CLOUDERA

Top Three Issues Facing the Modern Data Warehouse

Grow with More Users, More Use Cases, and More Data Types

Table of Contents

The Challenges in Addressing Modern Business Analytics Requirements

Organizations' central IT departments are facing three main issues when it comes to modern data and analytics.

They need to:

1. Support multiple new users and use cases in a cost effective way. Keep data under central IT's management for reasons of traceability, security and governance

2. Transform new data into actionable insights, accelerating better decision-making through the use of data lakes

3. Support increased business needs for analytics with a flexible infrastructure and approach leveraging cloud that enables quick, secure onboarding of new users and use cases and handles massive scale

To successfully meet those needs, the modern data warehouse needs to deliver three fundamental things:

1. Multifunction analytics on a single platform

2. Quick time to value

3. Multi-vendor hybrid and multi-cloud

Without each of these three pillars, your modern data warehouse will not live up to its full potential. At best, it will create frustrations for your users and central IT; at worst, it will hamper your business's ability to compete effectively.

With all three in place, you will empower your business to thrive in the future.
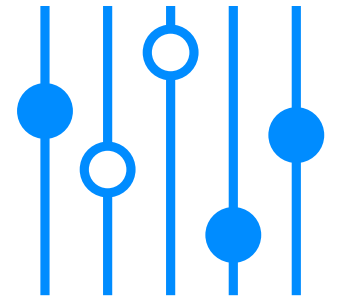
Multifunction Analytics on a Single Platform

Companies today desperately need help making better decisions. This means data professionals require deeper insights utilizing better information, delivered more quickly.

Central IT is good at ensuring their organizations' analytics have:

- **Traceability**—Analytics means little if you do not trust your source. You need to know basic things about it, including where your data originated and if it's up to date. Without that knowledge and proof, you will never be able to build trust in your conclusions.

- **Governance**—Well-governed data from a high-quality source ensures the quality, integrity and completeness of the data you are using in your analytics.

- **Security**—This means your data is and has been protected from any form of violation; that the right people have access to it at the right time, and the wrong people do not. It means you are not exposing your clients' and partners' data. According to Gartner, "By 2020, we expect that companies that are digitally trustworthy will generate 20% more online profit than those that aren't."[1]

"By 2020, we expect that companies that are digitally trustworthy will generate 20% more online profit than those that aren't."[1]

Top 10 Strategic Technology Trends for 2020, October 21, 2020
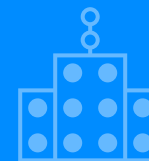
The challenge comes when business units and leaders or functions require new types of analytics for new categories of users. Data scientists, architects, and business analysts need data in different ways for different uses. Central IT is struggling to support them all, often limited by traditional data warehouses, usually on-premises, and capable of executing a limited amount of traditional analytics with known, typical and structured data.

When business units and functions bump up against the limits of the traditional data warehouses of central IT, they find other ways to get the answers they need. They may turn to a multitude of point solutions, many of which are very good, but none of which can handle all the business and users' needs. As a result, many different point solutions, or instances using point solutions, are engaged, resulting in multiple silos. This creates holes in security because you may be putting secure data into the point solution without guaranteeing that each instance or point solution reflects the security model central IT has worked so diligently to create. Certainly not. These silos open the business up to traceability gaps and breaks in governance because you can't easily trace the data back to its source. "Transparency and traceability are critical elements to support digital ethics and privacy needs, Gartner asserts."[2]

The downfall of using point solutions as a short-term fix is twofold:

- These point solutions can't enable different users to use the same data for different types of analytics in a single instance, and

- The impossibility of adequately addressing the increasing scale of complex organizations' analytics needs

The only answer that can effectively address a large organizations' need to support a myriad of users with a variety of use cases at massive scale, while maintaining traceability, governance and security, is multi-function analytics on a single platform.

The only answer that can effectively address a large organizations' need to support a myriad of users with a variety of use cases at massive scale, while maintaining traceability, governance and security, is multi-function analytics on a single platform.

Get Quick Time to Value with a New Approach and Modern Architecture

One of the biggest issues with IT and analytics today is the time it takes from when new data arrives to when it becomes available for actionable insights. The longer it takes, the lower the value of that insight.

Knowing what stock is trading at now is more valuable to a trader or investor that what it was trading at two weeks ago. A retailer gets more value from knowing a product has sold out today than a month ago. Obviously, the faster we can transform new data into business insight, the more value we have provided to the business. We keep this in mind when we build dashboards and design data warehouses today.

The faster we can transform new data into business insight, the more value we have provided to the business.
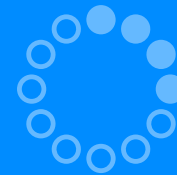
It breaks down when a problem arises. Sales of a certain product is dropping off. Inventory is piling up. Manufacturing is delayed. We don't know why, and we need to explore it and figure out the reasons why. This involves unknown questions on unknown data. It may require us to get new data, raw data that is more granular, data from somewhere else that may shed light on the problem. We need to explore new things, and dig deeper. We need to bring new queries online quickly, and get responses quickly so we can ask deeper questions and determine which threads to pull on. We need fast response from our systems.

Or maybe we need to bring new data to light, say from new censors we've added to our manufacturing equipment. We need to add new IoT data on our dashboards. We need better ad hoc reporting, better forensic reporting, fast response, new types of data, and we need it as quickly as possible. But we don't have months or years to wait.

The truth is, we need to rethink the entire data warehouse infrastructure.

We need a new architecture that allows us to land all forms of data regardless of its structure: machine log data. Social media. Documents. Emails. Pictures. Video. IoT. We need to take it all in quickly, land it in a data lake, then transform it. Extract, load, transform as opposed to extract, transform, load. All data at scale, ingested immediately with next to zero latency.

Known data doesn't go away. It is an island of well-managed data in the lake, surrounded by all the raw data we don't need for day-to-day reports and dashboards.

This approach dramatically reduces time to value. The second pillar is complete.

The truth is, we need to rethink the entire data warehouse infrastructure. We need a new architecture that allows us to land all forms of data regardless of its structure: machine log data.
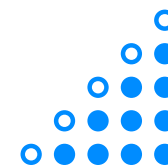
Scale with Multi-Vendor Hybrid and Multi-Cloud

The final piece of the puzzle is a very flexible infrastructure that goes beyond the traditional appliance and on-premises model, allowing us to onboard users and use cases on a massive scale. This, of course, is hybrid and multi-cloud.

Imagine enterprises with central IT serving dozens to hundreds of business units, with dozens to hundreds of use cases, all working in teams supporting their business unit. Now imagine that these enterprises have highly specialized security and governance, such as companies in financial services, healthcare, government, sophisticated retail, automotive, energy or manufacturing companies. It would be difficult to "lift and shift" data from these types of organizations to any one cloud.

As a result, these organizations may feel locked into an on-pre world. In this scenario, if someone needs something outside of your typical dashboards, you have a choice. The first is to turn on a high-priority VIP workload, which could slow everything else down and result in missing SLAs. The second is for central IT to say no.

26%
and growing

Global executives confirmed through survey have indicated that they've moved to hybrid cloud.

Central IT saying no to the business leads us to shadow IT.

We should not confuse shadow IT with business unit IT. In many situations, there are great reasons for business units to own, manage, and curate their own data separate from central IT. Basically, when there is no central value to the analytics, no cross-purpose, and it meets needs unique to the business unit, business unit IT makes sense.

Shadow IT involves data that should be centralized and centrally managed, largely for security reasons. However, when IT has said "No", the business has taken a copy of the data and put it into their own solution. They have broken the rules, and they have exposed the business to significant risk because, "when business units can get whatever they need without the knowledge or approval of the IT team, it's hard to maintain an effective cybersecurity program across the organization."[3]

The rise of shadow IT does show that employees have a need to find solutions that central IT isn't supplying them. In fact, a recent study shows that 77 percent of IT employees believe that their organizations could have a competitive edge "if company leaders were more collaborative about finding solutions to shadow IT needs from both IT and non-IT employees."[4]

A hybrid and multi-cloud environment enables us to say "yes" to new users and use cases, enabling us to quickly and easily leverage the public cloud. Organizations can seamlessly replicate data, run it, ingest it, and add security and governance. Users can do what they want right away in a safe and secure environment–and maybe even billing business departments for it. Hybrid cloud lets you grow beyond on-premises immediately, in a way that is consistent with central IT's well-managed data.

Another key to this capability is a multi-vendor, multi-cloud approach, that enables you to work with one or more clouds and move data cloud to cloud–whichever one has the best economic model and works for your purpose, providing the ultimate in flexibility. This enables you to avoid vendor lock-in and retain control over costs. According to Capgemini's chief cloud officer Charlie Li, "organizations are recognizing the benefits of multi-cloud, such as improved organizational flexibility, efficiency, and performance, as well as avoidance of vendor lock-in."[5]

Microsoft, AWS, Google... It doesn't matter, and it all works. Cloudera is the Switzerland of clouds. We want you to be able to look at data anywhere and anyhow. We make sure you can move your data in full or in part to any cloud or vendor that makes sense.

77%

of IT employees believe that their organizations could have a competitive edge "if company leaders were more collaborative about finding solutions to shadow IT needs from both IT and non-IT employees."

Learn More

For more information about Cloudera Data Warehouse, please go to
www.cloudera.com/products/data-warehouse.html

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com | US: +1 888 789 1488 | Outside the US: +1 650 362 0488

Sources

[1] Top 10 Strategic Technology Trends for 2020, October 21, 2020

[2] Top 10 Strategic Technology Trends for 2020, October 21, 2020

[3] Shadow IT: You Can't Protect What You Can't See, August 14, 2019

[4] When properly managed, shadow IT can benefit your organization, October 16, 2019

[5] 5 enterprise cloud strategy trends for 2020, December 10, 2019

Cloudera, Inc. 5470 Great America Pkwy, Santa Clara, CA 95054 USA cloudera.com

Privacy Policy | Terms of Service

CLOUDERA