



Flash

PayPal Says More Fortune 2000 Firms Could Benefit from HPC for Big Data Analytics

Steve Conway

IN THIS FLASH

This IDC Flash discusses PayPal's views on high-performance computing (HPC) technology for Fortune 2000 firms, presented at the SC14 annual supercomputing industry conference held November 16-21, 2014, in New Orleans. PayPal, an eBay company that IDC estimates has saved well over \$700 million by adopting high-performance computing technology, said at the SC14 supercomputing industry conference that other Fortune 2000 companies could benefit from adopting HPC. In the conference session, the PayPal speakers outlined a next-generation big data analytics problem they said will be crucially important for Fortune 2000 firms to solve. They called on the HPC community to collaborate with PayPal and other Fortune 2000 firms to attack this type of mission-critical big data problem. Audience members recommended a method for collaborating and many expressed interest in doing this.

SITUATION OVERVIEW

A few years ago, PayPal upgraded from enterprise IT technology to HPC clusters with much larger central memories and turbo-charged I/O. The move allowed PayPal to detect credit card fraud in real time and perform predictive analytics, versus PayPal's earlier need to wait up to two weeks for enterprise servers to identify fraud in the midst of myriad transactions (by then the fraud had hit consumers' credit cards). IDC estimates that in their first year on the job, the HPC clusters saved PayPal more than \$700 million.

In a session the author moderated at the SC14 annual supercomputing industry conference in New Orleans (November 16-21, 2014), PayPal presenters Arno Kolster and Ryan Quick said that the PayPal example is not an isolated one. They argued that many Fortune 2000 companies have mission-critical, next-generation big data problems that could benefit from HPC. IDC has closely tracked the marriage of big data and HPC for more than five years and calls this fast-growing convergence market *high-performance data analysis* (HPDA). During that period, we've seen a growing movement of daunting enterprise analytics workloads to HPC technology. IDC forecasts that by 2018, the HPDA market will reach \$2.6 billion in server revenue and another \$1.6 billion in external storage.

Most of the 90-minute SC14 conference session was a brisk back-and-forth discussion between the PayPal speakers and an audience that included some of the HPC community's best and brightest big data experts, drawn from government, academia, and vendors. Kolster and Quick outlined a complex HPDA problem they said Fortune 2000 companies like theirs will absolutely need to solve in the next

five years. They then described which parts of the daunting HPDA problem commercial firms have become better at tackling than the HPC community, which parts the HPC community knows how to address better, and which parts of the mission-critical, next-generation problem no one, to PayPal's knowledge, has figured out yet.

The PayPal speakers said they are already working with Oak Ridge National Laboratory to slice through some of these Gordian knots and invited others in the HPC community to collaborate with them and other Fortune 2000 firms to find solutions. When an HPC expert in the audience who works with both classified and non-classified HPDA users said he believes most of the future problem has been solved in the classified sector but solving the problem in its entirety looked impossible, the PayPal speakers said that they had solved other intractable-seeming HPDA problems and their Fortune 2000 customers would need this problem to be solved, impossible or not, in the next five years.

The Next-Generation Problem

Quick characterized the next-generation HPDA problem by likening it to a dynamically changing graph that needs to be updated in about 50ms. That's typically the time PayPal and other merchant firms get to respond when, for instance, a mall shopper swipes a credit card and waits for the transaction to be validated and processed. Fortune 2000 companies know how to do this part well today. The challenges arise from future competitive requirements. For purposes of illustration, the PayPal speakers provided a list of hypothetical but realistic requirements for a problem of this type:

- Expand the graph to encompass an entire shopping mall, with all its stores and other public areas treated as a fixed infrastructure.
- Account for all the shoppers in the mall, including their identities, their changing geographic locations (including mall entries and exits), their social interactions, their interests (topics), and their situations (demographics and in situ shopping behaviors). What pathways do people follow in the mall? Do people grab a cup of coffee before heading to a department store? Does the same person behave differently at different times of day or on holidays? To what extent can past or present behavior predict future behavior? How do relationships with friends, family, and others figure into all this?
- To capture this dynamic complexity, assume a group of three or more very large, nested graphs with an aggregate 1 billion nodes and 600 million edges. Each graph tracks a set of relationships as they dynamically change in time.
- The coherence of the graph structure must be maintained (100% for atomic reads) and updates to all graphs must be accomplished in less than 1 second.
- Availability must be at 99.999% for 99% of the connected components of the graph structure. (This means that no more than 1 out of every 100 nodes can be lost, but losses cannot affect the coherence of the graph structure.)
- Overall performance (response time) must be less than 50ms to traverse three levels of nested graphs.
- The graph structure has to be reasonably partitionable so that customers from throughout the world don't need to rely on slower access to clusters located in a single country.

Commercial and HPC Big Data Analytics Competencies

As a prelude to inviting more collaboration between Fortune 2000 firms and the HPC community to attack this next-generation problem, the PayPal speakers described what they thought each community is better at today.

Commercial Enterprise Competencies

Things the commercial world is better at:

- Making (and spending) money
- Repeatable, reusable scaling models (hardware, software, and interconnect)
- Delighting customers
- Availability and resiliency
- Efficiency, including power and space
- General-purpose distributed computing (e.g., provisioning, management, and cloud use)

HPC Community Competencies

Things the HPC community does better:

- Acquiring new problems – HPC users are willing to tackle almost any problem
- Adopting new technology
- Pushing the performance envelope (except power and space efficiency)
- Parallelism in all its varieties
- Collaboration – no other IT community collaborates as well

Kolster stressed that Fortune 2000 companies like his know how to manage big, distributed systems. To wit: "Give me 10,000 computers and I'm happy. One big computer scares the heck out of me."

The Call for Collaboration

Kolster and Quick said the HPC and Fortune 2000 communities should work together to solve the aspects of the next-generation problem that are intractable today, especially since the audience agreed that mastering these capabilities will be important for both communities.

Multiple audience members recommended that PayPal and other Fortune 2000 companies employ a proven method for collaborating with the HPC research community. In this method, the company would offer a modest reward, perhaps \$100,000-150,000, for the researcher or research team that provided the best solution for the problem by a fixed date of six to nine months in the future. Aside from the chance to win the monetary reward, this method would incent researchers by giving each of them an opportunity to write a paper and submit it for publication.

FUTURE OUTLOOK

Especially in the past five years, IDC has seen a growing movement of daunting enterprise analytics workloads to HPC technology. IDC forecasts that by 2018, the HPDA market for existing data-intensive HPC problems and the new enterprise analytics problems will reach \$2.6 billion in server revenue and another \$1.6 billion in external storage. The main factors driving businesses to adopt HPC for big data analytics (i.e., HPDA) fall into a few main categories of high complexity, high time criticality, and high variability (large numbers of variables).

PayPal's own business serves as a prime example of the benefits of HPC adoption. IDC believes that more and more Fortune 2000 companies will be driven to exploit HPC for their most daunting, mission-critical problems and the accompanying need to deal with higher complexity, time criticality, and variability. The next-generation problem outlined by PayPal in the SC14 conference session exemplifies this market trend. PayPal is already working with a leading HPC site, Oak Ridge National Laboratory, to tackle presently intractable big data analytics challenges.

IDC believes that a growing number of Fortune 2000 firms and HPC sites will begin to combine their respective competencies to attack next-generation analytics problems. These partnerships should benefit not only the collaborating parties but also the economies and societies in which they exist.

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.

