

Hochverfügbarkeit im Data Warehouse

27.03.2017, 18:57 | IT, New Media & Software

Pressemitteilung von: *mip GmbH*

Presseagentur: *dieleutefürkommunikation AG*

Big Data, Small Data, Sensor- oder Log-Daten, strukturiert oder unstrukturiert – deutsche Unternehmen nutzen endlich ihre Informationen: Bei vier Fünfteln basieren Unternehmensentscheidungen auf Datenanalysen, so eine Studie (<https://www.bitkom.org/Presse/Anhaenge-an-PIs/2016/Juni/Bitkom-Research-KPMG-Mit-Daten-Werte-schaffen-10-06-2016-final.pdf>). In modernen Data-Warehouse-Architekturen sollten heute Informationen zentral, sicher, indexiert und bereits vorsortiert für weiterführende Analysen vorliegen. Doch was ist, wenn beispielsweise ein Server ausfällt oder ein Netzteil in Rauch aufgeht? Themen wie Hochverfügbarkeit und fehlertolerante IT-Architekturen werden dabei auch im Zusammenhang mit dem Data Warehouse (DWH) immer wichtiger und beschäftigen Big-Data- und DWH-Experten wie Markus Ruf von der mip GmbH.

„Die frühere Ansicht, dass Daten nicht unbedingt hochverfügbar sein müssen, da ja Unternehmen ohne sie bzw. ohne Auswertungen und Analysen weiter funktionieren können, zählt heute nur noch eingeschränkt“, erklärt Ruf. Zwar haben immer noch die tatsächlich wertschöpfenden Prozesse Priorität. Jedoch sind in der Zwischenzeit Informationen sowie Datenanalysen teilweise selbst zu wertschöpfenden Prozessen in Unternehmen geworden. Vortages- oder Tagesauswertungen gehören zu wichtigen Entscheidungshilfen für Vorstände, Management und Unternehmensführung, genauso wie für Abteilungen. „Daten etwa zum Kundenverhalten sind heute wesentlich für aktuelle sowie künftige Entwicklungen und Trends – Stichwort: Predictive Analytics“, so Ruf.

Hochverfügbares Data Warehouse: Welche Variante macht Sinn?

Auch wenn bei einem Ausfall die Firma nicht gleich zusammenbricht, ist ein hochverfügbares DWH sinnvoll. Welche Methode bzw. Hochverfügbarkeitsstrategie sich hierfür am besten eignet, muss von Fall zu Fall am besten zusammen mit DWH-Experten geprüft werden: Wie lange darf mein DWH im Jahr höchstens ausfallen? Müssen die Daten 24 Stunden am Tag, 7 Tage die Woche vorliegen? Damit hängt meist auch die Frage zusammen, wie viel das Thema Hochverfügbarkeit überhaupt kosten darf.

Sehr häufig wird ein Data Warehouse nur tagesaktuell genutzt. Aber selbst wenn bei einem Ausfall die Berechnungen und Auswertungen über Nacht nicht stattfinden können, weil kein paralleler bzw. gespiegelter DWH-Aufbau vorliegt, lassen sich die Informationen aus den operationalen Daten wiedergewinnen. „Eine Herausforderung ist jedoch, dass einige Firmen ihre Daten direkt im Data Warehouse archivieren und so nicht mehr alles aus dem operativen System rekonstruieren können“, erklärt Ruf. „Hier ist es deshalb entscheidend, dass die Informationen zum Beispiel auf einem parallel betriebenen DWH oder in der Cloud gesichert sind.“

Verschiedene Wege führen zu mehr Hochverfügbarkeit

Um Daten hochverfügbar zu machen, existieren mehrere Methoden, die jeweils von der im Vorfeld ermittelten höchstmöglichen Ausfallzeit und dem eigenen Geldbeutel abhängen. Dazu gehören:

1. eine vollständige (und bestenfalls automatisierte) Rekonstruktion der Daten aus den operativen Vorgängen (über Nacht) – allerdings mit einem Tag Verzögerung
2. die Virtualisierung des Data Warehouse über eine virtuelle Maschine (VM): Wenn eine DWH-Hardware-Komponente ausfällt, wird die VM einfach auf einen anderen physikalischen Server oder ins Storage Area Network (SAN) verschoben. So lässt sich das DWH wieder starten. Diese Variante setzt voraus, dass der Storage- bzw. VM-Server nicht ausgefallen ist. Erst eine Spiegelung auf der physikalischen Ebene macht das System hochverfügbar.
3. ein HADR-Cluster (High-Availability Data Replication) in der Datenbank selbst: Während der Datenaufbereitung

wird eine Kopie des gesamten DWH angelegt. Somit lassen sich die Informationen bereits dort auf einer zweiten physikalischen Ebene spiegeln. Das hat aber auch Performance-Einbußen zur Folge, weil die Daten jedes Mal doppelt geschrieben werden müssen.

Synchron oder asynchron – das ist hier die Frage

Beim Thema Hochverfügbarkeit im Data Warehouse muss zudem im Vorfeld entschieden werden, ob dieser Bereich synchron oder asynchron aufgebaut werden soll. „Bei angestrebter Synchronität muss man stets darauf warten, bis alle Informationen geschrieben worden sind, was ein DWH langsam macht“, so Ruf. „Beim asynchronen Schreiben der Daten können diese allerdings verloren gehen, während sie ins DWH übernommen werden.“ Die wenigen so eingebüßten Datensätze lassen sich jedoch wieder per Datenrekonstruktion vervollständigen. Hier müssen sich Unternehmen entscheiden, ob sie nicht zugunsten einer höheren Performance eventuelle Datenverluste verschmerzen können.

Cloud und Hochverfügbarkeit

Geht es darum, Daten im Notfall dynamisch verschieben zu können, kommt beim Thema Hochverfügbarkeit die Cloud ins Spiel. Hierbei sind aber wieder zwei Seiten zu betrachten. Ruf: „Zum einen liegen die bekannten Cloud-Vorteile vor, wie etwa mehr Flexibilität durch eine bedarfsgerechte Speicherplatznutzung, ein höherer Sicherheitsstandard, den sich Firmen aus dem Mittelstand so gar nicht leisten können, und die wegfallenden bzw. geringeren Wartungs- sowie Administrationsaufwände.“ Wenn das komplette Data Warehouse in der Cloud aufbereitet wird, etwa als Database-as-a-Service (DBaaS), werden bestenfalls nur die Deltas aus den operativen Prozessen in die Cloud übertragen, wodurch keine Performance-Probleme auftreten. Allerdings könnten im schlechtesten Fall beide DWHs asynchron auseinanderlaufen.

„Arbeitet die Firma jedoch mit einem lokal aufgebauten Data Warehouse, welches zunächst alle Daten verarbeitet, um dann das gebündelte Gesamtvolumen zur Absicherung in die Cloud zu übertragen, kann sich dies auf die Performance auswirken“, beschreibt Ruf. Übertragungsgeschwindigkeiten von einem Gigabyte pro Sekunde reichen bei einem DWH mit Massendaten-Transfer bis zu mehreren Terabyte pro Nacht nicht mehr aus. Außerhalb der von Providern gut abgedeckten Gebiete sind Gigabyte-Bandbreiten allerdings kaum bis gar nicht verfügbar.

Hochverfügbare DWH-Strategien verlangen genaue Planung

Bereits vor dem Aufbau eines Data Warehouse und einer entsprechenden IT-Infrastruktur sollte deshalb darauf geachtet werden, welche DWH-Strategie das Unternehmen verfolgen möchte. Die grundlegende Frage lautet dabei, wie sich das DWH-Architektur-Konzept so aufsetzen lässt, dass das lokale DWH und die DBaaS-Lösung synchron laufen, ohne durch einen Infrastrukturtteil wie eine langsame Cloud-Anbindung ausgebremst zu werden.

Eine hybride Cloud ist beispielsweise dann sinnvoll, wenn ein asynchrones Modell verfolgt wird, bei dem sich mit einem Versatz von ein oder zwei Stunden arbeiten lässt. Damit lassen sich immer so viele Daten verschieben, wie die jeweilige Bandbreite erlaubt, während die User im On-Premise-DWH auf die aktuellen Daten zugreifen können. Wenn nur in der Cloud gearbeitet wird, besteht immer ein Zeitversatz-Problem. Teams, die an verschiedenen Standorten am gleichen Projekt arbeiten, müssen so mit unterschiedlich aktuellen Informationen arbeiten, was zu Fehlern führen kann.

Übergangsphase von On-Premise in die Cloud

Ruf: „Wir bei mip gehen davon aus, dass es künftig für Kunden einfacher wird, mit Echtzeitdaten aus ihren DWHs zu arbeiten. Das Hauptproblem bezüglich Datentransfers ist ja, dass die meisten Unternehmen ihre großen ERP-Systeme heute noch lokal installiert haben und auf dieser Ebene die operativen Daten erzeugen.“ Deshalb müssen sich Firmen beim Thema Data Warehouse überlegen, ob sie die Daten aus der ERP-Lösung in zwei physikalische, parallel laufende und damit hochverfügbare DWHs verschieben oder alternativ VMs bzw. eine zeitversetzt arbeitende hybride Cloud-Architektur nutzen – mitsamt Performance-Einbußen oder asynchronen Datensätzen.

„In Zukunft werden aber verstärkt auch die großen IT-Infrastrukturen über die Cloud betrieben“, führt Ruf weiter aus. „Infrastructure-as-a-Service oder Cloud-Systeme werden gerade von den ERP-Anbietern vorangetrieben.“ Dabei ‚zwingen‘ SAP Hana Cloud Platform, Microsoft Dynamics Nav etc. Unternehmen dazu, mit ihrer gesamten IT in die

Cloud umzuziehen, da sie sonst keine Software-Unterstützung mehr erhalten. Dann werden sowohl operative Daten als auch DWHs in der Cloud liegen, also in großen Rechenzentren. Durch die kurzen Wege sowie den schnell erweiterbaren Speicherplatz verschwinden Performance-Probleme. Die jeweiligen finanziellen Möglichkeiten bestimmen dabei die Übertragungsgeschwindigkeit sowie die Ausfallsicherheit.

Diese Entwicklung wird von den Anbindungsmöglichkeiten der Unternehmen an die Cloud bzw. der Netzwerk-Topologie der Provider abhängen. An die Frage nach einem uneingeschränkten Datenzugriff schließt sich auch die Diskussion um den Ausbau des Glasfasernetzes in Deutschland an – andere Länder sind schon viel weiter. „Hier muss sich noch einiges tun, damit auch die Unternehmen aus Randgebieten besser angebunden werden“, so Ruf. „Denn eine Hochverfügbarkeit ist erst dann gegeben, wenn ich über zwei getrennte Internet-Leitungen von zwei unterschiedlichen Providern verfüge. Der berühmt-berüchtigte Bagger, die häufigste Ursache für einen Ausfall, lauert überall!“

Eine Infografik zum Thema unter: <https://mip.de/hochverfuegbarkeit-im-data-warehouse/>

Portrait

Über mip

Die mip Management Informationspartner GmbH agiert seit 1988 als innovatives Dienstleistungsunternehmen mit Erfahrung in den Bereichen Data Warehouse, Business Intelligence sowie Business- und Predictive Analytics. Dabei begleitet das Beratungshaus seine Kunden als Ideengeber von der Konzeption, über die Umsetzung, bis hin zur Nachbetreuung auf Lösungs- und Infrastrukturebene. Strategische Partnerschaften mit großen Herstellern und spezialisierten kleineren Häusern stellen sicher, dass Projekte den entsprechenden technologischen Unterbau bekommen. Die langjährige SAP-Erfahrung der mip reicht vom klassischen ERP bis hin zu SAP HANA und Cloud Lösungen. Dazu gehören die Beratung in den Themen Einführung, Migration, Konsolidierung, Release-Wechsel, Prozess- bzw. Modulberatung und Transformation.

News-ID: 944821 • Views: 543 (Stand: 17.04.2026)

Link zur Pressemitteilung:

<https://www.openpr.de/news/944821/Hochverfuegbarkeit-im-Data-Warehouse.html>