



INDUSTRY DEVELOPMENTS AND MODELS

High-Performance Data Analysis: When Big Data Needs More Firepower

Steve Conway

IDC OPINION

Big data arguably originated in the global high-performance computing (HPC) community in the 1950s for government applications such as cryptography, weather forecasting, and space exploration. High-performance data analysis (HPDA) (big data using HPC technology) moved into the private sector in the 1980s, especially for data-intensive modeling and simulation to develop physical products such as cars and airplanes. In the late 1980s, the financial services industry became the first commercial market to use HPC technology for advanced data analytics (as opposed to modeling and simulation). Investment banks began to use HPC systems for daunting analytics tasks such as optimizing portfolios of mortgage-backed securities, pricing exotic financial instruments, and managing firmwide risk. More recently, high-frequency trading joined the list of HPC-enabled FSI applications. The invention of the cluster by two NASA HPC experts in 1994 made HPC technology far more affordable and helped propel HPC market growth from about \$2 billion in 1990 to more than \$20 billion in 2013. More than 100,000 HPC systems are now sold each year at starting prices below \$50,000, and many of them head into the private sector. Leading commercial companies in a variety of market segments are turning to HPC-born parallel and distributed computing technologies – clusters, grids, and clouds – for challenging big data analytics workloads that enterprise IT technology alone cannot handle effectively. IDC estimates that the move to HPC has already saved PayPal more than \$700 million and is saving tens of millions of dollars per year for some others. In addition:

- Competitive forces will increasingly drive leading firms to follow the examples of PayPal and other commercial pioneers that are using HPC technology to move beyond today's static searches to exploit higher-value, dynamic pattern discovery.
- Knowledge-based businesses should explore HPC technology to see if it is a good fit for their evolving big data opportunities and challenges.

IN THIS STUDY

This study examines the growing adoption of HPC technologies by commercial firms to address mission-critical big data analytics problems. The study reviews the history of HPC use in the private sector, especially since NASA HPC experts conceived the cluster, and goes on to look at the HPC technologies in question and the factors driving businesses to adopt these technologies. Next in this document comes a series of real-world examples of commercial HPC use for big data analytics, along with economically important use cases that have emerged in this formative market. Finally, IDC offers essential guidance for firms facing big data analytics challenges that HPC technology may be uniquely able to tackle effectively.

SITUATION OVERVIEW

It's widely known that government and academic researchers have relied on high-performance computing for decades to solve scientific and engineering problems needing "big compute" and/or "big data" capabilities. It's no secret, either, that industrial firms of all sizes have adopted HPC to speed the development of products ranging from cars and planes to golf clubs and potato chips. But lately, something new is happening. Leading commercial companies in various sectors have begun turning to HPC technologies for mission-critical, big data analytics workloads that today's enterprise IT technology can't handle well enough alone.

This isn't totally surprising when you realize that some of the key technologies underpinning business analytics (BA) originated in the world of HPC. The evolution of these HPC-born technologies for business analytics has taken two major leaps and is in the midst of a third. The advances have followed this sequence:

- **Phase 1** was the advance from the mainframe mentality of running single applications on traditional SMP servers to modern clusters (i.e., systems that lash together homogeneous Linux or Windows blades to exploit the attractive economics of commodity hardware). The cluster was invented by two NASA HPC experts in 1994.
- **Phase 2** was the move to grids with the goal of supporting multiple applications across business units coherently. This enables enterprisewide management of the applications and workloads.
- **Phase 3** is the emerging move to cloud computing, which focuses on delivering generic computing resources to the applications and business units on an on-demand, pay-as-you-go basis. Clouds can be hosted within a company, by an external provider, or as a hybrid combination of both.

Why Businesses Turn to HPC for Advanced Data Analytics

High-performance data analysis is the term IDC coined to describe the formative market for big data workloads that exploit HPC resources. The HPDA market represents the convergence of long-standing, data-intensive modeling and simulation (M&S) methods in the HPC industry/application segments that IDC has tracked for more than 25 years and newer high-performance analytics methods

that are increasingly employed in these segments as well as by commercial organizations that are adopting HPC for the first time. HPDA may employ either long-standing numerical M&S methods, newer methods such as large-scale graph analytics, semantic technologies, and knowledge discovery algorithms, or some combination of long-standing and newer methods.

The factors driving businesses to adopt HPC for big data analytics (i.e., HPDA) fall into a few main categories:

- **High complexity.** HPC technology allows companies to aim more complex, intelligent questions at their data infrastructures. This ability can provide important advantages in today's increasingly competitive markets. HPC technology is especially useful when there is a need to go beyond query-driven searches in order to discover unknown patterns and relationships in data – such as for fraud detection, to reveal hidden commonalities within millions of archived medical records, or to track buying behaviors through wide networks of relatives and acquaintances. IDC believes that HPC technology will play a crucial role in the transition from today's static searches to the emerging era of higher-value, dynamic pattern discovery.
- **High time criticality.** Information that is not available quickly enough may be of little value. The weather report for tomorrow is useless if it's unavailable until the day after tomorrow. At PayPal, enterprise technology was unable to detect fraudulent transactions until after the charges had hit consumers' credit cards. The move to high-performance data analysis using HPC technology corrected this problem. For financial services companies engaged in high frequency trading, HPC technology enables proprietary algorithms to exploit market movements in minute fractions of a second, before the opportunities disappear.
- **High variability.** People generally assume that big data is "deep," meaning that it involves large amounts of data. They recognize less often that it may also be "wide," meaning that it can include many variables. Think of "deep" as corresponding to lots of spreadsheet rows and "wide" as referring to lots of columns (although a growing number of high-performance data analysis problems don't fit neatly into traditional row-and-column spreadsheets). A "deep" query might request a prioritized listing of last quarter's 500 top customers in Europe. A "wide" query might go on to analyze their buying preferences and behaviors in relation to dozens of criteria. An even "wider" analysis might employ graph analytics to identify any fraudulent behavior within the customer base.

Examples of HPDA in the Enterprise

The following examples illustrate the expanding range of HPC usage for advanced business analytics/business intelligence (BA/BI):

- The financial services industry was the first commercial market to adopt supercomputers for advanced data analytics. In the 1980s, large investment banks began hiring particle physicists from Los Alamos National Laboratory and the Santa Fe Institute to employ HPC systems for daunting analytics tasks such as optimizing portfolios of mortgage-backed securities, pricing exotic financial instruments, and managing firmwide, global risk. High-frequency trading is the newest major addition to the financial services industry's growing portfolio of HPC-driven applications.
- PayPal, a multibillion-dollar eBay company, adopted HPC hardware systems to perform sophisticated fraud detection on eBay and StubHub transactions in real time, before fraud hits credit cards. Enterprise technology had taken up to two weeks to identify fraud. IDC estimates

that using HPC to detect fraud has already saved PayPal more than \$700 million and has also enabled the company to perform predictive fraud analysis. Following this success, PayPal is extending HPC use to affinity marketing and management of the company's general IT infrastructure.

- Schrödinger, a global life sciences and materials science software company, with offices in Munich and Mannheim, Germany, hired Cycle Computing, a cloud computing services company, to test 21 million drug candidate molecules on the Amazon public cloud, using a new technical computing (HPC) algorithm Schrödinger developed. The successful run used 51,000 Amazon cores, took about four hours, and cost a little more than \$14,000. It would cost less than \$4,000 if run today. Schrödinger has completed even larger runs since this one, harnessing more than 150,000 Amazon cores.
- For cost and growth reasons, GEICO moved to automated insurance quotes on the phone. They needed to provide quotes instantaneously, in 100 milliseconds or less. They couldn't do these calculations nearly fast enough on the fly. GEICO's solution was to install an HPC system and every weekend run updated quotes for every adult and every household in the United States. That takes 60 wall-clock hours today. The phones tap into the stored quotes and return the correct one in under 100 milliseconds.
- A leader in the digital television services market, with 30 million U.S. and international subscribers, turned to HPC technology to help optimize revenue and customer satisfaction in the midst of escalating data volumes and business complexity. The company reports that the use of HPC-derived grid technology generates \$10 million in annual savings, along with higher customer satisfaction.
- One of the most respected firms in the global financial services industry updates detailed information daily on several million companies around the world. Clients use the firm's credit ratings and other company information in making lending decisions and for other planning, marketing, and business decision making. The firm uses statistical models to develop a company's scores and ratings. After moving to a multicluster HPC grid, the company can now maximize the use of its computing resources. The software automatically assigns jobs to server nodes with available capacity, instead of having users wait in queue for time on fully utilized nodes. The company executive estimates that it would have cost 30% more to purchase servers with enough capacity to handle these peak workloads on their own.
- One of the world's largest professionally managed vacation exchange and rental businesses uses a point system to describe trading power – that is, the value of each unit for exchange purposes. To drive this complex valuation model, the company has to produce a large number of high-volume supply-and-demand forecasts each day – even as the company's data volumes have been growing quickly. The company struggled to complete all the forecasts in one day with the system it was using until recently. An HPC grid allows the firm to generate all its forecasts at full resolution in eight to nine hours, about half the time these tasks were taking before. With this acceleration powering the exchange side of the business, the company is now free to spend more time on the rental side.

Economically Important HPDA Use Cases

It's easy enough to cite interesting examples, but much more important for the future of HPDA is determining which examples represent repetitive use cases that are evolving into pursuable market segments. After closely tracking the formation of the HPDA market for more than five years, especially actual sales of HPC compute and storage systems, IDC has added the following new commercial

HPDA applications to the established HPC segments we have reported on for more than 25 years. In 2014, for the first time, we have produced detailed five-year forecast for the new HPDA segments:

- **Fraud and anomaly detection.** This "horizontal" workload segment centers around identifying harmful or potentially harmful patterns and causes using graph analysis, semantic analysis, or other high-performance analytics techniques. The patterns may point to fraud, which is the deceptive exploitation or annotation of data for wrongful or illegal personal gain, or they may point to cybersecurity crime or insider threats, significant errors, or other anomalies that may deserve further investigation. HPC technology enables hidden fraud patterns to be detected in real time or near real time, something that standard enterprise IT technology typically cannot accomplish.
- **Marketing.** This segment covers the use of HPDA to promote products or services, typically using complex algorithms to discern potential customers' demographics, buying preferences, and habits.
- **Business intelligence.** This workload segment uses HPDA to identify opportunities to advance the market position and competitiveness of businesses by enabling businesses to better understand themselves, their competitors, and the evolving dynamics of the markets they participate in.
- **Commercial – other.** This catchall segment includes all commercial HPDA workloads other than the three just described in the previous bullets. IDC expects that over time, some of these workloads will become significant enough to split out of this "other" category and command their own segments. An example of such a high-potential workload is the use of HPDA to manage large IT infrastructures, ranging from on-premise datacenters to public clouds and Internet-of-Things (IoT) infrastructures.

FUTURE OUTLOOK

HPDA is a fast-growing, formative, worldwide market that is still heavily in motion. One thing is nearly certain; however, HPDA is evolving from static searches to an emerging era of higher-value, dynamic pattern discovery. The challenge presented by these problems is to discover hidden patterns and relationships – things you didn't know were there – and then to track patterns dynamically as they form and evolve or dissolve. Many of the commercial examples cited previously in this document – from fraud detection and BA/BI to marketing – already benefit from graph analytics and other pattern discovery methods.

Perhaps no field has stronger potential for benefiting from HPDA in general, and pattern discovery in particular, than bioscience. HPDA applications already in motion in this varied field range from advanced research – notably in genomics, proteomics, epidemiology, and systems biology – to commercial initiatives to develop new drugs and medical treatments, agricultural pesticides, and other bioproducts.

One of the world's most socially and economically important HPDA thrusts will almost surely be the multiyear transition from today's procedures-based medicine to personalized, outcomes-based healthcare. Identifying highly effective treatments in near real time (while the patient is still in the office) by comparing an individual's genetic makeup, health history, and symptomology against tens of millions of archived patient records poses enormous HPDA challenges that may take another decade

to master. The goal here is for the computer to process all this data and generate efficacy ratings for a range of treatment options. The options will eventually be highly personal: what constitutes a good outcome for a broken hand will vary, depending on whether the patient is an office worker or a concert violinist. When this capability matures, it will likely serve as a decision-support tool of unprecedented utility for the global healthcare community.

The HPDA vendor scene is becoming increasingly heterogeneous and vibrant. The analytics side of the formative HPDA market is where traditional HPC users and first-time commercial adopters are converging most rapidly. Established vendors that have served each of these customer groups are exploiting this convergence by following their buyers into the new HPDA analytics territory.

IDC forecasts that revenue for HPDA-focused servers will grow robustly (13.3% CAGR), increasing from \$743.8 million in 2012 to reach \$2.7 billion in 2018. HPDA storage revenue will approach \$1.6 billion in the latter year. The most serious technical challenge to liberating HPDA growth is data movement and management, although the HPDA market should be seen more fundamentally as a war among clever algorithms.

The growing market for high-performance data analysis – using HPC for big data challenges – is already enlarging HPC's long record of contributions to science, commerce, and society. HPDA promises to play a major role in helping commercial firms to address the major opportunities and challenges of the 21st century.

ESSENTIAL GUIDANCE

- **To stay competitive, knowledge-based businesses will need to out compute their rivals. HPC can help.** Competitiveness increasingly depends on a company's ability to exploit relevant data, and exploiting data increasingly depends on the ability to pose intelligent questions and obtain answers in near real time. As the examples in this study illustrate, HPC technology is achieving these goals for more and more businesses. Knowledge-based businesses should explore HPC technology to see if it is a good fit for them.
- **HPC technology is crucial for the transition from search to discovery.** The world of big data has already started the transition from today's static searches to the emerging era of higher-value, dynamic pattern discovery. In the new era, search will remain an important big data tool, but the most economically important business opportunities and challenges will increasingly rely on discovering hidden patterns in data and tracking these patterns as they evolve in real time. Fraud and anomaly detection is already becoming an important application for HPC-enabled pattern discovery. Innovative firms are also using HPC-powered pattern discovery for everything from affinity marketing to managing large on-premise and cloud-based IT infrastructures, including plans for Internet of Things.
- **HPC technology is more affordable and accessible than some people might think.** HPC clusters now start at under \$100,000 and rely heavily on standard technologies. Most vendors provide preassembled, pretested clusters and other HPC products, along with human experts. More than 100,000 HPC systems are sold around the world each year, and the vendor community has gained substantial experience helping even first-time users make the most of these products.

LEARN MORE

Related Research

Additional research from IDC in the high-performance data analysis program includes the following documents:

- *Worldwide High-Performance Data Analysis 2014-2018 Forecast* (IDC #248789, May 2014)
- *Perspectives on High-Performance Data Analysis: The Life Sciences* (IDC #248348, May 2014)
- *Perspectives on High-Performance Data Analysis: Government Applications* (IDC #248349, May 2014)
- *NVIDIA Unveils Technologies That Can Address HPDA Market Needs* (IDC #247780, April 2014)
- *IDC's Worldwide High-Performance Computing Predictions 2014* (IDC #WC20140211, February 2014)
- *Micron Demonstrates Technologies to Address Emerging Challenges in Big Data Applications* (IDC #244843, December 2013)
- *HPDA Pulse: 2013 Software and Consulting Market Analysis* (IDC #244513, November 2013)
- *HPDA Pulse Results: 2013 Hardware and Storage Market Analysis* (IDC #244493, November 2013)
- *Catalyst Supercomputer Heralds Shift to More Balanced Architectures* (IDC #lcUS24437513, November 2013)
- *China Eyes 10,000-Fold Data Reduction for Internet of Things* (IDC #lcUS24392513, October 2013)
- *High-Performance Data Analysis in the Life Sciences: HPC User Forum, September 2013, Boston, Massachusetts* (IDC #243774, October 2013)
- *High-Performance Data Analysis – PayPal Breaks New Ground: HPC User Forum, September 2013, Boston, Massachusetts* (IDC #243772, October 2013)
- *European Conference Underscores Movement of Enterprise Big Data Analytics to HPC Capabilities* (IDC #243618, October 2013)
- *HPDA: Importance of Hardware in Effective Big Data Solutions* (IDC #243495, September 2013)
- *eBay Deploys Innovative Green Power-Efficient Datacenter Technologies* (IDC #lcUS24350013, September 2013)
- *Worldwide High-Performance Data Analysis 2013-2017 Forecast* (IDC #241315, June 2013)
- *Changing Market Dynamics: HPC Meeting Big Data and IDC's Projected Evolution of the Market* (IDC #240365, March 2013)

- *High-Performance Data Analysis: HPC Meets Big Data* (IDC #DR2013_LSI1_SC_CD, March 2013)
- *High-Performance Data Analysis: The Visible Human Project* (IDC #238253, December 2012)
- *High-Performance Data Analysis at NASA JPL* (IDC #238254, December 2012)

Synopsis

This IDC study examines the growing adoption of HPC technologies by commercial firms to address mission-critical big data analytics problems. The HPC government market was arguably the birthplace of big data. HPC long ago moved into the private sector for applications based on data-intensive modeling and simulation, such as product development. In the 1980s, the financial services industry pioneered the commercial use of HPC for big data analytics.

"Today, more and more commercial firms are adopting HPC technology to tackle mission-critical advanced analytics problems that enterprise IT technology alone can't handle effectively. Some of these companies are already saving tens of millions of dollars per year," said Steve Conway, IDC research vice president for High Performance Computing/Data Analysis. "Competitive forces will drive more large companies, as well as SMEs, to exploit HPC technology as the market evolves from today's static searches to the emerging era of higher-value, dynamic pattern discovery."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2014 IDC. Reproduction is forbidden unless authorized. All rights reserved.

