Clusterpoint Server

A totally searchable and scalable database technology

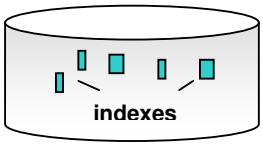
Clusterpoint Server technology enables to design, develop and operate totally searchable and scalable databases. It represents less complex and more affordable database technology for many contemporary Internet and corporate Web applications where instant data access, search and retrieval in large databases are essential.

What does it do?

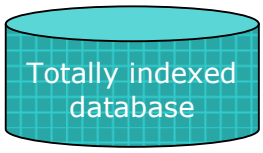
- Solves data growth problems
- Provides full content data access
- Radically improves data access speed
- Improves the customer experience
- Reduces complexity of IT systems
- Increases your productivity
- Saves time and money

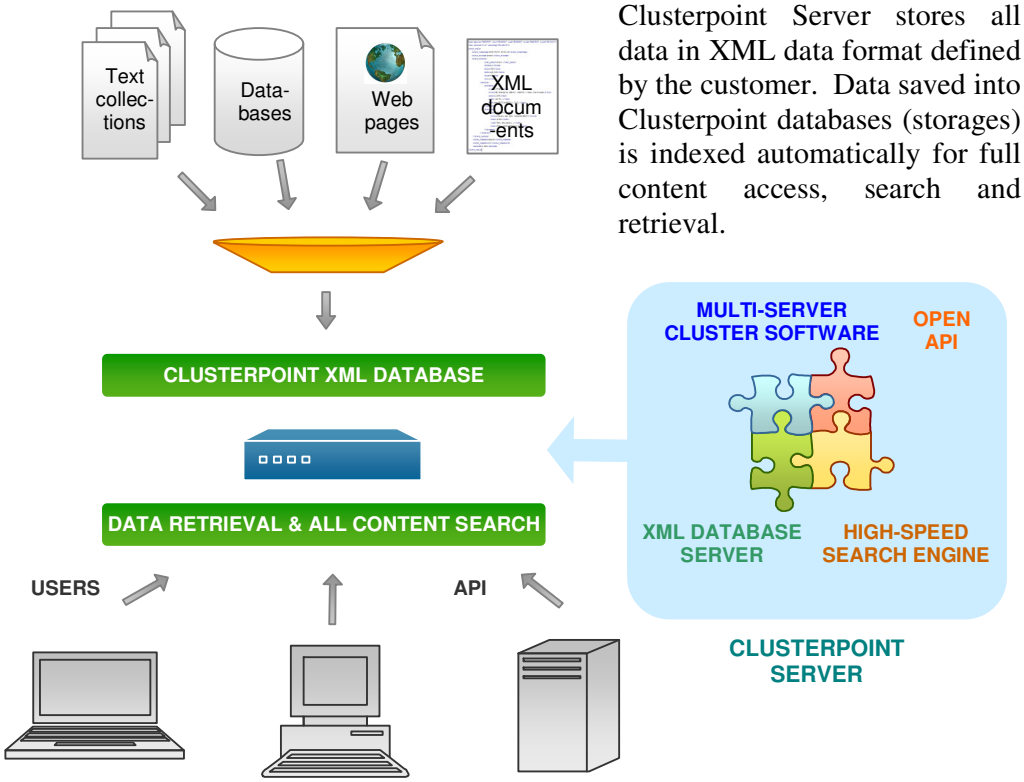
Build database software applications in a new, more simple and efficient way. For many types of existing database applications Clusterpoint's data retrieval tasks can outperform relational database systems by a factor of 1000 or so.

- In a relational database model, limited data are indexed for fast search:

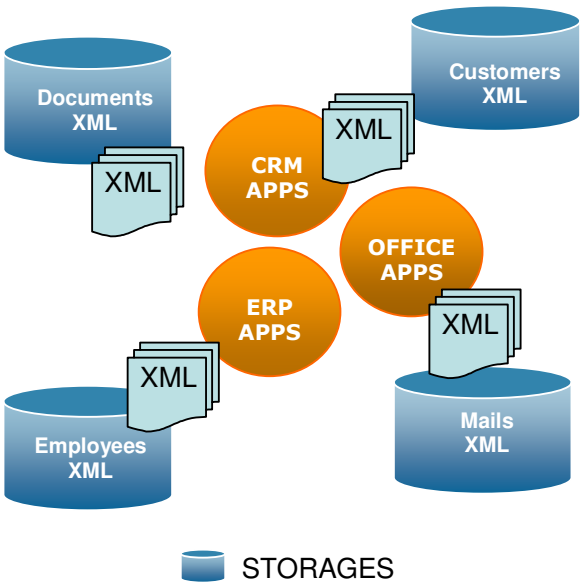


- With Clusterpoint's Server, all data is indexed for fast search:





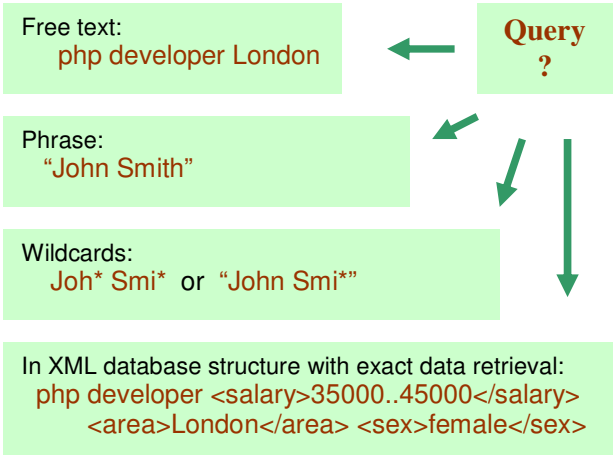
Clusterpoint Server stores all data in XML data format defined by the customer. Data saved into Clusterpoint databases (storages) is indexed automatically for full content access, search and retrieval.



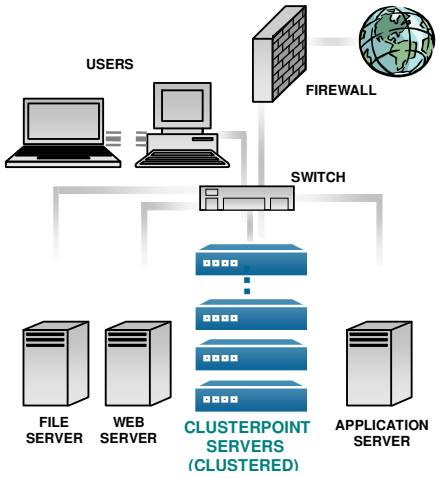
Remove the need to handle SQL database columns and indexes in your application software. Work only at storage level with your own real-life data objects in XML. Handle any custom XML database objects with simple storage-update-retrieval operations in your software.

Clusterpoint infrastructure DBMS software is engineered using best-of-breed industry standards: client-server database architecture, XML database server, uniquely designed index covering entire data content, high speed structured and unstructured data search engine, open XML based API messaging system using robust HTTP protocol and transparent cluster software for multi-server database scalability.

Searching Clusterpoint databases is made easy and understandable to most users who can search any database without the complexity of SQL. Simplicity and efficiency of data access is illustrated here by some query formats, for an example a CV database (from any keywords to structural search in XML).

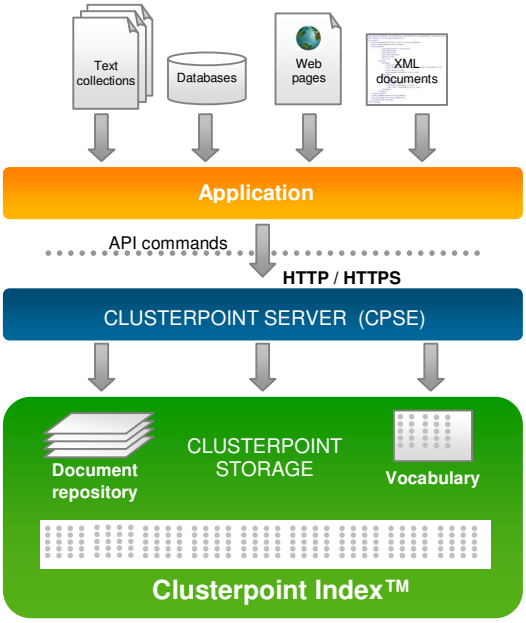


Clusterpoint's technology was designed for building and operating scalable databases of massive size. The platform distributes computing tasks across a cluster of networked servers, providing instant and relevant data retrieval from very large databases. The Clusterpoint Server acts as a grid database software utilizing total power of all available processors, RAM and disk systems. As a result, customers can scale massive databases in a cost-efficient way. Customers also benefit from radically better data access times in a cluster database. The key benefits of this approach are as follows:



- Outperforms RDBMS in ad hoc queries, typically from 10 to 100 times;
- Query response times below 0.5 seconds across a multi-server cluster database;

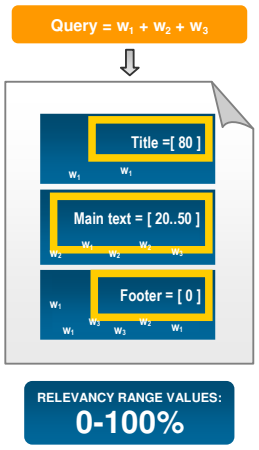
- Linear scalability addresses data growth and performance needs cost-effectively;
- Simplified database life-cycle management and ease of information access compared with more complex legacy database technologies;
- Centralized administration of all databases and servers through a Web browser-based interface, allowing users to configure and manage cluster storages productively.

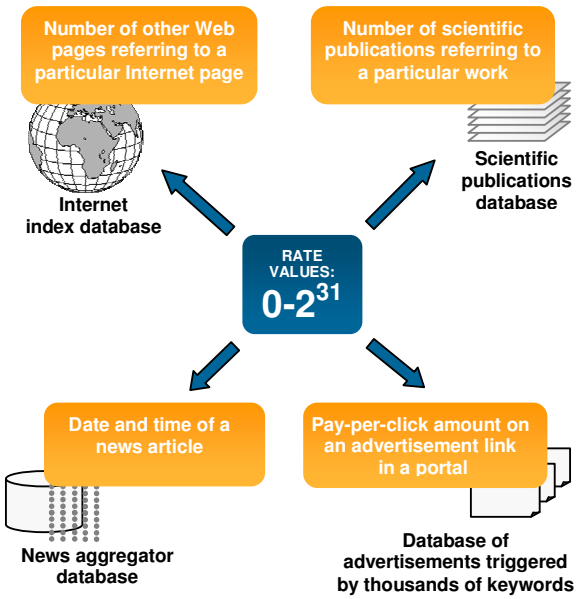


Clusterpoint Server automatically creates full content database index: Clusterpoint Index™, which is used for high performance data access and search in all stored documents.

Clusterpoint Index™ is based on multiple search quality dimensions: relevancy and rates.

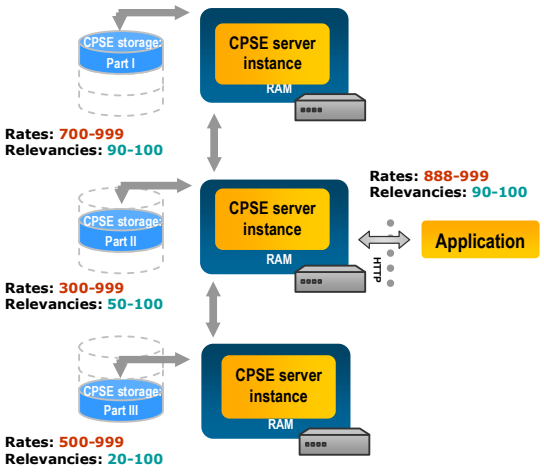
Relevancy is used to define custom search ordering rules for any part of the XML document. It brings search results listed in the exact order that a specific application may need. Unlike SQL queries having a flat relevancy for all results, Clusterpoint technology allows users to define flexible custom ordering algorithms for query terms matches to specific XML document parts, for example, to treat search hits in document titles higher than hits in the main text or a footer. Another example - any hits in an e-shop product names list could be treated more relevant than hits in product specifications or user comments etc. Relevancy is based on relative weights in the Clusterpoint system.





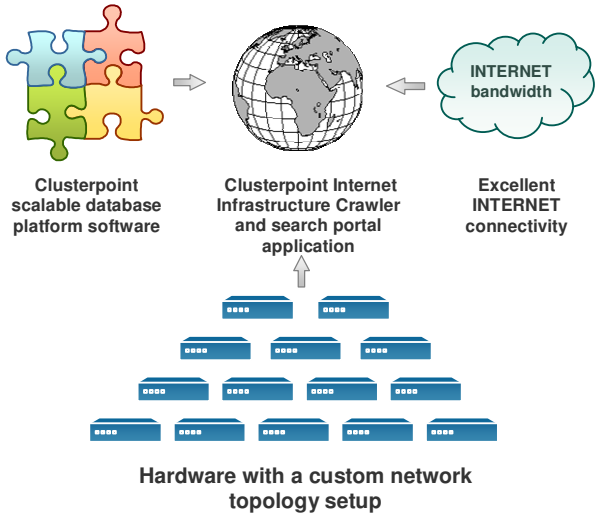
Rate dimension provides another scale for precision data quality definition. Rate is a custom integer value used for scalable and fast data retrieval from massively clustered databases. End users can instantly access the most relevant data from very large databases, containing many billions of XML documents each having a custom rate value.

Multiple data quality dimensions of the Clusterpoint Index™ create the foundation for the totally searchable and scalable DBMS platform. Users can access and search massive databases almost instantly.



The unique design of the Clusterpoint Index™ allows users to organize distributed database storage in a multi-server cluster environment using commodity hardware components. Clusterpoint DBMS handles total data volume in a scalable way where each cluster node runs its own small part of the database. Clusterpoint Server core software was engineered in C/C++ to deliver fast data access and filtering in this massively scalable distributed database architecture. The system works mostly on pre-sorted data which also minimizes disk access needs. Any queries are performed in parallel on all cluster nodes and search results are automatically merged by Clusterpoint DBMS software. The resulting system delivers sub-second search performance in a managed transaction processing environment required to run large scale mission-critical database applications.

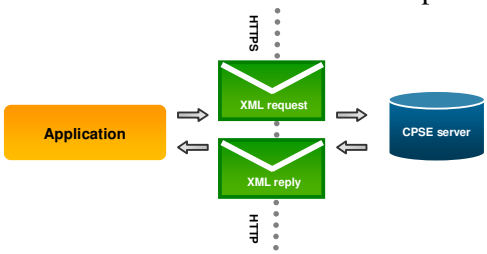
Clusterpoint databases can scale to run on hundreds and thousands of computers in a single cluster. Clusterpoint DBMS software can be used as a scalable platform in the Web infrastructure projects. For example, Clusterpoint Internet Infrastructure Crawler application may be used for automatic Web link spidering. All that data can be stored on the Clusterpoint servers connected into a custom network topology to create and operate large scale Internet search indexes.



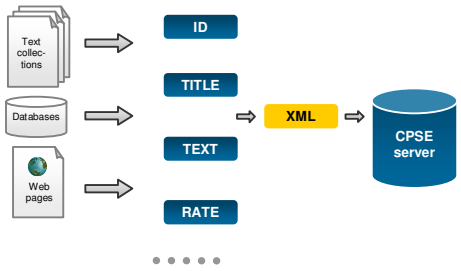
One of the most cost-effective data processing approaches today is to use a scalable database technology at infrastructure level to run large cluster databases using expendable commodity hardware. By distributing computing tasks on simple and standard infrastructure components the cost of maintaining and operating the total system is fixed per user and per gigabyte of data. The costs do not escalate when system capacity needs to be increased. The total cost of system ownership can be planned using a fixed costs model, representing an attractive return on investment. This approach has been proven by some of the world's largest Internet companies providing Internet services to tens and hundreds of millions of users. Clusterpoint technology is an open and affordable database infrastructure software platform

providing similar performance, scalability, functionality and feature set for any custom data structure.

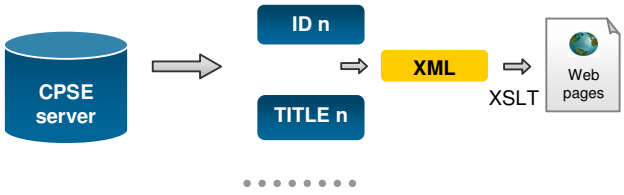
There are no barriers to testing and using Clusterpoint's technology for large, scaleable databases. There is an open standards-based and easy to use API available for developers. The Clusterpoint API uses robust 'request'-'reply' XML messaging: a simple, fast & secure transaction system over HTTP.



The Clusterpoint Server stores any custom user XML document and automatically creates an index on the XML document's internal data when Clusterpoint XML '**request**' message with API '**insert**' or '**update**' command is sent to the Clusterpoint Server. In order to search, the Clusterpoint XML '**request**' message with API '**search**' command is sent to the Clusterpoint Server again – this time with the content part of the Clusterpoint XML envelope contains the user's search query.



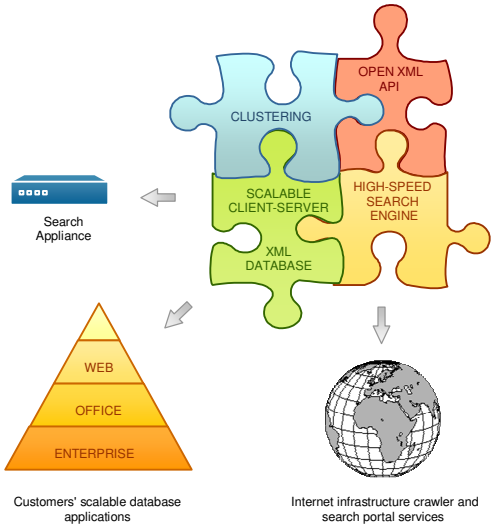
Transaction results are always returned from the Clusterpoint server as XML '**reply**' messages. Replies are formatted for easy parsing by any programming language (Java, PHP, .NET etc.) and contain technical parameters for easy construction of multi-page navigation systems for Web database applications.



In order to create a Clusterpoint database, the customer chooses a unique document <id> (e.g., URL, file name or database object identifier), assigns a <title> tag to be shown in search results listing, and assigns a custom <rate> value for results ordering according to specific business needs. Despite Clusterpoint's simple and robust API, there is a rich feature set available for developers to exploit advanced details and functionality of the Clusterpoint Server: more than 150 software developer options and system configuration options. The platform also has language support enabling to store and query data in multiple languages in the same storage in UTF-8, or to use multi-language terms in a single search query.

Innovative Clusterpoint XML database technology could power many applications built with the latest Web technologies. It could also provide a more cost-efficient, safe and straightforward way to manage databases. Clusterpoint can help to reduce the complexity of customers' IT systems:

- Use existing corporate IT knowledge
- No learning necessary for basic data access
- Work with real-life data objects in XML
- Build reusable software code for XML objects
- Change database structure on-the-fly
- Perform evolutionary data migration as needed
- Reduce workload on RDBMS servers
- Implement cross-platform data objects and use them in any programming language
- Platform-independent long-term storage in XML format can last for decades
- Rapidly develop new application software versions for a live database
- System administration is cost-efficient and does not require special training
- Easily build aggregated XML databases from many different data sources and implement instant ad hoc enterprise knowledge discovery without costly data warehousing
- Integrate new functionality into existing applications using just XML and HTTP



About Clusterpoint:

Clusterpoint was founded in 2006 by technology entrepreneurs based in the UK and software developers based in Latvia. Clusterpoint's team of IT professionals has a proven track record of expertise in developing, deploying and operating Internet and corporate database and search solutions.

Clusterpoint distributes its software products and related services through local partners: system integrators, IT solution providers, application services providers, OEMs and equipment manufacturers (through software licencing). The company has a broad international strategy to expand its partnerships in global markets.

Contacts:

Mr: Gints Ernestsons
Phone: +371-29479905
Email: info@clusterpoint.com
Website: www.clusterpoint.com
Skype: clusterpoint

Business development office:
Imprimature Capital
2nd Floor, Staple Hall
Stone House Court
87 – 90 Houndsditch
London
EC3A 7NP
United Kingdom
Email: tm@impcap.com

Research and software development office:
Clusterpoint SIA
Gertrudes 37-8
Riga, LV-1011
Latvia
Phone: +371 29479905
Email: info@clusterpoint.com

