glassbeam™

Glassbeam White Paper

Glassbeam Rules & Alerts

www.glassbeam.com

Contents

Introduction

The Glassbeam Big Data Analytics platform offers the uniqueness of being one of the most powerful and full-featured platforms highly optimized for machine log data for the industrial IoT market today.

As a cloud based solution, Glassbeam's core intellectual property is contained within the LCP Engine (Loader / Compiler / Parser), also known as Glassbeam SCALAR™, and Glassbeam's core invention of a new domain specific language called SPL™ (Semiotic Parsing Language). SCALAR is Glassbeam's cloud-based analytics server that consumes incoming source data and outputs parsed and indexed data from which high value analytics are derived and rules applied.

Glassbeam Rules

The specific topic of this paper is to discuss Glassbeam's 'rules-based' engine and its complex event-processing capabilities as key components for enabling machine learning delivered by the Glassbeam cloud-based, Big Data Analytics platform.

Before we jump into the Rules and Alerts (R&A) component of Glassbeam Analytics, we'll do a quick recap of the Glassbeam platform. Every analytics platform (cloud or on-premises based) begins with the need for source data input.

While many (if not most) of the available analytics platforms depend on structured data sources, Glassbeam's capability for processing a variety of incoming source data formats is unrivaled in the industry. These formats include:

- Structured and semi structured text files
- Binary files when provided a file decoder or details for stripping file header information to reveal clear text
- Binary streams, for example, message bus
- ASCII Streams, for example, syslogs and text streams

The data capture techniques for importing these source file formats into Glassbeam's cloud are accomplished through:

- On-demand bulk uploads
- Scheduled 'call-home' uploads
- Email
- Streamed data

Streamed data is inherent to Glassbeam's architecture in the sense that Glassbeam processes all data as if it were streamed, regardless of whether the source data was received as a bulk upload or as a real-time data stream.

The entire Glassbeam system platform is built on the premise that data is processed as soon as it becomes available – not processed as batches. Acting upon data as soon as it becomes available is a critical requirement for many Use Cases, including the near real-time and real-time analytics demanded by IoT – the Internet of Things.

Glassbeam goes beyond traditional search methods for analytics by also approaching analytics as measures acting upon recognition of patterns associated with a set of behavioral requirements. Behavioral requirements are described as Use Cases. Use Cases define how the user and the system will interact – where the user can be a human or a machine. Glassbeam parses the incoming source file based on the patterns and measures defined for each Use Case. The results of the parsed data analytics can be displayed as parametric search results and as dashboards, graphically illustrating a Use Case to supplement the textual results, whenever a chart offers additional context.

Use Cases alone can provide a rich set of dashboards that illustrate the resulting analytics. But Glassbeam extends the value of Use Cases by applying Rules. The Rules and Alerts engine can be applied to further process the Use Case source data to extract additional elements of contextual information. The complex rules defined enable proactive action based on a specific circumstance or set of circumstances. Use Cases plus the application of the Rules and Alerts Engine advance the role of analytics into the realms of machine learning with predictive and prescriptive analytics.

The first version of Glassbeam Rules handled 'simple events' only. There was no Complex Event Processing (CEP). The newly released Glassbeam Rules & Alerts 2.0 introduces three new capabilities that offer a huge leap in both the analytical capability and the scalability of 'distributed' Big Data Analytical Computing.

Specifically, R&A 2.0 offers:

1. Complex Event Processing (CEP)
2. Finite State Machine (FSM)
3. Distributed Edge Analytics

Before we jump into the details, it's important to note that the roadmap for enhancing Glassbeam R&A capabilities has already been identified with following innovations:

1. Add 'time boundaries' to Complex Event Processing, which will lead to analyzing Hierarchical Events; a higher level in complexity compared to analyzing the flat events in version-2.
2. Enable greater comparisons with historical data to increase the accuracy of identifying trends as source input to predictable and prescriptive analytics.
3. Further enhance Glassbeam's current Machine Learning capabilities by integrating R&A with MLlib. MLlib is Apache Spark's scalable machine learning library consisting of common learning algorithms and utilities including classification, regression, clustering, collaborative filtering, dimensionality reduction, and optimization primitives

Complex Event Processing

Glassbeam's R&A 2.0 works seamlessly with SCALAR™ LCP parsed data to convey event notification messages as alerts to the User. Each Rule that is defined is evaluated in real-time in the streamed data path. Rules can be setup at the following levels:

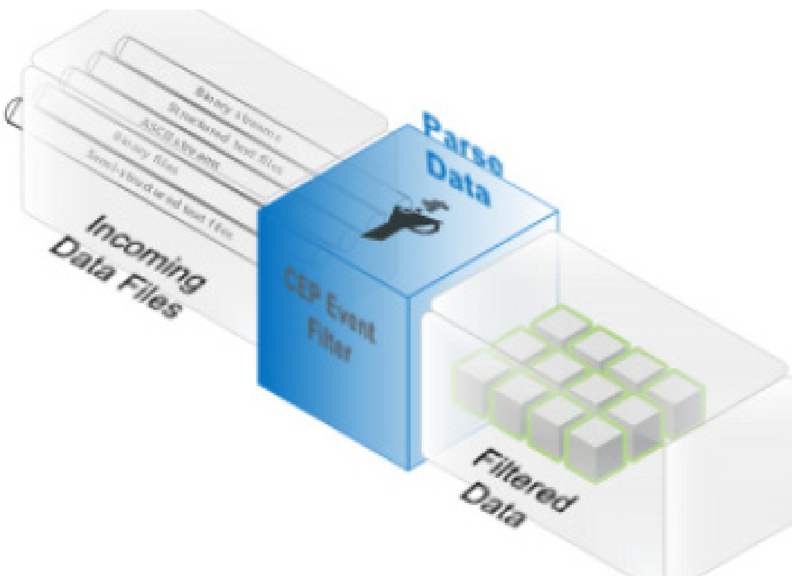
1. Event
2. File
3. Bundle
4. System

Complex Event Processing combines data from multiple sources to infer events or patterns that suggest more complicated circumstances that deserve greater inspection by the User. The goal of complex event processing is to identify meaningful events and respond to them as quickly as possible.

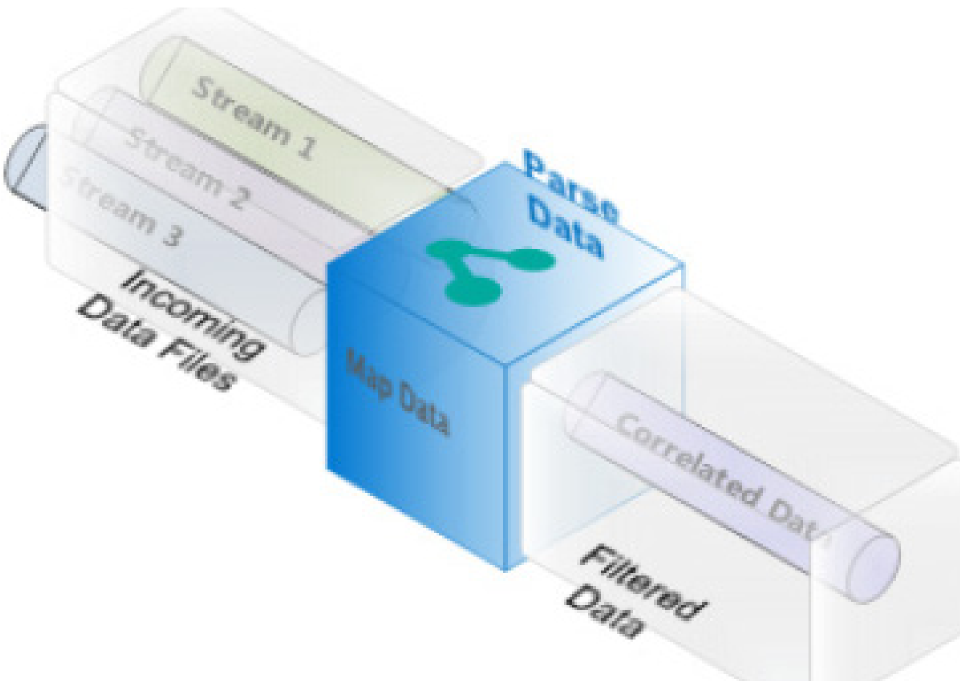
Glassbeam CEP is enabled by the set of building blocks the Glassbeam Analytics platform employs for creating the complex design patterns used to recognize event-oriented data contained within the source data supporting customer defined Use Cases. The Glassbeam platform parses the incoming source data to describe complex design patterns in a usable format. Rules can then be applied to this parsed data for creating alerts.

Glassbeam R&A 2.0 enables the following levels of CEP:

- Filtering is the simplest design pattern to apply for event processing. Filtering is easy to implement with non-CEP products and with custom-built applications. Most commercial search-oriented analytics platforms and homegrown analytics applications inside the enterprise use filtering as their core analytics capability. Even though filtering represents the simplest form of complex event processing, it creates the framework for describing other, more complex patterns for processing events.



- Correlation represents the next level in CEP design patterns. While filtering design patterns are typically applied to one stream of data at a time, correlation tackles more advanced data patterns by correlating and mapping events across multiple source data streams. CEP correlations are either CPU-intensive or, more often, memory-intensive conditionals applied against the streamed data parsed and processed by the Glassbeam analytics engine.



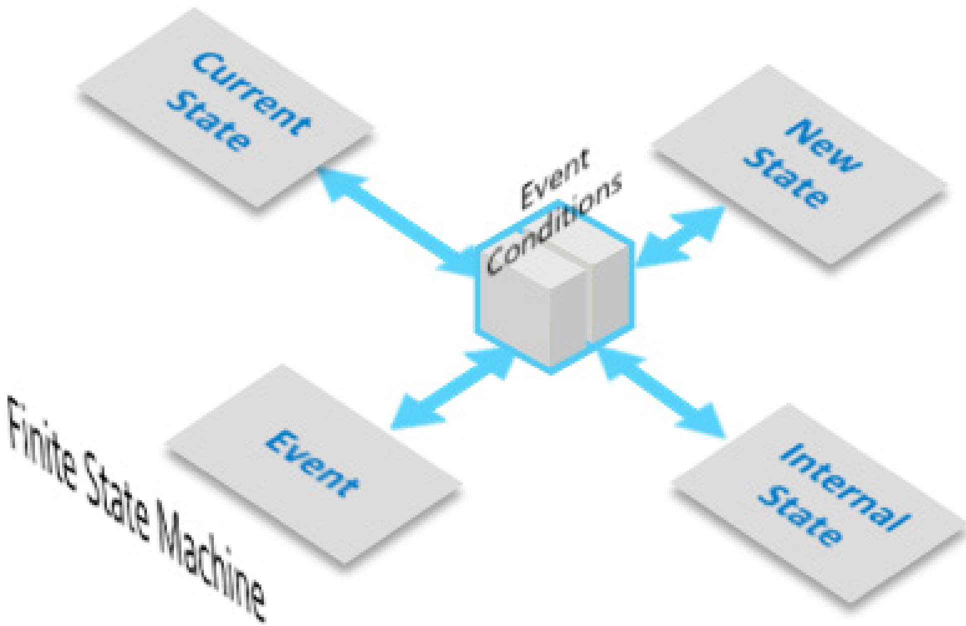
Finite State Machine

A Finite State Machine (FSM) is another powerful capability enabled by R&A 2.0. Glassbeam's FSM is used where complex behavior and process workflows need to be modeled and tracked.

Glassbeam's finite state machine defines a set of internal states for a process, together with the event conditions that define transitions from one state to another.

The new state depends on the old state and the input or event that prompted a change in state. Glassbeam's FSM design includes the following capabilities:

- Metadata: stored in a database and also in memory,
- Rules: kept in memory and have access to metadata in memory,
- Multiple processes and instantiations of Rules that can be tracked, and
- Capability to cope with imperfections and recover from failures.

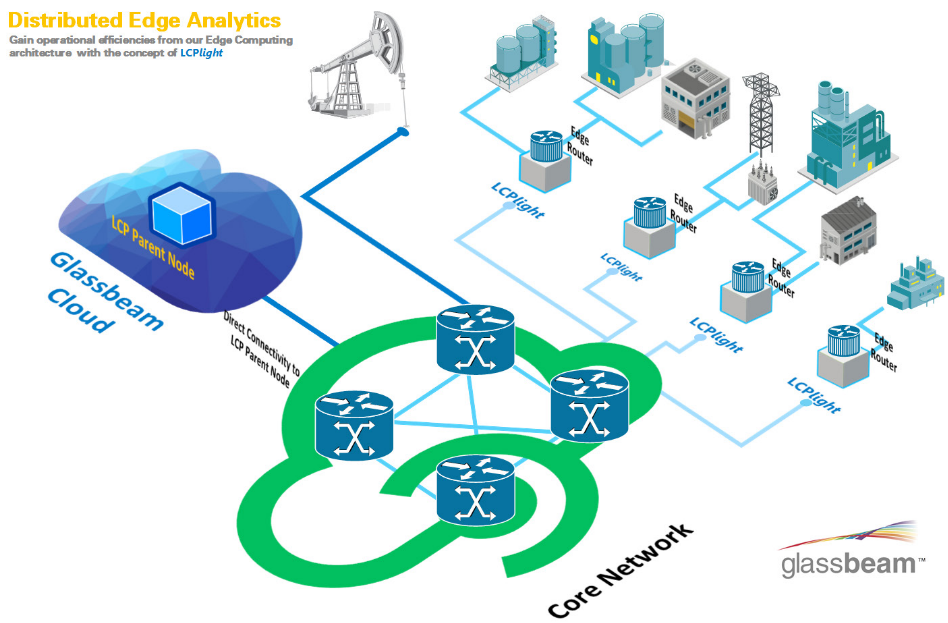


Distributed Edge Analytics

R&A 2.0 offers a huge leap in the scalability of cloud-based Big Data Analytics by extending the reach, performance, and capacity of the Glassbeam analytics platform via the introduction of a new Edge Computing Architecture and introducing the concept of a 'LCP Leaf Node' or '*LCPlight*'.

The *LCPlight* has the same functionality of the SCALAR LCP Parent Node, but in its role as a "Distributed Edge Computing Appliance", it does not connect to the Glassbeam datastores; LogVault, Cassandra, and Solr.

Serving as an "Appliance", the *LCPlight's* role is focused on computing analytics at the local customer site. It establishes an upstream message bus to pass its high quality, computed analytics results from the customer site, residing on the edge of the network, to the cloud-based SCALAR LCP Parent Node that manages the analytics datastores.



This distributed analytics capability is a result of Glassbeam's new distributed Message Bus; an intelligent transport layer that comes as part of R&A 2.0. LCP Parent Nodes, LCP Aggregator Nodes, and *LCPlights* can now communicate across the wide area network all the way to the network edge servicing the last mile of the network.

Together, with the LCP Parent Node, the *LCPlight* Appliance Node and the LCP Aggregation Node are connected as a cluster using TCP as the communications protocol and Glassbeam's new intelligent transport message bus.

This LCP cluster enables Glassbeam's analytics computing power to be distributed to the edge of the network – i.e. distributing analytics processing to the Remote Office / Business Office (ROBO). The benefits of creating a cluster of LCPs distributed to the network edge include:

- Simple Concurrency with multi-threaded message architecture glassbeam™

- Asynchronous I/O for scalable multi-core message-passing apps
- Distributed network architecture by design
- High-level abstractions like Actors and Futures.

The Glassbeam *LCPlight* can be deployed at any customer location. This virtual appliance is a single, compact system with a complete instance of Glassbeam analytics installed. The *LCPlight* does not connect to cloud data storage. It processes analytics and rules locally and then connects to a Parent LCP or to a LCP Aggregator. Parent LCPs and LCP Aggregators can connect many *LCPlights*. Each Edge LCPlight will pass its results to the next LCP higher in the network hierarchy – pass its results to the SCALAR LCP Parent node.

The *LCPlight* will parse streaming data from log files and real-time streamed data directly from local customer devices, and provide a local level of analytics and Rules. Upon completion of edge data parsing and processing, the Edge deployment will employ the new capabilities in R&A 2.0 to then push its results data to Glassbeam's SCALAR LCP Parent Node in the cloud-core.

These new capabilities include:

1. Rules to create a new event
2. Rules to aggregate an event
3. Rules to forward an event

Glassbeam's Edge deployment is resilient by design, fault tolerant, and self-healing.

The benefits of distributing the computing of analytics to the edge of the network are many. The last mile network serves end-users of all application types. The quality of the last mile network infrastructure and its proximity to the edge of the network varies greatly. Computing at the edge overcomes the physics of distance, the data pipe, shared bandwidth, host CPU contention with simultaneous users, and the number of hops to reach a parent host.

Glassbeam's Distributed Edge Computing is not unlike Map Reduce but it is geographically distributed. It takes ownership of the heavy processing required for sophisticated big data analytics, and passes its computed results on to its parent host as compared to the traditional way of passing data to the parent host for full compute processing. Glassbeam's Distributed Edge Computing optimally utilizes the available network resources and bandwidth to ensure data is processed at the right place and time with the least amount of disruption to the user.

This network resource optimization is made possible by the new message bus (mentioned earlier) between the LCP nodes that provides the intelligence to manage issues like network congestion and failures.

The largest, most geographically dispersed enterprise organizations can now benefit from shared Big Data Analytics and the inherent benefits of collaborative analytics based on the performance boost resulting from the power of combining both CEP and Edge Computed Analytics into Glassbeam's new R&A 2.0.

Summary

The Internet of Things (IoT) analytics and related applications directly benefit from Glassbeam's CEP and distributed analytics via Edge Computing. IoT enabled devices are progressively getting smaller and becoming more widely dispersed every day. Each of these IoT assets , especially complex machines like mining equipment, oil & gas rigs, generate tremendous amount of machine data that provides valuable behavioral information whether the "user" is a human or a machine.

The Glassbeam *LCPlight* narrows the geographical distance between these IoT assets and Glassbeam's Big Data Analytics Engine that's tracking, monitoring, and analyzing the state of these devices. The new edge computing enabled by R&A 2.0 plus the fact that Glassbeam's Analytics engine, inherently, processes all data as streams makes Glassbeam a solid contender for processing near real-time and real-time IoT data analytics; in particular, those IoT assets that lack any type of local storage requiring the immediate transfer of data produced.

Contact us at sales@glassbeam.com

Glassbeam, Inc.
5201 Great America Parkway, Suite 360 • Santa Clara, CA 95054
Phone: 408-740-4600 • www.glassbeam.com