rackspace®
the open cloud company

corporation*wiki* ®

Hybrid Hosting Powers Big Data Processing for Corporation Wiki
RackConnect creates an optimal compute environment.

Corporation Wiki promotes corporate transparency by being a historical archive of corporate data, creating a valuable resource for people interested in discovering and understanding connections between corporations and executives. Corporation Wiki serves over 300 million page views per month to visitors and search engine web crawlers by connecting their dedicated Rackspace servers to Cloud Servers with RackConnect capability, allowing the company to serve its visitors quickly and at scale.

Mike Prince, Founder of Corporation Wiki, explains how a Rackspace Hybrid Hosting solution enables his company to conquer big data processing in a timely manner, "We process large amounts of data from sources all over the country. Spinning up dozens of cloud servers helps us get through big data processing jobs in hours instead of weeks.

"We process large amounts of data from sources all over the country. Spinning up dozens of cloud servers helps us get through big data processing jobs in hours instead of weeks. Rackspace's ability to manage and connect our dedicated servers to their demand-based cloud instances is unlike anything else out there."

Mike Prince,
Founder, Corporation Wiki

Rackspace's ability to manage and connect our dedicated servers to their demand-based cloud instances is unlike anything else out there."

Rackspace provided a big data processing solution for Corporation Wiki that allows the company the flexibility to grow and scale, creating an optimal compute environment for them. RackConnect rendered that freedom with a hybrid hosting solution, enabling them to unify their original managed dedicated server, clustered SQL Servers, coupled with Windows cloud servers, and topped off with a load balancer.

SCALING WITH RACKSPACE
Over the past few years, Corporation Wiki has grown into one of the largest corporate directories on the web. They have created a corporate directory of over 90 million

> "Corporation Wiki would not be where it is today were it not for the ability to help scale, support, and maintain a high performance infrastructure from Rackspace. This support allows the team behind Corporation Wiki to focus on the programming required to bring such a large site online."

Mike Prince,
Founder, Corporation Wiki

officers and executives and have successfully mapped and connected these companies and individuals to each other.

Initially, Corporation Wiki relied on a single server maintained and managed by Rackspace. "As traffic grew, we added another server in order to cluster the Microsoft SQL Server database and front-end web services were moved to the cloud and made to scale by employing a dedicated Big-IP F5 load balancer in front of the cloud web servers," says Prince. "This RackConnect hybrid network allowed us to run their database on a high-capacity and highly-available server cluster while scaling the web front-end servers through a high performance, dedicated load balancer."

Corporation Wiki obtains a majority of its data from public record sources which are large in nature. "These data stores add up to terabytes of data," says Prince. "In order to clean, standardize, and unify the data in a reasonable amount of time, we break these large jobs into smaller jobs and run them on dozens of cloud servers. These server instances are spun up as needed and the import and processing jobs complete by a factor of 100 times faster. Cloud instances are essential in the future growth of the data behind corporationwiki.com."

LOOKING INTO THE FUTURE

Corporation Wiki is adding new datasets in their effort to create a comprehensive historical archive of every company that ever existed. In addition to their Hybrid Hosting solution, Rackspace Cloud Files helps make this possible by providing a scalable data store for images. "Cloud Files nearly limitless storage capability and scalable performance make this much more cost effective than attempting to do this on dedicated hardware," says Prince.

Corporation Wiki search will soon be powered by an open source technology called Elastic Search that provides scalable and distributed full-text search. "This will allow the website visitors the ability to search through terabytes of data quickly," says Prince. "As the data grows in Corporation Wiki, Elastic Search allows us to maintain search performance by simply adding new cloud servers to the search cluster in order to serve increasing amounts of data to an increasing number of visitors."

"Corporation Wiki would not be where it is today were it not for the ability to help scale, support, and maintain a high performance infrastructure from Rackspace," concludes Prince. "This support allows the team behind Corporation Wiki to focus on the programming required to bring such a large site online, and continue to grow and scale into the future."

rackspace®
the **open cloud** company