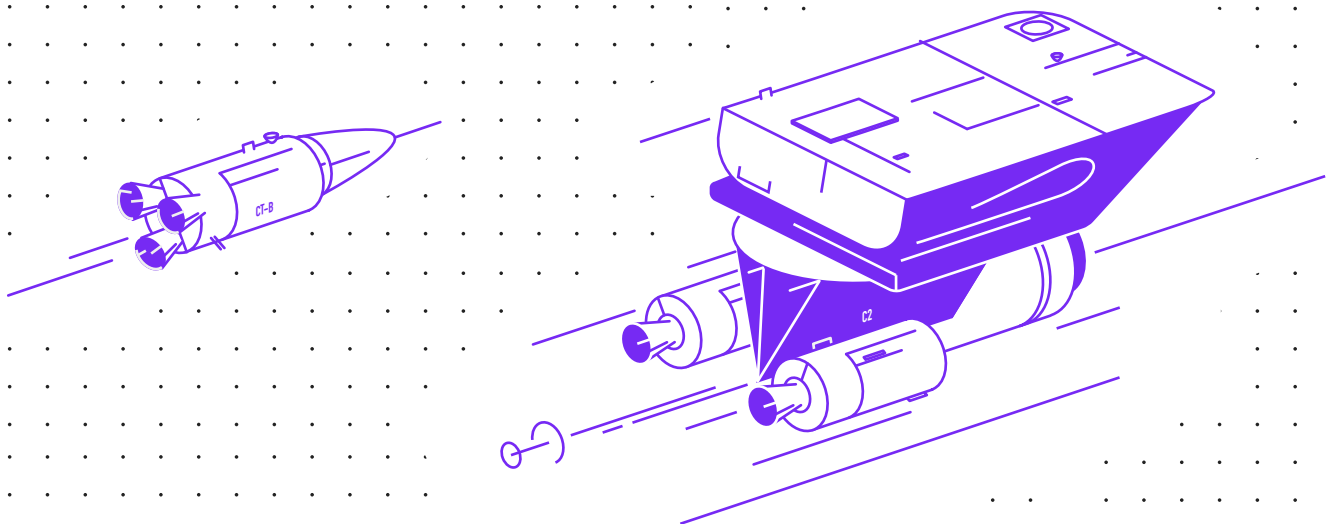


CLUSTERPOINT

WHY ARE ENTERPRISES BETTING BILLIONS ON DATABASE AS A SERVICE?



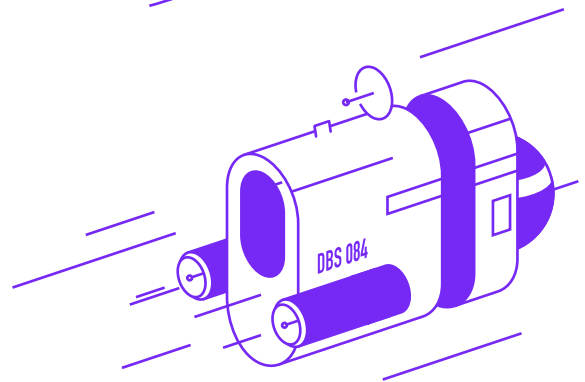
Girts Palejs, Clusterpoint

Copyrights © Clusterpoint Ltd.
info@clusterpoint.com
December 2015



FROM BIG MACHINES TO BIG DATA

Over the last few years, big data has become an increasingly important part of the enterprise landscape.



IN A 2015 SURVEY, GARTNER REPORTED THAT OVER 75% OF ENTERPRISES HAD EITHER INVESTED IN OR PLANNED TO INVEST IN BIG DATA PROJECTS OVER THE NEXT TWO YEARS.¹

The interest in big data is the result of two key trends:

1. The sheer volume of data being generated in the world today—IBM projects that data will grow 800% by 2017²
2. The desire of businesses to become more data driven across all of their operations.

Big data is generally defined as a data environment that meets the four criteria of volume, veracity, velocity, and variety. That is, a data environment that features a lot (volume) of different kinds of data including text and video (variety), where reliability (veracity) and speed of analysis (velocity) are paramount.

This environment is largely the result of recent developments in technology including "smart" mobile devices (e.g., smartphones, tablets), social media, and the rise of video-based communications, which collectively help to generate billions of gigabytes (or *exabytes*) of new data every day.

More than 50 years earlier, the computing industry set into motion a different kind of "big" data with the introduction of the world's first data storage devices: huge mainframe-sized appliances that held several million bytes (or *megabytes*) of data at the cost of about \$10,000 per megabyte (about \$100,000 per megabyte in today's currency).

By the 1980s, a one gigabyte (i.e., 1,000 megabytes) storage device had shrunk to the size of a small refrigerator and would cost you a mere \$40,000.³ The ubiquity of the personal computer (PC) accelerated the evolution of data technology, and by the 1990s most PCs featured their own built-in hard drive, which brought the cost per megabyte of storage below one dollar.

Fast-forward to today, and a multi-terabyte (i.e., 1,000 gigabytes) storage device can fit in your hand and costs you pennies per gigabyte. This shrinking cost of data storage is one of the key factors behind the growth of data, but it's not the only factor.

In the original definition of big data, one important V has been left out: Value. Enterprises today see tremendous value in analyzing and leveraging data to drive better decisions, improve processes, uncover new market opportunities, and anticipate the future.

THE DATABASE: DRAWING VALUE FROM DATA

In order to provide value to decision makers, data needs to be stored, organized, and analyzed. This is accomplished today through database management systems, which allow enterprises to efficiently store and organize data for subsequent analysis.

The first commercial database systems in the 1970s employed a hierarchical storage structure that allowed users to search and find data, but offered no true basis for comparative analysis. This changed in the 1980s with the introduction of relational database management systems (RDBMS) such as IBM DB2 and Oracle Database.

Relational databases store information in tables, which offers more flexibility in terms of data organization and analysis. Using Structured Query Language (SQL) commands, data analysts can compare and analyze the relationships between different data tables in a relational database. (The previously undiscovered correlation between diapers and beer in shopping carts is an early and famous example of this, which revealed that new fathers were likely to buy both items at the same time.)

In time, these analytic insights were expanded through the use of decision support, business intelligence and data mining technologies, some of which used highly sophisticated algorithms to discover hidden associations in enterprise resource planning (ERP) and customer relationship management (CRM) databases as well as, increasingly, purpose-built data repositories known as data warehouses and data marts.

Through this golden age of business intelligence, companies such as BusinessObjects, Cognos, SAS Institute, and Teradata eventually grew to become billion-dollar businesses.

Yet, for the most part, large-scale business intelligence was a luxury that only large enterprises could afford. It cost a significant amount of time and money to build and maintain a multi-terabyte data warehouse—and, in fact, it still does.

With recent advances in computing such as virtualization and Flash-based storage, however, small and medium enterprises are also able to store, access, and analyze large amounts of data.

It's all part of a new paradigm in information technology (IT)—dubbed the Third Platform by analysts IDC—characterized by the convergence of cloud technology, big data, social media, and mobility.

In the *Third Platform*, traditional client-server architectures give way to virtualized and distributed IT environments that leverage cloud-based computing, networking, storage, and applications to meet the increased demands for scalability, mobility, and simplicity.

DATA IN THE CLOUD

Cloud computing grew out of the need to scale IT systems efficiently to handle increasingly large and complex workloads. Instead of the traditional client-server architecture—in which IT resources such as networking, storage, and processing are dedicated to specific applications—cloud computing uses a "shared everything" approach that allows enterprises to allocate these resources where they're needed most.

The earliest commercial examples of large-scale, distributed computing environments were developed by companies like Amazon and Google, which in the 1990s began building massive server farms as part of their own internal IT systems.

It was Amazon that first realized the opportunity to monetize this investment by making their internal IT systems available to customers as a service. The early awareness of cloud's potential has paid off for Amazon, and today they're the leader in a growing cloud services market that includes Google, Microsoft, and many others.

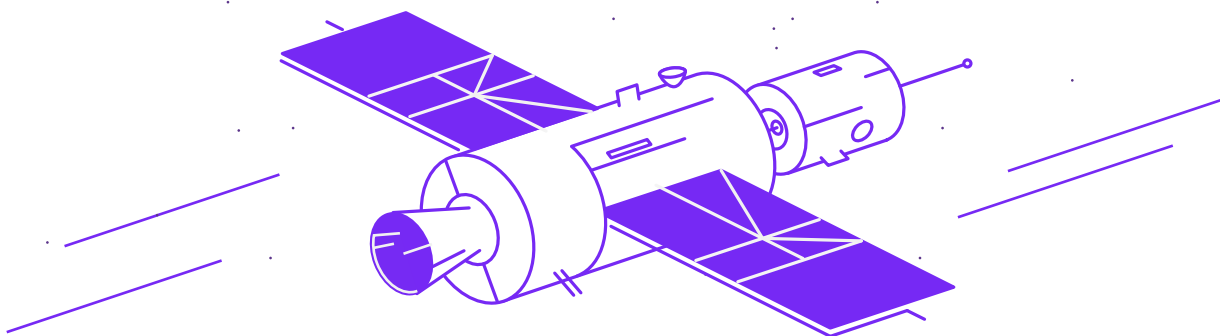
In the next five years, that market looks to get even more crowded as telecommunications companies and Internet service providers (ISPs) open their own massive networks to cloud services as well.

All of this has had a *Cloud Effect* on IT systems, beginning with the Software as a Service (SaaS) movement. In the SaaS model, for example, software is no longer loaded on a specific device; it is accessed over a network connection (typically, the Internet), which allows users to access the same application from multiple devices such as desktop computers, laptops, smartphones, and tablets.

The popular customer relationship management (CRM) application, Salesforce, is an early example of a successful SaaS model. In the future, many business software applications, including the popular Microsoft Office suite, are expected to be consumed primarily as a cloud-based service.

Storage as a Service (STaaS) is another popular cloud service, driven by the market need for consistent and mobile access to data from multiple devices. Particularly with the advent of smartphones, which feature limited storage capabilities, STaaS has played a significant role in helping enterprises think "outside the box" when it comes to business data.

With the success of SaaS and STaaS, it was only a matter of time before enterprises began moving their database management into the cloud as well.



DATABASE AS A SERVICE

There are currently several different ways that enterprises can leverage cloud technology for database management:

- 1 They can extend their internal database management system into the cloud for additional scalability/performance by leasing virtualized servers;
- 2 They can move their database into a hosted, managed cloud environment;
- 3 They can purchase Database as a Service (DBaaS).

Of these three options, DBaaS has the most long-term potential. In fact, the DBaaS market is expected to grow at an impressive compound annual growth rate (CAGR) of 67% over the next five years to become a \$14 billion market by 2019.⁴ The reason for the rise of DBaaS is several-fold.

First, it allows enterprises to consolidate and centralize their enterprise data, which is vastly more efficient than the siloed approach that characterizes most enterprise database environments.

Secondly, DBaaS enables enterprises to quickly spin off new databases as needed—in some cases, in minutes.

Thirdly, despite the feeling by some that internal equates with secure, DBaaS providers have made significant investments to meet industry compliance requirements (e.g., HIPAA, PCI DSS) as well as security policies (e.g., ISO 27001).

Lastly, DBaaS allows enterprises to easily bring together multiple data types (text, numeric, binary) in a single solution.

MOVING FROM RELATIONAL TO CONTEXTUAL

Before the Era of Big Data, most data was generated by internal business applications such as ERP and CRM systems. This data was relatively predictable in terms of format and could therefore be stored efficiently in a relational data model using a consistent set of pre-defined fields such as customer name, address, transaction type, date, etc.

While the RDBMS model works well for predictable and structured data like this, much of the data being generated today is semi-structured and characterized by volatility. This includes data generated from social media, text, videos, and machine-generated data, which can be highly dynamic and unique from source to source. Attempting to overlay a uniform structure on data like this results in added complexity, loss of context, and ultimately undermines efforts to analyze the data in meaningful ways.

To draw value from semi-structured data, the industry is looking beyond relational data to contextual data stored in NoSQL databases. Unlike a traditional RDBMS, NoSQL databases do not store data in tables but rather as a collection of objects or documents with nested relationships, depending on the type of NoSQL data model employed.

Currently, the most common type of NoSQL databases are key-value store databases—so named for the fact that a key is used to identify and locate specific data within the database—and [document-oriented databases](#).

Redis and DynamoDB are examples of a key-value store database, while MongoDB and Clusterpoint are examples of a document-oriented database.

The primary distinction between these two types of NoSQL databases is that a key-value model stores information as a generic "blob" of data while document-oriented databases store information in a language that the database understands, such as XML or JavaScript Object Notation (JSON).

The importance of this distinction comes into play when querying the data, as data analysts can query document-oriented data specifically rather than just universally.

	Relational Database (RDBMS)	NoSQL Database
Data Model	Tabular	Key Value or Document Oriented
Computational Model	Relational Algebra	Create, Read, Update, Delete (CRUD). Some vendors provide SQL view on data or computational capabilities.
Consistency	ACID-compliant transactions	Only a few vendors offer ACID-compliant, multi-document transactions.
Scaling	Vertical (i.e., more powerful HW)	Horizontal (i.e., more servers)

Table 1. RDBMS vs. NoSQL features

At the present time, NoSQL databases are more likely to supplement rather than replace an enterprise's RDBMS environment. According to a recent InformationWeek survey, approximately one in four enterprises (26%) have already deployed NoSQL technology in their business.⁵

Most industry experts believe that NoSQL will become increasingly important as the need for real-time analysis of semi-structured data increases; already, IBM and Oracle have built NoSQL interoperability into their RDBMS platforms in anticipation of this future.

WHY ENTERPRISES ARE SAYING “YES” TO NOSQL

Traditional RDBMS platforms are designed for static data—think accounting spreadsheets and reports with rows of numbers. They are not designed for less predictable and more complex data types such as geo-spatial information, large text files, videos, and so on.

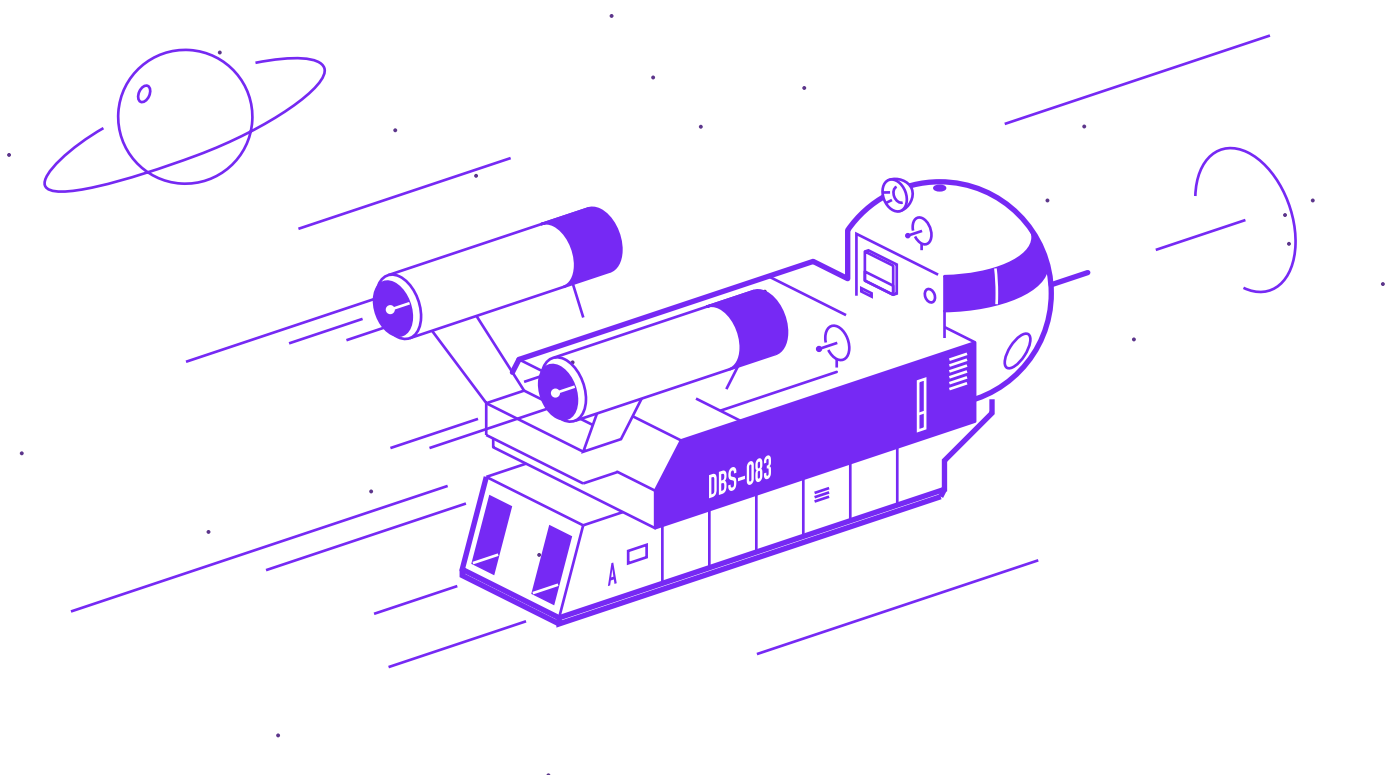
These data types often defy traditional data schemes and make it very difficult [for developers to create applications that effectively store and retrieve this kind of data in relational databases](#).

NoSQL databases, by contrast, are designed for the "anything goes" data demands of the future. Database administrators do not need to create a data schema in advance of loading data into a NoSQL database; they can simply load data and create context around it on the fly for later analysis. This gives enterprises a much broader degree of flexibility in both the kinds of data they collect and how they analyze it, particularly those enterprises that are combining data from multiple upstream sources for analysis.

Not surprisingly, enterprises that expect a wide degree of volatility in the data they collect (e.g., Web-centric businesses, enterprises exploring big data applications) have become the most passionate adopters of NoSQL technology.

Beyond flexibility, NoSQL technology also offers enterprises greater agility in how they draw value from their data. NoSQL systems are generally easy for developers to learn and leverage, as they use open and object-oriented technologies such as JSON. The ability to use open-source standards in NoSQL programming means that data analysts can choose from a wider array of analytic tools as well.

A third advantage that NoSQL offers is that of scalability. Performance in a traditional RDBMS environment degrades as the database scales. There are workarounds to this problem—e.g., horizontally or vertically scaling the environment through more hardware or software licenses—but these tend to drive up cost and complexity while offering minimal improvements. NoSQL databases, on the other hand, are designed to get faster as they get larger, making them ideal for big data applications.



NOSQL IN THE CLOUD: A NEW PARADIGM FOR ACHIEVING HIGHER DATA SCALE AND VALUE

Although NoSQL databases feature the ability to scale out horizontally across hundreds of servers, few small and medium enterprises will actually implement them this way.

Instead, most enterprise will opt to deploy NoSQL databases in a hosted Cloud environment where they can leverage the power of hundreds of servers without actually having to manage, maintain and (most importantly) pay for them. In fact, you could say that the cloud puts the power of NoSQL within reach of the average enterprise.

Particularly for applications such as big data that are compute- and data-intensive, a cloud-based solution allows enterprises to hand off the complexity and minimize the cost of deploying a NoSQL solution by offering:

RAPID DATABASE DEPLOYMENT

VIRTUALLY LIMITLESS SCALABILITY

HIGH PERFORMANCE WITH NO IMPACT TO OTHER DATA CENTER APPLICATIONS

EXCEPTIONAL SECURITY, AVAILABILITY, RELIABILITY, AND GRACEFUL
DISASTER RECOVERY

COMPLIANCE WITH INDUSTRY
REGULATIONS FOR SECURITY
AND PRIVACY

A COST-EFFICIENT,
PREDICTABLE,
PAY-AS-YOU-GROW MODEL

The cloud's inherent scalability allows enterprises to achieve significant performance improvements as well. For example, a NoSQL calculation that takes 30 seconds to perform on a single server would take less than one second to complete in a cloud environment that featured dozens of servers working in parallel.

If the cloud provider uses a CPU consumption model to base its price, the enterprise would actually pay the same amount whether the calculation is performed on one server or 100 servers—since the amount of CPU consumed is identical in either case—but would receive a much faster response time as a result.

While NoSQL databases in the cloud is nothing new, a new breed of DBaaS offering is emerging that is different. Many database services today allow users to provision a number of server instances in the cloud, but those services still require provisioning and resharding of the database as it grows over time. The new breed of DBaaS is built in the cloud and it scales instantly, so users never have to worry about the number of server instances on which their database is running. This is a subtle but important distinction.

8 THINGS TO LOOK FOR IN THE CLOUD

Enterprises today are more comfortable with cloud technology than in the past. They understand that it is not only a **cost-effective** path to the future, but also an open path to new technologies, seamless mobility, and unprecedented scales of efficiency.

That said, not all cloud environments offer the same levels of performance, security, scalability, and cost-efficiency. In evaluating a DBaaS cloud provider, enterprises should carefully weigh these eight criteria:



ACID (Atomicity, Consistency, Isolation, Durability) Compliance

ACID compliance is a requirement for many business applications, yet the majority of NoSQL DBaaS providers do not yet offer it.

As a relatively new technology, NoSQL lacks the standardization found in mature RDBMS platforms, but that doesn't mean that enterprises should sacrifice ACID compliance for innovation. Instead, for applications where ACID compliance is a requirement, enterprises should look for a DBaaS provider that offers this as part of their solution.



High Reliability and Availability

One of the key advantages of cloud computing is its redundancy. A well-architected cloud environment will offer geo-redundancy and stateful disaster recovery, which is especially important for large-scale analytic applications.



Instant Scalability

Another key advantage of cloud computing is instant scalability. With the cloud-based database model, developers and enterprises can distribute their data among many servers, thus providing on-demand scaling and low-latency. Users will have access to whatever computational resources they need, which is critical for computationally intense queries. Not only does a DBaaS scale on demand instantly eliminating the need to provision resources upfront, but it also runs much faster, due to massive parallelism and without a significant investment in hardware. Imagine computationally intense request being completed 100 times faster and cheaper due to massive parallelism that is possible in the cloud.



Real-Time Analytics

A common reason for cloud resistance among IT departments is the inability to control the network experience. Users expect real-time responsiveness, especially in the case of big data analytics. Enterprises should look for a DBaaS provider

that can deliver these requirements and scale their network's performance accordingly as user demands spike.



Minimal Learning Curve

As enterprises move to a mix-and-match technology approach, it's important that they leverage the same core skill sets across applications and environments. Not all DBaaS platforms are easy to use, nor do they all support the same standards. Enterprises should look for a DBaaS platform that aligns with the existing skill sets of their database analysts and application developers, such as XML, SQL, and JSON.



Scalability of Cost

Most enterprises assume that data-based initiatives will scale out over time, especially in lieu of recent data growth trends. Not all clouds scale out well from a cost perspective, however, depending on the type of cost model they use. Some DBaaS providers charge by gigabyte, others by the number of queries, and still others by computing cycles. Calculate the cost of the solution across multiple scenarios to make sure you enter the relationship with your eyes (and not your wallet) open.



Integration with Legacy Databases and Applications

As most enterprises have already learned, data doesn't work well in a vacuum; it needs to be consolidated, integrated, and validated before it can be truly appreciated. The most effective DBaaS platforms are those that allow enterprises to integrate data initiatives and applications across their enterprise.



Monitoring and Metrics

Some cloud providers are, frankly, *cloudy* when it comes to providing clear network performance metrics and usage statistics. Simply because a system isn't on premise doesn't mean that IT departments should be out of the loop. Instead, they should expect daily reports on their own usage metrics, query responses, and other key performance indicators.

INTRODUCING THE CLUSTERPOINT DBAAS SOLUTION

Clusterpoint's cloud-based Database as a Service delivers a robust, ACID-compliant, document-oriented database solution for enterprises and application developers. With hundreds of customers, Clusterpoint offers a proven, stable NoSQL database platform that can support real-time access and analysis of billions of documents.

The Clusterpoint DBaaS platform features rapid database deployment, scalability up to petabytes of data, sub-second database response times, powerful text searches, and much more—all at an attractive price-point based on a flexible, pay-as-you-grow model.

Clusterpoint's database platform is a complete NoSQL database system with best-in-class features and unique functionality...

It's 100% ACID compliant.

Clusterpoint's DBaaS platform delivers high-speed, ACID-compliant transactions to ensure data security and consistency. The majority of cloud-based NoSQL solutions cannot make this claim.

It has full-text search built right into the database.

Simply type in a keyword, and in milliseconds Clusterpoint's powerful text-based search displays the most relevant information from your database. Compared to SQL-based queries, Clusterpoint's text search is dramatically faster!

But it also supports SQL.

Data analysts can easily use SQL queries to retrieve data, organize it, analyze it, and package it into reports. Support for SQL, combined with text-based searches, gives Clusterpoint a powerful one-two punch that beats other cloud database platforms.

It's an OLTP database too.

Enterprises can use Clusterpoint as a business intelligence platform and a fully transactional OLTP database in one solution.

It's designed to minimize information overload.

Clusterpoint uses a unique feature known as *relevance ranking* that intelligently sorts data and displays the most relevant results first. It works by ranking and indexing data based on the current query, which reduces the amount of data that needs to be processed and displayed. The end result: much faster queries that provide more relevant answers.

It gives you a viable alternative to Hadoop.

Clusterpoint has the ability to execute arbitrary JavaScript code at any step during query processing, which allows enterprises to build powerful and extensible analytical applications that can run computations alongside the data using indices. This leads to ultra-fast query processing that can rival Hadoop in terms of analytical performance.

It's the best performer for the price.

Clusterpoint's pricing model is designed to give enterprises more big data bang for their buck without penalizing them when they need more performance or scalability. Some DBaaS providers tout their low entry price, but Clusterpoint has no entry price: the Clusterpoint database software license is free to use, with unlimited storage and number of replicas. You pay only for the cloud capacity you use, and can scale up performance requirements on demand without incurring penalties.

It's really, really, really safe.

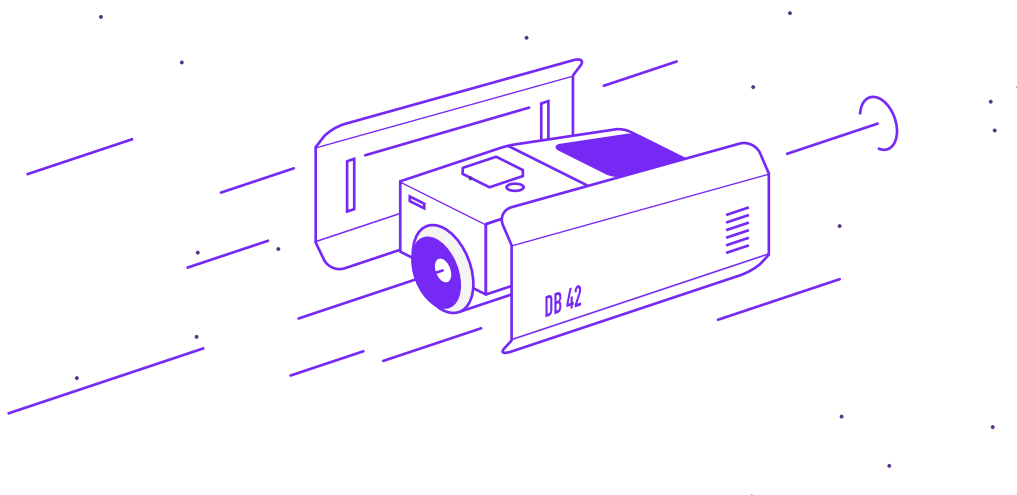
All cloud providers offer some kind of data redundancy. Clusterpoint offers *re-re-redundancy*: we always keep at least three replicas of your data in our secure cloud, because nothing is more important than your data.

It integrates with your premise-based systems.

You can install Clusterpoint's database software on your own hardware and integrate it with your data in the cloud. Both systems use the same application programming interfaces (APIs), so data and application sharing works the same inside and outside the cloud.

It's super scalable.

The Clusterpoint Cloud features a responsive multi-tenant model that quickly shifts compute power and storage capacity to your database applications on demand. Unlike many DBaaS providers, we don't "double charge" customers for storage and speed; we charge them only for the processing they use. This frees customers to tap into more processing performance when they need it, without fear of paying for it when they don't need it.



CLUSTERPOINT IS WHERE DATA IS MOVING

Big data may seem like a recent anomaly; it is, in fact, the new reality. As billions of connected devices contribute to the crush of data being created on the Internet of Things, more businesses will rely on large-scale data analysis to navigate the world and the real-time opportunities around them.

Cloud computing will become the *de facto* standard for business applications, communications, and data as IT leaders re-envision their role from managing information to managing innovation. And NoSQL databases will become more common as non-traditional data types become the norm.

In this new information landscape, Clusterpoint is poised to help enterprises evolve from data drowned to data driven. We offer a proven, no-risk path to big data storage and analysis based on advanced NoSQL database technology and tools that have been tested and validated in hundreds of enterprise engagements and developer hackathons around the world.

There's no cost to get started and minimal training required; most developers begin using Clusterpoint within 24 hours of downloading it.

To learn more or download your free copy today, visit us online at www.clusterpoint.com.

Footnotes

1. Gartner, Inc., "Survey Analysis: Practical Challenges Mount as Big Data Moves to the Mainstream," 2015.
2. IBM Corporation, "It's Time to Evolve the Way the World Works with Data" Infographic, 2014.
3. PC World, "Timeline: 50 Years of Hard Drives," 2006.
4. MarketsandMarkets, "Cloud Database and DBaaS Market by Databases (SQL, NoSQL), Software (Database Application Builder, Data Scaling and Replication, Back-Up and Recovery, Database Encryption), Deployment Models (IaaS, PaaS, DBaaS) and Service Models (Public Cloud, Private Cloud, Virtual Private Cloud, Hybrid Cloud) & Organization Size (Small and Medium, Large) – Global Advancements, Markets Forecasts and Analysis (2014-2019)," 2014.