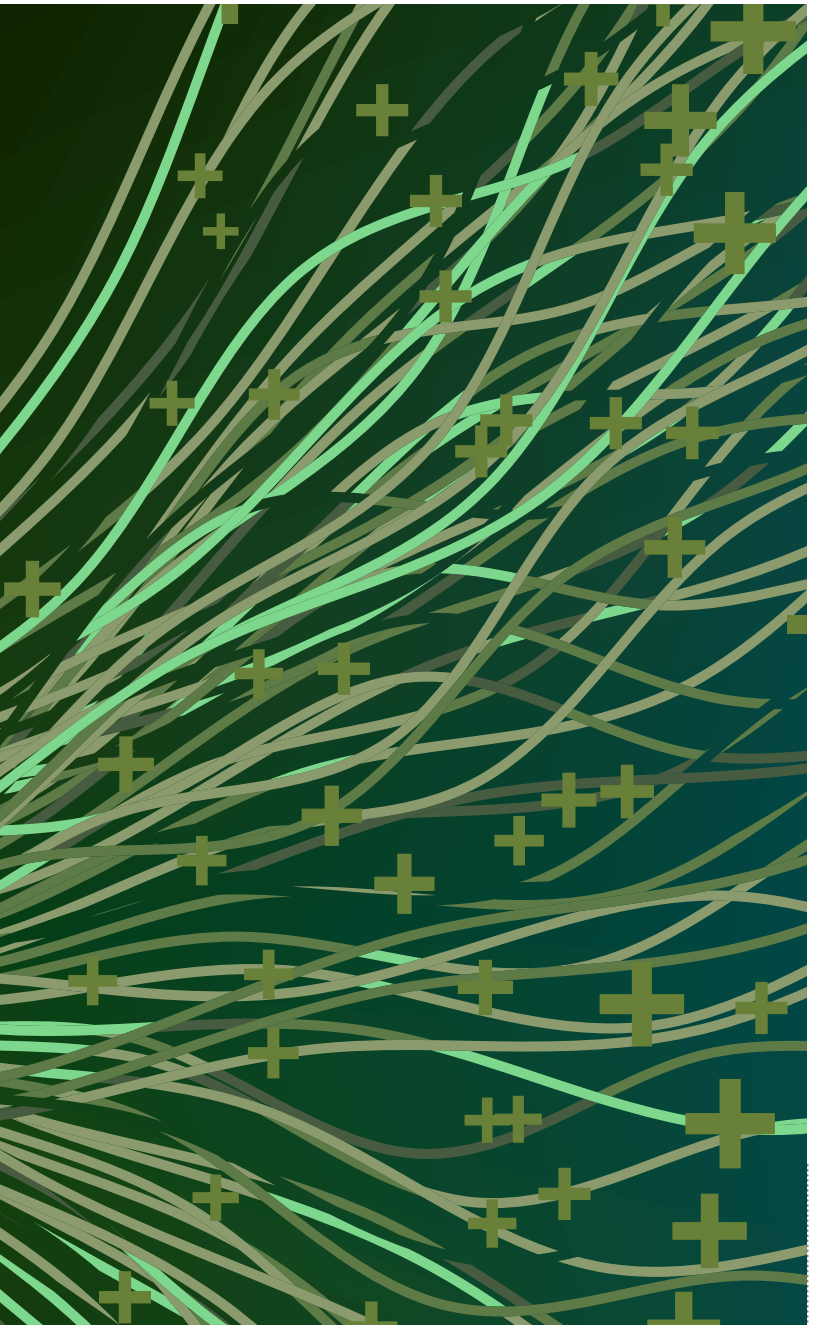

Special Report

TOP-TIPPS FÜR OPTIMIERTEN DATENSCHUTZ BEI BIG-DATA-PROJEKTEN

TechTarget | SearchSecurity.de

DER SCHUTZ PERSONENBEZOGENER Daten steht für viele Unternehmen im Mittelpunkt. Dies erfordert allerdings die notwendige Transparenz in der Datenverarbeitung, welche die Grundlage für modernen Datenschutz bildet. Die meisten Unternehmen können besonders bei Big-Data-Projekten den Schutz der Daten nicht immer gewährleisten. Dieser E-Guide erläutert Tipps für optimierten Datenschutz bei Big-Data-Projekten und erklärt, wie Sie diesen maximieren können.

DATENSCHUTZ BEI BIG-DATA-PROJEKTEN: TIPPS FÜR UNTERNEHMEN

Bereits Ende 2012 hat die Federal Trade Commission (FTC) insgesamt neun Datenhändlern um Informationen gebeten, wie sie die Daten ihrer Nutzer erheben, sammeln und nutzen. Diese Aktion des FTC war ein klares Signal, dass die Nutzung und Analyse von großen Datenmengen (Big Data) zwar vielversprechende Geschäftsvorteile bietet, aber auch einige Fragen in Bezug auf Datenschutz aufwirft.

WARUM BIG DATA?

Big Data unterscheidet sich vom bislang üblichen Data Warehousing, da damit Analysen auf Basis jeder Art von Datenformaten oder Dateien möglich sind, auch mit Bildern, Videos und Daten aus sozialen Medien. Big Data stützt sich nicht mehr auf eine „1:1“ -Beziehung zwischen Server und Datenspeicher, sondern auf eine Virtualisierungs-Architektur. Diese ist notwendig, da die zu analysierenden Inhalte aus großen Datenbanken und Archiven stammen.

Führungskräfte und Abteilungsleiter entscheiden sich für die Implementierung von Big-Data-Lösungen, um genauere, detaillierte Prognosen oder Vorhersagen

treffen zu können, die mit hoher Wahrscheinlichkeit Vorteile für das eigene Unternehmen bieten. Firmen können in verschiedenen Bereichen von Big Data profitieren, sei es bei der Entwicklung neuer, verbesserter Produkte, der optimalen Preisgestaltung, der Auswahl von Bewerbern oder der Entwicklung wirksamer Marketing-Kampagnen.

Auch die Politik nutzt die Vorteile von Big-Data-Analysen: Barack Obama setzte 2012 während seines Wahlkampfes gezielt auf die Big-Data-Analyse, nicht nur um potenzielle Wähler zu erkennen und diese zu beeinflussen, sondern auch um Spenden für die Finanzierung seiner Kampagne zu generieren und die Wähler dazu zu bewegen, ihre Stimme tatsächlich abzugeben. Diese Strategie erwies sich letztendlich als Schlüssel für den Wahlerfolg.

BIG DATA UND DATENSCHUTZ

Die anfangs beschriebene Aktion der FTC richtete sich speziell an Datenhändler. Diese Unternehmen sammeln und analysieren Daten rund um das Verhalten von Verbrauchern und verkaufen die Ergebnisse an andere Unternehmen. Diese nutzen die erworbenen Daten, um ihre Marketing- und Vertriebsaktivitäten zu verbessern. Wachsende Bedenken hinsichtlich der Privatsphäre und des Datenschutzes

bei Big-Data-Projekten beschränken sich aber nicht nur auf diese klassischen Datenhändler.

Die Economist Intelligence Unit, ein unabhängiges Unternehmen innerhalb der Economist Group, hat bereits vor einiger Zeit eine Studie zu den führenden Unternehmen bei der Nutzung von Big Data veröffentlicht. Die Untersuchung umfasst 19 Branchen wie Fertigung, IT und Technologie, Finanzdienstleistungen, Gesundheitswesen, Pharmazie, Biotechnologie und Konsumgüter. Es besteht kein Zweifel: Die Big-Data-Revolution hat begonnen.

Angesichts der Merkmale von Big Data und der Art und Weise der Nutzung gehören die Qualität und Genauigkeit der Daten zu den wichtigsten Aspekten des Datenschutzes. Schließlich kann die Datenqualität Entscheidungen einer Person negativ beeinflussen. Wie genau sind zum Beispiel personenbezogene Daten aus Social Medien? Kann man Informationen aus Social-Media-Kanälen oder anderen Quellen aus dem Internet nutzen, um Bewerbungen zu filtern und zu bewerten oder den Preis für eine Krankenversicherungs-Police zu erhöhen?

Grundlegende Daten zu Personen wie Alter, Familienstand, Bildungsabschluss oder Arbeitsstelle werden in der Regel nicht verifiziert. Gleiches gilt für kostenlose E-Mail-Services, bei denen sich der Kontoinhaber durch die Annahme der

Allgemeinen Geschäftsbedingungen dazu verpflichtet, ein gewisses Maß an Privatsphäre aufzugeben. Auch diese persönlichen Daten werden meist unzureichend geprüft.

Die Qualität wird außerdem beeinträchtigt, wenn Suchbegriffe oder Suchphrasen im Internet falsch interpretiert werden. Beispiele für den unpassenden Einsatz von Big Data können Suchbegriffe sein, welche die Preisgestaltung bewerten oder vielleicht gezielt potenzielle Kunden ansprechen sollen. So können beispielsweise mehrere Bewohner eines Haushalts einen einzigen Computer benutzen, und es gibt viele Gründe, warum jemand im Internet sich über ein Thema informiert, das für ihn oder sie nicht direkt relevant ist.

Diese Art der Datensammlung, Analyse und Nutzung kann zu fehlerhaften Ergebnissen und schlechten Entscheidungen führen, sprich einem doppelten Verlustszenario für Privatpersonen und Unternehmen, die auf Basis dieser Daten handeln. Dieser Mangel bei der Qualitätskontrolle von Big-Data-Informationen verweist auf ein anderes etabliertes Prinzip des Datenschutzes: die konsistente und sachgemäße Sammlung von personenbezogenen Daten im Sinne des vorgesehenen Zwecks.

BEST PRACTICES FÜR DATENSCHUTZ BEI BIG-DATA-PROJEKTEN

Noch gibt es keine umfassenden Best Practices für die Arbeit mit Big-Data-Lösungen, die zu Innovationen führen können ohne dabei den Schutz persönlicher Daten zu opfern. Aus den bisherigen Erfahrungen lassen sich aber bereits einige grundsätzliche Erkenntnisse ziehen.

Der erste Schritt zur effektiven Nutzung von Big Data sind ausgeprägte Kompetenzen bei der Beschaffung und Verwaltung von Cloud-Services, die als Voraussetzung für den kosteneffizienten Betrieb von Big-Data-Lösungen gelten: Die meisten Unternehmen können oder wollen nicht in den Aufbau einer eigenen IT-Infrastruktur investieren, die für den Betrieb einer Big-Data-Lösung erforderlich ist.

Sie setzen stattdessen auf Cloud-basierte Anwendungen, Infrastruktur und Rechenleistung. Selbst diejenigen, die eine eigene kostspielige Infrastruktur aufbauen, kaufen zusätzliche Ressourcen aus der Cloud hinzu, um flexibel zu bleiben und beispielsweise Lastspitzen abfedern zu können.

Doch genau hier liegt ein Schwachpunkt in vielen Unternehmen: Es fehlt in der Regel an Kompetenz, um die Sicherheit und Vertraulichkeit der Daten in der Cloud zu gewährleisten. Denn es reicht nicht aus, im Vertrag allgemeine und Standard-Klauseln rund um Sicherheit zu verankern.

Es muss klar definierte Verantwortlichkeiten sowohl auf Seiten des Anbieters der Cloud-Services und des Nutzers für die erforderlichen spezifischen Datenschutzkontrollen geben. Die Cloud-Services müssen auch permanent überwacht und mit Audits bewertet werden nach relevanten Kennzahlen wie Datenintegrität, Vertraulichkeit und Verfügbarkeit. Eine ausgezeichnete Quelle für den Datenschutz von Cloud-Computing-Services ist die Cloud Security Alliance, die Leitfäden veröffentlicht und auf ihrer Website zur Verfügung stellt.

Die bisherige Erfahrung zeigt, dass es am besten ist, Big-Data-Lösungen als Prototyp zunächst in der Public Cloud zu starten und anschließend in eine Private Cloud zu verschieben. Warum? Eine Public Cloud ist per Definition öffentlich und durch externe Unternehmen, Personen oder sogar wenig vertrauenswürdigen Organisationen zugänglich.

Private-Cloud-Implementierungen werden direkt von einer Organisation oder einem Unternehmen gesteuert und verwaltet, auch wenn sich die IT-Infrastruktur mit den Daten außerhalb der eigenen Geschäftsräume befindet. Zudem können ausschließlich vertrauenswürdige Instanzen auf die Private Cloud zugreifen.

Eine weitere Maßnahme für die effizientere Nutzung von Big Data ist der Einsatz von so genanntem Converged Storage, sprich einer vorkonfigurierte Hardware- und

Software-Lösung in einer einheitlichen Umgebung. Converged Storage ist effizienter und reduziert die Wahrscheinlichkeit von Fehlern, welche die Datenqualität und Genauigkeit der Daten beeinflussen.

Zu nennen ist hier die Daten-Deduplizierung, die zudem auch Kostenvorteile bietet. Deduplizierung sorgt dafür, dass identische Daten nur einmal abgespeichert werden. Die Technologie hilft Anwendern, Datenredundanzen zu erkennen und zu beseitigen. Ziel ist es, das Volumen der zu speichernden Daten zu reduzieren und damit Speicherplatz auf den Storage-Systemen einzusparen.

Zudem ist es wichtig, die Daten zu bereinigen, um die oben beschriebenen Datenschutz-Fragen oder -Probleme zu vermeiden. „Durch Filtern, Bereinigen, Beschneiden, Matching und Verbinden der Daten lassen sich bereits zu sehr frühen Berührungspunkten Diagnosen erstellen“, sagt Amy Dean, Data-Warehouse-Spezialistin an der Emory University in Atlanta.

Dean empfiehlt, die unterschiedlichen Datenquellen nach ihrer Datenqualität zu gewichten und in die Analyse miteinzubeziehen. Sie schlägt zudem vor, die Datenquellen zu verlinken beziehungsweise zu referenzieren, damit bei Bedarf jedes einzelne Datenelement zu seiner Quelle beziehungsweise seinem Ursprungsort zurückverfolgt werden kann.

Das beste Sicherheitsnetz für die Exaktheit und Richtigkeit persönlicher Daten (und damit auch bessere Datenschutz-Maßnahmen) ist es, die Benutzer oder Kunden zu ermutigen, und Ihnen nicht nur vorzuschreiben, dass sie auf alle Informationen, die über sie gesammelt wurden, zugreifen, diese überprüfen und korrigieren können.

Dieser Prozess muss sich einfach abwickeln lassen und für die Benutzer kostenfrei sein. Für viele Early Adopters von Big-Data-Lösungen stellt dies eine Herausforderung dar, weil sie häufig große Mengen an Daten sammeln, die sie noch nie genutzt und ausgewertet haben.

Zudem besteht die Furcht davor, den Kunden zu zeigen, wie umfangreich und ausführlich die Datensammlung ist. Aber dieses Maß an Transparenz ist der beste Weg, um das Vertrauen der Kunden grundsätzlich zu gewinnen sowie das Vertrauen für Entscheidungen, die auf Big Data beruhen.

Institutionen, die sich mit Kreditauskünften befassen, setzen bereits seit längerem Prozesse ein, die den Kunden den Zugriff auf ihre persönlichen Daten erlauben mit der Option, diese zu korrigieren. In den USA ist diese Option für die Branche gesetzlich vorgeschrieben. Auch Datenschutzhinweise auf Websites, die Kontaktdaten für Fragen oder Bedenken enthalten, sind eine weitere bewährte Methode für

höhere Transparenz und den Kampf gegen unkorrekte Daten.

DAS RÄTSEL RUND UM BIG DATA

Das Konzept, für die Erhebung und Nutzung persönlicher Daten die Zustimmung der Nutzer zu verlangen, ist in Unternehmen sehr umstritten. Wenn man die Zeit zurückdrehen und bei Null anfangen könnte, wäre dies eine ideale Grundregel.

Jedoch reicht es heute nicht mehr aus, Einzelpersonen um die Zustimmung für die Sammlung ihrer persönlichen Daten zu bitten, weil die Menge der bereits gesammelten und verbreiteten personenbezogenen Daten einfach zu groß ist. Die harte Wahrheit: Es ist nicht möglich, alle Organisationen zu ermitteln, die Informationen zu einer Person gesammelt haben.

Natürlich können einzelne Menschen wieder die Kontrolle über ihre persönlichen Daten erhalten, wenn sie bei Bedarf oder auf Wunsch ihre Daten löschen oder selbst bestimmen dürfen, welche Informationen über sie gesammelt werden. Es ist unwahrscheinlich, dass die Big-Data-Anbieter diese Funktion anbieten; es wäre zugleich die Feuerprobe dafür, ob die Verbraucher die Nutzung ihrer Daten erlauben, weil sie Vorteile davon erwarten.

Für die Regulierungsbehörden wäre die Möglichkeit, die Daten zu löschen, eine

wesentliche Voraussetzung für den Schutz der Privatsphäre der einzelnen Kunden. Bei Big-Data-Projekten sollte daher jedes Unternehmen künftig bereits beim Design und der Architektur der Lösung die Funktion integrieren, Daten oder einzelne Datenfelder einer individuellen Person löschen zu können.

Eine weitere Option für besseren Datenschutz stellt die Anonymisierung personenbezogener Daten dar. Doch leider hat sich das Konzept der Anonymisierung, das auf der Beseitigung der Personen identifizierenden Feldern und Attributen basiert, als nicht tragfähig erwiesen.

Bereits im Jahr 2000 bewies Latanya Sweeney, mittlerweile Professorin für Government and Technology in Residence an der Harvard University, dass sich 87 Prozent aller Amerikaner mit nur drei Informationen identifizieren lassen: Postleitzahl, Geburtstag und Geschlecht. All diese Daten finden sich in öffentlichen Aufzeichnungen. Die Ergebnisse zeigen, dass selbst mit Einsatz eines Anonymisierungssystems keine Garantie besteht, dass einzelne Personen unerkannt bleiben.

FAZIT

Der gemeinsame Nenner zum Schutz der Privatsphäre in Zeiten von Big Data ist es, persönliche Daten richtig und genau zu erfassen und sie sachgemäß zu

interpretieren. Unternehmen, welche die oben beschriebenen Grundsätze zum Datenschutz bei der Entwicklung ihrer Big-Data-Lösungen berücksichtigen, werden die besten Ergebnisse erzielen und möglicherweise auf den geringsten Widerstand bei ihren Kunden stoßen.

DATENSCHUTZ: BIG-DATA-ANALYSEN ERFORDERN MEHR ALS ANONYMISIERUNG

Big Data gehört zu den IT-Trends, bei denen ein hohes wirtschaftliches Potenzial und gleichzeitig ein großes Bedrohungspotenzial gesehen wird. Kongresse mit dem Thema „Big Data – Goldmine oder Dynamit?“, „Sicherheit von Daten und Identitäten angesichts NSA und Big Data“ oder „Big Data für Bond 2.0“ spiegeln die Sorgen der Datenschützer wieder, wenn große Mengen an Daten gesammelt, gespeichert und ausgewertet werden.

Big Data mache deutlich, dass Daten zunehmend zum Rohstoff des 21. Jahrhunderts werden, an dem viele partizipieren wollen, so der

Landesdatenschutzbeauftragte von Baden-Württemberg. Marktforscher sehen für 2014 einen globalen Umsatz mit Big-Data-Produkten und -Dienstleistungen von rund 73,5 Milliarden Euro, ein Plus von 66 Prozent im Vergleich zum Vorjahr.

Unternehmen sind gefordert, einen Weg zu finden, die wirtschaftlichen Chancen von Big-Data-Analysen für sich zu nutzen und gleichzeitig dem Datenschutz gerecht zu werden. Das ist keinesfalls unmöglich, wenn man die Vorgaben des Datenschutzes konsequent auf Big Data anwendet.

POTENZIALE VON BIG DATA LASSEN SICH NUTZEN

Eine Reihe von Initiativen versuchen Unternehmen zu helfen, Big Data wirtschaftlich zu nutzen, wie der Technologiewettbewerb „Smart Data - Innovationen aus Daten“ des Bundeswirtschaftsministeriums oder der BITKOM-Leitfaden „Big-Data-Technologien - Wissen für Entscheider“. Dabei wird immer auch auf die Bedeutung des Datenschutzes hingewiesen.

Für die datenschutzgerechte Nutzung von Big Data ist es wichtig zu verstehen, dass das Erfolgsrezept von Big Data gerade auf der Nutzung und Verknüpfung möglichst großer und heterogener, ursprünglich für andere Zwecke

erhobener Datenbestände beruht. Die Datennutzung zu anderen Zwecken erfordert aber eine rechtliche Grundlage beziehungsweise die Einwilligung der Betroffenen. Deshalb kommt der Transparenz und Aufklärung der Betroffenen die größte Bedeutung zu, wenn von Unternehmen Big-Data-Analysen geplant werden.

DATENSCHUTZPRINZIPIEN KONSEQUENT AUF BIG DATA ANWENDEN

Neben der Einwilligung der Betroffenen als Legitimierung der Datennutzung und Zweckänderung sollten Unternehmen an die Prinzipien der Erforderlichkeit und Datenvermeidung denken. Auch wenn dies scheinbar ein Widerspruch zur Ansammlung großer Datenmengen zu sein scheint: Weder aus Datenschutzsicht noch aus wirtschaftlicher Sicht ist es sinnvoll, möglichst alle Arten von Daten zu sammeln.

Es gilt dagegen, die Datensammlung und Datenspeicherung auf den Zweck der Big-Data-Analysen auszurichten. Dies erspart unnötige IT-Kosten und ermöglicht die richtige Aufklärung der Betroffenen über die geplante Datenverarbeitung.

Zu der Datenvermeidung und Datensparsamkeit gehört auch die

Anonymisierung und Pseudonymisierung der personenbezogenen Daten. Diese Datenschutzmaßnahmen werden meist zuerst genannt, wenn es um Big-Data-Analysen geht.

Dabei darf aber nicht vergessen werden, dass bei der Verwendung von Pseudonymen immer die Gefahr besteht, dass sich Nutzer in einem zweiten Schritt identifizieren lassen, wenn die Pseudonyme aufgelöst werden. Aus diesem Grund gelten die zuvor genannten Anforderungen des Datenschutzes auch dann, wenn vor den Big-Data-Analysen Verfahren zur Pseudonymisierung eingesetzt werden.

DATENSICHERHEIT BEI BIG DATA IST ANSPRUCHSVOLL

Ein weiterer Punkt, der bei Big-Data-Analysen nicht vergessen werden darf: Die Datenansammlungen müssen auch vor Missbrauch durch unbefugte Dritte geschützt werden. Letztlich wird damit verhindert, dass die Daten jenseits der Einwilligung der Betroffenen ausgewertet und damit ausgenutzt werden.

Die erforderlichen technisch-organisatorischen Maßnahmen (zum Beispiel Verschlüsselung und Backup, Datenlöschung) müssen allerdings erst noch in praktikable, für die Unternehmen handhabbare Lösungen einfließen

und integraler Bestandteil moderner IKT-Systeme werden (Privacy by Design), so das Bundeswirtschaftsministerium anlässlich des eingangs erwähnten Technologiewettbewerbs „Smart Data – Innovationen aus Daten“.

Tatsächlich zeigen verschiedene Studien, dass Lösungen für den technischen Schutz von Big Data noch eher selten auf dem deutschen Markt zu finden sind. Beispiele, wie sich Datenschutz und Big-Data-Analysen praktisch in Einklang bringen lassen, kommen unter anderem von Fraunhofer IAIS, darunter „Privacy-preserving Data Mining in der Medizin“ und „Mobility, Data Mining und Privacy“.

FAZIT

Es zeigt sich, dass Datenschutz bei Big-Data-Analysen zwar aufwändig, aber möglich und zudem lohnend ist. Erst dadurch können die Potenziale von Big Data wirklich genutzt werden.

KOSTENLOSE ONLINE-RESSOURCEN FÜR IT-EXPERTEN

TechTarget publiziert qualifizierte Medieninhalte im IT-Bereich, die Ihren Informationsbedarf bei der Suche nach neuen IT-Produkten und Technologien abdeckt, und Ihr Unternehmen somit gezielt in der Strategieentwicklung unterstützt. Es ist unser Ziel, Ihnen durch die Bereitstellung von Online-Ressourcen über die aktuellsten Themen die Kaufentscheidungen für IT-Produkte zu erleichtern und kostengünstiger zu gestalten.

Unser Netzwerk an Technologie-Webseiten gibt Ihnen die Möglichkeit, auf eine der weltweit größten Online-Bibliotheken zum Thema IT zuzugreifen, und anhand von unabhängigen Expertenmeinungen und Analysen, sowie auch zahlreichen Whitepapern, Webcasts, Podcasts, Videos, virtuellen Messen und Forschungsberichten zu einer ausgewogenen Kaufentscheidung zu gelangen.

Unsere Online-Ressourcen berufen sich auf die umfangreichen Forschungs- und Entwicklungskompetenzen führender Technologieanbieter, und ermöglichen es Ihnen somit, Ihr Unternehmen für künftige Marktentwicklungen und Herausforderungen zu rüsten. Unsere Live-Informationsevents und virtuelle Seminare geben Ihnen die Möglichkeit, Ihre täglichen individuellen Herausforderungen im Bereich IT mit den Experten der Branche zu diskutieren.

Außerdem können Sie in unserem Social Network, dem IT Knowledge Exchange, praxisnahe Erfahrungsberichte mit Fachkollegen und Experten in Echtzeit austauschen.

WAS MACHT TECHTARGET SO EINZIGARTIG?

Bei TechTarget steht die Unternehmens-IT im Mittelpunkt. Unsere Autoren und das Redaktions-Team sowie auch unser großes Netzwerk an Industrieexperten bietet Ihnen Zugriff auf die neuesten Entwicklungen und relevantesten Themen der Branche.

TechTarget liefert klare und überzeugende Inhalte und umsetzbare Informationen für die Profis und Entscheidungsträger der IT-Branche. Wir nutzen die Schnelligkeit und Unmittelbarkeit des Internets, um Ihnen in realen und virtuellen Kommunikationsräumen hervorragende Networking-Möglichkeiten mit Fachkollegen zu ermöglichen.