



# WATSON ANALYTICS

Prepared for Dataskill, Inc. by Andrew Hook

January 2012



[www.superwatson.com](http://www.superwatson.com)



Executive Introduction .....	2
Introduction .....	3
What is Watson?.....	3
What Watson must do to answer a question .....	3
Watson’s lessons for analytics.....	4
Watson now .....	6
Conclusion .....	7
About Dataskill .....	7



## Executive Introduction

It has been nearly a year since a talking computer stunned humanity by beating the world's best players at the TV game show "Jeopardy!".

IBM's Supercomputer Watson may have appeared to many as a glimpse of the future, but the future might not be as far in the distance as you think. There are lessons that can be learned from Watson and applied to your organization's analytics today.

Now IBM has decided it is ready to tackle something a little more involved. Watson is about to go to medical school, and will even study a specialty: Oncology. Sometime this year, after studying and even taking exams to prove what it has learned, Watson will be assigned to assist human physicians in the treatment of breast, lung and colon cancer.

Dataskill is working with IBM to bring together a set of transformational technologies to drive optimized outcomes.

## Introduction

In the past few years we have seen amazing advances in information analytics. But when IBM's Supercomputer Watson competed on *Jeopardy!* earlier this year, many thought we were seeing a glimpse of the future. It is perhaps surprising then to discover that there are lessons which can be learned from Watson's analytics that can be put into practice today.

This paper will outline those lessons and briefly look at how Watson's unique analytics abilities are being applied today in Healthcare.

## What is Watson?

Watson was named after IBM's founder Thomas John Watson (1874-1956). IBM describes Watson as:

*"an application of advanced Natural Language Processing, Information Retrieval, Knowledge Representation and Reasoning, and Machine Learning technologies to the field of open domain question answering"*

It is a supercomputer based on 90 IBM Power 750 servers with 2880 processor cores and 16 Terabytes of RAM. It uses pioneering Natural Language Processing (NLP) to focus in on keywords in a clue and searches millions of lines of text to find an answer.



This type of technology may not be in reach of many organizations today but lessons can be learned from Watson which can be applied to analytics today.

## What Watson must do to answer a question

Although many focus on the hardware or the "Natural Language Processing" of Watson the analytics plays a significant role in its success.

In order to consistently answer questions faster and more accurately than its competitors Watson must “decompose” a question and generate multiple hypotheses for an answer. It uses machine learning

algorithms to develop 3000 predictive models every time it tries to answer a question. It then gathers evidence from its 4 terabyte “knowledge base” of structured and unstructured data to support each hypothesised answer, before evaluating the probabilities associated with the collected evidence. It evaluates all of these hypotheses in parallel before proposing the final answer, all within a few seconds.

This is big data analytics.



## Watson’s lessons for analytics

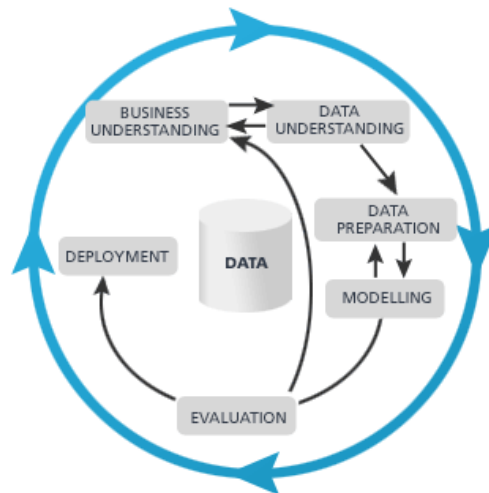
⇒ **Big Data Analytics.** Watson uses a large number of predictive models to analyze big data in real time. Rather than paralleling a single analysis algorithm Watson applies several different predictive and scoring algorithms concurrently. The fact that it uses a 4 terabyte knowledge base shows organizations that the amount of data doesn’t have to be restrictive to analytics.

While the data sets Watson was analyzing during Jeopardy were relatively small by Big Data standards, being able to quickly and effectively analyse unstructured data is representative of many big data analytics situations. The difference is perhaps that often you don’t always know what data you will need to analyze, where it will come from, how large each data set will be, how clean it will be and how long you will have to provide an answer.

⇒ **Despite the hype predictive analytics does not predict the future.** Part of Watson’s success lies in the confidence it gives to its predictions. Watson generates thousands of possible answers and then scores those answers using probabilities to arrive at its final answer. This is the heart of Predictive Analytics which uses similar techniques to help turn uncertainty about the future into usable probabilities, to help decision makers make the best possible decision.



⇨ **Predictive analytics is an iterative process.** Watson improved its answers over time. The answers it gave were analysed, and the techniques were refined in a feedback loop that allowed for better answers and better performances. The 3000 predictive models at the heart of Watson were “tuned” over several months in 134 sparring matches with previous jeopardy winners. Businesses successfully utilizing big



analytics can take this process of knowledge discovery even further, identifying questions, exploring the answers and asking new questions based on those answers. This iterative quality of data analysis, rather than incremental exploration, can lead to a deeper understanding of business and markets, and begin to answer questions never before considered.

⇨ **Watson demonstrates real time analytics is possible.** Speed to decision is the corner stone of real time analytics. Watson was not only able to come up with answers with a required level of confidence but also physically buzz in before its human competitors. For organizations with vast stores of data the value of information has diminishing returns as time goes by. If the information is not acted on, its value and usefulness plummets. For big data analytics speed to decision is just as important, even though the data stores might be huge, the value of information is reduced if decision makers are forced to wait for an answer to be returned. With answers being returned in as little as three seconds, Watson demonstrates that real time analytics is possible, even for very complex problems.

⇨ **Using different techniques to analyse data.** Watson uses different techniques to assess evidence and data. This is an important method which can be applied to analytics as best practice. By combining different techniques and methods to the same problem it is often possible to create an analytical model which is more accurate than an individual method or technique.

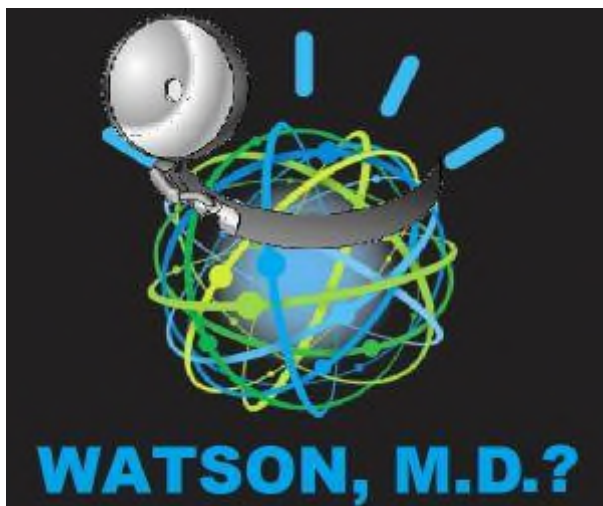
⇨ **A barrier to analytics is its complicated interface.** One of the barriers to adoption of analytics which is frequently quoted is a complicated and hard to understand interface which frightens off all but the most confident data analysts. The interaction with Watson through spoken language will be particularly important for the broader use of analytic applications by business users. Watson

demonstrated convincingly that natural spoken language interaction with computers is no longer science fiction or the result of Hollywood special effects teams.

- ⇔ **Success in analytics still requires asking the right questions.** IBM describes Watson as “a system designed for answers”. We shouldn’t underestimate the achievement of Watson to understand the question but ironically the key to success in using analytics effectively is to know the right questions to ask.

## Watson now

Watson’s use of analytics is a key asset in looking at the future uses of the project and going beyond *Jeopardy!* .



There are many potential uses for Watson in health care. According to the April, 2010 Harvard Business Review, doctors inaccurately or incompletely diagnose patient illness an estimated 15 percent of the time.

The health care area is growing rapidly with new technology forcing the pace of change and solutions becoming increasingly intricate and complex.

A medical Watson, with Natural Language processing, could help physicians find evidence from thousands of journals and research papers or similar cases and suggest diagnosis and treatment, in real time, for the physicians to review.

The analytics capability of Watson bring these to the real world and in the best tradition of analytics will help decision makers, in this case doctors and nurses, make better decisions.

In September this year IBM announced a partnership with WellPoint to use Watson technology for Health Care Analytics. Under the agreement, WellPoint will develop Watson based health care through the delivery of up-to-date evidence-based health care.



## Conclusion

IBM's Watson pushes the boundaries of analytics. There are many lessons we can learn from Watson's use of analytics, but ironically for the *Jeopardy!* winner, as well as being able to answer the questions being posed, analytics is also about asking the right questions.

In order to gain value from analytics results must be linked to decisions. Better decisions are the building blocks of high performance.

More insightful smarter decisions ⇨ Improved business outcomes ⇨ High Performance

But, in order to make better decisions you need to ask the **right** questions.

To find out more about how Dataskill can help your organization with its analytics program please contact Dataskill on (800) 481-3282 or visit the Dataskill website [www.dataskill.com](http://www.dataskill.com).

## About Dataskill

Dataskill, established in 1981, is a business intelligence, data integration and software engineering firm. Trusted by over 500 clients throughout the decades, Dataskill has been helping some of the world's leading companies create and implement technology-centric solutions that deliver business results.

Dataskill's team of highly-skilled technical and business consultants help clients to:

- ◆ Design and deliver custom software solutions
- ◆ Unlock the value of data by improving its quality, availability, and management
- ◆ Harness advanced analytics to fuel high-performance business processes.

Dataskill has been an IBM Business Partner for 15 years and in 2011 launched a WATSON technology-enabled practice focused on breakthrough solutions for Healthcare.

Dataskill is headquartered in San Diego, with offices in Mexico and the United Kingdom currently with 50 employees in total and annual revenues of \$15 million.