THE MODERN ANALYTICS STACK

The analytics stack has been undergoing fundamental AI-driven restructuring to compress superior performance into a robust, streamlined machine. These are the components, features, and vendors of the next generation stack, that will turn your analytics into a powerful competitive force.

anodot

INTRODUCTION

Analytics are increasingly becoming a core differentiating capability for data-driven businesses, with growing demands on data and analytics leaders.

While all analytics stacks are expected to perform their basic functions of making data accessible, secured and governed, there are increasing expectations within boardrooms and in-house data customers for analytics to enable, deliver and empower the insights that are crucial for strategic business decisions.

Driven by the proliferation of AI, with an emphasis on machine learning, the analytics stack has been undergoing fundamental restructuring. Increased implementation of autonomous and augmented processes has assimilated manual and code-based analytics operations into the stack, effectively decreasing the number of essential components and platforms.

The modern stack compresses superior performance into a more robust machine, built of fewer moving parts. With reduced integrations, transformations and manual data science processes comes a significant reduction of both data handling and errors, and of the overall resources needed to operate, manage and support business analytics. On the data customer end, increased transparency and usability creates unprecedented accessibility to data and insights across the business, for technical and business users alike.

There are many methods of cutting the analytics cake. In the on-going analytics stack discussion unique functionalities are often referred to by different names (i.e. collection/integration), and different stack elements are sometimes broken into sub-elements or, on the contrary, grouped into higher-level elements. For the sake of this discussion we will cover the five key stack building blocks:

- **Data sources**
- **Data pipelines**
- **Data warehousing**
- **Analytics & visualization**
- **Business monitoring & anomaly detection**

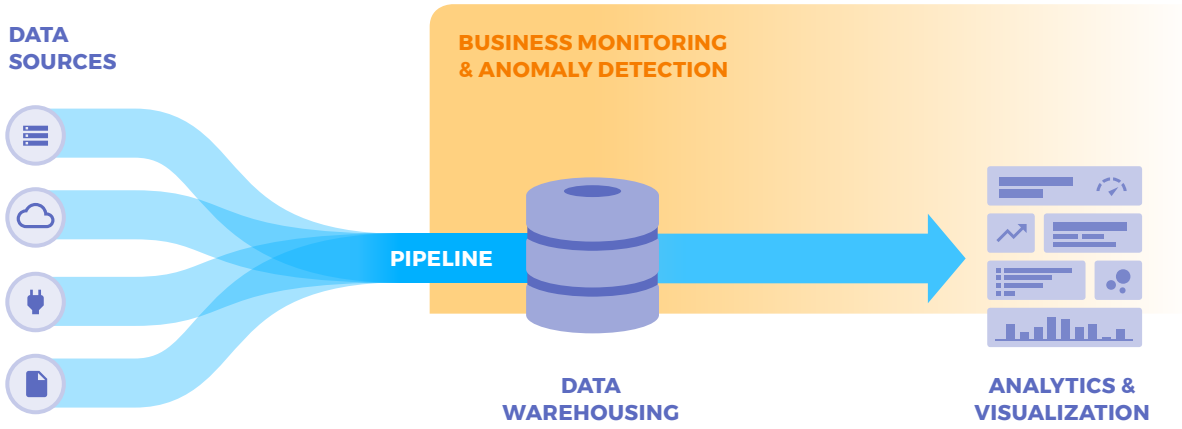
The analytics market offers end-to-end analytics, as well as solutions that cover more than a single stack functionality. Sisense, for example, started out as a single stack solution, but with the advent of anomaly detection extended its offering to

include monitoring. Panoply provides warehousing services, but as a next generation technology loops the data pipeline in one system, offering a hit the ground running approach to implementation and integrations.

The two key takeaways to keep in mind are the need to ensure that every element of the stack is covered, and that often the optimal solutions, especially for high-performing businesses, are best-of-breed platforms that excel in their core functionality, rather than broader offerings that often stretch at the seams.

Data and analytics leaders are hard pressed today to deliver solutions that satisfy both the needs of IT for infrastructure and ongoing support for production level solutions, and the requirements of business stakeholders and customer-facing units for realtime flexible reporting and insights. These solutions are also expected to withstand scalability, incorporate future data sources and formats, and enable mature workflows.

While traditional BI systems may accommodate some of these needs to an extent, the modern analytics stack is designed to create a seamless analytics environment that sits as close to the business as possible, in which the ability to identify opportunities and threats is available to all stakeholders business-wide, effectively turning analytics into a powerful competitive force.



DATA SOURCES

The overarching goal of BI analytics is to provide the most clear and comprehensible picture of operations across the business in order to support optimal business decisions. This goal is closely dependant on the ability to correlate between disparate data, so that, for example, a spike in customer support calls can be associated with marketing activities, app usage with sales and operations, server side with client side.

So, while your data sources are not actually part of your analytics stack, they are the fuel that drives it.

For most companies, however, data is siloed between separate departments, applications and systems, including but not limited to CRM, product analytics, ERP, web analytics, search analytics, marketing, advertising and customer support. In addition, data is stored as databases, files and events.

According to an American Management Association survey, 83% of executives think their companies have silos, and 97% think it's having a negative effect on business. The problem isn't limited to management—it affects every department and operation. For example, marketers worldwide say their primary challenge in executing a data-driven customer experience strategy is a fragmented system to deliver a unified view of the customer experience (38%), followed by silos of customer data that remain inaccessible across the entire organization (30%).

Individually, none of your data sources can offer a complete picture of what's going on. The more data sources you have the more (and deeper, and better) insights you can glean, but only pulling all your data together can truly reflect your business reality. To make use of this data, you'll need to pipe it into a data warehouse that organizes and stores it under one roof, making it easy to combine, query, and analyze.

But while many companies recognize that data silos are a problem, consolidating and integrating them is a major technological and organizational challenge.

It's up to the next component of your analytics stack—the data pipeline—to solve it.

DATA PIPELINES

Across the business's systems, data is stored in different types, formats and schemas. Duplicate data sources create redundancies. Metrics and key performance indicators lose consistent meaning, as do the calculations behind them. Reconciling data silos by integrating all company data into one platform depends on overcoming this data chaos.

Before it reaches the warehouse, all data needs to be cleansed, prepped, formatted, and occasionally enriched. This conversion is achieved through ETL processes, which involve the extraction, transformation and loading of data into the warehouse.

Traditional analytics rely on manual or semi-manual ETL transformations. These are labour intensive processes: it's estimated that in traditional environments, BI professionals spend 80% of their time cleaning and preparing raw data for analytics.

Businesses which strive to mitigate this bottleneck through automation by building their own ETL processes from scratch, quickly discover that it's an impractical undertaking. Engineering a custom pipeline from just one platform is complex; repeating that task for the 140 SaaS and other apps which make up the average company's data source landscape verges on the impossible.

A comprehensive solution to the problem comes from modern ETL tools, also known as Data Pipelines. These platforms autonomously sync, transform and load data from any source into the warehouse. Data pipelines can be roughly divided into two categories. Data Syncers have streamlined setups and workflows for quickly moving data from one source to another. Data Transformers not only sync data, but also transform and enrich it through more involved applications, while usually also supporting a more diverse set of data sources and offering enhanced functionality.

Evaluation criteria for data pipelines should involve considerations regarding data quality, completeness, integrations, usability, security and governance. The optimal solution will vary according to your business's unique use case, but core competency demands the ability to continuously, seamlessly and securely push data from any source, while protecting data integrity, quality and governance.

Leading Vendors

- Etleap is a fully managed ETL solution built specifically for Amazon Redshift. It allows data teams to easily set up pipelines that extract data from any source and load clean and structured data into Amazon Redshift. Etleap also has a powerful transformation engine that lets you clean, structure, and enrich your data without writing code.

- Fivetran is a fully managed data pipeline with a web interface that integrates data from SaaS services and databases into a single data warehouse. It provides direct integration and sends data over a direct secure connection using a sophisticated caching layer that effectively moves data from one point to another.

- mParticle is a real-time customer data pipeline that supports data collection from a wide variety of sources. The platform provides standardization; cleansing and deduping; tags; data enrichment via scoring, contextual or behavioral data; and segmentation and customer profile management.

- Tealium is a slightly different offering, focused on the unique needs of customer data. It provides end-to-end management of customer data across the analytics and marketing technology stack, including tag management, API hub, customer data platform, and data management solutions for data collection, standardization, transformation and enrichment, integration, segmentation, and activation via multiple execution channels, with turnkey integrations to over 1,000 client-side and server-side vendors and technologies.

DATA WAREHOUSING

At last, you now have all your data under one roof, stored in a central repository and ready to be queried and analyzed.

Getting your data into a warehouse you control, and can connect to any analytics tool of choice, is vital for keeping your business growing quickly. But a modern data warehouse is much more than a storage solution. To analyze big data sets, you need to store that data someplace that has the scale, performance, and power needed to get the job done—a modern warehouse that can scale to accommodate any size of data and any number of analysts.

With the rapid growth of data creation and ingestion, conventional on-premise data warehouses (DBMS) and big data solutions struggle to deliver on their fundamental purpose: making it easy to amass all data and ensure availability and reliability to enable rapid analytics for all business users.

The cloud storage paradigm has largely overcome these obstacles. As its name implies, cloud-based data warehouses don't require on-premise hardware and servers, and significantly reduce IT work and operational costs. The implications are low entry costs, faster setup, improved connectivity, and, of course, scalability. Cloud solutions enable businesses to manage data on a scale of just a few hundred gigabytes—to a petabyte or more, allowing businesses to grow without having to think about capacity, reliability or performance.

Other evaluation criteria for data warehouses should include speed, security and availability. Make sure that your vendor of choice offers solutions to meet your unique data transfer needs and get your data on the cloud quickly and securely, and ensure a very high up-time, making the warehouse reliable no matter what.

Leading Vendors

- Amazon Redshift is a leading enterprise-grade cloud data warehouse that handles petabyte-scale data while enabling analysts to query it in seconds. Redshift delivers fast performance by using machine learning, massively parallel query execution, and columnar storage on high-performance disk. Redshift offers quick setup, limitless scalability, and high concurrency on Amazon's industry leading architecture with no upfront costs.

- Google BigQuery is another enterprise-grade cloud based data warehouse that offers rapid SQL queries and interactive analysis of massive datasets. Like Redshift, it can run blazing-fast queries on datasets of petabyte scale. The platform utilizes a columnar storage paradigm that allows for much faster data scanning as well as a tree architecture model that makes querying and aggregating results easier and more efficient. BigQuery is serverless and built to be highly scalable thanks to its fast deployment cycle and on-demand pricing.

- Snowflake is a new generation cloud data warehouse built on top of the Amazon Web Services (AWS) cloud infrastructure, and is a true SaaS offering. There is no software or virtual or physical hardware to select, install, configure, or manage. All ongoing maintenance, management, and tuning is handled by Snowflake. Snowflake is built for speed, even with the most intense workloads. Its patented architecture separates compute from storage so you can scale up and down on the fly, without delay or disruption.

- Panoply is another new generation cloud data warehouse that can be effortlessly set up to gather all your data in one place, and seamlessly connect to any BI tool for lightning fast analytics with scalability, high availability and fast querying built in. It also provides end-to-end data management, automating all data preparation tasks. Panoply's proprietary self-optimizing architecture uses machine learning and natural language processing to model and streamline data from source to analysis to deliver insights in minutes.

ANALYTICS & VISUALIZATION

You have finally arrived at your destination: a solid and robust data infrastructure which continuously streams all processed business data into one centrally accessible platform. But stored data is meaningless if the business can't utilize it to make decisions. This is where your analytics and visualization layer comes into play.

This part of the stack combines interactive analysis tools, visualization tools, and a collaboration platform, and it is where you will analyze your data and build insightful dashboards that can drive your business. BI analytics is the client side of your stack, and should optimally enable all in-house analytics customers, including non-technical users, to ask and answer their own data questions.

Your analytics & visualization platform should allow you to:

- Query all your data with analytics engines that deliver fast query performance at scale for any kind of workload regardless of concurrency, size or complexity.
- Analyze and explore data for BI reporting and advanced analytics operations—for data experts and business users as well.
- Build dynamic visualizations and communicate insights more effectively with a wide range of visualization options.
- Share insights across the business in real-time with direct linking to live reports.
- Enable end users to explore and drill down to find their answers.

With the advance of AI and ML, BI analytics are becoming more intuitive, and more proactive. Augmented analytics, a paradigm that includes natural-language query and narration, automated advanced analytics and visual-based data discovery capabilities, is rapidly becoming a dominant driver of new purchases of business intelligence.

According to Gartner, by 2020, 50% of analytical queries will be generated via search, natural-language processing or voice, or will be automatically generated. Advanced BI platforms already offer these augmented capabilities, which are forecasted to deliver twice the business value compared to traditional BI.

Leading Vendors

- MicroStrategy is a leading pure-play enterprise business analytics platform that includes a range of data analytics, reporting and visualization applications and data connectivity and preparation tools. The platform offers augmented analytics with contextual BI-driven information, and enables users to develop BI apps on the Android and iOS platforms in a no-code/low-code environment

- Sisense is a new generation of end-to-end BI platform the handles preparing, loading, analyzing and visualizing big data using a drag-and-drop interface, built to eliminate the barriers between business users and business analytics software. Sisense uses a proprietary in-chip (as opposed to in-memory) architecture to speed up data tasks from ingestion and cleaning to visualization and dashboard construction, making it easy to work with large datasets.

- Tableau one of the most popular data analysis and visualization tools, built for data scientists, analysts and business teams. Its low-code interface makes it easy to generate a wide range of attractive visualizations and interactive dashboards. Tableau offers a set of over 150 built-in functions for calculations and data analysis, and a wide range of pre-built data connectors.

- Looker is a cloud-based BI platform known for its innovative approach to data modeling and exploration. Looker uses its own proprietary modeling language, LookML, which provides a data modeling layer that is both reusable and maintainable, yet its steep learning curve makes it less approachable than other alternatives. Looker connects to a database or data warehouse directly, without the need to extract data, and auto-generates a data model from your schema. Uniquely, it also works on fresh data direct from the source.

BUSINESS MONITORING & ANOMALY DETECTION

Your analytics stack is now set to provide data access and transparency across the organization, empowering truly data-driven business decisions.

But even tech-savvy companies using advanced analytics stacks manage to monitor only a small percent of the data created, collected and stored by the business. According to Forrester, on average, between 60% and 73% of all data within an enterprise goes unused for analytics. While your analytics & visualization platform empowers users across the business to slice, dice and deep dive into the data, the insights they provide are limited by time and resources, personal ability, and cognitive biases. The sheer volume and variety of data streams, KPIs, and unique metrics and dimensions, is humanly impossible to monitor.

Within these millions of data events occurring daily, some are symptoms of problems with a direct effect on operations, customer experience, and revenue. Spikes and drops in usage, payment processing failures, API glitches, server downtime, security loopholes either go unnoticed, or detected only after adversely impacting the business, resulting in lost revenues and resources.

Providing holistic data coverage requires a monitoring / anomaly detection platform that proactively audits the full range of your data, independently of the above mentioned human restraints, to alert you to any and all data anomalies. Advanced anomaly detection—the 5th element in the modern analytics stack—is crucial for securing any business against late detection and resolution of incidents occurring within their business's dark data.

Advanced anomaly detection platforms rely on AI and machine learning to monitor, analyze and correlate 100% of company data in real-time, all the time. By providing full data coverage, dynamic baselines and anomaly scoring systems, these tools turn billions of data points into single mission critical alerts, empowering early incident detection for every scenario across the business. This final analytics layer is responsible for ensuring that no insights that can impact your business go unnoticed.

A robust, innovative anomaly detection system is the gatekeeper of your business and the frontline protector of your customers and revenue.

Leading Vendors

- Anodot is an AI-based monitoring platform built from the ground up to monitor, analyze and correlate 100% of integrated data sources in real-time. The platform autonomously learns the normal behavior of every metric to billions of data events into the single, scored, timely actionable alerts. There's no need to define what data to look for or when, and no manual thresholds to set up or update. Anodot's patented machine learning significantly decreases time to detection and resolution of incidents across the business.

- Outlier is an artificial intelligence-based analysis software that collects and monitors business data from several SaaS services and databases to identify unexpected changes or patterns across data. While Outlier is a recent newcomer offering a limited amount of data integrations, it does connect to the major data warehouses such as Amazon Redshift, Google BigQuery and Snowflake.

- NewRelic APM is an analytics product for application performance monitoring that delivers real-time and trending data about web application performance and the level of satisfaction that end users experience. The tool offers end-to-end transaction tracing and a variety of color-coded charts and reports, and provides both current and historical information about memory usage, CPU utilization, database query performance, web browser rendering performance, app availability and error analysis, external services, and other useful metrics.

- Splunk is a data monitoring platform that collects, analyzes and visualizes machine data from all levels of the IT stack, including applications, infrastructure and wire data on the network, so organizations can make business-critical decisions tied to monitoring, troubleshooting and planning. Splunk delivers massive scale and speed to give the real-time insights needed to boost productivity, security, profitability and competitiveness of IT, security, and IoT operations.

CONCLUSION AND FUTURE TRENDS

In today's digitized economy, the ability to use data effectively represents a real and essential competitive advantage. It's a strategic goal for the entire company and, when addressed properly, will lead you to develop a future-proof data infrastructure. With the increased proliferation of AI technologies, analytics, the core of digital business, is coming to a critical inflection point. Forrester and McKinsey are now converged in their opinion that the BI space has been fundamentally changed, and that the traditional legacy BI approach is no longer adequate.

According to Gartner, AI-driven analytics is already a dominant driver of new purchases of business intelligence and analytics platforms. By 2020, the number of users of modern business intelligence and analytics platforms that are differentiated by AI-driven data discovery capabilities is predicted to grow at twice the rate — and deliver twice the business value — of those that are not.

As you consider your company's unique analytics needs, it's worthwhile keeping in mind prominent trends which are already shaping the future of data-driven businesses:

- **Increased automation** — cross the entire analytics workflow, machine-learning automation is already being leveraged for data preparation, data discovery and data analysis using advanced analytics models.
- **Advanced features** — modern BI analytics tools are ushering in AI-driven features, including proactive insights and conversational analytics, which enable users to generate queries and explore data in natural language.
- **Advanced forecasting** — next generation forecasting uses deep learning to automatically optimize forecasts by working on data streams in real time and adapting predictive models as internal or external conditions change.
- **Analytics for everyone** — as data and analytics become increasingly automated, conversational and accessible, adoption by non-technical users will increase by an order of magnitude, effectively relegating analytics from the realms of IT and turning it into a common language spoken throughout the organization.

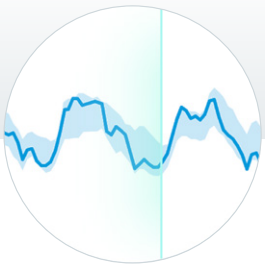
In the face of exponential data growth and the dynamic nature of the digital landscape, the modern analytics stack builds on these technological advancements to enable faster time to insight and, ultimately, faster time to action and impact on business outcomes. With data and business analytics solidifying their stance as one of the most critical elements in business today, the modern analytics stack is key in providing a perpetual competitive edge.

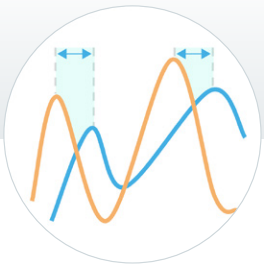
THE MOST ADVANCED BUSINESS MONITORING, STRAIGHT OUT OF THE BOX

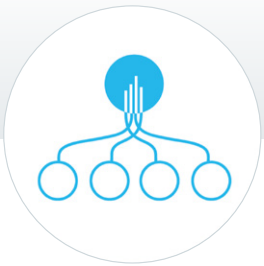
Anodot's autonomous business monitoring platform leverages advanced machine learning techniques to constantly analyze and correlate every business parameter, providing real-time alerts and forecasts, in their context, lowering time to detection and resolution.

Our leading-edge, patented technology is trusted by clients such as Facebook, Microsoft, Lyft, Waze, Pandora, Appnexus, Wix and King, in industries ranging from eCommerce to finserv, adtech, telco, gaming and more.

We have your back, so you're free to play offense and grow your business.







COMPREHENSIVE

All the data

Cross silo

Data agnostic

CONTINUOUS

Realtime

All the time

ADAPTIVE

Adjusts to changes

Autonomously learns baselines and seasonality

SPOT ON

Root cause guidance

Accurate

Actionable

anodot

www.anodot.com