

eBook

Data and Analytics Integration Hubs

Unifying Data and Analytics Tools

By Wayne Eckerson

June 2020

eBook reprinted by



Get more value from your data

This publication may not be reproduced or distributed without prior permission from Eckerson Group.

About Eckerson Group

[Eckerson Group](#) helps organizations get more value from their data through research, consulting, and education. Our experts each have more than 25+ years of experience in the field, specializing in business intelligence, data architecture, data governance, analytics, and data management. We provide organizations with expert guidance during every step of their data and analytics journey. Get more value from your data. Put an expert on your side. [Learn what Eckerson Group can do for you!](#)



Table of Contents

- Introduction** 3
- Chapter 1: The Modern Way to Manage Disparate BI and Data Environments**..... 4
 - Core Components 4
 - Benefits and Challenges 6
 - Getting Started 7
 - Summary 7
- Chapter 2: Unifying Disparate Data Sources** 8
 - Unifying Data and Workflows..... 9
 - Summary 10
- Chapter 3: Unifying Analytics Across Disparate BI Tools** 11
 - Universal BI Semantic Layer 11
 - Versatility and Pricing 12
 - Summary 13
- Conclusion** 13
- About Eckerson Group** 14
- About DataClarity**..... 15

Introduction

“It was the best of times, it was the worst of times...” That opening phrase from the Charles Dickens novel A Tale of Two Cities aptly describes the state of data in most organizations where pockets of governed data are surrounded by a sea of data silos and inconsistent reports. Data reflects the organization in which it resides. Unfortunately, most organizations are fragmented into numerous fiefdoms, each with their own data systems, analytical tools, and processes.

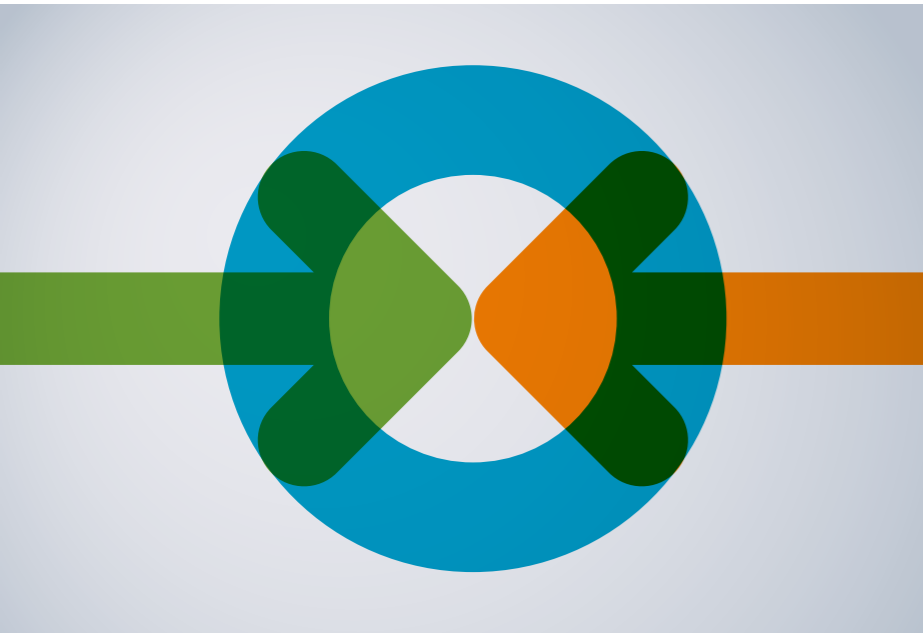
Unifying this unholy mess requires a large dose of integration. This can be done manually by expensive systems integrators or in a more automated fashion by humans armed with integration middleware. In the analytics space, we call this middleware a “data and analytics integration hub.” Rather than consolidating data, these hubs virtually connect data and models in disparate systems and business intelligence (BI) tools to create a unified view of data for business consumption.

This eBook introduces data and analytics integration hubs and shows how they are critical components in a self-service architecture. The eBook explores how this data middleware integrates data, models, and reports that are locked in various data and BI silos, making it easier for business users to gain a comprehensive view of information within their organization and beyond. The eBook has three chapters:

- **Chapter One: The Modern Way to Manage Disparate BI and Data Environments.** A data and analytics integration hub is an emerging technology that provides all-in-one functionality while unifying disparate BI and data environments.
- **Chapter Two: Unifying Disparate Data Sources via Data Virtualization.** Data and analytics integration hubs provide an all-in-one integrated environment for data analysts to discover, prepare, and analyze data.

- **Chapter Three: Unifying Analytics Across Disparate BI Tools.** Companies need to invest heavily in teams and people, both at the corporate level and in the business, if they want to become a data-driven organization.

There are several vendors that offer data and analytics integration hubs, and one is the sponsor of this eBook. If you need to integrate disparate data and BI silos, we highly recommend that you carefully evaluate various products in the space to determine which is best suited to your needs.



Chapter 1: The Modern Way to Manage Disparate BI and Data Environments

A data and analytics integration hub is an emerging technology that provides all-in-one functionality while unifying disparate BI and data environments.

Most organizations are riddled with data silos and saddled with a dizzying array of analytical tools. As a data analytics manager, your natural inclination is to consolidate: that is, consolidate data silos into an enterprise repository where data can be easily modeled, integrated, and governed; and standardize on a single business intelligence (BI) tool to streamline support and training and reduce costs.

If your only strategy is to consolidate, you'll suffer the same fate as the Greek king Sisyphus who was doomed by Zeus to perpetually push a boulder up a hill. To get the data analytics gods on your side, you need a more nuanced approach that accommodates political reality. Most organizations are not disciplined enough or centralized enough to succeed with consolidation strategies. Data and tools proliferate as quickly as you can consolidate them.

So, while it's important to have a data warehouse and BI tool standards, it's also

important to be able to connect to data sources and BI tools dynamically. This gives organizations the ability to adapt to continually changing data and analytics environments. Products that stitch together existing BI and data resources into a seamless whole are called *data and analytics integration hubs*.

Core Components

Data and analytics hubs give business users a single view of enterprise data assets as well as a single place they can go to access data, no matter where that data is located, what system it runs on, or what methods (i.e., BI tools) are required to access them. Data and analytics integration hubs are middleware that stitches together disparate data and analytics environments.

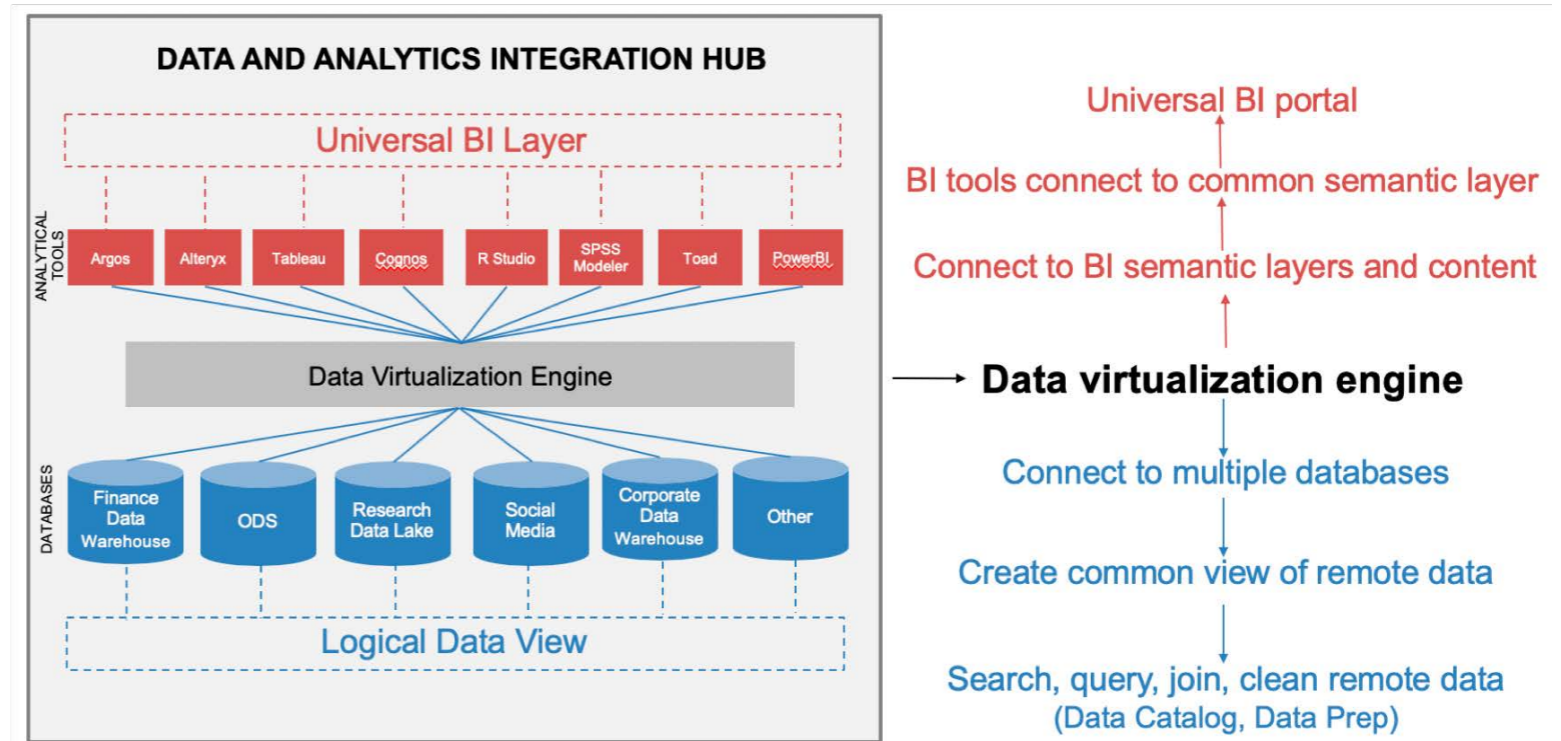
At the heart of a data and analytics integration hub is a data virtualization engine that gives business users real-time access to various resources without having to know where

data is located, what system it runs on, or what methods (i.e., BI tools) are required to access them. The engines “virtualize” all back-end data sources, including BI models, making them seem as if they are part of single, integrated system that resides locally on their desktop or server. They use bidirectional connectors to federate queries and intelligent caches to optimize performance.

Data virtualization engines are not new. But few of them connect to BI tools, such as those from MicroStrategy, IBM (Cognos), SAP (BusinessObjects), and Tableau. Nor do many provide business user functionality to search, prepare, analyze, visualize and report on data. A data and analytics integration hub not only connects users to diverse data and BI sources, it also enables them to join, format, clean, and transform data as well as develop reports, dashboards, visualizations, and predictive models. (See figure 1.)

Thus, a data and analytics integration hub not only creates a [data fabric](#) that connects your BI tools and data silos, it’s also an all-in-one data and analytics workbench that facilitates self-service. It contains everything business users need to find, manipulate, analyze, and visualize data. It also makes life easier for architects and

Figure 1. Data and Analytics Integration Hub



administrators to govern user access to data and analytical resources. Through granular security and single sign-on, administrators can provide a secure data analytics sandbox tailored to every business user.

Benefits and Challenges

Benefits. A data and analytics integration hub offers several benefits. First, it integrates the data and analytics resources you have, avoiding the need to consolidate or replace them. This saves the time and cost required to migrate existing assets to a new platform, repoint applications, and inform users of the change. Second, it gives users one place to go to get any information they need. They can use the analytics tool with which they're familiar to access logic and models built into another BI tool or data sets in remote systems. (See figure 2.)

Third, it gives new life to incumbent BI tools that are used as glorified data extraction tools to download data from a data warehouse or other source into a desktop spreadsheet or database. Fourth, it provides a common or logical view of distributed data and BI resources, making them easier to find and query. It also provides a consolidated view of all reports no matter which tool created them

Figure 2. Benefits of a Data and Analytics Integration Hub



Data and analytics integration hubs virtually integrate data and models stored in disparate databases and BI tools, saving organizations the time and expense of consolidating data.

in a single portal view, that can be tailored to each individual. And fifth, it supports the complete data analyst workflow required to facilitate self-service analytics.

Getting Started

When working with a data and analytics integration hub, most organizations start small. They use it to support a tactical use, such as the need to link two BI tools together. For instance, a company wants to give Tableau users access to data modeled in an incumbent reporting tool, such as IBM Cognos Analytics, without having to recreate that logic within Tableau.

Once organizations see the integration power of a data and analytics integration hub, they expand the footprint, adding other BI tools and data sources until they've created a data fabric that weaves together a majority of the company's data and analytics assets. They might then discover its business user functionality—data catalog, data preparation, data visualization, data distribution—that makes it easier to implement self-service

analytics. Given its broad platform and all-in-one capabilities, some companies decide to standardize on a data and analytics integration hub, reducing costs and duplication.

Summary

Data and analytics integration hubs virtually integrate data and models stored in disparate databases and BI tools, saving organizations the time and expense of consolidating data. By creating views on top of virtual data, the hubs speed time to market for analytics solutions while reducing costs, delighting business customers and chief financial officers, respectively.

Now that you have a broad understanding of data and analytics integration hubs, let's drill down into how these products unify data contained in disparate systems through an enterprise.

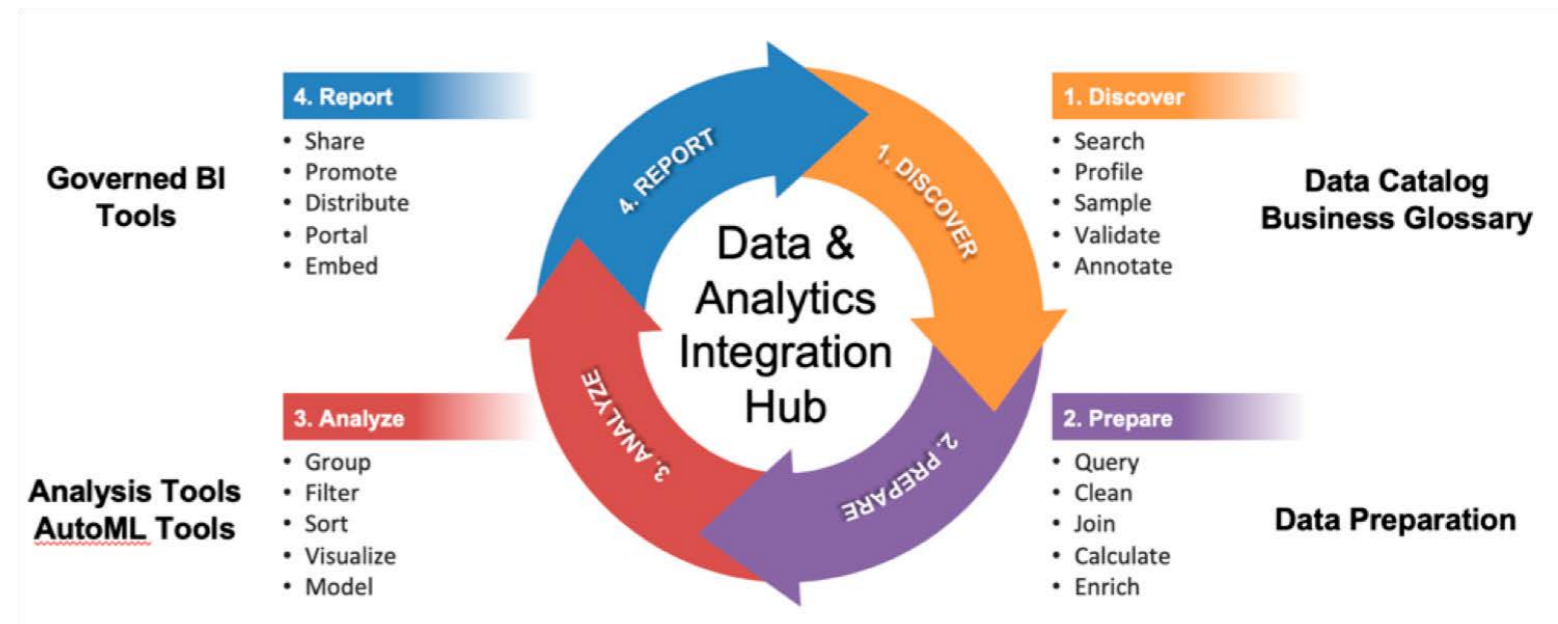
Chapter 2: Unifying Disparate Data Sources

Data and analytics integration hubs provide an all-in-one, integrated environment for data analysts to discover, prepare, and analyze data.

Data analysts and scientists are voracious consumers of data. They follow a highly iterative workflow to discover, prepare, and analyze data and share their insights with others. (See figure 1.) In most organizations, data analysts are left to their own devices to navigate this workflow. Without modern self-service tools and technology, they spend too much time finding and manipulating data rather than analyzing and interpreting it for others.

Discover. When given an assignment, the first thing a data analyst does is scour the organizations for relevant data assets to answer a business question. Some of those assets are centralized in a data warehouse or data lake, but many are not. For example, an analyst may want to clickstream data from a website, a social media platform, a marketing cloud application, or data from a human resources system. Hunting down relevant sources can take days or weeks. Then, for

Figure 1. Data Analyst Workflow



each source, the analyst may need to gain permission from the owner to access the data and might also be required to use a specific tool.

With data in hand, the real work begins—or not. First, the data analyst needs to examine the data to see if it meets their needs. How many

records? How many columns? What is the range of values in those columns? How many nulls? How are the metrics calculated? Where did the data come from? Who has used this source before and for what purpose? Tribal knowledge goes a long way to helping data analysts discover relevant data sources.

Data and analytics integration hubs provide an all-in-one, integrated environment for data analysts to discover, prepare, and analyze data.

Prepare. Once the analysts identify the proper data sets, then they must prepare them for analysis. This involves formatting, cleaning, and combining data sets to create a single table, either physical or virtual. From there, they apply logic to enrich the data set. They might create a new field by summing the values in two columns or create custom groups and benchmark their performance over time. They might also enrich the data set with local data (i.e., spreadsheets) or external data.

Analyze. The next step is to analyze the data using various mathematical, statistical, and machine learning functions as well as visualization capabilities that make it easy to spot the trends, anomalies, correlations within and between data sets. Visualization tools enable analysts to drill, sort, filter, and pivot data to see patterns from various perspectives and create predictive models used to score incoming data.

Report. In the final step, data analysts codify their insights in interactive reports, dashboards, or narrative “stories” designed for business consumption. If the report is deemed critical, they might pass it through a governance review process to get it “certified” so it can

be distributed to business users on a regular schedule or embedded in a corporate intranet or portal.

Today, there are self-service tools, such as data catalogs, data preparation tools, and automated machine learning (AutoML) tools that analysts can use to facilitate these tasks. However, these tools often come from different vendors and are not tightly integrated. Completing the workflow with a disjointed set of tools interrupts the analyst’s natural flow of work.

Unifying Data and Workflows

Vendors have recognized the limitations of existing toolsets to support data analyst workflows. Some now ship data and analytics integration hubs that help take the pain out of delivering business insights to business users. These all-in-one platforms contain all the functionality data analysts need to discover, prepare, analyze, and report on data.

Data Fabric. More importantly, data and analytics integration hubs create a “data fabric” that stitches together disparate data sources into a unified whole. A company might have 20 or 200 distributed data sources running on

different systems with different databases; a data and analytics integration hub makes those sources appear like they are running on a single, local data server. This capability dramatically greatly simplifies data access, making it easy for data analysts to discover, prepare, and analyze data.

Data Virtualization. The data fabric is made possible by a data virtualization engine. The engine gives data analysts real-time access to distributed data sources without them having to know the location of those sources or how to access them. The essential capabilities of data virtualization consist of data abstraction, zero data replication, real-time access, self-service data services, and security and metadata governance.

- **Data abstraction.** A data virtualization engine gives data analysts access to distributed data sources without them having to know the location of those sources or how to access them. It hides the data access complexity by virtualizing the disparate data resources and making them seem like they are part of a single repository.
- **Zero data replication.** A data virtualization engine does not move the data, but leaves

it at its source. It can cache only the query results that are frequently used by users. Advanced caching capabilities allow synchronizing the updates on the data sources with the cache results.

- **Real-time data access.** A data virtualization engine gives data analysts real-time access to any data. The engine dynamically queries the distributed sources, pushing native SQL to each source, and joining the results on the fly. It supports data transformation through calculations and data cleansing functions without the latency of a traditional data movement approach.
- **Self-service data services.** A data virtualization engine provides a universal semantic data layer that facilitates self-service data discovery for analysis for end-users. A standard data protocol allows data analysts to query and visualize the data in any of their preferred tools.
- **Metadata and security governance.** A data virtualization engine provides a single entry point for data access, security, and management. Additionally, it provides access to data and metadata catalogs to help users search and understand the data they access.

Summary

Data and analytics integration hubs provide an all-in-one, integrated environment for data analysts to discover, prepare, and analyze data. At the heart of a hub is a data virtualization engine that provides a supporting data fabric that stitches together distributed data into a logical data view accelerates the data analyst workflow. Next, we will examine how these hubs integrate diverse sets of BI tools to create a universal semantic layer and report catalog.



Chapter 3: Unifying Analytics Across Disparate BI Tools

If you've been frustrated by the lack of integration among BI tools, then you should consider implementing a data and analytics integration hub.

Universal BI Semantic Layer

Most companies are saddled with several BI and analytics tools. Although they may have an enterprise standard, not all departments and business units use it. Without enterprise support, these departments build business logic into their reports, creating data fiefdoms that, in effect, hoard data, insights, and knowledge. Unlocking these BI fiefdoms is one of the major advantages of a data and analytics integration hub.

The hub connects to every data and BI source via special connectors and a data virtualization engine. On the BI side, for example, these hubs allow Tableau users to query IBM Cognos Analytics Framework Manager models. The Tableau users see the Cognos models as

part of their workbook schema and can manipulate that data, just as if it was modeled and managed in Tableau directly. A security framework with granular permissions and single sign on provide business users with seamless access to distributed resources.

Besides connecting two BI tools, a data and analytics integration hub enables BI administrators to combine the models of all BI tools into a universal semantic layer. This way BI users can access objects from any of the company's BI tools, no matter which BI tool they're currently using. The universal semantic layer makes it easy for them to create dashboards that contain side-by-side visualizations from multiple BI tools. (See figure 4.)

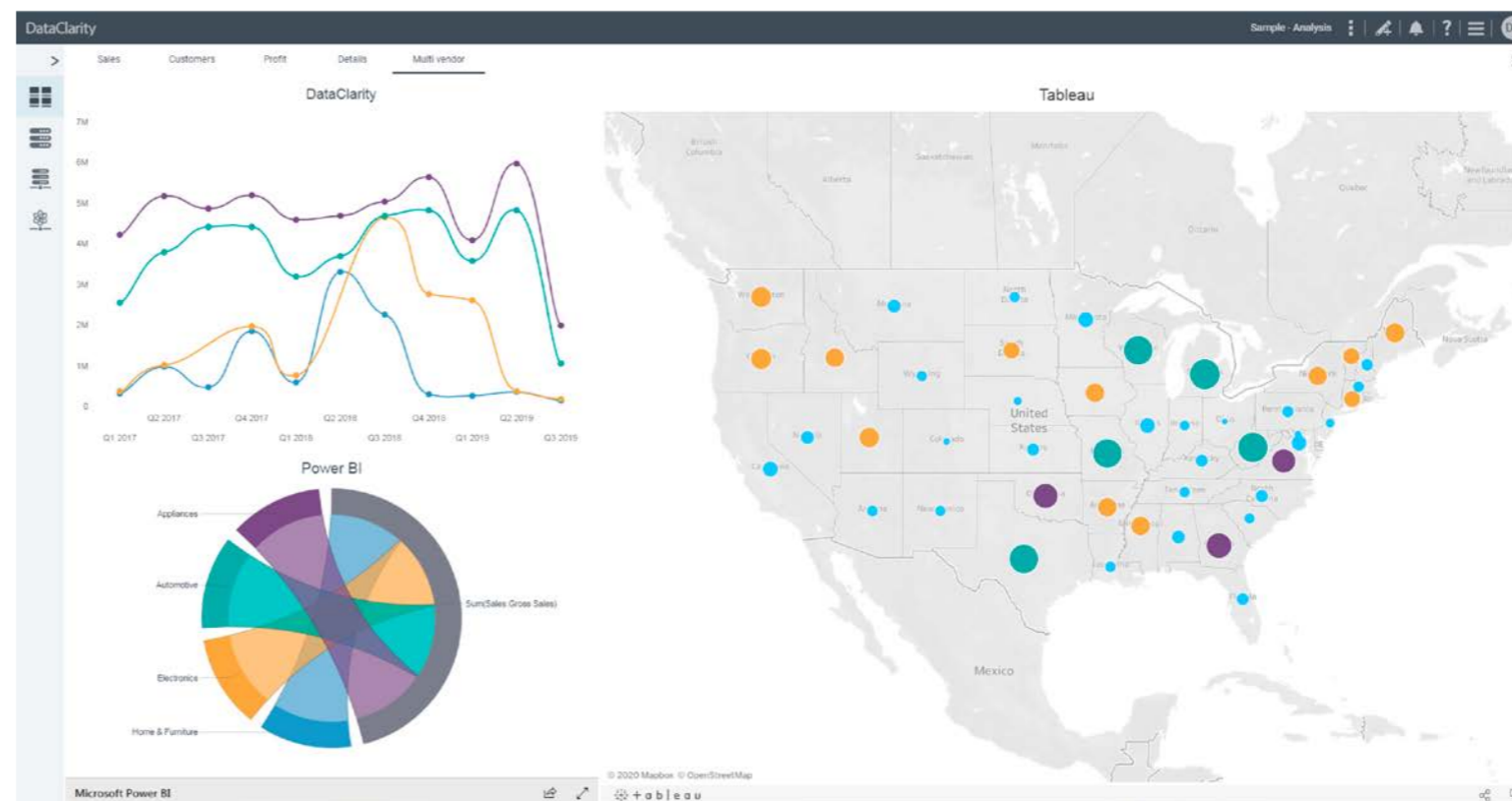
BI Portal. The hub also lets administrators create a universal BI portal that combines the reports of all BI tools into a single catalog that is personalized based on a user's role, department, and preferences. The portal displays thumbnails of reports and dashboards along with descriptive metadata. When users click on the thumbnail, it opens the report in the hub so users can view and interact with it.

BI and Analytics Functionality. More than just an integration engine, a data and analytics hub is a full-fledged BI and analytics tool in its own right. Although it can provide users access to Tableau, MicroStrategy, Power BI, and other BI tools, it offers built-in functions for dashboard creation, visual analysis and exploration, spatial analytics, and data science (i.e., R and Python coding and execution). Many companies eventually migrate to using the embedded BI and data science functionality of data and analytics integration hubs because the features are integrated with its data capabilities, providing a seamless workflow for data analysts.

Versatility and Pricing

Versatility. The versatility of a data and analytics integration hub gives customers

Figure 4. Visualizations from Multiple BI Tools on One Page



options. They can purchase the platform as a BI tool for business users and then layer in its data catalog and preparation features for power users. Alternatively, they can buy the hub for its data connectivity and preparation features, and then expose power users to its visual analysis, geospatial, and data science functions

to support deep analysis of their data sets. Any combination is possible.

Pricing. The best data and analytics integration hubs offer simple user-based pricing that doesn't punish organizations for the number of BI tools or data sources they have, or number of queries they execute.

They also offer affordable pricing, making it a no-brainer for organizations to expand usage of the platform as new needs arise.

Summary

Many organizations have made sizable investments in BI tools that they don't want to abandon. As valuable as the tools themselves are, the dozens or hundreds of custom metrics, models, and reports that the organization has built using its legacy BI tools are also significant investments. Data and analytics integration hubs enable organizations to leverage these legacy investments as they migrate towards next-generation BI tools.

Conclusion

The trend in data analytics today is towards convergence. Data tools are converging into data platforms and BI and analytics tools are converging into analytics platforms. Data and analytics integration hubs integrate these two types of platforms into a seamless whole, providing end-to-end support for business user workflows. Organizations that want one-stop-shopping for all their business user data and analytics needs should consider a data and analytics integration hub.



About Eckerson Group



Wayne Eckerson, a globally-known author, speaker, and consultant, formed [Eckerson Group](#) to help organizations get more value from data and analytics. His goal is to provide organizations with expert guidance every step of their data journey.

Today, Eckerson Group helps organizations in three ways:

- **Our thought leaders** publish practical, compelling content that keeps data analytics abreast of the latest trends, techniques, and tools in the field.
- **Our consultants** listen carefully to craft tailored solutions that translate your business requirements into compelling strategies and solutions.
- **Our educators** share best practices in consulting workshops and seminars tailored to each organization's needs.

Eckerson Group is a global research and consulting firm that focuses solely on data analytics. Our experts each have more than 25+ years of experience in the field and specialize in data governance, self-service analytics, data architecture, data science, data management, and business intelligence.

Our clients say we are hard-working, insightful, and humble. It all stems from our love of data and our desire to help you get more value from your data. We see ourselves as a family of continuous learners, interpreting the world of data and analytics for you.

Get more value from your data. Put an expert on your side. [Learn what Eckerson Group can do for you!](#)





About DataClarity

DataClarity helps people easily access, prepare, consume, and visualize any data from anywhere, and transform it into actionable information by fully-integrating self-service data virtualization, data science, and data visualization together in a modern digital experience. Purpose-built as both end-to-end enterprise BI platform and integration hub with analytics fabric architecture spanning on-premises and multiple cloud

environments, DataClarity unifies hundreds of unique data sources, models, visualizations, security, and web content across industry-leading BI and AI/ML tools, platforms, and applications. To learn how DataClarity can leverage existing technology investments and multiply your analytics value, please visit www.DataClarityAnalytics.com and www.DataClarityData.com.