Building blocks of a modern analytics platform

This whitepaper will explore the building blocks that comprise a modern analytics platform the business and IT can use—together—to deliver data, value, and decisions for the whole company. This includes the new, shiny tools of today as well as the traditional tools that have been the foundation for business intelligence for decades. We will show how each building block fits into the larger process of turning data into insight, whether it's the tools you need to capture and report on data, or new technologies to share interactive insights. We'll also discuss how Tableau can be both a foundation for a modern analytics platform and a catalyst for creating a new data-driven analytics culture.

Table of contents

What is "modern analytics," and why do we need it?

We live in an exciting time of accelerated innovation, increased global competition, and unprecedented opportunity to disrupt and reinvent. The exponential growth of digital technology, ubiquitous internet connectivity, and proliferation of smart devices—all of which create a deluge of data that holds a competitive advantage for anyone that can quickly and accurately understand it—is enabling a fourth industrial revolution.

The potential of modern technology cannot be understated. The concrete line between the physical and virtual world is being erased, reducing the barriers to market entry and creating entirely new business models. Every industry is ripe for disruption, even those that were the disruptors a decade ago. Consider the impact of crowd funding on small business lending, online retail in the broader retail market, or subscription media services on traditional cable television. Everywhere around you, this transformation is accelerating: Automated processes, messenger bots, and artificial intelligence are just a handful of new technologies paving the way towards an exciting, if uncertain, future.

To learn more about the changing role of BI's favorite catch-all term, consider **"Define Analytics,"** our whitepaper that highlights the key terms surrounding analytics today.

But there is one constant thread woven among all these trends: massive amounts of data.

If data is a raw material, then analytics is the process of refining it into useful information, and ultimately giving your business a competitive advantage. Data is more vital today than ever before. As businesses evolve to keep pace with shifting industries, they must rely on increasing amounts of accurate and timely data to make fast, smart decisions. But analytics technology has traditionally been slow and cumbersome in adapting to the growth and changing forms of data. It seems like there's a new type of database every week, data generated from a new suite of devices, and it's all built using entirely new technology. Analytics simply hasn't kept up.

To learn more about how IT and the business working together is critical in enabling the modern approach to enterprise analytics, please consider our whitepaper: **How to Build a Culture of Analytics.**

To create a data-driven culture for today's digital era, one that's ready to tackle today's business challenges with accuracy and speed, organizations must not only invest in new technology, but also new ways of delivering that information, including the people required to lead the charge. This parallel cultural shift is a fundamental change in the relationship between IT and the business; they are two partners working to collect and mine data, but also to refine it and deliver the right information on demand. It's when IT and the business work together that organizations can turn the pipe dream of a self-service analytics culture into a reality.

> When IT organizations lead the transformation to self-service analytics, they can ensure governance and security at scale. And by empowering the business to be data-driven and agile, IT becomes a trusted partner to the business.
>
> - COLIN REES, CIO, DOMINO'S PIZZA GROUP LIMITED

1. The three data challenges of today

Data is everywhere and generated abundantly, by the second. For example, a toothbrush is now a smart device, capable of logging when you brush, for how long, and the status of its internal parts. It can also send all this information to your dentist. A simple task is now thousands of data points. Multiply this one example by millions of devices, and you now have a single, small industry creating mountains of data that didn't exist just a few years ago. Add in event logging, APIs, social media, website tracking, and a host of other Internet technologies, and data explodes every time you turn your head.

This modern ecosystem presents three business challenges, which we'll look at now.

The first challenge: Data is everywhere

Historically, most organizations kept data on-premises. They worked hard to control all the data that was both created and then stored in a seemingly well-defined data warehouse. If there was data you didn't capture, it was insignificant and probably not worth the effort.

That mentality can end a business in today's era, where websites, mobile devices, and cloud applications all generate data outside of an organization, consider Google Analytics, Splunk, ServiceNow, and Salesforce, to name just a few. This trend is only going to accelerate, with more and more useful data being generated in the cloud, by third-party providers. To complement this cloud bias, some organizations are moving their own on-premises infrastructure to the cloud as well.

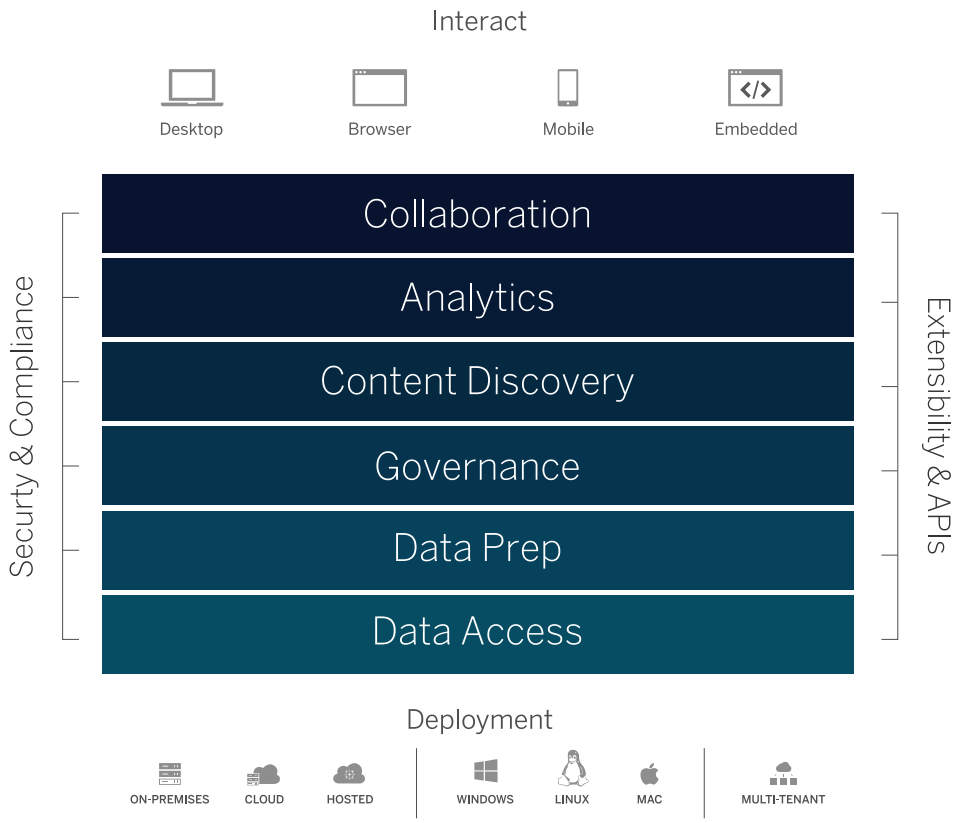

Figure 1 Tableau Connects to Data Everywhere

The second challenge: Everyone needs data

Within the last two decades, we've seen a market shift towards digital business and the cloud. This is one key in the modern analytics revolution. The other is the shift to enabling a data-driven culture with self-service business intelligence. A culture of analytics permeates throughout today's most innovative companies, where the business user with the question is also the one who can discover the answer on his or her own. This turns organizations into masters of refining data—digital gold—into information at rapid speed. To fully create a culture of analytics, an organization must bring together its people and its data, arguably an organizations greatest assets, giving everyone access to the right data and encouraging them to explore and collaborate.

With a modern approach to analytics, IT and the business work together. IT provides a centralized environment where business users can find trusted data and content, and enables everyone to securely use it, ask questions, experiment, and make decisions at the speed of thought. This is a bottom-up methodology comprised of subject matter experts creating metadata, business rules, and reporting models that provide fluid agility and expedite continuous improvement.

> When we first started with Tableau, we were just thinking about dashboarding and reporting. We never thought Tableau would fundamentally change the DNA of the organization. It's not just about a solution or a technology, it's about how the culture towards data has changed.
>
> - ASHISH BRAGANZA, DIRECTOR OF GLOBAL BUSINESS INTELLIGENCE, LENOVO

The third challenge: Data is always changing

As we all know, the only true "constant" is "change." A modern analytics platform prioritizes flexibility; the ability to move data across platforms, adjust infrastructure on demand, take advantage of new data types, and enable new use cases. Additionally, it seems that every day a new technology for analyzing data is released to the world— technologies like machine learning, voice assistants, and natural language queries. Some of these may seem more fiction than practical, at least for now, but new methods and techniques are bound to mature and prove their worth to your customers and your company.

In a world of rapidly evolving data, flexibility is paramount, both for your expanding infrastructure needs and for new technologies. Flexibility is critical to creating and maintaining your distinct competitive advantage. When you are considering a future-friendly analytics architecture, avoid vendor lock-in to proprietary architecture as that can significantly hinder your ability to be agile in the future.

2. The building blocks of a modern analytics platform

Those three challenges facing businesses today are not as insurmountable as they may seem. If data is the common thread in today's evolving business world, then a modern analytics platform is the key to unlocking its potential. But a modern analytics platform is not one singular piece of infrastructure: It comprises multiple, independent building blocks. Some of them are traditional staples of business intelligence, just brought into the modern era (e.g. data warehouses). Others, are entirely new concepts that have revolutionized the way businesses approach data analysis in the first place (e.g. visual analytics). Together, both types of building blocks form an analytics platform that can help any organization face the challenges of today's businesses.

A modern analytics platform can be boiled down into two distinct halves:

- **IT Enabled**, which includes the gathering, curation, and preparation of data
- **Author & Consume**, which includes analyzing data and communicating insights to the right stakeholders.

Traditionally, both halves were combined into a singular process in the domain of IT and IT only. We now see the first half—the creation and processing of data sources—as IT enabled. The second half—analysis and delivery—are still empowered by IT, but run by business users themselves.

This combination of the two is a true partnership between business and IT, and the modern way to run any organization that desires to make data-driven decisions quickly. It's sometimes referred to as bimodal BI, keeping the best of traditional BI and operational reporting, while adopting the self-service mechanism of modern analytics.

In this relationship between IT and business, IT designs the data architecture and facilitates proper data security and access controls. Business subject matter experts then create the analytical assets they need, when they need them. The result is IT enabling everyone to answer their own critical questions efficiently, and business users that can answer a question the moment it's asked, creating an agile organization ready to attack the challenges of the modern business landscape.

We will take a high-level look at the various building blocks that comprise each side of a modern analytics platform, some of the top trends within each, and important concepts to keep in mind. For a deep dive into specific components—including market-leading vendors who provide them—and to see whether they apply to you, please see each building block's corresponding section in the Appendix.

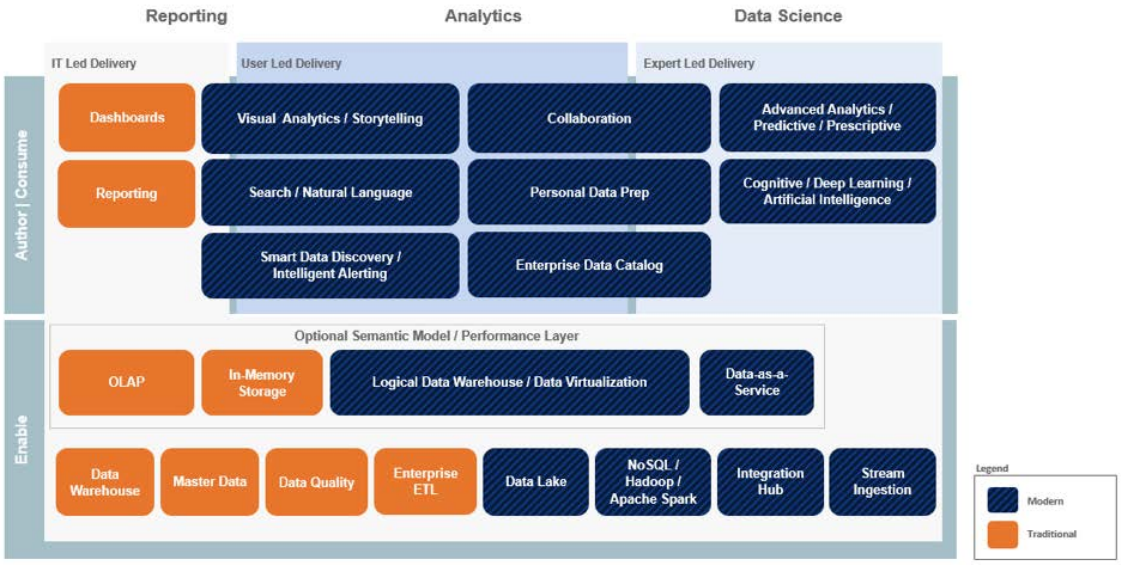

Figure 2 Basic Building Blocks

IT enabled

Unlike traditional IT-led business intelligence, today's most effective IT organizations focus on enabling analytical data sources by orchestrating, organizing, and unifying data for users and experts to author and consume. This role is obvious, but still cannot be overstated. Collecting data, managing its sources, and processing it so that it can be used by others, has always been crucial to business intelligence, and remains at the heart of a modern analytics platform. What insights are there to discover, if there is no raw material to refine from?

The difference in a modern analytics platform is the partnership between business and IT. When business users are given tools to analyze data on their own, they are free to answer questions on the fly, knowing they can trust the data itself. This leads to accurate, agile reports and dashboards. And IT, free from dashboard and change requests, can finally prioritize the data itself: safeguarding data governance and security, ensuring data accuracy, and establishing the most efficient pipelines for collecting, processing, and storing data.

Prioritizing data couldn't come at a better time. Your business, no matter its size, is already collecting data and most likely analyzing a small portion of it—the rest is dark data. There are a billion places to gather data, and more tools are coming to market to help you collect as much of it as possible. Today, you'll find an assortment of technologies being used to handle various characteristics, such as high volume, data location, and a variety of data source types. Every organization is truly unique, and you should take the time to prioritize which components are the most applicable to you today and in the future.

Here are some points to consider.

For more details in specific technologies, such as streaming data and data-as-a-service, including specific vendor choices, please see the Appendix below.

Databases and data warehouses

Databases and data warehouses have been the foundation of business intelligence for decades. Some of them continue to have their place in modern analytics architecture, while others are beginning to lose their relevance.

Some of the oldest databases are known as OLAP (online analytical processing). They began as a response to database technology that was slow, and used aggregations and caching to speed up response times to predictable queries. But as a company's questions have become more complex and harder to predict, OLAP can't keep up, often requiring completely new aggregations be built altogether. It is also losing relevance in the face of more improved database technology.

Today's databases take advantage of computing advances, such as in-memory and massive parallel processing (MPP) technology. This enables databases to deliver extremely fast performance with linear scalability, while optimizing for data storage, hardware memory usage, and sometimes even include built-in computational and data science functions.

Additionally, the arrival of the cloud has breathed new life into database technologies that on-premises versions simply cannot match. This includes the ability to start without procuring hardware, to scale elastically as business demands change, and without needing to build a team to manage the infrastructure.

There will always be a place for databases and data warehouses in modern analytics architecture, and they continue to play a crucial role in delivering governed, accurate, conformed dimensional data across the enterprise for self-service reporting. Even companies who adopt other technologies (e.g. Hadoop, data lakes) typically retain relational databases as a part of their data source mixture.

NoSQL, unstructured data, and data lakes

Databases and data warehouses are particularly powerful in supporting analytics when the data comes from predictable sources and formats. Of course, not all data is predictable. In modern analytics architecture, NoSQL databases are becoming part of every organization's arsenal due to the benefits of being able to quickly load data from any source, including data sources that do not have well-defined schemas or formats. NoSQL databases—sometimes called non-SQL or not-only-SQL—provide alternate types of data storage compared to traditional relational databases, including column, document, key-value, and graph storage types.

Related to unstructured data is the concept of big data and data lakes. Data is generated everywhere, sometimes in random places, and collecting it all and putting it into a usable format can be frustrating. Technologies have been developed to allow analytical tools to connect to the raw data as it is, instead of forcing the data to fit a certain format first.

One of these is called a data lake, a storage repository that can hold vast amounts of data in its native format, structured or otherwise. People can then analyze the data using optimized processing mechanisms like APIs or SQL-like languages to transform the data on the fly, without having to pre-process all of it into a specific format.

All these tools are often used in projects related to the Internet of Things, data science, streaming data, and other unstructured analytics use cases where the creation of data is unpredictable, both in volume and in location.

For a list of data technologies related to NoSQL, Hadoop, and data lakes, please see the Appendix.

Flat files

The prime days of Excel and CSV files are far from over. Whether you're a small or large organization, these flat files that seemingly appear out of thin air will continue to exist, probably forever. In fact, they're in more places than ever before. They used to just exist on someone's physical computer. Now, they sit in cloud storage systems, such as Google Drive or Dropbox. Third-party organizations produce flat files as part of data-as-a-service. They're useful as legends for various data fields, additional customer research, or any small bit of information that augments an existing data set because they are quick to create.

Additionally, apply the right security measures to flat files, at the right time. Encourage their use when necessary, particularly in one-off scenarios. If certain files grow in popularity, apply the right security protocols, ensuring that they're secure and accessible only by the right individuals.

Author and consume

The hallmark of modern business intelligence is the introduction of the business user to the business intelligence platform. Gone are the days when a decision-maker needed to request a report from IT and wait several days, only to receive an outdated report that didn't quite answer the question he or she had in mind. Today, the decision maker who has the question is also the one who can use the tools to answer it themselves. Because IT has enabled the entire organization to trust the available data, business users can make smart, data-driven decisions on demand without having to program.

The critical building block here is the actual analysis tool, something you can start using on day one. Even before you figure out which building blocks in this section are most relevant to your organization, you should start connecting to your data wherever it exists, if only to validate and rapidly explore your data sources as you prepare to build out your entire architecture.

While there are various components to analyzing data, we believe the core is visual analytics, a means by which anybody in your organization—whether they have programming experience or not—can connect directly to a data source and draw insights from it. Putting these kinds of tools in the hands of teachers, doctors, and sales members turns your organization from a change request slog to a well-oiled machine.

New to business intelligence is the additional focus on how you share insights with others. People are no longer confined to dashboards and reports; they can build full interactive applications, long-form articles weaving in data, text, and pictures, or even mobile-optimized viewing experiences.

Additionally, as businesses and their smaller departments grow, they turn to smart productivity tools to share information quickly, discover data sources, stay up to date with dashboards, and follow the most important metrics.

In this section, we look at a few components the best modern analytics tools provide to empower data delivery. For even more, such as Storytelling and Alerting, please see the Appendix.

Visual analytics

The human visual system is one of the most powerful tools in the world. And today, it is finally an integral way of analyzing data. Based on pattern recognitions the brain already uses daily, visual analysis can similarly reveal patterns in data, such as trends that slope upward and downward, irregular spikes in activity, or specific records that are outliers.

Traditional spreadsheets required you to analyze data in rows and columns, select a subset to share, and then create a chart. Whether through awkward wizards or text-based commands, these charts sometimes answered the question, and sometimes raised entirely new ones, but were always the final dead-end step. In contrast, visual analytics provides an elegant, intuitive analytical experience with simple drag-and-drop actions that make visuals a part of the process, and offers more than just a chart as the end result—the journey to insights is as valuable as the answer.

Visual analytics is not just pretty visualization tools. It is a language allowing you to combine data, spot anomalies, and augment data with calculations, groupings, what-if conditions, and much more, all without requiring programming assistance.

Traditional BI and Reporting

There's still a place for traditional BI, dashboards and reporting, though their creation process looks different today. Many static reports, such as executive dashboards or financial audits, required technical development skills to create, with analytical queries asked way in advance and often requiring the changing of underlying data models. All of this could take days, weeks, or months to develop.

In a modern analytics platform, many of these dashboards and reports begin as ad-hoc questions, which due to the nature of the questions they answer, are hardened and verified by IT and data stewards, and ultimately replace traditional static reports. This updated process takes advantage of business users' domain expertise as they dig through the data to find the right answers as questions evolve, change, and lead to entirely new questions. The flexibility of modern analytics is simply replacing traditional tools, even as those traditional report requirements persist.

Personal data prep tools

Not to be confused with ETL, data prep tools are lightweight applications designed to help non-IT users make powerful and precise manipulations to the data. They are built on the principle of ease of use, speed, and agility as visual analytics tools, which allow everyday business users to combine data sets, automate joins, rename fields, and make other improvements to data with the goal of getting it analytics-ready.

These should be slightly further along on your priority list, as it's often hard to know how you may need to modify data until you start using it. But over the lifetime of an organization, much of the time spent trying to answer a question is allocated to wrangling the data into the right form. Personal data prep tools are a powerful way to alleviate that time investment, without necessarily involving, nor negating, the need for an IT-built, predefined reporting semantic layer.

According to **"What's Your Data Strategy?"** from Harvard Business Review, "80% of analysts' time is spent simply discovering and preparing data."

Advanced analytics

Advanced analytics has emerged as a critical component of modern analytics architecture, with companies turning to statistics, predictive algorithms, and machine learning to maximize the value of very large data sets.

In the past, advanced analytics was only accessible to trained data scientists, often requiring programming language experience. Today, improvements in visual analytics have introduced more complex built-in features to all self-service analytics users, such as box plots, tree maps, clustering, basic predictive modeling, and correlations.

There are still many use cases for dedicated statistical analytics tools. Organizations looking to produce constantly evolving algorithms to offer customers "what to watch next," or perhaps write functions to determine when fraud has been detected on a customer's credit card, may want to investigate specific, dedicated tools to supplement their core visual analytics tool belt. Deploying these tools requires the requisite training and tool specific programming experience, which can take months to acquire.

Additionally, some integrations provide a middle ground where more complex statistical results from another tool can be pulled back into your visual analysis. Read more about advanced analytics extensibility in the Appendix.

Sharing and collaboration

In modern BI platforms, sharing, collaborating, and socializing insights is a key capability. From evaluating context to prioritizing the next best action, the impact of insights is maximized through collaboration. Modern analytics platforms provide discussion forums, annotations, commentary, favorites, likes, and other social concepts adapted from leading productivity and portal apps. Being able to communicate the insights gained from an analytics platform, directly within the tool itself, makes the flow of analytics much simpler, and encourages additional exploration and discussion of valuable findings. Collaboration can also be expanded into external applications and portals through embedding.

Embedded analytics

One of the most powerful, and often overlooked, business concepts is flow. Rather than take business users out of their standard operating processes to look up answers to data, insert those insights seamlessly into their established flows, and incorporate them into existing processes.

With modern analytics, you will find data and dashboards embedded directly into company portals, other applications, or integrated with productivity tools. The best analytics platforms support all these scenarios with mature APIs, software developer kits, and flexible delivery mechanisms, allowing you to easily jump from one tool to the next, and even combine all your tools together into a single portal.

Flow also extends to physical location. Today's workforce is often on the road, without access to internal resources behind firewalls. Modern analytics supports accessing data from anywhere, on any kind of device. This means a sales person can make an informed, data-driven decision from their mobile device, without having to pull out a laptop. It also means a construction manager can be on-site, accessing critical information without having to VPN into a corporate network. Mobile, and the cloud, have forever changed how a business can run its operations, and true modern analytics platforms must also empower businesses to leverage those advantages.

3. Tying it all together

These building blocks are the foundation for a modern analytics platform that empowers businesses to tackle any challenge head-on. Combined with a true partnership between business and IT, this platform provides everyone in your company with the confidence of knowing that for any decision they need to make, they have the tools to make the decision, and know that their decision is made on trusted data.

Putting together an entire modern analytics platform can seem like a challenge in itself. The good news is that you don't need to completely build out the entire ecosystem before you get started. In fact, today's businesses are most successful when they don't. Instead, they start small, make incremental transformations that ultimately inform where the business should invest next. Companies can begin with pilot tests before rolling technology solutions out to broader departments. You do not need to integrate every single component for an entire strategy to get off the ground. For example, you can use your visual analytics tool to discover holes in your data pipeline before completing your entire warehouse. This helps you create immediate value from analytics, find gaps and errors in your data, and ultimately build a more accurate and functional data warehouse.

The key is to work with tools that empower these types of incremental changes and a modern analytics platform is precisely a suite of building blocks that can be put together one at a time, making your organization more accessible, agile, and able to draw insight from a diverse range of data sources. This is exactly what world-class analytics leaders are doing, strategically leveraging visual analytics in combination with other best-of-breed big data analytics, Internet of Things, and data science solutions.

For example, Netflix has built a comprehensive big data platform and data lake to support the enormous amounts of data they generate from their operations. Tableau is the essential component that allows them to combine their disparate tools like S3, EMR, and Spark into a cohesive analytics platform that supports their business.

Regardless of where you are on the digital business transformation path, it's critical to immediately start using the data you have today. Businesses are only going to have to act faster in the face of the next big market disruption. Pick up a component of modern analytics and empower the business to make data-driven decisions and be the disruptor.

4. Appendix

There are many different analytics technology and solutions options, each with their narrow purposes and advantages. In this Appendix, we break down options within each one that were not covered in the overview.

Stream ingestion

Stream data is generated continuously by connected devices and apps located everywhere, such as social networks, smart meters, home automation, video games, and IoT sensors. Often, this data is collected via pipelines of semi-structured data. While real-time analytics and predictive algorithms can be applied to streams, we typically see stream data routed and stored in raw formats using Lambda Architecture and into a data lake, such as Hadoop, for analytics usage.

Lambda architecture is a data-processing architecture designed to handle massive quantities of data by taking advantage of both batch and stream processing methods. The design balances latency, throughput, and fault-tolerance challenges.

A variety of options exist today for streaming data including Amazon Kinesis, Storm, Flume, Kafka, and Informatica Vibe Data Stream.

Integration hub orchestration

Hub-and-spoke integration patterns are an easily understood and widely used data architecture design. Hubs decouple data sources located anywhere, and target enabling more flexible integration by reducing the number of point-to-point interfaces to manage. Integration hubs publish/subscribe capabilities foster data reuse, and deliver central control for the purposes of optimization, data standards, and governance. Centralized management brings improved visibility across all orchestrated data movement pipelines that span data sources everywhere.

New generation data integration hubs extend traditional capabilities to self-service analytics users. Anyone can publish or subscribe to modern integration hub data feeds with minimal involvement from IT. Data consumers can leverage certified data, get visibility into lineage, and integration processes. Other benefits of the modern data integration hub include seamless data quality functions, expedited data source on-boarding and just-in-time delivery of small or massive data sets.

Informatica and Cisco are market leaders in data integration hub technology. Tableau's deep integration with Informatica allows you to combine hundreds of different data sources into Tableau Data Extracts, stored and kept up to date on Tableau Data Server for use by anyone in the organization.

More on unstructured data, NoSQL, and data lakes

Data lakes support modern big data analytical requirements through faster, more flexible data ingestion and storage for anyone to quickly analyze raw data in a variety of ways. Data lakes do not replace data warehouses.

In modern ingest-and-load design patterns, the destination for raw data of any size or shape is often a data lake. A data lake is a storage repository that holds a vast amount of data in its native format—structured, semi-structured, and unstructured. Data lakes also provide optimized processing mechanisms via APIs or SQL-like languages for transforming raw data with "schema on read" functionality.

Although Hadoop has been used for data lakes since the first Hadoop Distributed File System (HDFS) due to its resilience and low cost, it is not the only data lake implementation option. Object stores, such as Amazon Web Services Simple Storage Service (S3) and NoSQL databases with flexible schemas can also be used as data lakes. Tableau now supports Amazon's Athena data service to connect to Amazon S3, and has various tools that enable connectivity to NoSQL databases directly.

In modern analytics architectures, NoSQL databases are becoming the norm due to the benefits of quick data loads from data anywhere and schema-less database concepts. NoSQL, non-SQL or not-only-SQL databases provide alternate types of data storage. Common NoSQL storage types include column, document, key-value, and graph.

Examples of NoSQL databases that are often used with Tableau include, but are not limited to, MongoDB, Datastax and MarkLogic.

While Hadoop is often used as a big data platform, it is not a database. Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, massive processing power, and the ability to handle extreme volumes of concurrent tasks or jobs.

In a modern analytics architecture, Hadoop provides low-cost storage and data archival for offloading old historical data from the data warehouse into online cold stores. It is also used for IoT, data science, and unstructured analytics use cases.

Within the Hadoop framework, related technologies for loading, organizing, and querying data include, but are not limited to, the following:

- **Apache Spark** – Open-source cluster computing framework with highly performant in-memory analytics and a growing number of related projects
- **Apache Impala** –The open-source, analytical MPP database for Apache Hadoop. This is the data connection most commonly used in successful Hadoop related projects with Tableau
- **Apache Presto** – An open-source distributed SQL query engine for running interactive queries across data sets of all sizes. Tableau added Presto support in Version 10
- **MapReduce** – A parallel processing software framework that takes inputs, partitions them into smaller problems and distributes them to worker nodes
- **Hive** – A data warehousing and SQL-like query language. Hive 2.0 also includes LLAP (Live Long and Process) which dramatically improves Hive query performance.
- **Hadoop Distributed File System (HDFS)** – the scalable system that stores data across multiple machines without prior organization
- **YARN (Yet Another Resource Negotiator)** – Provides resource management for the processes running on Hadoop
- **Ambari** – A web interface for managing Hadoop services and components
- **Cassandra** – A distributed database system
- **Flume** – Software for streaming data into HDFS
- **HBase**– A non-relational, distributed database that runs on top of Hadoop

- **HCatalog** – A table and storage management layer
- **Oozie** – A Hadoop job scheduler
- **Pig** – A platform for manipulating data stored in HDFS
- **Solr** – A scalable search tool
- **Sqoop** – Moves data between Hadoop and relational databases
- **Zookeeper** – An application that coordinates distributed processing

Notably, over the past two years, Apache Spark has moved from being a component of the Hadoop ecosystem to the stand-alone big data analytics platform of choice for a number of enterprises. Spark provides dramatically increased data processing speed compared to Hadoop. Spark itself has many related projects including the core Apache Spark runtime, Spark SQL, Spark Streaming, MLlib, ML, and GraphX. It is now the largest big data open source project with 1,000+ contributors from 250+ organizations.

Tableau is a market leader in big data specific analytics connectivity and visual data analysis. Best-in-class big data analytics programs are using Tableau with Cloudera, Spark SQL, Amazon EMR, Hortonworks, Microsoft HDInsight/Data Lake, and MapR. Many other big data technologies can be connected to Tableau through these overtly supported technologies, or their drivers.

Data as a service

In a digital world where data is gold, data is also a product for anyone to consume. Customer, financial, market, weather, geographical, and demographical data is already being offered as a service for purchase in data markets and trading platforms.

Data-as-a-Service applies a flexible Service Oriented Architecture (SOA) pattern for data delivery via the cloud. This approach offers extreme agility since the SOA architecture is simplistic. Today we see ISVs, CRMs, and ERPs provide standard data-as-a-service REST APIs for integrating or external reporting scenarios.

Tableau's Web Data Connector SDK allows people to build connections to data that lives outside of the existing connectors. Self-service analytics users can connect to almost any data accessible over HTTP including internal web services, JSON data, and REST APIs.

Logical data warehouse

Leading analytics organizations are delivering flexible, logical, and unified dimensional views of data everywhere via data virtualization technologies from vendors such as Cisco and Denodo. For analytics users, a logical data warehouse looks and acts just like a relational data warehouse. Tableau users can connect to these tools using off the shelf ODBC drivers.

One of the key capabilities of data virtualization is optimization of remote distributed heterogeneous queries to a variety of different data sources and REST APIs. The logical data warehouse also serves as a semantic layer buffering reporting applications from data source changes. Logical data warehouses are often used with an enterprise data catalog.

Master data management

Analytics is only as good as the quality of the data being used, and people can only make the right decisions if their data is accurate. With more subject matter experts creating data in a bottoms-up approach, we are seeing renewed interest in traditional data quality and master data management that are once again vital for ensuring reporting data sources are current, cleansed, consistent, and accurate.

Popular master data management offerings include, but are not limited to, Informatica, IBM, and Stibo. Several data quality solutions most often used with Tableau are Trillium, Informatica Data Quality, Talend Data Quality 6.0, and Tamr Eisenhower.

Enterprise data catalog

Another emerging technology is the enterprise data catalog. Enterprise data catalogs allow self-reliant reporting users to easily find the right data for decision making from approved data sources. Enterprise data catalogs exist within visual analytics solutions and are also available as stand-alone offerings designed for seamless integration with Tableau.

Enterprise data catalogs are populated with metadata from tables, views, and stored procedures by scanning ingested data sources. With automated discovery of new data sources, intelligent data classification, and cross-data source entity mapping, a data catalog essentially serves as an enterprise business glossary of data sources and common data definitions. Subject matter experts further enhance catalog data source context by adding annotations, versions, and documentation.

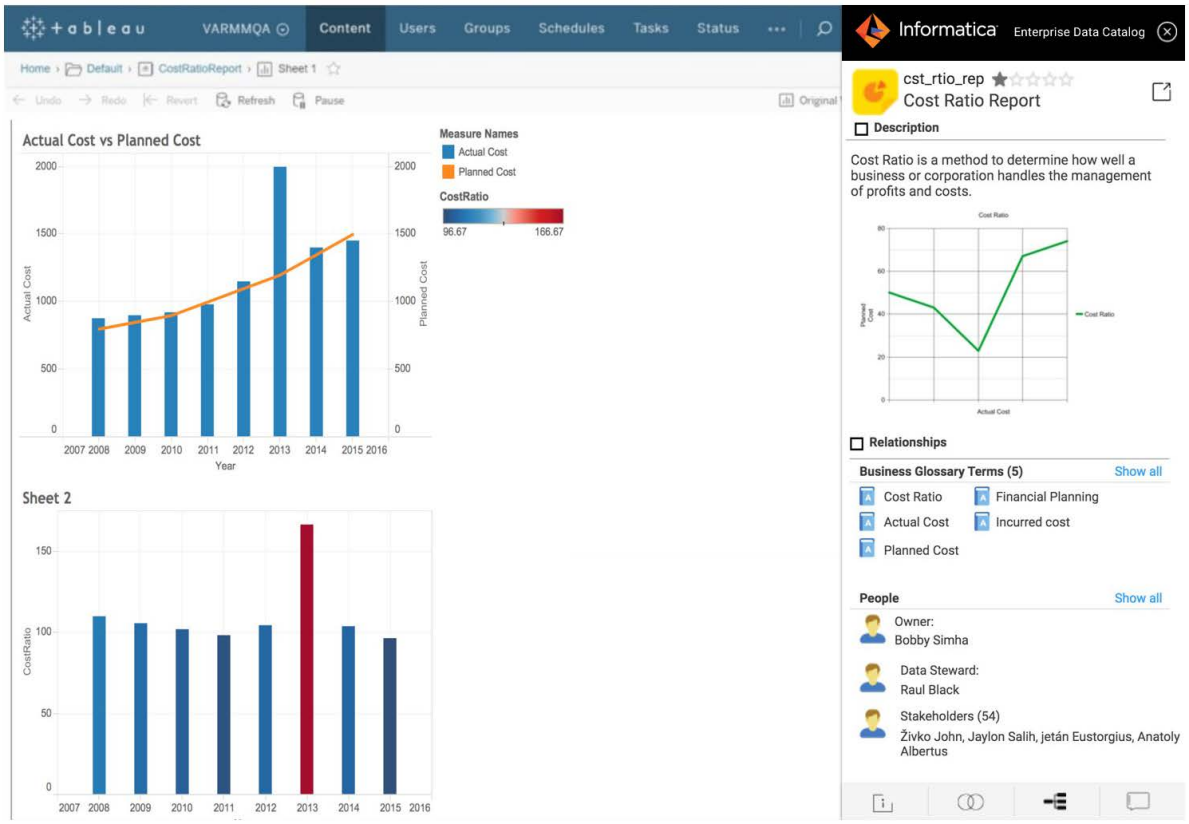

Figure 3 An Enterprise Data Catalog from Informatica

Data catalog solutions foster curation of data and efficient reuse of existing data. They also provide highly desirable data lineage and an additional level of data governance, security, logging, and auditing.

Vendors with rich data catalogs that integrate quite nicely with Tableau include Informatica, Alation, Unifi, Collibra, and Waterline.

Machine learning

Taking advanced analytics even further, cognitive, deep learning, and artificial intelligence make inferences from existing data and patterns, draw conclusions based on existing knowledge bases, and then insert this back into a knowledge base in perpetual, continuous, self-learning loops.

Consumption of this type of analytics typically means reviewing output via an embedding API in a report or integrated application. Tableau is used today to visualize output from CognitiveCode, Digital Reasoning, and other vendors.

In Tableau 10.3, we debuted recommended tables and smart joins to save you time when connecting to and preparing your data. Backed by machine learning, recommendations improve over time as databases are used more often.

Natural language

Taking data storytelling further, natural language and speech query are beginning to bring data discovery to everyone in new and more flexible ways. Natural language makes analytics more accessible on any platform by getting automated contextual descriptions of key findings, asking for forecasts, or analyzing volumes of text documents.

Today, Tableau's visual analytics can be combined with leading natural language generation (NLG) solutions such as Yseop, Narrative Science, and Automated Insight. Since these technologies are largely interpreting the context of Tableau visualizations, the integrations are most frequently performed in the Natural Language tool itself, or as an extension through JavaScript. Further, the acquisition of ClearGraph will enable smarter data discovery and analysis directly in Tableau, making it easier to interact with data through natural language.

Search

Modern analytics architectures allow users and experts to search for and find data with Google-like ease, regardless of location. Rather than modeling data, analytical search engine indexing technology automatically relates different data sources based on field names, data types, and machine learning intelligence. Over time, dynamic search suggestions are generated based on historical queries and reported usage. Recently, with the addition of speech technologies such as Siri and Alexa, we are beginning to see voice query capabilities combined with analytical search.

Advanced analytics extensibility

SMachine learning algorithms and statistical analysis can provide much deeper analytical ("what will happen") and prescriptive ("how to optimize") capabilities with data formatted for that purpose. Tableau not only connects to file outputs from MATLAB, R, SAS, and SPSS as data sources, but supports direct integration with R and Python. You can run code directly inside Tableau, as well as visualize and manipulate model results from predictive services such as Rserve and TabPy.

Alerting and subscriptions

Some tools provide snapshots at regular intervals; others actually trace logs to see if numbers have crossed certain thresholds. There are reasons for both. Some dashboards are informative, and you simply want to check them every day. Others are the basis for critical action, but checking a dashboard every day, without any actionable insight, is not an efficient use of time.

A modern analytics architecture includes configurable, intelligent, data-driven alert notifications that continuously monitor for a valuable signal in the digital ocean of data. It is impossible for a human to check each and every important value 24 hours a day, 7 days a week. This is where automation, alerting, and subscriptions are wonderful assets in a modern analytics arsenal.

In Tableau, you can stay on top of your business with data-driven alerts for Tableau Server. Just choose a threshold to receive an email alert for yourself or your whole team. Subscriptions are incredibly flexible, allowing email snapshots at any desired interval. Plus, there's an empty-view option that only sends emails when data exists in a view—a good choice for high-priority alerts.

Storytelling

Sometimes, the insight—or the "what"—isn't enough. People also want to know the "why" behind the data. Why did sales increase? What caused the spike in web traffic? Why are we struggling to keep medical supplies stocked?

Businesses have long tried to solve this problem by combining analytics with other forms of communication: text, pictures, even videos. Analysts build presentations using PowerPoint, or write long reports as PDFs, or even more cumbersome, print out pages and pages of documents, and put them together in a binder.

Today's modern analytics tools take the best of these storytelling concepts and actually integrate them as first-class features. They let you build interactive dashboards, send specific snapshots of data that automatically update in the background when new data is added, or even creating reports that combine interactive charts with text and images. Storytelling empowers users to explain the analysis of the data, instead of just providing a number.

Tableau fundamentally values the importance of choice and open standards. We invest heavily in research and development to make analytics faster and easier, and that also means innovating with our ecosystem of partners. This ensures that as the analytics world evolves and new technologies come into play, analytics leaders will always be able to integrate Tableau with their choice of current and future data technologies.

About Tableau

Tableau helps people see and understand their data, no matter how big it is, what channel it's coming from, or what database it's stored in. Quickly connect, blend, and visualize your data with a seamless experience from the PC to the iPad. Create and publish marketing dashboards with automatic data updates, and share real-time insights with colleagues, teams, executive leaders, partners or customers—no programming skills required. Try it for free today!

Additional resources

Modern BI Evaluation Guide

Define Analytics

Approach to Analytics Webinar

Advanced Analytics with Tableau