First Mile Observability

An Overview

Shift Left to First Mile Observability

The purpose of this paper is to provide context around First Mile Observability and to answer as many questions as you might have about the concept. What we've found over the last thousand or so discussions with developers and observability practitioners is that First Mile Observability is self-explanatory at a high level, but as people start to think through how they might realize the many and significant benefits in their own environment, they have a lot of specific questions.

So I'll start by addressing things like:

1. What exactly is First Mile Observability
2. How it works
3. How much data can be processed
4. How fast that data can be processed
5. How much money it could save you on your Splunk bill

Then, I'll provide ways for us to continue the conversation and discuss your specific goals and objectives.

Before We Start: Solution Definition

But first, let me define the core solutions that make up First Mile Observability. We'll be referencing these solutions throughout the post.

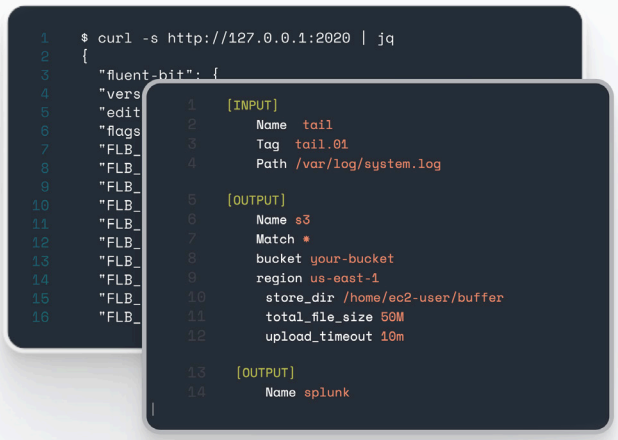
Fluent Bit and Fluentd

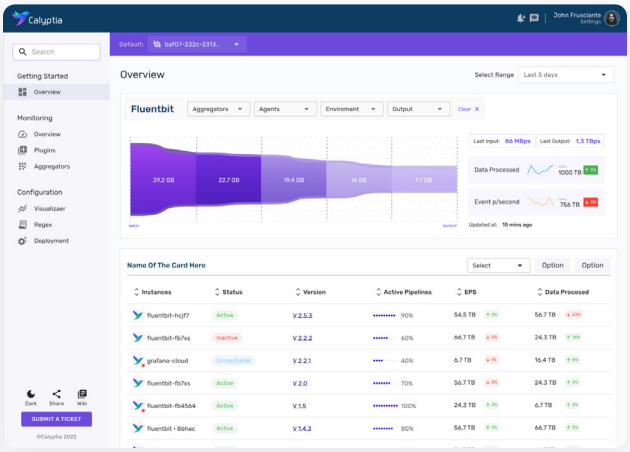
These open-source projects are super fast, lightweight, and highly scalable logging and metrics processors and forwarders. They act as agents at both the data source and destination, connecting to transport layers such as Kafka or Confluent as well as to analytic backends like Splunk, Datadog, and Elasticsearch. They also ensure that while data is collected reliably, it is also parsed, formatted, and routed as defined. The Fluent projects have been deployed over **one billion times.**

Calyptia Enterprise for Fluent Bit

This is the platform for enterprise First Mile Observability. It provides all the additional functionality and support needed to enable any enterprise to implement end-to-end First Mile Observability. It is developed and maintained by Calyptia, which was founded by the creators and maintainers of Fluent.

Bottom Line: Fluent is the engine for first mile observability and Calyptia enables the enterprise to realize all the possible benefits.





Fluent Bit, Fluentd

- Comprehensive support for **log and metrics** collection, Prometheus and OpenTelemetry compatible
- Optimized **data parsing and routing** to improve security and reduce overall cost
- **Stream processing** functionality; perform data selection and transformation via SQL queries
- Robust, lightweight, and **portable architecture** with built-in reliability and error-handling capabilities
- Pluggable and **extensible design,** to add support for inputs, filters, and outputs; more than 80 plugins available

Calyptia Enterprise

- **Integrated management** command line for complete visibility and control
- Scalability to process **petabytes of data** across thousands of servers, containers, or network devices per day
- Freedom to leverage your existing analytics tools and storage destinations with **no vendor lock-in**
- Comprehensive **multi-cloud support** to ensure extensibility and choice

What is First Mile Observability?

Observability, of course, is a well-understood concept within IT, generally meaning “to understand a system by analyzing the event data it creates.” In terms of managing this event data to optimize system performance, a pretty standard process has evolved. It starts when logs or metrics, for example, are created by each application, network component, and infrastructure device in an environment. That event data then flows through IT systems via predefined paths on data pipelines and eventually ends up in back-end observability systems. Once stored in the back-end repositories, the data is finally processed and analyzed, with reports created and alerts sent.

“First Mile” refers to that first step in this process, where and when event data is created and collected.

“First Mile Observability” then, is about developers and practitioners being able to get insights about their systems as soon as possible. The quicker an organization can diagnose, troubleshoot, and respond to any issue, the better their systems will perform – and the business as a whole will benefit.

Legacy Observability Solutions Provide Insight at the End of This Process, After Data Transport and Processing in a Data Back-end

First Mile Observability From Calyptia Shifts Insight Left to When and Where Data is Created, and Improves Back-end Analysis

What Are Some Concrete Benefits from First Mile Observability?

There are a number of significant benefits from this "shift left" to First Mile Observability, to providing insight and analysis at the moment data is created and collected. I'll group them into two categories:

1. Immediate, actionable intelligence
2. Optimized data flow and backend analysis

Immediate, Actionable Intelligence

A major advantage Calyptia Enterprise has over traditional observability systems is that it can process event data at scale—up to petabytes of data daily across thousands of servers in many of our deployments—when and where it is created, to provide an organization with immediate insight and actionable intelligence. The status quo had been that observability could occur only after data was transported to a back end and centrally analyzed.

Calyptia Enterprise: Core Use Cases

Calyptia benefits extend beyond superior data parsing and routing

From the perimeter to the core, Calyptia Enterprise drives, core DevOps, Security, and AIOps processes

Performance Monitoring

Enable real-time insight into system performance via log and metrics analysis at point of collection

Cost Reduction

Analyze streams to dedupe and remove unneeded data; reroute less critical data to lower-cost backends

Stream Processing

Select, aggregate and transform data via SQL for analysis and prediction. Create streams from query results

Data Security – Future

Enforce privacy requirements by restricting delivery of specific data types. Block identified at-risk data sources

System Migration

Orchestrate execute and monitor transfer of data during upgrade, migration or recovery; validate completeness of transfer

Data Enrichment

Append or enhance collected data with source metadata to add value during backend analysis

PII Detect and Redact

Identify personally identifiable information within collected data and remove it from data stream

Data Replay – Future

Reproduce a data stream in the exact way it would have initially been recorded; facilitates archiving in lower-cost backend

This capability is valuable in itself, as it enables an organization to respond more quickly to troubleshoot or optimize system performance, but it also opens up a whole lot of new use cases that provide even more benefit.

For example, **stream processing** enables you to select, aggregate, and transform event data via SQL for analysis and prediction. You get immediate insight into device performance and you can create new streams from the query results and route them for storage and further analysis.

You can also perform **data enrichment**, appending and enhancing collected data with source metadata that would no longer be available further down the data pipeline. You can enrich event data with context such as operating system and version, CPU usage, or user state, to add value during backend analysis.

For **data privacy**, you can perform **PII detect and redaction**, locating personally identifiable information in real-time within collected data and removing it from the data stream, enabling you to easily comply with data security and privacy requirements.

And we keep learning from our customers and expanding the solution to cover more use cases. For example, you'll soon be able to block identified at-risk data sources, and enforce data location policies by restricting delivery of certain data types and sources to defined locations.

The Calyptia Advantage

Calyptia Enterprise for Fluent Bit is purpose build for First Mile Observability in the dynamic, distributed enterprise

Legacy tools have limited scale and minimal control – and require expensive, time-consuming, data transport and centralized analysis in a back end.

Speed and Scale

Enable real-time insight into system performance via log and metrics analysis at point of collection

Control

Granular data parsing and routing. Filtering and enrichment to optimize security and reduce cost.

Efficiency

Lightweight, asynchronous design optimizes resource usage: CPU, memory, disk I/0, and network.

Extensibility

Integration with all your technology - cloud services, containers, streaming processors, and data backends.

>1pb

Data throughput across thousands of sources and destinations daily.

1b+

Sources managed by Fluent Bit from IoT Devices to Windows and Linux servers.

~450kb

Minimal footprint maximizes asset support. Zero external dependencies.

80+

Plugins for inputs, filters, analytics tools and outputs.

Fluent Bit is:

1. Fast, enabling greater than 1PB data throughput daily across thousands of sources and destinations for many of our customers
2. Flexible, allowing granular management of data parsing and routing
3. Efficient, with a less than 450kb footprint and no dependencies
4. Extensible, integrated with all your cloud native services, containers, streaming processors, data sources, and backends

Optimized Data Flow and Backend Analysis

But there's more, of course, as Calyptia Enterprise for Fluent Bit not only provides insight and control over the data that is in the flow, it also gives you control over the flow itself.

Even though First Mile Observability enables immediate insight, it does not replace traditional observability solutions. You are still going to want to transport data, store it in a back-end, and perform long-term analysis. Calyptia Enterprise works seamlessly with your existing solutions for traditional observability and actually makes them more effective.

As Fluent Bit connects both

1. The data source to a transport mechanism
2. The transport mechanism to the destination

You now have a single solution to manage and validate the entire flow.

You have end-to-end visibility of your data pipeline to ensure all data ends up where it is supposed to — and you get real time insight into the process, with built-in buffering and error-handling to handle any interrup-tions. Calyptia Enterprise can optimize the data flow, through granular parsing and routing, and can enable greater long term analysis by filtering data and adding contextual metadata.

Bottom Line: First Mile Observability not only provides value you didn't have before, in terms of improved management of data routing and immediate insight into event data, it also improves your existing systems.

Why Is First Mile Observability Possible Now?

Another question that comes up frequently is "Why are you able to do this now? It seems like an obvious problem with significant benefit, why hasn't anyone been able to do it before?"

Historically, a challenge with the analysis of event data across large, distributed, and dynamic systems

stems from the amount of data. Because the volume of data created across a typical enterprise system is now so large, the difficulty has been how to quickly and efficiently parse and analyze the data to uncover any valuable insight within all the noise.

Solutions have evolved to support this requirement, and they have added a lot of value, but due to the sheer amount of compute power required to analyze the data, they all required that data be transported and analyzed centrally within a back-end system. Little intelligence could be provided until after this expensive and time-consuming process.

One significant reason Calyptia Enterprise and Fluent Bit can enable First Mile Observability is that they can leverage the dramatic increase in compute power at the edge - to then analyze event data at the edge. The growing amount of power in devices such as servers, mobile devices, laptops, and workstations - and the skyrocketing number of those devices in an organization's environment, have led to an unprecedented amount of distributed compute power. This increase in power is then enabling more processing to be done in a distributed manner. With Calyptia Enterprise, instead of having to bring data back to storage and then processing it in a centralized manner, we can leverage distributed processing to process data where and when it is created.

Bottom Line: Calyptia leverages edge computing to enable First Mile Observability.

Why Is Calyptia Enterprise for Fluent Bit So Important Now?

First Mile Observability has become a top enterprise priority due primarily to widespread adoption of dynamic, distributed, cloud-native IT infrastructures. Moving to the cloud has a ton of benefits, but it does create numerous challenges for observability:

1. More compute in more places
2. More data coming from all that compute
3. More complexity

All of this makes it more difficult to know what you have in your environment, let alone to try and manage it.

For example, with Kubernetes, users have the ability to spin up an endless number of compute environments in a matter of seconds, meaning more services, more applications, and more data load. Additionally, IoT means more devices in more places, all creating data and needing to be monitored, maintained, and secured. This avalanche of data makes it more difficult to determine:

1. What is working and what is not
2. What is secure and what is not

Bottom Line: The amount of data is exploding, complexity is increasing, and existing systems that require centralized processing in a backend data store can't respond quickly enough to enable you to run your organization effectively. That's why First Mile Observability is now so important.

Can First Mile Observability Also Save Me Money?

Here's where we can address the "Can you help me save money on Splunk," question we are often asked. The short answer is "Yes, and in many cases the savings can be significant."

It is expensive to route, store, and analyze data in a back end like Splunk. Splunk is versatile and valuable, but becomes cost-prohibitive, especially as distributed, dynamic IT explodes the amount of data you need to store and process.

First Mile Observability can help. Many Calyptia customers quickly realize a 30-40% decrease in their spending on back-end systems like Splunk. By understanding the event data more quickly and more comprehensively with Calyptia Enterprise for Fluent Bit, you can identify what data is essential, and what is not. Non-essential or duplicative data can be routed to a lower cost back-end like S3 for long term storage and immediate cost savings. If needed, that data could be migrated to Splunk at a later time.

Calyptia Use Case

Save money on Splunk with Calyptia

Splunk is versatile and valuable, but becomes cost prohibitive, especially as distributed, dynamic IT explodes the amount of available data.

Four ways Calyptia can help you save:

Validate and filter data to reduce amount sent

Remove data with null content in required fields

Filter and exclude data with defined attributes – "trace or bug" for example

Re-route tagged data to a lower-cost backend

Tagged data can be sent to S3 for example, and archived instead of excluded

Soon, you will be able to replay data and analyze it later if needed

Aggregate logs before sending them

Send analysis of aggregate data rather than individual messages

Summary of errors or Average, Max, Min of response time over 1000 messages

Enhance data and preserve context

Search for specific content and, upon encountering, send contextual speed debugging

Ensure you're not only sending errors to Splunk

Bottom Line: Calyptia Enterprise for Fluent bit is one of those solutions that not only adds a lot of value, but can also save you a lot of money.

Our Commitment To Open Source Software and Vendor Neutrality

One more topic I wanted to cover is the benefit provided by Calyptia's commitment to open-source software and vendor neutrality.

Calyptia Enterprise for Fluent Bit is a vendor and technology neutral solution. It can leverage all major cloud platforms such as AWS, Google Cloud, and Microsoft Azure, cloud-native services including Kubernetes and Prometheus, and data backends like Splunk, New Relic, Elasticsearch, and Datadog. Unlike competitive offerings that require vendor lock-in, Calyptia customers can realize all the benefits of First Mile Observability while using whatever technology they want. You can send data from anywhere to anywhere. Calyptia Enterprise for Fluent Bit can easily be integrated into your existing environment without expensive rip and replace or retraining. There is no need for a large investment in a new technology stack and time-consuming migration to realize the benefits of First Mile Observability.

Calyptia solutions are also based on proven open source technology, deployed in some of the largest and most complex organizations and embedded into industry-standard technology like Kubernetes and OpenShift. Fluent was only the 6th project ever to graduate from the Cloud Native Computing Foundation after industry-standard technologies such as Kubernetes and Prometheus. It is supported by a community of thousands of active users and contributors and is proven in production environments, being deployed over two million times per day.

Bottom Line: When you deploy Calyptia Enterprise for Fluent Bit you are using a solution that has been deployed in production over one billion times and can work with and improve your existing IT infrastructure.

Contact Us to Discuss Further

We’d Love to Hear About Your Specific Requirements

Email

hello@calyptia.com

Slack

https://launchpass.com/fluent-all