

# How Watson Works

Dave Mobley

Watson Solutions Architect,  
Watson Technical Sales

1/22/14

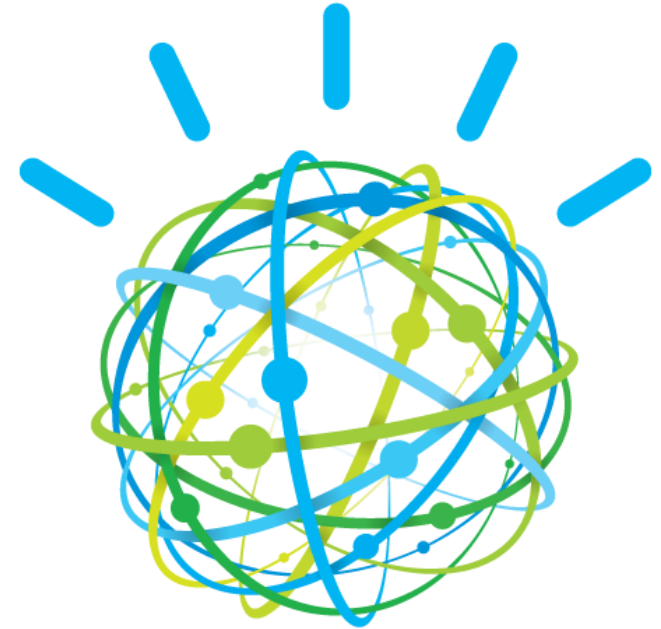


## What Watson isn't

- Search engine
- New-fangled database system
- Skynet or HAL 9000

## What Watson is

- Cognitive system
- Combines information retrieval and natural language processing (NLP)
- Builds its domain knowledge from sources comprising structured and unstructured data
- A core set of technologies that can be customized and targeted to specific industries
- Runs on Apache UIMA (Unstructured Information Management Architecture) technology



## Watson at Play

## Watson at Work

1 User	→	10s of thousands concurrent users
Max. input was two sentences	→	Pages of input (e.g. medical record)
5+ days to retrain	→	Dynamic content ingestion
Evidence not present	→	Supporting evidence integral
Text-only input	→	Text, tables and images as input
Q&A model	→	Both Q&A + Conversation model
Basic security	→	High security (e.g. HIPAA)

# Traditional approaches to engaging with customers **IBM WATSON™** come up short



**270B**

Calls made  
annually to call  
center costing  
\$600B



**1 in 2**

incoming calls  
require  
escalation or go  
unresolved



**61%**

of all calls could  
have been resolved  
with better access  
to information



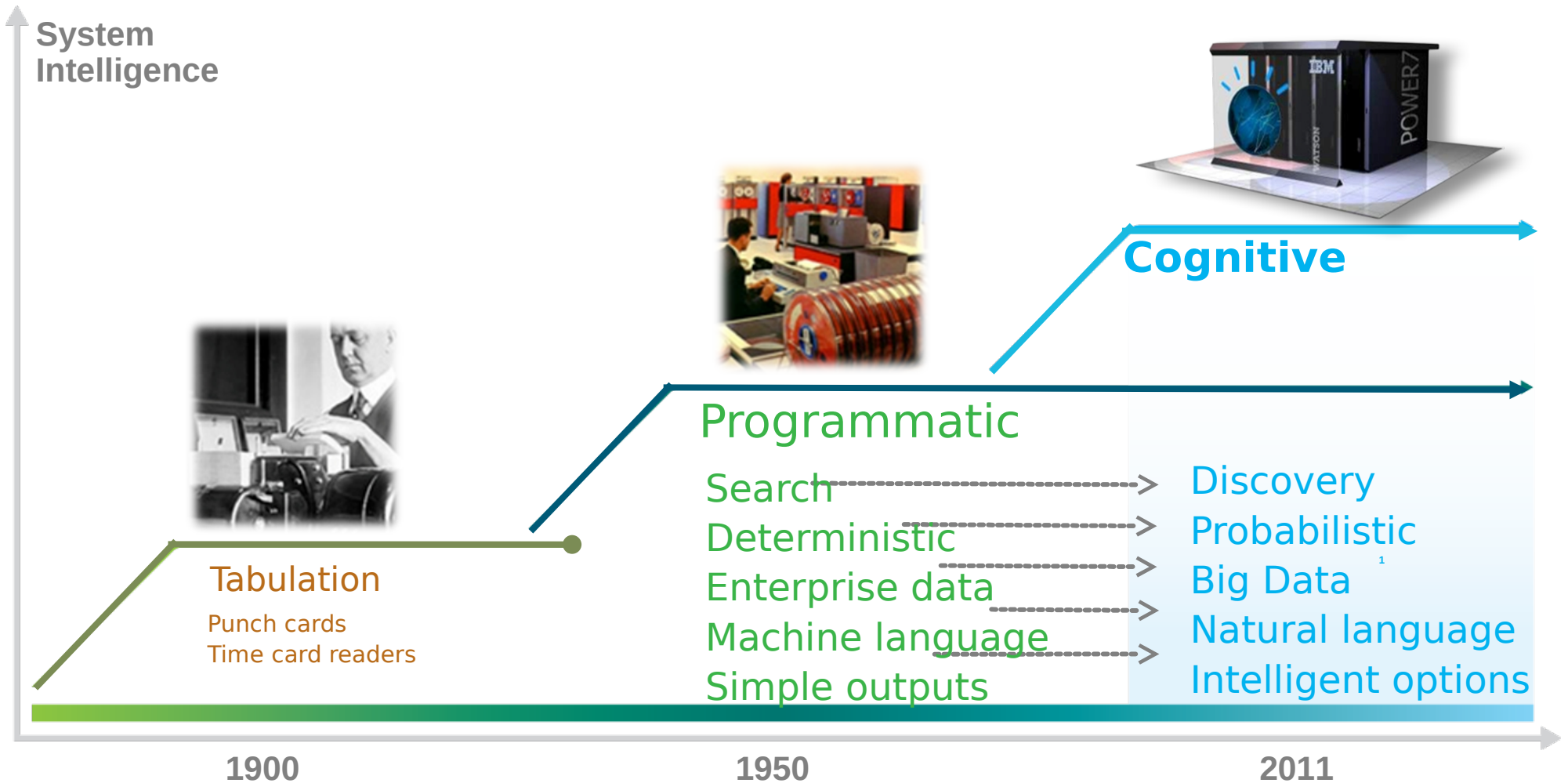
**4.6%**

Market value gain  
from a single point  
customer sat gain

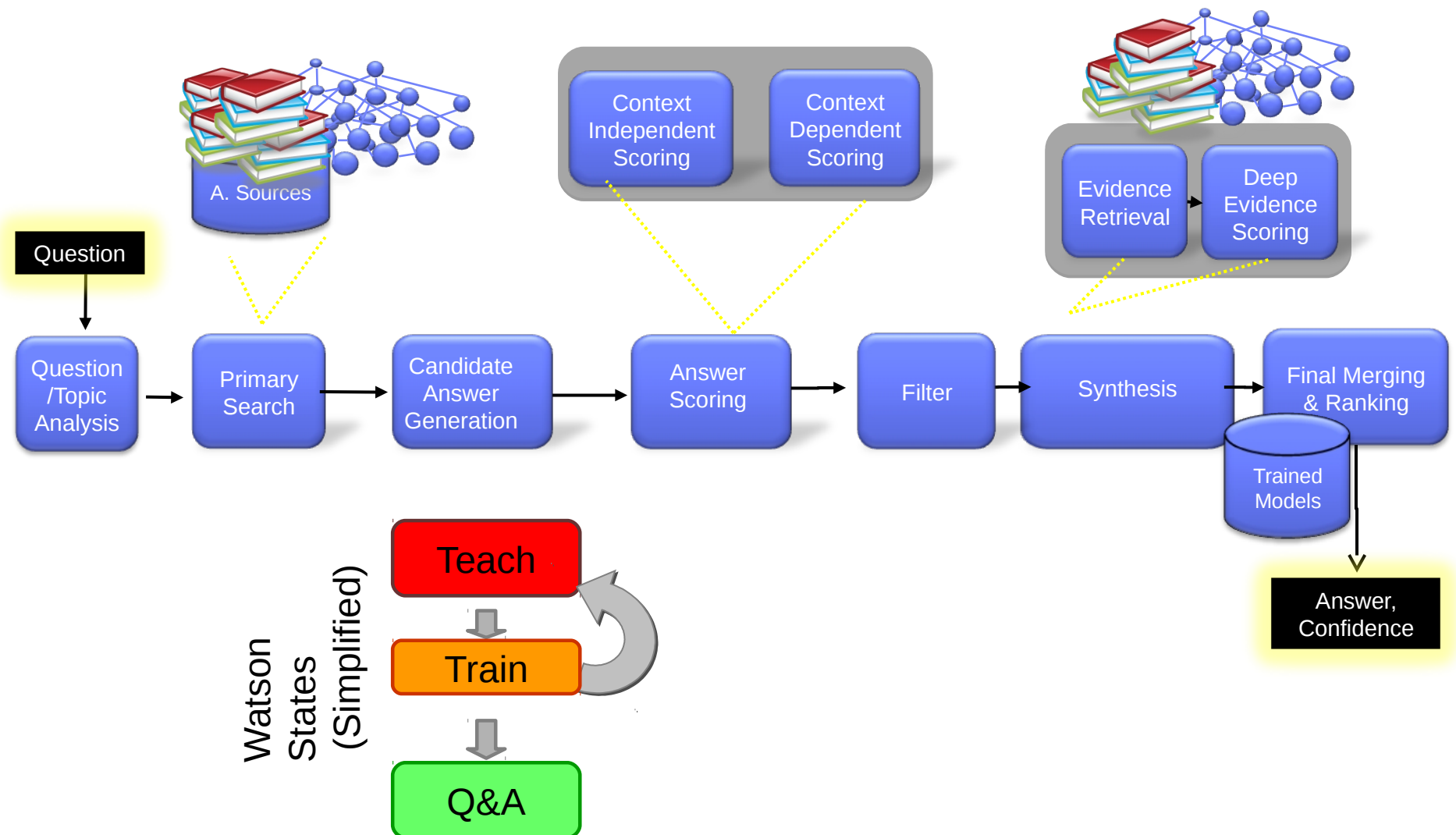


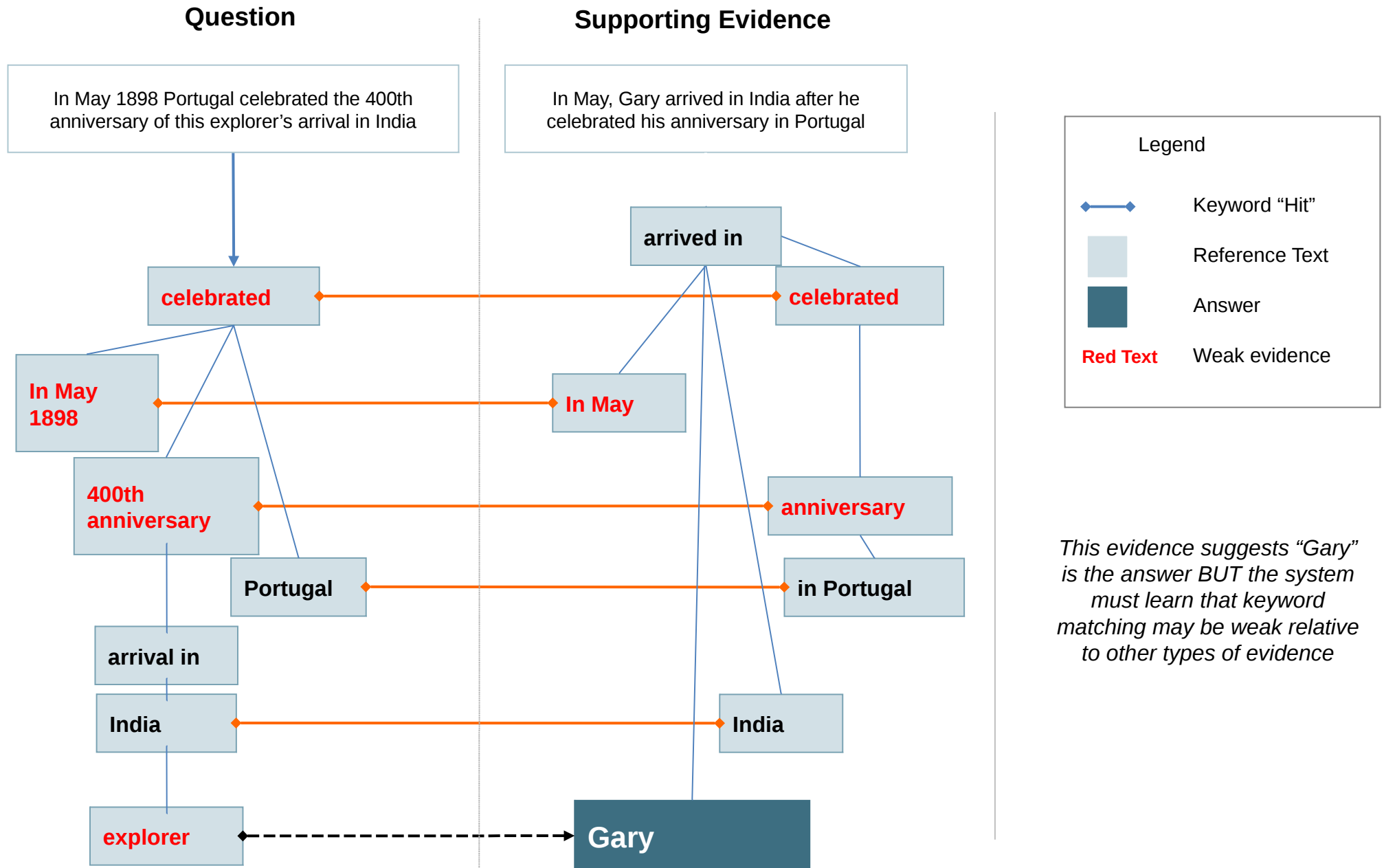


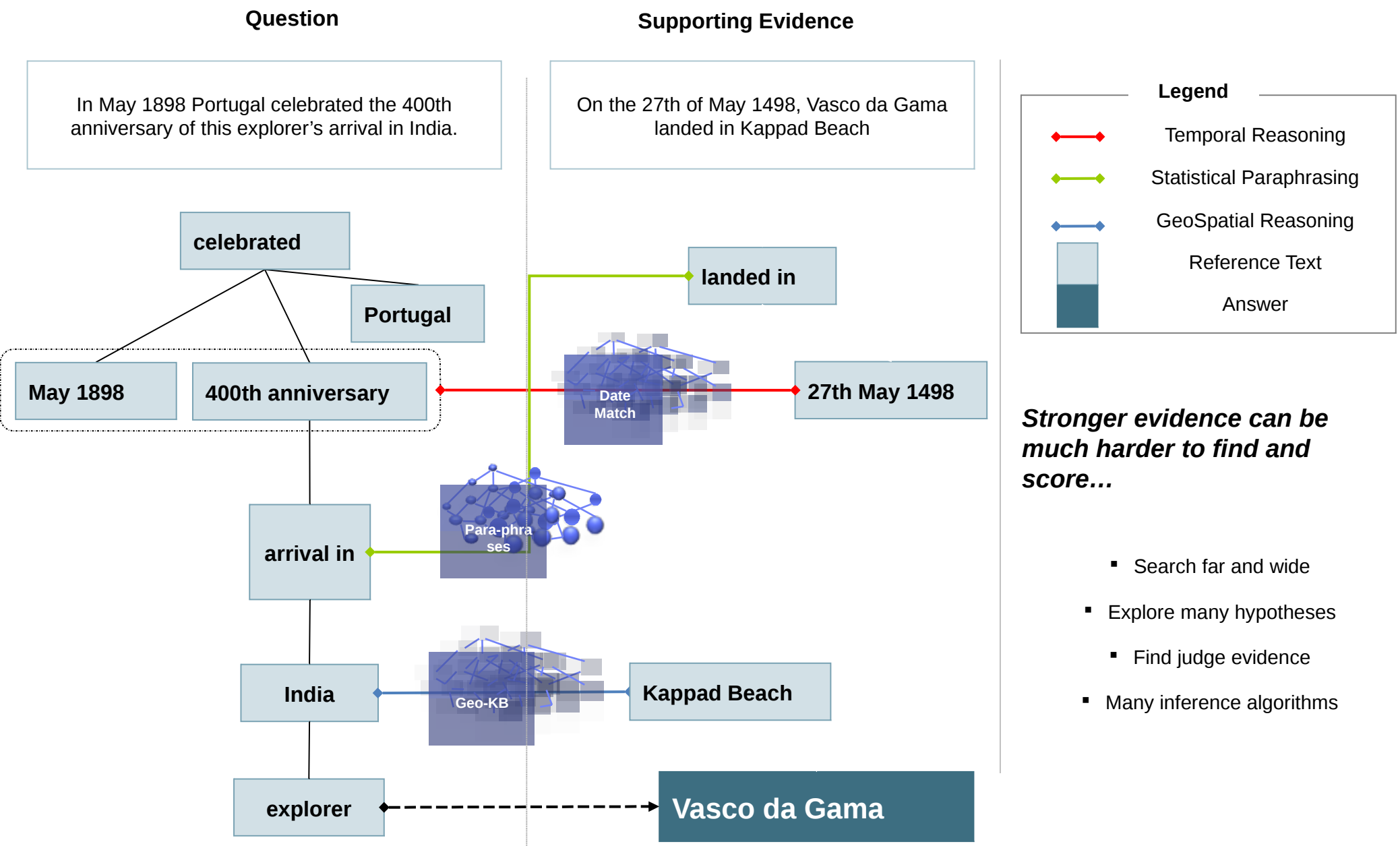
# IBM Watson represents a bold step into a new era of computing



...enabling new opportunities and outcomes







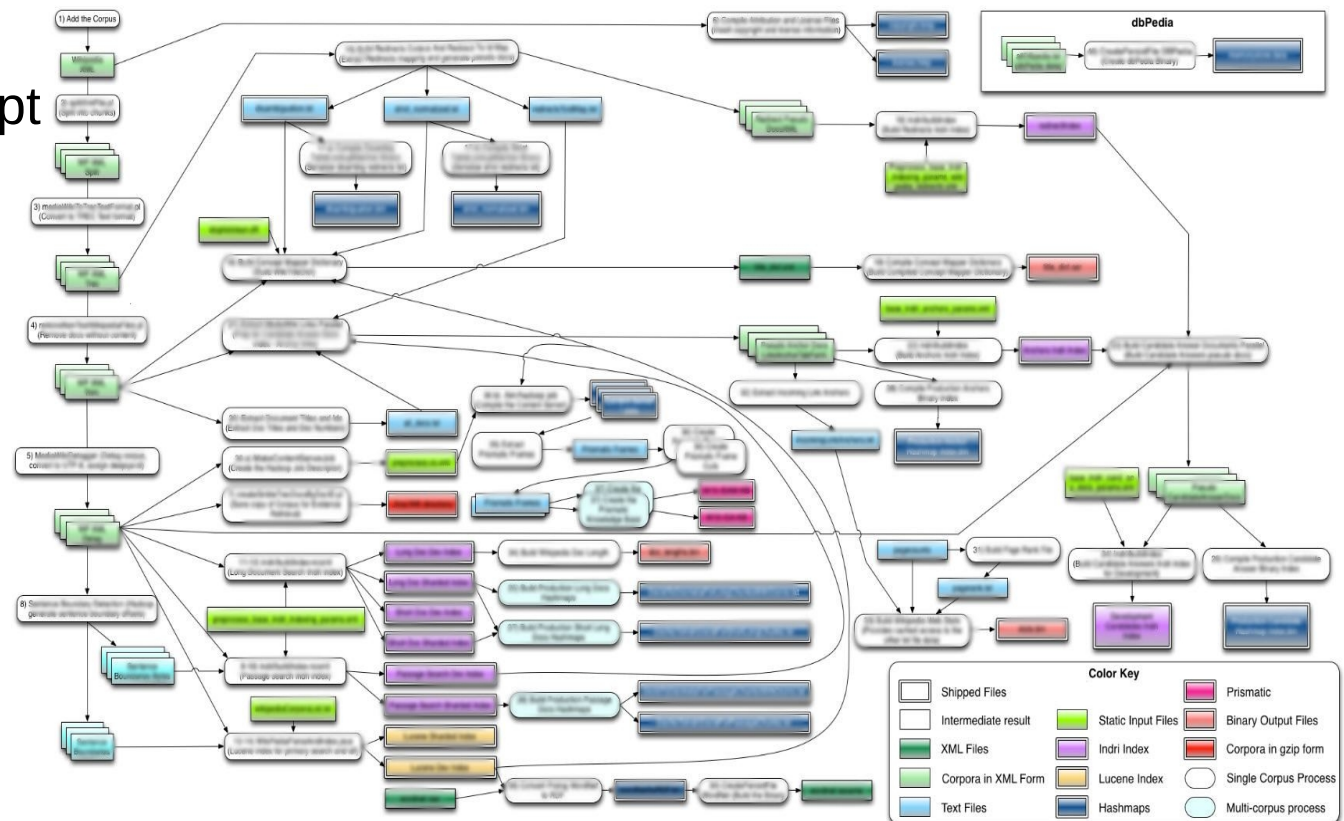
*Stronger evidence can be much harder to find and score...*

- Search far and wide
- Explore many hypotheses
- Find judge evidence
- Many inference algorithms

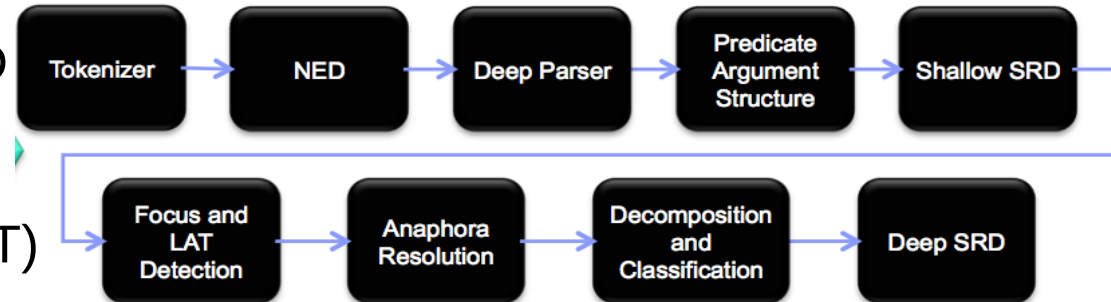
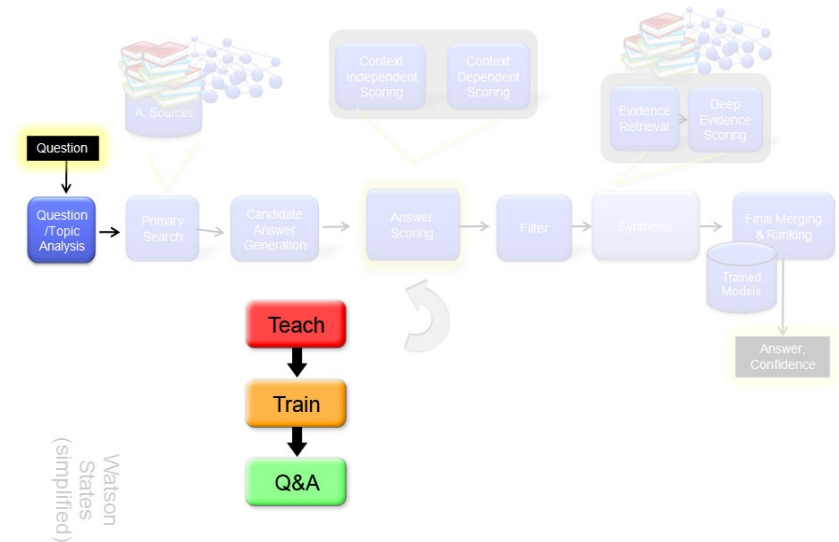
- Data must be preprocessed into TREC (Text Retrieval Conference) format
- Does allow for multiple corpora to be generated and used by a single pipeline
- Process for ingestion is its own pipeline which can be run via LiteScale

•Creates Indexes, and dictionaries such as Concept Annotator

- Future:
  - Frequent ingestion



- Rounds of teaching and training
- Core NLP
- Named entity recognizers/Detectors (NER/NED)
  - Type identification (places, people, dates, and so on)
  - Slot grammar parsers (XSG)
- Relationship detection
- Conference/Anaphora (pronoun) ID
- Keyword identification
  - Term/Lexical answer type (LAT) identification
- Machine learning to determine most likely LATs to consider further
- Multiple queries formed, based on full question, LAT, and terms, or inferences



Category/Topic: MICHIGAN

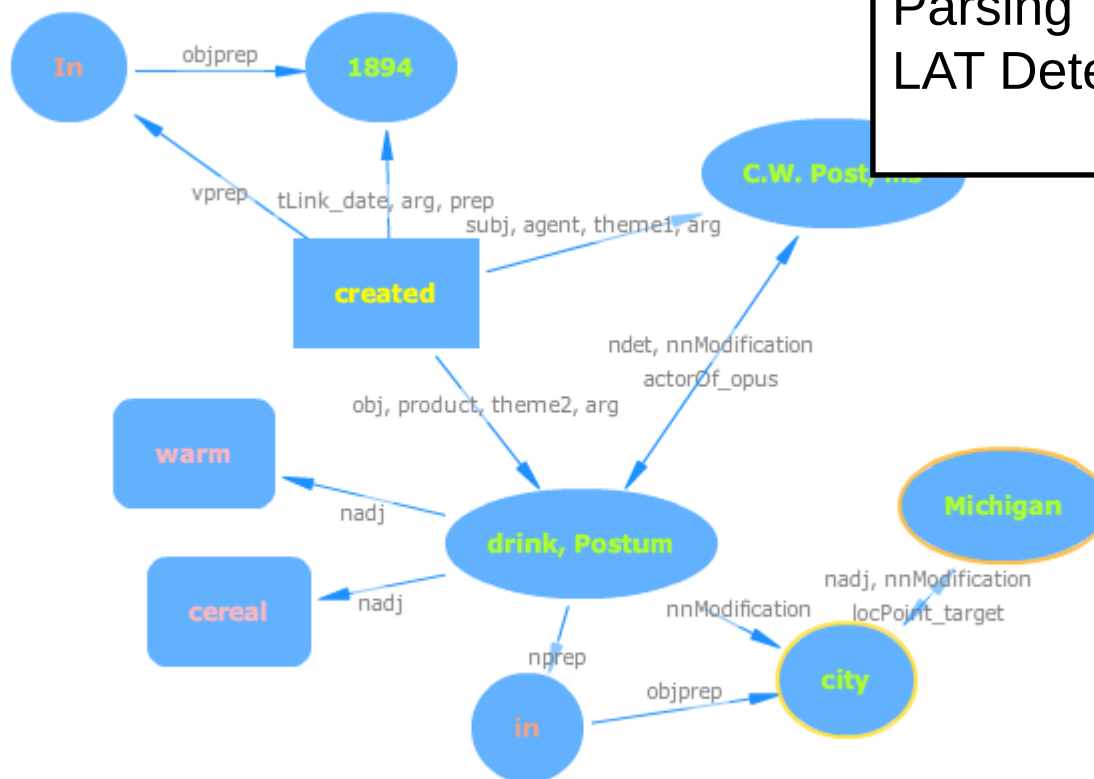
Question: In 1894 C.W. Post created his warm cereal drink Postum in this Michigan city

Focus: this Michigan city

LAT: Michigan city

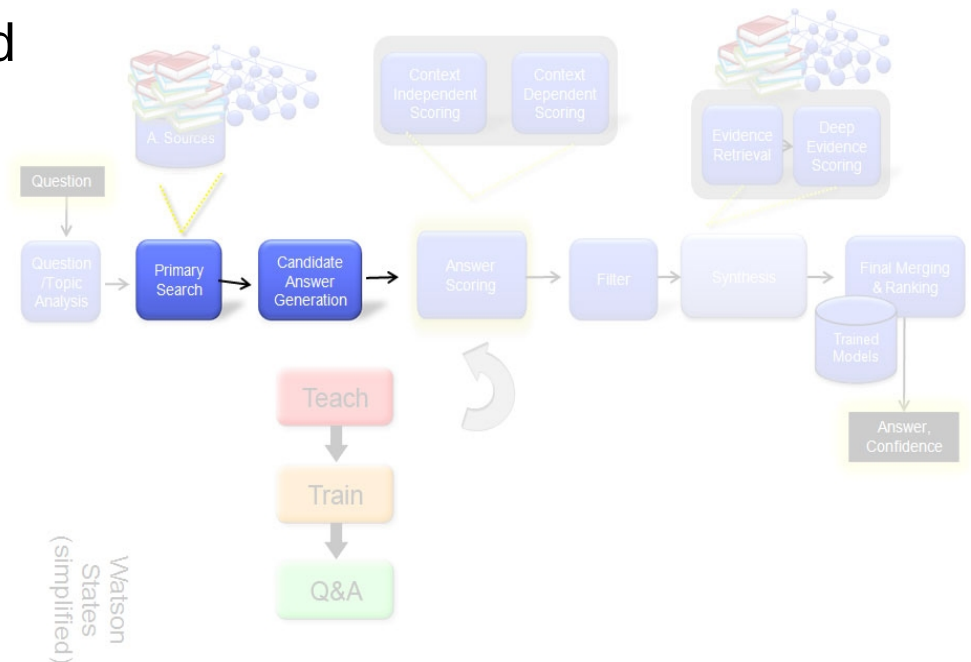
Keywords:

1894  
C.W. Post  
created  
warm  
cereal  
drink,  
Postum  
Michigan  
City





- Primary search (PS)
  - Take previously constructed queries and search among many available sources.
    - - Lucene
    - - Indri (multiple index types)
- Candidate answer generation
  - Parse PS results to build candidates of possible answers based on:
    - Titles
    - Anchor text
    - Passages and their parts: headwords, numbers, dates
    - Checking candidates against constraints
    -



# Step 2: Primary search

The keywords (**1894, C.W. Post, created, warm, cereal, drink, Postum, Michigan, city**) are used to search over millions of documents to find relevant hits. 55 documents are found, and 30 passages are found.

Indri Passage Search		Passage Search Results	
Lucene Passage Search		Rank	Passage
		0	C.W. Post came to the Battle Creek sanitarium to cure his upset stomach. He later created Postum, a cereal-based coffee substitute
		1	The caffeine-free beverage mix was created by The Postum Cereal Company founder C. W. Post in 1895 and produced and marketed by Postum Cereal Company as a healthful alternative to coffee
		2	1895: In Battle Creek, Michigan, C.W. Post made the first POSTUM , a cereal beverage. Post created GRAPE-NUTS cereal in 1897, and POST TOASTIES corn flakes in 1908
		3	1854 C. W. Post (Charles William) was born. He founded the Postum Cereal Co. in 1895 (renamed General Foods Corp. in 1922) to manufacture Postum cereal beverage
		4	The company was incorporated in 1922, having developed from the earlier Postum Cereal Co. Ltd., founded by C.W. Post (1854-1914) in 1895 in Battle Creek, Mich. After a number of experiments, Post marketed his first product-the cereal beverage called Postum-in 1895
		5	...

### General Foods

From Wikipedia, the free encyclopedia

**General Foods Corporation** was a company whose direct presence in the USA by **Charles William Post** as the **Postum Cereal Company**. General Foods was adopted in 1929, after several corporate acquisitions. General Foods was acquired by Philip Morris Companies (now Altria), the largest non-oil acquisition to that time. In December, 1999, Kraft, Inc., and in 1990 combined the two food companies as Kraft Foods. "General Foods" was dropped from the corporate name in 1995 as a brand name for a flavored coffee-based beverage, *General Foods*.

**Contents** [hide]

- 1 History
- 2 Major acquisitions
  - 2.1 Purchase of Birdseye
- 3 See also
- 4 Notes and references
- 5 Timeline of selected later events

**History**

[1] C.W. Post established his company in 1891, when he was a patient at an asylum in Battle Creek, Michigan. Dr. Kellogg, with his brother John, was part of their patients' diet, but a roasted, cereal-based coffee during his time in the sanitarium, which he copied.

### Will Keith Kellogg

**Britannica Concise Encyclopedia:** Will Keith Kellogg (born April 7, 1860, Battle Creek, Mich., U.S. — died Oct. 14, 1956, Battle Creek, Mich.) American food manufacturer and philanthropist. After working with his brother John Harvey Kellogg, he founded (1906) the W.K. Kellogg Co. to manufacture cereal. It soon became a leading U.S. producer of cereal. In early years, its annual sales exceeded \$9 billion. The W.K. Kellogg Co. is now a philanthropic institution.

For more information on **Will Keith Kellogg**, visit [Britannica Online](#).

### Post Foods

From Wikipedia, the free encyclopedia

**Post Foods, LLC**, also known as **Post Cereals** (formerly **Postum Cereals**) was founded by C.W. Post. It began in 1895 with the first Postum, a "cereal beverage", developed by Post in Battle Creek, Michigan. The first cereal, Grape-Nuts, was developed in 1897. Post has its headquarters in the Bank of America Plaza in Downtown St. Louis, Missouri.[2]

The Postum Cereals company, after acquiring Jell-O gelatin in 1925, Baker's chocolate in 1927, Maxwell House coffee in 1928, and other food brands, changed its name to General Foods Corporation in 1929. General Foods was acquired by Philip Morris Companies in 1995.

In 1989, Philip Morris merged General Foods with Kraft Foods, which it had acquired in 1987 to form the Kraft General Foods division. The cereal brands of Nabisco were acquired in 1993. In 1995, Kraft General Foods was reorganized and renamed Kraft Foods.

On November 15, 2007, Kraft announced that it would spin off Post Cereals and merge that business with Railcorp Holdings. [3] That merger was completed August 4, 2008. [4] The official name of the company became Post Foods, LLC.

**Contents** [hide]

- 1 Cereal brands – present cereals
- 2 Discontinued cereals
- 3 External links
- 4 References

**Cereal brands – present cereals**

- 100% Bran - Currently Only Available in the United States
- Honey Bunches of Oats - with Real Oats
- Selects Cranberry Applesauce

Category/Topic: MICHIGAN

Question: In 1894 C.W. Post created his warm cereal drink Postum in this Michigan city

Candidate Answers (possible answers to the question) are identified in the search results. They are found by looking at document titles (including a variety of title variants and expansions) and possible answers in the text of the documents and passages, such as *named entities*, noun phrases, anchor text, dates, etc. The Candidate Answers are get their first evidence feature scores from their corresponding document search rank and passage search rank.

The screenshot displays search results for the query 'Will Keith Kellogg'. On the left, a Wikipedia article for 'General Foods' is visible, with a red circle around the title. Below it, a 'Post Foods' section is also circled. On the right, a Britannica entry for 'Will Keith Kellogg' is shown, with a red circle around the title. A table titled 'Candidate Answers' and 'Evidence Feature Scores' is overlaid on the right side of the image. Red arrows point from the circled titles in the search results to the corresponding rows in the table.

Candidate Answers	Evidence Feature Scores	
	Doc Rank	Pass Rank
General Foods	0	1
Post Foods	2	1
Battle Creek	1	2
Will Keith Kellogg	3	
Grand Rapids		
1895		0

- Responsible for confidence of answers

- Indexes used

- PRISMATIC (relationship search)
- Semantic relations (DBpedia)

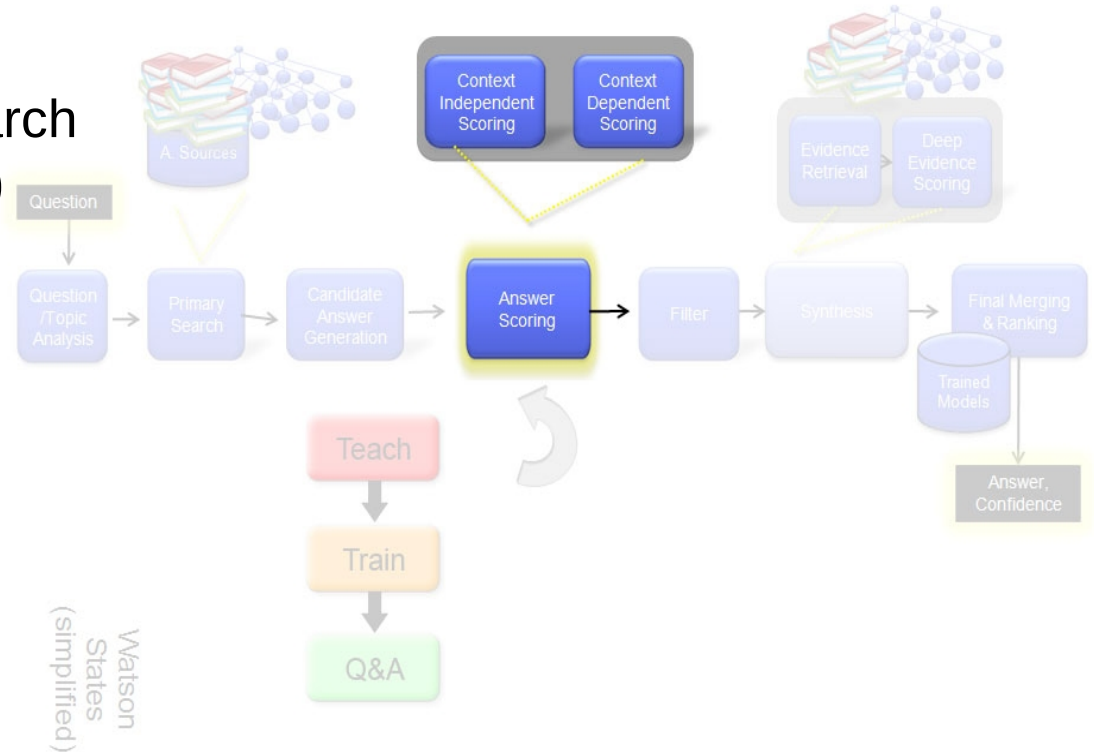
- More than 50 scoring components:

- Taxonomic
- Geospatial (location)
- Temporal
- Source reliability
- Gender
- Name consistency
- Relational
- Passage support
- Theory consistency

- Context dependent (deep evidence)

- Context independent

- Features for machine language



## Category/Topic: MICHIGAN

**Question: In 1894 C.W. Post created his warm cereal drink Postum in this Michigan city**

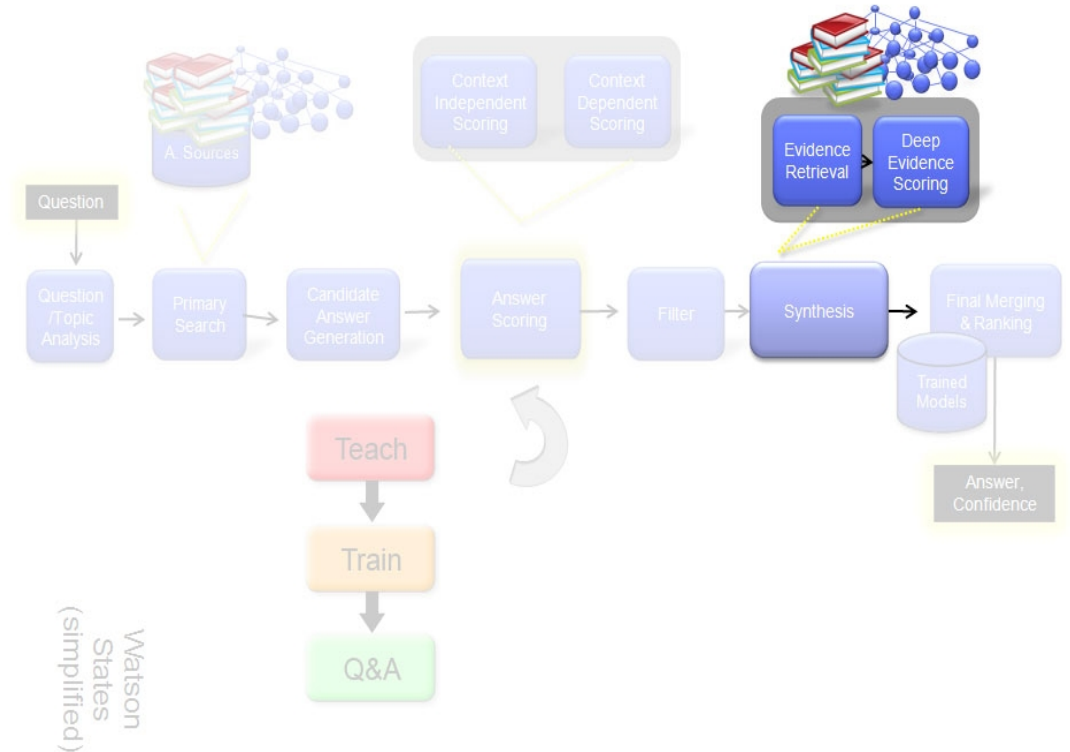
Next, the Candidate Answers are scored using a large number of answer scoring analytics. Some of the analytics use only the candidate answer and the question, along with a large amount of general background knowledge, e.g., the ensemble of Type Coercion (TyCor) scorers. The TyCor scorers estimate the likelihood of a candidate answer being an instance of the Lexical Answer Type (LAT) in the question. In this example, the LAT is “city”, i.e., the correct answer will be a city.

isA(“General Foods”, “city”) = 0.1  
 isA(“Post Foods”, “city”) = 0.1  
 isA(“Battle Creek”, “city”) = 0.8  
 isA(“Will Keith Kellogg”, “city”) = 0.1  
 isA(“Grand Rapids”, “city”) = 0.9  
 isA(“1895”, “city”) = 0.0

Candidate Answers	Evidence Feature Scores		
	Doc Rank	Pass Rank	Ty Cor
General Foods	0	1	0.1
Post Foods	2	1	0.1
Battle Creek	1	2	0.8
Will Keith Kellogg	3		0.1
Grand Rapids			0.9
1895		0	0.0

# Step 5: Supporting Evidence

- Passage search
- Much like a primary search, but requires candidate answer as a term
- Further scored to ensure candidate answer context
- Shared scoring solutions:
  - Passage term match
  - Skip-bigram
  - Text alignment
  - Logical form answer candidate scoring



- Merging

- Due to candidate count usually duplicates exist
- Requires normalizing scores per feature to make merger

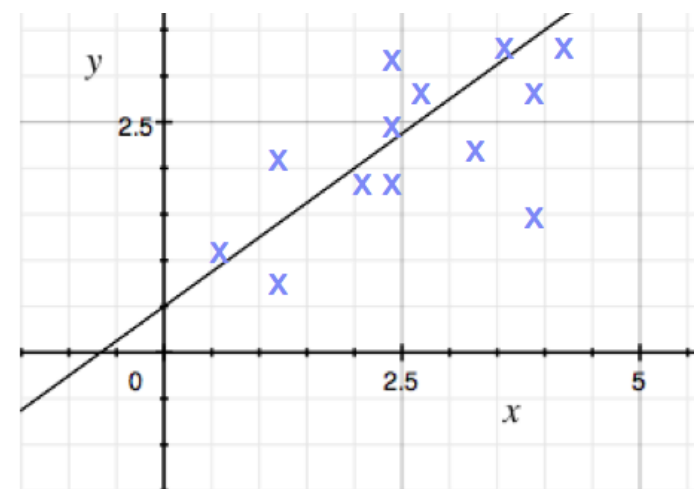
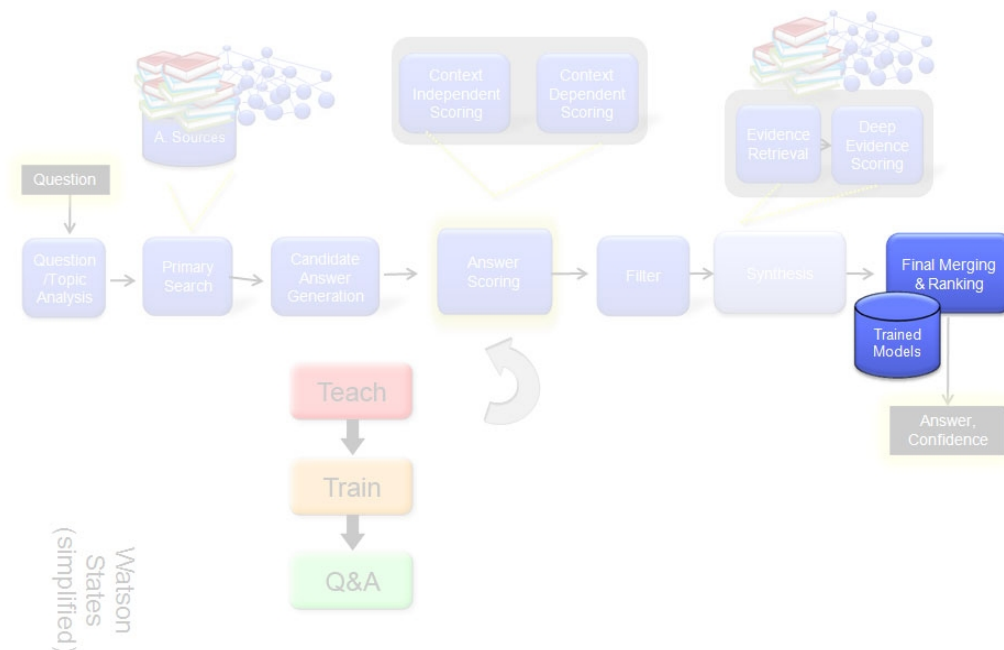
- Ranking

- Use of ML and IBM® SPSS® over training data to create the model to rank future results
- Linear and logistic regression techniques

- Teach-train-execute cycle

- 10,000 training questions and 2000 test questions
- Estimate 48 hours with 11 blade subordinates

—





# Step 6: Merging candidate answers and scoring the confidence IBM WATSON™

**Category/Topic: MICHIGAN**

**Question: In 1894 C.W. Post created his warm cereal drink Postum in this Michigan city**

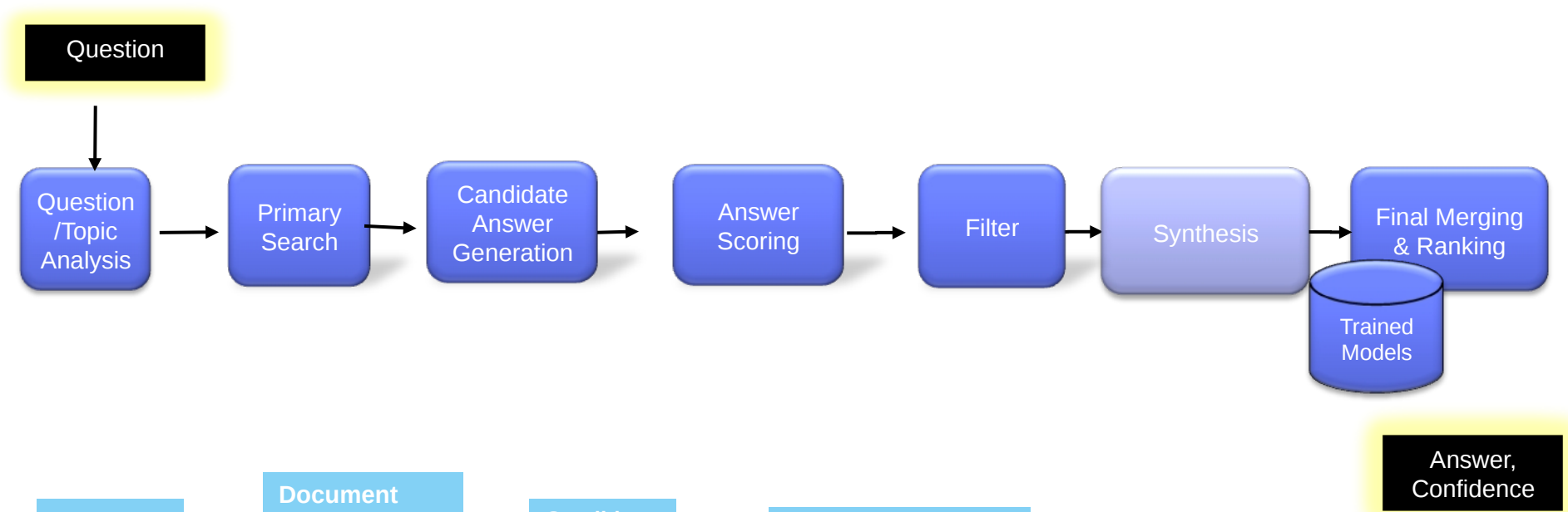
In the final processing step, Watson detects variants of the same answer and merges their feature scores together. Watson then computes the final confidence scores for the candidate answers by applying a series of Machine Learning models that weight all of the feature scores to produce the final confidence scores.

Candidate Answers	Evidence Feature Scores						
	Doc Rank	Pass Rank	Ty Cor	Geo	LFACS	Term Match	Temporal
General Foods	0	1	0.1	0	0.2	22	1
Post Foods	2	1	0.1	0	0.4	41	1
Battle Creek	1	2	0.8	1	0.5	30	0.9
Will Keith Kellogg	3		0.1	0	0	23	0.5
Grand Rapids			0.9	1	0	10	0.5
1895		0	0.0	0	0	21	0.6

**Correct Answer**

Machine Learning Model Application

Final Answers	Confidence
Battle Creek	0.946
Post Foods	0.152
1895	0.040
Grand Rapids	0.033
General Foods	0.014



LAT
Michigan City

Document Search Results	
R	Title
0	General Foods
1	Battle Creek
2	Post Foods
3	Will Keith Kellogg

Candidate Answers
General Foods
Post Foods
Battle Creek

Evidence Features	
Ty Cor	Geo
0.1	0
0.1	0
0.8	1

Final Answers	Confidence
Battle Creek	0.946
Post Foods	0.152
1895	0.040

Answer, Confidence

# Example



**Watson**  
Engagement Advisor

IBM **WATSON**















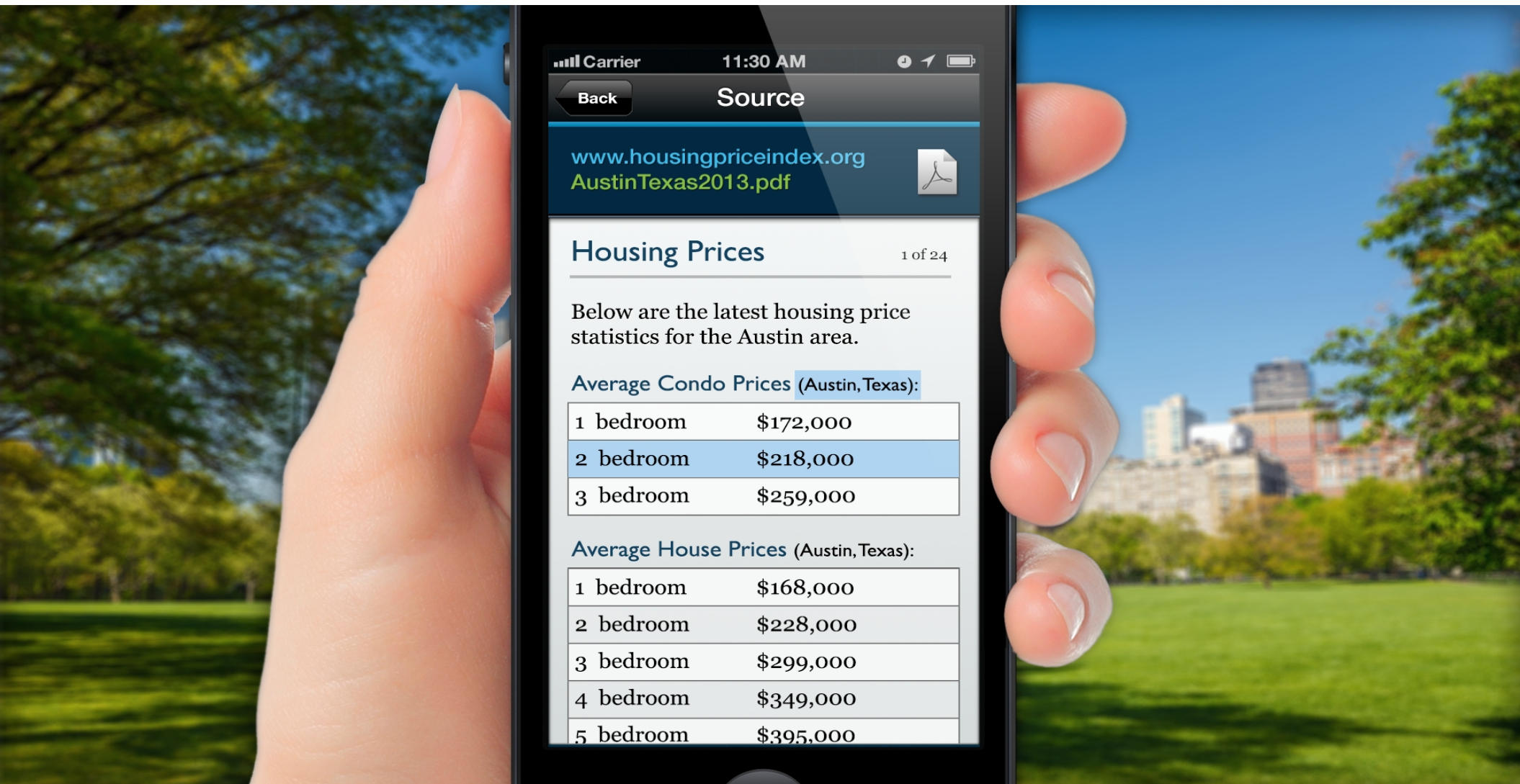






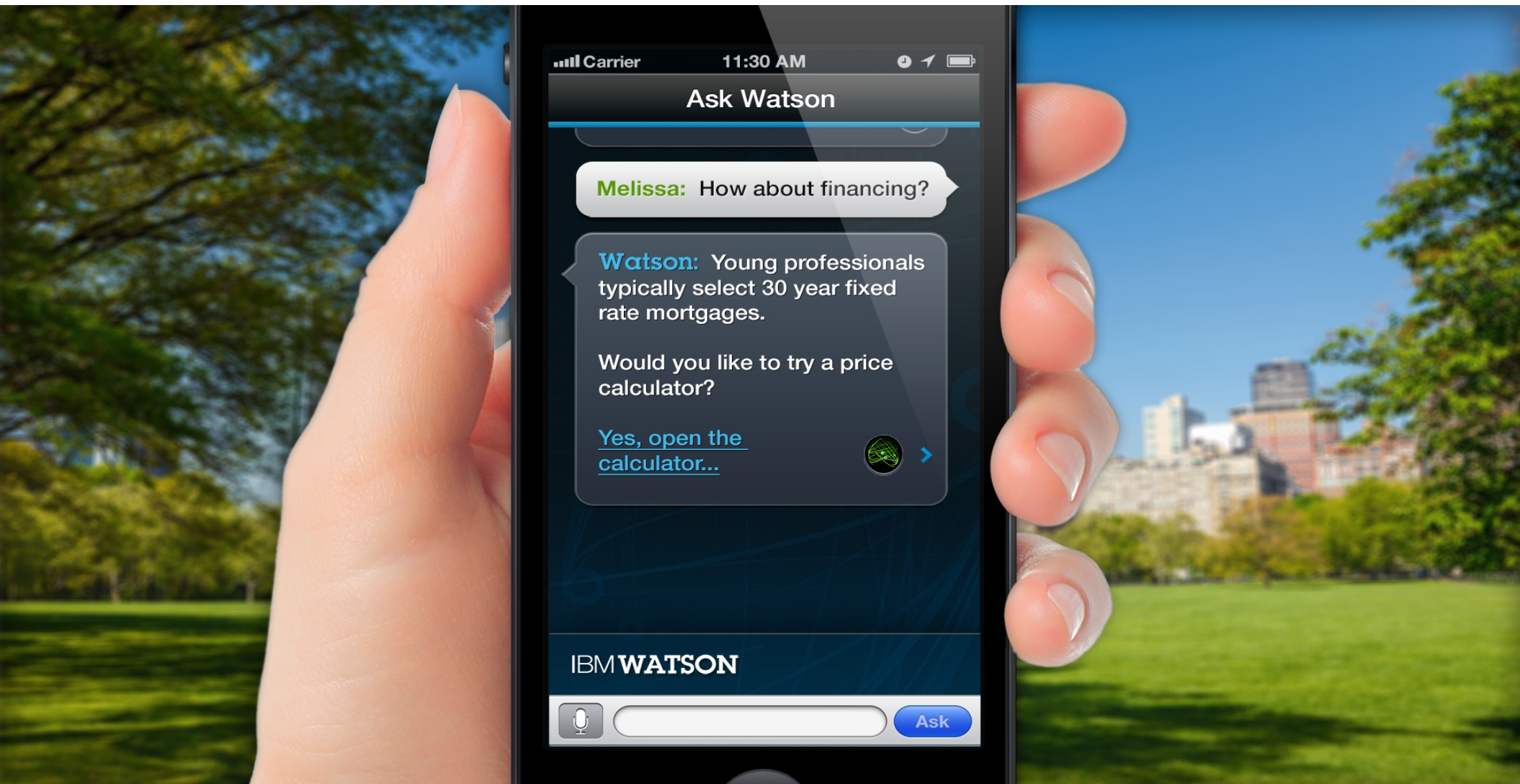


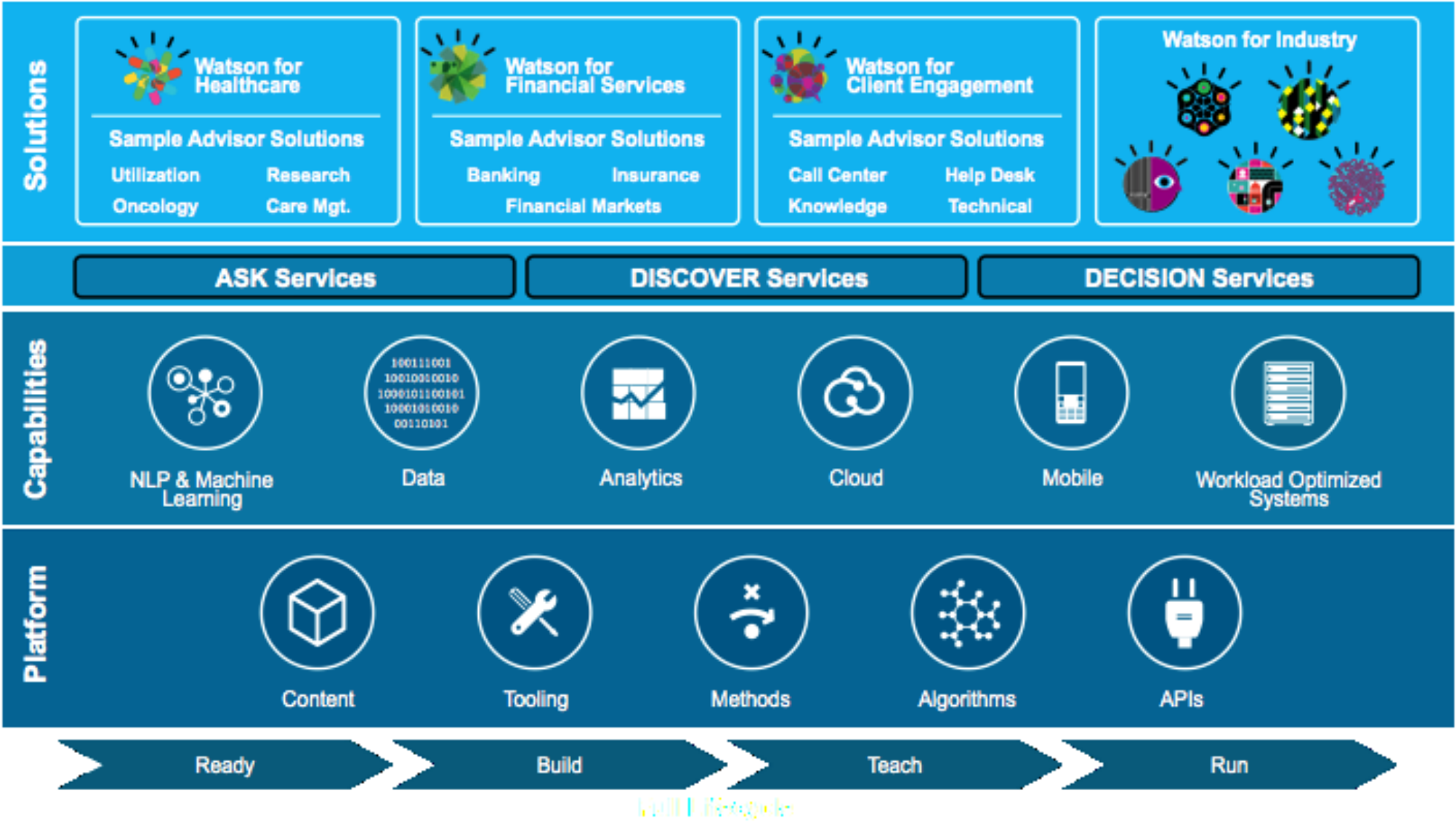












## What it does:

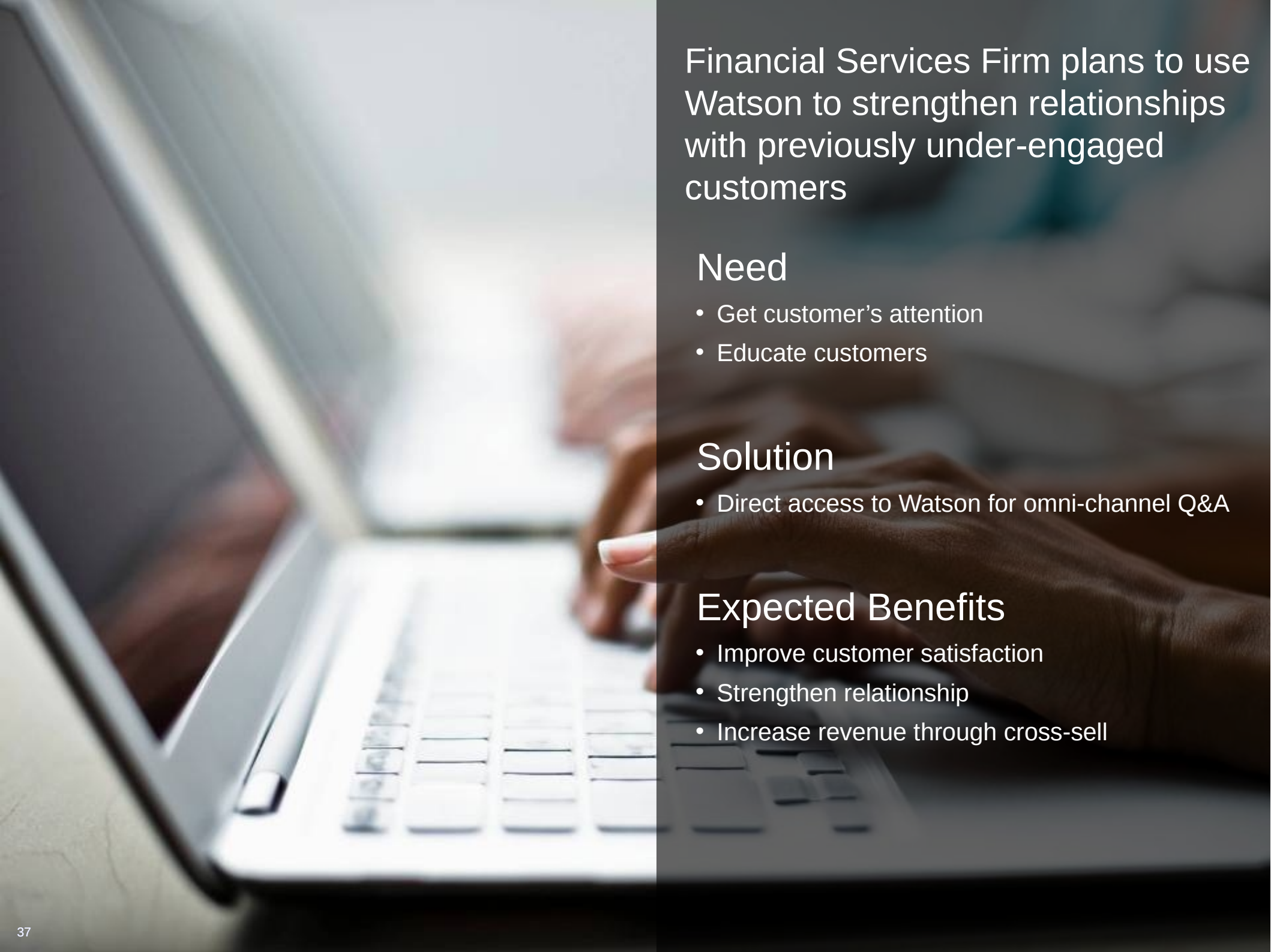
- **Transforms client engagement** by knowing, engaging and empowering clients where they are
- **Develops client relationships** by reaching out to clients who do not leverage traditional channels
- **Empowers consumers and contact center agents** to take informed action with confidence

## How it does it:

- **Answers questions and guides users** through processes with plain-English dialogue
- **Leverages natural language** to interact with users and build knowledge and expertise
- **Utilizes evidence evaluation and learning** to provide informed and effective responses to users







Financial Services Firm plans to use Watson to strengthen relationships with previously under-engaged customers

## Need

- Get customer's attention
- Educate customers

## Solution

- Direct access to Watson for omni-channel Q&A

## Expected Benefits

- Improve customer satisfaction
- Strengthen relationship
- Increase revenue through cross-sell



Mobile Phone Provider plans to use Watson to differentiate the company with personalized service and support

## Need

- Meet changing expectations
- Reduce churn
- Beat competition

## Solution

- Omni-channel self-service
- Guide through processes

## Expected Benefits

- Increase loyalty
- Decrease churn
- Grow customer base

# IBM is working with industry leaders to address this opportunity IBM **WATSON**



*"We believe Watson is going to be a key facilitator to this critically important priority."*

*"Watson can help us make better use of the abundance of information to give higher value response to our customers."*



*"We expect Watson to have a significant impact on our customer's experience."*



*"We believe technology, like Watson, can create a competitive differentiator for us."*



*"We envision Watson as a key strategy for engaging our customers in dialog."*

Questions or comments?  
dmobley@us.ibm.com Or  
dave.mobley@uky.edu



Further reading

IEEE collection:

<http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=6177717&punumber=5288520>



IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at <http://www.ibm.com/legal/copytrade.shtml>.

Other product and service names might be trademarks of IBM or other companies.