Building a Big Data center of excellence

Advice from Criteo's experience

Big Data analytics maturity comes when a company begins to run its business based on data and analytics. It is the transformation point for many organizations and shows a maturity level in the operationalization of data. Creating an analytics center of excellence (CoE) is a hallmark of corporate adoption. A CoE is typically one person who leads the analytics initiatives and provides a central source of standardized products, expertise, and best practices for a particular functional area. The role can also provide visibility into quality and performance parameters of the delivered product, service, or process. This helps to keep everyone informed and aligned with long-term business objectives.

Criteo, a performance marketing technology company, created a CoE model with the intention to help identify and spread best practices around Big Data analytics and to drive business- and customer-valued results. Criteo enables e-commerce companies to effectively engage and convert customers by using large volumes of granular data. Based in Paris, it has one of the biggest European R&D centers dedicated to performance marketing technology and an international R&D hub in Palo Alto. By selecting **HPE Vertica**, Criteo gains deep insights across tremendous data loads, enabling it to optimize the performance of its display ads delivered in real-time for each individual consumer across mobile, apps, and desktop.

Data to empower everything

Criteo logs 50 billion total events per day. Three billion banners are served every day and more than one billion unique users visit its advertisers' websites each month. Its Hadoop cluster ingests more than 25 TB a day. The system makes 15 million predictions per second out of seven data centers running more than 15,000 servers, with more than 5 petabytes under management. That's a staggering amount of data; but it's a fairly simple stack. Criteo started by using Hadoop for internal analytics, but soon found that its users were unhappy with query performance, and that direct reporting on top of Hadoop was unrealistic.

Adding HPE Vertica on top of the existing Hadoop framework helped Criteo generate real intelligence and profit from Big Data. The company has experienced double-digit growth since its inception, and HPE Vertica allows it to keep up with the ever-growing volume of data. Criteo uses HPE Vertica to distribute and order data to fine tune for specific query scenarios. Its HPE Vertica cluster is 75 TB on 50 CPU heavy nodes and growing.

Building a CoE culture

The most effective CoE has a mix of people and technical skills. It's an operational client-facing role so the right person will enjoy providing value by quickly analyzing why something is or isn't working. Look for engineers interested in seeing things work in action, and making users happy. Another good candidate is an analyst who shows more technical acumen along with people skills. Regardless of who you choose, members of the CoE have to be really good at what they do and how they interact with internal clients, because they have really broad authority.

f 𝕏 in ✉

Sign up for updates

A fairly simple stack

Tableau and ROLAP Cube for Structured Data Access — Tableau, Cube

Hive and Vertica for Data Warehousing — Hive, Vertica

Cascading, Scalding, and Hive for Data Transformation — Cascading Scalding Hive

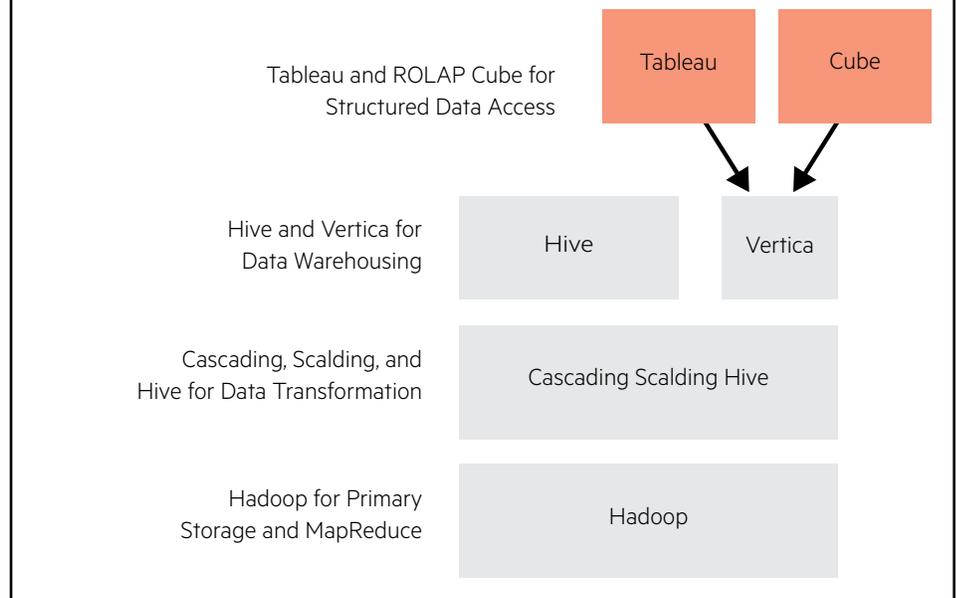Hadoop for Primary Storage and MapReduce — Hadoop

Figure 1. The performance marketing technology company's Big Data analytics stack.

Even if you don't call it a CoE, having a central team dedicated to making sure all activities around Big Data analytics follow best practices will help keep business on the right Big Data path. A mix of professionals who understand how databases work along with those who understand how people use data in their business will create a high-functioning team. The goal of that team is to respond quickly to business needs within the technical constraints of the architecture and to act deliberately and accordingly to create a tighter feedback loop on how the analytics stack performs.

Goals of a Big Data CoE:

• Defining a common set of best practices and work standards around Big Data

• Assessing (or helping others to assess) whether they are utilizing Big Data and analytics to best advantage

• Providing guidance and support to assist engineers, programmers, end users, data scientists, and other stakeholders to implement these best practices

Best practices

Without question, the most important thing is to simplify, simplify, simplify. Sole-sourcing data for HPE Vertica from Hadoop, for example, provides a defined backup process. It also allows for easy replication to multiple clusters. Vertica facilitates a simplified backup process and leaves little room for error.

Becoming mature with Big Data is not simply about using diverse data and building models that are operationalized in your organization. Big Data maturity is also about building a culture that supports analytics and is able to execute on it. It involves new organizational models, new leadership roles, and ensuring new development and deployment models.

Learn more at
hpe.com/software/bigdata

Hewlett Packard Enterprise