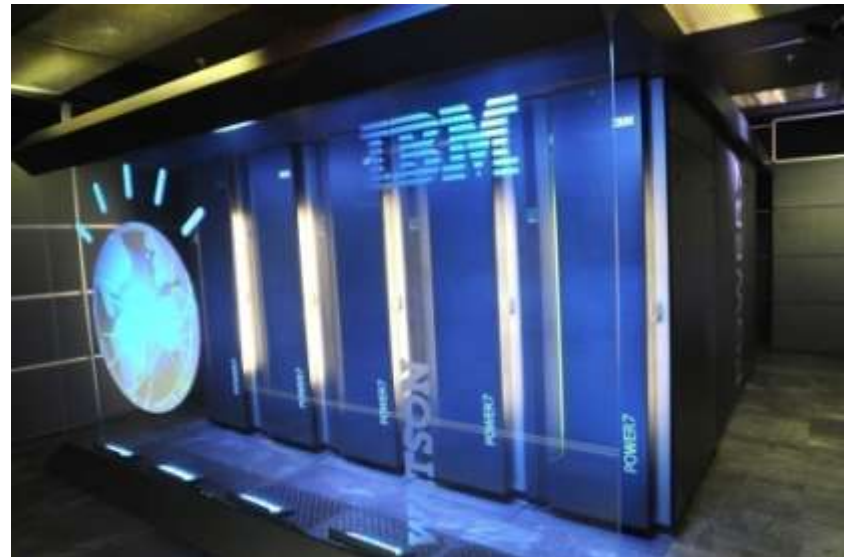


# IBM Watson: НОВЫЙ ВИТОК ЭВОЛЮЦИИ суперкомпьютеров IBM



## История

- 1944 Mark I**  
Первый программируемый арифмометр
- 1952 IBM 701**  
Первый компьютер на вакуумных трубках
- 1954 IBM NORC**  
Вычисление числа Пи с 3089 знаками за 13 минут
- 1955 IBM STRETCH**  
Конвейер, предвыборка, memory interleaving
- 1991 IBM + Thinking Machines**  
Идея о массивно-параллельных компьютерах
- 1997 IBM Deep Blue\***  
Шахматный суперкомпьютер
- 2000 IBM ASCI White**  
Впервые система IBM - № 1 в Top 500
- 2004 IBM Blue Gene/L**
- 2008 IBM Roadrunner**
- 2010 IBM Watson\***  
DeepQA



## Кто такой Watson?

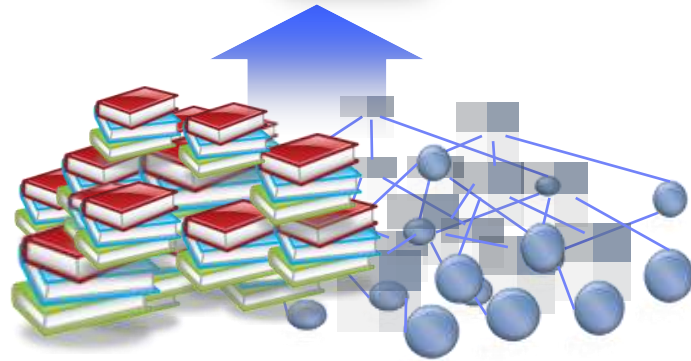
### ▪ Предпосылки для появления

- Лавинообразный рост количества обрабатываемой информации
- Необходимость в экспертных системах, обрабатывающие запросы на естественном языке
- Использование существующих знаний накопленных в форме естественных языков

### ▪ Какие задачи решает Watson?

- Обработка текстов на естественном языке
- Извлечение информации
- Ответы на вопросы

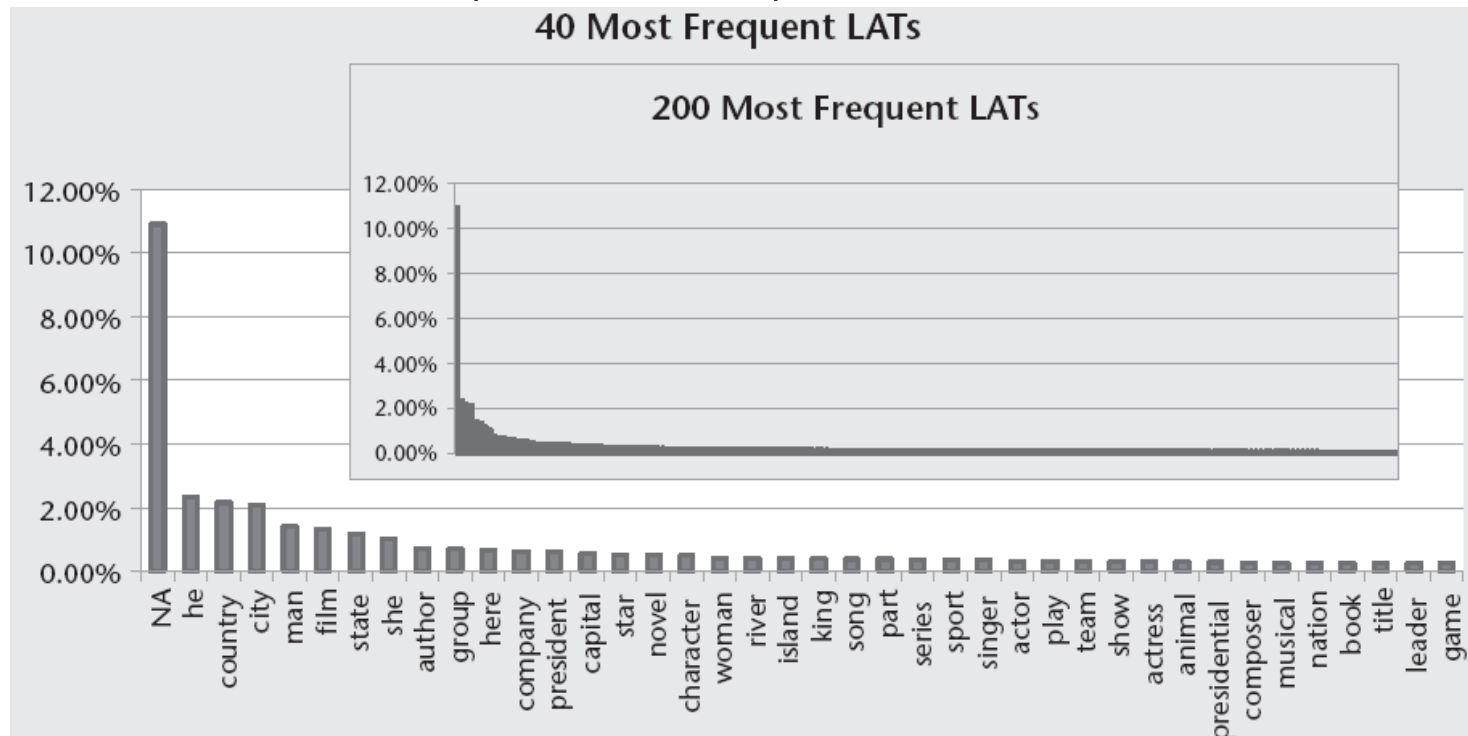
# DeepQA: как создается ответ?



**Огромные объемы текстов  
из разнообразных областей  
знания**

## DeepQA: определение области поиска

- Загрузка источников
- Анализ вопросов к системе
- Анализ области поиска (LAT-анализ)



## DeepQA: анализ вопроса

- Классификация вопроса
- Поиск LAT
- Поиск фокуса вопроса
- Поиск связей
- Декомпозиция вопроса



## DeepQA: генерирование гипотез

- Поиск источников для ответа
- Выделение вариантов ответа
- Watson генерирует несколько сотен вариантов
- Существование правильного ответа
- Гипотеза = вопрос + вариант ответа



## DeepQA: оценка гипотез

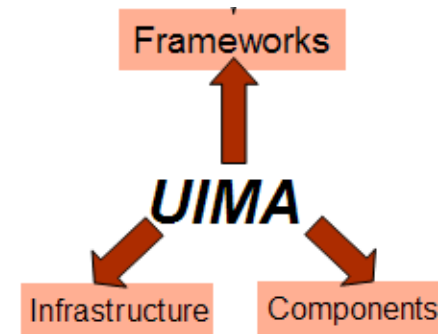
- Сбор доказательств
- Численная оценка результата
  - Доля общих терминов между вопросом и гипотезой
  - Оценка соответствия последовательностей
  - Анализ семантических сетей вопроса и гипотезы
- **Пример:** He was presidentially pardoned on September 8, 1974
- **Правильный ответ:** Nixon
- **Одна из неправильных гипотез:** President Ford pardoned Nixon on Sept. 8, 1974
- Оценка терминов и последовательностей одинакова
- На правильный ответ указывает анализ семантических сетей





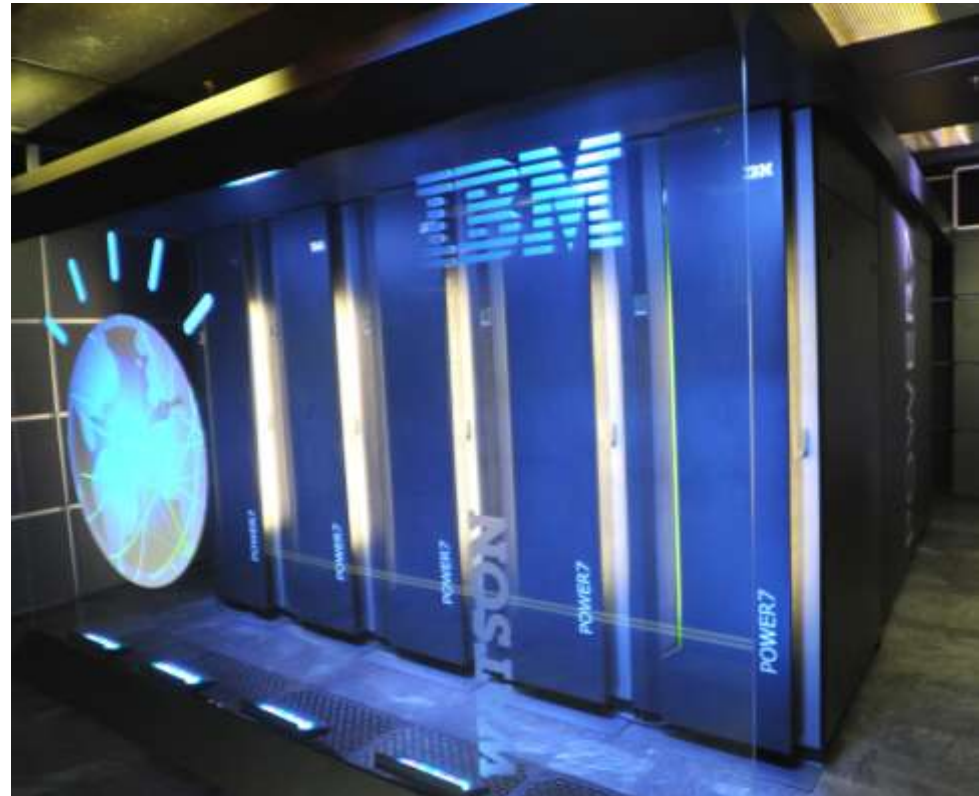
## DeepQA: особенности программной реализации

- Основано на PIQUANT
- Нет интернет-соединения
- Использование UIMA
  - <http://uima.apache.org/>
- Распределенная база знаний
- 3 секунды на ответ (Watson)

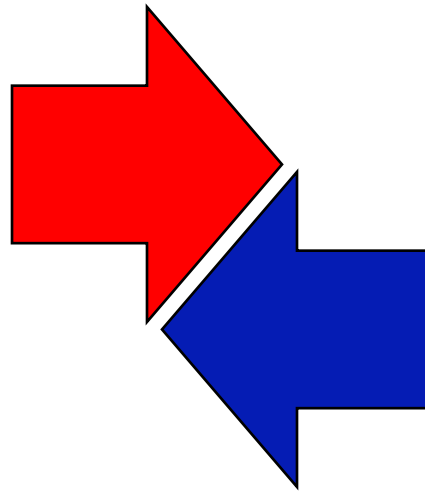
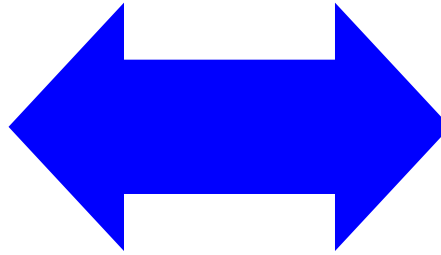


## Watson: аппаратная платформа

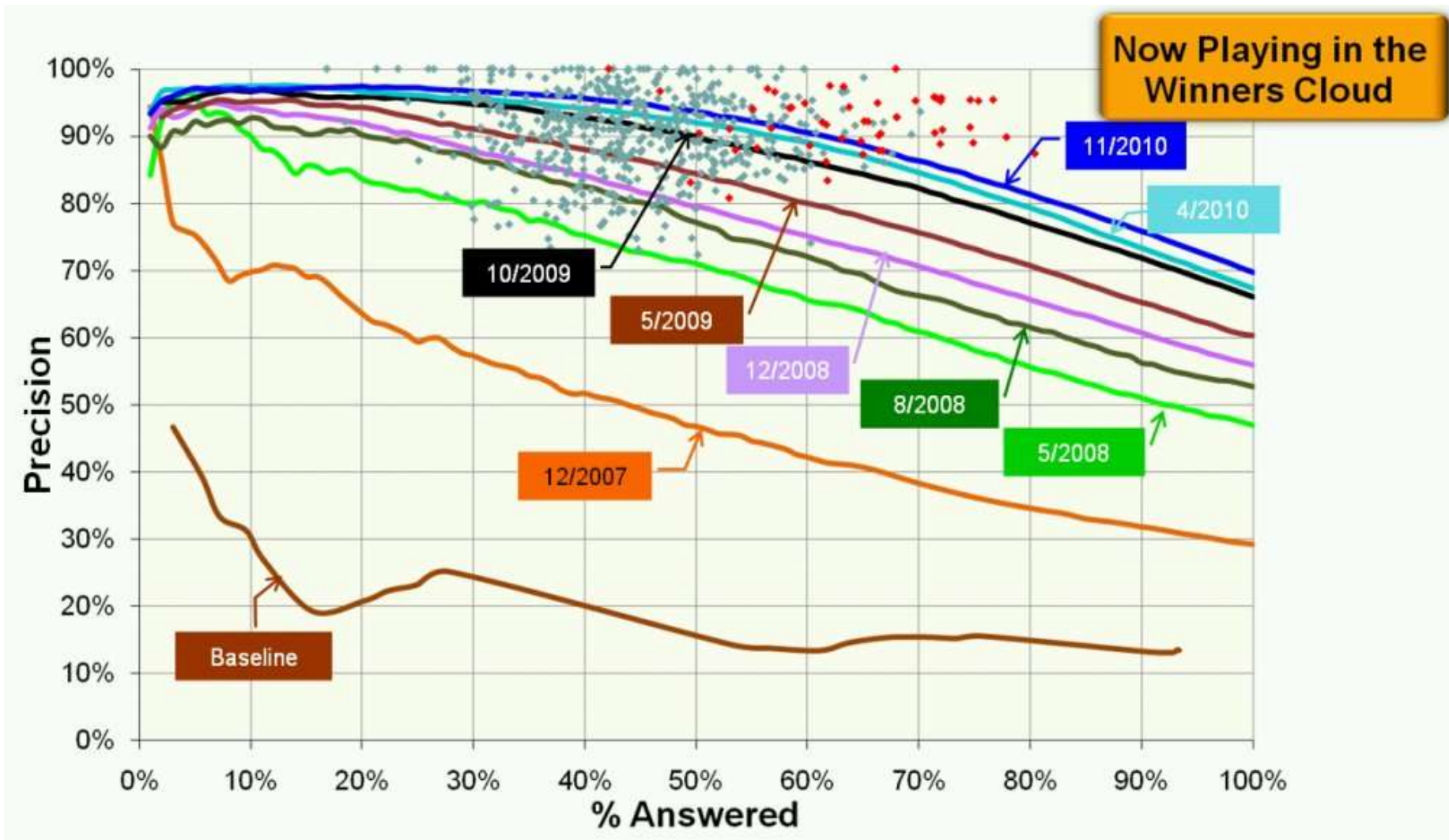
- 10 серверных стоек
- 90 IBM Power 750 серверов
- 2880 процессорных ядер
- Общая память – 15 ТБ
- Высокоскоростная сеть
- В 100х раз быстрее чем DeepBlue



# Watson VS Jeopardy



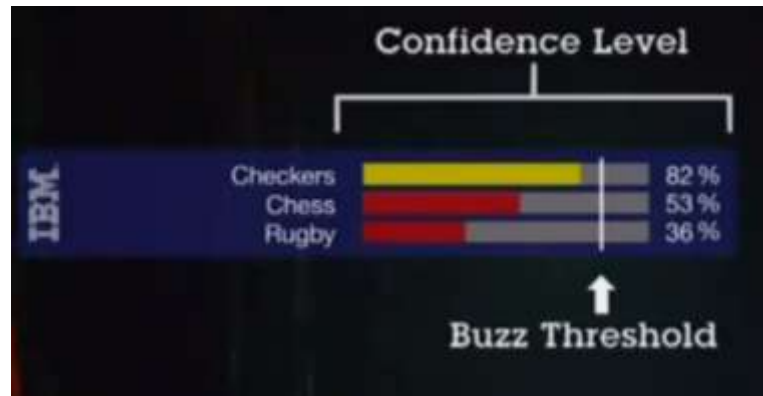
# Watson: подготовка к игре



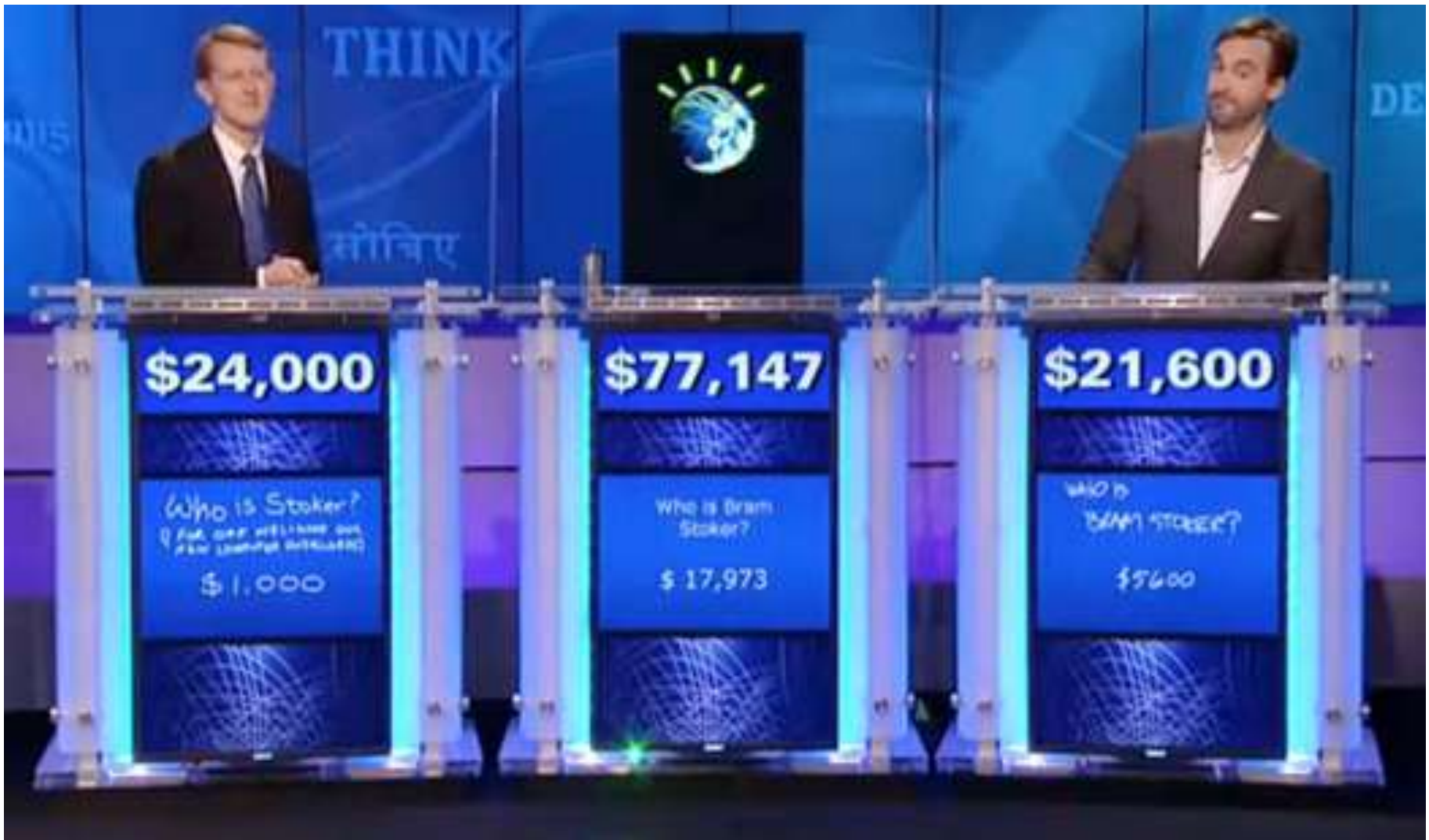
## Watson: особенности матча

- Без аудио/видео вопросов
- Без специальных вопросов
- Watson тоже нажимал на кнопку
- Ожидание до окончания чтения вопроса
- Watson не отвечал, если не уверен

THE FONT USED BY  
JEOPARDY! CLUE  
SCREENS AS WELL AS  
THE **CAPCOM** LOGO



## Watson: результаты матча



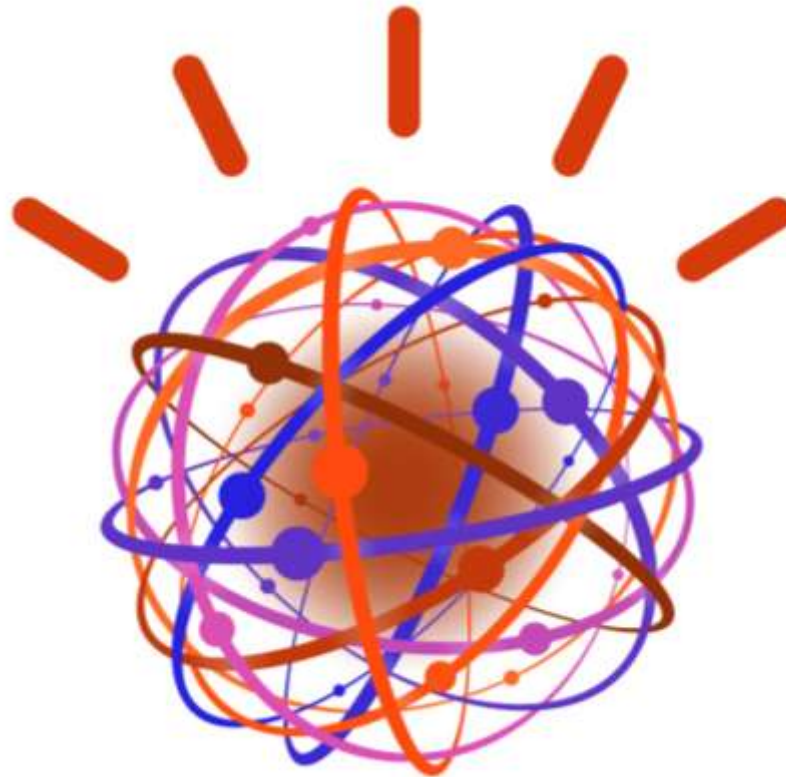
## DeepQA: игры закончились

### ■ Области применения:

- Финансы
- Здравоохранение
- Call-центры
- Юриспруденция



Спасибо за внимание!



[artyom@ru.ibm.com](mailto:artyom@ru.ibm.com)