Charter Solutions, Inc.

A Big Data Solution for Time-Series Data

Charter Solutions, Inc.
Atria Corporate Center
3033 Campus Drive, Suite N160 | Plymouth, MN 55441
763.230.6100 (Main) | 763.230.6110 (Fax)

Thinking & Technology
That Move You Forward

Big Data Solution for Time-Series Data

Introduction

Charter Solutions recently worked with a global company interested in putting together a solution to continuously collect, interpret and act on time series data key to their managed service offering. While they had attempted to create a solution using traditional approaches, they were not able to achieve the desired overall reliability and up-time for this key product offering. They were looking for alternative technologies to further broaden capabilities and expand business value to customers. In order to validate a new approach, a proof of concept was architected and designed. This case study looks at this challenge, the solution identified, and the proven value achieved for the organization.

Charter Solutions Background

Charter Solutions is a Minnesota-based consulting firm specializing in Management Consulting, Analytics and Information Management Services, and Application Development. We make a difference by working in partnership with our clients to deliver leading-edge IT solutions aligned to the business need. From small advisory or development projects to large transformations, our staff of highly experienced business and technology professionals are known for crafting simple, elegant solutions for tough business problems.

We pride ourselves in helping our clients achieve their desired results in a timely and cost effective manner by applying technology appropriately to address the business need. Our overall goal is to work on projects that create value to our clients that can directly benefit the bottom line. We provide services at various organizational levels from strategic to tactical for many Fortune 1000 clients.

Our Big Data offering, an Analytics and Information Management Service, is aimed at helping our customers find value from datasets that were previously inaccessible due to size, transaction volume, or lack of clear structure. Our solutions leverage big data technologies, cloud infrastructures, and innovative visualization technologies.

Challenge

The challenge was to review current technology, which supports the client's portfolio of services, to ensure that they can scale to growing market demands, continue to meet their clients' Service Level Agreements, and to easily and quickly add services to their customers.

"…[the customer] anticipates collecting approximately 600 million data points per day which will result in over 40 terabytes of data each year."

The first challenge was overcoming the belief that standard IT software applications, hardware and design practices were sufficient to meet the key business needs. There is a comfort level in the fact that existing technology and support approaches are a known quantity. However, while this is highly effective for management of internal systems, it is not always the best solution for all business needs – as is the case here. A new approach needed to be championed in order to meet the business objectives.

The design of the current system, while functional, required higher levels of developer effort to implement the desired changes. The new system needed to minimize the development time and effort in order to remain competitive.

The current technology is scalable, but reaching its performance and growth limits. Due to the client's current and projected growth, the technology needed to scale to tens-of-thousands of customer sites worldwide while maintaining response times and SLAs. In addition the volume of data collected and analyzed daily continues to grow. For example, in the near future the customer anticipates collecting approximately 600 million data points per day which will result in over 40 terabytes of data each year.

In addition to the challenges of growth, the business wanted to provide new and improved services to its customers, leveraging other data such as energy usage and weather patterns. As these new ideas are vetted and determined to be viable services, the organization will need to bring them to market quickly in order to maintain its market position. These services will require that their systems be easily adaptable to new and complementary data sources and allow them to continue to provide existing analytics but also provide new data and insights to its customers.

Along with faster time to market with new services, the client's customers demand that their data collection be available at all times and always working; this includes their UI, APIs, and analytics. With the ever-expanding growth in data volume a new technical solution was needed.

Solution

The client needed to architect a new solution for its customers. However, a customer-driven platform sets a very high bar as all issues become highly visible to all customers. Any and all outages, regardless of response time to correct the issue, have the possibility of negatively affecting the client brand. Any changes to the existing technology needed to prove that this transparency would improve brand image, not detract from it.

To achieve this goal the key architecture principles need to include:

- Redundancy
- No single point of failure
- Design assuming components will fail
- Increased performance

- Increased access to data

There are a number of technical challenges to overcome when looking at any new architecture. The client collects time series data at varying data collection intervals. In order to successfully analyze this type of data it would need to be normalized for analysis and graphing, requiring unique architecture principles to support it. Along with the nature of the data, the velocity and volumes collected were becoming greater than the client's current system could handle, making it a classic Big Data problem.

"[The new architecture incorporated] cloud based infrastructure services [and] a big-data database with time-series awareness…"

The key to the solution designed for this proof of concept was to build a simplified architecture that leveraged:

- cloud-based infrastructure services
- a big-data database with time-series awareness
- architecture layering based on elasticity needs
- light-weight, simple Java code to collect and process the data
- meta-data driven data structures and processing instructions to handle new and varying data sources

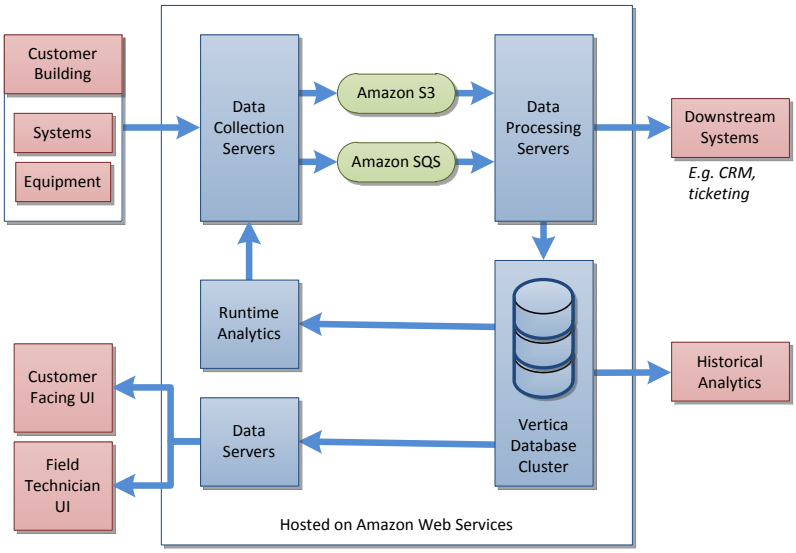
A cloud-based infrastructure provided the flexibility and speed to market that the client demanded, is scalable both up and down as the organizational needs change, has high availability, and would reduce costs. While there are a variety of technologies that could be used, Amazon Web Services (AWS) was chosen. AWS is the market leader in cloud-based infrastructure with a global footprint and is built using technology that would meet the client's other needs as well. AWS offers their Platform as a Service (PaaS) including their S3 data storage, SQS message queuing, and SNS notification which ultimately will lead to reduced development time.

Along with AWS' infrastructure the solution required a high performance analytic database. Vertica Database from HP was chosen which easily fit with AWS' cloud based infrastructure and was also highly scalable. Vertica allows users to add servers and capacity in a manner which would go unnoticed to the client's customers. Another key feature is its built-in time series gap filling capability.

Once the technical components were chosen, the Proof of Concept focused on building a data-collection system using this architectural approach and validating it under high volumes of load.

ARCHITECTURE DESCRIPTION

The diagram below shows the key elements of the new architectural approach.



In this architecture, new data is sent via HTTP requests. The data collection layer then accepts and stores the data to Amazon S3 and places processing requests on an Amazon Simple Queue Service (SQS) message queue. Using this approach allowed us to separate the capture of data from the processing and loading into the database. It also provided the ability to scale this layer independently of all other parts of the architecture, allowing it to absorb spikes easily. Amazon's S3 highly reliable/redundant file storage system essentially guaranteed that the received data will always be available for loading into the database. Additionally, Amazon SQS high-availability message queuing system was used for internal notifications and requests between architectural layers.

Once the data is collected and stored processing requests are picked up off SQS and processed by a separate layer in the architecture. The data files, primarily XML, are parsed and the measurements are stored in the Vertica database. Scalability is improved in this layer by assigning more or removing servers which also easily manages throughput and maintenance costs. In addition, processing of the data formats does not impede the communication between the installed devices and the data collection service maintaining high performance and availability.

An additional challenge for the client was their need to allow customers and field technicians who were sent to resolve issues to visualize the data collected. The architecture chosen for the proof of concept and the data servers provided easy, secure, and user friendly access to the data as well as caching to increase end-user performance and availability.

Benefit

"…over 60x performance improvement…"

"…cost improvements… of over 50% annually…"

"…productivity increase of at least 40%..."

The Proof of Concept was deemed highly successful by the client's senior management and end users. There were a number of factors that contributed to this success. Development time was rapid and low cost with only a single FTE (two half-time resources) needed to build the Proof of Concept in just two months.

The greatest sign of success however was the over 60x performance improvement with near linear scalability. The cost improvements were also tremendous when looking at the current technology versus the proof of concept. The client estimates that they will see ongoing cost improvements with the new system of over 50% annually, along with a productivity increase of at least 40% vs. existing technologies.

In addition to the monetary benefits there are other benefits that should be considered. Amazon AWS' development environment enabled higher speed development because of the ability to create and de-commission servers and environments on demand. This ability eliminated much of the wait time in infrastructure resource requests that frequently happen when working on development projects. The Amazon AWS infrastructure can be configured to automatically add servers when loads get too heavy due to increased demand and automatically remove servers when the load decreases ensuring just-in-time capacity. This is anticipated to yield additional benefits by controlling costs based on need. Rather than buying enough servers to handle estimated peak loads--capacity that would sit unused most of the time—the client can scale server capacity up and down as needed, only paying for actual server time used.

Due to the tremendous success of the Proof of Concept the client has moved forward with production implementation of this approach. They see this solution as an innovative approach to a customer facing opportunity – a goal they have been reaching for and something they would like to see more of within their organization.

Contact Info

Charter Solutions

Dee Thibodeau

763-230-6100
dee.thibodeau@chartersolutions.com

3033 Campus Drive, Suite N160
Plymouth, MN 55441