



BUSINESS
APPLICATION
RESEARCH
CENTER

WHITE PAPER

Abfrageperformance: Erfolgsfaktor für Business Intelligence



Autoren



Timm Grosser
Analyst
BARC GmbH



Carsten Bange
CEO
BARC GmbH



Melanie Mack
Leiterin
Marktforschung
BARC GmbH

Business Application Research Center – BARC GmbH

Steinbachtal 2b
97082 Würzburg
+49 (0)931 880651-0
info@barc.de



Die Studie wurde völlig unabhängig vom neutralen Marktanalysten BARC erstellt. Dank eines Sponsorings durch Netezza kann die Studie kostenfrei zur Verfügung gestellt werden.

Erfolgsfaktor Abfrageperformance

Anforderungen an Business-Intelligence- und Data-Warehouse-Systeme wachsen beständig. Datenvolumina steigen exponentiell, immer mehr Anwender haben umfangreichere Informationsbedarfe und die Toleranz für Wartezeiten auf Datenabfragen sinkt in einer immer schneller werdenden Geschäftswelt rapide. Die Steuerungssysteme eines Unternehmens müssen die wachsende Dynamik und Komplexität der operativen Prozesse abbilden können – bei hohen Anforderungen an Kosteneffizienz, Skalierbarkeit und Performance.

Gerade die Performance gilt nachweislich als kritischer Faktor für erfolgreiche Business Intelligence (BI). Am deutlichsten zeigt dies die weltweit größte Befragung von BI-Anwendern, der BI-Survey. Die Umfrage zeigt, dass schlechte Abfragegeschwindigkeit seit 2004 das Top-Problem von Business-Intelligence-Projekten ist, noch vor schlechter Datenqualität und Unternehmenspolitik. Dass sich hier trotz wachsender Leistungsfähigkeit von Hardwarekomponenten und BI-Software keine Verbesserung eingestellt hat verschiedene Gründe: Erstens wächst häufig die Komplexität von BI-Systemen hinsichtlich Anwenderzahl, Nutzung und Datenmengen mindestens ebenso schnell wie Soft- und Hardwareverbesserungen, andererseits existiert auch eine „subjektive Performanceerwartung“ der Anwender, die immer anspruchsvoller wird. Ein relativ hoher Verarbeitungstakt von Information im Berufsalltag (man denke nur an die Anzahl von E-Mails, die jeden Tag bearbeitet werden) sowie der Umgang mit Systemen wie der Google Suchmaschine, die den Zugriff auf riesige Datenmengen in Sekundenbruchteilen ermöglicht, wird als selbstverständlich empfunden. Das setzt den Benchmark auch für analytische Abfragen an BI-Systeme.

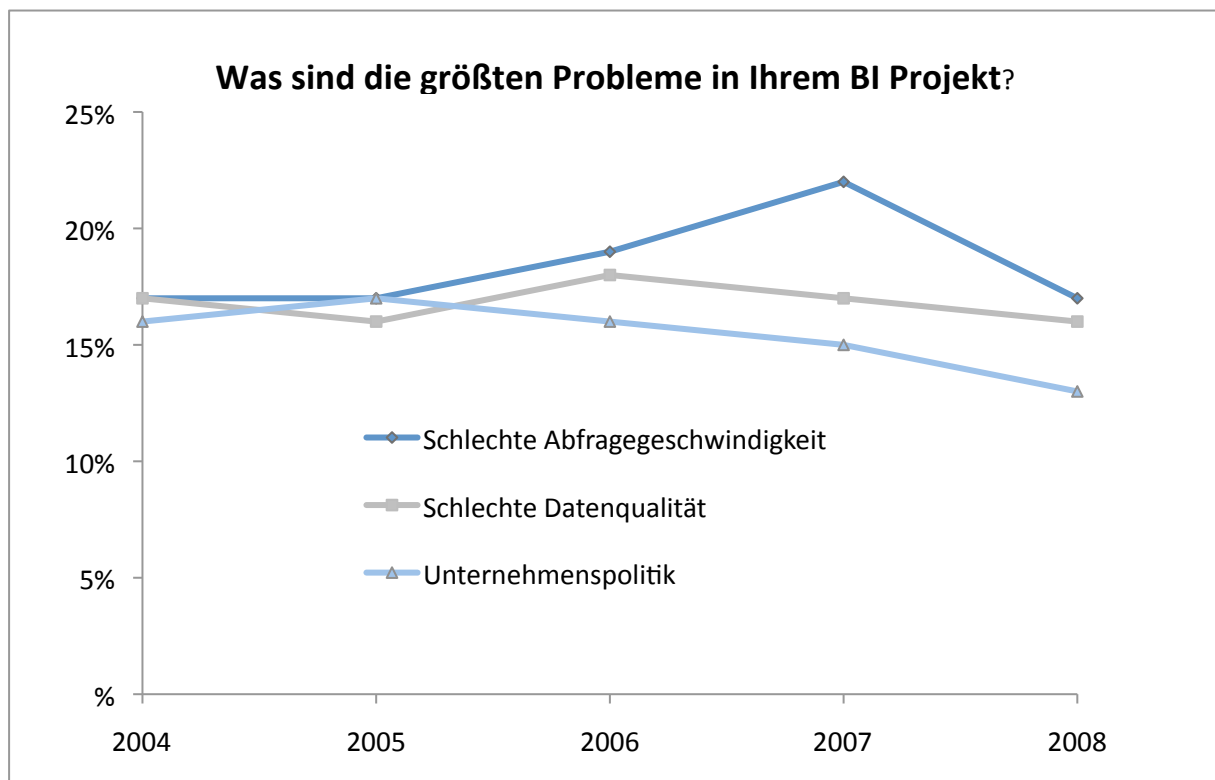


Abbildung 1: Schlechte Abfragegeschwindigkeit ist das Top-Problem in Business-Intelligence-Projekten (Quelle: BI-Survey 8, n=1965)

Performance eines Business-Intelligence-Systems

Die Performance eines Business-Intelligence-Systems setzt sich aus zwei Bereichen zusammen:

- Die **Abfrageperformance** oder auch Antwortzeit/-geschwindigkeit bezeichnet die Zeitspanne zwischen dem Absenden einer Anfrage und der vollständigen Ergebnisaufbereitung für den Anwender.
- Die **Ladepformance** bezeichnet weitergefasst die benötigte Laufzeit von der Datenextraktion aus Vorkontrollsystemen bis zur Bereitstellung der Daten im BI-System. Der Ladeprozess beeinflusst auch unmittelbar die Abfragegeschwindigkeit, wenn notwendige Daten für eine Abfrage zunächst durch einen Ladeprozess beschafft werden müssen.

Aufgrund der Globalisierung und der damit verbundenen weltweiten Abfrage und Integration von Daten rund um die Uhr, fallen im Gegensatz zu klassischen Systemen heute oftmals Abfragezeit und Ladezeit zusammen.

Zu lange Wartezeiten frustrieren Anwender und mindern die Akzeptanz des BI-Systems. Erst eine hohe Nutzerfreundlichkeit und das Vertrauen der Anwender in eine schnelle und qualitätsgesicherte Informationsversorgung erzeugen den gewünschten Nutzen aus analytischen Systemen.

Auswirkungen schlechter Performance

Eine **schlechte Performance** hingegen

- **senkt die Nutzung** von BI-Systemen und führt zu abteilungsspezifischen Datenhaltungs- und Auswertungssystemen, oftmals Excel. Damit droht ein Wildwuchs an Datensilos und Kennzahldefinitionen.
- **verhindert die Verbreitung** von BI im Unternehmen – sowohl aus technischer als auch aus Anwender-Perspektive. Systeme mit Performanceengpässen müssen erst verbessert werden, bevor mehr Last durch weitere Anwender, Daten und Applikationen erzeugt wird.
- **senkt den Nutzen** einer BI-Lösung, da neue Erkenntnispotentiale unentdeckt bleiben, wenn Anwender wegen schlechter System-Performance tendenziell weniger und einfachere Fragen an das BI-System stellen. Komplexere Fragestellungen mit potentiell interessanteren und nützlicheren Ergebnissen können nicht bearbeitet werden.

Quintessenz: Schlechte Performance gefährdet den Erfolg eines BI-Systems und kann damit zu Investitionsruinen führen.

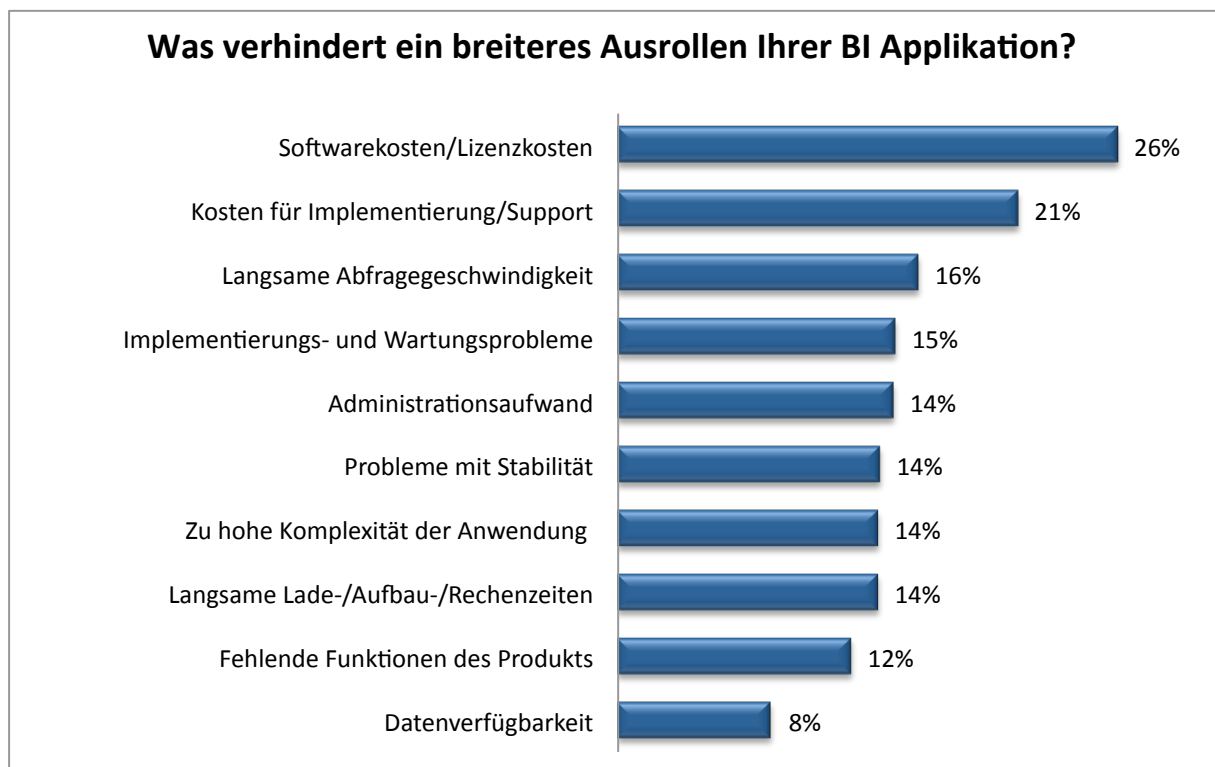


Abbildung 2: Sowohl schlechte Abfrageperformance als auch die Ladeperformance wurden von den Umfrageteilnehmern des BI-Survey 8 als Hindernis für die Verbreitung von BI im Unternehmen benannt. (Quelle: BI-Survey 8, n=1623)

Einflussfaktoren auf Performance

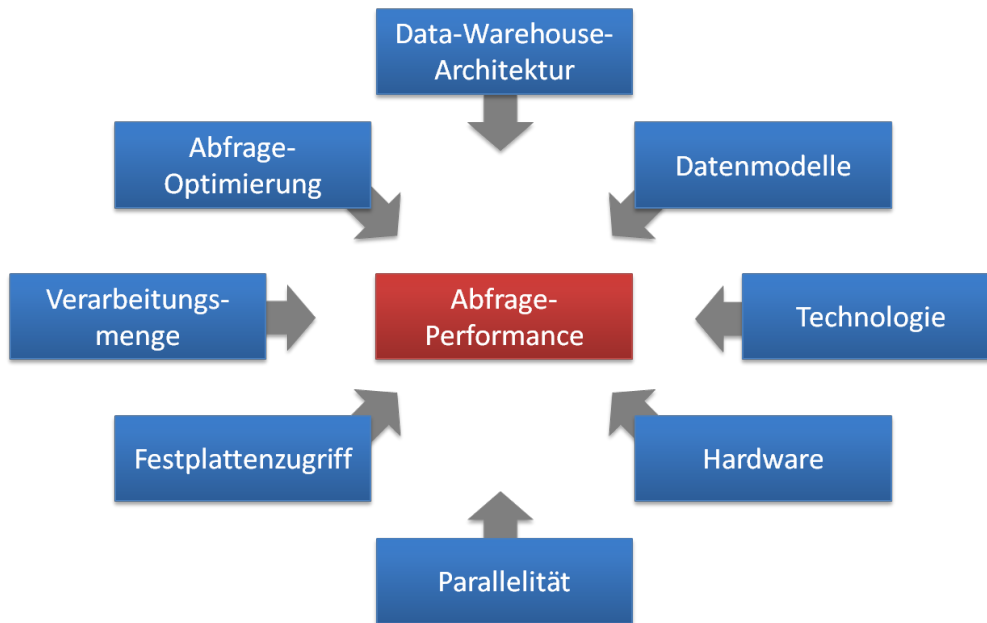


Abbildung 3: Einflussfaktoren für Abfrageperformance (Copyright BARC)

Zahlreiche Faktoren beeinflussen die Abfrageperformance und machen das Thema hoch komplex. Die wichtigsten Stellschrauben sind folgend kurz erläutert.

- Die **Data-Warehouse-Architektur** im Sinne einer Topologie definiert die Informationsflüsse und Datenbereitstellung im Unternehmen konzeptionell und wirkt damit auf die Performance und auch viele der nachfolgenden Punkte. Bekannte Data-Warehouse-Architekturen sind das zentrale Enterprise Data Warehouse (EDW), abhängige oder unabhängige Marts, föderierte Data-Warehouse oder Mischarchitekturen.
- Spezielle **Datenmodelle** eignen sich für unterschiedliche Applikationen und beeinflussen Erweiterbarkeit, Flexibilität und Performance. Für die analytische Informationsverarbeitung eignen sich besonders dimensionale Modellierungsansätze wie das Star- oder Snowflake-Schema in relationalen Datenbanken sowie die Modellierung von multidimensionalen Würfeln.
- Die eingesetzten **Technologien** entscheiden über die Leistungsfähigkeit des gesamten BI-Systems. Das Zusammenspiel von Datenintegration, Datenbank und Business-Intelligence-Werkzeug definiert die Performance.
- Die unterstützten **Hardware-Architekturen** determinieren Performance und Skalierbarkeit der Systeme. Vor allem die Unterstützung von Symmetric-Multi-Processing (SMP)- oder Mas-

sive Parallel Processing (MPP)-Architekturen ermöglicht Performancesteigerung durch Parallelität.

- **Parallelität** subsumiert Funktionen und Vorgehen für die Performancesteigerung im Sinne einer parallelen Verarbeitung.
- Der **Festplattenzugriff** gilt als Engpassfaktor im Kontext von Prozessorleistung und Arbeitsspeicher. Funktionen für Kompression sowie Caching-Mechanismen beeinflussen die Lese- und Schreibgeschwindigkeit von Daten.
- Wachsende **Verarbeitungsmengen** erhöhen die Komplexität die passenden Daten zu finden und zu verarbeiten. Funktionen wie Indizierung, Aggregate oder materialisierte Sichten helfen bei der Reduktion dieser Mengen.
- Die Art und Weise der Datenbankabfragen beeinflusst die Performance und variiert mit den Abfragetypen. Eine **Abfrageoptimierung** wird manuell oder aber von Datenbankfunktionen – Query Optimizer – durchgeführt.

Ein weiterer wichtiger Einflussfaktor für Abfrageperformance ist die Komplexität des Gesamtsystems in Form von sehr vielen Datenspeichern, Datenintegrationsprozessen und einer Kombination vieler Werkzeuge verschiedener Hersteller, häufig für ähnliche Aufgaben. Viele Unternehmen arbeiten hier an einer Konsolidierung solch stark heterogener BI-Systeme um eine einfachere Wartbarkeit, schnellere Anpassungsfähigkeit und bessere Performance zu ermöglichen.



Datenbank-Technologien für Business Intelligence

Business-Intelligence-Lösungen stellen typischerweise besondere Anforderungen an Datenbanksysteme, die von transaktionsorientierten Datenbanken nicht immer gut erfüllt werden können. Herausforderungen liegen in der Abfrage und Berechnung sehr großer Datenmengen, gemischten Abfrage- und Ladelasten oder den besonderen Skalierbarkeits- und Performanceanforderungen. Aus diesen Gründen haben sich in den letzten Jahren zahlreiche Lösungsansätze auf den Markt etabliert, die besonderen Anforderungen analytischer Systeme besser abdecken zu können. Häufig werden sie auch kombiniert eingesetzt.

- **Data-Warehouse-Appliances** sind vorinstallierte Datenbanksysteme mit abgestimmten Hardwarekomponenten, die speziell für analytische Aufgaben optimiert sind. In einigen Lösungsansätzen übernehmen dedizierte Hardware-Bausteine spezielle Datenbankaufgaben und steigern dadurch die Performance.
- **In-Memory-Datenbanken** halten die gesamten Daten dank durchdachter Kompressionskonzepte im Hauptspeicher. Der Flaschenhals „Festplattenzugriff“ wird nur noch für die Sicherung der Datenbestände in festgelegten Intervallen notwendig.
- **Solid-State-Disks (SSD)** sind Speichermedien, die einen direkten wahlfreien Zugriff auf einen Datensatz ermöglichen. Der langwierige Suchprozess nach passenden Datensätzen auf Festplatten entfällt. Erste Datenbanken unterstützen bereits SSD-Speicher. Sie sind aus Performance-Sicht zwischen Festplatte und Hauptspeicher angesiedelt.
- **Multidimensionale Datenbanken** sind speziell für analytische Fragestellungen optimierte Datenhaltungssysteme. Daten werden in einen mehrdimensionalen Matrizen („Würfeln“) abgespeichert, dessen Kanten die Dimensionen repräsentieren. Kennzahlen können direkt adressiert und ausgelesen werden ohne zeitaufwendige Tabellenverknüpfungen (Join-Abfragen) ausführen zu müssen. Im Gegensatz zu SQL-Abfragen auf relationalen Datenbanken werden Anfragen an Würfel in proprietären Sprachen oder in Multidimensional Expressions (MDX) formuliert.
- **Spalten-orientierte relationale Datenbanken** speichern die Daten im Gegensatz zu den weit verbreiteten zeilen-orientierten Datenbanken in Spalten. Gerade in Datenbeständen mit sehr vielen unterschiedlichen Attributen pro Entität bieten diese Datenbanken zum Einem performante Zugriffe aufgrund höherer Kompressionsraten, zum Anderen beinhaltet jede Spalte eine Indizierung in sich und ermöglicht dadurch eine schnelle Filterung/Sortierung der Daten.
- **Caching-Mechanismen** halten häufig nachgefragte Daten oder Ergebnisse im Hauptspeicher vor und reduzieren damit die zeitintensiven Festplattenzugriffe sowie redundante Berechnungen. Query-Optimizer bestimmen in der Datenbank wo, wann und mit was Caches befüllt werden und leiten eingehende Abfragen entweder direkt gegen die Daten oder auf einen Cache. Caching-Mechanismen werden nicht nur von Datenbanken angeboten, auch in Business-Intelligence- oder Datenintegrationswerkzeugen können diese zur Performancesteigerung des Gesamtsystems eingesetzt werden.
- **Weitere Technologien** wie assoziative Datenbanken nutzen spezielle Indizierungsformen zur Performancesteigerung.

Empfehlungen

Performance ist einer der wichtigsten Aspekte für Business-Intelligence-Projekte. Wegen der zahlreichen Einflussfaktoren auf Performance sollten sie alle in einer Systemauswahl oder -optimierung berücksichtigt werden. Für die Auswahl des geeigneten Systems sollten Sie folgende **Empfehlung** beachten:

- Ebenso wie die Performance ist die **Skalierbarkeit** des Systems von Bedeutung um auch zukünftig angemessen auf neue Fragestellungen flexibel und schnell reagieren zu können. Die Skalierbarkeit beschreibt die Fähigkeit des Systems auf die wachsende Zahl an Anwender und Abfragen, die steigenden Datenvolumina, auf die komplexere Abfragen und Datenmodellen zu reagieren.
- Die Bedeutung von Performance und Skalierbarkeit erfordert bei der Auswahl von Lösungen diese Aspekte an eigenen, repräsentativen **Testscenarien** gründlich zu untersuchen. An Aussagekraft gewinnt der Test, wenn er mit eigenen Daten in einer authentischen IT-Umgebung abläuft.
- Wertvolle Informationen erhalten Sie von **Referenzkunden** der Anbieter, die oftmals gerne bereit sind über den Einsatz, Performance und Skalierbarkeit der Software Auskunft zu geben.
- Berücksichtigen Sie neben den traditionellen Datenbanken **moderne Technologien** in Ihrem Auswahlprozess um die bestmögliche Lösung für Ihr Einsatzszenario zu finden.
- Gerade im Datenbankumfeld mit **Vorsicht** zu genießen sind **Marketingaussagen** der Softwarehersteller sowie die bekannten Benchmarks. Achten Sie darauf, dass die untersuchten Benchmark-Szenarien auf Ihr Unternehmen zutreffen.



Die Studie wurde völlig unabhängig vom neutralen Marktanalysten BARC erstellt. Dank eines Sponsorings durch Netezza kann die Studie kostenfrei zur Verfügung gestellt werden.

Studiensponsor:



Mit der Entwicklung der Data Warehouse Appliance hat Netezza den Markt revolutioniert.

Netezza kombiniert gängige Komponenten wie Intel Prozessoren, IBM-Blades, Linux und quell-offene Datenbank und erhöht die Analyse-Performance signifikant.

Netezza liefert das 10-100fache an Leistung, bei der Hälfte der Kosten und minimiert damit die laufenden Betriebskosten erheblich. Skaliert von unter 1 Terabyte bis zu Petabyte linear.

Neben höchster Performance schätzen unsere Kunden besonders die Einfachheit und Eleganz im täglichen Einsatz, sowohl bei der Entwicklung als auch im laufenden Betrieb des Data Warehouses.

Sprechen Sie mit uns! Wir lösen Ihr BI Performance-Problem!

Susanne Winter / Netezza

+49 221 99 202 653

swinter@netezza.com

www.netezza.com

Business Application Research Center

BARC steht für Neutralität, Kompetenz und Qualität

Neutralität

Das Business Application Research Center (BARC) ist aus den komparativen Produktanalysen am Lehrstuhl für BWL und Wirtschaftsinformatik der Universität Würzburg, Prof. Dr. R. Thome hervorgegangen. In seiner Stellung als unabhängiges Institut ist BARC strikt neutral gegenüber Software-Anbietern. Dies bedeutet, dass keine Gebühren für die Aufnahme in Software-Evaluationen oder Provisionen bei der Empfehlung von Software erhoben werden. BARC bietet auch keine Implementierung von Software an, um keine internen Interessen zu erzeugen.

Kompetenz

BARC-Mitarbeiter sind seit 1994 in der Evaluation von Produkten für Business-Intelligence und Enterprise Content Management sowie der Beratung von Unternehmen tätig. Dabei vereinen die BARC-Analysten Markt-, Produkt- und Einführungswissen. Know-how-Basis sind die seit Jahren ständig durchgeführten Marktanalysen und Produktvergleichsstudien, die ein umfassendes Detailwissen über den Leistungsumfang aller marktrelevanten Software-Anbieter und neueste Entwicklungen im Markt sicherstellen.

Qualität

BARC-Beratungsprojekte sind hoch effizient und gewährleisten ein Höchstmaß an Auswahlssicherheit. BARC-Studien bieten einen qualifizierten Marktüberblick über alle Anwendungsfelder des Informationsmanagements. BARC-Tagungen und Seminare geben einen konzentrierten Eindruck aller relevanten Anbieter im Business-Intelligence- und Dokumentenmanagement-Markt. BARC-Guides enthalten umfassende Marktübersichten und Beiträge zu BARCs Themenkomplexen. Über 90 Prozent der DAX100-Unternehmen nutzen in vielfältiger Weise das Know-how der BARC-Analysten.





Copyright © BARC GmbH 2010. Alle Rechte vorbehalten.

Business Application Research Center - BARC GmbH
Steinbachtal 2b
97082 Würzburg
+49 (0)931 880651-0
www.barc.de

Quelle Titelbild: www.istockphoto.com

