 ASICs</td>
</tr>
<tr>
<td>Hardware</td>
<td>Highly Specialized</td>
</tr>
<tr>
<td>Processing</td>
<td>Mathematical / Structured</td>
</tr>
</tbody>
</table>

### Data Analysis
- Finite number of possible moves and countermoves
- Structured data
- Mathematical probability
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 drawers & 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

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
</table>
| 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.
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.....

<table>
<thead>
<tr>
<th>Structured Data</th>
<th>Unstructured Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Easy Questions…..

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

Select Payment where Owner=“David Jones” and Type(Product)=“Laptop”,
Hard Questions....

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

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

<table>
<thead>
<tr>
<th>Person</th>
<th>Birth Place</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Einstein</td>
<td>ULM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Person</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. Welch</td>
<td>GE</td>
</tr>
</tbody>
</table>

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.*
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. This evidence suggests “Gary” is the answer, but the system must learn that keyword matching may be weak relative to other types of 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.

Stronger evidence can be much harder to find and score.

The evidence is still not 100% certain.
Each dot represents an actual historical human Jeopardy! game.

Top human players are remarkably good.

Computers?

Winning Human Performance

Grand Champion Human Performance

2007 QA Computer System

More Confident

Less Confident

% Answered

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

IBM Watson
Playing in the Winners Cloud

![Graph showing precision vs. % answered for different versions of DeepQA from 2007 to 2010.](image)
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 – 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
Real-Time Game Configuration

Clue Grid

Human Player 1

Human Player 2

Decisions to Buzz and Bet

Strategy

Text to Speech

Clues & Category

Answers & Confidences

Insulated and Self-Contained

Watson’s QA Engine

Power 750 Compute Cores

Data

Jeopardy! Game Control System

Clues, Scores & Other Game Data
Trivia: Who is....

Who is Ed Toutant????

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

**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

- **Workload Optimized Systems**
  - Integrated, Optimized by Workload
Applying Watson’s capabilities for business

Helping clients with deployments which tend to follow three phases:

1. **Manage Data**
   - Instrument physical assets and processes to capture and get your arms around oceans of data

2. **Analyze Patterns**
   - Cull deep insight out of all this data, using business analytics to reveal trends, patterns and new business opportunity

3. **Optimize Outcomes**
   - Transform your business by bringing it all together into a smarter system

**Unique value realized**

**Analyze Patterns**

**Optimize Outcomes**
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

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.

**Evidence Profile for UTI Diagnosis**

- Symptoms
- Family History
- Personal History
- Medications
- Findings

**Overall Confidence**

0 0.5 1
Continuous Evidence-Based Diagnostic Analysis

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

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

Diagnosis Models

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renal Failure</td>
<td></td>
</tr>
<tr>
<td>UTI</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
</tr>
<tr>
<td>Influenza</td>
<td></td>
</tr>
<tr>
<td>Hypokalemia</td>
<td></td>
</tr>
<tr>
<td>Esophagitis</td>
<td></td>
</tr>
</tbody>
</table>

Most Confident Diagnosis: Influenza
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
The Evolution of Science

Era of Natural Philosophy

- Platonic Academy (387 BC)
- Astronomy (Babylon, 1900 BC)
- Mathematics (India, 499 BC)

Scientific Revolution (1543 AD)

Industrial Revolution

- Newton’s Laws (1687 AD)
- Computing (1946 AD)
- Quantum Physics (1925 AD)

Era of Modern Science

- Relativity (1905 AD)
- DNA (1953 AD)

Learning Systems (XXI Century)

Evolution of Science

Time
The Evolution of Technology

Achieve understanding of natural language, images and other sensory information
Device and Technology Roadmap

Non-Von Neumann Architectures

- Computing Without Programming
  - Algorithms
  - Subsystem Integration (3D, SCM, Photonics)
  - Neuromorphic Devices & Circuits

Von Neumann Architectures

- InAs
- Spintronic & Magnetic
- Existing Architectures
  - Computing with Programs: Fetch Instructions, Decode, Execute & Repeat.
  - Low V Devices
  - Reconfigurable Logic
  - FPGA’s

New Devices, Architectures & Computing Paradigms

- Low V Devices
- Neuromorphic & Quantum Devices
- Quantum Devices & Circuits
- New Architectures Leveraging 100’s of Billions of Low-Voltage Devices

Increasing Efficiency

- Bio-Inspired Computation
- New Architectures
- Computing Without Programming
- Computing with Programs
- Existing Architectures
- Non-Von Neumann Architectures
- Von Neumann Architectures
- Time
- Today
- 22 nm, 15 nm, 11 nm, 8 nm, 5 nm, <5 nm?
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

Achieve understanding of natural language, images and other sensory information

- Abacus (Circa 3500 BC)
- Antikythera Astronomical Computer (ca 87 BC)
- Napier’s rods (Circa 1600)
- Counting Machine (Circa 1820)
- ENIAC (Circa 1945)
- System/360 (1964)
- Deep Blue (1997)
- Watson (2011)
- Learning Systems 20XX

New IT Frontier
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.

Keywords Search
Delivers lists based on keywords & human filters

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

Dynamic Learning Systems
- 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

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

1985

Today

Future

Greater Autonomy
The Future ....

Watson

It is just the beginning...