

Native Visual Analytics for Big Data:

Scaling Analytics on Apache Hadoop and the Cloud

In this white paper, we explore how enterprises can utilize native visual analytics software using Arcadia Enterprise to provide real-time insights by unifying visual analysis, data discovery, and business intelligence (BI) within modern data platforms such as Apache Hadoop, NoSQL, and the cloud. In addition to using analytics internally for understanding and optimizing business processes such as how to incorporate and respond to customer feedback, or reduce errors in the manufacturing process, organizations need to move from “analytics” to “data applications” which drive competitive advantage and better customer service in real time, to impact business as it happens.

In the big data age, the requirement “to disrupt or to be disrupted” with digital business transformation requires organizations to build a data strategy which utilizes data-centric applications to provide value-added services back to customers and partners. For example, companies such as Fitbit and Strava provide analytics and dashboards back to users and allow them to compare their performance over previous attempts as well as against their peers. Without these data services, these stand-alone companies would be commoditized by companies such as Apple and Samsung who now provide basic data applications in their devices. The same differentiation through digital services is happening in automotive, media networks, healthcare, and more.

The future of scalable enterprise software and applications is tightly bound with Hadoop and other massively-parallel distributed systems. The big data movement (volume, velocity, variety) has required scale-out data storage and platforms, but analytics systems have not kept up, with the exception of data-engineering-led organizations who can wrangle and build data applications and analytics with data science, math, and code. Therefore, in order for the modern enterprise to keep up with the scale of big data, visual analytics and business intelligence (BI) systems must re-architect to a similar distributed, scale-out approach,



Arcadia Data lets you easily create visual applications to solve business problems such as cyber security using big data analytics.

which is core to the next phase of the big data movement and digital transformation. Ideally, this distributed BI and analytics approach should also be "data native" and run where the data resides, leveraging existing storage and compute resources to accelerate time to insights while reducing system complexity. This paper will go into more details on what it means to be data native.

Arcadia Data provides the first native visual analytics software that runs within modern data platforms for the scale, flexibility, performance and security users need to glean meaningful and real-time business insights and design data-centric applications in the era of big data. Arcadia Enterprise is a fully distributed, massively-parallel processing (MPP) analytics platform purpose-built to analyze large volumes of data without extracting it, therefore filling the gap between self-service BI tools and data-scientist-level advanced analytics for use cases like the connected car, cyber security, and trade surveillance.

Arcadia Enterprise provides a comprehensive, integrated analytics package for accessing, analyzing, designing, and sharing insights and data-centric applications, running directly inside ("in-cluster or in-cloud") big data environments and modern storage engines such as HDFS-compliant data platforms, NoSQL databases, MPP RDBMSs, and object storage such as Amazon S3.

Arcadia Enterprise was built from the ground up, using the latest web technology as well as native HTML5 that's extendible with custom CSS and JS. The result of this data-native approach? Ultra-fast business insights with radically lower cost because no data movement is required to get high-definition access to all granular data in its native state. This is the only way to keep up with the scale and speed of our modern big data world.

Agility

The Arcadia Enterprise native visual analytics architecture lets you explore data quickly and directly without having to start with extracts, cubes, or data marts.

With Arcadia Enterprise, the intuitive visual interface makes it easy to do business-driven exploration and semantic modeling directly on all of your data. The native web technology is completely server-based and both managed and used via a web browser on desktop or mobile devices with no client or driver installation required. This browser-based approach means that users simply point to a URL and everyone has direct, live access to the latest information, enabling collaboration across teams without having to email extracts and report artifacts across desktops.

Self-service visual analytics and BI

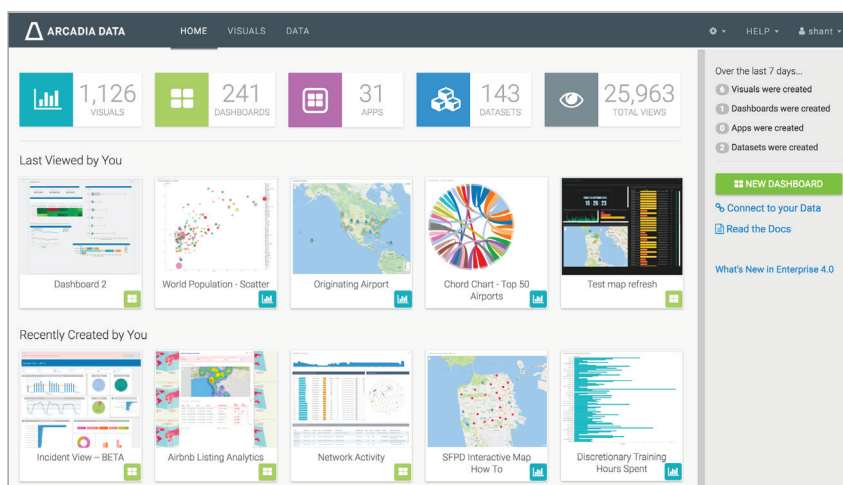
In traditional BI and analytics systems, data cubes or extracts are basically pre-aggregated results tables that provide faster application response times, since they don't need to scan every row of a base fact table when providing the data needed by the application. But because all of the data joined in a cube is already aggregated, it prevents analysts from getting a complete picture of their data; you lose the opportunity to gain insight from seeing all the data in one place across functional silos.

With Arcadia Enterprise, you have direct access to all of your data. Arcadia Enterprise tackles high-definition big data analytics on a dramatically simplified and real-time modern BI platform, so you can get instant, deep insights on real-time streaming, and historical data at scale. Self-service visual analytics and BI at scale give you the ability to drill down to all details without needing to go back to IT for more data extracts or views. Arcadia Enterprise continuously models data based on usage for fast concurrent access. Unlike traditional BI and OLAP cube architectures, Arcadia Enterprise enables you to explore all data in modern platforms regardless of format or structure to discover information without limitations of what has been pre-modeled in a cube.

Cutting-edge visualizations

Use the Arcadia Enterprise web-based visual designer to easily create dashboards and applications without needing to write any code. You have access to a library of over 28 visual types to choose from, including advanced analytics for network graph, micro-segmentation, time-series, and correlation analyses. Arcadia Enterprise also provides pixel-aware data aggregation that results in pixel-perfect visualizations. You can connect directly to a wide variety of relational, real-time, and NoSQL data stores, including Apache Hadoop (HDFS), Amazon S3, Apache Spark results, Apache Kudu, Apache Solr, MapR-FS, and more.

You can also extend visuals from D3 and other libraries.



A wide variety of visuals including advanced analytic types provide pixel-perfect insights to your data

Modern UI based on material design principles

Material Design, which was pioneered by Google Design in 2015, provides a seamless workflow experience. Arcadia Data leveraged Google's Material Design principles to design a reactive and responsive interface that is goal-oriented and uncluttered. Instead of being distracted by a plethora of extra tabs or pop-ups, you simply navigate and edit visuals in context within Arcadia Enterprise, simplifying and streamlining the user experience; it suggests actions and next steps for the you to take, so you can focus on the task at hand.

Connect and blend data

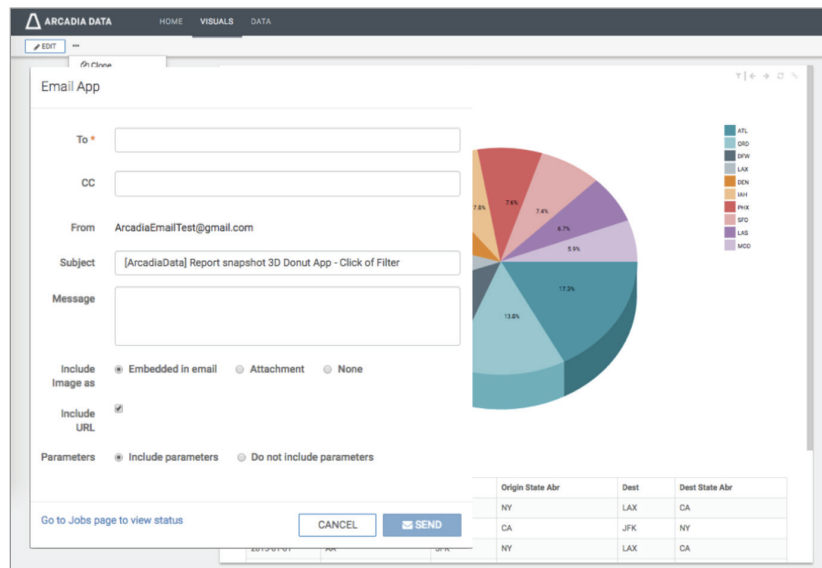
Self-service data preparation, also known as "data blending" or "data wrangling," is comprised of various steps to turn raw data into a consumable asset that can be used by AD's real-time, native visual analytics and BI platform.

Visuals are coherent and permit interaction across a myriad of data sources. These data sources run the gamut from historical resources such as relational data, delimited files, Excel documents, and business applications to more complex, real-time multi-structured data from NoSQL databases and cloud sources such as Amazon S3.

Accelerated insight-to-action

Proactive alerting and scheduling features drive next steps with dashboard alerts and email notifications based on conditional thresholds on real-time data.

Arcadia Enterprise visuals and applications are easily shared in email with links back to live data for real-time insights and collaboration.



Applications

Develop your own advanced visual analytic applications with no coding required

Build visuals, dashboards, and applications with point-and-click ease

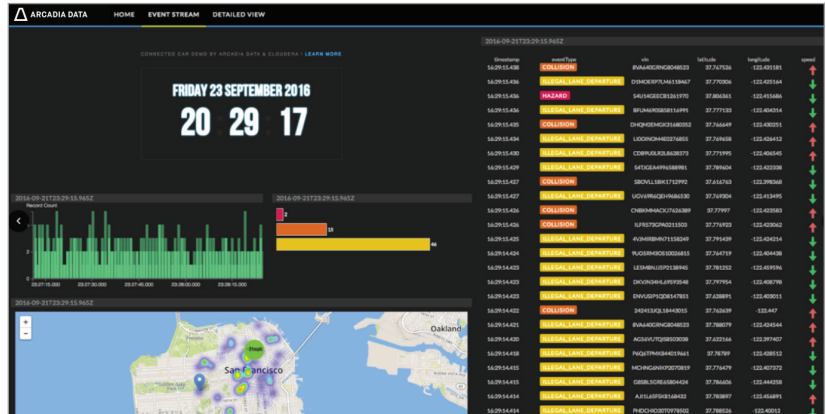
By including a multitude of visual types which include advanced analytics for network graph, geo-location analysis, behavioral analysis, time-series, and correlation analyses, Arcadia Enterprise makes it easy to build production-quality dashboards and applications. Connect directly to a wide variety relational, real-time, and NoSQL data stores including Apache Hadoop (HDFS), Amazon S3, Apache Spark, Apache Kudu, Apache Solr, MapR-FS, and more. You can also add custom extensions from D3 and other libraries.

Increase your productivity with these powerful features:

Point-and-click rapid application development

Expanding both the access and use of all data sources across the organization for historical and real-time, the Arcadia Visual Designer provides a responsive visual designer for users to easily define workflows and customize applications to meet company brand standards. Additionally, application deployment features mean analysts can quickly develop and deploy from test to production environments with simplified application migration to accelerate time to market.

Example connected vehicle application combines real-time streaming data with rich historical analysis for proactive fleet management.



Advanced analytics

Arcadia Enterprise makes it possible to go beyond simple BI and reporting and present complex relationships across time and sequence, through easy-to-create graphics that show segmentation and correlation across any dimensions and metrics. Create any number of advanced analytics visualizations, including customer engagement flows, geo-location analysis, network graph analysis, conversion funnels, behavioral analysis, and correlation analysis.

Easily embed Arcadia Enterprise into your other web-based applications

There are many types of visualizations that Arcadia Enterprise offers out of the box, with over 28 different types of visuals that include advanced analytic capabilities for micro-segmentation, time-series, and correlation analyses. With Arcadia Enterprise, you can easily design data applications to meet your brand standards. Native web technology built from the ground up using HTML5, and extensible with custom CSS & JS, means complete customization control and the ability to embed easily into other web-based applications to power data-driven services to your customers, suppliers, partners, and internal users.

Architecture

Simplified, in-cluster and in-cloud architecture runs analytics directly within Hadoop nodes, cloud instances, or other scale-out modern data platforms

There are lots of visual analytics and BI tools in the market, but only Arcadia Enterprise provides a data-native architecture to accelerate insights while greatly reducing complexity and cost. Some key aspects of being data native include:



In-cluster/cloud execution — compute where the data resides, with no data movement to speed up time to insight. Moreover, allow end users to analyze granular data directly and quickly without requiring upfront data modeling, ETL, data structuring or “cubing”.



Unifying real-time, interactive, and batch analytics across modern data platforms — use the right data and modern storage for your use-case. For example, whether it’s HDFS for historical analysis or Solr for search based near-real time indexing, you should be able to visualize this data in a single view with the ability for end users to drill down and through to more details.



Leverage existing security — inherit role-based access privileges and authentication from existing data platforms and systems for single point of security administration.



Open data access — analyze directly on open file formats such as Parquet with no vendor-proprietary data formats which reduces lock-in and the overhead of data transformation and duplication

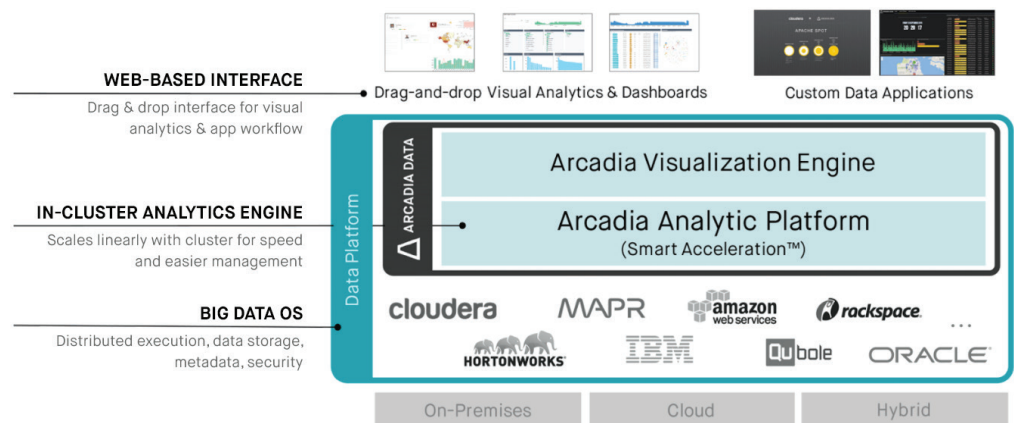


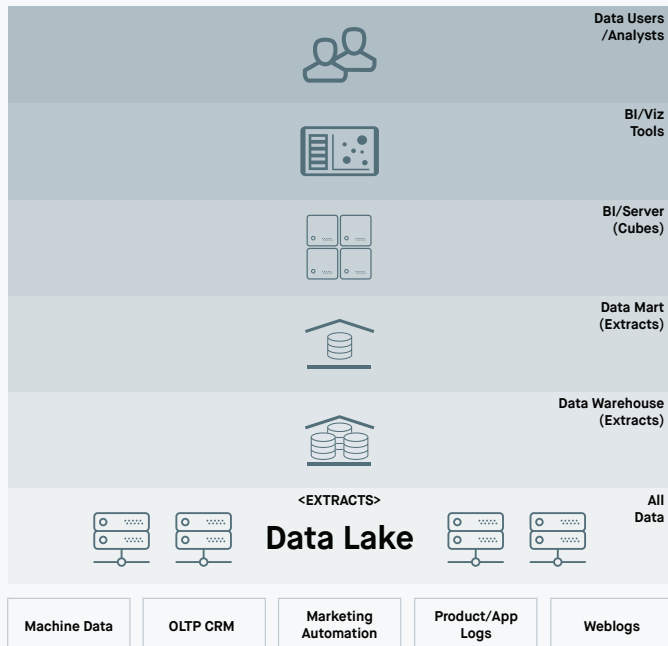
Flexible deployment architecture — deploy agnostic to cloud or on-premises, Arcadia Enterprise runs wherever you want your data to reside.

Put your analytics where your data is

Get deeper insights from modern data platforms with the simplicity and power of direct in-cluster distributed processing – without moving data. Administration and system costs are much lower because Arcadia Enterprise runs on existing hardware nodes or cloud instances and inherits native security privileges and administration in Apache Hadoop and cloud platforms.

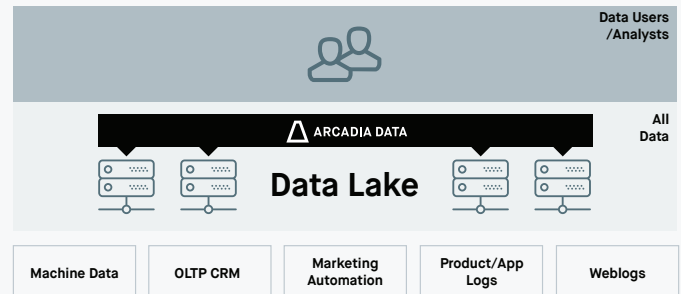
Arcadia Enterprise runs within modern data platforms to provide secure, high-performance, and scalable analytics.





Traditional data pipelines with legacy BI and analytics tools

Connect the end user directly to the data and radically reduce the complexity of traditional data pipelines for analytics.



Native Visual Analytics within Modern Data Platforms

Arcadia Data is the only native visual analytics platform for big data with a distributed execution engine

Traditional BI is complicated with stacks of technologies between data and end user. Arcadia Data eliminates the need to have a separate BI server or to build data cubes or extracts before analyzing the data and developing BI applications — you can start exploring raw data immediately. No need to submit a ticket to IT to build a cube first.

“For three years, we’ve been evaluating the market for a BI product... Arcadia Enterprise is the first product we found that provides truly on-cluster Hadoop BI... Its execution model and user self-service approach deliver performance at Hadoop scale, and lets us develop our analytics quickly.”

- Terry McFadden, Associate Director



Linear performance and scalability

User concurrency of business analysts on big data platforms is a barrier to widespread adoption. The Arcadia Smart Acceleration™ framework provides an in-cluster analytic engine that directly scales with the data platform (e.g., HDFS, the elastic tier of Amazon S3 or Google Cloud Platform) for speed and easier management. Achieve linear scalability to petabytes of data and billions of records. With Arcadia Smart Acceleration, you can support hundreds to thousands of concurrent users with sub-second response times by continuously monitoring and routing queries to optimized Analytical Views™, which are derived from data access, usage heuristics, and recommendations.

With Arcadia Smart Acceleration, the system gives you recommendations on what queries, visuals, dashboards, and applications are most important to accelerate based on actual usage. This data-driven capability analyzes visuals, recommends Analytical Views, and stores them on the data platforms (e.g., Hadoop Distributed File System (HDFS) or Amazon S3) to make performance pre-tuned for faster visualization.

Secure extranet deployments for 1000s of users

Enterprises have a growing need to share data externally with customers, suppliers, and partners. Blending principles from website design and data-centric applications, secure extranet capabilities in Arcadia Enterprise gives organizations an easy way to publish data applications externally, while securely provisioning and controlling the exact sets of data that are safely published and accessible for deeper insights.

Powerful Security with data-native, unified security with the underlying data platform

Arcadia Enterprise is deployed in leading financial services and healthcare companies with strict security requirements. Because AE runs within the data platform, it reduces security complexity by leaving data in source systems, integrating with existing user directories, and inheriting role-based access controls in the underlying data platform.



Integrated, role-based access control — Data movement outside of data platforms and databases breaks role-based access. Arcadia Enterprise delivers role-level security for mature enterprise deployments on data directly in clusters, without sacrificing granular access control. Use role definitions, such as from Apache Sentry or Apache Ranger for role-based access controls at the BI tier, without defining privileges more than once. With Arcadia Enterprise, you overcome the complexities of extraction, data movement, and fragmented permission models inherent to BI/analytics architectures that run outside Hadoop or other modern data platforms. The Arcadia Enterprise security architecture is designed to extend selective control of granular data access from the platform straight through to the UI, for business users and not just for data scientists.



Single copy of data — Multiple copies of data raises governance concerns, and are often generated in separate silos for different purposes such as data backup, disaster recovery, test, development, analysis, snapshots or migrations. And with that comes multiple security models, which means increased complexity. Arcadia Enterprise offers a single copy of data to secure, which reduces the footprint of data copies with the same or summarized information.



Single security policy definition — The Arcadia Enterprise role-based access control can automatically import group membership from underlying directory sources based on Active Directory, Kerberos, LDAP or SAML, as well as role membership and privilege information from Apache Sentry or Apache Ranger. This integration with security services centralizes policy definition and removes redundancy for administrators. Arcadia Enterprise provides full authentication (Kerberos, LDAPS/AD, PAM and SAML) and offers a single sign-on for end users



Easier compliance — The Arcadia Enterprise data-native visual analytics capabilities deliver more immediate business impact. Because Arcadia Enterprise runs entirely inside a clustered data platform, proper security and auditing of that cluster is all that's needed for HIPAA compliance.



Ease of administration — Administrators can control groups, roles, and permissions through the Arcadia Enterprise web-based GUI.



Easy software administration for users — The Arcadia Enterprise web-based architecture enables your users to perform self-service analytics on your data directly from their browser, with no individual desktop licenses required. Use drag-n-drop authoring (no coding required) to focus on BI and exploratory analytics, create a full range of charts and graphs, and edit and publish — all from the same place. No management of client software or drivers is required.

Benefit from a lower total cost of ownership (TCO)

No need for additional hardware or software — Since Arcadia Enterprise is deployed directly on data nodes, there's no need for additional hardware or cloud instances. You can leverage the data infrastructure for scale, and increase scale in tandem with Hadoop or other data platforms.

Lower upfront costs with subscription licensing — With Arcadia Enterprise, you pay as you go with a yearly subscription license, instead of a perpetual license with annual support and maintenance fees, which results in a much larger upfront cost.

Fewer licensing components and no per-user licensing — With the simple licensing model, you are licensed per data node, which means you can have an unlimited number of users with no financial constraint if you decide to expand users in the future. With other analytics platforms, you typically pay for two license components: a BI server and per user licenses, which puts a financial constraint on your ability to increase the number of users.

Arcadia Enterprise — Built for Scale, Speed, and Visual Analytics

Arcadia Enterprise is the de facto standard for unifying visual analytics, data discovery, and business intelligence on big data. You'll benefit from an immense scale of analysis, with access to billions of records in near real-time via a simple, intuitive drag-n-drop interface. Build custom data-centric applications that can connect multiple data sources together into rich, interactive applications which can be securely shared across and outside your organization. Our unique in-cluster execution architecture gives you direct access to the data that's sitting in the nodes of Hadoop or other scale-out modern data platforms, so you can gain cutting-edge insight directly from your data. The end result?

With Arcadia Enterprise, you can now solve your complex big data problems with the scale, flexibility, performance, efficiency, and security you need to glean meaningful and real-time business insights in the era of big data.

Contact Us

Arcadia Data

999 Baker Way, Suite 120

San Mateo, CA 94404

Phone: (415) 680-3535

Email: info@arcadiadata.com

