Big Data as a Service for Data Science

Data science is revolutionizing analytics. Highly-skilled teams explore diverse datasets to discover unexpected correlations, potential algorithms and new analytics applications. This discipline requires smart people, innovative ideas and an agile data infrastructure. However, many data scientists struggle getting the compute resources they need to do their jobs. Big Data as a Service is changing this equation.

The Data Science Resource Challenge

Many data science teams today have an infrastructure challenge. Teams that rely on other groups for infrastructure often wind up in a frustrating wait state. But, teams that manage their own data infrastructure must spend valuable time building, supporting and maintaining systems – as well as manually collecting and moving data into one place for analysis. This takes highly-skilled resources away from their primary job, delaying data discoveries and business impact.

Introducing Big Data as a Service

Big Data as a Service is an emerging category of services for data processing in the cloud. These cloud services free data scientists from delay and wasted time by providing pre-built analytic processing environments (e.g. data marts and data lakes), which are fully managed “as a service” by the vendor.

Services vary, but often include pre-configured storage, database engines, connectors for BI/analytics tools and other features, all bundled and provided for a single fee. The vendor is responsible for managing, monitoring and supporting the service, while data science teams just use it.

Not only does the cloud enable this helpful delivery model, using Big Data as a Service offers many advantages for data scientists: It’s ideal for modeling and experimentation across large, diverse datasets, especially if those datasets originate in the cloud. The cloud also offers the flexibility required for iteration, with resources available on-demand.

Cazena leads the Big Data as a Service category, offering many unique capabilities including multiple database engines, data movers, enterprise security features and support for a wide variety of tools and analytic methods. Cazena’s data movers makes it easy to move multiple datasets into a single analytic environment, then connect to analyze data using tools such as RStudio, Tableau or most others. When you’re ready to move discoveries into production, Cazena makes it easy to move analytic workloads into new cloud data marts, or back to existing on-premises data warehouses or other systems. Cazena provides a robust, flexible foundation for data science, helping teams stay focused on analytics, not infrastructure.

Cazena Benefits for Data Science

Quickly Implement and Intelligently Provision New Cloud Data Marts and Data Lakes: Cazena installs and manages everything, setting up a secure cloud environment, with encrypted tunnel for data movement and gateway for easy integration with other on-premises systems.

Handle Any Analytic Workloads, with Multiple Engines - MPP SQL, Hadoop... and whatever comes next: Analyze any data with Cazena, including structured application data or semi-structured big data. You’ll get new cloud data marts or cloud data lakes in minutes, while behind the scenes, Cazena intelligently selects and provisions the ideal processing engine and infrastructure for each workload, choosing between best-of-breed options.

Benefit from the Newest Processing Engines, Without the Implementation Time: Cazena regularly benchmarks and incorporates the latest data processing engines into the platform. Since our staff manages everything – enterprises get the benefits of new engines without the hassle of implementing and configuring them.

Collect and Move Any Data Securely into the Cloud: Easily move datasets into Cazena with built-in data movers for local, cloud or third-party sources. Cazena makes it simple to move and run models in a powerful cloud environment, improving iteration and cycle times as you experiment.

Connect to Popular Analytics Tools and Use RStudio to Manipulate Data with R: Connect to Cazena with your favorite business intelligence and analytic tools. The service supports most popular tools, so team members can continue using familiar methods. It’s also easy to access any data in Cazena with RStudio Server. For larger workloads, Cazena makes it simple to leverage R across a Hadoop-based data lake as a service. The RStudio Server console is automatically provisioned and integrated with Cazena’s data lake as a service, supporting faster analytic iteration.

Maintain Enterprise Security and Integration Standards: Cazena’s unique architecture ensures secure communication between enterprise datacenters and the cloud, and makes it easy to integrate with on-premises data warehouses or other systems. Data can also be encrypted at motion and at rest.

Focus – while Cazena Manages, Optimizes and Supports the Cloud: Cazena is delivered “as a service,” handling everything from installation to ongoing optimization of the platform to ensure that you get the best price-performance for your workloads. One fee covers everything. Focus on data science, while Cazena manages the cloud for you.
Cazena Use Cases for Data Science

Sandboxes
Create a separate sandbox environment for data exploration and discovery that’s flexible and designed for easy iteration. Cazena will provision the ideal analytic environment, based on your preferred tools and typical workloads. Use built-in data movers to move data into Cazena from local, cloud or third-party sources. Cazena supports fast iteration, leveraging cloud power and scale to run multiple models, instead of doing this locally. You’ll get results quickly, determining which models to progress and which to discard.

Cazena also helps you put data and discoveries into production faster, making it easy to move workloads from the sandbox into new cloud data marts in Cazena — or back into enterprise data warehouses or other systems.

Data Lakes
Collect raw data in a cloud-based data lake that can scale as data sources and volumes grow. Cloud data lakes are typically optimized for storage and batch analytic processing, ideal for collecting diverse datasets in one place. This provides an ideal environment for exploring, pre-processing and storing data cost-effectively in a repository that is easily accessed and queried.

Cazena’s intelligent provisioning ensures that you’ll get the ideal configuration and best price-performance for your unique requirements. And with a Cazena data lake, subsets of data can easily be moved back into on-premises data warehouses or other systems, into production data marts or into new sandboxes designed for advanced analytics.

Data Pipelines
Better manage incoming data by building data pipelines in the cloud. Collect and store raw data in a data lake, where you can explore, analyze and pre-process data. Then easily move subsets into other data marts, data science sandboxes or back to production data warehouses.

Cazena’s platform has multiple processing engines, which enables you to use the ideal technology for each stage of an enterprise data pipeline.

Now you can get the analytic functions and performance required for each step, without compromising. Cazena’s data movers make it simple to build and automate pipelines, improving data delivery and access across the enterprise.

About Cazena
Cazena’s mission is to radically simplify big data processing in the cloud, for faster business outcomes at a fraction of the cost. Founded by former Netezza leaders, Cazena leads the emerging Big Data as a Service category with end-to-end solutions for cloud data marts and cloud data lakes. Cazena is backed by Andreessen Horowitz, Formation 8 and North Bridge Venture Partners. The company is headquartered in Waltham, MA.

© 2015 Cazena. Cazena is a trademark of Cazena.