Big Data on the Open Cloud

Rackspace® Private Cloud, Powered by OpenStack®, Helps Reduce Costs and Improve Operational Efficiency

Written by Niki Acosta, Cloud Evangelist,
Rackspace®

rackspace
the open cloud company

Big Data on the Open Cloud | Cover
RACKSPACE® HOSTING | 5000 WALZEM ROAD | SAN ANTONIO, TX 78218 U.S.A

Table of Contents

rackspace
the **open cloud** company

1. Introduction

Rackspace® Enterprise Business Intelligence group (EBI) is a central team that aggregates, manages and provides business intelligence on data from several business-critical data sources. To keep up with Rackspace's customer growth and technology infrastructure, EBI wanted to consolidate the rapidly-growing volumes of data for reporting, trending, and analytical purposes. This white paper highlights how EBI used Rackspace Private Cloud Software to power a cloud-based big data solution while reducing costs and improving operational efficiency.

2. Turning Bytes into Business Intelligence

EBI's legacy data warehouse consists of commercial database vendor solutions on dedicated servers. Data points included customer account data, usage and billing information, with business intelligence toolset interoperability from Informatica and Qlikview. From an operational level, the overall data became unmanageable once important information like monitoring, response, and support metrics came in from dedicated, virtual, and cloud devices.

Daily reporting became a time consuming and resource-intensive process, only occurring nightly and with a 24-hour data point lag time. Commercial database licensing and hardware costs were rising in a disproportionate manner as the EBI team worked with database administrators to quickly increase capacity during peak hours. Finally, the legacy set up did not handle unstructured data very well, and the team wanted to be able to apply different best-of-breed technologies (e.g. columnar, noSQL, SQL) alone or in combination depending upon the type and size of data they wanted to store and analyze.

To continue serving the business efficiently and effectively, EBI put together requirements for a new solution. Named the Analytic Compute Grid (ACG), the solution would act as the backbone for EBI and needed to be able to:

- House an ever-growing set of data collected in different formats, structured and unstructured, from multiple business units within Rackspace

- Rapidly and dynamically scale resources up and down to efficiently meet business demands

- Add new resources on the fly without waiting for new hardware provisioning during peak hours

- Run different, best-of-breed, big data technologies for storing, managing, analyzing and distributing data on one technology platform

- Enable the EBI team to move away from rising commercial database licensing fees

- Utilize open APIs to facilitate integration and programmatic access with other enterprise systems and BI tools

- Support Rackspace security and compliance requirements

- Embrace open cloud and open source technologies

rackspace
the *open cloud* company

With those requirements in mind, the Rackspace EBI team then evaluated the following options:

Requirement/Options	Current System	MPP Appliance	Legacy on Virtualized Platform	Open Technologies Stack
Support massive data volumes	◔ (quarter)	● (full)	◑ (half)	● (full)
Agility: scale up/down — add resources on the fly	○ (empty)	◑ (half)	◑ (half)	● (full)
Save licensing costs	○ (empty)	○ (empty)	◔ (quarter)	● (full)
Support multiple data formats and technologies	◔ (quarter)	◑ (half)	◔ (quarter)	● (full)
No vendor lock	○ (empty)	○ (empty)	○ (empty)	● (full)
Open APIs to support integration with enterprise systems	○ (empty)	○ (empty)	○ (empty)	● (full)
Lower migration costs	● (full)	◔ (quarter)	◑ (half)	○ (empty)
Support compliance and privacy requirements	● (full)	● (full)	● (full)	● (full)

Option 1: Stay the Course

- Pros
  - Short-term minimal interruption to existing projects and end-users
  - No additional training necessary
  - Could continue to leverage vendor support

- Cons
  - Licensing costs that spiked as data volume increase
  - Database administration (DBA) support for resources spread across multiple OLTP databases and BI databases.
  - Scalability of systems – to grow the current system is very time consuming in conjunction with growing data volumes
  - Current technologies offer no support for big data
  - Legacy commercial database products do not scale performance with data volume. Making these products scale would require complex clustered footprints of servers. In addition, both vendors recommend their own proprietary infrastructure and database technology.

rackspace
the *open cloud* company

Option 2: Purchase an MPP (Massively Parallel Processing) Appliance

- Pros
    - o High-performance
    - o Purpose-built for BI workloads
    - o Interoperability with existing BI toolsets
    - o Large BI customer base with a rich feature set provided by vendors
- Cons
    - o High costs relative to current environment, including cost to acquire appliance, set up fees, licensing, maintenance, training, etc.
   - o Proprietary hardware configurations and database engines

Option 3: Running Legacy BI Apps on Commercial Virtualization Software

- Pros
    - o More efficient than running on physical hardware
    - o Some elasticity to "scale up" the VMs and expand footprint
    - o Relatively easy migration of legacy BI apps to virtualized infrastructure
- Cons
    - o Limited "scale out" capabilities and resource-sharing as compared to a cloud environment
    - o Additional licensing costs
    - o Concerns of building on and getting locked into proprietary and licensed commercial virtualization software

Option 4: End-to-end Open Source Solution on Rackspace Private Cloud

- Pros
    - o Enables scaling out and back faster than siloed hardware or virtualized servers
    - o An entire open source technology stack – avoiding vendor lock-in
    - o Ability to leverage commodity hardware
    - o No software licensing costs
    - o Take advantage of faster innovation in open source platforms due to community participation and contribution
    - o Ability to leverage public cloud resources where appropriate
- Cons
    - o Training developers and end users on new technologies
    - o Large migration
    - o Must build, buy, or find adaptors for BI tools

3. The Choice: End-to-end open source solution on Rackspace Private Cloud
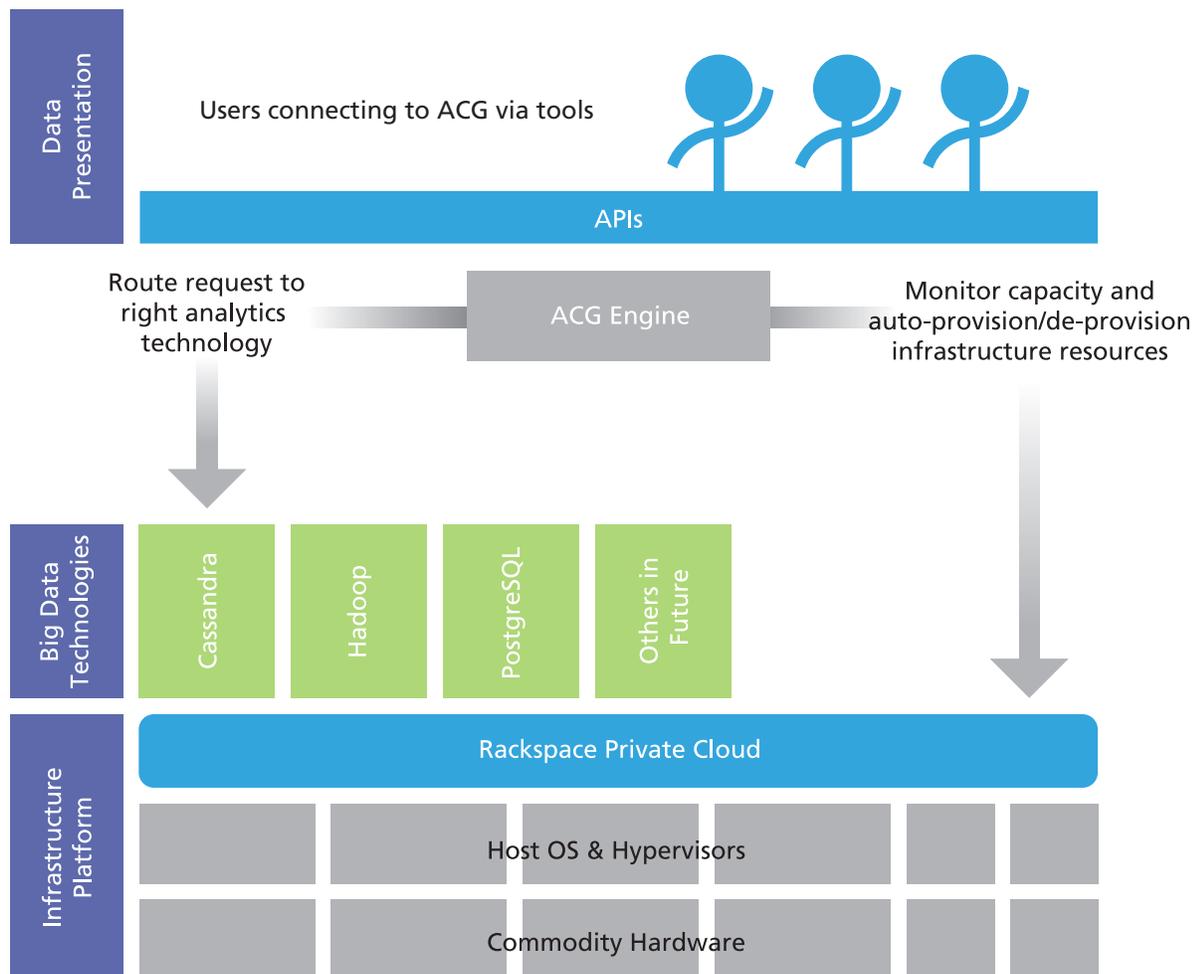
These requirements led EBI to design and build a stack based on open source technologies – from infrastructure to big data software – to allow for rapid growth and scale. The underlying infrastructure platform they selected was Rackspace Private Cloud, powered by OpenStack®, in tandem with Cassandra, Hadoop, and PostgreSQL. The solution was dubbed as Analytic Compute Grid or ACG.

ACG is a big data management software platform built on Rackspace Private Cloud software. As a key benefit, it provides a consolidated and flexible solution to store, analyze, distribute and present the data based on the type of the data (structured or unstructured), operation (storing or analyzing the data) and the consumer's skillset (data scientist accessing via APIs or a marketing analyst using BI tools to run reports.)

rackspace
the **open cloud** company

4. The Results

- The EBI can now process terabytes of data per day in real-time or on-demand

- Processing tasks that took six days on the legacy system have been reduced to three hours

- Existing BI tools can be leveraged by custom ANSI SQL APIs, and additional technologies can be easily added via extensions

- The ACG reduced the need for two additional administrators

- Improved trending and reporting data is currently being utilized to enhance support capabilities and the Rackspace customer experience

5. Conclusion

By creating a single holistic platform utilizing open source technologies, the Enterprise Business Intelligence team's Analytic Compute Grid can handle the storage, analysis and distribution of data at scale in a timely manner. The big data tools available today helped solve the problem but required new ways of thinking about the underlying infrastructure, processes and data structures to make it a reality. Built using Rackspace Private Cloud, powered by OpenStack, Hadoop, Cassandra, and other tools, the ACG has resulted in improvement in data processing speeds and a significant reduction in overall capex and opex. Multiple business units at Rackspace can now make near real-time decisions that can directly benefit Rackspace customers.

rackspace®
the **open cloud** company

6. What Do You Need to Solve?

Rackspace Private Cloud, powered by OpenStack is free software that allows you to run a Rackspace Cloud in your data center. The fastest and most cost-effective way for your enterprise to leverage open cloud technologies at scale is to choose a knowledgeable cloud provider that understands and uses it every day — and is standing ready to help match your business needs with the appropriate open cloud solution.

Additional information on Rackspace Private Cloud is available at www.rackspace.com/cloud/private.

DISCLAIMER

This Whitepaper is for informational purposes only and is provided "AS IS." This Whitepaper does not represent an assessment of any specific compliance with laws or regulations or constitute advice. We strongly recommend that you engage additional expertise in order to further evaluate applicable requirements for your specific needs.

RACKSPACE MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, AS TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS DOCUMENT AND RESERVES THE RIGHT TO MAKE CHANGES TO SPECIFICATIONS AND PRODUCT/SERVICES DESCRIPTION AT ANY TIME WITHOUT NOTICE. RACKSPACE RESERVES THE RIGHT TO DISCONTINUE OR MAKE CHANGES TO ITS SERVICES OFFERINGS AT ANY TIME WITHOUT NOTICE. USERS MUST TAKE FULL RESPONSIBILITY FOR APPLICATION OF

ANY SERVICES AND/OR PROCESSES MENTIONED HEREIN. EXCEPT AS SET FORTH IN RACKSPACE GENERAL TERMS AND CONDITIONS, CLOUD TERMS OF SERVICE AND/OR OTHER AGREEMENT YOU SIGN WITH RACKSPACE,

RACKSPACE ASSUMES NO LIABILITY WHATSOEVER, AND DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO ITS SERVICES INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT.

Except as expressly provided in any written license agreement from Rackspace, the furnishing of this document does not give you any license to patents, trademarks, copyrights, or other intellectual property. Rackspace, RackConnect and Fanatical Support are either registered service marks or service marks of Rackspace US, Inc. in the United States and/or other countries. All other product names and trademarks used in this document are for identification purposes only to refer to either the entities claiming the marks and names or their products, and are property of their respective owners. We do not intend our use or display of other companies' tradenames, trademarks, or service marks to imply a relationship with, or endorsement or sponsorship of us by, these other companies.