



A Toshiba Group Company

White Paper

Unterstützung von Big-Data-Anwendungen mit Flash-basierten Speicherlösungen

Teil 1: Welche Möglichkeiten bieten Speicher im Blick auf Big Data?

Scott Harlin

Published June 2014

OCZ Storage Solutions, Inc. – A Toshiba Group Company

| | | |
|---|--|---|
| 1 | Einleitung | 2 |
| 2 | Big Data-Übersicht | 3 |
| 3 | Big Data-Anwendungsbeispiele | 5 |
| 4 | Big Data-Analyse | 6 |
| 5 | Vorteile von Flash-basierten SSDs für Big Data | 8 |
| 6 | Zusammenfassung | 9 |



1 Einleitung

IT-Manger entdecken derzeit die diversen Möglichkeiten, um Flash-basierte Speichermedien im Unternehmensumfeld zu implementieren, wobei besonders das Anwendungsgebiet rund um das Thema Big Data erwähnenswert ist. Dabei handelt es sich um große Mengen strukturierter und unstrukturierter Daten, die sich schnell bewegen und deren Anforderungen die Möglichkeiten traditioneller Datenbanken und Software-Techniken überschreiten. Im Grunde genommen handelt es sich um umfangreiche Datenmengen, die konventioneller oder digitaler Natur sind. Erzeugt werden diese Daten unternehmensintern und -extern, wobei das Gesamtvolumen stetig wächst. Diese erzeugten Terabytes oder sogar Petabytes sind typisch für diese Anwendungen.

Betrachtet man Big Data aus einer analytischen Perspektive, dann lassen sich aus diesen umfangreichen Datensätzen Datenmuster und Zusammenhänge ableiten, die auf den ersten Blick nicht sichtbar sind. Daraus entstehen Wettbewerbsvorteile, indem fundierte Geschäftsentscheidungen getroffen werden können, was wiederum ein effizientes Marketing und höhere Einnahmen zur Folge hat. Einem Bericht des „McKinsey Global Institute“ (MGI) zur Folge, kann Big Data auf fünf unterschiedlichen Wegen zu Wertsteigerungen im Unternehmen führen³:

- Information wird transparent und vielseitiger nutzbar
- Das Sammeln von genaueren und detaillierteren Leistungsinformationen wird ermöglicht, um die Variablen zu ermitteln und die Leistung zu erhöhen
- Kundenstämme können detaillierter segmentiert werden, sodass Produkte exakter auf deren Bedürfnisse zugeschnitten werden können
- Es können differenzierte Geschäftsentscheidungen, basierend auf einem gesteigerten Verständnis gefällt werden
- Die Qualität von Produkten und Services der nächsten Generation wird gesteigert



Nach Erwartungen von IDC wird der Markt mit Technologien und Servicedienstleistungen rund um Big Data von 9.8 Milliarden US-Dollar im Jahr 2012 auf 32.4 Milliarden US-Dollar im Jahr 2017 anwachsen, was einem durchschnittlichen Wachstum von 27 Prozent pro Jahr

3. Services: Beinhaltet Unternehmensberatung, Integrationservice, Speicherservice, Sicherheitservice, Hardware- und Softwaresupport, Training, Outsourcing sowie jegliche Dienstleistungen im Zusammenhang mit Big Data-Implementierungen;

Nach Erwartungen von IDC wird der Markt mit Technologien und Servicedienstleistungen rund um Big Data von 9.8 Milliarden US-Dollar im Jahr 2012 auf 32.4 Milliarden US-Dollar im Jahr 2017 anwachsen. Das jährliche Wachstum soll durchschnittlich 27 Prozent betragen, damit das Sechsfache im Vergleich zum übrigen Informationstechnologiemarkt.

Definiert man den Begriff Big Data, ist es wichtig die Unterschiede zwischen unstrukturierten und multi-strukturierten Daten zu verstehen, aus denen Informationen zusammengesetzt sind⁵:

- - **Unstrukturierte Daten:** Informationen, die nicht sortiert sind und dementsprechend nicht problemlos von herkömmlichen Datenbanken interpretiert werden können. Das sind typischerweise textintensive Metadaten, wie z.B. Twitter-Tweets oder Posts in sozialen Medien. Der Großteil unstrukturierter Daten befindet sich in Textdateien, die mindestens 80 Prozent des Datenaufkommens in einem Unternehmen verursachen. Indem diese beachtliche Datenmenge ungenutzt in Speichersystemen verbleibt, steigen die Infrastrukturkosten und Potential wird verschwendet.
- **Multistrukturierte Daten:** Eine Vielfalt an Datenformaten und Datentypen, die üblicherweise aus der Interaktion von Personen oder Computersystemen resultieren. Darunter können sich beispielsweise Web-Anwendungen/Transaktionen oder soziale Netzwerke befinden. Ein gutes Beispiel für solche Daten sind Web-Log-Files, die eine Kombination aus Text und Bild enthalten.

Diese Daten beinhalten variable Muster und Informationen, die in der Vergangenheit unentdeckt blieben, da der damit verbundene Zeit- und Kostenaufwand für die Analyse in keinem Verhältnis stand. In den Rechenzentren von heute ist Big Data bereits Realität. Es gibt nun auch kosteneffiziente Ansätze, mit denen man die diversen Herausforderungen von Big Data in den Griff bekommt. Dabei handelt es sich um die folgenden drei Begriffe⁶:

- **Volume:** Steht für die Datenmenge
- **Geschwindigkeit:** Beschreibt wie schnell Daten verarbeitet werden können
- **Variabilität:** Steht für die Unterschiedlichkeit der vorkommenden Datentypen und -quellen

3 Big Data Use Case Examples



Anwendungsbeispiele Bei Big Data kann es sich um Chats in sozialen Netzwerken, Log-Files von Webservern, Verkehrsüberwachungssensoren, Satellitenbilder, Übertragung von Streams (Audio, Video oder eine Kombination von beidem), Finanztransaktionen, Internet-Downloads, Empfang von Dokumenten, Dokumentenscans, GPS-Tracking, Telemetrie-Daten im Automobilsektor, Marktdaten oder so gut wie alle anderen Datenarten handeln, die schnell generiert werden und höchste Anforderungen an aktuelle Datenverarbeitungssysteme stellen. Im Folgenden einige Big Data-Beispiele und wie wir im Alltag damit konfrontiert werden⁷:

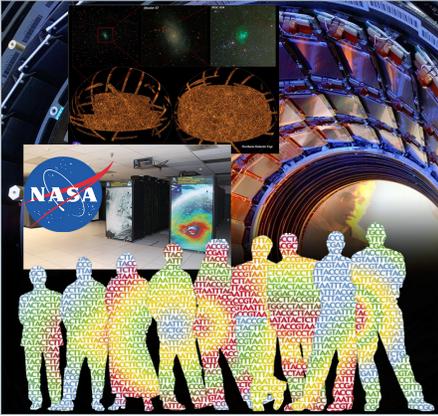
Privatsektor



- Facebook verwaltet täglich über 50 Milliarden Bilder von seinen Benutzern
- Walmart verarbeitet stündlich mehr als eine Million Kundentransaktionen
- Amazon bewältigt täglich Millionen von Back-End-Operationen sowie mehr als eine halbe Million Anfragen von Dritten
- eBay verwendet zwei „Datawarehouses“, die Datenmengen von 7.5 bis 40 Petabyte verarbeiten

Regierungssektor

- Die Big Data-Forschungs- und Entwicklungs-Initiative, welche 2012



- von der Obama-Administration ins Leben gerufen wurde, ermittelt die Verwendung von Big Data bei Problemstellungen auf Regierungsebene
- Big Data-Analysen spielten bei der erfolgreichen Wiederwahl von Barack Obama 2012 eine wichtige Rolle
- Die US-Regierung besitzt sechs der weltweit zehn leistungsstärksten Supercomputer
- Die National Security Agency (NSA) entwickelt Rechenzentren, um die umfangreichen Datenmengen zu verarbeiten, die aus dem Internet gewonnen werden

Sektor Wissenschaft und Forschung

- Die leistungsstärksten Teilchenbeschleuniger erlauben die Überwachung von bis zu 150 Millionen Sensoren. Basierend auf der Big Data Analyse dieser Daten werden Theorien oder Vorhersagen erstellt, die im Zusammenhang mit der Teilchen- und Hochenergie-Physik stehen (als Teil des Large-Hadron-Collider-Projekts). Dabei werden Daten bis zu 40 Millionen Mal pro Sekunde übertragen.
- Die ‚Sloan Digital Sky‘-Umfrage (SDSS) sammelt Astronomie-Daten, wobei in einer Nacht 200 Gigabyte neuer Daten erzeugt werden. Seit dem Beginn im Jahr 2000 häuften sich um die 140 Terabyte Daten an.
- Die Dekodierung des menschlichen Erbguts nahm 10 Jahre in Anspruch und kann mittlerweile mit Big-Data-Verarbeitung innerhalb von weniger als eine Woche durchgeführt werden.
- Das NASA-Zentrum für Klimasimulationen (NCCS) speichert 32 Petabyte an Beobachtungs- und Simulationsdaten.

4 Big Data Analysen

Fortschritte bei der Analyse von Big Data bringen alternative Möglichkeiten mit sich, um Entscheidungsprozesse für wichtige Entwicklungen, die unser jetziges und zukünftiges, tägliches Leben beeinflussen, zu verbessern. In seinem Artikel „The Awesome Ways Big Data is used today to change our world“ präsentiert Bernard Marr eine Vielfalt von Einsatzmöglichkeiten für Big-Data-Analysen.

- **Verstehen und ansprechen von Kunden:** Big Data-Analysen werden dazu verwendet, Kunden und deren Verhalten besser zu verstehen. Zudem lässt sich das Kaufverhalten analysieren. Kreative Modelle helfen bei Vorhersagen. Bei diesem Ansatz handelt es sich um die am weitesten verbreitete Methode, umfangreiche Datensätze gewinnbringend auszuwerten.



10 Einsatzmöglichkeiten für Big Data-Analysen:

- *Verstehen und ansprechen von Kunden*
 - *Optimieren von Geschäftsprozessen*
 - *Quantifizierung individueller Ziele*
 - *Verbesserung des Gesundheitswesens*
 - *Steigerung der sportlichen Leistung*
 - *Verbesserung von Wissenschaft und Forschung*
 - *Optimierung der Leistung von Maschinen*
 - *Verbesserung der Sicherheit*
 - *Verbesserung von Städten/Ländern*
 - *Optimierung von Finanztransaktionen*
- **Verstehen und verbessern von Geschäftsprozessen:** Big Data-Analysen optimieren Geschäftsprozesse, wie z.B. die Überwachung von Lagerbeständen eines Versandhauses, basierend auf Vorhersagen. Durch die Verwendung von Verkehrsdaten in Echtzeit lassen sich Auslieferungsrouten direkter gestalten und Verzögerungen minimieren.
 - **Optimierung und Quantifizierung individueller Leistung:** Big Data-Analysen optimieren und quantifizieren Personalanforderungen. Sie können Daten von tragbaren Geräten wie Smartwatches oder Armbändern zu sammeln und zu verarbeiten. Anhand der Analyseprozesse kann beispielsweise der tägliche Kalorienumsatz eines Trägers, dessen Bewegungsgrad, Schlafmuster und vieles mehr überwacht und/oder ermittelt werden. Darüber hinaus verwenden die meisten Dating-Webseiten Big Data-Analyse-Verfahren, um passende Partner zu evaluieren.
 - **Verbesserung der Gesundheit und des Gesundheitswesens:** Big Data-Analysen können gesamte DNA-Stränge innerhalb weniger Minuten dekodieren und neue Heilmittel für bislang unheilbare Krankheiten erforschen. Auch diverse, mit Krankheiten im Zusammenhang stehende Muster können erkannt werden. Durch die Überwachung der Herz- und Atemfrequenzmustern eines frühgeborenen Babys, lassen sich physische Symptome vorhersagen und präventiv bekämpfen, bevor diese sich ausbilden. Aber auch Grippe-Epidemien können in Echtzeit überwacht werden.
 - **Steigerung sportlicher Leistung:** Setzt man auf Video-Aufnahmen, lässt sich mittels Big Data-Analysen die Leistung von Athleten verfolgen. Das hilft dabei Trends aufzuzeigen und die Konstanz zu fördern. Sensoren in Ausstattungsgegenständen von Sportlern (Bällen, Golfschlägern, Tennisschlägern, ...) bieten eine zusätzliche Dimension hinsichtlich Optimierungsprozessen. Viele der besten Sportteams weltweit verwenden Big Data-Analysen, um beispielsweise die Ernährung zu überwachen.
 - **Verbesserung von Wissenschaft und Forschung:** Beim bereits angesprochenen ‚Large-Hadron-Collider-Projekt‘ werden Big Data-Analysen verwendet, um Wissenschaft und Forschung zu verbessern. Am CERN-Labor für Nuklearphysik in der Schweiz werden zur Analyse von Big Data 65.000 Prozessoren in Rechenzentren verwendet, um 30 Petabyte an gespeicherten Daten zu analysieren. Diese befinden sich auf tausenden unterschiedlichen Computern, an weltweit 150 verschiedenen Orten. Ein weiteres Anwendungsbeispiel beschäftigt sich mit der Analyse von



Big-Data-Anwendungen verursachen gemischte Lasten, mit gleichermaßen Schreib- und Leseprozessen. Voraussetzung für bestmögliche Leistung sind Speichersysteme mit geringen Latenzzeiten und hoher IOPS-Leistung. Enterprise-SSDs sind hier wesentlich besser

seismischen Daten, die anschließend dazu führen, einen geeigneten Ort für Öl- oder Gasförderung zu entdecken.

- **Optimierung der Leistung von Maschinen und Geräten:** Maschinen und Geräte werden derzeit immer intelligenter und operieren vermehrt unabhängig von Big Data-Analysen. Beispielsweise kann Googles selbstfahrender Toyota Prius so sicher wie ein Mensch navigieren, Dank Kamera, GPS und Computer
- **Verbesserung der Sicherheit:** Die „National Security Agency“ (NSA) verwendet Big Data-Analysen um Cyber-Attacken aufzuspüren und terroristische Aktivitäten aufzudecken, bevor diese überhaupt stattfinden können. Anhand dieser Tools lassen sich kriminelle Aktivitäten nicht nur vorhersagen, sondern Kriminelle können anhand derartiger Daten auch gefasst werden. Kreditkartenunternehmen setzen bei der Bekämpfung von Betrug auf Big Data-Analysen.
- **Verbesserung und Optimierung von Städten und Ländern:** Die Optimierung des Verkehrsflusses ist ein gutes Beispiel dafür, wie Städte von Big Data-Analysen profitieren. Basierend auf Echtzeit-, Verkehrs- und Wetterdaten und anhand von Posts in sozialen Netzwerken, lassen sich dahingehend Muster erkennen. Durch die Vernetzung öffentlicher Infrastruktur, kann die Verspätung eines Busses z.B. dazu führen, dass der Zug später losfährt.
- **Optimierung von Finanztransaktionen:** Der Großteil aller Aktienhandel wird basierend auf Big Data-Analysen durchgeführt, die soziale Netzwerke und News-Webseiten überwachen, sodass informierte Kaufs- oder Verkaufsentscheidungen gefällt werden können.

Diese Kategorien zeigen Anwendungsgebiete, in denen Big Data-Analysen bereits heute eingesetzt werden. Den möglichen Anwendungsszenarien von Big Data-Analysen setzt derzeit höchstens die menschliche Vorstellungskraft Grenzen, sodass noch viel Potenzial ungenutzt ist und abgeschöpft werden kann.

5 Vorteile von Flash-basierten SSDs für Big Data

Um Big Data gewinnbringend zu verarbeiten und die Rentabilität (ROI) zu maximieren, suchen IT-Abteilungen in Unternehmen derzeit nach alternativen Wegen die umfangreichen Datensätze zu analysieren und zu

verarbeiten. Big Data stellt Unternehmen und IT-Infrastrukturen von heute vor große Herausforderungen, da die damit verbundenen Anwendungen eine ultra-schnelle Datenverarbeitung erfordern.

Big Data-Anwendungen verursachen gemischte Lastmuster, wobei Lese- und Schreibprozesse simultan ausgeführt werden. Das erfordert wiederum eine hohe IOPS-Leistung (Input/Output-Operationen pro Sekunde) und minimale Latenzzeiten. Derartige Zugriffsmuster können von herkömmlichen Festplatten nur ineffizient verarbeitet werden. Setzt man aber auf Enterprise Solid State Drives (SSDs), erhält man einen leistungsfähigen Speicher, dem diese Anforderungen auf den Leib geschnitten sind. Herkömmliche Festplatten sind schlicht und ergreifend nicht in der Lage, mit den Anforderungen von Big Data-Anwendungen Schritt zu halten. Während bereits ein einfacher Server Hunderttausende IOPS verarbeiten könnte, vermag eine einzelne Festplatte lediglich zwischen 100 und 300 IOPS zu liefern. Hier sieht man klar das Leistungsungleichgewicht. Immer wenn eine Festplatte Daten liefern muss, wird der Schreib/Lesekopf physisch bewegt, wodurch die Zugriffszeit mechanisch begrenzt wird.

Festplatten können konstante Datenströme effizient verarbeiten, also sequenzielle Schreib- und Lesevorgänge bei anspruchsvollen Geschwindigkeiten bewältigen. Moderne Betriebssysteme können dank Mehrprozessorarchitekturen, zufällig anfallende Daten schnell verarbeiten. Herkömmliche Festplatten sind dieser Entwicklung nicht gewachsen. Im direkten Vergleich mit Festplatten sind die NAND-Flash-Zellen von SSDs wesentlich dichter gepackt. Zudem kommen auch keine rotierenden Platten oder Magnetköpfe zum Einsatz, die Daten mechanisch lesen und schreiben. Es gibt lediglich einen Controller, der über den Verbleib von Daten stets informiert ist. Daraus ergibt sich eine gesteigerte Schreib- und Leseleistung. Da keine beweglichen Teile verwendet werden, sinkt auch die Ausfallwahrscheinlichkeit, die I/O-Leistung steigt drastisch und die Latenzzeiten sinken erheblich.

Flash-basierte SSDs erfreuen sich im Big Data-Umfeld wachsender Beliebtheit, da sie gegenüber Festplatten eine stark gesteigerte IOPS-Leistung bieten. Darüber hinaus unterstützen die Laufwerke auch hohe Kapazitäten, brauchen weniger Strom und sind mit diversen Formfaktoren sowie Schnittstellen verfügbar.

Kontaktieren Sie uns für weitere Informationen

OCZ Storage Solutions
6373 San Ignacio Avenue
San Jose, CA 95119 USA

P Vertrieb Deutschland
E Vertrieb_Deutschland@ocz.com
W ocz.com/enterprise

[EMAIL SALES TEAM >](#)

[VISIT OCZ ENTERPRISE >](#)

6 Zusammenfassung

Vertraut man auf Vorhersagen führender Marktforschungsinstitute, welche die Entwicklungen in der Speicherindustrie verfolgen, wird das Marktvolumen von Big Data in den kommenden fünf Jahren um mehrere Milliarden US-Dollar anwachsen. Big Data setzt auf eine neue Generation von Technologien und Architekturen, die darauf ausgelegt sind, umfangreiche Datensätze gewinnbringend zu extrahieren.

Flash-basierte SSDs erfreuen sich bereits einer breiten Akzeptanz, da sie den Anforderungen von Big Data-Anwendungen gewachsen sind und die notwendige I/O-Leistung bieten, die herkömmliche Festplatten schlicht überfordert. Zudem sind die Laufwerke mit unterschiedlichen Formfaktoren und Schnittstellen erhältlich. Ein weiteres Plus: sie konsumieren weniger Strom. Big Data macht eine ultra-schnelle Datenverarbeitung erforderlich, die datengetriebene Analysen beschleunigt. Durch das Vorkommen von strukturierten als auch unstrukturierten Daten, die zu groß sind, sich zu schnell bewegen oder aktuelle Verarbeitungskapazitäten überfordern, profitieren Anwendungen von Werkzeugen zur Überwachung von Datenaktivitäten und Flash-Ressourcen.

OCZ bietet ein komplettes Portfolio an SSD-Hardware und Speicherlösungen, speziell auf die Verwendung mit Big Data ausgelegt. Im zweiten White Paper „Auslieferung von Performance und Management-Lösungen, die für Big Data Anwendungen benötigt werden“ wird dies thematisiert.

Referenzen und Quellen aus der Industrie

- 1 „What is Big Data“ – eine Definition von Webopedia.com
- 2 “What is Big Data”, von Lisa Arthur, Forbes Magazin, 15. August, 2013
- 3 “Big Data: The next frontier for innovation and productivity”, von James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, McKinsey Global Institute (MGI), McKinsey & Company, Mai 2011.
- 4 “Worldwide Big Data Technology and Services 2013-2017 Forecast”, von Dan Vessel, Rob Brothers, Steve Conway, Matthew Eastwood, John Grady, Brian McDonough, Henry D. Morris, David Schubmehl, Mary Johnston Turner, Melissa Webster, Ashish Nadkarni, Christian Christiansen, Mukesh Dialani, Maureen Fleming, Tim Grieser, Rohit Mehra, Carl W. Olofson, Kuba Stolarski, Mary Wardley und Ali Zaidi, IDC, Report #244979, Dezember 2013.
- 5 “What is Big Data,” von Lisa Arthur, Forbes Magazine, 15. August, 2013.
- 6 “What is Big Data,” von Edd Dumbil, Strata-O’Reilly Media, 11. January, 2012.
- 7 “What is Big Data,” – eine Definition from Wikipedia.com.
- 8 “The Awesome Ways Big Data Is Used Today To Change Our World,” von Bernard Marr, LinkedIn.com, November 13, 2013.

Haftungsausschluss

OCZ kann jederzeit Änderungen an den Spezifikationen und Produktbeschreibungen vornehmen, ohne darüber im Vorhinein zu informieren. Die Informationen, die in diesem Dokument präsentiert werden, sind ausschließlich zu Informationszwecken gedacht und können Ungenauigkeiten, Weglassungen und typographische Fehler enthalten. Alle Leistungswerte und Bewertungen wurden mit Systemen gemessen, die ungefähr die Performance von den in den Tests verwendeten OCZ-Produkten wiedergeben. Jegliche Unterschiede bei Software- oder Hardware-Konfigurationen haben einen Einfluss auf die erzielte Leistung und OCZ kontrolliert weder das Design noch die Implementierung von Benchmarks oder Webseiten Dritter, auf die aus diesem Dokument verwiesen wird. Die in diesem Dokument enthaltenen Informationen unterliegen Veränderungen und können aus unterschiedlichen Gründen an Genauigkeit verlieren. Unter diesen Gründen können sich beispielsweise Anpassungen an Roadmaps und/oder Revisionen an Hardware befinden sowie die Einführung neuer Modelle, Software-Anpassungen, Firmware-Änderungen oder Vergleichbares. OCZ ist nicht verpflichtet Aktualisierungen an diesen Informationen durchzuführen, diese zu korrigieren oder zu revidieren.

OCZ MACHT MIT DEN IN DIESEM DOKUMENT ENTHALTENEN INFORMATIONEN KEINERLEI ANGABEN ZU GARANTIE UND NIMMT AUCH KEINE STELLUNG ZU BESTIMMTEN PUNKTEN. DARÜBER HINAUS ÜBERNIMMT MAN KEINERLEI VERANTWORTUNG FÜR INAKURATE INFORMATIONEN, FEHLER ODER UNGENAUIGKEITEN, DIE IN DIESEN INFORMATIONEN ENTHALTEN SIND. .

OCZ LEHNT JEDLICHE HAFTUNG VON IMPLIZIERTER GARANTIE ODER MÄNGELGEWÄHRLEISTUNG AB. UNTER KEINEN UMSTÄNDEN KANN OCZ FÜR INAKURATE INFORMATIONEN, FEHLER ODER UNGENAUIGKEITEN ZUR RECHENSCHAFT GEZOGEN WERDEN, DIE IN DIREKTEM ODER INDIREKTEM ZUSMMENHANG MIT DEN IN DIESEM DOKUMENT ENTHALTENEN INFORMATIONEN GEBRACHT WERDEN, SELBST WENN OCZ AUF DIE MÖGLICHKEIT SOLCHER SCHÄDEN HINGEWIESEN WURDE.

ZUWENDUNG

© 2014 OCZ Storage Solutions, Inc. – A Toshiba Group Company. Alle Rechte vorbehalten

OCZ, das OCZ Logo, OCZ XXXX, OCZ XXXXX, [Produktname] und Kombinationen davon sind eingetragene Markenzeichen von OCZ Storage Solutions, Inc. – A Toshiba Group Company. Alle anderen Produktnamen und Logos sind lediglich zu Referenzzwecken eingebunden und sind Warenzeichen der entsprechenden Eigentümer.