White paper *Business*

Exasol

Die schnellste analytische Datenbank der Welt

Inhalt

Die schnellste analytische Datenbank der Welt

01

Einleitung 3

Merkmale 4

02

Was genau ist Exasol? 6

Core Architektur 8

Weitere Merkmale 9

03

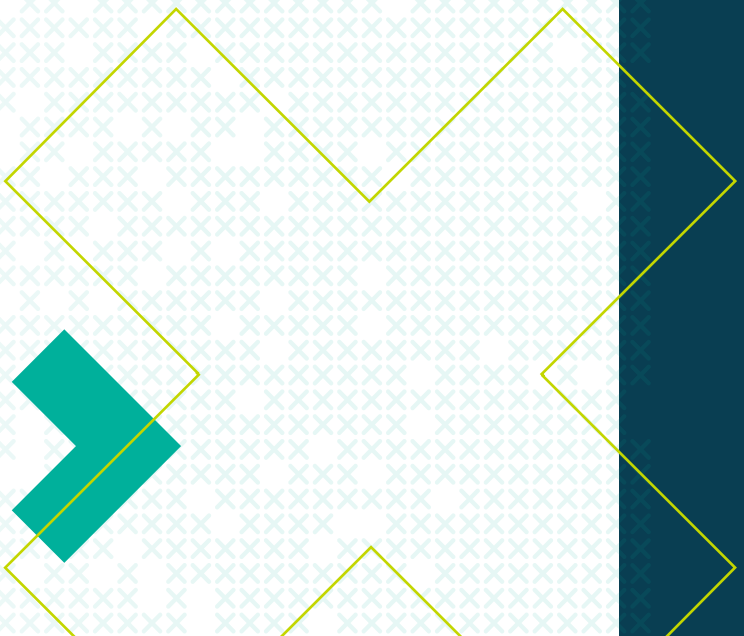
Nutzen 11

04

Einsatznutzen 13

05

Fazit 15



01

Einleitung

Exasol wurde 2000 in Nürnberg gegründet und unterhält Niederlassungen in den USA, Frankreich und Großbritannien. Am Hauptsitz Nürnberg stehen Aaron Auld (CEO), Mathias Golombek (CTO) und Michael Konrad (CFO/COO) sowie das Team für Qualität, Leidenschaft und partnerschaftliche Kundenbeziehungen.

Aushängeschild des Unternehmens ist das Kern-Produkt Exasol, eine hochleistungsfähige, In-Memory MPP-Datenbank, die speziell zu Analysezwecken entwickelt wurde. Bereits 2011 hat die Datenbank von Exasol im TPC-H Benchmark einen Rekord aufgestellt. 2014 konnte dieses Ergebnis in puncto Leistungsfähigkeit und Preis-Leistungs-Verhältnis noch weiter verbessert werden.
Exasol ist somit der einzige In-Memory-Datenbank-Anbieter, der Datenmengen von 100 GB bis zu 100 TB verarbeiten kann.

Mit Unterstützung der Exasol-Standorte in Deutschland, Großbritannien, Frankreich, den USA sowie weltweiten Partnern optimieren über 300 Organisationen ihr operatives Geschäft und bieten dank Exasol einen exzellenten Kundenservice. Exasol ist branchenunabhängig, d.h. die Lösung ist in den unterschiedlichsten Märkten zu finden, u.a. in den Bereichen Digital Media, Einzelhandel, Telekommunikation, Finanzdienstleistungen, produzierendes Gewerbe und Forschung.
Die In-Memory-Datenbank von Exasol wurde von Grund auf neu entwickelt und läuft auf kostengünstigen x86 Prozessoren. Das tuning-freie System ist schnell einsetzbar und spart erhebliche Kosten im operativen Betrieb. Die Hochleistungsdatenbank bietet umfassende Lösungsszenarien für Business Intelligence, Customer Insight und Data Warehousing bis hin zu komplexen Predictive Analytics in Echtzeit.

Anwender profitieren vor allem bei komplexen und anspruchsvollen Analysen sowie bei großen Datenmengen (Big Data) von den Vorzügen von Exasol.

01

Exasol ist ein spaltenorientiertes, massiv paralleles auf In-Memory-Technologie basierendes Datenbankmanagementsystem für Data Warehousing und Analyseanwendungen und zeichnet sich durch folgende Merkmale aus:

In-Memory-Technologie

Innovative In-Memory-Algorithmen ermöglichen eine wesentlich schnellere Verarbeitung großer Datenmengen im Hauptspeicher.

Spaltenorientierte Speicherung und Kompression

Durch spaltenorientierte Speicherung und Kompression wird die Anzahl der I/O Operationen und die zur Verarbeitung im Hauptspeicher benötigte Datenmenge verringert und die Leistung erhöht.

Massiv-Parallele Datenverarbeitung

Exasol wurde als Parallelsystem entwickelt, das auf einer Shared-Nothing-Architektur basiert. Abfragen werden auf alle Knoten eines Clusters verteilt. Dies geschieht unter Verwendung eines optimierten parallelen Algorithmus, der die Daten lokal im Hauptspeicher eines jeden Knotens verarbeitet.

Hohe Anzahl gleichzeitiger User

Tausende Benutzer können gleichzeitig auf eine große Datenmenge zugreifen und diese analysieren ohne dass die Leistungsfähigkeit der Abfrage beeinträchtigt wird.

Skalierbarkeit

Durch die lineare Skalierbarkeit kann das System jederzeit erweitert werden und die Leistungsfähigkeit kann durch das Hinzufügen zusätzlicher Knoten erhöht werden.

Tuningfreie Datenbank

Intelligente Algorithmen überwachen kontinuierlich die Nutzung und passen sich eigenständig an, wodurch die Leistungsfähigkeit des Systems optimiert und der administrative Aufwand minimiert werden.

01

Schneller Zugriff auf sämtliche Datenquellen

Mit dem Datenvirtualisierungs-Framework „Virtual Schemas“ sowie dem hochperformanten Datenimport-Framework können neue Datenquellen leichter angebunden und noch schneller analysiert werden.

Umfassende Hadoop-Integration

Daten, die in beliebigen von HCatalog unterstützten nativen Datenformaten vorliegen, können direkt aus HDFS geladen werden, sodass Hochgeschwindigkeitsanalysen von strukturieren und unstrukturierten Daten einfach und reibungslos durchgeführt werden können. Der Datentransfer erfolgt dabei hochperformant und parallelisiert.

Advanced In-Database Analytics

User Defined Functions (UDF) ermöglichen es, dass Advanced Analytics innerhalb der Datenbank durch den Einsatz von R, Python, Lua und Java problemlos läuft.

Erweiterte Konnektivität

Zusätzlich zu den bisherigen JDBC, ODBC und .NET- Schnittstellen wird darüber hinaus eine Web-Socket-basierte SQL-Schnittstelle unterstützt. Durch diese neue Schnittstelle kann fast jede Plattform problemlos auf Exasol zugreifen, auch wenn kein dedizierter Treiber verfügbar ist. Exasol enthält bereits einen auf diesem API basierenden Python-Adapter.

Beliebige Einsatzmöglichkeiten

Als sehr flexible Lösung kann Exasol in unterschiedliche Geschäftsmodelle eingebunden werden. Folgende Möglichkeiten stehen zur Auswahl: als reine Software-Lösung, als Appliance oder in der Cloud (EXACloud, Microsoft Azure oder Amazon Web Services).

02

Was genau ist Exasol?

Core Architektur

Die Architektur von Exasol ist in Abb. 1 dargestellt. Exasol verfügt über eine benutzerfreundliche, webbasierte grafische Benutzeroberfläche EXAOperation, ein eigenes Cluster-Management-System EXAClusterOS sowie über eine eigene Storage-Mangement-Komponente EXAStorage. Exasol unterstützt den ANSI Standard SQL 2008 (einschließlich aller analytischen Funktionen) sowie einen großen Teil des gebräuchlichen Oracle SQLDialekts. Die Unterstützung von Teilen des Oracle-SQL-Sprachumfangs ist insbesonders bei der Migration von Oracle-angebundenen Applikationen von Vorteil. Ein Code-Refactoring wird dadurch deutlich vereinfacht oder sogar unnötig.

Ausfallsicherheit und Redundanz sind durch zusätzliche „Hot Standby" Server im Cluster gewährleistet. Sollte ein Server ausfallen, übernimmt automatisch einer der „Hot Standby" Server dessen Rolle und das Cluster arbeitet weiter. Der defekte Server kann ohne die Außerbetriebnahme von Exasol entfernt und ersetzt werden (er wird dann zum neuen „Hot Standby" Server).

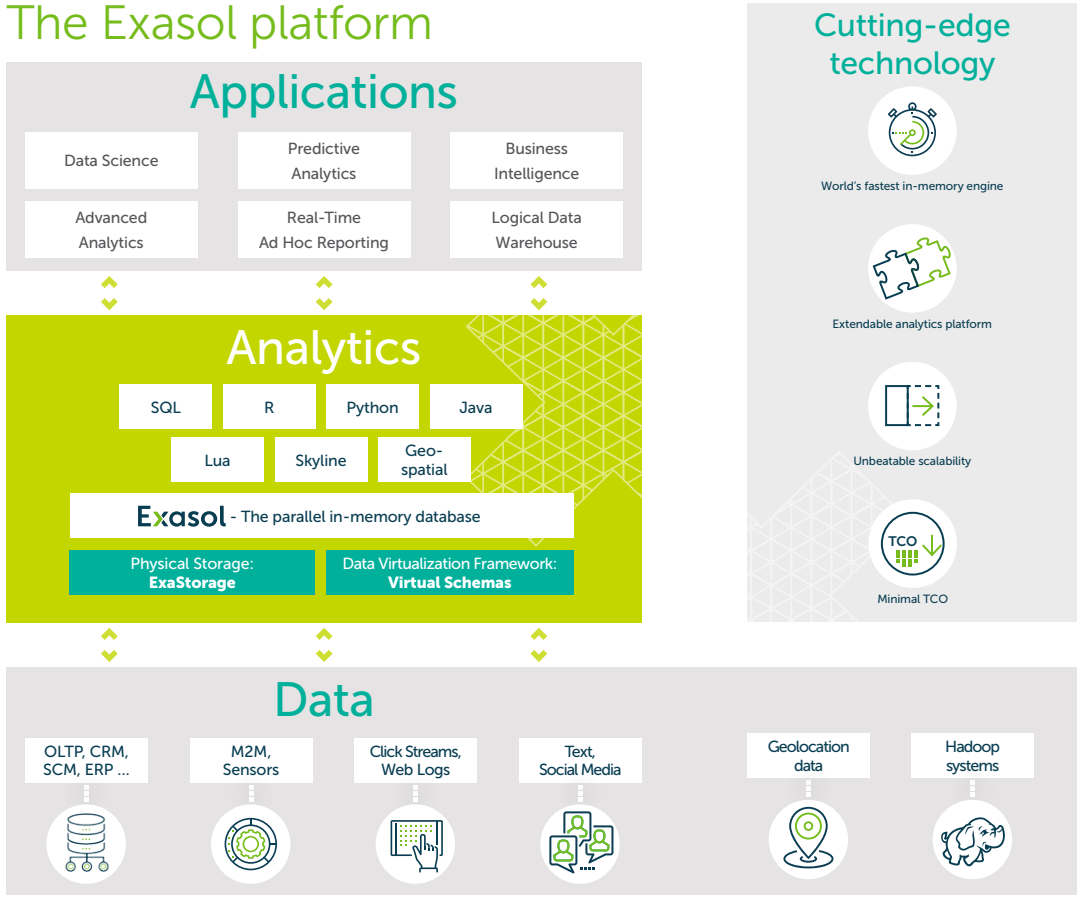


Abb. 1: Logische Architektur von Exasol

02

Was genau ist Exasol?

Core Architektur

Exasol verfügt über einen hohen Automatisierungsgrad und sichert damit einen niedrigen Administrationsaufwand. Einige Automatisierungsschwerpunkte sind z.B.:

Gleichmäßige automatische Verteilung der Daten

auf alle Server im Cluster.

Automatische Daten-duplizierung,

um im Falle eines Serverausfalls die Datenintegrität sicherzustellen.

Automatische Anwendung innovativer Kompressions-algorithmen,

die datentypspezifisch und für die In-Memory-Verarbeitung optimiert sind. Diese Algorithmen arbeiten auch unabhängig voneinander an jedem Knoten, um eine optimale Performance zu garantieren.

Automatische Datenkompression auf Spaltenebene

mit identischen Speicherbildern sowohl im Hauptspeicher als auch auf persistenten Speicher zur Optimierung der Performance.

Automatische Überwachung und Aufzeichnung der Systemressourcen

(RAM, Harddisk, CPU) zur Unterstützung der Kapazitätsplanung.

02

Was genau ist Exasol?

Core Architektur

Die fallenden RAM-Preise haben viele Lieferanten angeregt, In-Memory-Optionen für ihre bestehenden Datenbankprodukte zu entwickeln. Exasol hingegen wurde von Anfang an als In-Memory-Datenbank geplant und entwickelt. Dies ist kein „add-on" Feature und im Gegensatz zu einer Reihe von Wettbewerbsprodukten braucht man nicht die gesamte Datenbank In-Memory zu speichern. Die Erfahrung zeigt, dass selbst schon bei einer Hauptspeichergröße in der Größenordnung von lediglich 10 % des Rohdatenvolumens optimale Performance erzielt wird. Wie auch bei der persistenten Speicherung auf Disks ist Kompression auch hier hilfreich. So ergibt sich eine hinreichende Flexibilität um Kosten (Speicher, Server, etc.) sowie die benötigte Performance gezielt auf die eigenen Bedürfnisse anzupassen.

Exasol enthält einen intelligenten, kostenorientierten Query Optimizer, der vollautomatisch Optimierungen, wie etwa die Erstellung von Indizes oder die Replikation kleiner Tabellen vornimmt, wodurch in der Praxis die aufwendige manuelle Optimierung des Systems entfällt.

Außerdem überwacht Exasol als Teil des Workload Managements die Ressourcenauslastung und protokolliert sie. Wenn sich also das Arbeitsaufkommen erhöht (z.B. mehr Daten, mehr Nutzer, immer komplexere Abfragen) und die Leistungsfähigkeit der Datenbank zu sinken droht, hilft die Information des System-Monitorings bei der Festlegung wie viel mehr an Speicher pro Servers benötigt wird (Upscaling) oder falls nötig, um wie viele neue Knoten (Server) das Cluster ergänzt werden muss (Outscaling), um das Leistungsniveau zu halten.

Das Upscaling und die Versorgung der Datenbank eines jeden Servers mit mehr Speicherkapazität werden durch die EXAOperation GUI gesteuert. Das Hinzufügen neuer Server ist ebenfalls unkompliziert. Ist die neue Hardware erst einmal in das Cluster eingebunden, werden die Daten automatisch über die bestehenden und die neuen Knoten im Hintergrund umverteilt und der User kann während dieses Prozesses weiter Abfragen vornehmen.

Für spezielle Fälle, wenn die Auslastung sehr unterschiedlich ist (etwa weil sie zu verschiedenen Tageszeiten stark variiert), kann mit Hilfe der Priorisierung von Benutzern und Rollen die Ressourcenverteilung sowie das Scheduling den Bedürfnissen angepasst werden.

02

Was genau ist Exasol?

Weitere Merkmale

Exasol unterstützt Standardschnittstellen zur Integration von Upstream- (Datenintegration) und Downstream- (BI)-Tools. Die von der Datenbank genutzten Standardschnittstellen sind u.a. ODBC, JDBC, .NET sowie eine Web-Socket-basierte SQL-Schnittstelle (JSON). Die Datenbank unterstützt alle gängigen am Markt befindlichen BI- und Datenintegrations-Tools wie z.B. Informatica, Talend, Pentaho, Tableau, Alteryx, Business Objects, Cognos, Birst und MicroStrategy.

Des Weiteren ermöglicht die Unterstützung der nativen Oracle Schnittstelle OCI extrem schnellen und parallelisierten Datenaustausch mit Oracle-Datenbanksystemen.

Ein weiteres Unterscheidungsmerkmal ist, dass ein SQL-Präprozessor die Transformation bestehender proprietärer Queries in ANSI-standardkonformes SQL erlaubt, ohne die ursprünglichen Queries anpassen zu müssen. Durch diese Einbindung der bestehenden SQL-Anwendungen wird Zeit gespart und das Risiko beträchtlich minimiert.

Exasol besitzt weiterhin einen Bulk-Loader, der sich einfach mit Datenintegrationswerkzeugen verwenden lässt. Ein leistungsfähiges Merkmal ist u.a. die Fähigkeit, komprimierte Daten zu verarbeiten, z.B. im zip-Format, die noch schnelleren Datentransfer ermöglichen. Da Unternehmen Analysen immer mehr zur Unterstützung ihres operativen Geschäftes verwenden, müssen Daten im Data Warehouse regelmäßig angepasst und aktualisiert werden. (Batchload) Anwendungsfällen zu verarbeiten.

Die Datenbank ermöglicht Daten in Echtzeit aufzubereiten und einzelne Row Inserts parallel zu traditionelleren

Die Advanced Version von Exasol stellt zusätzlich ein leistungsfähiges Analytics-Framework zur Verfügung. Nutzer können Code, der in den Programmiersprachen R, Python, Lua oder Java geschrieben ist, parallelisiert mit extrem hoher In-Memory-Performance in der Datenbank als User-Defined-Functions (UDFs) ausführen. Zusätzlich können beliebige weitere Programmiersprachen integriert werden, so dass Anwender nicht mehr auf den bereits sehr umfangreichen Standard-Sprachumfang beschränkt sind.

02

Weiterhin können Berechnungen analog dem Map-Reduce-Prinzip von Hadoop-Systemen direkt in der SQL Engine durchgeführt und innerhalb eines SQL Statements mit Standard SQL kombiniert werden (beispielsweise über JOINs). Darüber hinaus unterstützt Exasol zahlreiche räumliche Funktionen zur Verarbeitung von Geo-Daten.

Die Advanced Edition von Exasol beinhaltet ein leistungsfähiges Datenvirtualisierungsframework (Virtual Schemas) und ein erweiterbares hochflexibles Integrationsframework (ETL-UDFs). Dies ermöglicht eine sehr flexible und performante Integration von Hadoop-basierten Daten in Exasol.

Ein weiteres Alleinstellungsmerkmal ist die analytische Skyline-Funktion für Preference Analytics. Skyline stellt quasi eine völlig neue Art von Analytics dar. Preference Analytics greift die grundlegenden Probleme der traditionellen Herangehensweise des Data Minings auf. Denn die ständig wachsende Datenflut und die Vielfalt der Variablen bedeuten, dass traditionelle Ansätze wie Scoring, Sortieren und Filtern zu suboptimalen Analysen führen.
Ein Beispiel aus der Praxis ist die Auswahl der besten Investmentfonds. Hierbei ist eine kontinuierliche objektive Analyse unter Berücksichtigung des Risikos, der Rendite und zahlreicher weiterer Kennzahlen mit herkömmlichen Ansätzen alles andere als einfach. Skyline berechnet dabei die in der Regel recht kleine Teilmenge derjenigen Fonds, die tatsächlich auf Basis der definierten Kriterien in die engere Wahl kommen können.

03 Nutzen

Exasol liefert High Performance Analysen auf einem hoch skalierbaren, leicht zu bedienenden System, das schnell einsatzbereit und extrem kostengünstig ist.

Für den Einsatz von Exasol müssen Sie Ihr bestehendes System nicht ersetzen. Es lässt sich als Ergänzung in die bestehende Infrastruktur implementieren und liefert hoch-leistungsfähige Analysen, die mit Ihrem vorhandenen System nicht möglich wären. Durch diese Ergänzung können Sie Ihre bestehende Infrastruktur weiterhin vollumfänglich nutzen und beweisen, dass Ihr Unternehmen hochleistungsfähige Analysen ohne den Eingriff in bestehende Prozesse durchführen kann. Falls gewünscht, können Sie dann im Laufe der Zeit eine schrittweise Umstellung der vorhandenen Analyseanwendungen auf die In-Memory Datenbank planen.

Der Vorteil, den Exasol in Bezug auf das Preis-/Leistungsverhältnis bietet, wurde durch den TPC-H-Benchmark-Test der unabhängigen Organisation Transaction Processing Performance Council deutlich. Exasol nimmt mit großem Abstand eine Spitzenposition gegenüber anderen Lösungen sowohl bei den zu verarbeitenden Daten als auch beim Preis für Datenvolumen von 300 GB bis 100 TB ein.
Das Diagramm in Abb. 2 verdeutlicht den Leistungsvorteil von Exasol bei allen Faktoren. Die ausführlichen Ergebnisse des TPC-H-Tests finden Sie auf der Website des Transaction Processing Performance Council unter: www.tpc.org

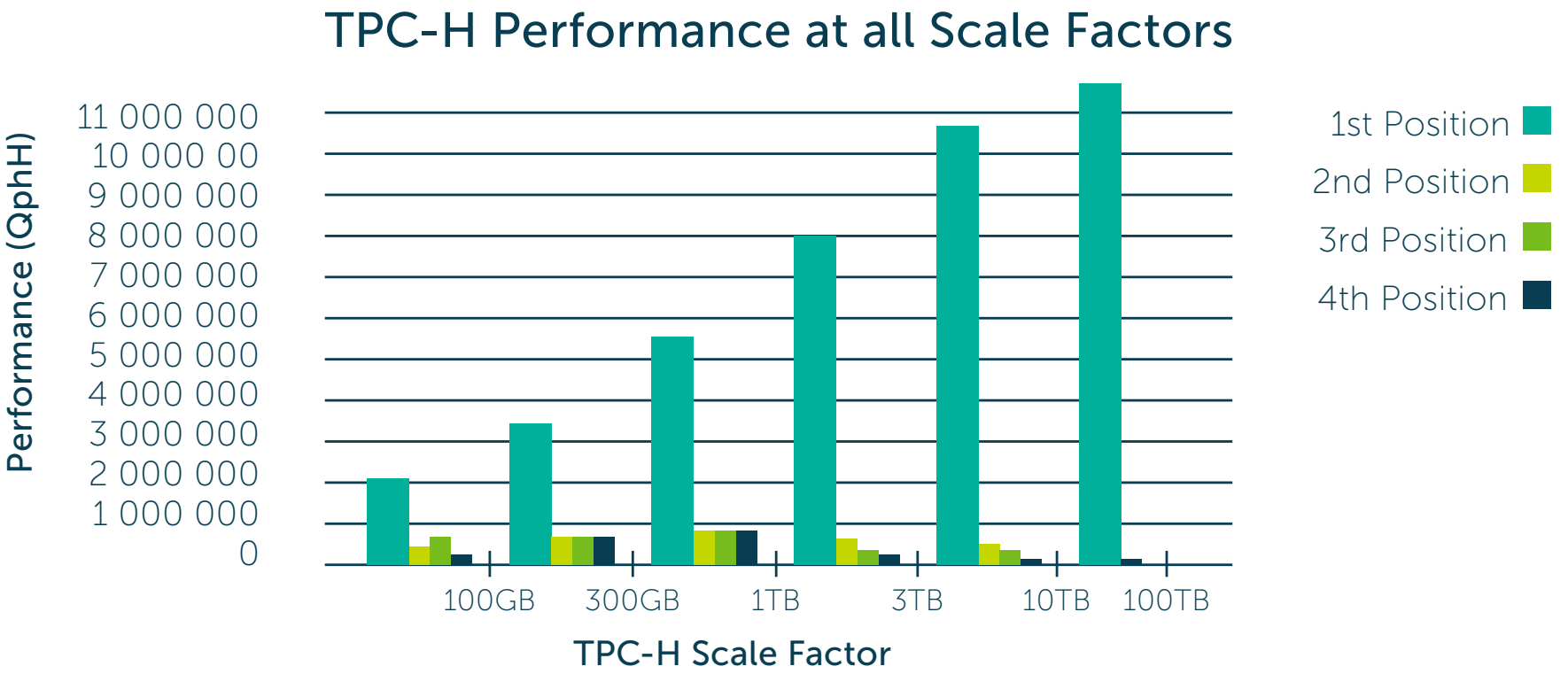


Abb. 2: Leistungsvorteile von Exasol

03 Nutzen

Die erweiterten Analytics-Fähigkeiten der Advanced Edition eröffnen vielfältige leistungsfähige Möglichkeiten für neuartige Analysen Ihrer Business-Daten. Diese lassen sich grob in zwei Kategorien einordnen:

Erweiterung, Integration oder gar Ersatz von traditionellen plattform-spezifischen Analyse-Anwendungen,

bei denen die maximal verarbeitbare Datenmenge durch die Plattformarchitektur stark begrenzt ist. Dies ist beispielsweise oft bei Systemen wie MATLAB oder SAS der Fall.

Erstellung neuer High Performance Computing (HPC) Anwendungen,

die nur mit MPP-Technology wie Exasol samt seiner skalierbaren Datenanalysefähigkeiten realisierbar sind.

Diese Anwendungen sind charakterisiert durch extrem große Datenmengen, die in kurzer Zeit mit komplexen Algorithmen analysiert werden müssen. Die erweiterten leistungsfähigen Fähigkeiten der Advanced Edition basieren auf offenen Frameworks für Datenintegration und analytischer Anwendungsentwicklung. Die Integration dieser so geschaffenen Funktionalitäten kann einfach über die Standard-SQL-Schnittstelle durch UDFs (User Defined Functions) erfolgen. Diese offene Herangehensweise ermöglicht es Unternehmen, Lösungen zu planen und zu gestalten, die bisherige technologische Investitionen optimal nutzen. Dies unterscheidet Exasol von vielen Wettbewerbern, die häufig geschlossene proprietäre Ansätze verfolgen, die zu einem klassischen Vendor-Lock In führen.

04 Einsatznutzen

Exasol ist sehr flexibel einsetzbar und hocheffizient:

Exasol läuft auf Standardhardware

Neue und teure High-End Server sind nicht erforderlich.

Die Exasol-Lizensierung basiert wahlweise auf dem der Anwendung zugeordneten Arbeitsspeicher oder der maximal verarbeitbaren Rohdatenmenge.

Dies ermöglicht eine flexible Lizensierung, die auf den Einsatzzweck zugeschnitten ist.

Die Leistung von Exasol ist abhängig vom Verhältnis zwischen Datenbankgröße, Systemauslastung sowie dem verfügbaren Arbeitsspeicher.

(Hinweis: Es besteht keine Notwendigkeit, ausreichend Speicher für die gesamte Datenmenge vorzuhalten.) Wenn alle zur Verarbeitung der Anfragen notwendigen Daten stets im Hauptspeicher liegen, ist die Performance optimal. Die Leistung mit weniger RAM (und geringeren Lizenzkosten) erfüllt jedoch üblicherweise auch die SLA-Anforderungen.

04 Einsatznutzen

Höhere Workload-Anforderungen können flexibel

durch Aufstockung des Hauptspeichers der existierenden Knoten (Scale-up) und/oder durch die Hinzunahme neuer Datenbankknoten (Scale-Out) erfüllt werden.

Unterschiedliche Datenbanken können auf einem einzigen Hardware-Cluster laufen

– beispielsweise ideal zur Unterstützung mandantenfähiger Anwendungen.

Exasol unterstützt einen Großteil des Oracle-Sprachdialekts,

wodurch Oracle-Umgebungen schnell, effizient und ohne große Aufwände migriert werden können.

Exasol enthält einen SQL Präprozessor,

der eine schnelle Migration von Applikationen ermöglicht, ohne die Applikationen selbst anpassen zu müssen.

Das leistungsfähige Analytics-Frameworks von Exasol

erlaubt eine nahtlose Integration von und mit anderen Technologien.

Datenvisualisierungs-Framework

Mit dem Datenvisualisierungs-Framework „Virtual Schemas" sowie einem hochperformanten Datenimport-Framework können neue Datenquellen leichter angebunden und noch schneller analysiert werden.

05

Fazit

Sie möchten Ihre bestehende BI-Plattform oder Ihre Business-Analysen beschleunigen? Sie wollen über Predictive Analytics komplexe Zusammenhänge vorhersagen? Oder müssen Sie Business Intelligence-Anwendungen und Analytics einer großen Anzahl an Nutzern zur Verfügung stellen? Wenn Sie Ihrem Unternehmen wirklich nutzbare „Big Data"-Lösungen anbieten möchten, dann ist Exasol die Plattform, die es Ihnen erlaubt, dies schnell, einfach und kostengünstig zu realisieren.

Überzeugen Sie sich selbst von den Vorteilen von Exasol, indem Sie unsere Lösung kostenfrei testen:

www.exasol.com/de/download/

Wenn Sie Ihre Anforderungen mit einem unserer Experten besprechen möchten, schreiben Sie einfach eine E-Mail an

info@exasol.com

Exasol AG
Neumeyerstr. 22 – 26
90411 Nuremberg
Germany
www.exasol.com

Tel: +49 911 23991-0
Email: info@exasol.com

Follow us for the latest content:

About this Whitepaper: Information listed here may change after the data sheet has been printed (June 2018). Exasol is a registered trademark. All trademarks named are protected and the property of their respective owner. © 2018, Exasol AG | All rights reserve