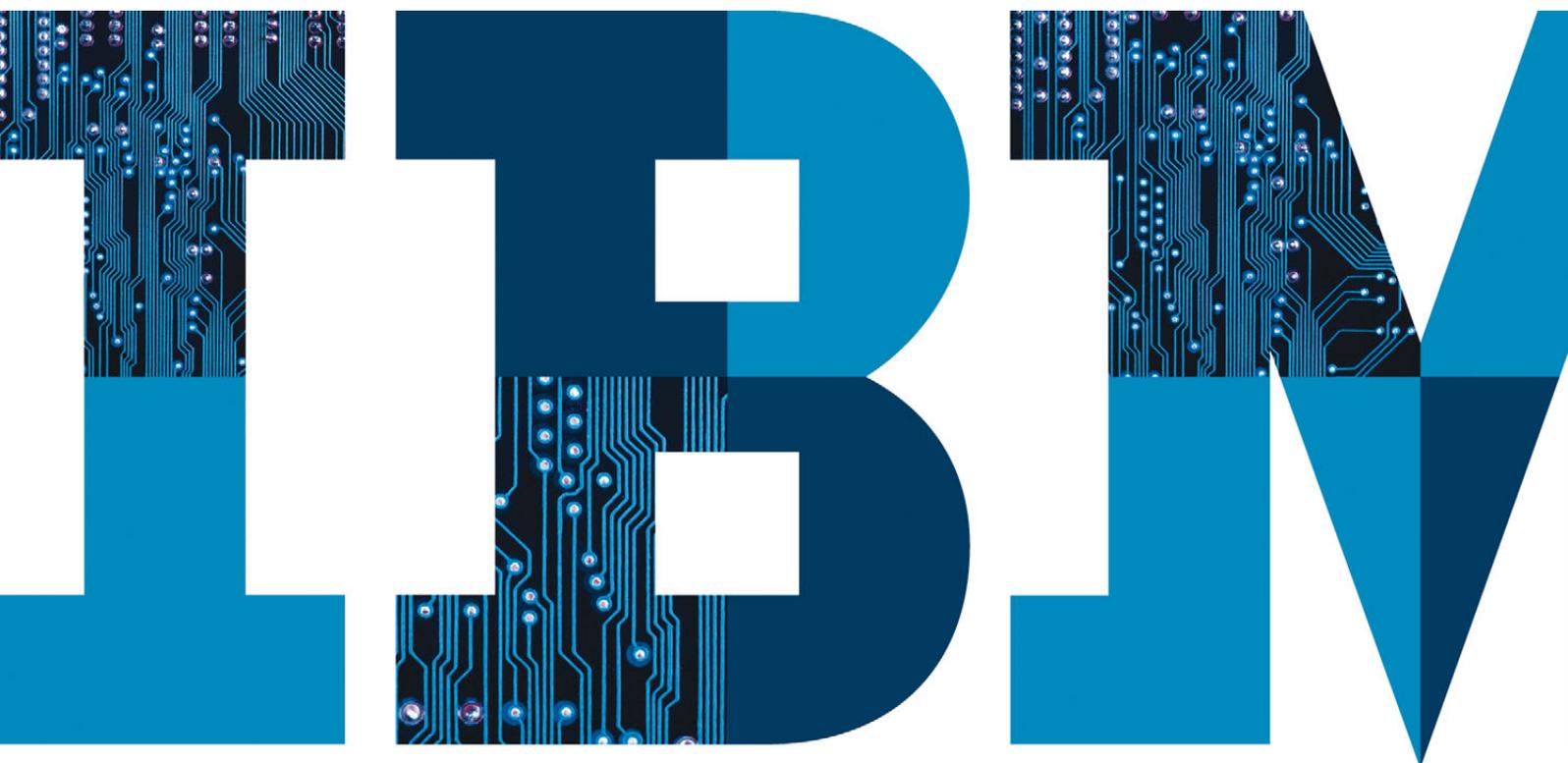


Watson – Um Sistema Projetado para Respostas

*O futuro do design de sistemas otimizados para
carga de trabalho*



Resumo

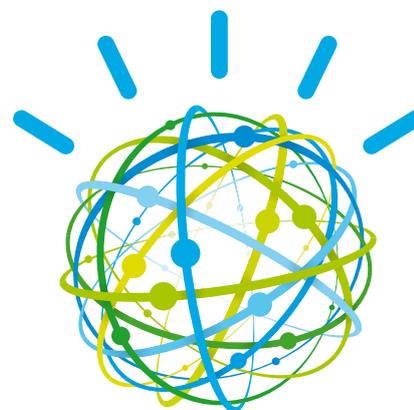
No último século, a IBM realizou várias inovações científicas graças ao seu compromisso com a pesquisa e sua tradição de Grandes Desafios. Esses Grandes Desafios – como o Deep Blue®, projetado para rivalizar com o campeão mundial de xadrez Gary Kasparov – são um esforço para impulsionar a ciência de maneiras que antes não eram consideradas possíveis. O Watson é o mais novo Grande Desafio de Pesquisa da IBM, projetado para desenvolver a ciência do processamento de linguagens naturais através de avanços na tecnologia de perguntas e respostas.

O Watson é um sistema otimizado para carga de trabalho, baseado na arquitetura IBM DeepQA e executado em um cluster de servidores baseados em processadores IBM® POWER7®.

Depois de quatro anos de pesquisa e desenvolvimento intensos por parte de uma equipe de pesquisadores da IBM, o Watson competiu no programa de TV Jeopardy! em fevereiro de 2011, competindo no mesmo nível de especialistas humanos em termos de precisão, confiança e velocidade contra dois dos mais conhecidos e bem-sucedidos campeões de Jeopardy!, Ken Jennings e Brad Rutter. Este informe explica o design de sistema otimizado para carga de trabalho do Watson, como ele é um marco no futuro do design de sistemas e por que ele representa um novo paradigma computacional.

Jeopardy! O desafio da IBM

Em 1997, Deep Blue, o sistema computacional para jogos de xadrez desenvolvido pela IBM Research, chamou atenção no mundo todo ao competir com sucesso contra o campeão mundial de xadrez Gary Kasparov. Foi o auge de um grande desafio para avançar a ciência da computação de uma maneira que criasse grande interesse popular.



Hoje em dia, com as empresas cada vez mais captando informações essenciais aos negócios em documentação de linguagem natural, há um interesse crescente em sistemas otimizados para carga de trabalho que analisam profundamente o conteúdo de perguntas em linguagem natural para responder a elas com precisão. Avanços na tecnologia de respostas a perguntas (QA, question answering) vão ajudar cada vez mais profissionais na tomada de decisões críticas e pontuais em áreas como atendimento médico, business intelligence, descoberta de conhecimento, gerenciamento de conhecimento corporativo e atendimento ao cliente.

Tendo a QA em mente, a IBM se impôs o desafio de desenvolver um sistema computacional chamado “Watson” (em homenagem a Thomas J. Watson, fundador da IBM), que pudesse competir ao nível de campeões humanos em tempo real no programa de perguntas e respostas da TV dos EUA *Jeopardy!* O programa, transmitido nos EUA há mais de 25 anos, coloca três concorrentes humanos para competir entre si, respondendo a perguntas de linguagem natural rica em uma ampla gama de tópicos, com penalidades para respostas erradas. Nessa



competição de três pessoas, confiança, precisão e velocidade de resposta são de suma importância, pois os concorrentes geralmente pensam na resposta nos poucos segundos que o apresentador demora para ler uma pista. Para competir no jogo no mesmo nível dos campeões humanos, um sistema de computador teria de responder cerca de 70 por cento das perguntas com precisão maior que 80 por cento, em três segundos ou menos.

O Watson representa um avanço impressionante no design e análise de sistemas. Ele executa a tecnologia DeepQA da IBM, um novo tipo de capacidade analítica que pode executar milhares de tarefas simultâneas em segundos para fornecer respostas precisas a perguntas. Ativado pela tecnologia dos processadores IBM POWER7, o Watson é um exemplo das cargas de trabalho de análise complexa que estão se tornando cada vez mais comuns e essenciais para o sucesso e a competitividade dos negócios no ambiente atual de grande fluxo de dados.

Watson competiu contra dois dos mais conhecidos e bem-sucedidos campeões de *Jeopardy!* – Ken Jennings e Brad Rutter – em uma competição de duas partidas, transmitida em três noites consecutivas a partir de 14 de fevereiro de 2011.

IBM DeepQA

DeepQA é uma arquitetura probabilística paralela maciça baseada em evidências. Para o desafio *Jeopardy!*, mais de 100 técnicas diferentes são usadas para analisar a linguagem natural, identificar fontes, encontrar e gerar hipóteses, encontrar e pontuar evidências, e mesclar e classificar hipóteses. Muito mais importante do que qualquer técnica em particular é a maneira em que todas essas técnicas são combinadas no DeepQA, de forma que abordagens sobrepostas podem unir forças e contribuir para melhorias na precisão, confiança ou velocidade.



A DeepQA é uma arquitetura com uma metodologia complementar, mas não é específica para o Desafio *Jeopardy!* A IBM começou a adaptá-la para diferentes aplicações comerciais e outros problemas desafiantes a serem explorados, inclusive nas áreas de medicina, pesquisa corporativa e jogos.

Os princípios predominantes da DeepQA são:

1. **Paralelismo maciço:** o paralelismo maciço é explorado na consideração de várias interpretações e hipóteses.
2. **Muitos especialistas:** facilitam a integração, aplicação e avaliação contextual de uma ampla gama de análises probabilísticas de perguntas e conteúdos fracamente acopladas.
3. **Estimativa universal de confiança:** não há um único componente que se compromete com uma resposta; todos os componentes produzem características e confianças associadas, pontuando interpretações diferentes de perguntas e conteúdos. Um substrato subjacente de processamento de confiança aprende a empilhar e combinar as pontuações.
4. **Integração de conhecimento superficial e profundo:** equilibra o uso de semântica restrita e semântica superficial, aproveitando-se de muitas ontologias formadas livremente.

Velocidade e escala horizontal

O DeepQA é desenvolvido usando a Apache UIMA, uma implementação de estrutura da Unstructured Information Management Architecture (Arquitetura de Gerenciamento de Informações Não-Estruturada). A UIMA foi projetada para dar suporte à interoperabilidade e escala horizontal de aplicativos de análise textual e multimodal. Todos os componentes da DeepQA são implementados como *anotadores UIMA*. Esses componentes analisam o texto e produzem *anotações* ou asserções sobre ele. Com o tempo, Watson evoluiu de forma que o sistema agora tem centenas de componentes. A UIMA facilitou a integração, teste e avaliação rápida dos componentes.

As implementações iniciais do Watson eram executadas em um único processador, que exigia duas horas para responder a uma única pergunta. Mas a computação da DeepQA é embaraçosamente paralela; portanto, ela pode ser dividida em várias partes independentes, cada uma das quais podendo ser executada por um processador separado. A UIMA-AS, parte da Apache UIMA, permite a escala horizontal de aplicativos UIMA usando mensagens assíncronas. O Watson utiliza a UIMA-AS para se escalar em 2.880 processadores POWER7 em um cluster de 90 servidores IBM Power® 750. A UIMA-AS gerencia toda a comunicação entre processos, usando o padrão aberto JMS. A implantação da UIMA-AS para POWER7 permitiu ao Watson fornecer respostas em um a seis segundos.

O Watson tem cerca de 200 milhões de páginas de conteúdo em linguagem natural (equivalente à leitura de 1 milhão de livros). O Watson utiliza a estrutura Apache Hadoop para facilitar o pré-processamento em grandes volumes de dados para criar conjuntos de dados internos à memória, usados em tempo de execução. Os anotadores UIMA DeepQA do Watson foram implementados como mapeadores na estrutura de redução de mapa da Hadoop, que os distribuiu pelos processadores no cluster. A Hadoop contribui para a utilização otimizada das CPUs e também fornece ferramentas convenientes para implantar, gerenciar e monitorar o processo de análise de dados.

Aproveitando-se do POWER7

O Watson se aproveita do desempenho de processamento paralelo maciço dos seus processadores POWER7 para executar suas milhares de tarefas da DeepQA simultaneamente em núcleos de processadores individuais. Cada um dos 90 servidores IBM Power 750 em cluster do Watson conta com 32 núcleos POWER7, executados a 3.55 GHz. Executando o sistema operacional Linux®, os servidores são abrigados em 10 racks, juntamente com os nós de E/S e hubs de comunicação associados. O sistema tem um total combinado de 16 terabytes de memória e pode operar a mais de 80 teraflops (trilhões de operações por segundo).

Com seu design inovador de oito núcleos, o POWER7 é ideal para o processamento paralelo maciço dos algoritmos analíticos do Watson. O POWER7 também conta com 500 gigabytes de largura de banda de comunicação interna, contribuindo para uma eficiência excepcional na utilização tanto da memória quanto do processador. E como cada servidor conta com 32 núcleos POWER7 de alto desempenho com até 512 GB de memória, o Power 750 é a plataforma ideal para os processos Java do Watson, que utilizam muitos recursos de processador e de memória.

Projetar o Watson em servidores Power 750, que estão disponíveis comercialmente, foi uma escolha deliberada para garantir uma adoção mais rápida de sistemas otimizados em setores como os de serviços de saúde e financeiros. Esse objetivo é uma diferença fundamental entre o Watson e o Deep Blue, que era um computador altamente personalizado. O Deep Blue era baseado em uma geração anterior da tecnologia de processadores Power, contando com um sistema RS/6000 SP de 30 nós, cada um deles contendo um único processador POWER2 de 120 MHz. Mas além dos processadores POWER2 normais, o desempenho do Deep Blue era aprimorado com 480 processadores especialistas em xadrez.

O mesmo servidor Power 750 usado pelo Watson já é implantado hoje por milhares de empresas em sistemas otimizados que oferecem processamento complexo de análises e transações. A Rice University em Houston, Texas, por exemplo, utiliza sistemas IBM Power 750 para acelerar o entendimento da base molecular do câncer através da aplicação de tecnologias de análise de genoma. Os sistemas POWER7 deram à Rice mais flexibilidade e eficiência, permitindo a eles enfrentarem com um único sistema uma gama mais ampla de desafios de pesquisa do que era possível antes. E a GHY International, uma firma de corretagem alfandegária no Canadá, migrou para um Power 750 novo, executando AIX®, Power i e Power Linux, para dar mais suporte às crescentes transações de comércio internacional dos seus clientes. Com a virtualização PowerVM™, agora a GHY é capaz de implantar capacidades novas no tempo de cinco minutos para dar suporte às necessidades em mudança dos seus clientes.



Um sistema projetado para respostas

Depois de quatro anos de pesquisa e desenvolvimento intensos por parte de uma equipe de pesquisadores da IBM, o Watson demonstrou suas habilidades de competir no *Jeopardy!* contra campeões humanos, com desempenho em nível de especialistas humanos em termos de precisão, confiança e velocidade. O projeto promoveu os campos da análise de dados não-estruturados, do processamento de linguagem natural e do design de sistemas otimizados para cargas de trabalho. Além do *Jeopardy!*, a tecnologia por trás do Watson pode ser adaptadas para resolver problemas comerciais e sociais – por exemplo, diagnosticando doenças lidando com perguntas de suporte técnico on-line e analisando grandes quantidades de documentos jurídicos – e para promover o progresso em vários setores.

A capacidade do Watson para entender o significado e o contexto da linguagem humana e para processar rapidamente informações para encontrar respostas precisas para perguntas complexas guarda um potencial enorme para transformar a forma na qual os computadores podem ajudar as pessoas a realizar tarefas nos negócios e nas suas vidas pessoais.

Para obter mais informações

Para saber mais sobre o Watson, o POWER7 e sistemas otimizados para carga de trabalho, entre em contato com o seu representante de marketing ou o Parceiro de Negócios IBM, ou visite os seguintes sites:

- ibm.com/systems/power/advantages/watson
- ibm.com/systems/power



© Copyright IBM Corporation 2011

IBM Systems and Technology Group
Route 100
Somers, NY 10589

Produzido nos Estados Unidos da América
Fevereiro de 2011
Todos os direitos reservados

IBM, o logotipo IBM, ibm.com Power, POWER7 e DEEP BLUE são marcas registradas da International Business Machines Corporation nos Estados Unidos e/ou em outros países. Se estes e outros termos com marca IBM aparecerem em sua primeira ocorrência nestas informações com um símbolo de marca comercial ou registrada (® ou ™), esses símbolos indicarão marcas comerciais comuns ou registradas nos EUA de propriedade da IBM no momento da publicação destas informações. Essas marcas comerciais também podem ser marcas registradas ou de leis comuns em outros países. Uma lista atualizada das marcas registradas IBM encontra-se disponível na Web em “Copyright and trademark information” (“Informações de copyright e marca registrada”), em ibm.com/legal/copytrade.shtml

Outros nomes de empresas, produtos e serviços podem ser marcas comerciais ou marcas de serviço de terceiros.



Por favor, recicle
