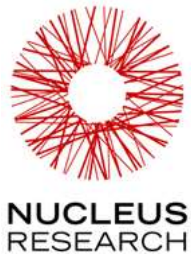


# THE PIVOTAL BIG DATA SUITE – REDUCING THE RISKS OF BIG DATA



## THE BOTTOM LINE

The explosion of big data and the rapid evolution of big data tools and technologies is challenging IT to meet the demands of business users with limited resources. As companies budget for big data initiatives, traditional data strategies such as overbuying to avoid business disruption or culling data that isn't perceived as useful are impractical and unrealistic. To leverage the returns from big data, companies need technologies that enable them to experiment while limiting their financial exposure. In looking at the Pivotal Big Data Suite, Nucleus found its integrated toolset, subscription model, core-based pricing, and unlimited Hadoop enabled organizations to tackle big data challenges while reducing risk.

IT departments are increasingly challenged to do more with fewer resources, and the challenge is even greater when it comes to big data. In the traditional data warehousing world, companies had to balance the cost of loading new volumes of data versus the value of being able to analyze it. Building a data warehouse meant organizing and structuring the data to store it based on the way it would be analyzed – and if business users later wanted to analyze it in a different way, they were out of luck (unless they had the budget to rebuild their data warehouse).

Exponential growth in data from the social Web, enterprise applications, and machine systems like network logs, along with the hype around big data, has put pressure on IT to tackle big data, and with good reason: Nucleus has found that enterprises are able to increase their ROI from analytics by an average of 241 percent with the addition of big data analysis (Nucleus Research *m17 - The stages of an analytic enterprise*, February 2012). However, that means storing all the data. Data that was once deleted because its storage was deemed too costly is now valuable, and IT is struggling to find ways to make these greater volumes of data accessible to users at the speeds they demand for analysis and decision making.

The traditional data warehouse doesn't work for big data, and cost is not the only reason. Unlike traditional data in warehouses, big data is:

- Unstructured. Big data sets can include all kinds of data, and it is structured at search and analysis, not when stored.
- Dynamic. Big data sets grow and change, and so do the analytical demands of business users. Additionally, as big data tools and technologies continue to evolve, IT has to know which model or tool is best for the job, or risk making a costly mistake, and an initial analytical approach may uncover the need for other tools or approaches.
- Multi-purpose. Unlike the data in a traditional data warehouse, different subsets of big data may be analyzed with different tools at different times depending on business demand.

These characteristics, as well as organizations' understanding of big data and business users' evolving demands for analysis are creating new challenges for data managers. They said:

- *"We're just starting to look at big data, and we don't know where it's going. The biggest challenge is identifying where to start."*
- *"We've had more requests coming in, but I can't just go out and buy a new tool every time I get a new request."*
- *"Our CFO continues to tell us we need to reduce the cost but deliver more computing power to the business."*

Even with relatively inexpensive Hadoop storage strategies, CIOs are challenged to meet the new business demands of big data. At the same time, as big data technologies and techniques continue to evolve, the timelines and budgets of traditional data warehouse projects are unrealistic and risky for big data efforts.

## PIVOTAL'S APPROACH

---

In looking at Pivotal's Big Data Suite, Nucleus found it enabled companies to improve their big data approaches and strategies while minimizing risk. The Pivotal Big Data Suite includes:

- The Pivotal Greenplum database, which uses a shared-nothing, massively parallel processing (MPP) database with flexible column and row oriented storage to support interactive analysis.
- Pivotal HD, which is Pivotal's commercial enterprise Hadoop distribution, to support real-time, interactive, and batch analysis using GemFire HD and HAWQ, Pivotal's SQL query engine.
- Pivotal Gemfire and GemFire XD, in-memory data grids for real-time analysis of massive quantities of data.

Beyond the technology itself, Pivotal's subscription pricing model is based on the number of core processors rather than the volume of data stored. This approach enables users to move data between the HDFS (Pivotal HD) and applications, and only pay for what they analyze. Additionally, Pivotal provides support for unlimited data storage.

**REDUCED TECHNOLOGY RISK**

Pivotal is providing a more flexible structure for organizations to consume the solution in a way that best meets their maturing analytic needs, whether that is using components such as GemFire or SQLFire standalone, or shifting data from Greenplum for more powerful analytics, then offloading to Pivotal HD for storage. If the business is looking for more SQL processing, the licensing can be adjusted to match these requirements, and an increase in the number of HAWQ licenses can be implemented. If there is an increasing need from the business to perform more real-time analysis the number of GemFire XD licenses can be increased, with complete flexibility and no need to renegotiate.

Organizations also have to deal with technology integration risks as data demands can result in infrastructure and platform changes. The Pivotal Big Data Suite's components are fully integrated, and users can change or add components within the suite as needed, accelerating time to value and reducing integration risk.

**REDUCED FINANCIAL RISK**

When organizations are faced with big data challenges, key considerations for resolution include storage costs, data volumes, and usage patterns. The old model of overbuying storage to avoid business disruption is no longer realistic. With Pivotal's new packaging and subscription-based licensing, the pricing is based on the value that customers gain from the software rather than the volume of data processed or stored.

This subscription model lets customers provide additional capabilities and functionality to their users without having to go to a number of vendors, or worry about integration challenges. This licensing structure reduces the risk of long term vendor lock-in and the challenges of traditional licensing models that could limit change, forcing organizations to maintain inflexible technology infrastructure that no longer meets their needs.

**UNLIMITED HDFS (PIVOTAL HD)**

It is difficult to predict how business needs and requirements will change as awareness and insight into the business value in the data changes. Data that was once deleted because of storage costs is now considered valuable, and organizations are struggling to find ways to make the greater volumes of data accessible to users at the speeds they require for decision making, in the format they require, and at a storage cost that is affordable. Nucleus has found that analytic deployments achieve returns of more than 1000 percent by tapping into big data sources that are large, contain a broad variety of data sets, are available on-demand, and change rapidly (Nucleus Research, *m20 – The big returns from big data*, March 2012).

To help organizations deal with the storage, volume, and cost concerns associated with big data, Pivotal has provided Pivotal HD as an unlimited and fully supported product, with no restrictions on the volume of data or the number of machines upon which a customer

may install it. The lack of restriction on data volumes also reduces the risk of experimenting with big data analysis, as companies incorporate new data sources such as social media, structured and unstructured data, streaming data, and machine data. By providing unlimited Pivotal HD, Pivotal is giving customers the ability to grow their data environment based on business needs, without having to worry about managing storage costs.

## CONCLUSION

---

Data volumes continue to grow and companies are struggling with ways to balance the costs of storing data, as well as meeting the analytic needs of their users. Overbuying storage is unrealistic, but so is discarding data that may later be meaningful. Pivotal's Big Data Suite enables companies to pay for analytics as they need it, and support and evolve their big data strategies over time as the market and their demands evolve, while reducing technology and financial risk.