

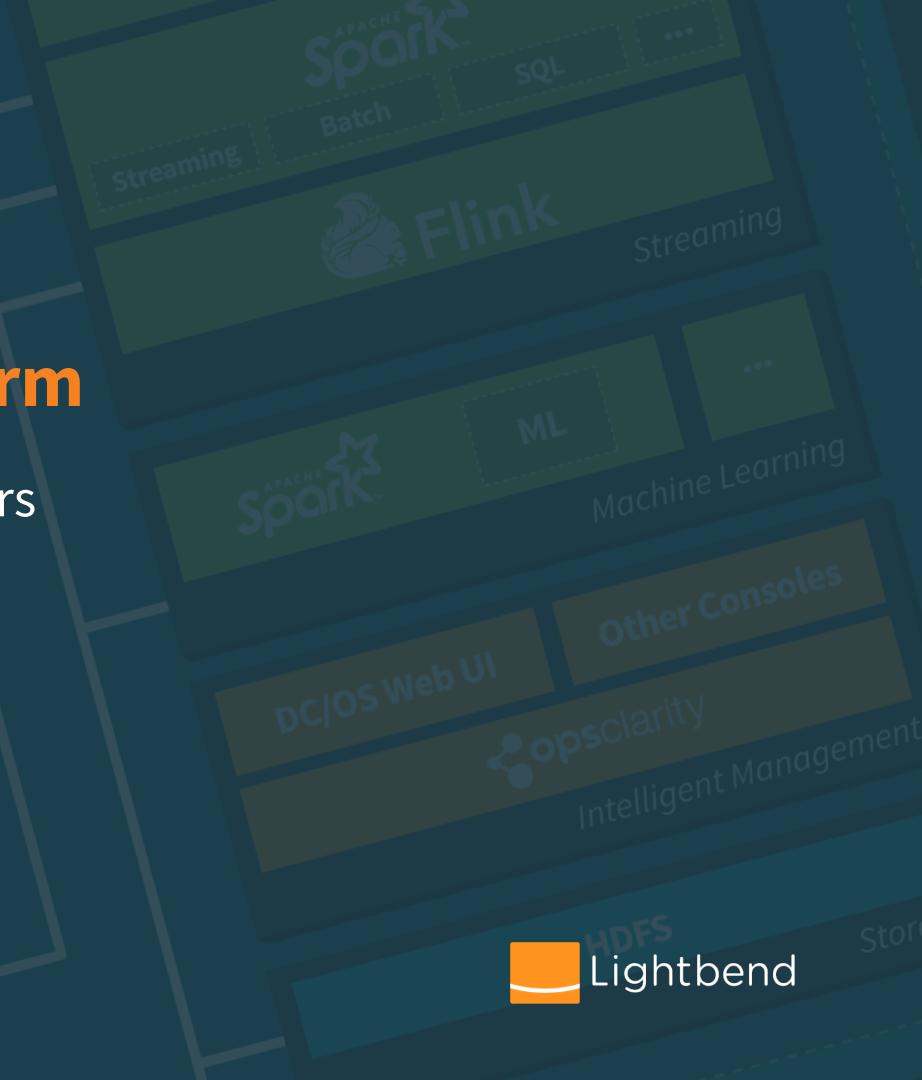
Lightbend Fast Data Platform

A Technical Overview For Decision Makers



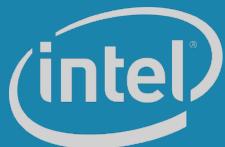
Kafka
Streams

Streaming



“Mobile and IoT use cases are driving enterprises to modernize how they process large volumes of data. Lightbend provides the fundamental building blocks for developing, deploying and managing today’s large-scale, distributed applications.”

—Intel Corporation



Executive Summary

Stream data processing is about more than just extracting information faster. It's about embracing wholesale change in how we build data-centric applications. The demands for availability, scalability, and resilience is forcing Fast Data architectures to become like microservice architectures. Conversely, successful organizations building microservices find their data needs grow with their organization. Hence, there is a unification happening between data and microservice architectures.

[Lightbend Fast Data Platform](#) is uniquely positioned to help businesses overcome these challenges because of our deep expertise in both worlds, microservices and streaming data.

Lightbend Fast Data Platform gives you everything you need to demystify the complex and continually evolving Fast Data ecosystem, so you can successfully design, build and run your Fast Data applications:

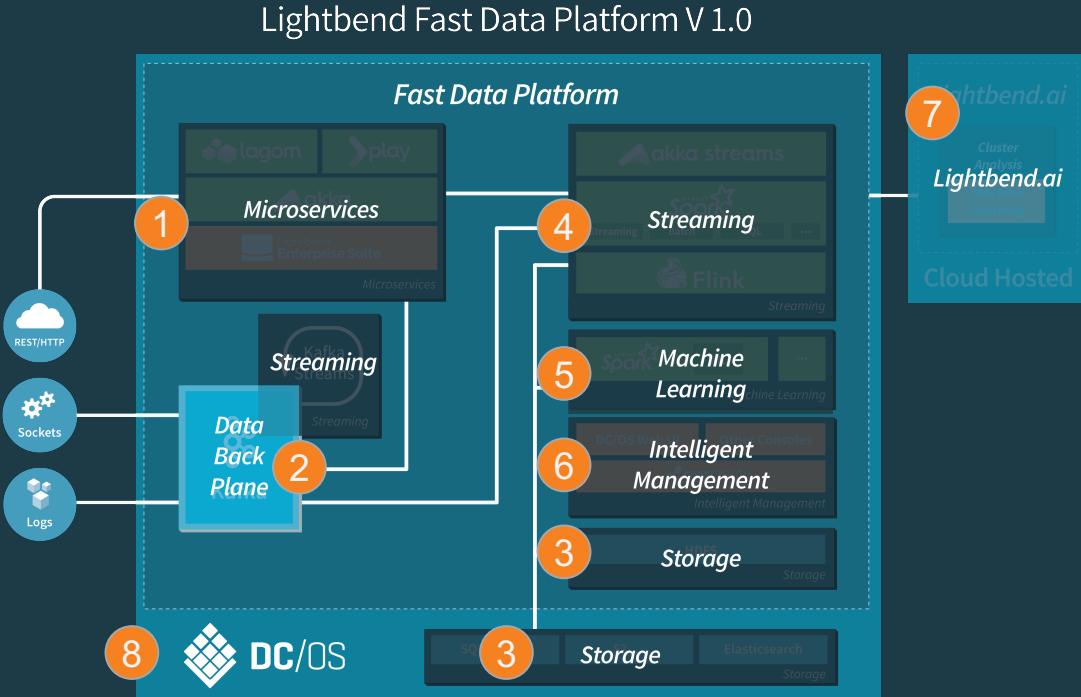
- Guided choices so you pick the streaming components best suited for your project
- An integrated platform and deep enablement to accelerate your developers' velocity
- Intelligent monitoring and management to keep your running clusters resilient, scalable and responsive, to ensure assure reliability at scale
- One-stop-support for all your Fast Data application components for complete peace of mind

This Technical Overview explores the architecture and design considerations behind Lightbend Fast Data Platform.

Lightbend Fast Data Platform

The simple way to successfully design, build
and run Fast Data applications

Sample Architecture For Streaming & Fast Data Applications



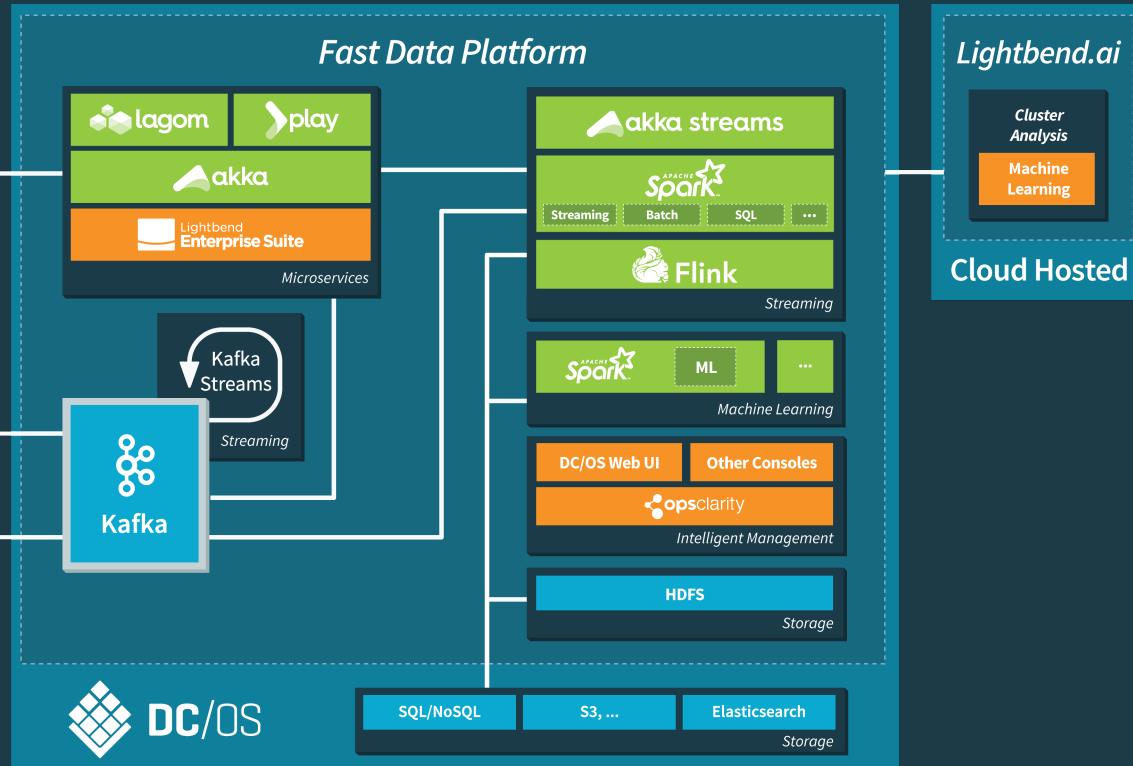
1. Microservices (Web, Mobile, IoT)
2. Durable Messaging Backplane
3. Data Persistence & Storage
4. Stream Processing
5. Machine Learning & Deep Learning
6. Intelligent Management
7. Cluster Analysis
8. Infrastructure (On-Premise, Cloud, Hybrid)

Lightbend Fast Data Platform

Meet The Components

Sample Architecture For Streaming & Fast Data Applications

Lightbend Fast Data Platform V 1.0



1

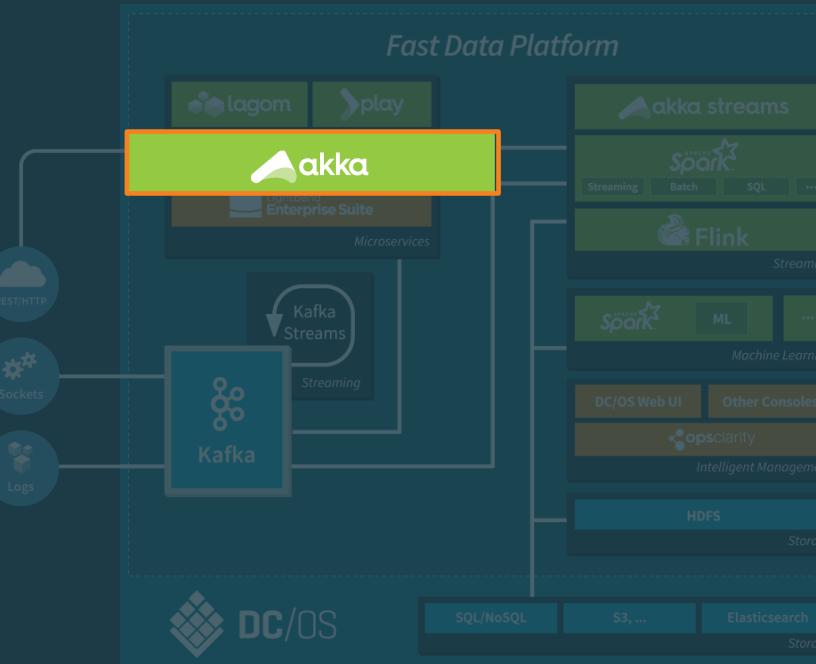
Microservices (Web, Mobile, IoT)



Akka is an advanced toolkit and message-driven runtime based on the Actor Model that helps development teams build the right foundation for successful microservices architectures and streaming data pipelines.

In Akka, the communication between services uses messaging primitives that optimize for CPU utilization, low latency, high throughput and scalability—hardened from years of enterprise production use and contributions from the open source community.

Akka's message-driven runtime drives responsiveness, resilience and elasticity, the three key principles of the Reactive Manifesto, and incorporates those into all Fast Data applications built using Lightbend Fast Data Platform.



1

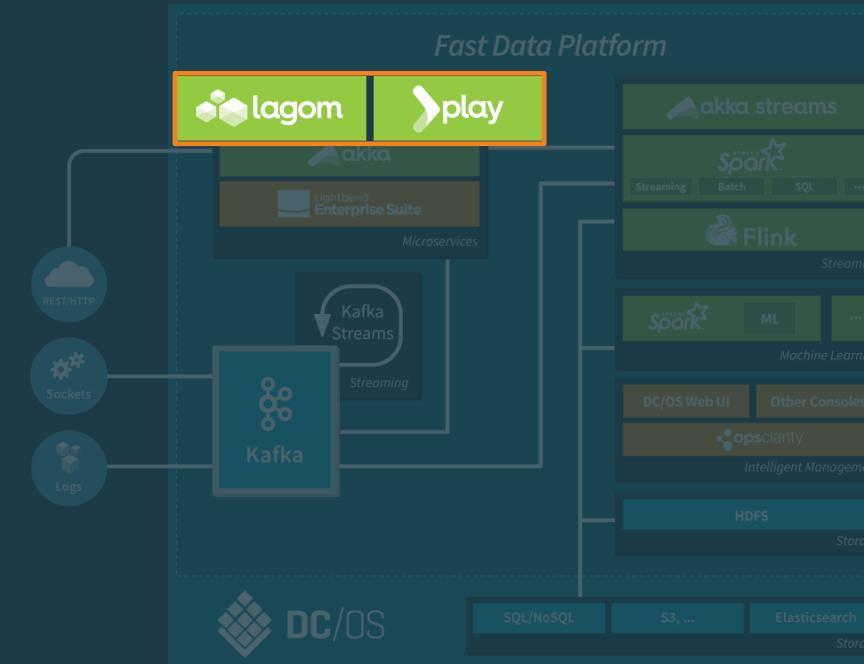
Microservices (Web, Mobile, IoT)



Lagom is an "opinionated" microservices framework that encapsulates state-of-the-art techniques for building Reactive microservices. Built using real-world experience with proven technologies (e.g. Akka, Play, Kafka, Cassandra and Lightbend Enterprise Suite) Lagom abstracts away the complexities of building, running, and managing Reactive microservices.



Play Framework is an asynchronous, non-blocking web framework that uses an underlying fork-join thread pool that enables work stealing for network operations. Built on Akka, Play provides predictable and minimal resource consumption (CPU, memory, threads) to deliver lightning-fast, highly scalable applications. For increasing developer productivity, Play features a hot-reload feature that enables any code change to be immediately visible for rapidly validating and testing applications.



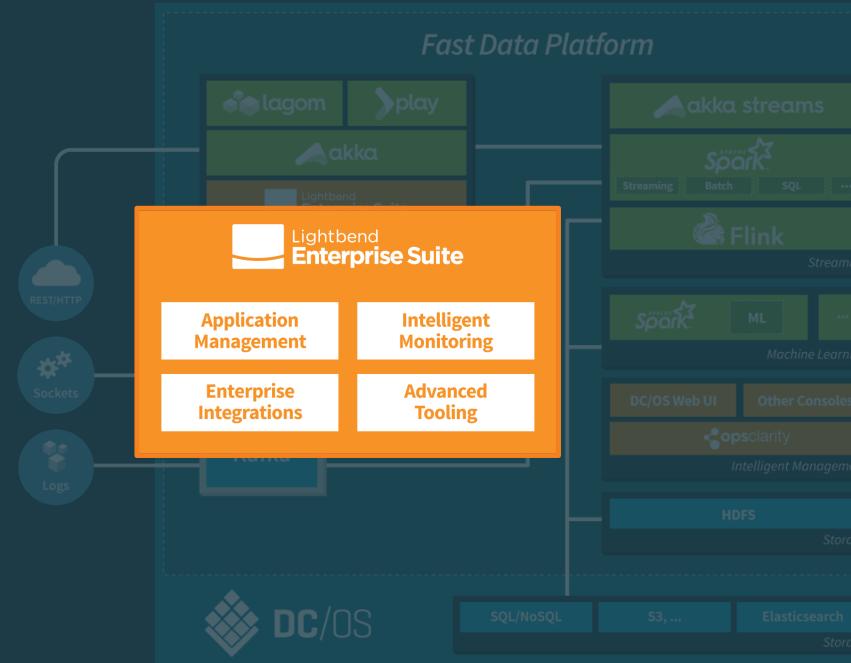
1

Microservices (Web, Mobile, IoT)



Lightbend Enterprise Suite helps businesses design, build and run streaming, microservices and Fast Data applications with the following feature categories:

- Application Management:** Safely orchestrates key deployment and management processes, helping users automate cluster management, network partition resolution, dynamic load balancing and other essential aspects of managing Reactive applications.
- Intelligent Monitoring:** Helps users assure the health, availability and performance of their applications with deep metrics and telemetry, automated discovery, configuration, topology visualization and intelligent, Machine Learning-driven anomaly detection.
- Enterprise Integrations:** Deploy to Microsoft Azure, Mesosphere DC/OS, Google Cloud Platform, AWS and others. Choose Docker containers or OCI (coming soon), and integrate with a growing number of 3rd-party tool integrations.
- Advanced Tooling:** Provides enterprises with advanced add-ons, such as automated application configuration checking, cluster diagnostics recording, thread starvation detector and a powerful developer sandbox to facilitate rapid local development.



2

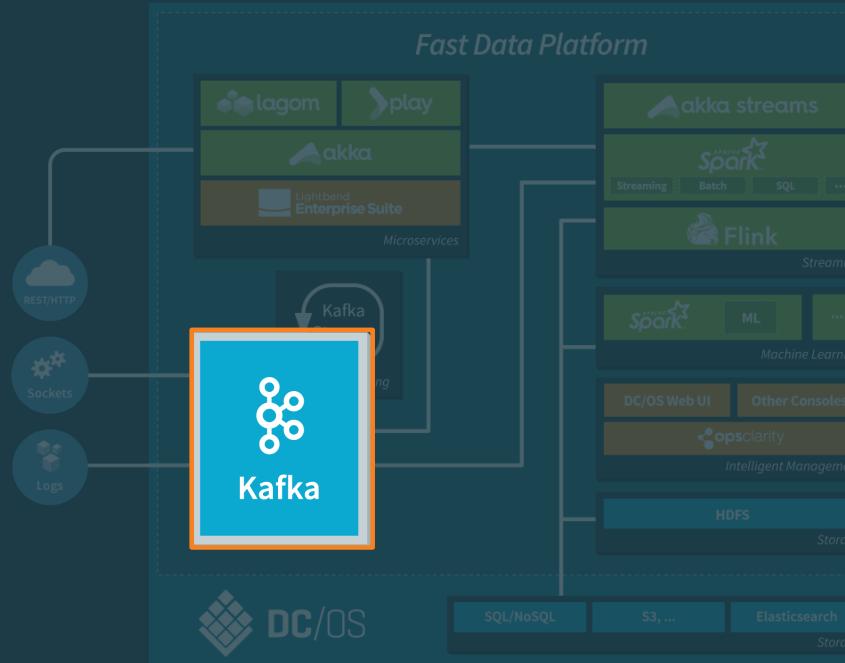
Durable Messaging Backplane



Apache Kafka provides scalable, reliable, and durable short-term storage of data, organized into topics (like traditional message queues), which can be consumed by downstream applications.

In Lightbend Fast Data Platform, Kafka functions as the messaging backplane and can easily ingest streams of data from logs and sockets from the outside (e.g. from IoT devices, social feeds, streaming application feeds, etc.).

This data can then be persisted into long-term storage using Kafka Connect and consumed by the four streaming engines, Akka Streams, Spark, Flink, and Kafka Streams, which might write the results to new Kafka topics.



3

Data Persistence & Storage

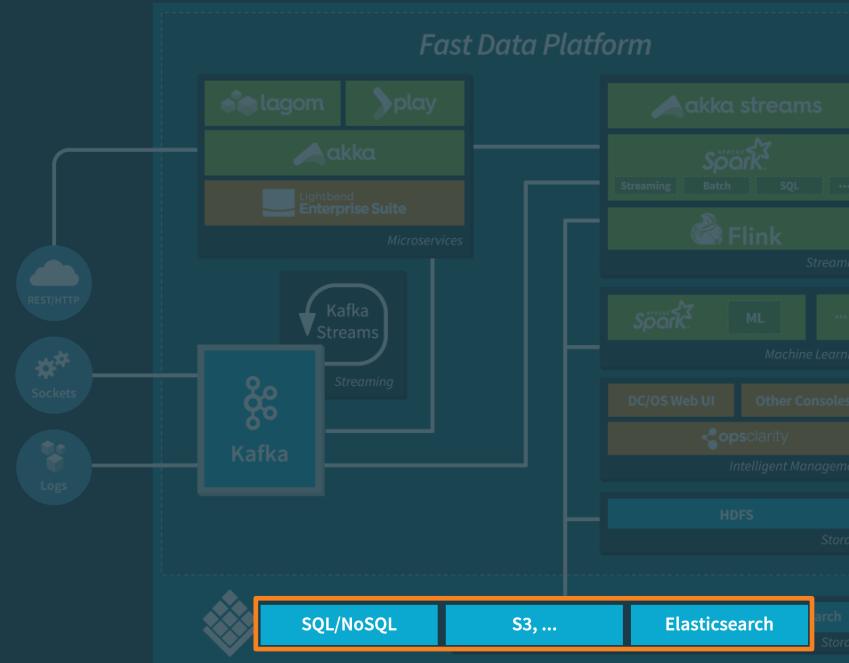


As enterprises rarely use just a single data persistence or storage technology, Fast Data Platform supports numerous persistence tier options for long-term, durable storage, depending on business requirements.

Fast Data Platform's flexible architecture allows for various SQL and NoSQL databases, distributed file systems such as Hadoop Distributed File System (HDFS), cloud-based object stores such as AWS S3, and search technologies such as Elasticsearch.

Data can be moved between storage and microservices using Akka and related Lagom and Play APIs, and between long-term storage and Kafka using Kafka Connect.

NOTE: Lightbend Fast Data Platform includes HFDS, and will be fully supported by Lightbend just like all other Fast Data Platform components.



Flexible Stream Processing For Different Needs

For most enterprises, there is rarely a “one size fits all” technology that can handle all streaming use cases. There are several considerations when choosing the right tool for the job, which is why Lightbend Fast Data Platform packages together four best-in-class streaming technologies: Akka Streams, Apache Spark Streaming, Apache Flink, and Kafka Streams.

Before diving into each technology provided inside Fast Data Platform, let’s take a look at some of the trade-offs and considerations:

What Are The SLAs For Latency And Volume?

Some tasks require a latency that is under a few milliseconds, while others can tolerate higher latencies, especially if you need to perform sophisticated or expensive computations, like training machine learning models, writing to databases, etc.

At high volumes, scalable tools with amortized, high performance-per-event processing are optimal. For complex event processing, it's possible to employ low overhead per event processing at lower volumes.

What Types Of Data Transformation?

It's important to decide the types of data processing & analytics your system requires. Will processing be on an individual, per event basis, or in bulk? Also, does the use-case need complex event processing (CEP), aggregations, or ETL?

For example, CEP is often done best with a tool that processes each event individually, whereas other kinds of data can be processed “en masse” and it's more efficient to do so.

What Level Of Data Connectivity?

Which tools and data sources/sinks does your chosen streaming tool interoperate with? This is a key consideration because nothing can live in isolation in a complex, modern system.

In the real world, data sources and sinks need to be connected with various streaming engines and the data persistence/storage layer. And microservices need to interoperate easily with analytics tools and systems.

Selecting The Right Streaming Technology

The table below outlines the characteristics of the four selected stream processing technologies: Akka Streams, Spark Streaming, Apache Flink, and Kafka Streams.

These have been chosen for their ability to meet all the needs described previously, for their relative maturity and for their active communities so users can be assured of ongoing enhancements.

	Latency	Volume/Stream	Processing Use Case
 akka streams	Low (milliseconds)	Medium	Complex Event Processing (CEP), flow graphs
	Medium (> 0.5 seconds)	High	Bulk (<i>en masse</i>) data flows, SQL, Machine Learning
	Low (milliseconds)	High	Bulk (<i>en masse</i>) data flows with sophisticated handling of late-arriving data
	Low (milliseconds)	Medium	Per event flows and aggregations

4

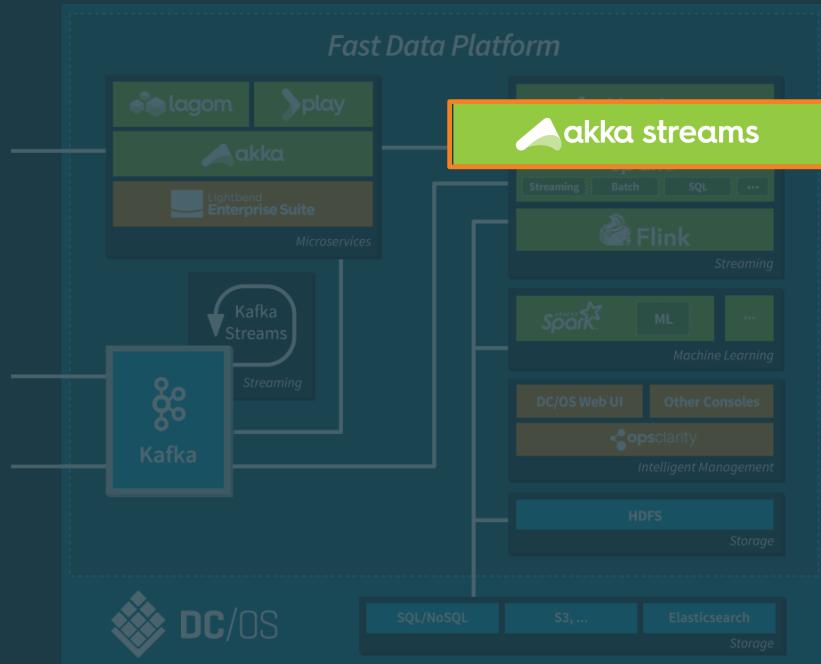
Streaming Data

akka streams

Akka Streams is a streaming dataflow abstraction on top of AkkaActors, giving developers a better way to define their workflows. As a founding implementation of the [Reactive Streams](#) specification, Akka Streams adds benefits of back-pressure and type-safety to developing streaming applications.

Akka Streams and the underlying Akka Actor model are ideal for low-latency processing of data streams, and therefore play a vital role in Lightbend Fast Data Platform.

Sample use-cases and applications where Akka Streams is critical include complex event processing, such as scoring events as anomalous, managing sessions, requesting other services, such as real-time traffic updates or location-based services, etc.



4

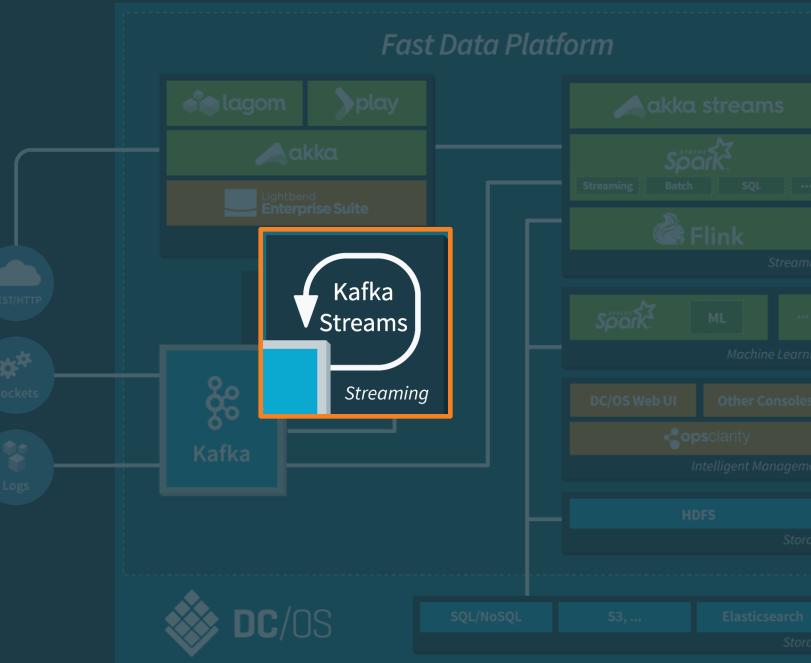
Streaming Data

kafka streams

While Apache Kafka holds data reliably, it does so temporarily, before it's written to longer-term storage or consumed by downstream processors. If some processing is required first, such as filtering, transformation, reformatting, or aggregation, then Kafka Streams is convenient for these tasks.

Compared to Akka Streams, which focuses on scalable, distributed, low-latency, complex event processing (CEP) with rich options for connectivity to other systems, Kafka Streams is better used for low-overhead processing that doesn't require the sophistication or flexibility provided by other stream processing tools.

For example, Kafka Streams is ideal for common data processing scenarios like reading records from one Kafka topic, then filtering, transforming, or aggregating them before writing new records to another topic for downstream consumption.



4

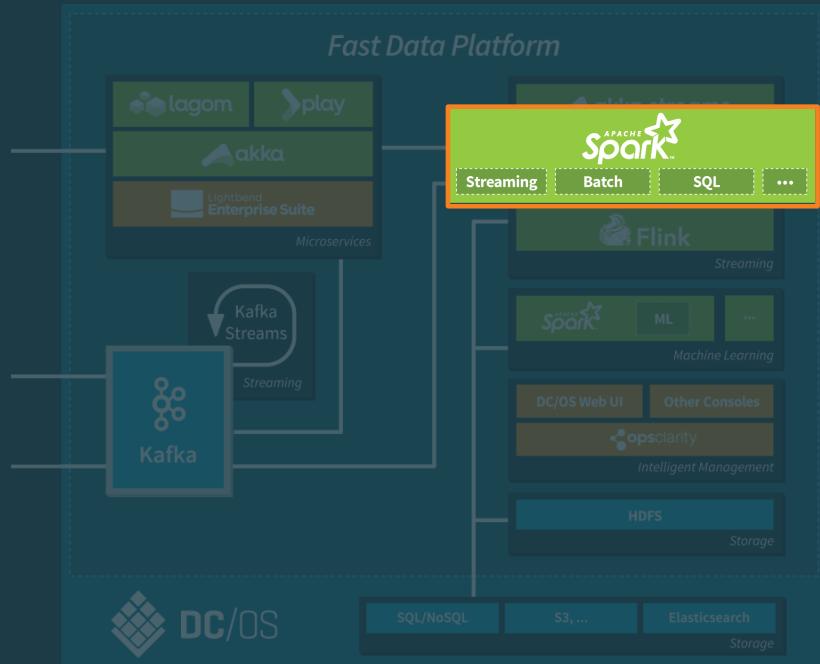
Streaming Data



Apache Spark is a widely used, highly flexible engine for batch-mode and stream data processing that is well suited for scalable performance at high volumes.

Spark's capabilities include SQL-based analytics, dataflow processing, such as for ETL (extract, transform, and load), a rich library of built-in machine learning algorithms (Spark MLlib) and third-party libraries, such as Intel's BigDL for deep learning.

Spark's streaming model is a medium latency "mini-batch" model, where data is captured in fixed time windows, then each batch is processed at once. This model works well for expensive operations, like incremental training of machine learning models and database CRUD operations, but isn't suitable for low-latency processing.



4

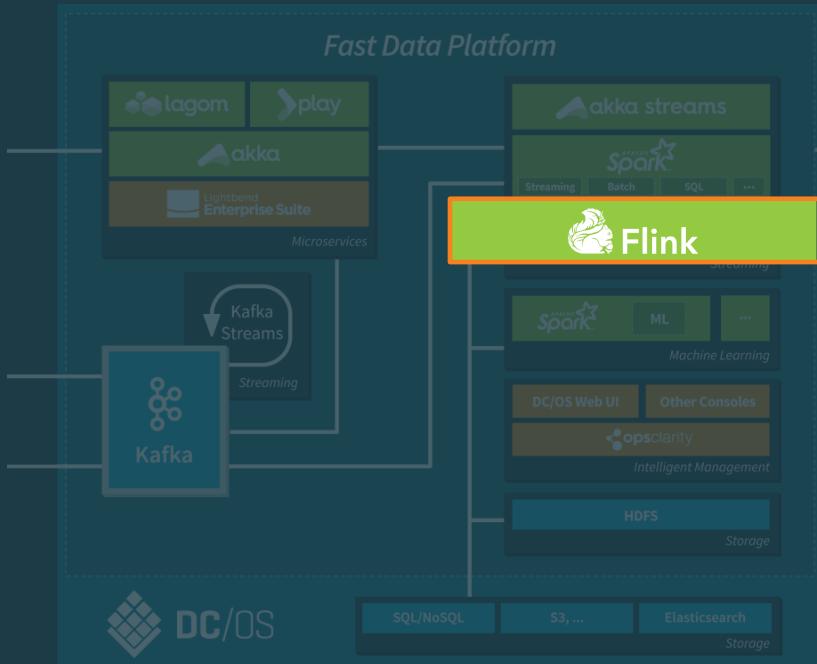
Streaming Data



Apache Flink is an open-source stream processing framework for distributed, high-performance, always-available, and accurate data streaming applications.

Apache Flink adds value in two areas that Spark (currently) can't support. First, Flink provides lower latency processing than Spark's mini-batch streaming model. Second, Flink provides state-of-the-art semantics for more sophisticated stream processing, especially when high accuracy (as opposed to approximate values) is important. Note that these semantics are defined by Apache Beam (a.k.a., Google Dataflow).

Flink, akin to Spark, also provides excellent, scalable performance at high volumes and is generally used for building analytics services.



5

Machine Learning & Deep Learning

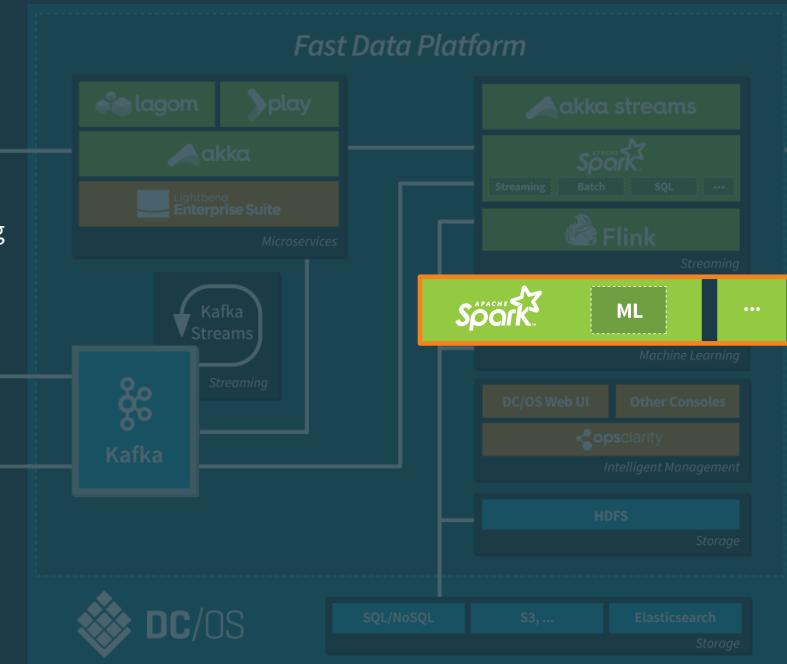


Machine Learning (ML) and its subset Deep Learning (DL) have evolved in the last decade to become an often hidden part of the everyday infrastructure. From self-driving cars to real-time credit card fraud detection, organizations are using ML and DL to build systems that can train themselves using algorithms and historical data to actively manage complex scenarios without being explicitly programmed to do so. Many providers are now embracing ML/DL to create advanced offerings that set them apart from their competitors, such as:

- Anomaly detection for discovering outliers in telemetry data (i.e. banking, home security, etc.)
- Voice interfaces that respond to commands (Apple Siri, Amazon Echo, etc.)
- Real-time image classification from camera systems
- Personalized recommendations that reflect usage patterns and interests
- Automatic tuning of an IoT environment to improve overall system robustness

Spark MLLib provides tools for common learning algorithms (classification, regression, clustering, and collaborative filtering), feature extraction (transformation, dimensionality reduction, and selection), and tools for constructing, evaluating, and tuning ML pipelines.

Fast Data Platform v1.0 will support Spark MLLib. Additional ML libraries are currently under evaluation for inclusion in Fast Data Platform.



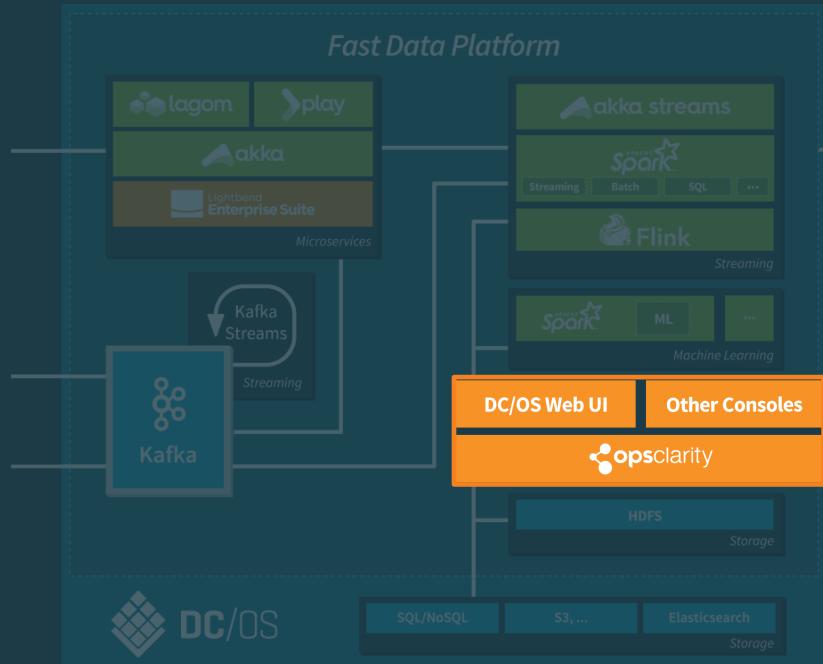
6

Intelligent Management



OpsClarity (included in Lightbend Enterprise Suite) provides unparalleled visibility into systems built with Akka, Spark, Kafka, and other components. Through the telemetry provided by other parts of the Reactive Platform, OpsClarity includes automated discovery, configuration, deep visibility into metrics, topology visualization and intelligent/data science-driven anomaly detection.

OpsClarity leverages metrics and telemetry generated by Mesosphere DC/OS and running frameworks such as Akka, Apache Spark, Apache Kafka and Apache Flink to derive meaningful features used for anomaly detection, auto-scaling and potential workload-specific misconfiguration.



7

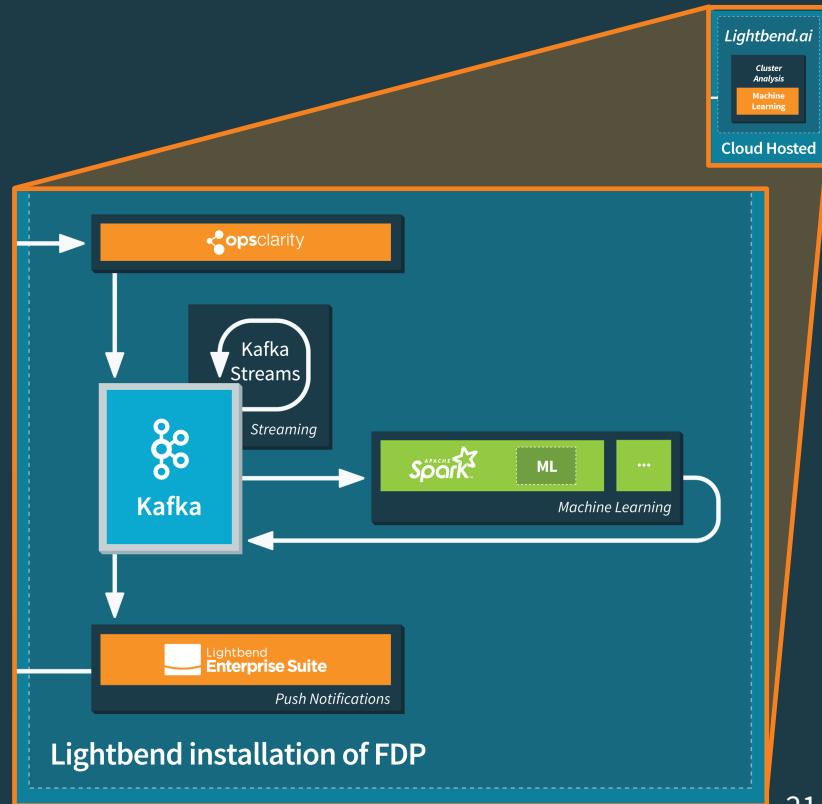
Cloud-Based Cluster Analysis - Lightbend.ai



Lightbend.ai is an intelligent, machine learning-driven cluster management system that collects telemetry and other real-time system health/performance data.

Lightbend.ai uses this telemetry and system data to train Machine Learning models for anomaly detection, predictive maintenance, auto-scaling, and other capabilities designed to keep your Fast Data Platform application clusters running reliably with minimal hands-on maintenance.

Users can experience this continually growing and improving list of capabilities without having to manually upgrade any aspect of their application cluster, since Lightbend.ai runs as a cloud-hosted service.



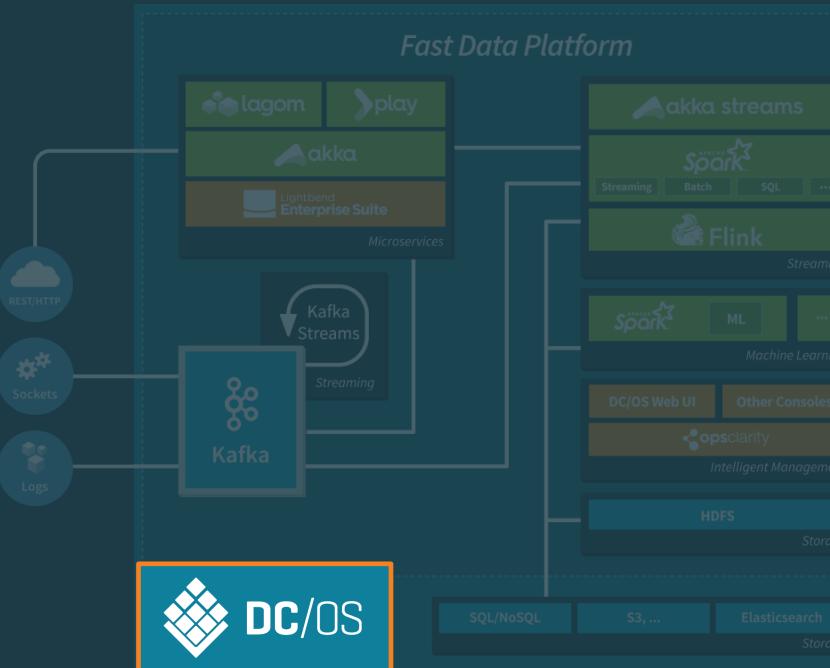
8 Infrastructure - On-Premise, Cloud, Hybrid



Mesosphere DC/OS is an enterprise ready framework designed to deliver Apache Mesos, the Marathon Scheduler and support the ecosystem of DC/OS applications. DC/OS provides infrastructure management and monitoring tools that accelerate cluster creation, evolution, and application management.

Compared to the Hadoop-YARN platform, DC/OS fixes many of the limitations of the former, making DC/OS more capable of running all the services and applications in Fast Data Platform, especially with regards to the unique requirements of stream-oriented architectures.

Through Mesosphere DC/OS, Fast Data Platform supports enterprise integration for Amazon Web Services (AWS), Microsoft Azure Container Service, and on-premise and hybrid-cloud deployments.



Lightbend Fast Data Platform

Business Benefits And Success Stories

Business Benefits

Get Rapid Time To Value

By providing an integrated, ready-to-use platform, and on-demand access to Lightbend expertise in the form of Dev Assist and Dev Support, Lightbend Fast Data Platform accelerates your Fast Data project development so your business starts to see results in just weeks, not months or years.

Improve Customer Retention And Growth

Lightbend Fast Data Platform provides an easy, guided path for your business to unlock the value of your data streams with timely, real-time insights so you can create new, innovative offerings or rapidly enhance your existing ones. This helps you delight existing customers to increase retention, and attract new ones to fuel growth.

Launch and Iterate More Often

Whether you're considering IoT, real-time decision making, real-time personalization or other Fast Data applications, Lightbend Fast Data Platform lets your team rapidly build, launch, test and iterate. In turn, this helps your business launch new offerings, enter new markets and expand across new segments more easily.

Lightbend lets us do fast prototypes and fast changes to better serve the millions of IoT devices touching our platform every day.

SAMSUNG

Started receiving new Fast Data insights for corporate customer services in just weeks after launching in production.

 **UniCredit**

Business Benefits

Reduce Cost And Wasted Resources

Lightbend Fast Data Platform is lightweight and resource-efficient, and uses intelligent auto-scaling to help you reduce cloud & infrastructure expenses when not in use. Built-in resilience and self-healing features further minimize the risks of expensive downtime and the indirect costs to your brand from any unexpected or prolonged outages.

Increase Team Productivity

Engineers and teams feel engaged and motivated with Lightbend Fast Data Platform thanks to the integrated, hassle-free development environment that makes it easy to design, build and test their applications. The included access to Lightbend expertise, best practices, tutorials, sample applications and how-tos lets them continually enhance their skills and enhance team productivity.

We've been using Lightbend technologies for over two years in production without a single crash, and we deploy 400% more frequently than before.



Lightbend is a fantastic asset, letting us make sure that Swisscom stays close to the cutting edge of technology in the Scala and Spark ecosystem.



Success Stories



Enterprise Fast Data



IoT Platform



Consumer IoT

In order to leverage the high demand on streaming engines to process immense amounts of data from a rapidly growing set of disparate data sources, Intel turned to Lightbend for their GearPump streaming platform. Intel created a feature-rich, scalable, low latency, resilient and highly available platform that doesn't trade performance for features: GearPump can handle 11 million messages per second on just four nodes.

[**READ MORE**](#)

In just 24 months, Samsung built and launched SAMI, a cloud-based IoT data broker to handle any data from any device or application. This innovation allows IoT developers to focus on building their business logic and user experience while leveraging SAMI to handle data transformation, storage, and security for millions of devices per day.

[**READ MORE**](#)

To create the next generation of cloud-based WiFi that delivers sub-second response time and always-on availability through mesh networks, eero used Lightbend technologies. Setting its sights on continuous product innovation over time, eero released over 10 software updates just three months after launching and is now a top seller on Amazon.com and available at BestBuy.

[**READ MORE**](#)

Success Stories



Consumer IoT

When obsolete software infrastructure began to threaten the company's vision, Norwegian Cruise Lines (NCL) approached Lightbend to create a new customer experience and convert its monolithic legacy architecture to a more modern development model. This new UX resulted in a YoY increase from \$650m to over \$1b in revenues in 2015, and has protected their users from experiencing any downtime across 65 deployments in 2 years.

[READ MORE](#)



Corporate Banking

Facing an inability to easily access and rapidly analyze decades of historical data, UniCredit Group started on their Fast Data project with Lightbend. Delivering a new production system in under 6 weeks, UniCredit was able to unlock the value in this massive quantity of never-before-seen data in order to understand the needs of future customers.

[READ MORE](#)



Urban Planning

Tasked with enriching Switzerland's urban planning, Swisscom created a new "Mobility Insights" platform to derive value from network intelligence and optimization using Lightbend technologies. Going from concept to commercial availability in less than nine months, Swisscom now provides 60% of Switzerland's mobile users with real-time, accurate population density mapping, city insight planning and intelligent traffic routing services.

[READ MORE](#)



Lightbend Fast Data Platform is available by subscription. Contact us to schedule a brief demo.

lightbend.com/contact

The background of the slide features a photograph of a rocky, mountainous landscape under a clear sky. A large, translucent red diagonal banner runs from the top left towards the bottom right, containing the title and author information.

O'REILLY®

Dean Wampler

O'REILLY®

Fast Data Architectures for Streaming Applications

Getting Answers Now from Data Sets that Never End

By Dean Wampler, Ph. D., VP of Fast Data Engineering

[DOWNLOAD NOW](#)

Fast Data
Architectures
for Streaming Applications
Getting Answers Now from Data Sets that Never End



Lightbend

Lightbend ([Twitter: @Lightbend](#)) provides the leading Reactive application development platform for building distributed applications. Using microservices and fast data on a message-driven runtime, enterprise applications scale effortlessly on multi-core and cloud computing architectures. Many of the most admired brands around the globe are transforming their businesses with our platform, engaging billions of users every day through software that is changing the world.

Lightbend, Inc.
625 Market Street
10th Floor
San Francisco, CA 94105
www.lightbend.com