

High Performance Computing und Big Data

Nach Einschätzung von IDC beläuft sich das Datenvolumen derzeit auf ca. 1,8 Millionen Zetabytes und wird sich alle zwei Jahre verdoppeln. Auf Unternehmen und Organisationen kommen also große Datenmengen zu, die es zu speichern, verwalten, analysieren und auswerten gilt, um geschäftliche Vorteile daraus zu generieren.

Das Management und die Analyse dieser rasch wachsenden Datenbestände zwingt somit viele IT-Verantwortliche, sich mit datenintensiven Computing-Ansätzen zu beschäftigen. In anderen Worten: „Big Data“ verlässt die Nische High Performance Computing (HPC) und trifft auf die unterschiedlichsten Anforderungen und Szenarien „normaler“ IT-Abteilungen. Daraus folgen vielfältige Implikationen: Klassische HPC-Ansätze werden nun von einer zunehmenden Zahl von Unternehmen und von Usergruppen (Business User) wahrgenommen. Damit einher geht aber auch zumeist die Erkenntnis, dass die im konventionellen HPC genutzten Verfahren kaum übertragbar auf andere Unternehmensbereiche sind.

HPC Business Case und Workloads sind häufig sehr spezifisch angelegt und auf eine konkrete Anforderung (z.B. Simulation, wissenschaftliche Analysen etc.) ausgerichtet. Die eingesetzte Technologie (Hardware, Software, Architektur) ist dann folgerichtig speziell für den jeweiligen Case „gebaut“ und somit relativ starr hinsichtlich Erweiterungen und Anpassungen. Dieser Spezialisierungsgrad wird sich in einigen Fällen noch verstärken.

Aus Sicht von IDC unterliegt HPC derzeit einer hohen Dynamik. Folgende Entwicklungen lassen sich aktuell nachzeichnen:

- Weiterentwicklungen und Neuerungen (z.B. Prozessorleistung, Datenmanagement) finden Eingang in die Systeme und Installationen
- Stark wachsende Datenmengen und die Vielfalt der Datenquellen erfordern neue Lösungsansätze
- Standardisierung von Systemen lässt HPC ein Stück weit Commodity werden (und senkt die Eintrittsschwelle für neue Anbieter)
- High End HPC ist ein Prestigeobjekt von Firmen, Institutionen und Staaten

Die Grenzen von High Performance Computing sind derzeit mehr oder minder im Fluss. HPC bewegt sich immer stärker in der Spanne zwischen Aufwand und Ertrag (Petascale-Exascale). Veränderte Rahmenbedingungen und Datenszenarien erfordern (neue) adäquate Lösungsansätze. Hier kann HPC lediglich einen Teil abdecken bzw. der Einsatz von klassischem HPC wäre wirtschaftlich nicht sinnvoll, denn beachtliche Ressourcen für Projekt- und Lösungs-Design, Architektur gehen Hand in Hand mit hohen Anschaffungskosten. Unternehmen und Organisationen sind nach wie vor bereit, hohe Beträge einzusetzen. IDC schätzt, dass die globalen Ausgaben für HPC-Ökosysteme bis 2015 jährlich um mehr als 6 Prozent wachsen werden. Trotz des Ausgabenzuwachses werden Unternehmen Alternativen zur Lösung datenintensiver Business Cases prüfen. Zunehmend verbreiten sich Lösungsansätze, die auf Standardtechnologien (Hardware und Software) basieren und häufig von Standardsoftwareanbietern vorangetrieben werden. „Big Data“ ist somit nicht mehr eine HPC-Domäne. Daraus lassen sich zwei Tendenzen ableiten:

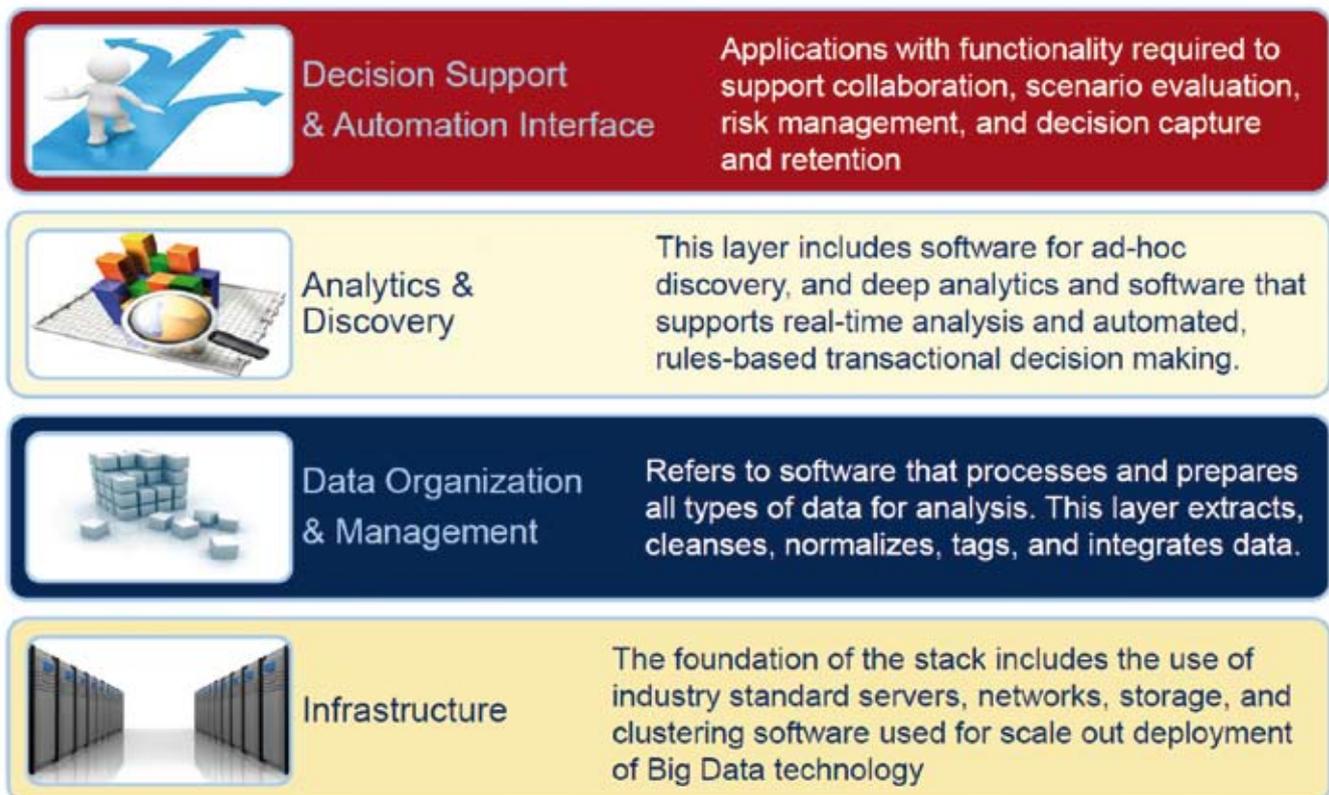


Abbildung 1: IDC Big Data Technology Stack

1. HPC im Rampenlicht:

Das ständige „mehr“ an Daten in unterschiedlichsten Formen, Formaten und aus einer Vielzahl von Quellen hat das allgemeine Verständnis für große Datenmengen geschärft. Das Volumenverständnis für Big Data schwankt dabei zumeist zwischen wenigen Terabyte bis hin zum einstelligen Petabyte-Bereich. Aber auch HPC-Sites mit einem Datenvolumen von 10 bis 25 Petabyte sind keine Ausnahme mehr. IDC erwartet hier alle zwei bis drei Jahre eine Verdoppelung des Datenvolumens. High Performance Computing erschließt sich neben den „klassischen“ Anwendungsfällen Klimaforschung, Pharma und Life-Sciences sowie Energiewirtschaft zunehmend weitere Themen unter zu Hilfenahme von semantischen Verfahren und Algorithmen zur Wissenserschließung. IDC erwartet folgende Entwicklungen in diesem Segment für 2012 und 2013:

- Verbindung etablierter (Modellierung und Simulationen) und neuerer (Analytics und Graph Analytics) Ansätze
- Annäherung von HPC und Standard-Business-Intelligence-Ansätzen. Solche Anbieter adressieren Lösungen für High Performance Analytics und haben entsprechende Units aufgebaut
- Verändertes Systemdesign ermöglicht unterschiedliche Herangehensweisen. Beschleunigung der Daten aufgrund höherer Bandbreite und geringerer Latenz bzw. Verringerung der Datenbewegung durch solche Ansätze wie In-Memory, Nearline Storage und massive Multithreading

Somit wird deutlich: HPC „öffnet“ sich ein Stück weit anderen technologischen Ansätzen, Zielsetzungen und Herangehensweisen.

2. Big Data für Alle

High Performance Computing und Big Data lassen sich per se nicht trennen. Die Handhabung großer Datenmengen außerhalb klassischer HPC-Szenarien wird aber für eine wachsende Zahl von Unternehmen unumgänglich. IDC geht davon aus, dass mit Big-Data-Technologien und Lösungen von Datenintegrations- und BI-Anbietern geeignete Lösungsansätze zur Verfügung stehen, um den Umgang mit Daten zu verbessern, zu erweitern und neu zu gestalten. Insbesondere Big Data Analytics kann einen Wertbeitrag zur Steigerung der Unternehmensperformance leisten, denn Daten werden immer stärker zur strategischen Ressource, sowohl für das aktuelle Business als auch für künftige Geschäftsszenarien. IDC empfiehlt Unternehmen, die sich mit Big Data auseinandersetzen (müssen), unabhängig vom im Detail gewählten Business-Szenario und der damit einhergehenden Lösung, einen Big Data Technology Stack für ihr Unternehmen zu entwickeln. Solch ein Stack unterstützt IT-Organisationen in der Ausgestaltung der erforderlichen technologischen Ebenen, unabhängig davon, ob die Lösungen On Premise implementiert werden oder ob externe Ressourcen (z.B. Dienstleister oder Cloud Services) hinzu gezogen werden.

Die Ansätze aktueller Big-Data-Technologie sind sehr vielfältig und betrachten die Thematik zum Teil aus konträren Positionen zwischen Appliances oder In-Memory auf der einen Seite und verteilten Systemen und parallelem Processing auf der anderen Seite.

Fazit

Von HPC und Big Data gehen derzeit Impulse für einen sich wandelnden Umgang in der Speicherung und Analyse von Daten aus. Beide Ansätze greifen zudem interessante und vielfältige technologische Entwicklungen der letzten Jahre auf, die Unternehmen in die Lage versetzen, ihre Daten besser als bisher als Produktivfaktor zu nutzen. Dieser an sich banalen Einschätzung wohnt aber umfassendes Potenzial inne, deren sich Business und IT bewusst werden sollten.



Matthias Zacher
Senior Consultant
IDC in Frankfurt