Open-source–based Big Data technology platform development for a leading Software-as-a-Service provider

About the client

The client is a leading global Cloud Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS) provider. Its core solution is an integration platform that enables applications to network and connect with a variety of on-premises and cloud services.

Technology problem

The client was facing the following technology issues:

- Rapid increase in customer base and transaction volumes was affecting the technology choices and hampering the scalability of the architecture. They were anticipating the year on year (YoY) load to increase at five times the prevailing rate and this had become an immediate challenge.
- The client was also facing high annual licensing costs and insufficient instrumentation for rich reporting.

Technology solution

Cybage provided the following technology solution:

Cybage provided the following technology solution:

- We developed a highly scalable (25 times the current load), multi-class, workload-supported data processing and analytical big data platform, capitalizing on open-source technologies.
- We provided a system that supported massive parallel processing and horizontal scaling across scale vectors.
- The system enabled continuous computation (real rime analytics) on high velocity and high volume stream data.
- It provided near-real–time, batch, and interactive reporting for data-driven insight.
- We implemented extensive tools for platform monitoring, management, and operational support.
- We ensured comprehensive regulatory compliance across varied compliance regions.
- Our technology solution enabled a growth spurt (25X) with incremental provisioning. The details are as follows:

1. Data velocity: More than 250,000 messages per second.
2. Data volume: 150 MB per second to 10 TB per day.
3. Data processing requirements: Data analytics: one minute — hourly, monthly, and historic ranges.
4. Interactive system response time: three seconds.
5. Raw data retention period: 90 days and after that, in offline data systems.

Execution strategy

The execution strategy of Cybage comprised the following:

- Conceptualization and development of a big data solution from the ground up to support the predicted scale of 25 times the current load.
- Lab experimentation to verify the proposed architecture for key NFRs (Non Functional Requirements).
- Adherence to Scrum Agile methodology for more predictable release cycle, higher solution stability, and accurate project visibility.
- Usage of development tools such as Atlassian FishEye and Crucible for better code quality and management of release processes.
- Collaboration using Atlassian Confluence, JIRA, and HipChat to support cross-functional and geographically scattered teams.

Value realized

Cybage's solution provided the following benefits:

- Ability to identify key trends and behaviors, obtain deep data traffic understanding, perform meaningful Key Performance Indicator (KPI) reporting, and recognize additional digital revenue streams.
- A synergistic architecture for future-proof Big Data management infrastructure.
- Access to richer data points and analytics.

Tools and technologies

Cybage used the following tools and technologies:

Development Cloudera Hadoop and its services, Storm, Spark, Kafka, Cassandra, Java

Testing   JMeter, Data mart, Java

DevOPS Puppet, Vagrant, Amazon Web Services (AWS) EC2 deployment

Application Lifecycle Management (ALM) Integrated Atlassian tools – Confluence, JIRA, HipChat, Stash, FishEye or Crucible, Gliffy plug-in

Monitoring and alerts   Graphite, Grafana, Nagios

Cybage services utilized

Architectural Services, Development, Testing, Platform Monitoring, Management and Operational Support, Cloud, Big Data, ALM, DevOps Capabilities

Technology