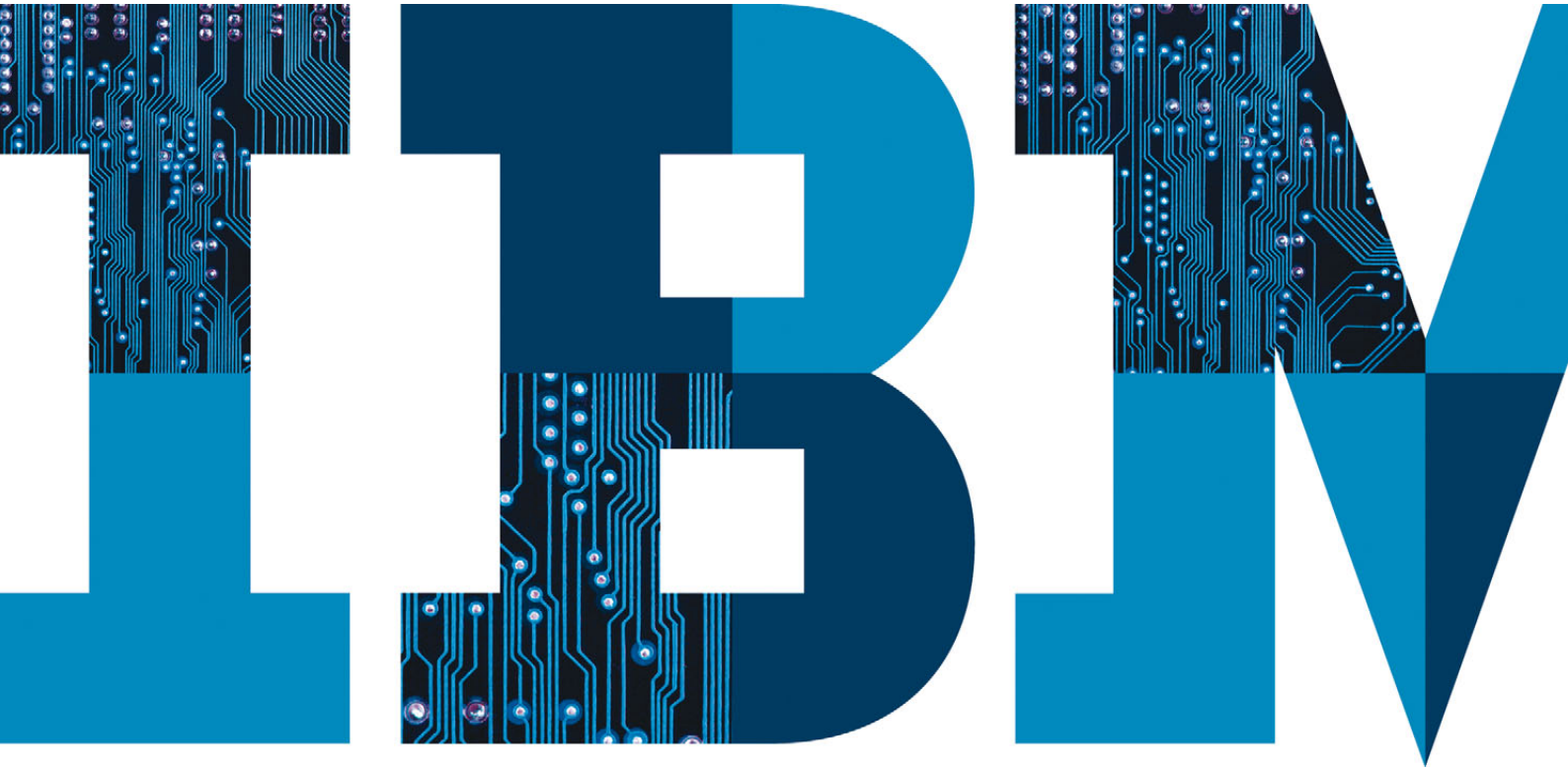
IBM Systems and Technology
Ein IBM Whitepaper

Februar 2011

Watson – Antworten mit System

Die Zukunft des workloadoptimierten Systemdesigns



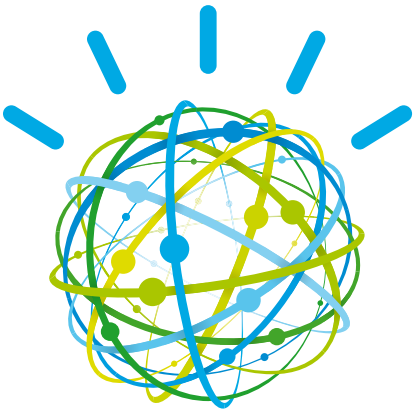
Kurzübersicht

Im vergangenen Jahrhundert hat IBM zahlreiche wissenschaftliche Durchbrüche erreicht – dank seines Forschungsengagements und der Tradition große Herausforderungen anzugehen. Diese großen Herausforderungen, „Grand Challenges“ genannt, brachten vielfach die Wissenschaft in besonderem Maße voran. Beispielhaft für ein solches Projekt ist etwa der Deep Blue. Er wurde entwickelt, um den Schachweltmeister Gary Kasparov herauszufordern. Mit seiner innovativen Technologie eröffnete er der Wissenschaft neue Wege, die vorher als undenkbar galten. Das System Watson ist die jüngste „Grand Challenge“ von IBM. Es soll das Wissen über die natürliche Sprachverarbeitung durch fortschrittliche Frage-Antwort-Technik voranbringen.

Watson ist ein workloadoptimiertes System. Es stützt sich auf die IBM DeepQA Architektur, die in einem Server-Cluster mit IBM POWER7 Prozessoren arbeitet. Nach vier Jahren intensiver Forschungs- und Entwicklungsarbeit durch ein Team aus IBM Wissenschaftlern trat Watson im Februar 2011 im Fernsehquiz *Jeopardy!* gegen die bekannten und erfolgreichen Kandidaten Ken Jennings und Brad Rutter an. Dabei zeigte das System die Genauigkeit, Geschwindigkeit und das Vermögen nach Wahrscheinlichkeit abzuwägen, die den Fähigkeiten eines menschlichen Experten entsprechen. In diesem Whitepaper wird das workloadoptimierte Systemdesign von Watson, seine Vorbildfunktion für die Zukunft des Systemdesigns im Allgemeinen und seine Rolle als Repräsentant eines neuen Computerkonzepts erläutert.

Jeopardy! Eine Herausforderung für IBM

1997 zog Deep Blue, der Schach spielende Computer aus der Forschungsabteilung von IBM, die Aufmerksamkeit der ganzen Welt auf sich, als er gegen den Schachweltmeister Gary Kasparov antrat und diesen besiegte. Dies war der Höhepunkt einer großen Herausforderung, die zum Ziel hatte, die Computerwissenschaft auf eine Weise voranzubringen, die großes öffentliches Interesse wecken würde.



Heute dokumentieren Unternehmen zunehmend mehr kritische Geschäftsinformationen in natürlicher Sprache. Daher wächst das Interesse an workloadoptimierten Systemen, die die Inhalte von Fragen in natürlicher Sprache bis an ihre Wurzeln analysieren, um präzise Antworten liefern zu können. Fortschritte in der Fragenbeantwortungstechnologie („Question Answering Technology“, QA) werden zukünftig immer stärker dazu beitragen, kritische und zeitnahe Entscheidungen auf Gebieten wie dem Gesundheitswesen, der Business Intelligence (BI), der Wissenserhebung, dem Unternehmenswissensmanagement und der Kundenbetreuung zu treffen.

Vor dem Hintergrund von QA setzte sich IBM das Ziel, ein Computersystem mit dem Namen „Watson“ (nach dem IBM-Gründer Thomas J. Watson benannt) zu entwickeln. Es sollte in der amerikanischen Quizshow *Jeopardy!* auf dem Niveau menschlicher Experten in Echtzeit antreten können. In der Sendung, die seit über 25 Jahren in den USA erfolgreich ist, treten drei menschliche Kandidaten gegeneinander an, um komplexe Fragen zu einem breiten Themenspektrum zu beantworten. Für falsche Antworten gibt es Abzüge. In diesem Drei-Personen-Wettkampf sind Wahrscheinlichkeit, Genauigkeit und Reaktionsschnelligkeit von entscheidender Bedeutung.

Normalerweise beantworten die Kandidaten die Fragen bereits, während sie vom Moderator noch vorgelesen werden. Um in diesem Spiel gegen menschliche Spitzenkandidaten antreten zu können, muss ein Computer rund 70 Prozent der gestellten Fragen mit einer Genauigkeit von über 80 Prozent in maximal drei Sekunden beantworten können.

Watson stellt einen beeindruckenden Sprung in der Weiterentwicklung von Systemdesign und -analytik dar. Dabei arbeitet es mit der DeepQA Technologie von IBM. Sie ermöglicht eine neue Form der Analyse, bei der mehrere tausend Aufgaben gleichzeitig in Sekunden verarbeitet werden, um Fragen korrekt zu beantworten. Watson basiert auf der IBM POWER7 Prozessortechnologie. Das System ist beispielhaft für die komplexen Analyseaufgaben, die in einer datenintensiven Welt wie der heutigen stetig zunehmen. Gleichzeitig werden diese Aufgaben mehr und mehr ausschlaggebend für den geschäftlichen Erfolg und die Wettbewerbsfähigkeit von Unternehmen und Organisationen.

Watson trat in zwei Turnieren, die an drei aufeinander folgenden Abenden ab dem 14. Februar 2011 ausgestrahlt wurden, gegen Ken Jennings und Brad Rutter, zwei der bekanntesten und erfolgreichsten *Jeopardy!*-Champions, an.

IBM DeepQA

DeepQA ist eine Massenparallel-Architektur, die auf der Grundlage des probabilistischen Nachweisprinzips arbeitet. Für *Jeopardy!* werden über 100 unterschiedliche Methoden zur Analyse natürlicher Sprache, Identifizierung von Quellen, Ermittlung und Generierung von Hypothesen, Feststellung und Bewertung von Nachweisen und zur Verbindung und Einstufung von Hypothesen verwendet. Weit wichtiger jedoch als irgend eine bestimmte Technik ist die Art und Weise, wie diese Techniken in DeepQA kombiniert werden: Dabei können sich überschneidende Prinzipien ihre Stärken verbinden und zu höherer Genauigkeit, Wahrscheinlichkeit oder Geschwindigkeit beitragen.

DeepQA ist eine Architektur mit einer begleitenden Methodik, aber nicht spezifisch auf *Jeopardy!* zugeschnitten. IBM hat bereits begonnen, das System an unterschiedliche Geschäftsanwendungen und weitere wissenschaftliche Aufgabenstellungen anzupassen, darunter in der Medizin, für Suchmaschinen und Spiele.

Die übergreifenden Prinzipien von DeepQA sind Folgende:

1. **Umfassende Parallelverarbeitung**: Anwendung der massiven Parallelverarbeitung auf multiple Interpretationen und Hypothesen
2. **Viele Experten**: Integration, Anwendung und kontextbezogene Bewertung umfangreicher, lose gekoppelter probabilistischer Frage- und Inhaltsanalysen
3. **Pervasive Konfidenzbewertung**: Kein Einzelelement ist mit einer Antwort verknüpft, sondern alle Elemente erzeugen Kennzeichen und zugehörige Wahrscheinlichkeiten (Konfidenzen), die eine Bewertung unterschiedlicher Frage- und Inhaltsinterpretationen ermöglichen. Dem zugrunde liegt ein Wahrscheinlichkeitsverarbeitungsträger, der lernt, die Bewertungen zu schichten und zu kombinieren
4. **Integration von oberflächlichem und tiefreichendem Wissen**: Ausgleich zwischen der Verwendung strikter und flacher Semantik unter Ausnutzung vieler lose gebildeter Ontologien.

Geschwindigkeit und Skalierbarkeit

DeepQA basiert auf Apache Unstructured Information Management Architecture (UIMA), einer Rahmenimplementierung von UIMA. UIMA wurde konzipiert, um die Interoperabilität und Skalierbarkeit von Text- und multimodalen Analyseapplikationen zu unterstützen. Alle Komponenten in DeepQA werden als *UIMA* Annotatoren implementiert. Hierbei handelt es sich um Komponenten zur Textanalyse und zur Erzeugung von Annotationen oder Assertionen über den Text. Im Laufe der Zeit hat sich Watson weiter entwickelt und hat nun mehrere hundert Komponenten. UIMA unterstützte eine rasche Integration, Prüfung und Bewertung der Komponenten.

Die ersten Implementierungen von Watson arbeiteten mit einem einzelnen Prozessor, der zwei Stunden zur Beantwortung einer Frage benötigte. Die DeepQA Rechenvorgänge arbeiten jedoch hoch parallel. Sie lassen sich in mehrere unabhängige Teile aufspalten, die jeweils von einem eigenen Prozessor verarbeitet werden können. UIMA-AS ermöglicht als Bestandteil von Apache UIMA die Skalierung von UIMA-Applikationen mittels asynchroner Übermittlung. Mithilfe von UIMA-AS skaliert Watson 2.880 POWER7 Kerne in einen Cluster aus 90 IBM Power 750 Servern. UIMA-AS verwaltet die gesamte Kommunikation zwischen Prozessen über den offenen JMS-Standard. Nach der Implementierung von UIMA-AS auf POWER7 war Watson in der Lage, Fragen innerhalb von maximal sechs Sekunden zu beantworten.

Watson verfügt über rund 200 Millionen Seiten Wissen in menschlicher Sprache (entspricht etwa einer Million gelesener Bücher). Watson verarbeitet mithilfe des Apache Hadoop Frameworks große Datenmengen in einem ersten Schritt, um daraus speicherinterne Datensätze zu erzeugen, die in Laufzeit genutzt werden. Die DeepQA UIMA Annotatoren von Watson wurden als Mapper im Hadoop MapReduce Framework eingesetzt. Das Framework verteilte sie über die unterschiedlichen Prozessoren im Cluster. Hadoop sorgt für eine optimale Auslastung der CPU und liefert darüber hinaus wertvolle Tools für die Implementierung, Verwaltung und Überwachung des Datenanalyseprozesses.

Die Power von POWER7

Watson nutzt die massive Parallelverarbeitungsfähigkeit seiner POWER7 Prozessoren, um die mehreren tausend DeepQA Tasks gleichzeitig auf den einzelnen Prozessorkernen auszuführen. Jeder der 90 geclusterten IBM Power 750 Server von Watson enthält 32 POWER7 Kerne mit einer Geschwindigkeit von 3,55 GHz. Die Server arbeiten mit Linux® als Betriebssystem und sind zusammen mit den zugehörigen I/O-Knoten und Kommunikationshubs in 10 Racks angeordnet. Das System verfügt über insgesamt 16 TB Speicher und kann mehr als 80 Teraflops (mehrere Billionen Operationen pro Sekunde) ausführen.

Mit seinem innovativen 8-Core-Prozessor-Design eignet sich POWER7 ideal für die massive Parallelverarbeitung der Analysealgorithmen von Watson. POWER7 verfügt zudem über eine On-the-chip-Kommunikationsbandbreite von 500 Gbit/s, die für eine außergewöhnlich hohe Effizienz von Speicher- und Prozessorauslastung sorgt. Mit jeweils 32 POWER7 Hochleistungskernen und bis zu 512 GB Speicher pro Server liefert Power 750 die notwendige Leistung, um die prozessor- und speicheraufwändigen Java™-Prozesse von Watson zu bearbeiten.

Die Entscheidung, Watson auf handelsüblichen Power 750 Servern aufzubauen, wurde gezielt getroffen, damit optimierte Systeme in Branchen wie dem Gesundheitswesen oder dem Finanzsektor schnell eingeführt werden können. Dieses Ziel machte einen wesentlichen Unterschied zwischen Watson und Deep Blue, einem hoch spezialisierten Superrechner, aus. Deep Blue stützte sich auf die Power Prozessortechnologie einer früheren Generation. Diese Technologie arbeitete mit einem RS/6000 SP System mit 30 Knoten, von denen jeder einen einzelnen 120 MHz POWER2 Prozessor enthielt. Zusätzlich zu den regulären POWER2 Prozessoren erhielt Deep Blue Extraleistung von 480 speziell auf den Schacheinsatz ausgelegten Prozessorchips.

Die von Watson genutzten Power 750 Server hingegen arbeiten heute bereits in vielen tausend Unternehmen in optimierten Systemen, die für komplexe Analytik und Transaktionsverarbeitung zuständig sind. Die Rice University in Houston, Texas, z. B. nutzt IBM Power 750 Systeme, um die Erforschung der Molekularbasis von Krebs durch die Anwendung von Genomanalyseverfahren zu beschleunigen. POWER7 Systeme haben bei Rice für mehr Flexibilität und Effizienz gesorgt. Somit konnte die Hochschule ein breiteres Spektrum an Forschungsaufgaben auf einem einzigen System wahrnehmen, als zuvor möglich war. Auch der kanadische Zollabwickler GHY International entschied sich für den Umstieg auf Power 750 mit Power AIX, Power i und Power Linux, um die zunehmende Aktivität seiner Kunden im internationalen Handelsgeschäft besser unterstützen zu können. Dank PowerVM Virtualisierung ist GHY heute in der Lage, neue Funktionalitäten innerhalb von nur fünf Minuten für veränderte Anforderungen seiner Kunden umzusetzen.

Antworten mit System

Nach vier Jahren intensiver Forschungs- und entwicklungsarbeit durch ein Team aus IBM Wissenschaftlern bewies Watson seine Fähigkeit, gegen Rekordchampions im Fernsehquiz *Jeopardy!* anzutreten. Dabei bewegte sich das System auf demselben Genauigkeits-, Wahrscheinlichkeits- und Geschwindigkeitsniveau eines menschlichen Experten. Das Projekt hat die Gebiete der unstrukturierten Datenanalyse und der natürlichen Sprachverarbeitung sowie die Gestaltung workloadoptimierter Systeme weit vorangebracht. Über *Jeopardy!* hinaus lässt sich die Technologie, die hinter Watson steckt, zur Lösung geschäftlicher und gesellschaftlicher Fragestellungen anwenden – z. B. für Krankheitsdiagnosen, die Beantwortung technischer Fragen online und den Vergleich umfangreicher Rechtsschriften – und so der Fortschritt in vielen Branchen vorantreiben.

Watsons hat die Fähigkeit, die Bedeutung und den Kontext menschlicher Sprache zu verstehen. Indem das System sprachlich transportiere Informationen schnell verarbeitet, ist es in der Lage, komplexe Fragen nahezu umgehend zu beantworten. Damit verändert sich die Art und Weise, in der Computer künftig Menschen unterstützen grundlegend – ganz gleich, ob es sich dabei um geschäftliche oder private Aufgabenstellungen handelt.

Für weitere Informationen:

Weitere Informationen zu Watson, POWER7 und workloadoptimierten Systemen erhalten Sie von Ihrem IBM Vertriebsbeauftragten oder IBM Business Partner. Oder besuchen Sie die folgenden Webseiten:

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

IBM erteilt keine Rechts-, Rechnungsführungs- oder Auditberatung oder sichert zu oder garantiert, dass seine Produkte oder Leistungsangebote zwangsläufig den jeweiligen gesetzlichen Bestimmungen entsprechen. Für die Einhaltung der entsprechenden Gesetze und Bestimmungen, einschließlich nationaler Gesetze und Bestimmungen, sind die Kunden selbst verantwortlich.

IBM Deutschland
IBM-Allee 1
71139 Ehningen
ibm.com/de

IBM Österreich
Obere Donaustraße 95
1020 Wien
ibm.com/at

IBM Schweiz
Vulkanstrasse 106
8010 Zürich
ibm.com/ch

Die IBM Homepage finden Sie unter **ibm.com**

IBM, das IBM Logo, ibm.com, AIX, DEEP BLUE, POWER, POWER2, POWER7, PowerVM und RS/6000 sind Marken oder eingetragene Marken der International Business Machines Corporation in den USA und/oder anderen Ländern. Sind diese und weitere Markennamen von IBM bei ihrem ersten Vorkommen in diesen Informationen mit einem Markensymbol (® oder ™) gekennzeichnet, bedeutet dies, dass IBM zum Zeitpunkt der Veröffentlichung dieser Informationen Inhaber der eingetragenen Marken oder der Benutzungsmarken (Common Law Trademarks) in den USA war. Diese Marken können auch in anderen Ländern eingetragene Marken oder Benutzungsmarken sein.

Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite „ Copyright and trademark information“ unter **ibm.com**/legal/copytrade.shtml

Linux ist eine eingetragene Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Java und alle Java-basierten Marken und Logos sind Marken oder eingetragene Marken von Oracle und/oder ihrer Tochtergesellschaften.

Marken-, Produkt- und Servicebezeichnungen anderer Unternehmen oder Hersteller werden anerkannt.

Hinweise auf Produkte, Programme oder Dienstleistungen von IBM in dieser Veröffentlichung bedeuten nicht, dass IBM diese in allen Ländern anbietet, in denen IBM vertreten ist.

Der Hinweis auf Produkte, Programme oder Dienstleistungen von IBM bedeutet nicht, dass nur Produkte, Programme oder Dienstleistungen von IBM verwendet werden können. Funktional gleichwertige Produkte, Programme oder Dienstleistungen können alternativ verwendet werden.

IBM Hardwareprodukte werden fabrikneu hergestellt. Sie können neben neuen auch wiederverwendete Teile enthalten. Unabhängig davon gelten in jedem Fall die IBM Gewährleistungsbedingungen.

Diese Veröffentlichung dient nur der allgemeinen Information. Änderungen vorbehalten. Aktuelle Informationen zu IBM Produkten und Dienstleistungen erhalten Sie bei Ihrem zuständigen IBM Vertriebspartner oder Reseller.

Fotos zeigen möglicherweise Konzeptstudien.

© Copyright IBM Corporation 2011
Alle Rechte vorbehalten.

Bitte der Wiederverwertung zuführen

POW03061-DEDE-00