Watson Developers Cloud: Build a new generation of apps

Swami Chandrasekaran
Executive Architect - CTO Office
IBM Watson Innovations
swamchan@us.ibm.com
@swamichandra

Red Hat Summit
San Francisco
April 14-17 2014
Please Note

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.
Topics

• What is Watson?

• How Watson works?

• Watson Ecosystem & Watson Developer Cloud

• Live Demonstration
  – Watson Tools
  – Watson QA API

• Recap & Q&A
On February 14, 2011 IBM Watson made history.
Big Data: More than just volume

Volume:
Terabytes to exabytes of existing data to process

Velocity:
Streaming data, milliseconds to seconds to respond

Variety:
Structured, unstructured, text & multimedia

Veracity:
Uncertainty from inconsistency, ambiguities, etc.
Consider this Natural Language Question

Microsoft? When you have frequent BSOD; Quite possible 😊

Would MS inhibit my nerve function?

Reference to Nerve function. Could it be Multiple Sclerosis?

Master of Science?

Keyword search and expert systems are not able to recognize the subtleties, idiosyncrasies, and ambiguities inherent in common human language.
This is how Watson would respond with Multiple Sclerosis content ingested as Corpus.

Watson: How can I help you today?

Common Questions

Swami: Would MS inhibit my nerve function?

Watson: 34. www.mstrust.org.uk In MS, only nerves in the brain and/or spinal cord are damaged. However, because these nerves control the functions of the whole body, MS can cause a wide variety of symptoms. The type, severity and duration of symptoms will vary from person to person. How MS affects an individual depends on where damage occurs and which nerve messages are interrupted or blocked.

Reference | Here is what I'm thinking...

High

Ask a question in the same way as you'd ask a person.
Deep natural language processing is NOT keyword search

*Watson is not a rules-based search or speech processing*

- Traditional systems use keyword and synonym matching to find documents
- Expert systems need rules to demonstrate expertise
- Keyword search and expert systems trade off precision for accuracy outside of what they have been programmed to recognize
- Cognitive systems, like humans, know how to apply context, fill in the gaps, resolve ambiguity and uncertainty, and accurately interpret language

Keyword search and expert systems are not able to recognize the subtleties, idiosyncrasies, and ambiguities inherent in common human language
Watson is cognitive computing

Watson understands me.
Watson engages me.
Watson learns and improves over time.
Watson helps me discover.
Watson establishes trust.
Watson has endless capacity for insight.
Watson operates in a timely fashion.

Understands natural language
Generates and evaluates hypotheses
Adapts and learns
Topics

• What is Watson?

• How Watson works?

• Watson Ecosystem & Watson Developer Cloud

• Live Demonstration
  – Watson Tools
  – Watson QA API

• Recap & Q&A

© 2014 IBM Corporation
How Watson works: DeepQA Architecture

Inquiry

Multiple Interpretations of a question

Answer Sources

Inquiry/Topic Analysis

100’s sources

Inquiry Decomposition

100’s Possible Answers

Primary Search

Candidate Answer Generation

Hypothesis Generation

Hypothesis and Evidence Scoring

Synthesis

Final Confidence Merging & Ranking

Responses with Confidence

Learned Models help combine and weigh the Evidence

Hypothesis

Hypothesis and Evidence Scoring

1000’s of Pieces of Evidence

100,000’s Scores from many Deep Analysis Algorithms

Balance & Combine

Evidence Sources

Deep Evidence Scoring

100’s Possible Answers
Building a Watson based app – How does it all come together?

Sources
- Partner Content
- Unstructured Data From Enterprise Apps
  - CRM
eComm
  - Workforce
HR
  - ECM
KMS
  - Product Catalog
Trouble Ticket
- Public Domain & Social
- Proprietary Content
  - .pdf
  - HTML
  - XML
  - Word
  - TXT
  - Excel
  - Email, Collaboration, Open Data

Preparation
- Discover
- Contract
- Acquire
- Cleanse
- Aggregate
- Curate
- Verify

Ingestion
- Ingest

Leverage
- Build Apps w/ QA API's
  - Application Partners

End Users
- App Users

Training & Domain Adaptation
- Q&A Pairs
- Lexicons
- Answer Types
- Interaction History

End Users / Crowd
- Domain Experts
Knowledge Base for Watson – Types of Unstructured Data that can be Ingested in Watson

• HTML
  – Websites
  – Domain Independent Content
  – FAQ pages
  – Forums
  – Product Descriptions
  – Product Catalog Export
  – Product Reviews
  – Social Network Extracts

• Microsoft Word
  – Operational Runbook’s
  – Policies & Procedures
  – Contracts & Agreements
  – Terms & Conditions

• PDF
  – Product Manuals
  – Financial Reports (sans graphs)
  – Analyst Research
  – User Guides
  – Troubleshooting Guide
  – Terms & Conditions
  – Setup & Configuration Guides
  – Magazines & Journals

• Text
  – Extracted Interaction Logs from CRM, Trouble Ticket, Care, Media Asset Management applications
Content Ingestion

• Before IBM Watson can answer user questions, it must have a knowledge base / corpus that contains information from which to formulate correct answers

• Content Ingested into the IBM Watson means that the content is processed by the system in such a way that the correct segments of that data can be quickly retrieved when needed

• Content can come from a variety of sources and will have to be physically made available

• Content requires a level of cleansing and curation for Watson to be effective
Based on this content one might be interested in understanding the following.

- Various domain (telecom in this case) terms or concepts such as equipment types (gateway, router, receiver), problems (connection issues, no audio etc.), associated conditions (correct cabling, receiver configuration, etc.), abbreviations etc.
- Concepts or terms expressed in different ways (receiver Vs STB Vs DVR, router Vs gateway etc.)
- Timing in terms of when a particular problem is more prevalent post service install & activation
- Statistics in terms of how many people were affected over the years.
- Locations and other geo references made in the document with respect to the concepts or terms.
- Co-references or relationships or associated service problems with other conditions or issues (DVR Recording & Memory Capacity, Data Features & Connectivity etc.), upgrades, value added options etc.
Topics

- What is Watson?
- How Watson works?
- **Watson Ecosystem & Watson Developer Cloud**
- Live Demonstration
  - Watson Tools
  - Watson QA-API
- Recap & Q&A
Watson has opened up a world of new possibilities

2011
- 2880 cores
- Single user system
- 2-3 sentences input
- 5+ days to retrain
- Wikipedia, general corpus

2013
- Single Power 750, 240% faster
- 1000s of users
- 20 pages of input
- < Day to ingest and train
- Medical corpus

2014
- Watson cloud service
- Millions of users
- Dialog chaining
- Few hours
- Broad industry corpus
IBM Watson Ecosystem is composed of mutually benefiting, interconnected partners

**App Partners**

**Develop** cognitive powered applications powered by IBM Watson

**Content Partners**

**Fuel** the ecosystem with content to drive cognitive applications

**Talent Partners**

**Support** application partners with developing and designing their solutions
IBM Watson Ecosystem is composed of mutually benefiting, interconnected partners

**Watson Talent Hub** bridges App Partners resource gaps by providing a marketplace for critical cognitive, big data, UI, and mobility skills.

**Watson Developer Cloud** delivers the tools, methodologies, software developer kits and API(s) for App Partners to build the next generation of cognitive applications.

**Watson Content Store** provides sources of free and fee based content including public, subscribed and enterprise content.
What does a partner get through the IBM Watson Ecosystem

- **Use the Watson Developer Cloud**
  - Use the Watson Experience Manager and other tools
  - API to access the QA service to Watson
  - Documentation
  - Ability to enrich Watson using content and train and test an application
  - Customizable user interface to embed in an application
  - Guidelines on using Watson elements in the application

- **Obtain or provide content in the Watson Content Store for use in applications**

- **Access to a network of partners, customers, content and talent through the IBM Watson Ecosystem**

- **Use of a collaborative community including IBM’s technical, marketing and sales**
## Typically Phased Infrastructure

<table>
<thead>
<tr>
<th>Delivery Environment</th>
<th>Delivery Phase</th>
<th>Ready</th>
<th>Configure and Train</th>
<th>Deploy and Manage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deploy</strong></td>
<td></td>
<td>16 core / 250 GB</td>
<td>16 core / 250 GB</td>
<td>16 core / 250 GB</td>
</tr>
<tr>
<td><strong>Pre-deploy</strong></td>
<td></td>
<td>16 core / 250 GB</td>
<td>16 core / 250 GB</td>
<td>16 core / 250 GB</td>
</tr>
</tbody>
</table>
Characteristics of a good Watson use case

A great use case leverages Watson’s cognitive capabilities below

- Exploits the 4 Vs of data - Volume, Velocity, Variety, & Veracity
- Leverages evidence based insights with weighted confidence
- Benefits from a continuously learning system
- Takes advantage of deep natural language processing
- Requires transparency to the source of information
- Transforms user experience with contextual relevance
- Transforms user experience with active dialoging
How to become a IBM Watson Ecosystem Application Partner?

• Use of the Watson Developer Cloud, including:
  ✓ Watson Experience Manager and other tooling
  ✓ API to access the QA service to Watson
  ✓ Documentation and tutorials
  ✓ Workspace to enrich Watson with content
  ✓ Workspace to train and test your application
  ✓ Customizable UI to embed into your application
  ✓ Guidelines on using Watson elements in the app
• Upload content into the Watson Content Store or source through Ecosystem partners
• Access to a network of partners, customers, content and talent
• Support from a collaborative community of IBM technical, marketing and sales resources

Take the Watson Developer Challenge

- Opportunity for Mobile Developers to create the next great cognitive mobile app – announced at Mobile World Congress by Ginni Rometty

- Finalists get access to Watson for one business week to prototype their Watson instance and mobile app

- The top 5 prototypes present live to IBM Executives

- Winners join the Watson Ecosystem and get access to IBM Interactive design Services

Submissions: 2/26/2014 – 4/15/2014
Finalist selection: 4/28
Winners Announced: 5/31
Topics

• What is Watson?
• How Watson works?
• Watson Ecosystem & Watson Developer Cloud
• Live Demonstration
  – Watson Tools
  – Watson QAAPI
• Recap & Q&A
LIVE DEMO
Key Tools & Capabilities – Watson Experience Manager

- Provides a role based set of tools for Developers, SME’s, Watson administrators, and Domain Experts
# Watson Experience Manager – Corpus Management

- Corpus Upload & Management
- View Ingested Corpus
- Allows for on-demand corpus pre-processing & ingestion

### Corpus Management Interface

<table>
<thead>
<tr>
<th>Name</th>
<th>Status</th>
<th>Size</th>
<th>Type</th>
<th>Uploaded On</th>
<th>Uploaded By</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.html</td>
<td>New</td>
<td>5 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
<tr>
<td>805.html</td>
<td>New</td>
<td>5 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
<tr>
<td>805877.html</td>
<td>New</td>
<td>5 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
<tr>
<td>805850.html</td>
<td>New</td>
<td>4 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
<tr>
<td>80584.html</td>
<td>New</td>
<td>4 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
<tr>
<td>805721.html</td>
<td>New</td>
<td>5 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
<tr>
<td>809711.html</td>
<td>New</td>
<td>4 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
<tr>
<td>80670.html</td>
<td>New</td>
<td>4 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
<tr>
<td>806707.html</td>
<td>New</td>
<td>4 KB</td>
<td>html</td>
<td>1/13/2014</td>
<td>tridentbasert1</td>
<td></td>
</tr>
</tbody>
</table>

- Number of Docs Ingested as Corpus
- View Ingested Document & CAS
- Tag Each Document
- Upload & Delete Documents
- Ingest into Corpus on-demand
Watson Experience Manager – Expert Training

- Allow SME's to enter & manage “Question & Answer” pairs
- Review & Approve Q&A Pairs
- Forms the basis for training data to help train Watson
- Goal is to build a representative set that enables Watson to understand the domain's language characteristics

Sample Q&A Pair Instance

Approve, Edit, Add Comments

Drive towards a goal
Watson Experience Manager – Testing Watson

• Test system by asking questions
Integration with Watson – Embedding Watson into an Existing Self-Service Application

Mortgage Calculator

House Price: $218,000
Down Payment: $45,000

Rate Type: Fixed
Frequency: Monthly
Interest Rate: 5.0%
Amortization: 30 Years

Ask Watson
Melissa: How about financing?
Watson: The most popular financing option is the 30-year fixed mortgage. Would you like to try a price calculator?
Yes, open the calculator...

© 2014 IBM Corporation
Integration with Watson – Question & Answer RESTful API’s

- The Watson Question and Answer API (QAAPI) is a REST styled service interface that allows applications to interact with Watson. Uses JSON Payload
- You can pose questions to Watson, retrieve responses, & submit feedback on those responses
- Provides transparency with its conclusions (confidence, evidence etc)
- Response, confidence evidence & other response metadata can be manipulated, enriched
- Can be used part of a conversation; or to a component in the system downstream for further processing

![Diagram showing integration with Watson QAAPI](image-url)
Integration with Watson – Sample JSON request / response

REQUEST

```

"question": {
  "questionText": "How to treat diabetes? What can "formattedAnswer": false,
  "items": "5",
  "evidenceRequest": {
    "items": "1",
    "profile": "Yes"
  },
  "filters": [
    "filterType": "metadataFilter",
    "fieldName": "indexedKey.gender",
    "values": ["male"
    ]
  ]
}
```

REQUEST

```

"question": {
  "questionText": "How to treat fruity breath: how i treat fruity breath : Dr. Daniel H. Ziegler",
  "evidence": "No copyright specified for this supporting passage or document.",
  "termsOfUse": "No license specified for this supporting passage or document.",
  "metadataMap": {
    "filename": "watsondata/ingestion/201403191425/eng/di/output/creo/Intermediate
    /xml-splitTrecTrim/33717246.xml",
    "title": "how i treat fruity breath : how i treat fruity breath : Dr. Daniel H.
    Ziegler",
    "corpusName": "PB",
    "DOCNO": "4D8DB38126AB052ACB590ED23BFD6B",
    "deepqaid": "33717246"
  },
  "synonymList": ["
  ],
  "pipelineId": "1267528343",
  "formattedAnswer": false,
  "category": "",
  "items": 5,
  "status": "Complete",
  "id": "D2CE13EDC9449F262C699E309CE025",
  "questionText": "How to treat fruity breath?",
  "evidenceRequest": [],
  "answers": ["
  ],
  "errorNotifications": [],
  "offTopicConfidence": 0,
  "passThru": ""
}
```
Topics

• What is Watson?
• How Watson works?
• Watson Ecosystem & Watson Developer Cloud
• Live Demonstration
  – Watson Tools
  – Watson QAAPI
• Recap & Q&A
Watson’s intelligence and capabilities are expanding

Explores

Reasons

Visualizes
This will be Watson

- Understands natural language
- Generates and evaluates hypotheses
- Adapts and learns

- Reasons
- Explores
- Visualizes

- Sees
- Hears
- Experiences
Summary

• Watson is a type of **Cognitive System**
• What makes Watson different are **three** powerful technologies
  – natural language
  – hypothesis generation, and
  – evidence based learning
• Before IBM Watson can answer user questions content must be **ingested** into a **corpus** to formulate correct answers
• Watson produces **multiple** answers and each of them contains a probabilistic **confidence** and supporting **evidence**
• Adapting Watson to a new domain is done through **Training**
• Integrating with Watson can be via **RESTful QAAPI** (or) embedding Watson UI
• **IBM Watson Ecosystem** is composed of mutually benefiting, interconnected partners to collaborate and create the next generation of cognitive apps.
• **IBM Watson Developer Cloud** provides developers access to technology, tools and APIs, including training and ways to develop and test applications built with Watson.
Some Interesting Links

- http://www.ibmwatson.com

- Apply to the Watson Ecosystem program

- Developer Q&A Forum
  - https://www.ibmdw.net/watson/

- Growing social dialogue
  - Twitter: @IBMWatson
  - LinkedIn: Watson Advocates Group
  - Facebook: facebook.com/ibmwatson

- IBM Watson Mobile Developer Challenge
  - Submissions: 2/26/2014 – 4/15/2014
Don’t forget to complete the session survey (available via mobile app)

Earn the “contributor badge” if you complete 8 session surveys
LEARN.NETWORK.
EXPERIENCE OPEN SOURCE.

April 14-17, 2014
San Francisco, CA

www.redhat.com/summit

10 YEARS and counting