DEVO
Data. Evolved.

WHITE PAPER

Speed, Simplicity and Scale

Devo Data Operations Platform Overview

Table of Contents

Introduction

Measuring, tracking, and analyzing the full breadth of a business's operations has never been more critical for increasing business performance, attracting and retaining customers, and staying competitive. Obtaining this level of business insight requires the collection, analysis, and visual exploration of both real-time and historical machine data from across the entire organization: business operations, sales, marketing, customer support, IT, and even partners.

Traditional open source big data frameworks have largely been focused on managing large data sets for batch processing with distributed queries. Commercial platforms analyze a more limited set of event data and have mostly been focused on IT use cases, with a heavy emphasis on real-time processing that is very resource intensive.

The digitization of nearly every aspect of the modern business has created massive amounts of highly diverse data from across the entire organization. To understand how a business operates, and to stay competitive, organizations need an integrated, scalable, and economically viable solution that can analyze massive amounts of streaming data in real-time, as well as huge amounts of historical data, with consistent and predictable performance.

This white paper explains how Devo™ meets these demanding requirements with a Data Operations Platform that is fast, massively scalable, and yet elegantly simple to manage and operate.

The Devo Approach

Devo was born of the frustration our founders felt while attempting to manage the IT and security operations of a large financial institution. The more data they added, the more time and money was spent on keeping the system up and running at the expense of actual analysis.

They also found that users of these platforms required deep technical knowledge and significant training to get any actual insights.

Devo was founded to change this broken model and to develop a new data management and analytics platform with the speed to deliver blazing fast insights, the simplicity to ensure non-technical users can uncover their own insights, and the scale to meet the data volume and query demands of the world's largest organizations.

FUNDAMENTAL COMPONENTS OF THE DEVO APPROACH INCLUDE

Index-Free Ingestion

Indexing has long been the Achilles heel of data ingestion. These massive, complex, and computationally expensive indexes are built at ingest time, and have been key to providing optimized queries. Indexes are based on data models associated with each data source. Any change to the data source requires an expensive re-indexing that can affect the availability of the platform itself and any related queries. These indexes also require large amounts of additional CPU and storage, and are often only computed for a limited set of "hot" or "warm" data because of their computational and data requirements.

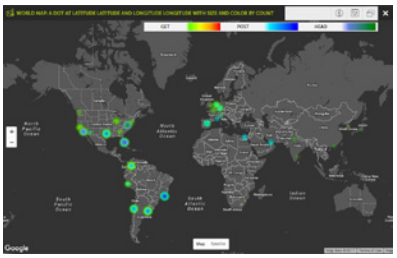
Devo takes a fundamentally different approach. As data enters the Devo platform it is classified with a small set of tags, based on the data source. Data is compressed, then stored in its unaltered, original state. No data model is applied and no complex indexes are built. All data models are applied at query time, allowing the platform to adapt to any change in data format without requiring changes to existing data or queries. Adding new data sources is simple and only requires defining a simple set of tags for parsing.

Flexible Data Models

Data models map raw data available on disk to what is available to the system for analysis, visualization, etc. Devo data models provide the ultimate flexibility in how customers want to use the underlying data in their analysis. Multiple data models can be built for a given data source,

allowing administrators to control which users have access to which data. This is especially important when data from a wide range of sources is processed and might contain personally-identifiable information (PII) or other sensitive information. Data models can also be built that integrate data from multiple data sources.

In the Devo UI, data models are represented by the various tables available to a user. As users interact with the data, remove columns, or enhance and correlate data, they are essentially creating new virtual tables represented by user-built data models.



The foundational components of Devo allow us to deliver a data management and analytics platform with the speed, simplicity, and scale required by the modern enterprise. The Devo technology stack delivers this breakthrough performance while also saving significant operational costs.

Always Hot Data

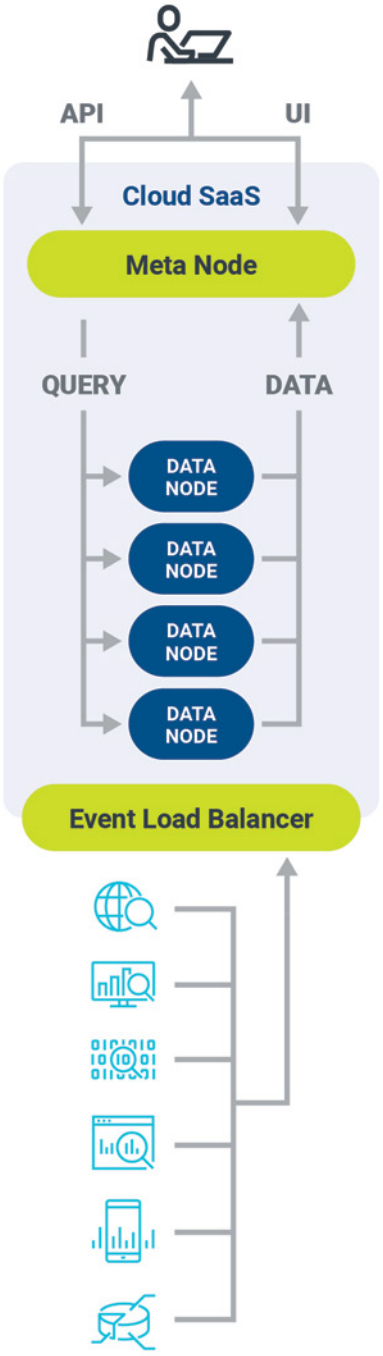
To fully leverage the benefit of IT and operational data across a number of use cases, organizations must be able to look at both real-time and historical data seamlessly. Existing solutions require organizations to designate how much hot, warm, and cold data they want to use. The colder the data is, the larger the performance penalty is paid to access it. Typically, only a few months of data is kept hot, optimizing real-time and recent data analysis, but introducing significant performance problems for historic analysis.

In Devo there is no need to designate different categories of data. All data is hot, all the time. With no reliance on expensive indexing, and utilizing industry-leading data compression, Devo is able to provide high-speed, uniform access to data, regardless of its age. Whether for forensic analysis for security, or identifying key customer usage trends, Devo provides seamless access to data of any age–one second to three years ago–with no performance penalty.

Visual Data Exploration

As the speed and volume of data continues to explode, so too does the size and diversity of people seeking to benefit from business insights. Traditional open source and commercial platforms require a high level of technical expertise to construct queries. This bottleneck significantly slows the ability of a company to leverage the full value of its data.

Users must be able to extract value from data by exploring and interacting, without requiring technical training or deep platform expertise. From its inception, Devo was built to empower users to visually explore and interact with their data. There is no need to learn complex query languages. Queries are automatically built as users visually interact with data. Exploration trees show users every aspect of their interactions, allowing them to easily retrace their steps and share the resulting queries.



Devo Architecture Overview

In this section, we review the Devo architecture, its major components, and how they work together.

Relay

A relay is software that receives, encrypts, and transmits customer data to the Devo platform. Usually, but not always, relays are deployed close to where data is generated. Data can be generated anywhere, so a relay can be on a customer premise, in a public cloud, a private cloud, a data center, or anywhere events are created. Customers can deploy as many relays as needed, wherever they need.

Event Load Balancer

Load balancers accept incoming data from Devo relays and directly from customer data sources and forward it to one or more data nodes. The load balancer's function is to ensure even distribution of data across data nodes. Larger deployments may use multiple load balancers.

Data Nodes

Data nodes are the workhorses of the platform, performing two essential services – data ingestion, and data querying. As data arrives at a data node it is classified, compressed, and stored. As query requests arrive from the meta nodes, data nodes perform the requested query and operations on the appropriate data and return the results to the meta nodes for post-processing.

A single data node can simultaneously:

- Ingest over 1.2 TB of data per day and collect up to 1 million events per second*
- Access 32 million events per second in queries*
- Run over 4,000 concurrent queries*

* These numbers scale linearly as more data nodes are added to the platform.

Meta Nodes

When a query is made via UI or API, it is translated and sent to a meta node. Meta nodes distribute the query across associated data nodes. Queries are then executed by the query engine on each data node. Responses to these queries are sent back to the meta node. The meta node combines all results from each data node into one set of results that is sent back to the UI or API.

Aggregation Engine

One of the most powerful aspects of the Devo solution is the ability to look at data in aggregate. Devo automatically turns queries from the UI into aggregations where appropriate. For example, a customer may want to aggregate login attempts or website visits into five-minute intervals, then graph and analyze trends over time. The aggregation engine tracks, manages and stores all statistics associated with aggregates to optimize and speed up query performance.

Back End Services

The back end is a set of services that provides traditional services including the web server for the GUI, access to the file system, time synchronization, and other services.

Correlation Engine

Devo not only provides interactive query, it also provides the ability to perform complex event flows and processing. Whether mixing data from multiple sources, or correlating a complex stream of events related to a particular business process, the correlation engine performs real-time queries against the data nodes. Results are typically used in conjunction with the alert engine to notify users of particular conditions.

Alert Engine

Administrators and users can define conditions and thresholds they deem important to be notified about. Alert conditions can be based on simple queries or on complex correlations. When user-defined conditions are met, the alert engine notifies the appropriate users via a variety of optional contact methods.

Conclusion

The Devo Data Operations Platform enables companies to gain insight into their digital business operations. The Devo architecture provides the speed, simplicity, and scale to empower enterprises to extend operational insights past IT and into their full digitized business.

For more information about Devo, visit **www.devo.com**

About Devo

Devo delivers real-time operational and business insights from analytics on streaming and historical data to operations, IT, security and business teams at the world's largest organizations. The Devo Data Operations Platform collects, enhances and analyzes machine, business and operational data, at scale, from across the enterprise.