

Hartmut Wedekind

Rezension zum Buch:

Hasso Plattner: Lehrbuch In Memory Data Management - Grundlagen der In-Memory-Technologie, Springer 2013

Eins ist nur schwer vorstellbar: Der VW-Aufsichtsratsvorsitzende Ferdinand Piech entwirft und baut mit seiner Truppe einen Boliden (gr. βολιζ = Geschoss) für die Formel 1- Rennen. Was im Automobilbau unmöglich erscheint, in der Informationstechnik geht das. Hasso Plattner, Aufsichtsratsvorsitzender von SAP, entwirft und baut mit seinem HPI in Potsdam eine Hochleistungs-Datenbank, genannt HANA (**H**igh Performance **A**nalytic **A**pliance), mit der er ganz erstaunliche Performance-Werte gerade im Bereich Big Data erzielt. Man sagt „[Appliance](#)“, weil es sich um eine Kombination aus Hardware (Multicore-Architektur) und Software handelt. Ob das „dort geht es nicht“ und „hier geht das“ am Mann oder an der Materie liegt, soll nicht interessieren. Es gilt, ein Lehrbuch für Studenten des SAP-Aufsichtsratsvorsitzenden zu besprechen. Wir staunen, weil Aufsichtsratsvorsitzende im Allgemeinen keine Lehrbücher schreiben. Es handelt sich genauer gesagt um ein Lehrbegleitbuch, weil es zum einen als Lehrbegleittext zur Vorlesung des Autors auf der Online-Plattform „openHPI“ eingesetzt wird. Zum anderen dient es als Hinführung zur 2012 veröffentlichten Monographie „Plattner/Zeier, In-Memory Data Management, Technology and Application“, auch bei Springer erschienen.

Das Lehrbegleitbuch von Plattner kann nicht voraussetzungsfrei gelesen werden. Erforderlich sind schon Kenntnisse, wie sie z.B. im [Härder/Rahm](#), Datenbanksysteme, Konzepte und Techniken der Implementierung, Springer, 2. Aufl 2001, bereitgestellt werden. Hier findet der Leser z.B. Ausführungen über die Systempufferverwaltung klassischer, plattenbasierter Datenbanksysteme. „Idealerweise sollte dieser „unendlich“ groß sein“, sagt Härder, erreicht aber nur eine Größe von GBytes (2001!). Performancekurven in Abhängigkeit von der Puffergröße fallen exponentiell ab. Bei In-Memory-Datenbanken ist der Härdersche Idealzustand heute erreicht; es gibt den Systempuffer plattenbasierter Systeme nicht mehr. Die Hauptspeicherkapazitäten sind bei starkem Preisverfall so dramatisch gestiegen, dass das technologisch neue In-Memory-Computing heute ökonomisch betrieben werden kann.

Eins dürfte auch sicher sein: Das Fach „Datenbanksysteme“ wird sich aufspalten in platten-basiert (disk-based) und Hauptspeicher-basiert (in-memory) mit ihren spezifischen Anwendungsgebieten z.B. in der heute noch gängigen Prozessdatenverarbeitung und im hoch-performanten BigData-Bereich. Keiner behauptet, dass sie „alten“ Datenbanken untergehen werden. Interessant ist festzustellen, dass die „neuen“ Datenbanken schon ziemlich präzise 2004 von Jim Gray (1944-2007) vorausgesagt wurden, und zwar in seinem Aufsatz „The Next Database Revolution“ (SIGMOD 2004). Im Jahre 2006 begannen die Entwicklungen zu HANA. Herausgestellt wird von Gray u.a. die bekannte Spaltenorientierung. Dünn besetzte Tabellen, wie sie häufig vorliegen, lassen sich spaltenweise als dreistellige Relation (Key, Attribute, Value) hoch komprimiert abspeichern, meistens als Bitmap. Zentral für moderne In-Memory-Datenbanken ist aber die Verlegung der Geschäftslogik in die Datenbank. Eine alte Streitfrage der Datenbanken-Architekturen, ob die Geschäftslogik „inside -the-database“ oder „outside- the-database“ angesiedelt sein soll, ist somit aus Performancegründen für In-Memory- Datenbanken entschieden. Das alte Konzept der „stored procedures“ (gespeicherte Prozeduren), in einer „full fledged programming language“ (voll ausgebaute Programmiersprache) abgefasst, feiert fröhliche Urstände. Die Datenbanken erscheinen so nach außen als kompakte Systemteile, bereit, um auf massiv paralleler (massive parallel) Weise Anweisungen auszuführen.

Von Jim Gray stammt auch der Ausspruch: „The disks of today are the tapes of tomorrow“. Systemplatten aber verschwinden, weil sich alle Daten im Hauptspeicher befinden (In-Memory). Nur noch Nebenfunktionen wie „logging“ (Protokollieren), „recovery“ (Wiederherstellen) und „archiving“ (Archivieren) werden über Platten abgewickelt.

Die geplanten Anwendungen für HANA liegen natürlich nicht im BigData-Bereich der Astronomie wie bei Jim Gray, sondern, wie das zu SAP gehört, im Bereich der betriebswirtschaftlichen Standardanwendungen (ERP, Enterprise Resource Planning). Im Akronym HANA steht das erste A für Analytik. Wikipedia sagt uns zum Thema „Analytik“: „Eine systematische Untersuchung eines Sachverhaltes oder eines Gegenstandes hinsichtlich aller bestimmenden Faktoren oder Komponenten“. Ein alter Weisheitsspruch sagt uns auch: „Forschung ist nur dann etwas wert, wenn sich mit ihr eine Geschichte erzählen lässt.“ Das Herbeischaffen eines Sachverhaltes in der Analytik leistet im

Rechnungswesen (Accounting) in historischer Sicht nach den Worten des Betriebswirtes Eugen Schmalenbach (1873-1955) eine Grundrechnung, die dann in der Datenbanklehre später zu **Online Transaction Processing (OLTP)** präzisiert wurde. Die systematische Untersuchung eines Sachverhalts in der Analytik nannte Schmalenbach „Sonderrechnungen“, von denen in der Regel viele existieren. Aus „Sonderrechnungen“ wurde später in der Datenbanklehre **OLAP (Online Analytical Processing)**. Die rapide Entwicklung einer „Accounting Technology“, die als Teil einer allgemeineren „Process Technology“ aufzufassen ist, konnte Schmalenbach nicht erahnen. Wer konnte das schon? Wichtig ist aber für alle Fächer, den laufenden Fortschritt zu verarbeiten und aktiv zu fördern, weil sonst in den Fächern Verkümmern eintritt. Die Erstellung und Pflege einer Offenen-Posten-Debitoren Datei, z.B., ist eine Grundrechnung (OLTP), eine mögliche Auswertung nach Fälligkeit z.B. in einem Mahnlauf ist eine Sonderrechnung (OLAP). Das Mahnen (Dunning) wird vom Autor mehrfach als Beispiel hervorgehoben und auch im umfangreichen Glossar in wohlmeinender pädagogischer Absicht erklärt. Im Zentrum des Interesses eines SAP-Mannes stehen natürlich die Sonderrechnungen (OLAP). Denn die Grundrechnung bis zum Belegniveau hinunter muss inhaltlich der Kunde liefern. Plattner sagt es in seinem online-Kurs aus seiner Sicht treffend: Anwendungen, das sind Aggregationen (aggregates), Zusammenfassungen im Sinne einer Sonderrechnung. Und die soll ermöglicht werden nicht nur auf Jahresbasis, wie in der klassischen Papierwelt, sondern über die Monatsbasis hinaus auch auf der Tagesbasis. Das ist, stellt man sich Großunternehmen vor, eine gewaltige Herausforderung. Geschieht das Ganze auch noch abstrakt mit den dazu gehörigen Prozessen in den „Wolken“ (in the clouds), dann ist das Entwicklungsprogramm komplett.

Herauszuheben ist zum Ende dieser Rezension der Lehrbuchcharakter des Werkes. Am Ende eines jeden Abschnitts gibt es Selbsttest-Fragen, früher sagten wir Übungsaufgaben. Es freut die Studenten, dass man hinten auf Seite 237 im Buch nachschauen kann, wie die Lösung aussieht. Durch Vorführen von Lösungen lernt der Mensch ja auch. Vorbilder sind Lösungen und ermutigen den Menschen ungemein.

Eine Ironie der Geschichte: Zu meiner Zeit bei der IBM 1969 wurde das „Unbundling“, die preisliche Trennung von Hard- und Software, gerichtlich geboten. Heute im Zeitalter der aufkommenden „[Appliances](#)“ handelt es sich

um technologische Innovationen. „Die IT-Industrie kehrt zu ihren Ursprüngen zurück“, meint Wikipedia leicht ironisierend. Auch die Weisheit der Gerichte ist offenbar sehr beschränkt.