

September 2011

KURZÜBERSICHT FÜR DAS MANAGEMENT

Die entscheidenden Vorteile von IBM InfoSphere Warehouse



International Technology Group

609 Pacific Avenue, Suite 102
Santa Cruz, California 95060-4406
Telefon: + 831-427-9260
E-Mail: Contact@ITGforInfo.com
Website: ITGforInfo.com

INHALTSVERZEICHNIS

| | |
|--|-----------|
| Die entscheidenden Vorteile von IBM InfoSphere Warehouse | I |
| INHALTSVERZEICHNIS | I |
| Abbildungen | i |
| KURZÜBERSICHT | 1 |
| Die wichtigsten Informationen im Überblick | 1 |
| Preisgestaltung | 2 |
| Paketierung | 4 |
| Technologie | 5 |
| Fazit | 5 |
| TRENDS | 7 |
| Übersicht | 7 |
| Wachstum | 7 |
| <i>Allgemeine Erfahrungswerte</i> | 7 |
| <i>Praktisches Beispiel</i> | 8 |
| Workloadkomplexität | 9 |
| Datenaktualität | 10 |
| Auswirkungen | 11 |
| LÖSUNGEN | 11 |
| InfoSphere Warehouse | 11 |
| <i>Allgemeiner Überblick</i> | 11 |
| <i>Datenkomprimierung</i> | 13 |
| <i>Leistungsoptimierungsprogramme (Performance Accelerators)</i> | 14 |
| <i>Workload-Management</i> | 14 |
| <i>Automatisierung</i> | 15 |
| <i>SQL-Kompatibilität</i> | 16 |
| <i>XML-Unterstützung</i> | 17 |
| Oracle-Lösungen | 17 |
| BERECHNUNGSGRUNDLAGEN | 19 |
| INFORMATIONEN ZUR INTERNATIONAL TECHNOLOGY GROUP | 21 |

Abbildungen

| | |
|--|----|
| 1. Abbildung 1: Softwarekosten über einen Zeitraum von drei Jahren für IBM InfoSphere Warehouse und Oracle Database 11g – Data-Warehouses zwischen 10 TB und 50 TB | 1 |
| 2. Abbildung 2: Softwarekosten über einen Zeitraum von drei Jahren für IBM InfoSphere Warehouse und Oracle Database 11g – Data-Warehouses zwischen 1 TB und 5 TB | 2 |
| 3. Abbildung 3: IBM InfoSphere Warehouse und Oracle Data Warehouse – Paketierung | 4 |
| 4. Abbildung 4: Data-Warehouse-Wachstum – am Beispiel eines Fertigungsunternehmens | 8 |
| 5. Abbildung 5: Variablen für eine Standardanalyse von Kundendaten | 9 |
| 6. Abbildung 6: Verteilung von Data-Warehouse-Workloads – Beispiel aus der Finanzdienstleistungsbranche | 10 |
| 7. Abbildung 7: Die zentralen InfoSphere Warehouse-Komponenten | 12 |
| 8. Abbildung 8: Weitere InfoSphere Warehouse Advanced Enterprise Edition-Komponenten | 13 |
| 9. Abbildung 11: Kostenvergleich – Installationen zwischen 10 TB und 50 TB | 19 |
| 10. Abbildung 12: Kostenvergleich – Installationen zwischen 1 TB und 5 TB | 20 |

KURZÜBERSICHT

Die wichtigsten Informationen im Überblick

Die Implementierung von Data-Warehouse-Lösungen setzt sich mit unverminderter Geschwindigkeit fort. Da Informationen für Unternehmen zu einem immer wichtigeren Kriterium für den geschäftlichen Erfolg werden, ist der Bereich Data-Warehousing zu einem der am schnellsten wachsenden Segmente in der IT-Welt geworden.

Diese erfolgreiche Entwicklung bringt jedoch auch neue Herausforderungen mit sich. Sobald Data-Warehouse-Lösungen installiert und in Betrieb genommen werden, wachsen auch die Anforderungen der Endbenutzer mit einer erstaunlichen Geschwindigkeit. Unternehmen, die zunächst mit kleinen, kompakten Systemen einsteigen, stellen nach einigen wenigen Jahren möglicherweise schnell fest, dass sie es mit Datenmengen im Terabyte-Bereich (TB), den unterschiedlichsten Anwendungen und Workloads und Hunderten oder gar Tausenden von Benutzern zu tun haben.

Sobald Data-Warehouses wachsen und damit komplexer werden, können die damit verbundenen Kosten in den Folgejahren schnell ansteigen. So stellt sich die Frage, wie sich diese Entwicklung am besten steuern lässt? Die Antwort hängt zu einem großen Teil davon ab, welche Data-Warehouse-Lösung implementiert wird. Kriterien wie Preismodell und Paketierung der Anbieter sowie Technologiegehalt können zu deutlich unterschiedlichen Kostenstrukturen führen.

Im vorliegenden Bericht werden die kostenspezifischen Auswirkungen der Unterschiede zwischen zwei der größten softwarebasierten Data-Warehouse-Frameworks untersucht: IBM InfoSphere Warehouse und Data-Warehouses auf Basis von Oracle Database 11g sowie der zugehörigen Tools.

InfoSphere Warehouse bietet deutlich niedrigere Softwarekosten. Bei Data-Warehouses mit 10 TB bis 50 TB Benutzerdaten liegen die Kosten für Lizenzen und Support über drei Jahre für InfoSphere Warehouse Enterprise Edition um durchschnittlich 39 Prozent unter den Kosten für Oracle Database 11g Enterprise Edition. Abbildung 1 verdeutlicht dies.

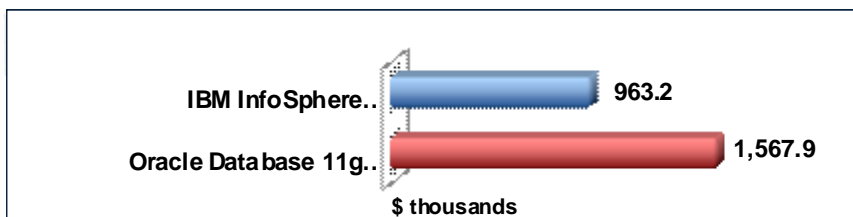


Abbildung 1: Softwarekosten über einen Zeitraum von drei Jahren für IBM InfoSphere Warehouse und Oracle Database 11g – Data-Warehouses zwischen 10 TB und 50 TB

Bei Data-Warehouses mit 1 TB und 5 TB Benutzerdaten sind die Kostenunterschiede noch größer. Wie in Abbildung 2 gezeigt, liegen hier die Kosten für InfoSphere Warehouse Departmental Edition im Durchschnitt 69 Prozent niedriger.

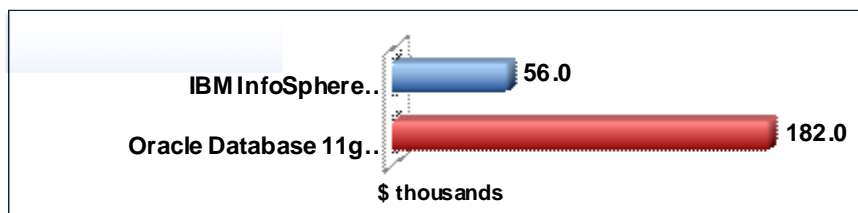


Abbildung 2: Softwarekosten über einen Zeitraum von drei Jahren für IBM InfoSphere Warehouse und Oracle Database 11g – Data-Warehouses zwischen 1 TB und 5 TB

Diese Vergleiche basieren auf sechs repräsentativen Data-Warehouse-Installationen in großen und mittelständischen Unternehmen. Die InfoSphere Warehouse-Kosten wurden mithilfe der neuen **Pro-Terabyte**-Preisoption von IBM auf Basis der Terabyte-Anzahl an Benutzerdaten errechnet. Die Kosten für die Oracle Database 11g-Lösungen wurden auf Basis der vom Unternehmen angebotenen kostengünstigsten Option – pro Prozessor oder pro benanntem Benutzer – berechnet.

Für die Lizenz- und Supportgebühren für die InfoSphere Warehouse Enterprise- und Departmental Edition-Lösungen wird ein Nachlass von 40 Prozent, für die Oracle Database 11g Enterprise Edition-Lösungen ein Nachlass von 80 Prozent gewährt. Je nach Kunde können die von den beiden Unternehmen gewährten Nachlässe variieren.

Einzelheiten zu Installation, Konfiguration und Preisgestaltung finden Sie im Abschnitt „Berechnungsgrundlagen“ in diesem Bericht.

Preisgestaltung

Oracle bietet für Oracle Database 11g-Warehouselösungen zwei Preismodelle. Im Vergleich zum Pro-Terabyte-Preismodell von IBM weisen diese beiden Modelle mögliche Nachteile auf. Sie lassen sich wie folgt zusammenfassen:

1. **Prozessorbasiert.** Das gängigste Oracle-Preismodell basiert auf der Anzahl der physischen Prozessorkerne. Der Trend beim Mikroprozessordesign geht jedoch dahin, die Anzahl der Kerne pro Chip zu erhöhen. Während dadurch das Preis-Leistungs-Verhältnis bei den Servern insgesamt zwar verbessert wird, erfolgt dies bei der Leistung pro Kern nur langsamer. Niedrigere Hardwarekosten ergeben nicht automatisch niedrigere Softwarekosten pro Kern.

Die schrittweisen Upgrades nehmen hingegen zu. Ein x86-Server mit zwei Sockets und der Intel Westmere 5600 Series-Technologie der neuesten Generation beispielsweise verfügt über 12 Kerne. Beim Wechsel zu einem Server mit vier Sockets und der letzten Generation der Intel Nehalem EX 7500 Series-Technologie ergibt dies eine Steigerung auf 32 Kerne.

Grundsätzlich lassen sich Kosten durch die Implementierung in Partitionen mithilfe von VMware und vergleichbaren Produkten senken. Mit der Ausnahme des eigenen VM-Hypervisors sieht Oracle jedoch x86-Hypervisoren als „softe“ Partitionierungstechnologien, für die Software pro Partition nicht lizenziert werden kann. Im Vergleich dazu unterstützt IBM InfoSphere Warehouse aus Gründen der Preisgestaltung diese Technologien.

Oracle-Support für seine Produkte auf VMware ist möglich. Das Unternehmen erbringt Supportleistungen für Probleme, von denen man weiß, dass sie entweder auf dem nativen Betriebssystem auftreten werden, oder die nachweislich nicht durch die Ausführung unter VMware verursacht werden. Bei einem komplexen, sehr schnell wachsenden Data-Warehouse ist dies als hoch riskantes Unterfangen zu sehen.

„Harte“ Partitionierungstechnologien wie logische Partitionen (LPARs) auf IBM Power-Servern werden von Oracle auf Pro-Partitions-Preisbasis unterstützt. Die Pro-Prozessor-Preise von Oracle für solche Plattformen sind jedoch deutlich höher als für x86-Server. Das Unternehmen wendet einen Kernfaktor von 0,75 bis 1,0 für RISC-Systeme der neuesten Generation an, im Vergleich zum Faktor 0,5 für x86-Prozessoren.

- 2. Benutzerbasiert.** Das Oracle-Preismodell auf Basis der Anzahl benannter Benutzer kommt hauptsächlich bei kleinen Installationen zum Tragen. Die Erfahrung hat jedoch gezeigt, dass Data-Warehouse-Benutzergruppen mit der Zeit immer größer werden. Auch wenn Systeme in erster Linie spezielle Benutzergruppen wie Führungskräfte und Analysten unterstützen, versuchen andere Gruppen in der Regel ebenfalls, auf dieselben Datenressourcen zuzugreifen.

Ein Data-Warehouse mit Analysedaten, das beispielsweise 30 Benutzer unterstützen soll, kann innerhalb von wenigen Jahren so anwachsen, dass es Hunderten von Managern, Spezialisten und Mitarbeitern im Kundenservice wie Vertriebsmitarbeitern und Kundendienstmitarbeitern als Quelle dient. In Großunternehmen sind Benutzergruppen mit mehreren Tausend Benutzern durchaus keine Seltenheit mehr.

Das Pro-Terabyte-Preismodell von IBM bietet eine Vielzahl von Vorteilen. Die Data-Warehouse-Kosten lassen sich wesentlich besser an den Wachstumsraten ausrichten. Die Benutzergruppen können zunehmen, ohne dass die jeweiligen Lizenzgebühren steigen. Zudem werden die Benutzer nicht für gleichzeitig ablaufende oder komplexe Abfrageworkloads „abgestraft“, die eine hohe Verarbeitungsleistung erfordern.

Zwei weitere Faktoren tragen ebenfalls zur Senkung der Kosten bei. Erstens beinhaltet die Definition von IBM des Begriffs „Benutzerdaten“ keine Indizes, Protokolle, temporären Systembereiche und andere Datenstrukturen, die in der Regel 30 bis 50 Prozent des gesamten Datenvolumens in einem Data-Warehouse ausmachen. Zweitens erlaubt dieses Datenmodell den Einsatz von Datenkomprimierungsfunktionen. Ein Data-Warehouse mit 2 TB, das auf 1 TB komprimiert wird, erfordert somit auch nur eine 1-TB-Lizenz.

Im vorliegenden Vergleich wurde bei den IBM InfoSphere Warehouse-Berechnungen eine konservative Komprimierungsrate von 50 Prozent zu Grunde gelegt. Im Allgemeinen lassen sich aber deutlich höhere Raten erzielen. Bei Oracle Database 11g sind Datenkomprimierungen ebenfalls möglich. Dies hat jedoch keine Auswirkungen auf das Preismodell.

Die Kostenunterschiede zeigen sich noch deutlicher, sobald Standby-Systeme verwendet werden. In einer Aktiv-Passiv-Konfiguration berechnet IBM nur 1TB an Daten auf dem passiven Server, während bei Oracle eine volle Pro-Prozessor- oder Pro-Benutzer-Lizenz erworben werden muss. Bei Aktiv-Aktiv-Konfigurationen berechnen beide Anbieter eine Full-Capacity-Lizenz. Die InfoSphere Warehouse-Preise liegen jedoch im Allgemeinen aufgrund der oben beschriebenen anderen Faktoren niedriger.

IBM bietet für InfoSphere Warehouse-Lösungen auch die Preisoptionen „Pro Prozessor“, „Pro Benutzer“ und „Pro Socket“. Es gibt jedoch bestimmte Einschränkungen, wodurch nicht alle Optionen für alle Versionen verfügbar sind.

Paketierung

In Bezug auf die Paketierung gibt es zwischen den IBM InfoSphere Warehouse- und Oracle-Angeboten deutliche Unterschiede. Diese sind nachfolgend in Abbildung 3 dargestellt.

| INFOSPHERE WAREHOUSE ENTERPRISE EDITION | ORACLE DATABASE 11g ENTERPRISE EDITION |
|--|--|
| Im Lieferumfang enthalten: <ul style="list-style-type: none"> • Partitioning • Deep Compression • Cubing Services • Intelligent Miner • Management-Tools Listenpreis pro TB: \$70,000 | Nicht im Lieferumfang enthalten, separate Berechnung: <ul style="list-style-type: none"> • Partitioning • Advanced Compression • OLAP • Data Mining • Diagnose- und Optimierungs-Packs Listenpreis pro Benutzer: \$2,530 Listenpreis pro Prozessor: \$126,500 |
| INFOSPHERE WAREHOUSE DEPARTMENTAL EDITION | ORACLE DATABASE 11g STANDARD EDITION |
| Im Lieferumfang enthalten: <ul style="list-style-type: none"> • Partitioning • Deep Compression • Cubing Services • Intelligent Miner • Management-Tools Listenpreis pro TB: \$35,000 | Nicht unterstützt: <ul style="list-style-type: none"> • Partitioning • Advanced Compression • OLAP • Data Mining Nicht im Lieferumfang enthalten, separate Berechnung: <ul style="list-style-type: none"> • Diagnose- und Optimierungs-Packs Listenpreis pro Benutzer: \$550 Listenpreis pro Prozessor: \$27,500 |

Abbildung 3: IBM InfoSphere Warehouse und Oracle Data Warehouse – Paketierung

Sowohl die Enterprise Edition als auch die Departmental Edition von IBM InfoSphere Warehouse beinhalten im Basispaket die Datenbank DB2 for Linux, UNIX und Windows (LUW) 9.7 sowie Tools für Partitionierung, Komprimierung, Online Analytical Processing (OLAP), Data Mining, Management und andere Tools.

Oracle berechnet vergleichbare Module separat. Der Listenpreis für Oracle Database 11g Enterprise Edition liegt bei \$ 950 pro benanntem Benutzer und \$ 47.500 pro Prozessor. Berücksichtigt man die genannten Module steigt der Listenpreis auf \$ 2.530 pro benanntem Benutzer und \$ 126.500 pro Prozessor.

Ein weiterer wesentlicher Unterschied liegt darin, dass IBM eine kostengünstige Version, IBM InfoSphere Warehouse Departmental Edition, im Angebot hat, die das gesamte Data-Warehouse-Funktionsspektrum für Systeme mit bis zu 16 Serverkernen, 64 GB Hauptspeicher pro Instanz und 15 TB Benutzerdaten bietet.

Im Vergleich dazu unterstützt die Oracle-Einstiegsversion (bis zu vier Sockets) Database 11g Standard Edition keine Partitionierung, Advanced Compression, OLAP und Data Mining. Während sich die Einstiegsversion Standard Edition vielleicht für Business Intelligence-Basisanwendungen eignet, stellt sie keine ernsthafte Alternative für das unternehmensweite Data-Warehousing dar.

Aus diesem Grund beziehen sich die oben angeführten Kosten für Oracle Database 11g-Lösungen für Data-Warehouses mit 1 TB bis 5 TB bzw. 10 TB bis 50 TB auf die Oracle Enterprise Edition.

Technologie

Die Unterschiede zwischen den IBM InfoSphere Warehouse- und Oracle Database 11g-Lösungen zeigen sich neben den Preismodellen und der Paketierung auch in anderen Bereichen. Die Unterschiede im Technologiegehalt zwischen den Datenbanken IBM DB2 9.7 und Oracle Database 11g sowie zwischen den zugehörigen Tools wirken sich ebenfalls auf den Kostenvergleich aus.

So ermöglicht DB2 9.7 beispielsweise eine höhere Datenkomprimierung. Hinzu kommen die leistungsfähigen Leistungsoptimierungsprogramme, die Managementfunktionen für gemischte Workloads sowie die Verwaltungs- und Optimierungstools, die eine effizientere Nutzung der Systemressourcen ermöglichen. Die umfassenderen Automatisierungsmöglichkeiten resultieren in einer höheren Produktivität der Datenbankadministratoren und einer höheren Systemverfügbarkeit.

Beim Vergleich des Technologiegehalts der InfoSphere Warehouse- und Oracle Database 11g-Lösungen kommen einige Punkte immer wieder zum Vorschein. Das Design von DB2 9.7 ist so ausgelegt, dass eine hohe Systemeffizienz und Administratorproduktivität in der gesamten Infrastruktur aus Servern, Speichersystemen und Software, die Datenbankanwendungen unterstützen, gewährleistet ist.

Hinzu kommt, dass die Stärken von DB2 9.7 insbesondere in Umgebungen mit komplexen Datenbankschemata, großen Datenmengen, unterschiedlichen Workloads und häufigen Änderungen und Erweiterungen zum Tragen kommen. Die Stärken passen hervorragend zum grundlegenden Trend, dass Data-Warehouses mittlerweile in den unterschiedlichsten Branchen und Unternehmen aller Größenordnungen implementiert werden.

Ein weiteres typisches Merkmal der InfoSphere Warehouse-Lösungen ist die schnelle Implementierung dieser Lösungen. Die Departmental Edition beispielsweise lässt sich mithilfe eines herunterladbaren virtuellen Images auf einfache Weise installieren und konfigurieren. Komplexere Enterprise Edition-Implementierungen profitieren ebenfalls vom hohen Integrationsgrad und von den umfangreichen Tests. In beiden Fällen kann dadurch wesentlich schneller ein geschäftlicher Nutzen und ein höherer ROI erzielt werden.

Fazit

Aus Geschäftsperspektive stellt die fortschreitende Entwicklung des Data-Warehousing-Konzepts in Unternehmen wohl das größte Spektrum an neuen Anwendungschancen seit der Verbreitung des Internet dar. Die neuartige Anwendung von Informationen auf die Prozesse im gesamten Unternehmen – und auch auf Interaktionen mit Kunden und Partnern – wird Umwälzungen ermöglichen, deren Potenzial deutlich über das hinausgeht, was bis heute möglich war.

Aus IT-Perspektive stehen diese Data-Warehousing-Trends für den „Perfect Storm“. Wachsende Datenmengen, Anwendungen und Benutzergruppen gehen einher mit zunehmend komplexeren Datenstrukturen und unterschiedlicheren Workloads. Der Druck, Abfrage- und Aktualisierungszyklen in immer kürzeren Zeitabständen abwickeln zu müssen, trägt ebenfalls zu einer größeren Nachfrage nach Prozessorleistung, E/A-Durchsatz und Netzwerkbandbreite bei.

Gleichermaßen kann davon ausgegangen werden, dass sich der Technologiegehalt verändern wird. „Cloudnetzwerke“ werden zu gängigen Technologien für die Bereitstellung von Data-Warehouse-Services werden. Die Virtualisierung wird eine immer wichtigere Rolle in Client-, Server- und Speicherinfrastrukturen spielen. In Datenbanken werden neue Medientypen (z. B. Text und Image) und Standards (z. B. XML) zu finden sein. Unternehmensweite Suchprogramme und andere neue Technologien werden auf der Bildfläche erscheinen.

Umfang und Entwicklungsgeschwindigkeit eines solchen Wandels lassen sich nur schwer vorhersagen. Sicher ist jedoch, dass dieser Wandel neue Herausforderungen mit sich bringen wird. Data-Warehouse-Frameworks müssen nicht nur in der Lage sein, auf lange Sicht das anhaltende Wachstum und die zunehmende Workloadkomplexität auf kosteneffiziente Weise zu bewältigen, sondern auch breiter gefassten, qualitativ hochwertigeren Formen von Veränderungen gerecht zu werden.

Die Auswahl des richtigen Data-Warehouse-Frameworks ist folglich eine Entscheidung von besonderer Tragweite. Eine solche Entscheidung wird in nicht unerheblichem Umfang Auswirkungen darauf haben, wie gut Unternehmen Informationen nutzen können, um ihre Geschäftsanforderungen zu erfüllen. Die Konsequenzen aus der richtigen (oder falschen) Entscheidung werden auf jeden Fall langfristig zu spüren sein.

Was die Kosteneffizienz sowie bewährte und neue Technologien anbelangt, bietet InfoSphere Warehouse ein herausragendes Leistungs- und Funktionsspektrum.

TRENDS

Übersicht

Der Wettbewerbsvorteil, der für InfoSphere Warehouse spricht, liegt nicht nur darin, dass das IBM Preismodell und die Paketierung mehr Flexibilität und Kosteneffizienz bieten. Es sind vielmehr die unverwechselbaren Merkmale von InfoSphere Warehouse, die gezielt und auf effiziente Weise auf die wichtigsten Trends beim Data-Warehousing in Unternehmen eingehen. Dies gilt insbesondere für die folgenden drei Bereiche – Wachstum, Workloadkomplexität und Datenaktualität –, auf die wir nachfolgend näher eingehen wollen.

Das darauffolgende Kapitel zum Thema „Lösungen“ enthält weitere Informationen zu den IBM InfoSphere Warehouse- und Oracle Database 11g-Angeboten einschließlich Oracle Exadata Database Machine. Zudem gehen wir auf die wichtigsten Alleinstellungsmerkmale der InfoSphere Warehouse-Technologie ein und untersuchen deren Auswirkungen im Detail.

Der Bericht schließt mit dem Kapitel „Berechnungsgrundlagen“, in dem wir die Vorgehensweisen und Annahmen beschreiben, die bei den vergleichenden Kostenberechnungen zur Anwendung kamen. Hinzu kommt eine detaillierte Aufschlüsselung der über einen Zeitraum von drei Jahren anfallenden Kosten durch den Einsatz der InfoSphere Warehouse- und Oracle Database 11g-Lösungen.

Wachstum

Allgemeine Erfahrungswerte

Bei Data-Warehouses ist in der Regel ein schnelleres Datenwachstum festzustellen als bei allen anderen größeren Systemarten. Jährliche Wachstumsraten im hohen zweistelligen Bereich sind in den meisten Branchen und Unternehmen aller Größenordnungen mittlerweile zur Norm geworden.

Dieser Trend wurde durch eine Vielzahl von Faktoren hervorgerufen. Unternehmen sammeln heute wesentlich mehr Daten zu Kunden und Transaktionen. In den Anfängen des neuen Jahrtausends beispielsweise verwalteten Banken in der Regel zwischen 1 und 2 MB Daten pro Kunde. Hier liegt der Durchschnitt mittlerweile bei über 12 MB. In der Telekommunikationsbranche, im Einzelhandel und in anderen Branchen hat sich die Datenmenge in den letzten zehn Jahren mehr als verzehnfacht.

Eine ähnliche Entwicklung ist in Unternehmen mit komplexen Lieferkettenprozessen festzustellen. Mittlerweise ist es technisch möglich, Millionen von SKUs (Artikelpositionen) in deren gesamtem Lebenszyklus mit einem deutlich höheren Detaillierungsgrad zu verfolgen – von der Bedarfsfermittlung über Akquisition und Distribution bis zu Vertrieb und Retouren. Die zunehmende Implementierung der RFID-Technologie trägt ebenfalls dazu bei, die Datenmenge weiter zu erhöhen.

Neben der zunehmenden Anzahl und Größe der Datensätze sind auch die längeren Aufbewahrungszeiträume ein nicht zu unterschätzender Faktor. Daten werden mittlerweile in vielen Unternehmen routinemäßig über einen Zeitraum von sieben Jahren oder länger für Protokollanalysen und Compliancezwecke aufbewahrt.

Auch wenn der hohe Anstieg bei der Datenmenge die Schlagzeilen dominiert, ist der allgemeine Trend wesentlich breiter gefächert und komplexer. Die meisten Unternehmen, die Data-Warehouses implementieren, stellen außerdem einen Anstieg bei der Anzahl der Anwendungen und Benutzer über lange Zeiträume fest.

Die Erfahrung hat gezeigt, dass der Bedarf an qualitativ hochwertigen Informationen und Analysedaten rapide ansteigt, sobald deren Nutzen erkannt wurde. Sobald Data-Warehouse-Lösungen installiert und in

Betrieb genommen werden, nutzen auch zahlreiche andere Specialist User und andere große Benutzergruppen wie Manager, Experten und Mitarbeiter im Kundenservice diese Informationsquelle.

In einigen Branchen wurde der Data-Warehouse-Zugriff auch auf Distributoren, Agenten, Broker und anderen Geschäftspartner ausgedehnt, sodass auch dadurch die Benutzergruppen ständig größer wurden. Diese Zielgruppen haben festgestellt, dass Data-Warehouses durch die Zugriffsmöglichkeiten auf qualitativ hochwertige Informationen und die entsprechenden Analyse- und Auswertungstools eine hervorragende Quelle bilden, um einen hohen geschäftlichen Nutzen zu erzielen.

Praktisches Beispiel

Das Fertigungsunternehmen, dessen Erfahrungswerte in Abbildung 4 zusammengefasst sind, verdeutlicht die Wachstumstrends.

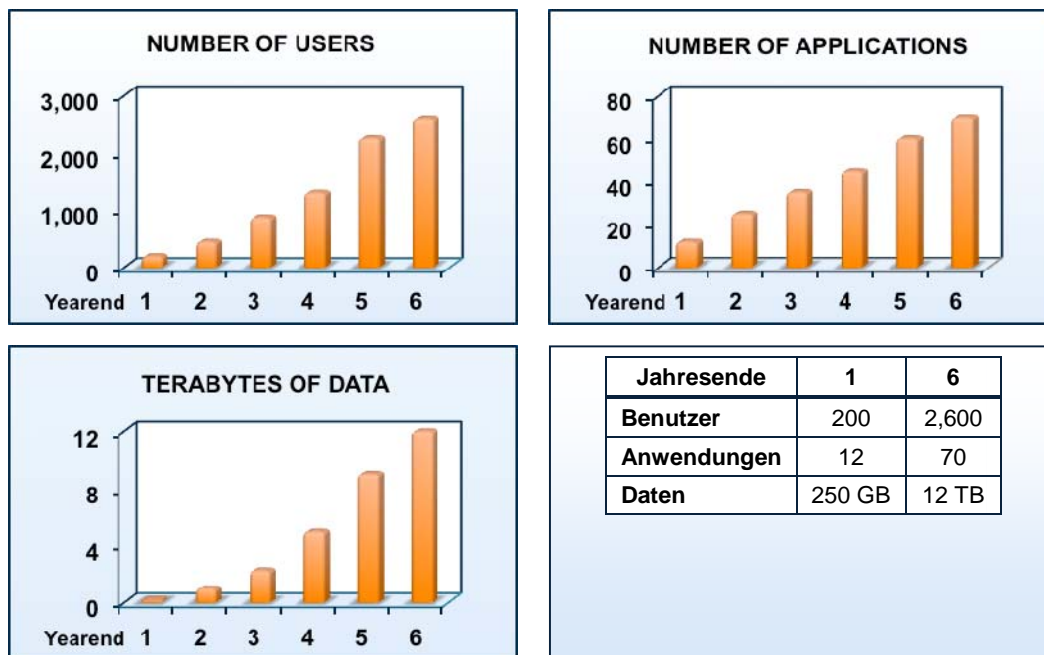


Abbildung 4: Data-Warehouse-Wachstum – am Beispiel eines Fertigungsunternehmens

Im vorliegenden Beispiel ist das Data-Warehouse des Unternehmens innerhalb von fünf Jahren von einem Datenbestand von unter 250 GB, einem halben Dutzend Anwendungen und 100 Benutzern auf über 12 TB, 70 Anwendungen und 2.600 Benutzer angewachsen – dies entspricht ca. 10 Prozent der gesamten Belegschaft des Unternehmens.

Einige Aspekte dieser Erfahrungswerte des Unternehmens sind sehr aufschlussreich. Zum einen hat sich gezeigt, dass Wachstum ein nicht linearer Prozess ist. Die jährlichen Wachstumsraten variieren von Jahr zu Jahr; zudem lassen sich auch kurzfristige Trends in manchen Fällen nicht vorhersagen. Bei bestimmten Anwendungen war ein geradezu explosionsartiges Wachstum bei der Nutzung in besonderes großen Segmenten im Unternehmen festzustellen. Bei anderen Anwendungen blieb die Benutzeranzahl größtenteils, auch langfristig gesehen, stabil.

Das Wachstum wurde durch die Konsolidierung von Datamarts weiter verstärkt. Als in den Jahren nach 2000 immer mehr BI-Tools eingesetzt wurden, implementierten viele Abteilungen anfänglich ihre eigenen Systeme. Die Probleme wurden jedoch schnell offensichtlich. Inkonsistenzen in Dateninhalten, -strukturen und -aktualität sowie Inkompatibilitäten auf technischer Seite zwischen den Abteilungen hatten zur Folge, dass auch die Ergebnisse häufig inkompatibel waren.

Die Kosten und die Schwierigkeiten für das separate Datenmanagement, Datenextraktionen, -transformationen und -ladevorgänge (ETL-Prozesse) und die Replizierung von Infrastrukturen für Dutzende separater Systeme erreichten bald inakzeptable Dimensionen. Auf der Grundlage bestehender Unternehmensrichtlinien wurden die meisten Datamarts auf dem Hauptsystem im Unternehmen konsolidiert.

Obwohl die Wachstumsraten differieren, sind die von den meisten Unternehmen, die Data-Warehouses implementierten, berichteten Erfahrungen in der Regel vergleichbar.

Workloadkomplexität

Mit der zunehmenden Nutzung von Data-Warehouses nahm nicht nur die Anzahl der reinen „zahlenverarbeitenden“ High-End-Anwendungen, sondern auch die der abteilungsspezifischen und individuellen Benutzertools zum Auffinden, Analysieren und Nutzen von Daten zu. Drill-down-, OLAP- und andere innovative Funktionen, die bisher nur Fachleuten zugänglich waren, werden mittlerweile in den Unternehmen routinemäßig vorausgesetzt.

Bei der Komplexität analytischer Prozesse ist ein progressiver Anstieg zu verzeichnen. Abteilungen, die beispielsweise mit der Ermittlung von Kunden, die in ein bestimmtes Profil passten, begonnen haben, analysieren mittlerweile produktübergreifende Kaufmuster oder versuchen, das Kundenverhalten vorherzusagen. Zunehmend werden auch Finanzanalysen durchgeführt, wobei versucht wird, Umsatz und Gewinn jeder Abteilung anteilig zu quantifizieren, um Strategien ausarbeiten zu können, die die Profitabilität im gesamten Unternehmen optimieren.

Selbst standardmäßige Analysen von Kundendaten umfassen mittlerweile die in Abbildung 5 gezeigten Variablen.

- ◆ Durchschnittl. Gewinn pro Kunde, Umsatz, Barzahlungen, Anzahlungen und Außenstände
- ◆ Durchschnittl. Dispositionskredit/Guthaben ◆ Kundengewinnungskosten ◆ Durchschnittl. Artikelpreis
- ◆ Durchschnittl. Gewinn pro Artikel ◆ Durchschnittl. Höhe der Warenkorbverkäufe ◆ Anzahl der verkauften Artikel
- ◆ Anzahl der Transaktionen ◆ Gewinnspanne ◆ Umsatzerlöse ◆ Transaktionszins
- ◆ Gesamtsaldenvereinbarung, Saldoveränderung, Anfangssaldo, Endsaldo
- ◆ Anzahl und Wert der angesammelten, eingelösten, stornierten Punkte, Punktestand
- ◆ Nachlass ◆ Verkaufswert der zurückgegebenen Waren
- ◆ Anzahl der gescannten, nicht gescannten, verkauften, zurückgegebenen und geleerten Warenkörbe
- ◆ Anzahl der Haushaltskunden ◆ Durchschnittl. Kundenalter und Einkommen pro Haushalt
- ◆ Durchschnittl. Anzahl der pro Tag verkauften Artikel ◆ Anzahl der Kundenbesuche ◆ Anzahl der Tage seit dem letzten Einkauf
- ◆ Durchschnittl. Anzahl der Vereinbarungen ◆ Durchschnittl. Kaufmenge ◆ Kundenbindung
- ◆ Anzahl der Beschwerden, durchschnittl. Anzahl der Beschwerden ◆ Anzahl der ertraglosen Konten
- ◆ Wertschöpfung, Marktvolumen, Prozentsatz ◆ Kundenmarktanteil, Prozentsatz
- ◆ Anzahl der Verzugstage, Verzugshöhe ◆ Anzahl der genutzten Kanäle

Abbildung 5: Variablen für eine Standardanalyse von Kundendaten

Dies ist eine der vielen Funktionen von IBM InfoSphere Warehouse Pack for Customer Insight, die nachfolgend beschrieben wird.

Die zunehmende Anwendungsvielfalt verstärkt diese Auswirkungen zusätzlich. Data-Warehouses in Unternehmen unterstützen mittlerweile eine Vielzahl von Abfragen mit variierenden Merkmalen, unterschiedlichen Bedeutungen für das Unternehmen und unterschiedlicher Zeitsensitivität für verschiedene Benutzergruppen.

Abbildung 6 beispielsweise zeigt die Schätzungen eines Unternehmens aus der Finanzdienstleistungsbranche zu den gesamten Data-Warehouse-Workloads in einem Zeitraum von 24 Stunden. Die Prozentangaben stehen für die Anzahl der CPU-Zyklen.

Im vorliegenden Fall, und dies gilt auch für viele andere Fälle, hat sich der Nutzungsumfang grundlegend gewandelt. Und zwar von relativ einfachen standardisierten Managementberichten und Adhoc-Abfragen zu wesentlich umfangreicheren Data-Mining-Jobs, OLAP, einer Vielzahl unterschiedlicher BI-Anwendungen für über 20 Geschäftsbereiche einschließlich Call-Center, Vertriebsmitarbeiter usw.

In solchen Umgebungen ist ein effektives Management für gemischte Workloads unabdingbar. Ohne eine differenzierte Jobzeitplanung und präzise Überwachung der Ressourcenauslastung kann es in den Unternehmen in Spitzenzeiten sehr schnell zu Engpässen und/oder zu einer Unterauslastung von Kapazitäten in anderen Zeitfenstern kommen.

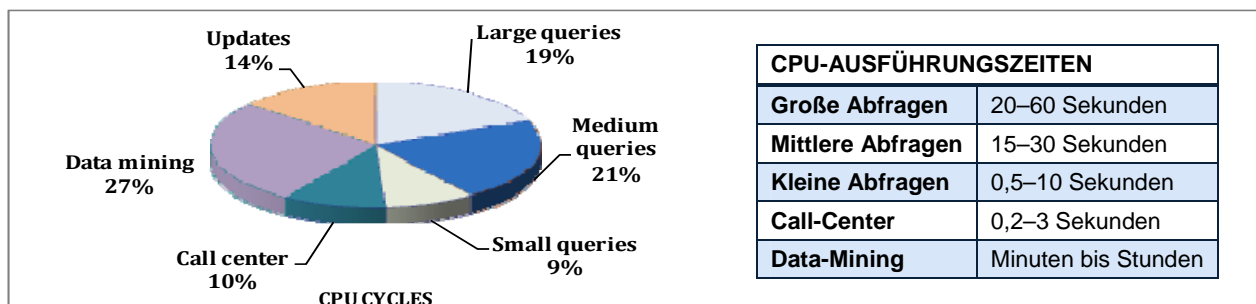


Abbildung 6: Verteilung von Data-Warehouse-Workloads – Beispiel aus der Finanzdienstleistungsbranche

Es hat sich gezeigt, dass die Möglichkeit, Workloads zu priorisieren und die verfügbaren Ressourcen zu steuern, ebenfalls ein kritischer Faktor für die Einhaltung von SLA-Zielen ist.

Datenaktualität

Bei herkömmlichen Data-Warehousing-Ansätzen wird davon ausgegangen, dass die Anforderungen an die Datenaktualität nicht so hoch sind wie bei den meisten anderen Systemarten.

In den letzten zehn Jahren jedoch war bei vielen Anwendungen eine deutliche und schnellere Veränderung hinsichtlich der Bedeutung der Datenaktualität festzustellen. Konzepte wie „Operational BI“ – also der Data-Warehouse-Zugriff durch Contact-Center-Mitarbeiter, Vertriebspersonal und andere Gruppen mit Kundenkontakt – haben zu einer erhöhten Nachfrage nach kontinuierlich aktualisierten Informationen geführt.

Weitere Beispiele hierfür sind Echtzeitanalysen von Verkaufs-, Preis- und Werbemitteldaten in Marketingunternehmen, Lieferkettenanalysen in Bereichen wie Fertigung, Einzelhandel, Distribution, Logistik, Transport, Gesundheitswesen und anderen Branchen sowie die kontinuierliche Analyse von finanzspezifischen und betrieblichen KPIs in den verschiedensten Unternehmen.

Dieser Wandel wurde durch den weltweiten wirtschaftlichen Abschwung noch verstärkt. Seit den Anfängen der Rezession im Jahr 2008 tendieren Unternehmen dazu, ihre Planungs- und Entscheidungszyklen zu verkürzen, um schneller auf sich ständig verändernde Geschäftsbedingungen reagieren zu können. Entsprechend hat die Bedeutung aktueller Daten zugenommen.

Neue Anwendungstrends haben zu immer kürzeren Aktualisierungszyklen bei den Data-Warehouses geführt. Im Jahr 2010 wurde im Rahmen einer Studie der International Technology Group festgestellt, dass 85 Prozent der Fortune-500-Unternehmen mindestens täglich bestimmte Data-Warehouse-Inhalte (in der Regel zwischen 15 und 30 Prozent) aktualisieren, während 29 Prozent mehrmals täglich Aktualisierungen durchführen.

Auswirkungen

Die Auswirkungen dieser Trends lassen den Schluss zu, dass die meisten Unternehmen, die Data-Warehouses implementieren, in absehbarer Zukunft einen anhaltenden Anstieg bei den Softwarelizenzierungs- und Supportkosten erleben werden. Die Kosten für Server, Speicher und Netzwerke für die Data-Warehouses sowie die Personalkosten für die Verwaltung dieser Ressourcen werden wohl ebenfalls steigen.

Dieser Prozess lässt sich jedoch gut steuern. Die Steuerungsabläufe erweisen sich bei IBM InfoSphere Warehouse-Lösungen deutlich effizienter als bei Oracle Database 11g-Lösungen.

Es sind nicht nur Kriterien wie das IBM Preismodell und die Paketierung, die mehr Wirtschaftlichkeit bieten als die Oracle Database 11g-Pendants. Die InfoSphere Warehouse-Komponenten bieten darüber hinaus eine einzigartige Funktionalität – wie z. B. Datenkomprimierung, Leistungsmessung, Automatisierung usw – die sich ebenfalls auf die Kosten auswirken.

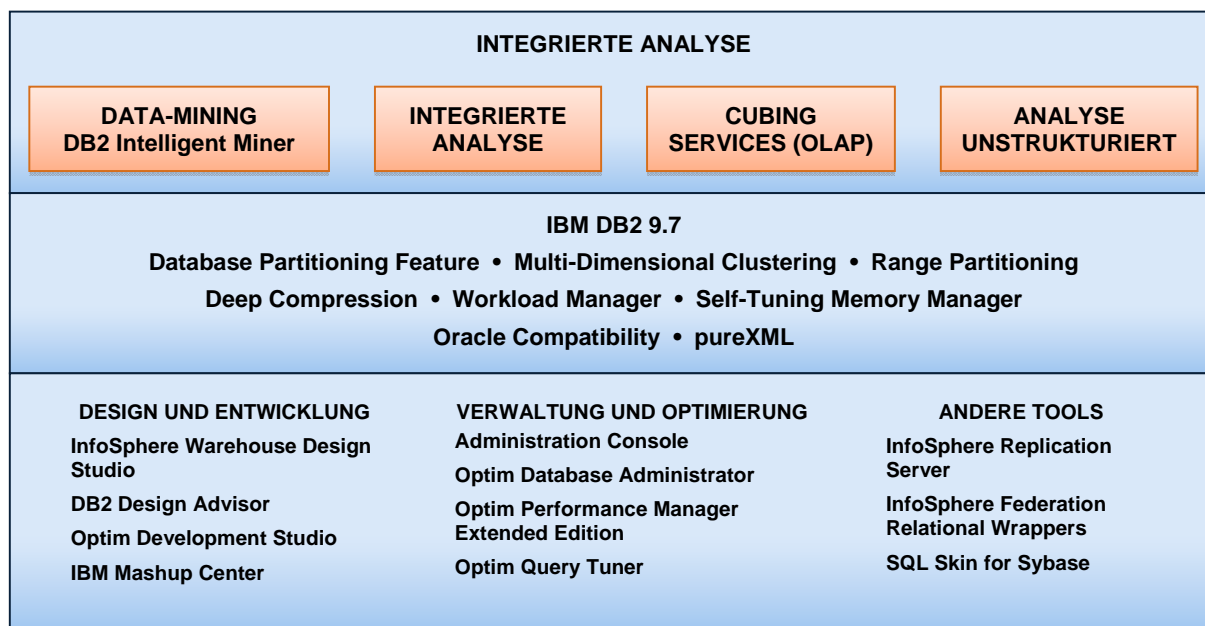
LÖSUNGEN

InfoSphere Warehouse

Allgemeiner Überblick

InfoSphere Warehouse ist ein preislich interessantes Lösungspaket bestehend aus IBM Lösungen, in deren Mittelpunkt das Datenbankmanagementsystem IBM DB2 9.7 steht. Es umfasst integrierte Analyseanwendungen (IBM Intelligent Miner, Cognos 8 BI und andere) sowie Tools für Design und Entwicklung, Verwaltung und Optimierung und andere gängige Data-Warehousing-Tools.

Diese Komponenten, die in Abbildung 7 dargestellt sind, spiegeln verschiedene IBM Marken wider.



Sie zeichnen sich durch Merkmale wie hervorragende Integrationsfähigkeit und Optimierungsoptionen, herausragende Leistung und hohe Funktionalität aus. Installation und Anpassung gestalten sich deutlich einfacher und die Mitarbeiterproduktivität ist wesentlich höher als bei einer Einzelimplementierung der Komponenten.

Abbildung 7: Die zentralen InfoSphere Warehouse-Komponenten

InfoSphere Warehouse wird in zwei wesentlichen Versionen angeboten: Enterprise Edition und Departmental Edition. Bei beiden Versionen werden dieselben Komponenten unterstützt, wobei Departmental Edition auf Konfigurationen bis maximal 16 Serverkerne für DB2, 64 GB Hauptspeicher pro DB2-Instanz und maximal 15 TB Benutzerdaten beschränkt ist. Bei der Partitionierung gibt es keine Beschränkungen.

Um die Implementierung zu vereinfachen und zu beschleunigen, kann InfoSphere Warehouse Departmental Edition als virtuelles Image auf x86-Servern installiert werden. Eine SUSE Linux Enterprise Server (SLES) 11-Lizenz ist im Virtual Image-Paket enthalten, das als Option beim Pro-Terabyte-Preismodell erhältlich ist.

IBM bietet darüber hinaus die Versionen InfoSphere Warehouse Advanced Enterprise und Advanced Departmental Edition, die mit weiteren Komponenten ausgestattet sind. Hierzu gehören u. a. das InfoSphere Data Architect-Tool für Datenbankdesign und -modellierung sowie die Advanced Recovery Solutions, die eine höhere Geschwindigkeit und mehr Präzision bei Datenwiederherstellungen ermöglichen – ein wesentlicher Vorteil dabei ist, dass sich Ausfallzeiten bei geplanten und ungeplanten Ausfällen deutlich reduzieren lassen.

Die Advanced Edition-Versionen umfassen auch Pakete mit drei gängigen Data-Warehouse-Anwendungsgruppen. Diese Pakete – für die Analyse von Kundendaten, Marketing- und Vertriebsdaten und Lieferkettendaten – sind Teil eines umfassenden Portfolios an IBM Branchen- und Prozessmodellen für Branchen wie Bankwesen, Finanzwesen, Gesundheitswesen, Versicherungen, Einzelhandel, Telekommunikation und andere.

Weitere Komponenten für Advanced Enterprise Edition sind in Abbildung 8 zusammengefasst.

| KOMPONENTE | BESCHREIBUNG |
|--|---|
| DESIGN UND MODELLIERUNG | |
| InfoSphere Data Architect | Eclipse-basiertes IBM Tool für Datenbankdesign, Datenmodellierung und -integration für die InfoSphere Warehouse- und Cognos BI-Anwendungsentwicklung. Unterstützt logische, physische und dimensionale Modelle. Lässt sich in andere IBM InfoSphere Warehouse-Komponenten und Optim-Lösungen integrieren. |
| ADVANCED RECOVERY-LÖSUNGEN | |
| Optim High Performance Unload for DB2 | Ermöglicht das Entladen, Extrahieren und Neupartitionieren großer Datenmengen bei Tabellenfehlern oder beim versehentlichen Löschen von Tabellen. Beschleunigt den Wiederherstellungsprozess durch das direkte Extrahieren von Daten aus Gesamt-, Teil- und Deltasicherungen, wodurch die Datenbankengine umgangen wird. Ermöglicht Entladeoperationen aus mehreren Datenbankpartitionen und erlaubt Neupartitionierungen in einem Arbeitsgang. Die Produktionsabläufe werden dadurch nicht beeinflusst. |
| IBM DB2 Recovery Expert | Beschleunigt die Problembhebungs- und Wiederherstellungsprozesse. Identifiziert Fehlerquellen durch die Analyse von Datenbankassets wie Tabellenbereiche, Tabellen, Indizes und Daten. Durchsucht DB2-Protokolle, um festzustellen, wo der Fehler aufgetreten ist, und empfiehlt szenarioabhängige Fehlerbehebungsmaßnahmen. Die Fehlerbehebung kann durchgeführt werden, ohne die gesamte Datenbank wiederherstellen oder Produktionssysteme in den Offlinemodus setzen zu müssen. |
| IBM DB2 Merge Backup for Unload for DB2 | Reduziert die Sicherungszeiten und ermöglicht die schnellere Datenwiederherstellung nach einem Ausfall. Erlaubt Datenbankadministratoren, Teil- und Deltasicherungen mit Gesamtsicherungen zu kombinieren, wodurch weitere Gesamtsicherungen unnötig werden. Die Verarbeitung erfolgt im Offlinemodus, d. h. die Produktionsabläufe werden nicht beeinträchtigt. |
| WAREHOUSE-PACKS | |
| InfoSphere Warehouse Pack for Customer Insight | Ermöglicht die Analyse von über 30 Arten von kundenbezogenen Daten wie Daten zu Verkäufen, Retouren, Kundengewinnungskosten, Krediten und Zahlungen, Rentabilität, Aktivitäten zu Kundentreueprogrammen, Kundenmarktanteilen, Kundenbindung, Demografie und Warenkörben. |
| InfoSphere Warehouse Pack for Market & Campaign Insight | Ermöglicht Datenanalysen für Anwendungen für das Marketing-, Vertriebs- und Kampagnenmanagement. |
| InfoSphere Warehouse Pack for Supply Chain Insight | Ermöglicht die Analyse von Messdaten zu Lieferanten, Lagerbestand, Distribution, Prognosen, Kosten usw. für die Lieferkettenprozesse. |

Abbildung 8: Weitere InfoSphere Warehouse Advanced Enterprise Edition-Komponenten

Auch wenn Oracle viele vergleichbare Funktionen bietet, sind die DB2 9.7-Datenbank und andere InfoSphere Warehouse-Komponenten eindeutige Differenzierungsmerkmale. Dies zeigt sich besonders in den folgenden Bereichen.

Datenkomprimierung

Die Datenkomprimierung wirkt sich massiv auf die Kostenstrukturen aus. Die Größe von Datenbanken kann damit deutlich reduziert werden. Dadurch lassen sich Einsparungen bei der Prozessor-, Speicher- und E/A-Kapazität bei Datenbankservern, Platten- und Bandspeichern und anderen Bereichen erzielen.

Weitere Einsparungen lassen sich bei Systemen, Medien und beim Administratorkaufwand für Sicherungen, Replizierungen und anderen Datenmanagement- und -verschiebungsprozessen realisieren. Zudem können Prozesse wie Replizierungen, Sicherungen/Wiederherstellungen und andere

Datenverschiebungsprozesse beschleunigt werden. Hinzu kommen die Kosten für WAN-Bandbreiten, die deutlich verringert werden können.

Da Softwareprodukte für Zeitpunktkopien, ferne Replizierungen/Wiederherstellungen, Sicherungen, Speichermanagement und andere Funktionen in der Regel auf Pro-Terabyte-Basis abgerechnet werden, lassen sich in diesem Bereich zusätzliche Einsparungen bei den Lizenz- und Supportkosten erzielen. Kleinere Konfigurationen tragen im Normalfall dazu bei, auch die Gebäudekosten wie Raummiete, Strom- und Kühlungskosten für das Rechenzentrum zu senken.

DB2 9.7 bietet eine der leistungsfähigsten Komprimierungstechnologien in der Branche. Die Algorithmen sind wesentlich leistungsfähiger als die Oracle Database 11g-Pendants und können sowohl für permanente als auch temporäre Tabellen, Indizes, Protokolldateien, große Objekte (LOBs), Werte, XML-Daten und andere Strukturen verwendet werden. Benutzer erreichen im Normalfall Komprimierungsraten von 55 bis über 85 Prozent.

Im Vergleich dazu stellen Benutzer von Oracle Database 11g Advanced Compression immer wieder fest, dass es selbst bei hohen Komprimierungsraten bereits zu einem frühen Zeitpunkt zu inakzeptablen Leistungseinbußen kommt. Die DB2 9.7-Komprimierung belastet den Prozessor deutlich weniger.

Advanced Compression funktioniert am besten, wenn Datenbanken gut strukturiert sind und nur wenig geändert werden. In dynamischeren Umgebungen ist die Oracle Database 11g-Komprimierung oft nicht besonders effizient, es sei denn, es werden häufige Datenbankreorganisationen durchgeführt – was jedoch für viele Benutzer nicht praktikabel ist.

Leistungsoptimierungsprogramme (Performance Accelerators)

Sowohl IBM DB2 9.7 als auch Oracle Database 11g bieten Funktionen für Partitionierung und Clustering, wodurch ein hohes Leistungsniveau bei der Parallelverarbeitung und der Skalierbarkeit erzielt wird. Die IBM DB2 9.7-Funktionen sind hierfür im Allgemeinen jedoch effizienter.

Zu den wichtigen DB2 9.7-Akzeleratoren gehören DB2 Database Partitioning Feature (DPF) und Multi-Dimensional Clustering (MDC). Data-Warehouse-Benutzer, die diese Tools nutzen, berichten von Durchsatzverbesserungen um das Acht- bis Neunfache bei Prozessen mit sich wiederholenden Abfragen und/oder großen Tabellenbereichen.

Das MDC-Tool, das seine Stärken insbesondere in Umgebungen mit großen Datenbanken und Data-Warehouses zeigt, erlaubt ein kontinuierliches, flexibles und automatisches Clustering von Daten in verschiedenen Dimensionen. Benutzer, die diese Technologie einsetzen, konnten die Abfrageleistung in der Regel um das Dreifache beschleunigen. Einige Benutzer erzielten sogar Verbesserungen um mehr als das Zehnfache.

Mit MDC kann der Zeitaufwand von Datenbankadministratoren für Aufgaben wie Reorganisation und Indexverwaltung ebenfalls deutlich reduziert werden. Das Design ist so ausgelegt, dass selbst beim Einfügen und Löschen von Zeilen kein Leistungsabfall beim Clustering festzustellen ist.

Bei Oracle Database 11g wird ein Tabellen-Clustering-Feature implementiert, das vergleichbare Funktionalität bietet, aber nicht die Leistung und den Funktionsumfang von MDC erreicht.

Workload-Management

IBM und Oracle bieten umfassende Funktionalität für das Workload-Management in DB2 9.7 bzw. Database 11g. Beim Design zeigen sich jedoch einige gravierende Unterschiede.

Bei den Oracle-Tools, die auf Oracle Database Manager und der Database Control-Komponente von Oracle Enterprise Manager aufbauen, kommt ein auf den Datenbankadministrator zugeschnittener Ansatz zum Tragen. Im Vergleich dazu werden bei DB2 9.7 Workload Manager technische Parameter angepasst, um die geschäftsspezifischen Prioritäten zu erfüllen.

Die Funktionen dieses Tools sind von der Mainframearchitektur abgeleitet und bauen auf den Mainframefähigkeiten beim Management unterschiedlicher, gleichzeitig auftretender Workloads auf, um die SLA-Ziele einzuhalten.

DB2 9.7 Workload Manager ermöglicht eine hoch differenzierte, automatisierte Priorisierung, Ressourcenzuordnung, Warteschlangensteuerung, Echtzeitüberwachung und Verwaltung von Workloads, die durch Hunderte oder Tausende separater Abfragejobs generiert werden. Prioritäten können auf Basis von Anwendungen, Benutzergruppen, Abfragetypen, Tageszeit und/oder anderen Variablen vergeben werden.

Dieses Tool lässt sich zudem in AIX Workload Manager auf IBM Power Systems integrieren, sodass Administratoren Server- und Datenbankressourcen detailliert zuordnen können. Linux on Power-Ressourcen können ebenfalls an den über DB2 Workload Manager festgelegten Prioritäten ausgerichtet werden.

Solche Funktionen können speziell in komplexen Workloadumgebungen nützlich sein und erfüllen insbesondere die hohen Anforderungen bei der Verwaltung von Data-Warehouses in Unternehmen.

Automatisierung

Die Möglichkeit, Aufgaben eines Datenbankadministrators zu automatisieren, bietet eine Vielzahl von Vorteilen. Der offensichtlichste Vorteil ist dabei die höhere Produktivität, die sich dadurch erzielen lässt. Eine Automatisierung kann auch zu Leistungsverbesserungen (die Nutzung von Systemressourcen wird in Echtzeit optimiert) und einer höheren Verfügbarkeit (geringeres Potenzial für Benutzerfehler) führen.

IBM und Oracle bieten beide entsprechende Automatisierungsfunktionen, wobei die Funktionalität der IBM Lösung allgemein als umfangreicher erachtet wird. Bestimmte DB2 9.7-Features, wie die in Abbildung 9 gezeigten, profitieren von den Stärken der IBM Autonomic Computing-Lösungen.



Abbildung 9: Wichtige Autonomic-Features in DB2 9.7

Autonomic Computing, d. h., die Anwendung innovativer Künstliche-Intelligenz-Technologien für IT-Verwaltungs- und Optimierungsaufgaben, ist schon seit über zehn Jahren ein Schwerpunktthema in der IBM Entwicklung. Das Unternehmen ist in diesem Bereich eindeutiger Branchenführer.

Benutzer haben festgestellt, dass sich DB2 Self-Tuning Memory Manager (STMM) hierbei besonders nützlich erweist. STMM passt Speicherkonfigurationsparameter und Pufferpoolgrößen automatisch an, sobald sich die Workloadmerkmale ändern. Dieses Feature, hinter dem eine der branchenweit innovativsten Selbstoptimierungstechnologien steht, sorgt für eine kontinuierliche Leistungsoptimierung, ohne dass der Datenbankadministrator eingreifen muss.

IBM Optim-Tools

Drei InfoSphere Warehouse-Tools bieten besonders offensichtliche Alleinstellungsmerkmale.

1. **IBM Optim Performance Manager Extended Edition** ermöglicht die schnelle Erkennung, Diagnose und Lösung von Engpässen bei der Datenbank- und Anwendungsleistung. Mithilfe der ermittelten Daten lassen sich zudem zu erwartende Probleme bereits im Vorfeld vorhersagen und es lässt sich verhindern, dass Fehler in neuen Anwendungen erneut auftreten.

Eine zentrale Funktion dieser Lösung ist Guided Problem Solving, mit der Datenbankadministratoren online Unterstützung von Experten zu dem jeweiligen Problem oder der betreffenden Konfiguration erhalten können.

Viele Benutzer haben berichtet, dass mithilfe dieser Funktion der Zeitaufwand für die Handhabung von Engpässen deutlich reduziert werden konnte (in einigen Fällen von mehreren Tagen auf wenige Stunden oder sogar Minuten). Zudem konnten weitere Vorteile wie höhere Verfügbarkeit (Reduzierung der Häufigkeit und Dauer von Ausfällen) und Mitarbeiterproduktivität erzielt werden.

2. **IBM Optim Query Tuner** bietet sowohl Datenbankadministratoren als auch Entwicklern die Möglichkeit, potenzielle Abfrageprobleme während des Entwicklungsprozesses bereits im Vorfeld zu erkennen und zu beheben, anstatt sie reaktiv zu beheben, wenn Leistungsengpässe oder Ausfälle auftreten. Auch dieses Tool bietet umfassende Möglichkeiten, Unterstützung von Experten online in Anspruch zu nehmen.
3. **IBM Optim Database Administrator** automatisiert zentrale Prozesse bei der Definition und Implementierung von Änderungen an Datenbankschemata. Mit diesem Tool können Änderungen mit bestehenden Konfigurationsdaten, Unternehmensstandards und Kompatibilitätsanforderungen abgeglichen werden. Zudem lassen sich Auswirkungen auf die Datenbank vorwegnehmen. Daraus ergeben sich ebenfalls Vorteile wie höhere Verfügbarkeit und höhere Produktivität beim Datenbankadministrator.

Die Funktionalität dieser Tools erweist sich insbesondere beim Data-Warehousing als außerordentlich hilfreich. Schnelles Wachstum und Veränderungen bei Anwendungen und Workloads machen die manuelle Leistungsoptimierung zu einem sehr aufwendigen und fehlerbehafteten Prozess. Gleichmaßen treten Änderungen bei Abfragetypen und Datenbankschemata deutlich häufiger auf als bei stabileren Systemen.

Einige der genannten Funktionen sind auch in den Oracle Diagnostics- und Change Management-Packs for Database 11g zu finden. Die Funktionen der Optim-Tools sind jedoch überzeugender und die Automatisierungstechnologien fortschrittlicher.

SQL-Kompatibilität

IBM DB2 9.7 umfasst Funktionen, mit denen die Benutzer Anwendungen von Datenbanken anderer Anbieter vergleichsweise einfach migrieren können. Die Risiken, die bei solchen Operationen normalerweise auftreten können, sind dabei als deutlich geringer einzustufen..

DB2 9.7 unterstützt Oracle Procedural Language/Structured Query Language (PL/SQL) und den Oracle SQL-Dialekt zusammen mit einer Vielzahl von Codes, Tools und Funktionen in Oracle-Umgebungen. Laut Angaben von IBM berichten Unternehmen, die Oracle-Anwendungen auf DB2 9.7 migriert haben, dass 90 bis 99 Prozent des Codes nicht verändert wurden. In den meisten Fällen waren nur wenige oder keine Änderungen an bestehenden Oracle-Entwicklungstools bzw. weitere Kenntnisse erforderlich.

Kompatibilitätsfunktionen sind in die DB2 9.7-Engine integriert und müssen nicht als Software-Overlays implementiert werden. Unternehmen erzielen in der Regel dieselben hohen Leistungswerte wie native DB2 9.7-Benutzer.

IBM SQL Skin for Sybase, das zum Lieferumfang der InfoSphere Warehouse-Pakete gehört, bietet vergleichbare Funktionen für die Migration von Sybase Adaptive Server Enterprise (ASE).

XML-Unterstützung

Extensible Markup Language (XML) wird mittlerweile in zahlreichen Branchen wie Finanzwesen, Versicherungswesen, Einzelhandel, Gesundheitswesen usw. eingesetzt. Für textbasierte Data-Warehouse-Inhalte ist XML mittlerweile praktisch zur Norm geworden.

Obwohl sowohl IBM DB2 9.7 als auch Oracle Database 11g die Integration von XML-Inhalten unterstützen, gibt es bei der Implementierung doch deutliche Unterschiede.

DB2 9.7 pureXML bietet Unterstützung für XML-basierten nativen Speicher, Indexierung, Abfragen, Aktualisierungen und das Datenmanagement. Das gesamte Funktionsspektrum von DB2 9.7 – einschließlich Partitionierung, mehrdimensionales Clustering, Komprimierung, Abfrageoptimierung und Automatisierung – steht nun auch für XML-Dateninhalte zur Verfügung.

Die Einbettung der XML-Funktionalität ist deutlich intensiver als bei Oracle Database 11g und zudem besser optimiert. Benutzer berichten insbesondere bei komplexen Datenbanken und Abfragen von einem deutlich besseren Leistungsverhalten bei DB2 9.7.

Die Erfahrungswerte mit DB2 9.7 lassen zudem den Schluss zu, dass Unterschiede bei der Produktivität von Datenbankadministratoren bei XML-spezifischen Aufgaben deutlicher zu Tage treten als bei reinen relationalen Umgebungen. Dies gilt insbesondere dann, wenn häufige Änderungen an Datenstrukturen vorgenommen werden müssen. Die XML-Funktionalität von DB2 9.7 XML wird in vollem Umfang von allen InfoSphere Warehouse-Versionen unterstützt.

Oracle-Lösungen

Neben Oracle Database 11g bietet das Unternehmen weitere Lösungen an, die in Data-Warehouse-Umgebungen verwendet werden können.

Oracle Business Intelligence Suite Enterprise Edition Plus (OBIEE) beispielsweise ist ein integriertes Lösungspaket mit Oracle-, Siebel- und Hyperion-BI-Tools. Die OBIEE-Module bieten Funktionen für Adhoc-Abfragen und -Berichterstellung, OLAP, Data-Mining, Überwachung von Geschäftsaktivitäten mit Alertausgabe, Finanzberichterstellung für Geschäftsaktivitäten, Dashboarderstellung usw..

Oracle bewirbt zudem ganz intensiv seine **Exadata Database Machine** für Data-Warehousing-Umgebungen. Laut Angaben des Unternehmens wurden bis jetzt zwei Drittel der über 1.000 Exadata Database Machine-Lösungen in diesen Umgebungen installiert (das verbleibende Drittel unterstützt nur Oracle-Datenbankkonsolidierungen sowie Oracle E-Business Suite, Oracle Retail und andere Anwendungen für die Transaktionsverarbeitung).

Obwohl Oracle Exadata Database Machine von Oracle auf breiter Front positioniert wird, ist diese Lösung doch in erster Linie für die Durchführung von Suchoperationen in großvolumigen sequenziellen Tabellen vorgesehen, die durch Anwendungen generiert werden. Diese sind in der Regel einfach strukturiert, erfordern jedoch sehr viel Verarbeitungsleistung, z. B. für die Erkennung und Sortierung von speziellen Variablen in großen Datensatzmengen.

Die Exadata Database Machine-Architektur sowie Technologien wie Exadata Hybrid Columnar Compression eignen sich für solche Workloads am besten.

Workloads mit intensiven Tabellensuchoperationen helfen im Allgemeinen nur einer kleinen Anzahl von Führungskräften und/oder Analysten weiter. Breit gestreutere Benutzergruppen, wie sie in den meisten unternehmensweiten Data-Warehouse-Umgebungen zu finden sind, nutzen diese Workloads relativ wenig. Die Oracle Exadata Database Machine-Funktionen für das Management gemischter Workloads sind vergleichsweise schwach.

Wie viele andere High-End-Data-Warehouse-Appliances ist Oracle Exadata Database Machine teuer in der Anschaffung. Unternehmen müssen nicht nur in Exadata Storage Server- und Database Machine-Hardware- und Software investieren, sondern auch in den Oracle-Software-Stack, wie in Abbildung 10 gezeigt.



Abbildung 10: Oracle Software-Stack für Exadata Database Machine

Aus einer früheren Studie der Verfasser des vorliegenden Berichts geht hervor, dass die kombinierten Anschaffungskosten selbst bei Einzelrack-Konfigurationen mit hohen Nachlässen in der Regel zwischen \$ 650.000 und \$ 4 Millionen lagen. Die Kosten über einen Zeitraum von drei Jahren gesehen, einschließlich Hardwarewartung und Software-Support, beliefen sich auf \$ 2 bis \$ 9 Millionen.

BERECHNUNGSGRUNDLAGEN

Die Berechnungen in diesem Bericht basieren auf repräsentativen Data-Warehouse-Installationen in großen und mittelständischen Unternehmen. Die Angaben zur Data-Warehouse-Größe beziehen sich auf unkomprimierte Daten und die Anzahl der Benutzer auf Systeme, die mindestens zwei Jahre alt sind.

Die Berechnungen beinhalten die anfänglichen Lizenzkosten sowie die Supportkosten über einen Zeitraum von drei Jahren und basieren auf den Listenpreisen der Anbieter. In den Übersichten zu den durchschnittlichen Kosten in den Abbildungen 1 und 2 wurden die Nachlässe auf die Listenpreise wie angegeben berücksichtigt. Bei der Aufschlüsselung der Installationskosten in den folgenden Abbildungen 11 und 12 wurden in die Kosten keine Nachlässe einbezogen. Die Kosten für Oracle Database 11g Enterprise Edition basieren auf der kostengünstigsten Option für jede Installation.

Die Kosten für Oracle Database 11g Enterprise Edition wurden für Oracle-Kernfaktoren angepasst – aufgrund des Preismodells des Unternehmens zählt Oracle Intel-Kerne als halben Prozessor und IBM POWER7-Kerne als jeweils ganzen Prozessor. Bei Berechnungen auf Benutzerbasis wurde beim Oracle-Preis das Minimum von 25 Benutzern pro Kern angenommen. Bei Installationen zwischen 1 TB und 5 TB liegt die Mindestanzahl der Benutzer über der tatsächlich zugrunde gelegten Anzahl.

Die IBM InfoSphere Warehouse-Preise basieren auf dem nächsthöheren 1-TB-Inkrement. Die Lizenz- und Supportkosten für 0,5 TB, 2,5 TB und 12,5 TB an komprimierten Benutzerdaten wurden auf 1 TB, 3 TB und 13 TB aufgerundet.

Die Oracle-Kosten können in der Praxis höher liegen als hier gezeigt. Um dieselbe Funktionalität wie Optim Database Administrator zu erhalten, ist in vielen Fällen Oracle Change Management Pack erforderlich, das einen Listenpreis von \$ 3.500 pro Prozessor oder \$ 70 pro benanntem Benutzer aufweist.

| DATA-WAREHOUSE | | | |
|--|------------------------------|-------------------------------|----------------------------|
| Benutzerdaten | 10 TB | 25 TB | 50 TB |
| Anzahl der Benutzer | 1,000 | 3,000 | 5,000 |
| Anzahl der Prozessoren | 2/16 x 3.2 GHz IBM POWER7 | 4/32 x 3,86 GHz IBM POWER7 | 8/64 x 4 GHz IBM POWER7 |
| ORACLE DATABASE 11g ENTERPRISE EDITION | | | |
| Mindestanzahl an Benutzern | 400 | 850 | 1,600 |
| Lizenzkosten pro Benutzer | \$ 2,530,000 | \$ 7,590,000 | \$ 12,650,000 |
| Kosten über 3 Jahre inkl. Supportkosten | \$ 4,199,800 | \$ 12,599,400 | \$ 20,999,000 |
| Lizenzkosten pro Prozessor | \$ 2,024,000 | \$ 4,048,000 | \$ 8,096,000 |
| Kosten über 3 Jahre inkl. Supportkosten | \$ 3,359,840 | \$ 6,719,680 | \$ 13,439,360 |
| IBM INFOSPHERE WAREHOUSE ENTERPRISE EDITION | | | |
| UNKOMPRIMIERTE DATEN | | | |
| Lizenzkosten pro Terabyte | \$ 700,000 | \$ 1,750,000 | \$ 3,500,000 |
| Kosten über 3 Jahre inkl. Supportkosten | \$ 1,120,000 | \$ 2,800,000 | \$ 5,600,000 |
| 50 % DATENKOMPRIMIERUNG | | | |
| Lizenzkosten pro Terabyte | \$ 350,000 | \$ 910,000 | \$ 1,750,000 |
| Kosten über 3 Jahre inkl. Supportkosten | \$ 560,000 | \$ 1,456,000 | \$ 2,800,000 |

Abbildung 11: Kostenvergleich – Installationen zwischen 10 TB und 50 TB

| DATA-WAREHOUSE | | | |
|--|------------------|-------------------|-------------------|
| Benutzerdaten | 1 TB | 2 TB | 5 TB |
| Anzahl der Benutzer | 60 | 150 | 400 |
| Anzahl der Prozessoren | 2/8 x Intel 5600 | 2/12 x Intel 5600 | 2/16 x Intel 7500 |
| ORACLE DATABASE 11g ENTERPRISE EDITION | | | |
| Mindestanzahl an Benutzern | 100 | 150 | 500 |
| Lizenzkosten pro Benutzer | \$ 253,000 | \$ 379,500 | \$ 1,265,000 |
| Kosten über 3 Jahre inkl. Supportkosten | \$ 419,980 | \$ 629,970 | \$ 2,099,900 |
| Lizenzkosten pro Prozessor | \$ 506,000 | \$ 759,000 | \$ 1,012,000 |
| Kosten über 3 Jahre inkl. Supportkosten | \$ 839,960 | \$ 1,259,940 | \$ 1,679,920 |
| IBM INFOSPHERE WAREHOUSE DEPARTMENTAL EDITION | | | |
| UNKOMPRIMIERTE DATEN | | | |
| Lizenzkosten pro Terabyte | \$ 35,000 | \$ 70,000 | \$ 175,000 |
| Kosten über 3 Jahre inkl. Supportkosten | \$ 56,000 | \$112,000 | \$ 280,000 |
| 50 % DATENKOMPRIMIERUNG | | | |
| Lizenzkosten pro Terabyte | \$ 35,000 | \$ 35,000 | \$ 105,000 |
| Kosten über 3 Jahre inkl. Supportkosten | \$ 56,000 | \$ 56,000 | \$ 168,000 |

Abbildung 12: Kostenvergleich – Installationen zwischen 1 TB und 5 TB

INFORMATIONEN ZUR INTERNATIONAL TECHNOLOGY GROUP

*Die International Technology Group (ITG) trägt dazu bei, Ihr Bewusstsein für aktuelle Entwicklungen und Ihre Wettbewerbsfähigkeit zu schärfen
... Voraussetzungen für zukünftiges Wachstum und Gewinnschätzungen*

Die im Jahr 1983 gegründete International Technology Group (ITG) ist ein unabhängiges Forschungs- und Management-Consulting-Unternehmen, das sich auf Investitionsstrategien im IT-Bereich, Kosten-Nutzen-Analysen, Infrastrukturstudien, Implementierungstaktiken, die Ausrichtung von Geschäftsprozessen und Finanzanalysen spezialisiert hat.

Die ITG leistete als innovatives Unternehmen bei der Entwicklung von Prozessen und Methodiken im Zusammenhang mit den Gesamtbetriebskosten und dem Return-on-Investment schon in einem frühen Stadium Pionierarbeit. Im Jahr 2004 erhielt das Unternehmen von der Information Technology Financial Management Association (ITFMA), dem führenden Fachverband für Schulungen und die Weiterentwicklung von Managementverfahren im Finanzwesen in IT-Organisationen den „Decade of Education Award“.

Das Unternehmen hat bereits mehr als 120 bedeutende Consultingprojekte durchgeführt, über 250 Managementberichte und White Papers veröffentlicht und über 1.800 Kurzinformationen und Präsentationen für Kunden, Benutzergruppen, branchenspezifische Konferenzen und Seminare auf der ganzen Welt verfasst.

Die Dienstleistungen des Unternehmens sollen den Kunden durch Fakten und zuverlässige Dokumentationen beim Entscheidungsfindungsprozess unterstützen. Die bereitgestellten Informationen bilden die Grundlage für die Erarbeitung taktischer und strategischer Pläne. Wichtige Entwicklungen werden umfassend analysiert und durch effektive praxisorientierte Hilfestellungen kann der Kunde schnell auf Veränderungen reagieren, die sich auf den komplexen Ablauf bei IT-Implementierungen auswirken können.

Das Unternehmen bietet ein breites Spektrum an Dienstleistungen an, durch die dem Kunden die Informationen bereitgestellt werden können, die er braucht, um sein internes Leistungsspektrum und Ressourcenpotenzial sinnvoll zu ergänzen. Individuell angepasste Kundenprogramme beinhalten verschiedene Kombinationen aus den folgenden Materialien:

| | |
|--|---|
| Statusberichte | Fundierte Studien zu wichtigen Problemstellungen |
| Kurzübersicht für das Management | Detaillierte Analyse signifikanter Entwicklungen |
| Management-Briefings | Regelmäßige interaktive Besprechungen mit dem Management |
| Präsentationen für die Geschäftsleitung | Zeitlich geplante strategische Präsentationen für Entscheidungsträger |

E-Mail-Kommunikation

Zeitgerechte Beantwortung von
Informationsanfragen

Telefonische Beratung

Sofortige Reaktion auf den Informationsbedarf
des Kunden

Der Kundenstamm des Unternehmens setzt sich zusammen aus einem Querschnitt von IT-Unternehmen aus dem öffentlichen und privaten Sektor, internationalen Unternehmen, Industriegesellschaften, Finanzinstituten, Serviceorganisationen, Bildungseinrichtungen, Behörden auf Bundes- und Länderebene sowie Anbietern von IT-Systemen, Software und Dienstleistungen. Zu den Kunden auf Bundesebene gehören Abteilungen im Verteidigungsministerium (z. B. DISA), Verkehrsministerium (z. B. FAA) und Finanzministerium (z. B. US Mint).

International Technology Group

609 Pacific Avenue, Suite 102
Santa Cruz, California 95060-4406
Telefon: + 831-427-9260
E-Mail: Contact@ITGforInfo.com
Website: ITGforInfo.com