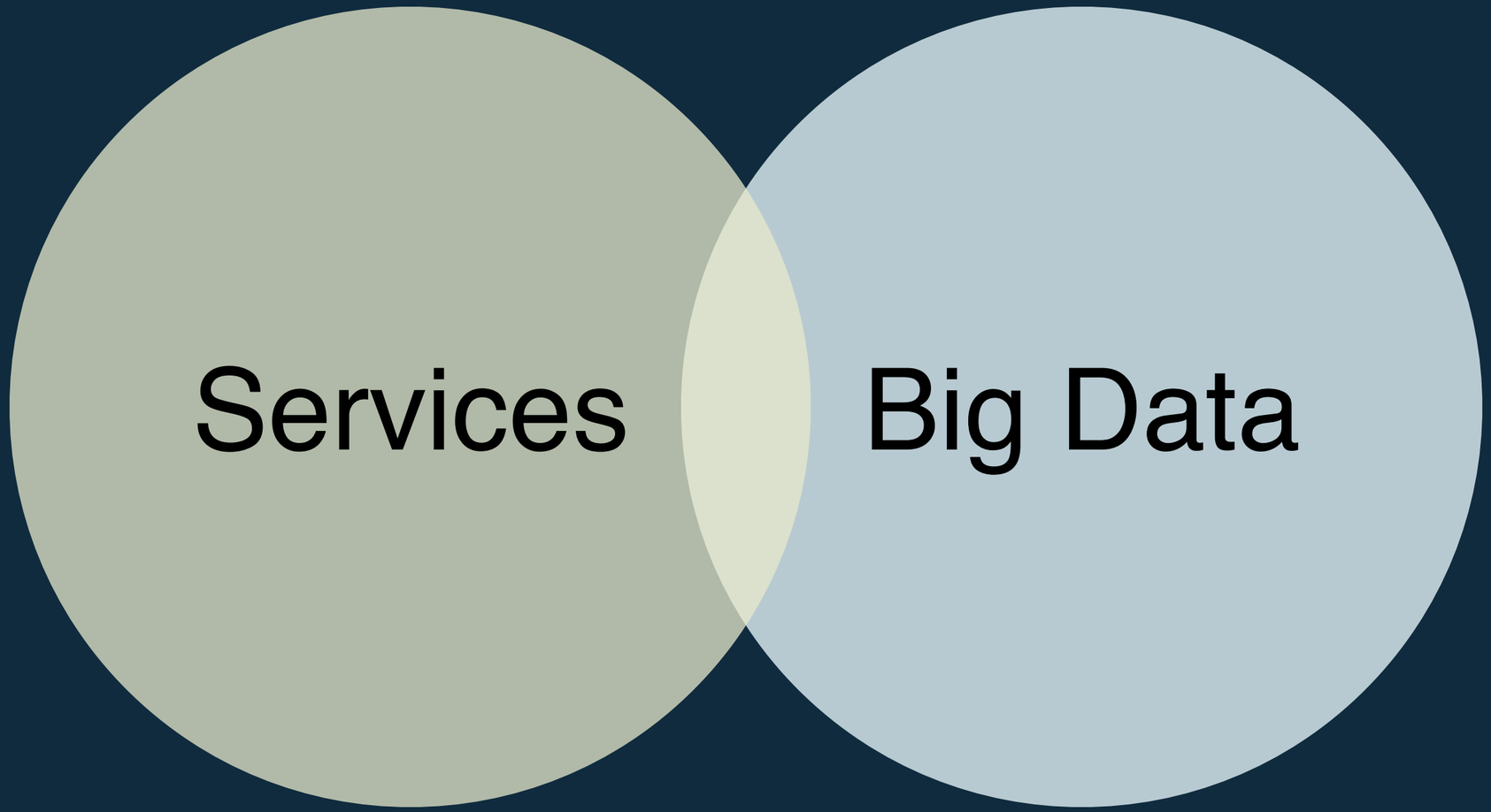
WEBINAR

Building Streaming and Fast Data Applications with Spark, Mesos, Akka, Cassandra and Kafka

Sean Glover (@seg1o), Senior Consultant at Lightbend

A Bit of History:
IT Services vs. Big Data

The Recent Past…

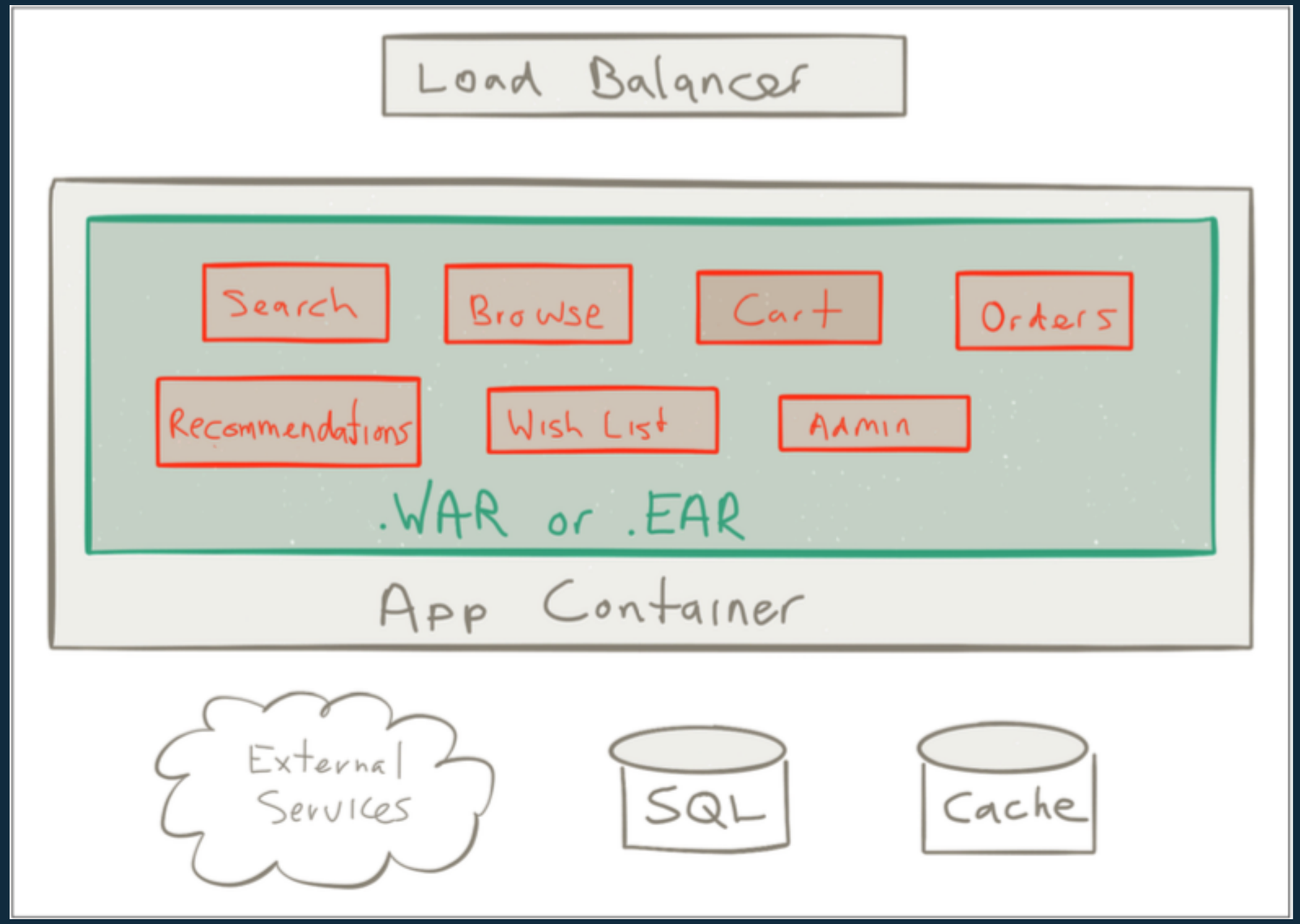


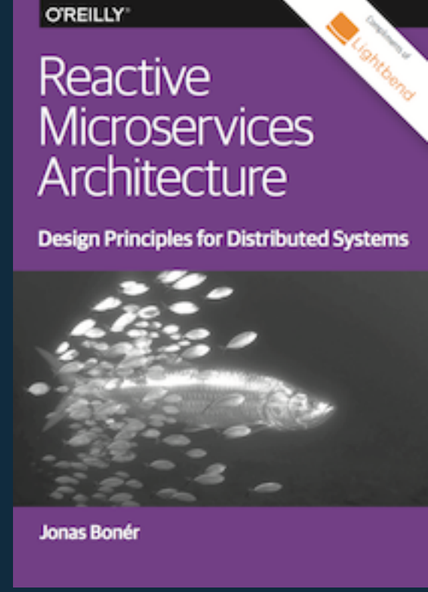
Some Overlap: Concerns, Architecture

The Future?

Much More Overlap

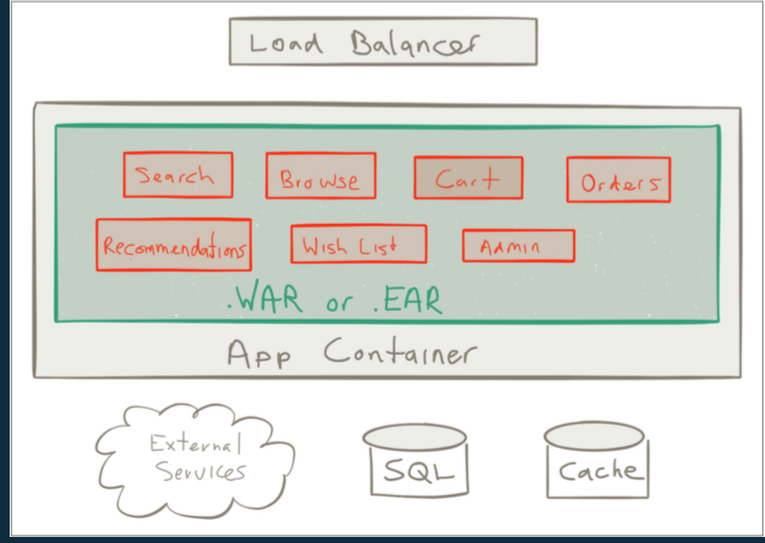
Drilling Down: Monoliths to Microservices



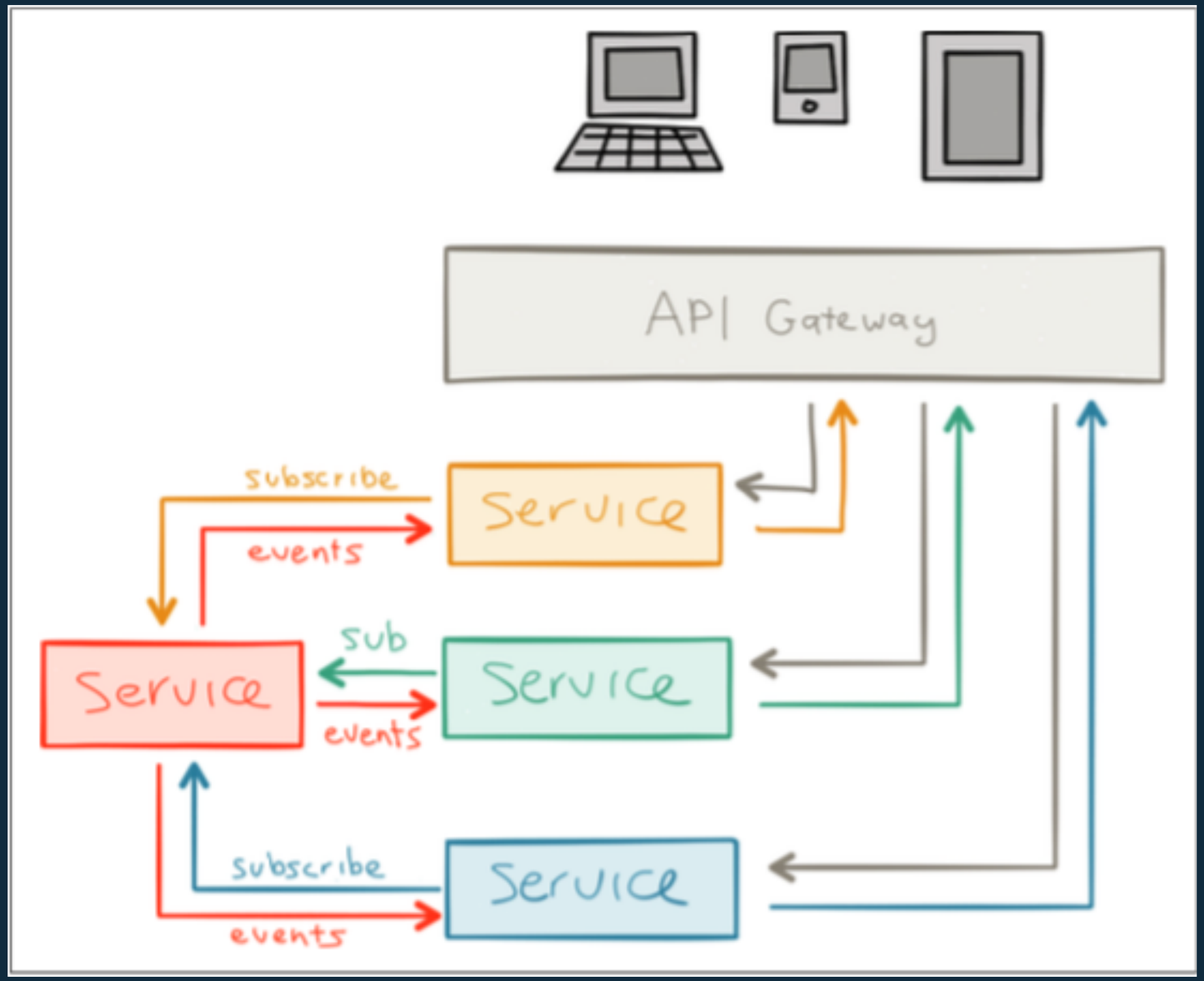


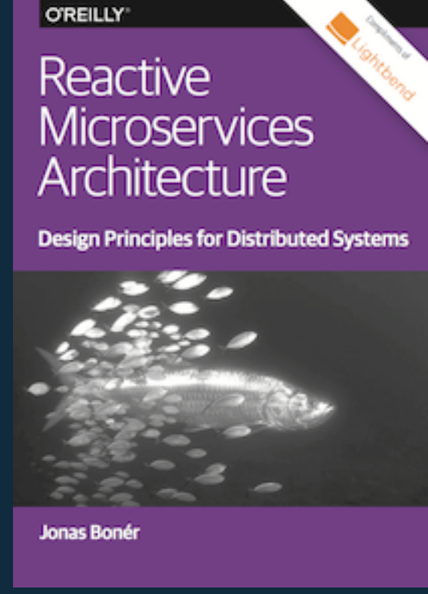
lightbend.com/reactive-microservices-architecture

Monoliths

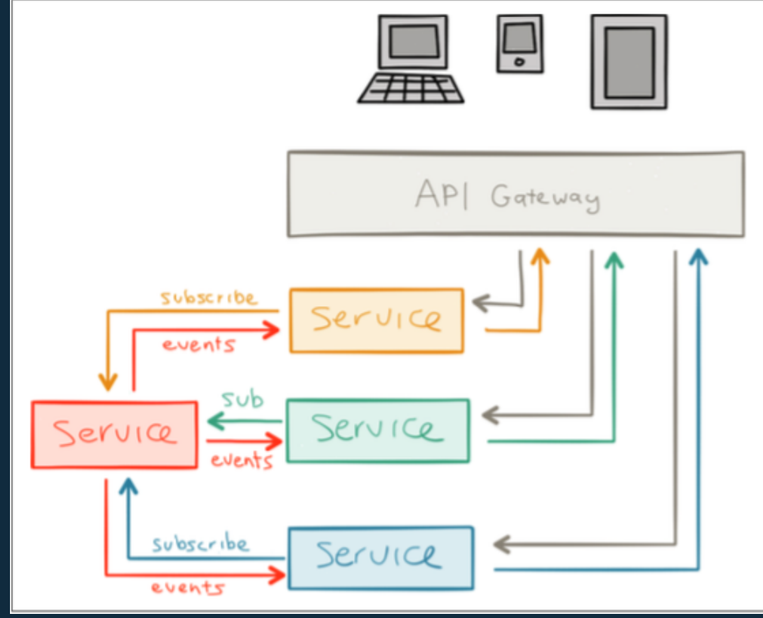


- Tangled responsibilities, lead to infrequent, “big-bang” deployments
- App lifetimes months to forever!


API Gateway
Service
subscribe
events
Service
sub
events
Service
subscribe
events
Service

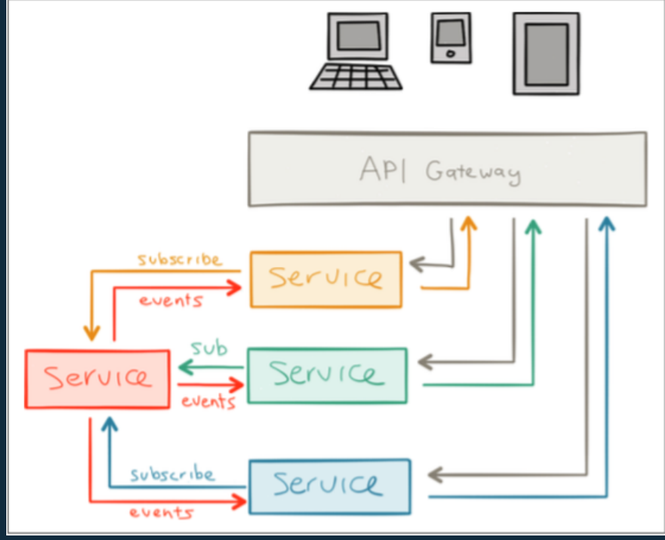

O'REILLY
Lightbend
Reactive Microservices Architecture
Design Principles for Distributed Systems
Jonas Bonér

Microservices



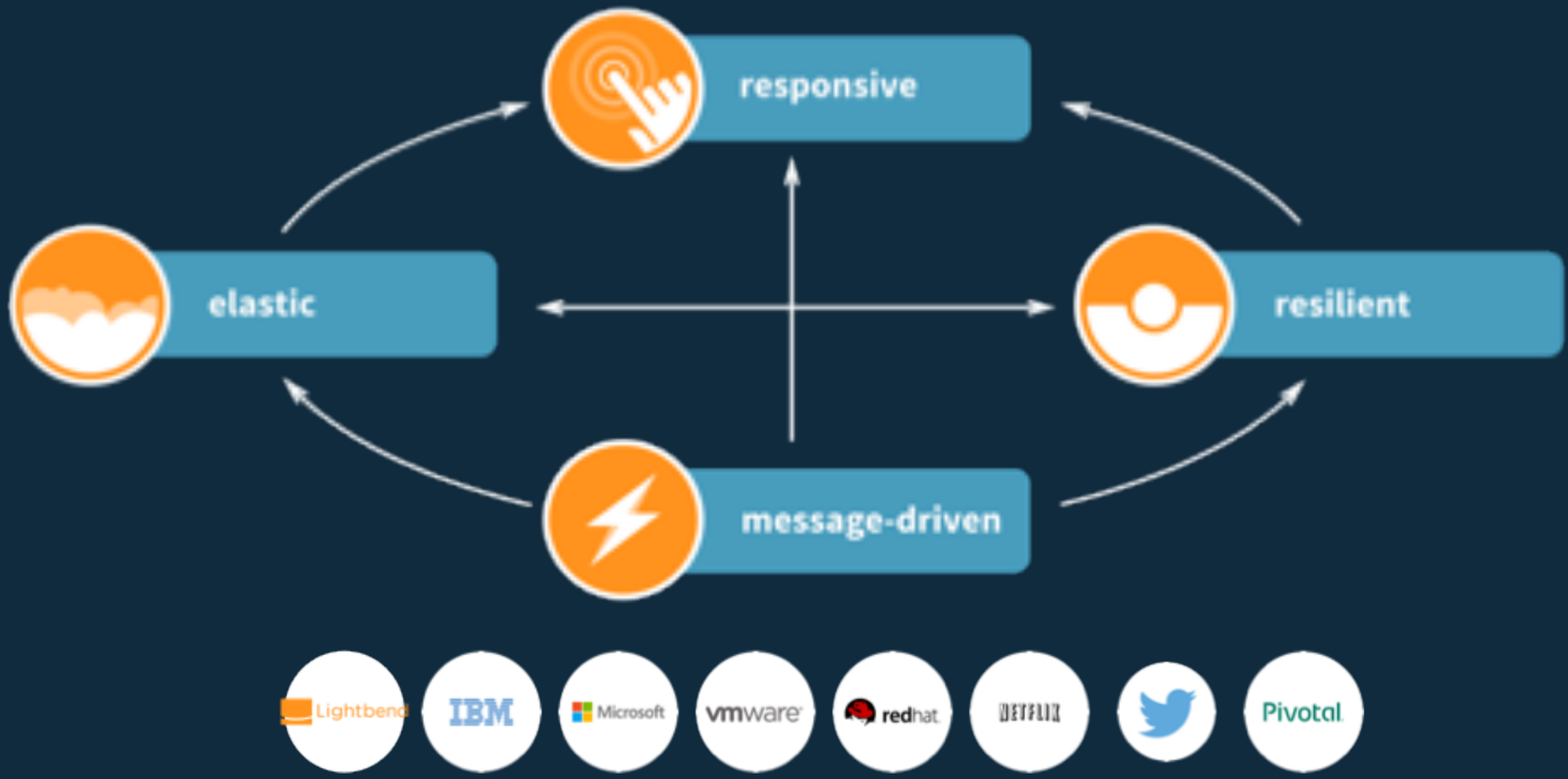
- Each does one thing, so must be message driven & asynchronous
- Updates easier, deployments frequent
- App lifetimes: minutes! to forever

Microservices



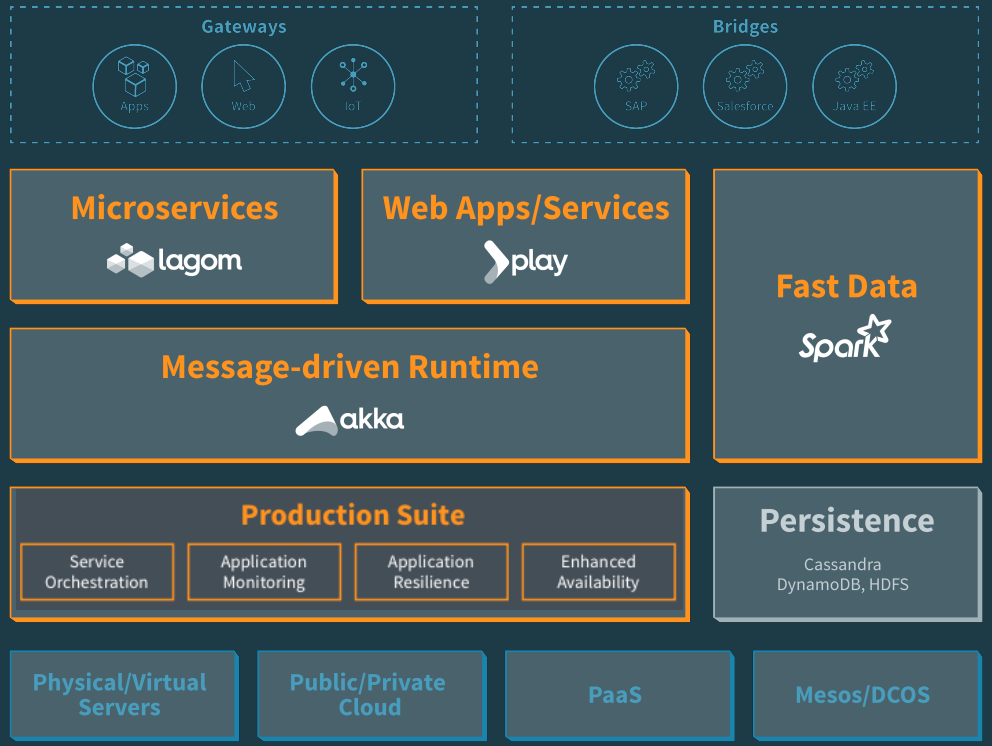
- But be careful:
  - Message overhead is much better than function calls!

Reactive Systems

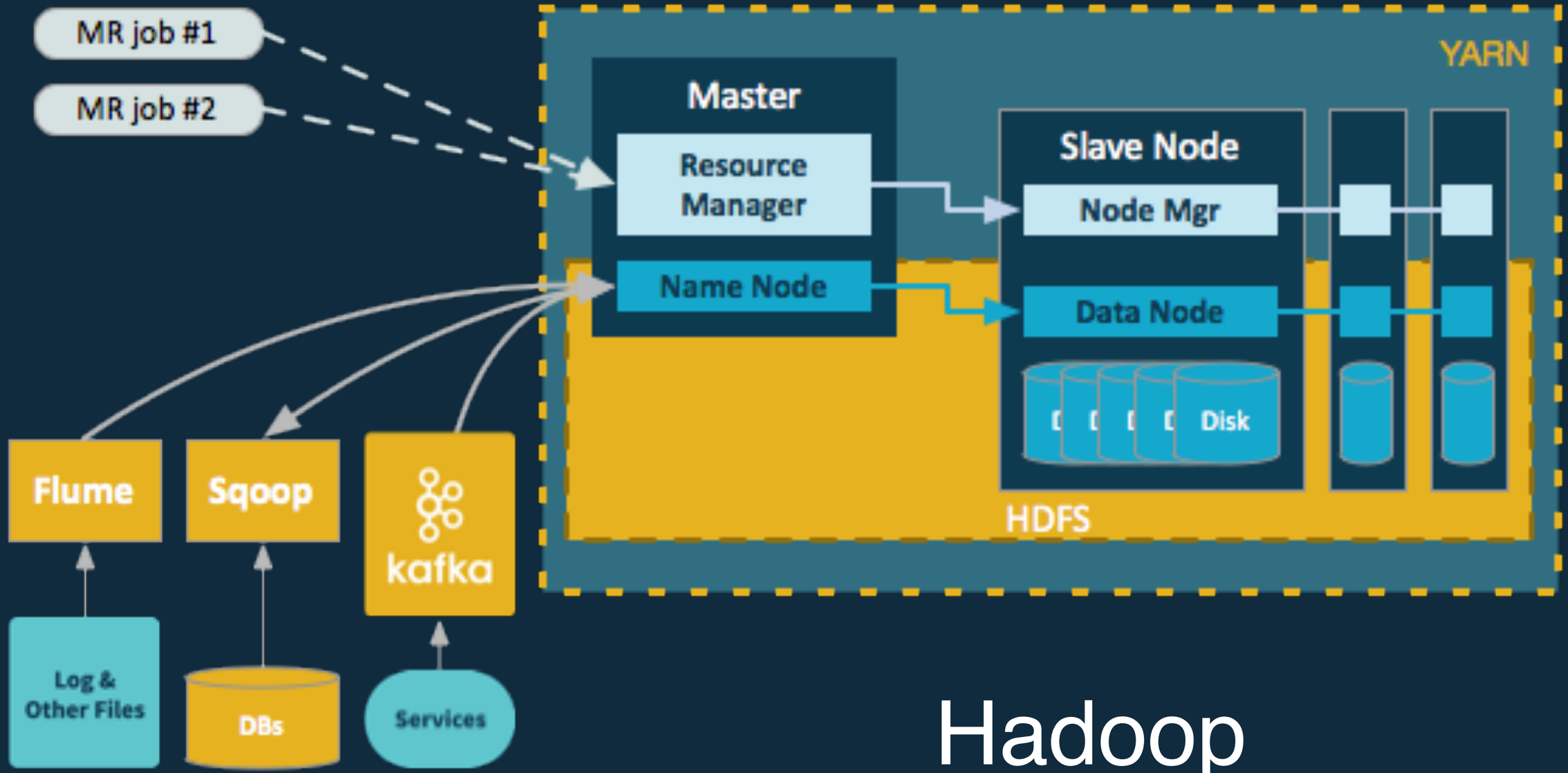


reactivemanifesto.org

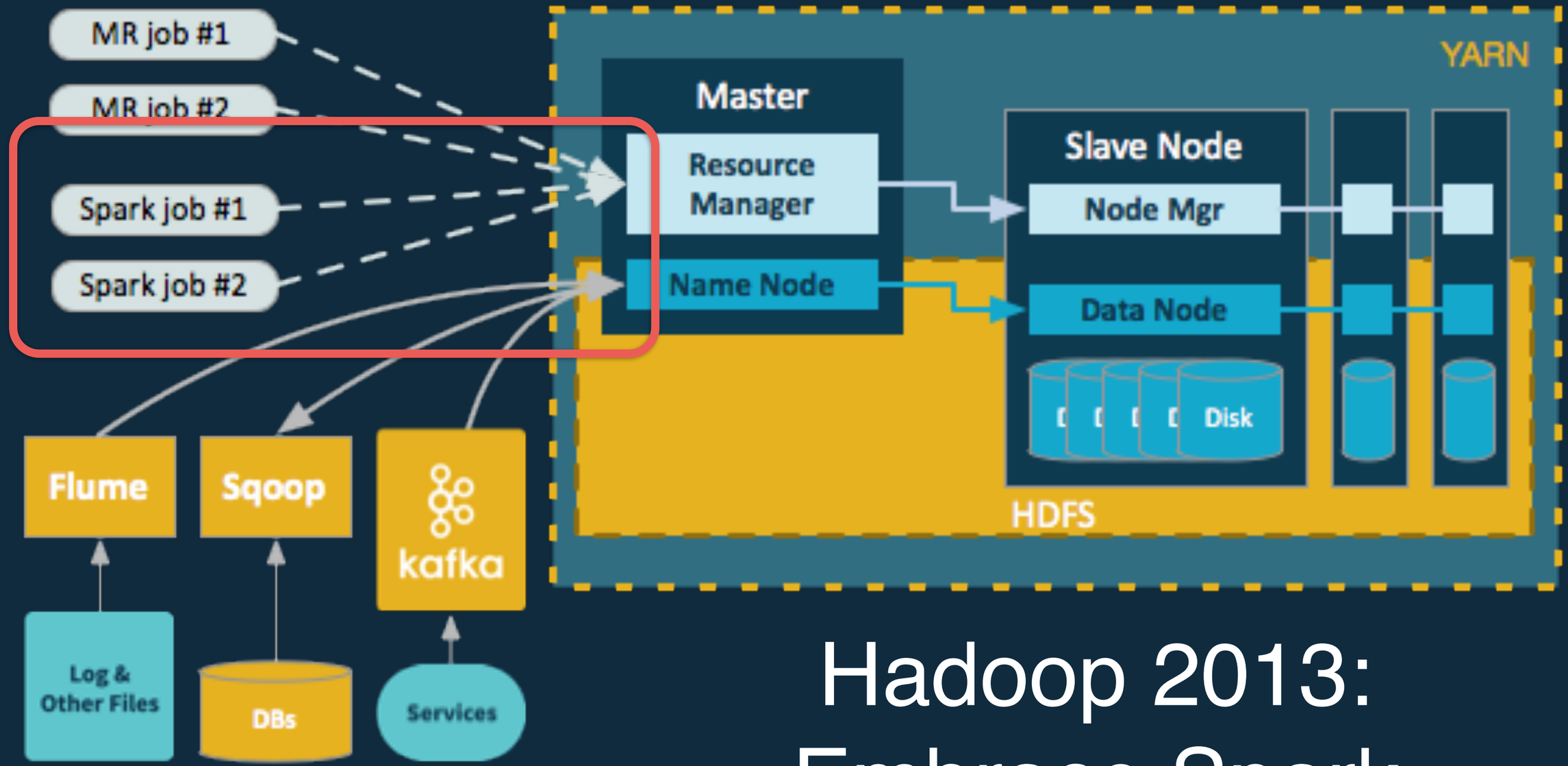
The Reactive Platform



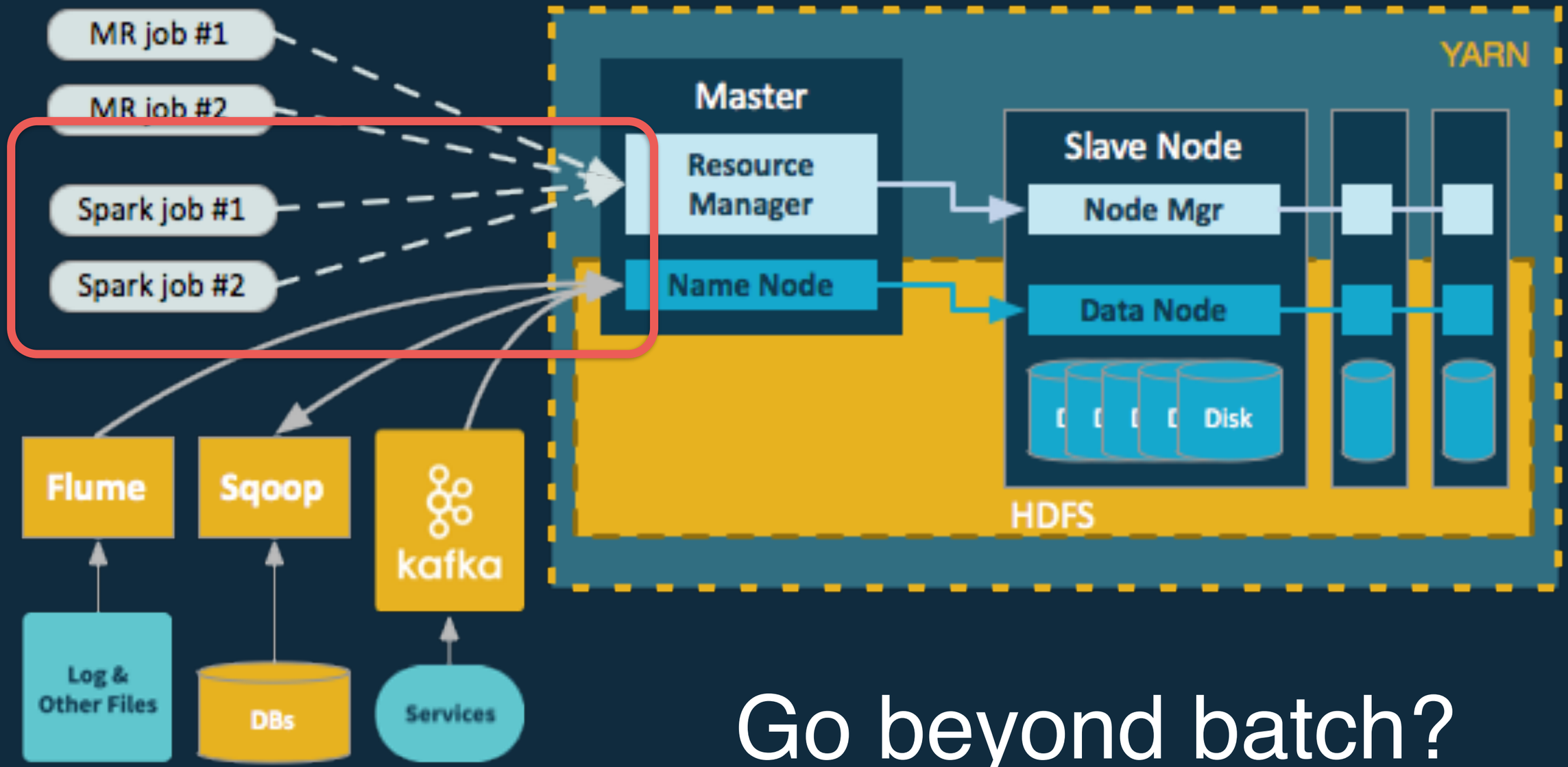
Drilling Down:
Big Data


MR job #1
MR job #2
YARN
Master
Resource Manager
Name Node
Slave Node
Node Mgr
Data Node
Disk
HDFS
Flume
Sqoop
kafka
Log & Other Files
DBs
Services

Hadoop



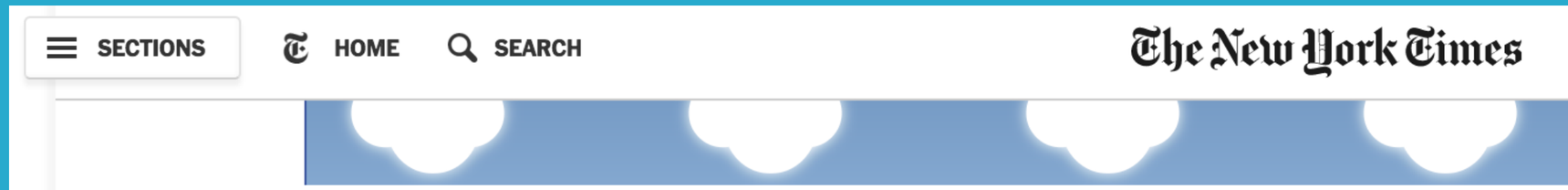
Hadoop 2013: Embrace Spark



Go beyond batch?

The Emergence of Fast Data: (Time Is Money)

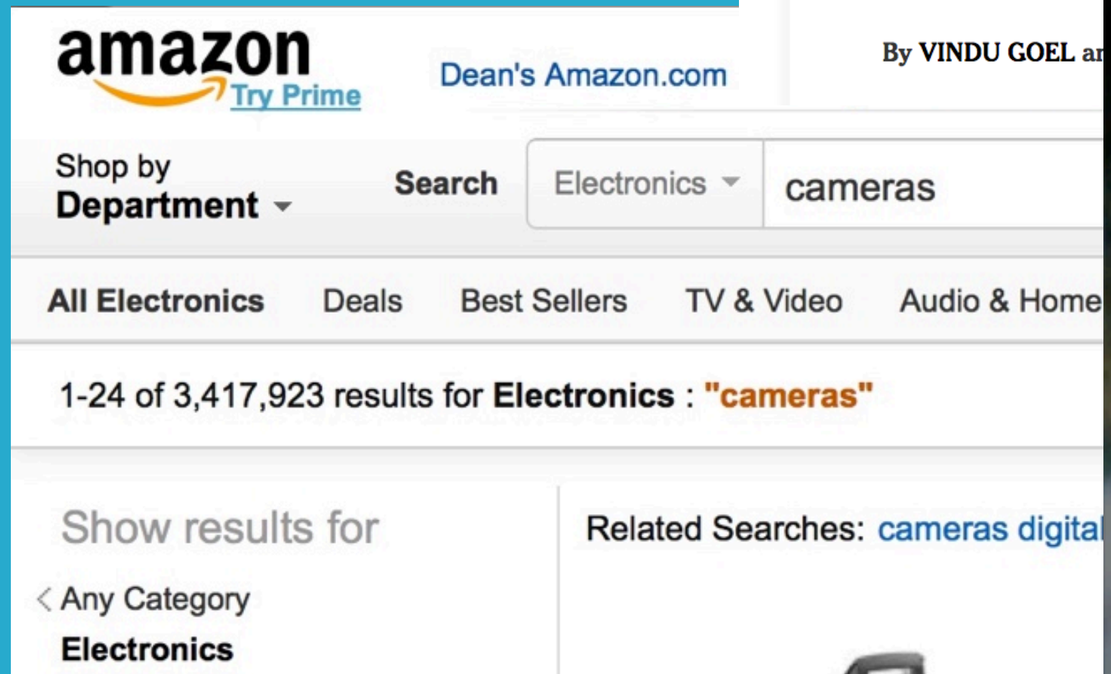
Respond to change



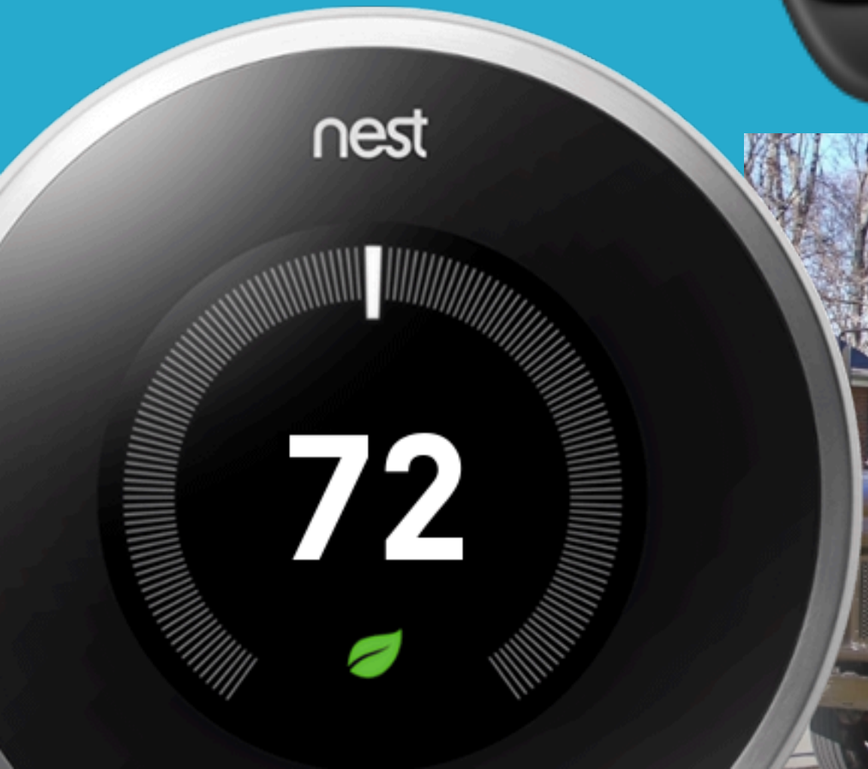
TECHNOLOGY

Yahoo Says 1 Billion User Accounts Were Hacke

By VINDU GOEL ar



IoT

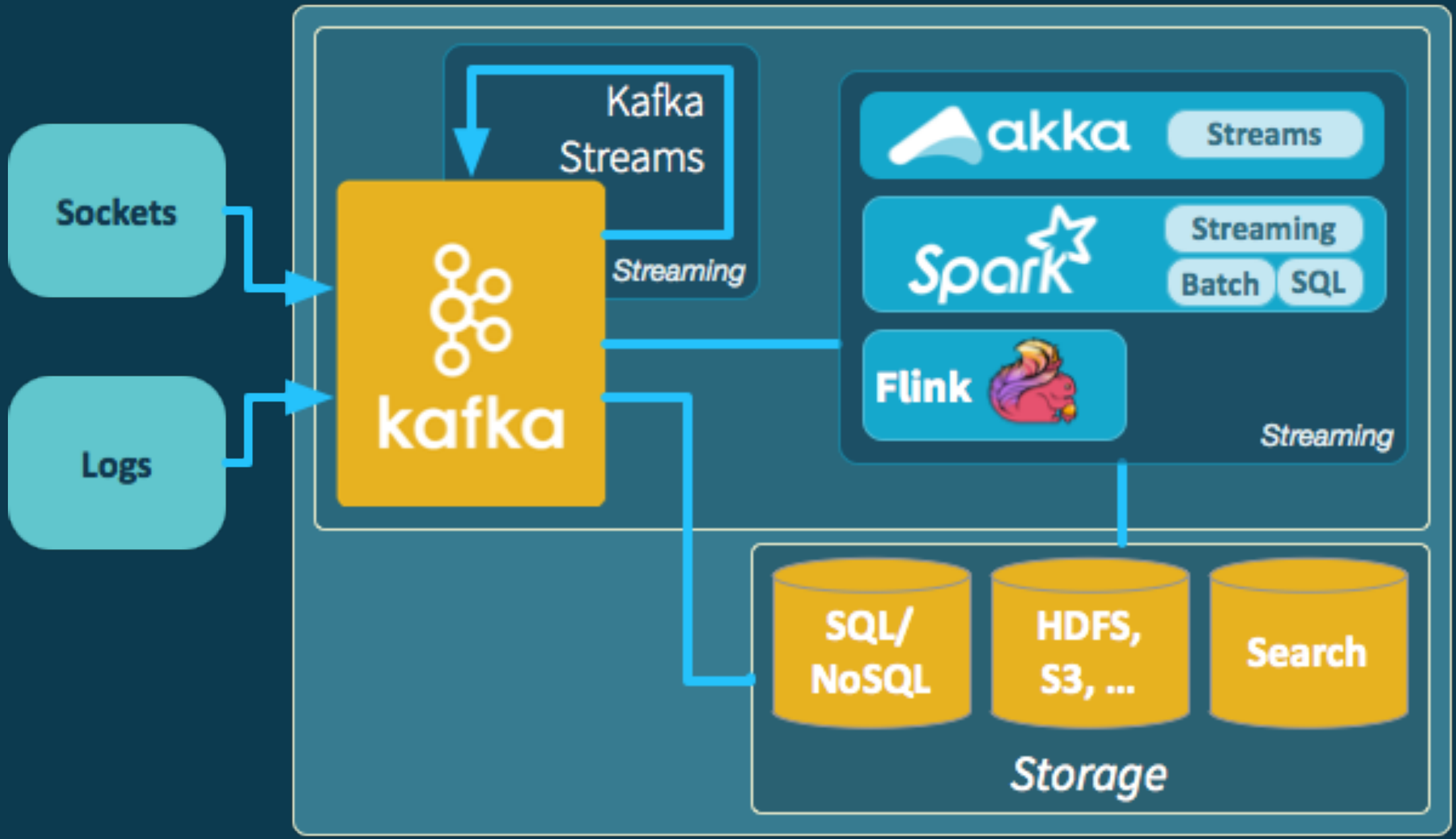

nest
72


4,718

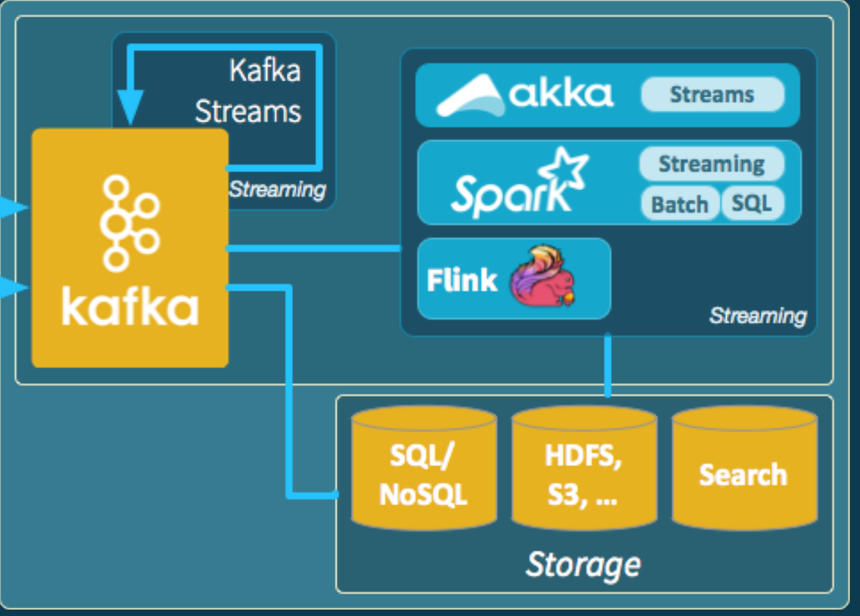

ups
Worldwide Services


9:41 AM
100%
Mail
Calendar
Photos
Camera
Maps
Clock
Weather
Wallet
Notes
Reminders
Stocks
Videos
iBooks
iTunes Store
App Store
Home
Health
Settings
Phone
Safari
Messages
Music

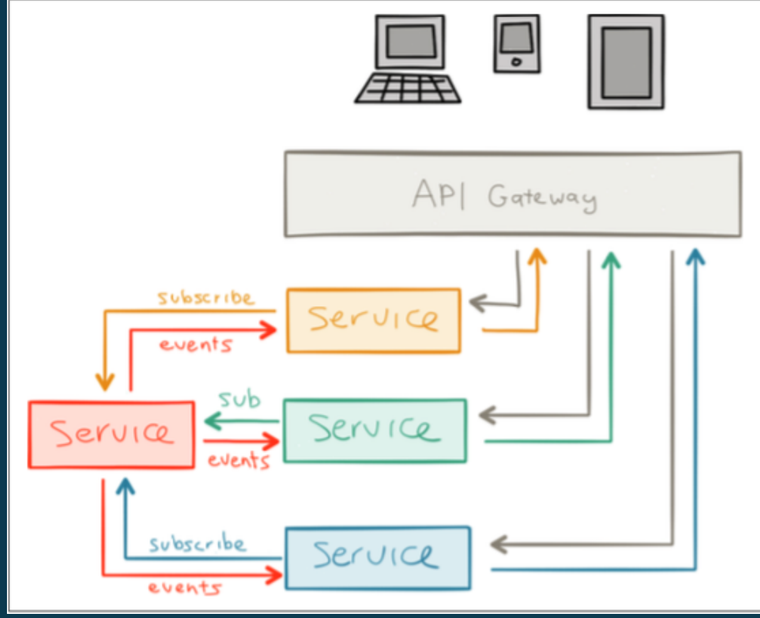

SYST
RPS
STAT
DUPLX
SPEED
MODE


Sockets
Logs
kafka
Kafka
Streams
Streaming
akka
Streams
Spark
Streaming
Batch
SQL
Flink
Streaming
SQL/
NoSQL
HDFS,
S3, ...
Search
Storage

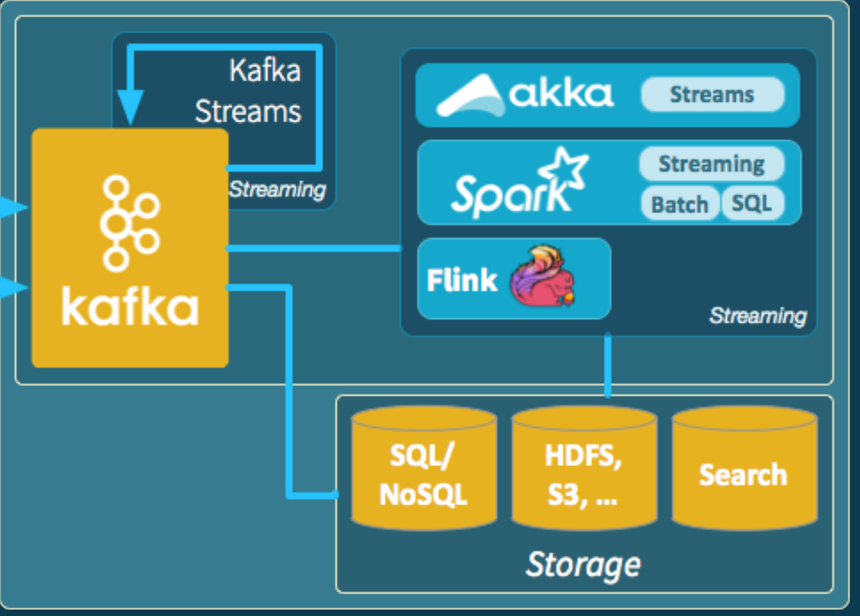
Fast Data and Microservices; Are they Converging?



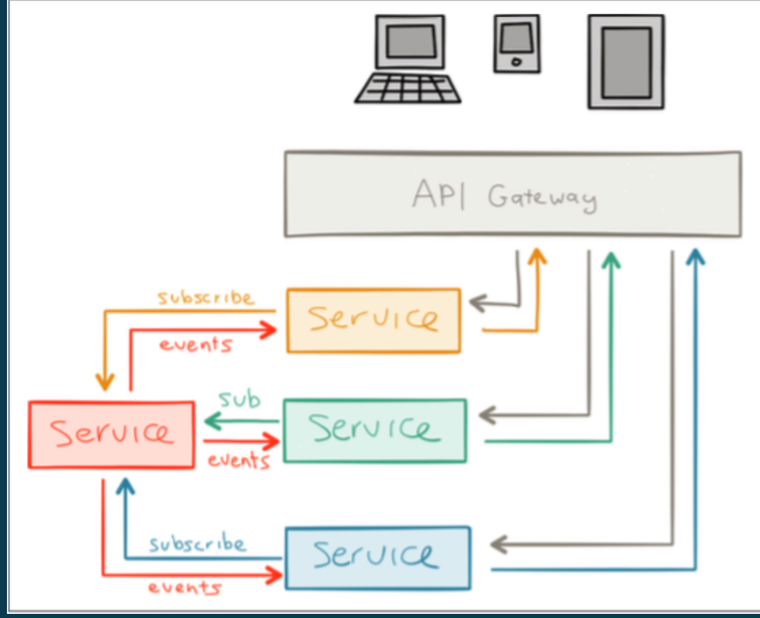
Synergies



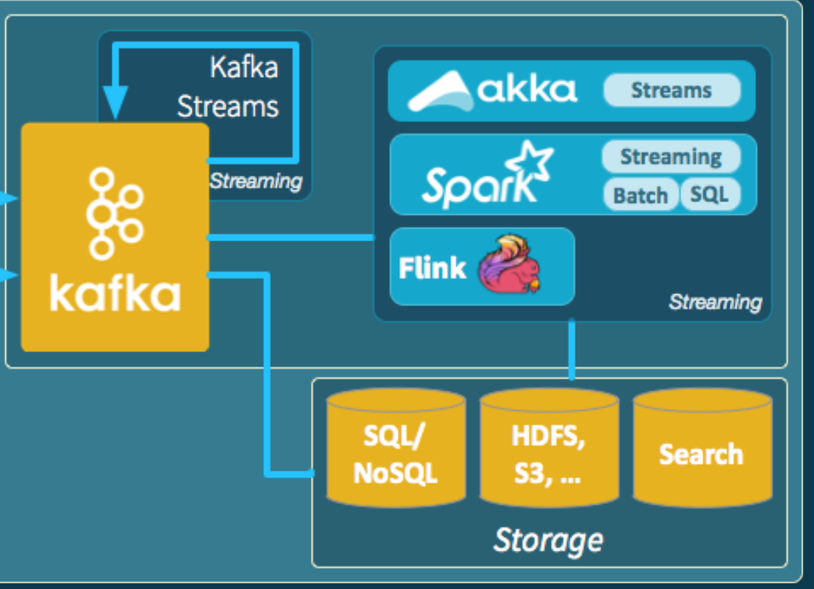
- Each [stream app or µservice]:
  - does one responsibility
  - ingests unending [data or messages]



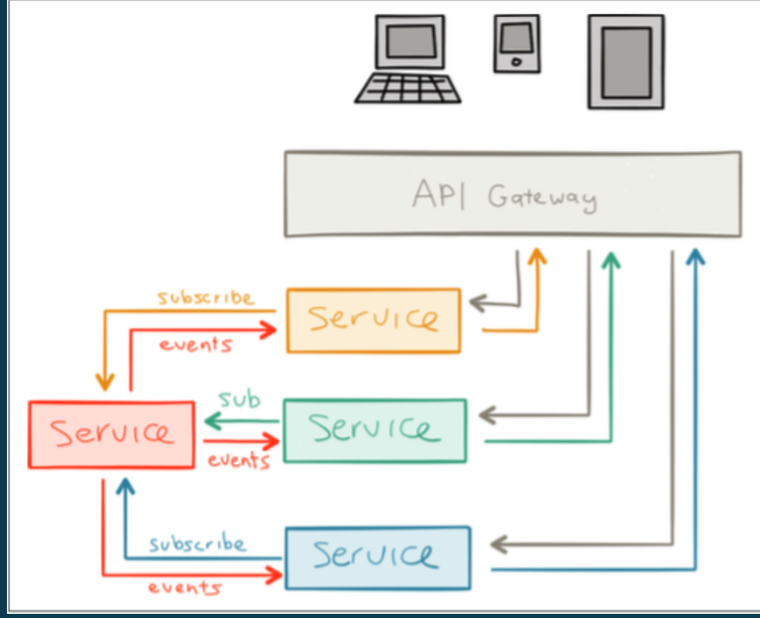
Synergies



- Each [stream app or µservice] must:
  - operate asynchronously
  - offer never-ending service



Thesis



- These architectures are converging:
  1. Similar design problems
  2. Data becomes dominant problem

Lightbend
Fast Data Platform

Value Three Ways

1. An accelerated on ramp for building streaming data systems, data applications, and other microservices.

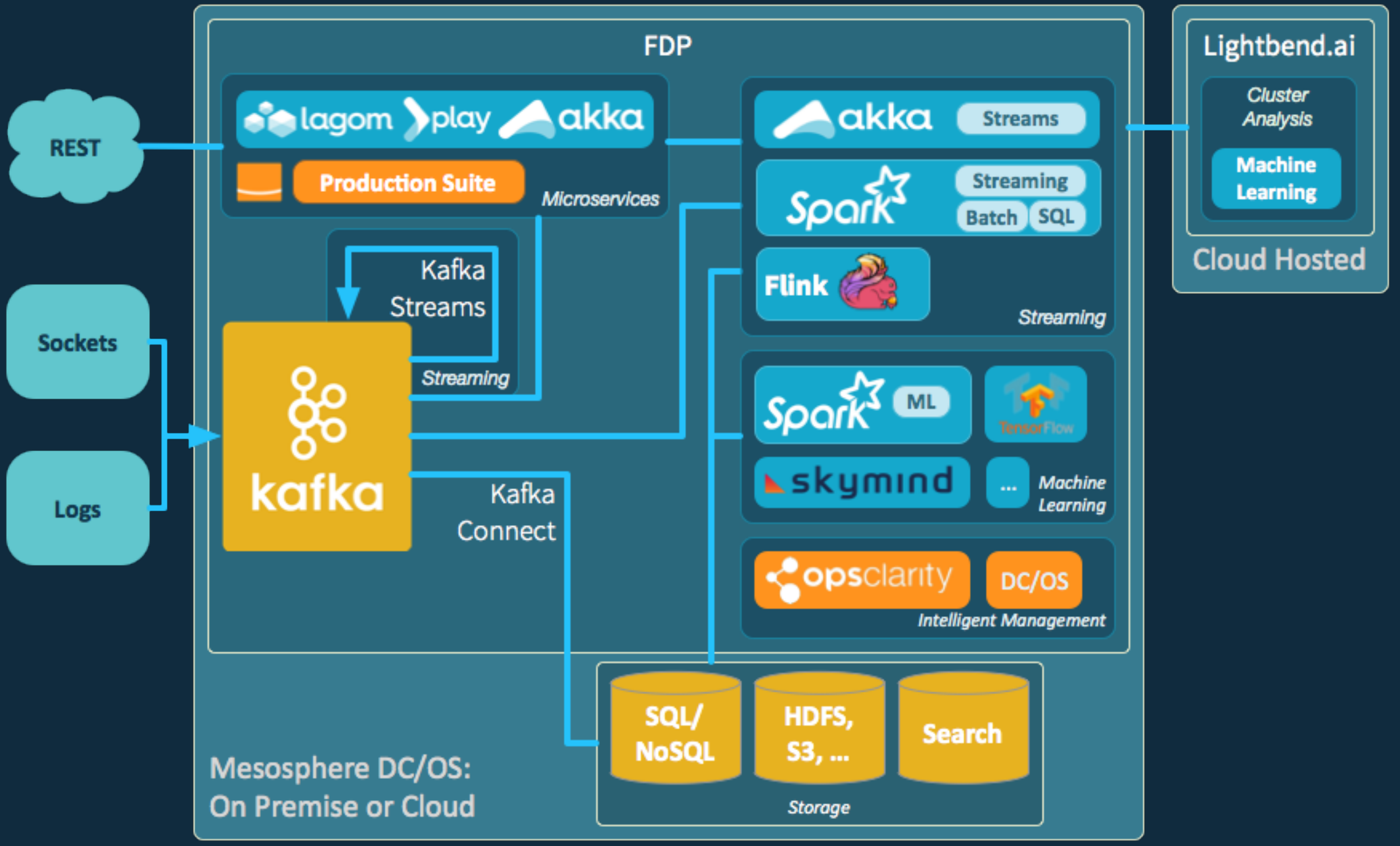
Value Three Ways

2. Best practices guidance for solving specific design problems:
    - Sample apps
    - Documentation
    - *Enablement services*

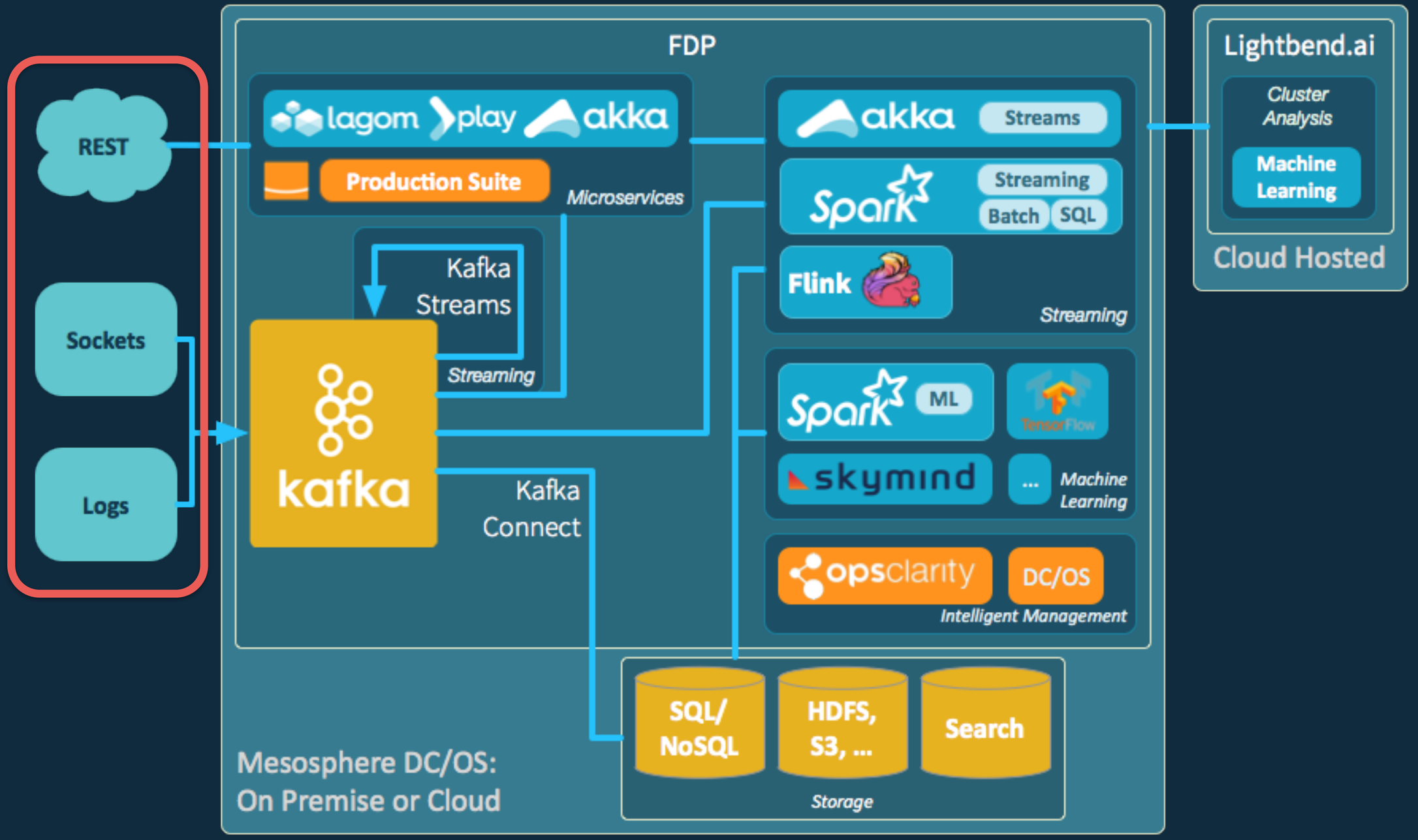
Value Three Ways

3. Machine learning-based monitoring and management:
    -Keep your systems resilient, scalable, and responsive with minimal user intervention.

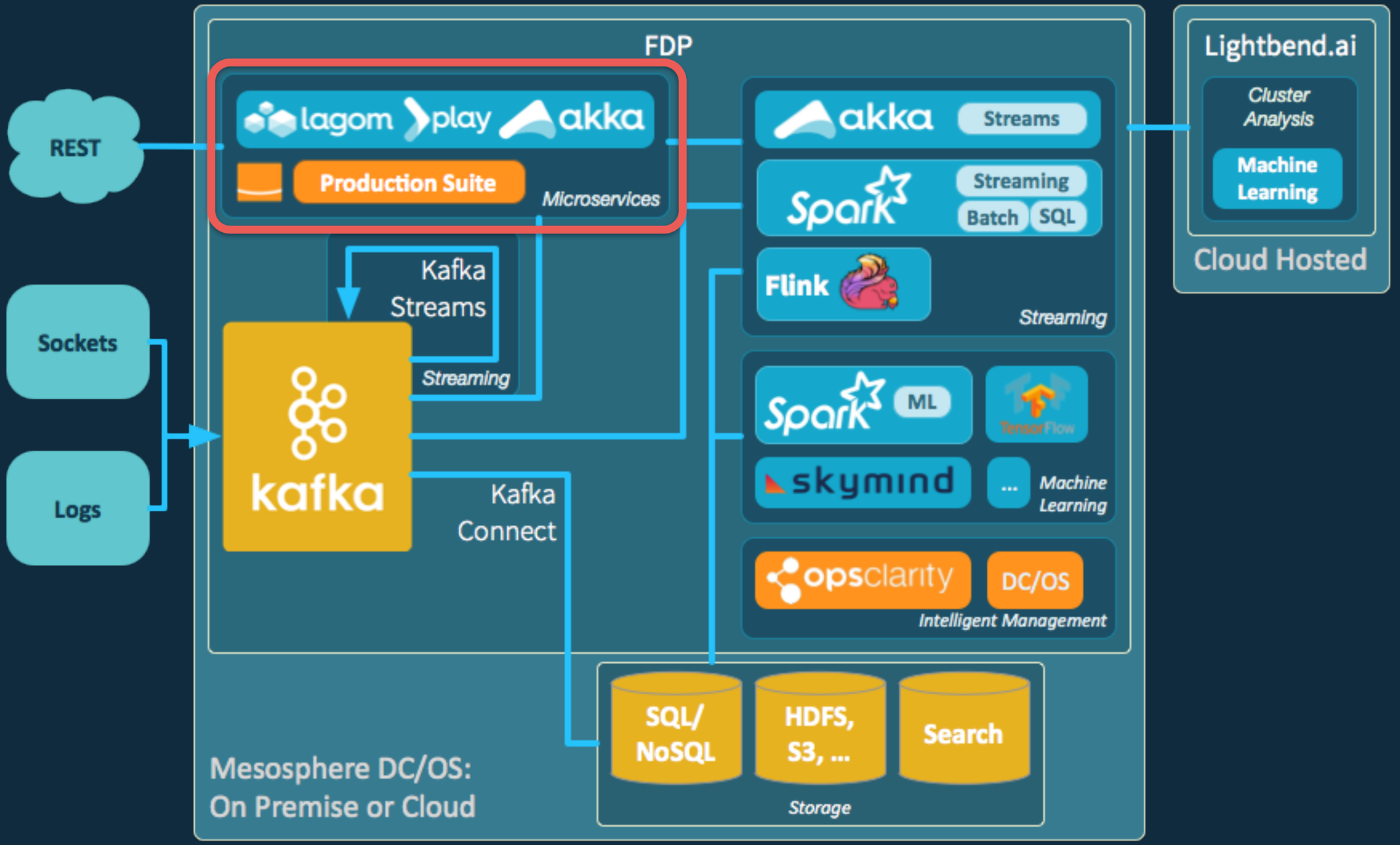
Lightbend Fast Data Platform V1.0



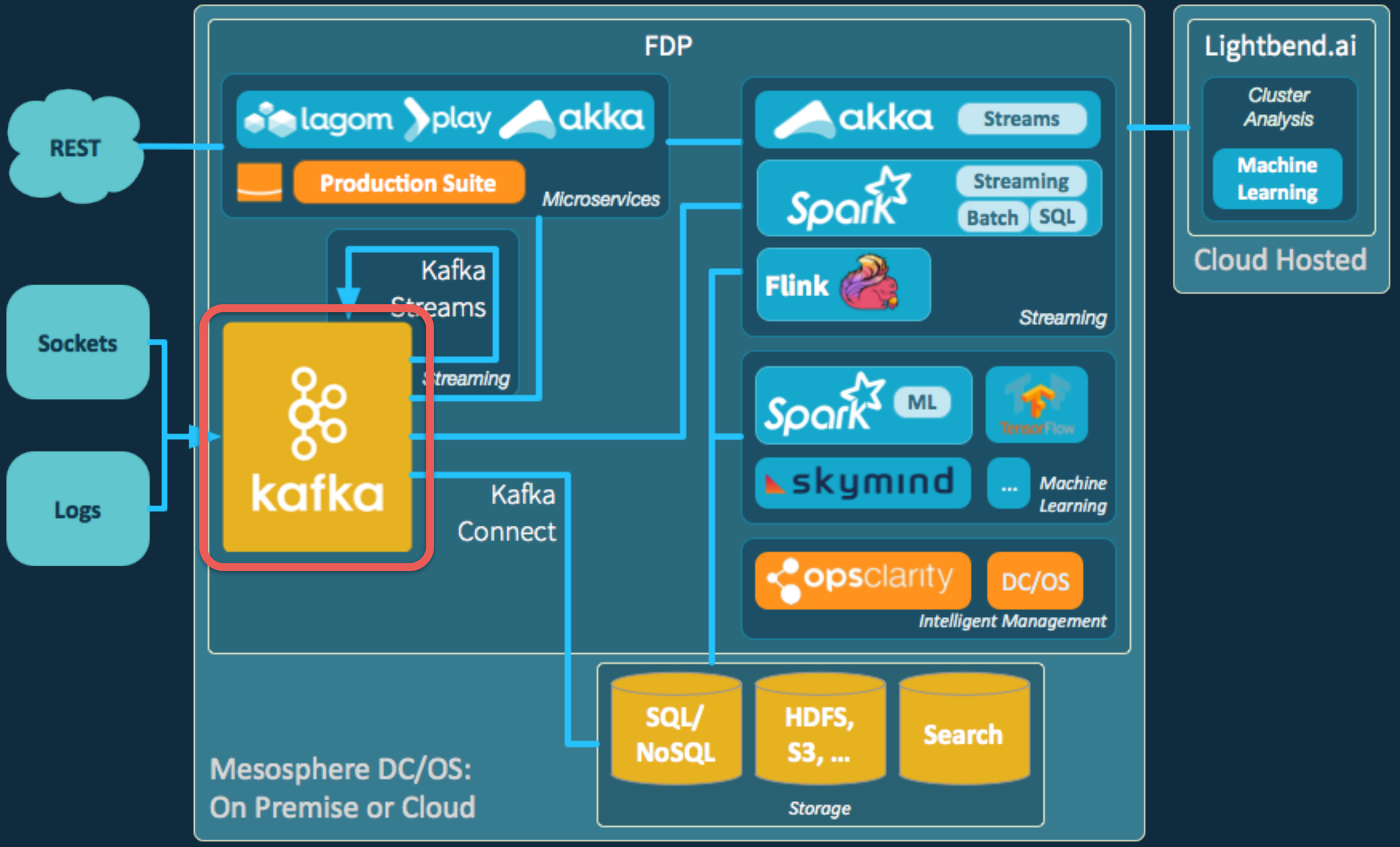
Lightbend Fast Data Platform V1.0

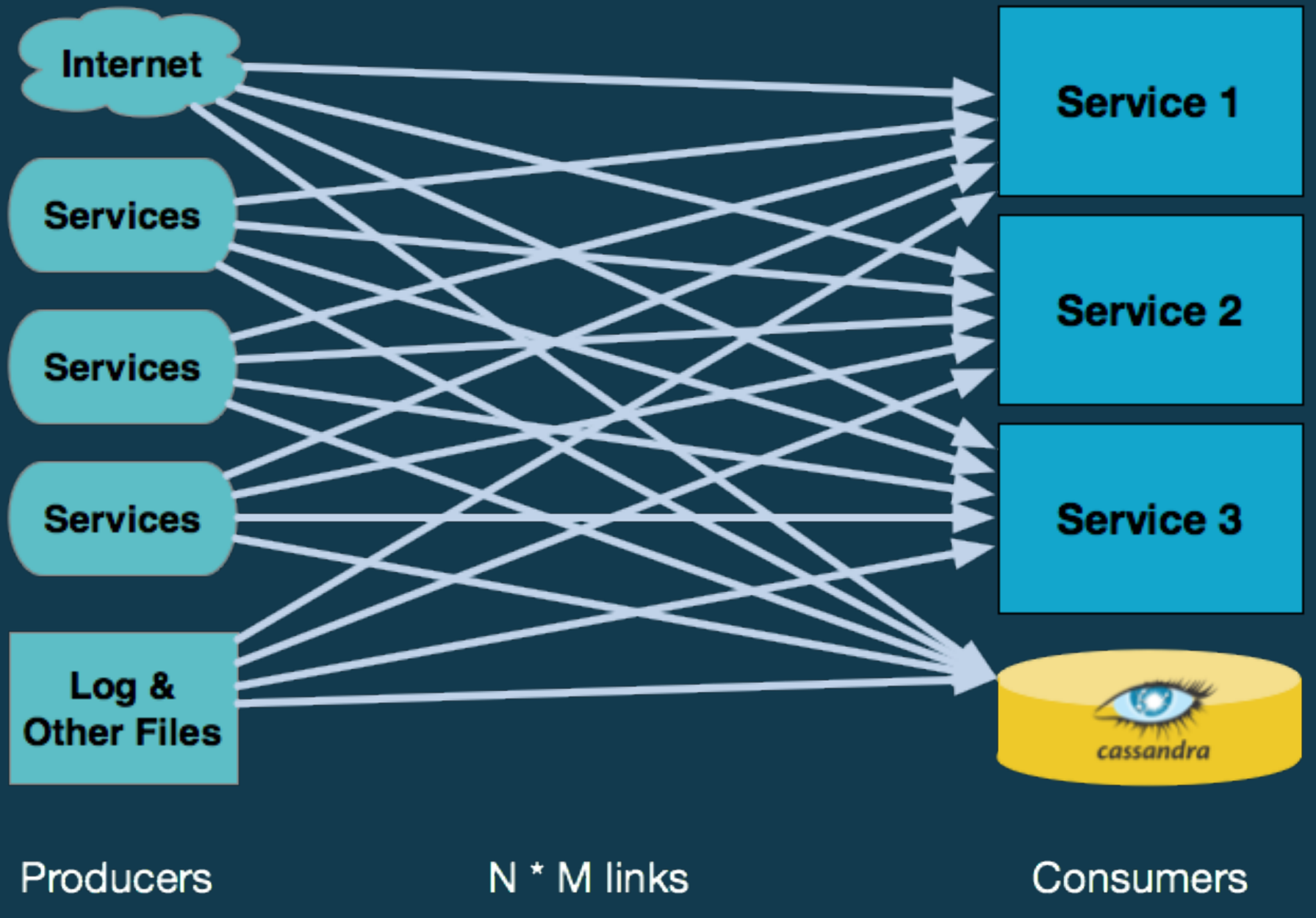


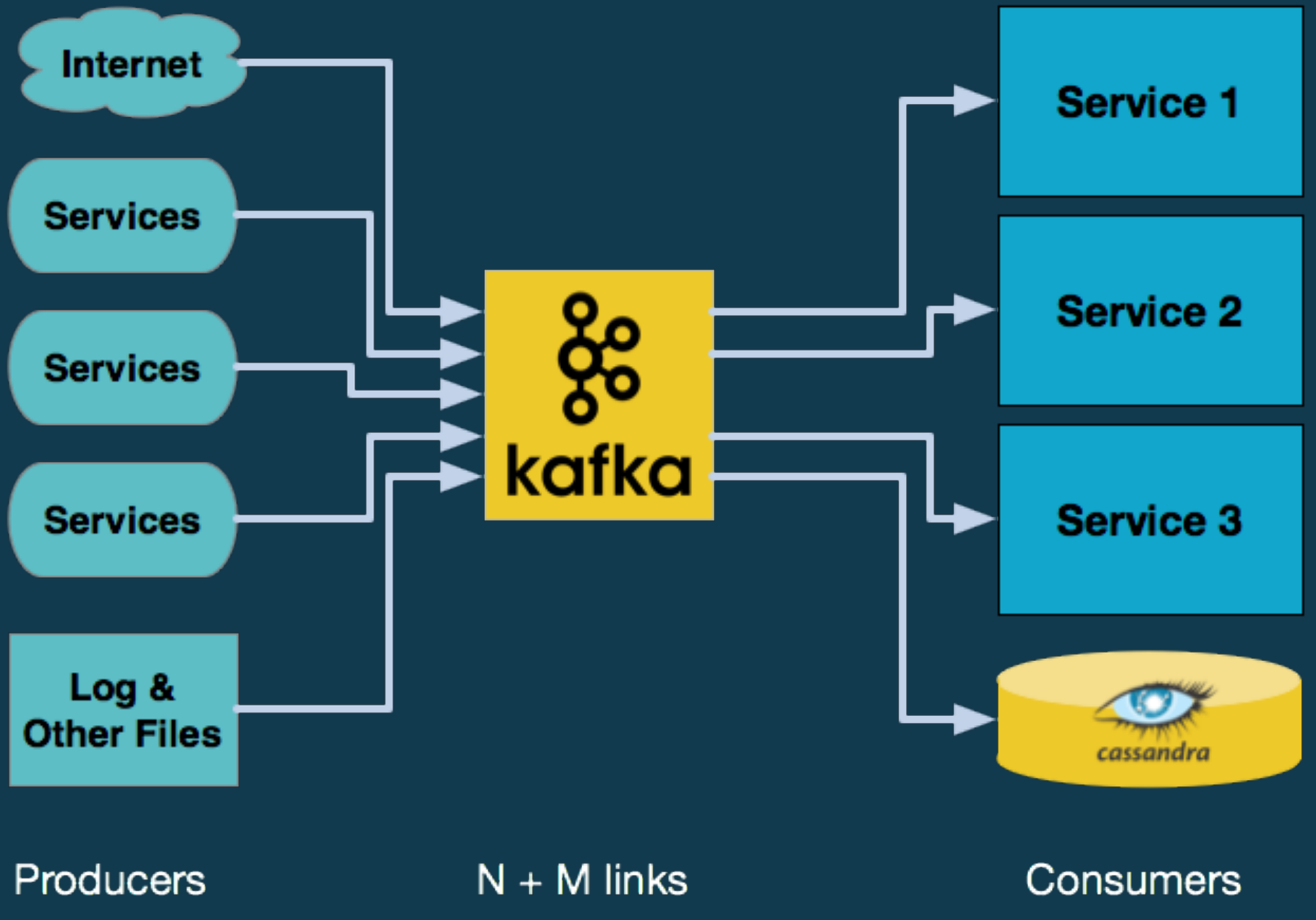
Lightbend Fast Data Platform V1.0



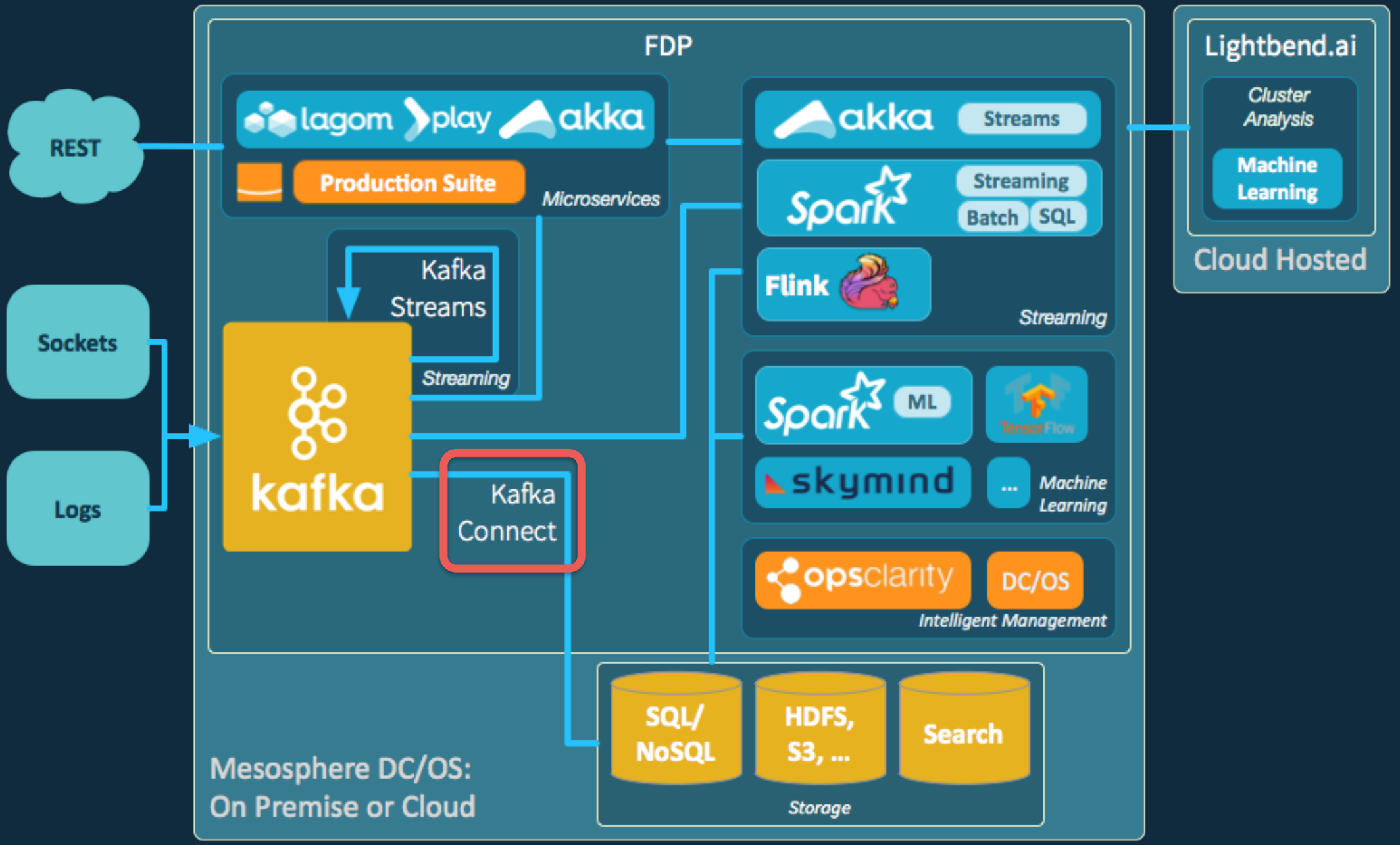
Lightbend Fast Data Platform V1.0



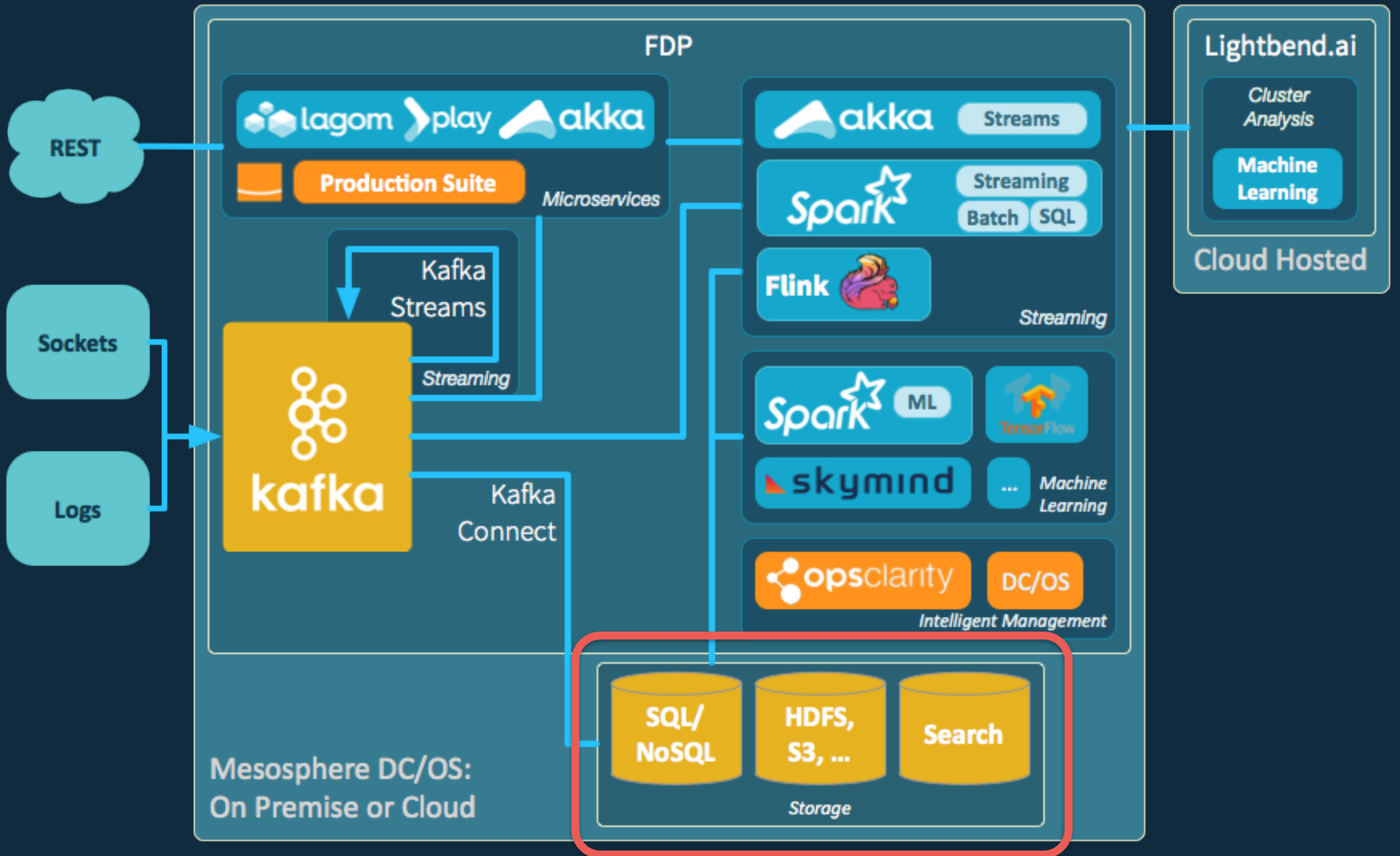

Internet
Services
Services
Services
Log &
Other Files
Service 1
Service 2
Service 3
cassandra
Producers
N * M links
Consumers


Internet
Services
Services
Services
Log &
Other Files
kafka
Service 1
Service 2
Service 3
cassandra
Producers
N + M links
Consumers

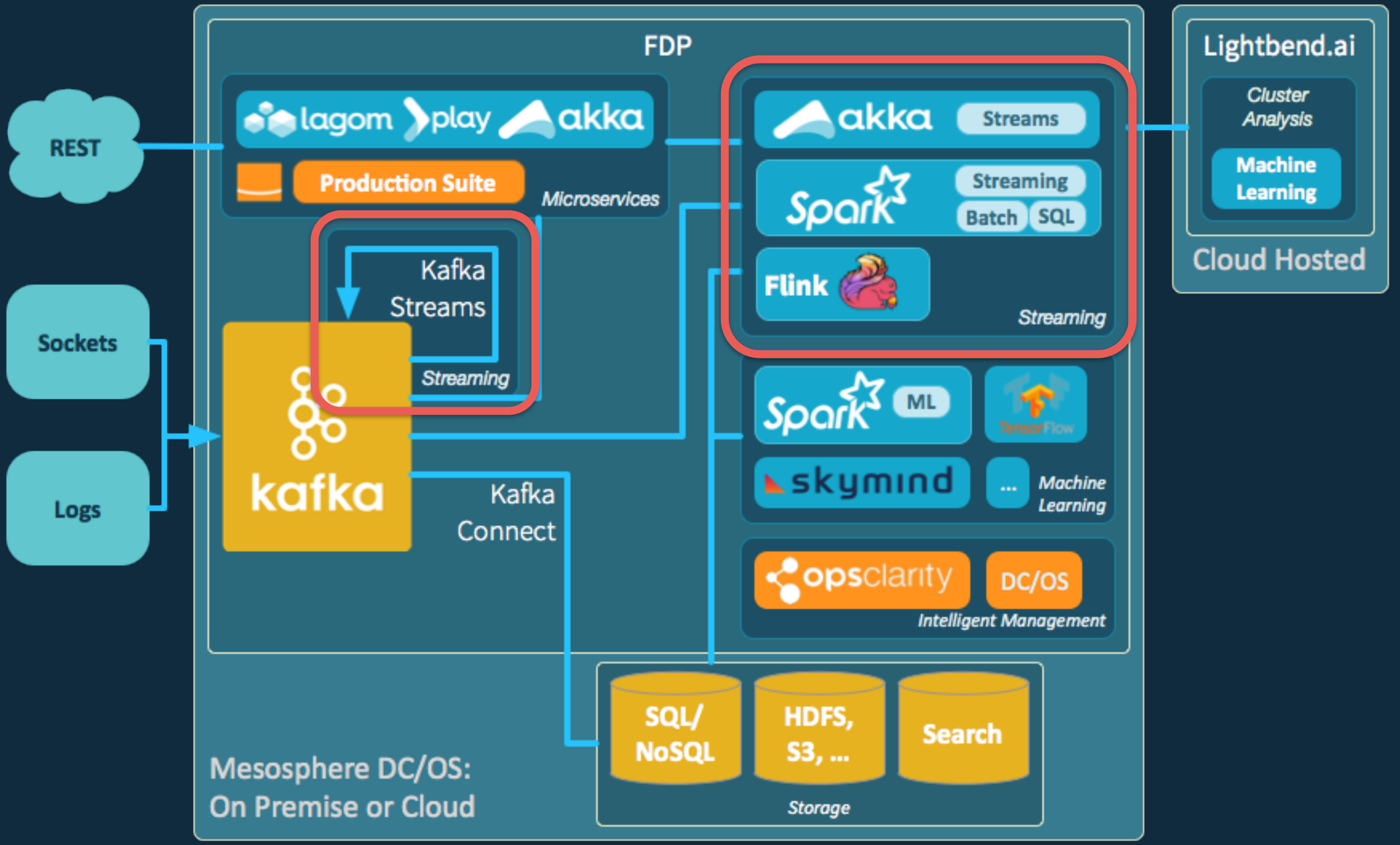
Lightbend Fast Data Platform V1.0



Lightbend Fast Data Platform V1.0



Lightbend Fast Data Platform V1.0



Streaming Tradeoffs (1/3)

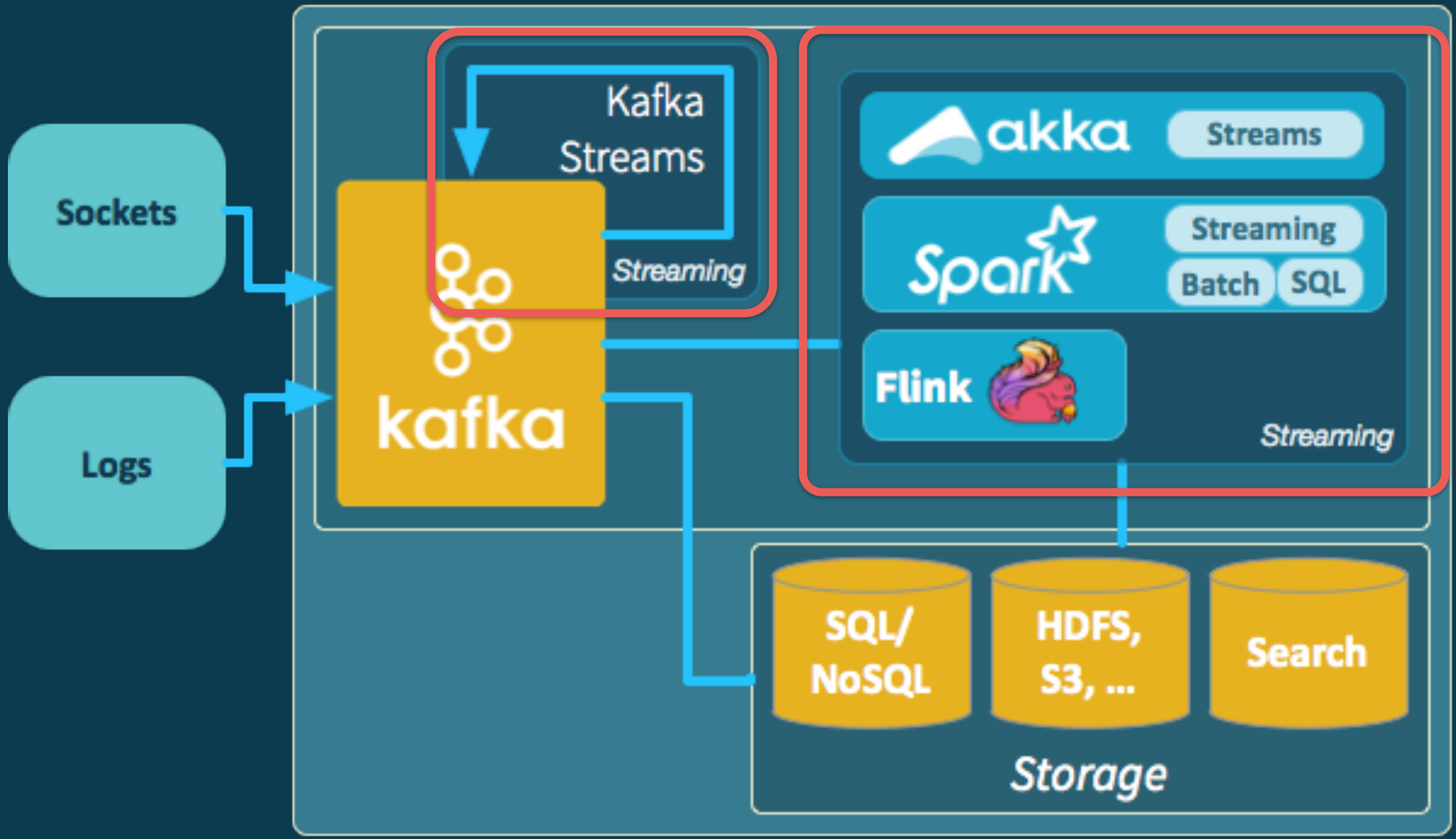
- Low latency? How low?
- High volume? How high?

Streaming Tradeoffs (2/3)

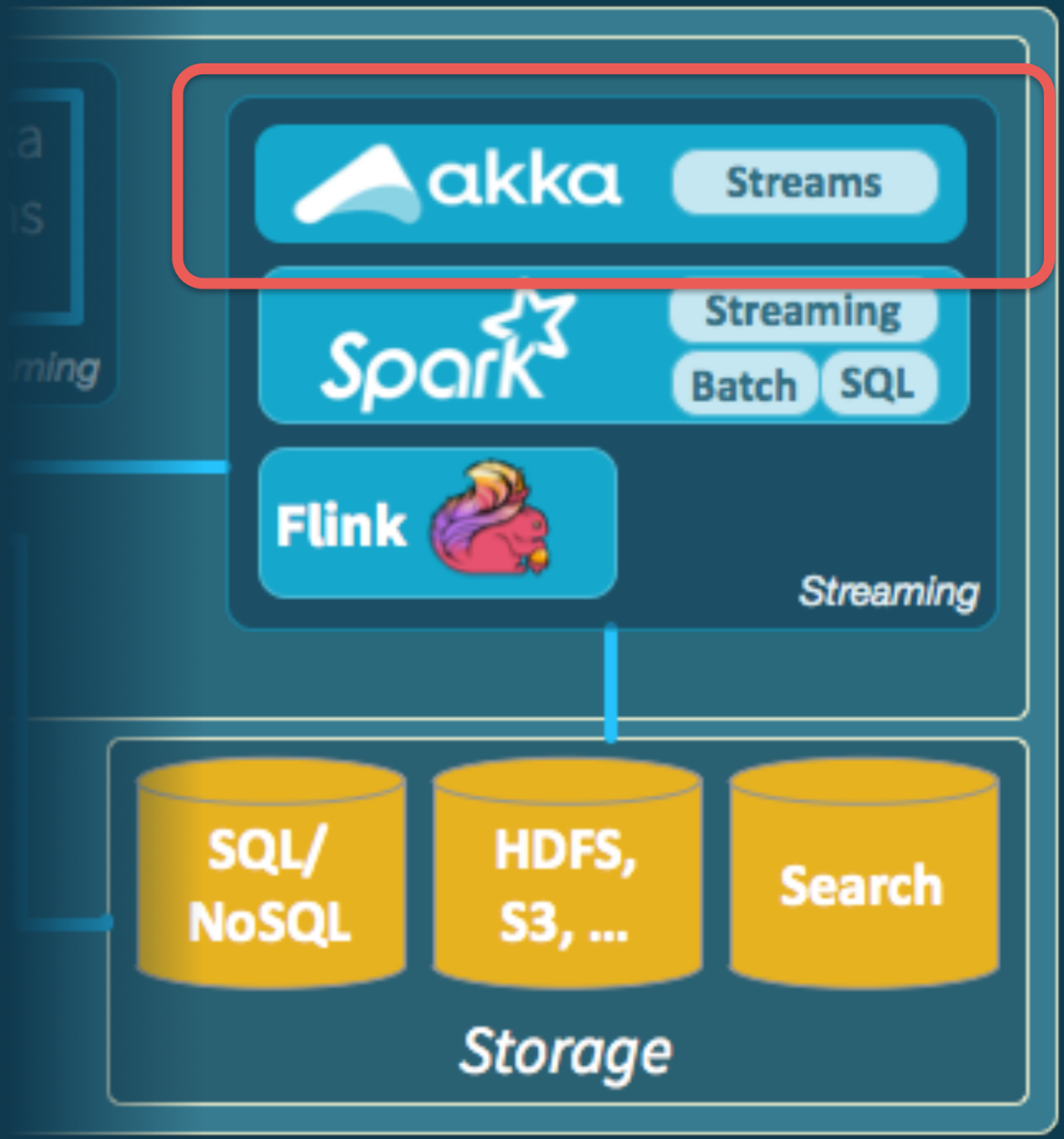
- Which kinds of data processing & analytics are required?
- How will this processing be done?
  - Individual processing of events?
  - Bulk processing of records?

Streaming Tradeoffs (3/3)

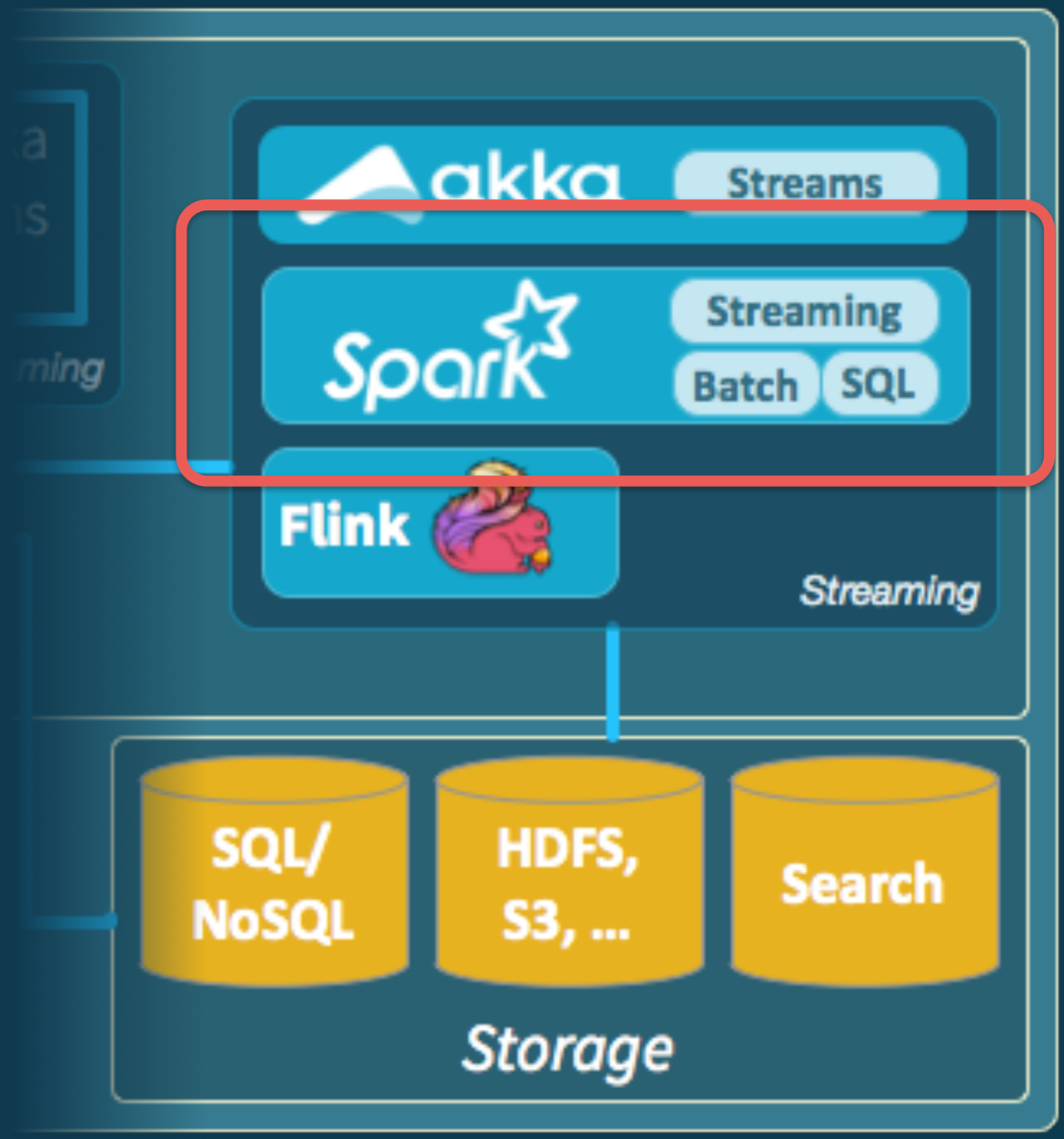
- Which tools and data sources/ sinks must interoperate with your streaming tool?


Sockets
Logs
kafka
Kafka
Streams
Streaming
akka
Streams
Spark
Streaming
Batch
SQL
Flink
Streaming
SQL/
NoSQL
HDFS,
S3, ...
Search
Storage

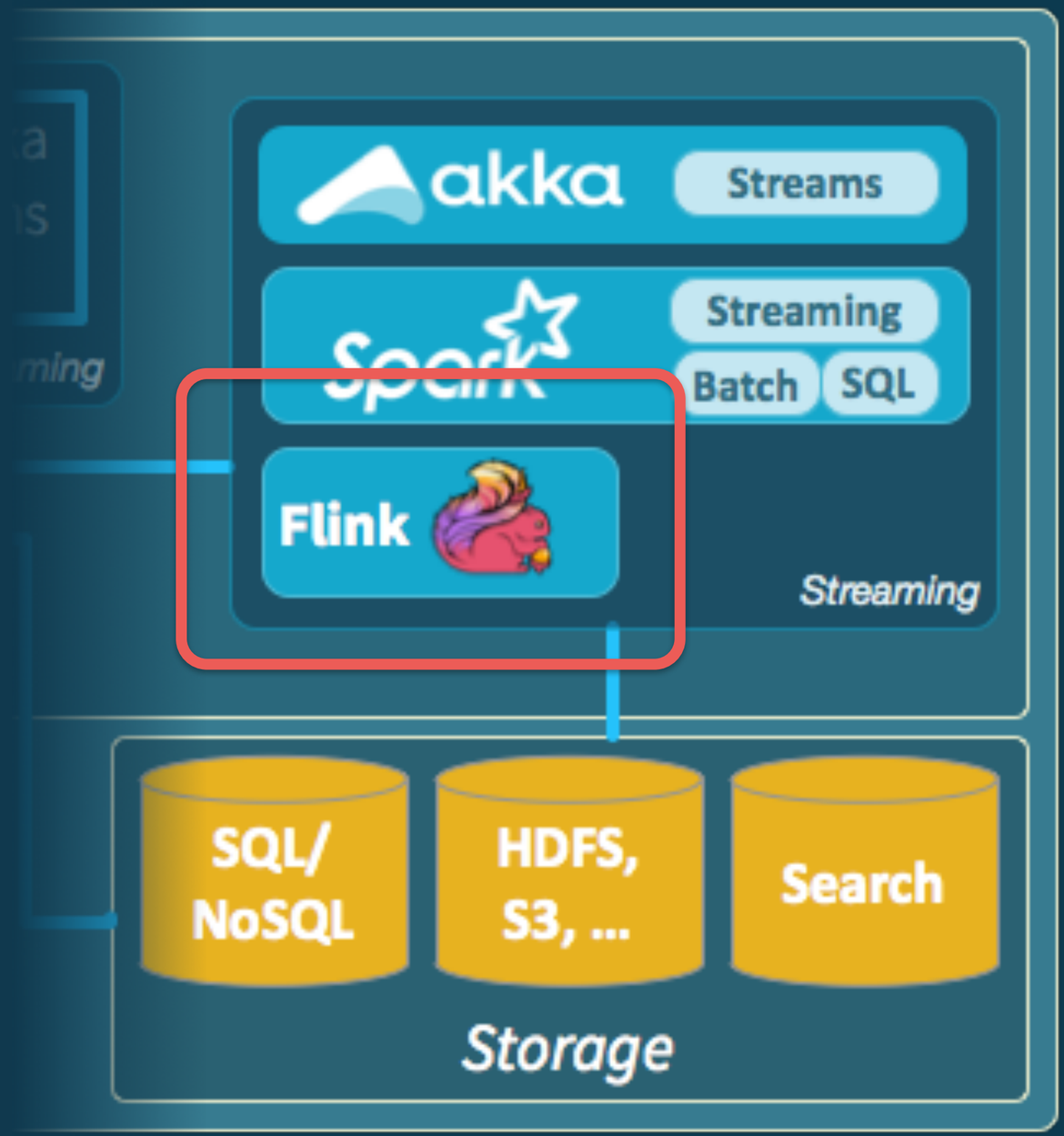
- Low latency
- Low volume
- Complex flows
- Complex Event Processing

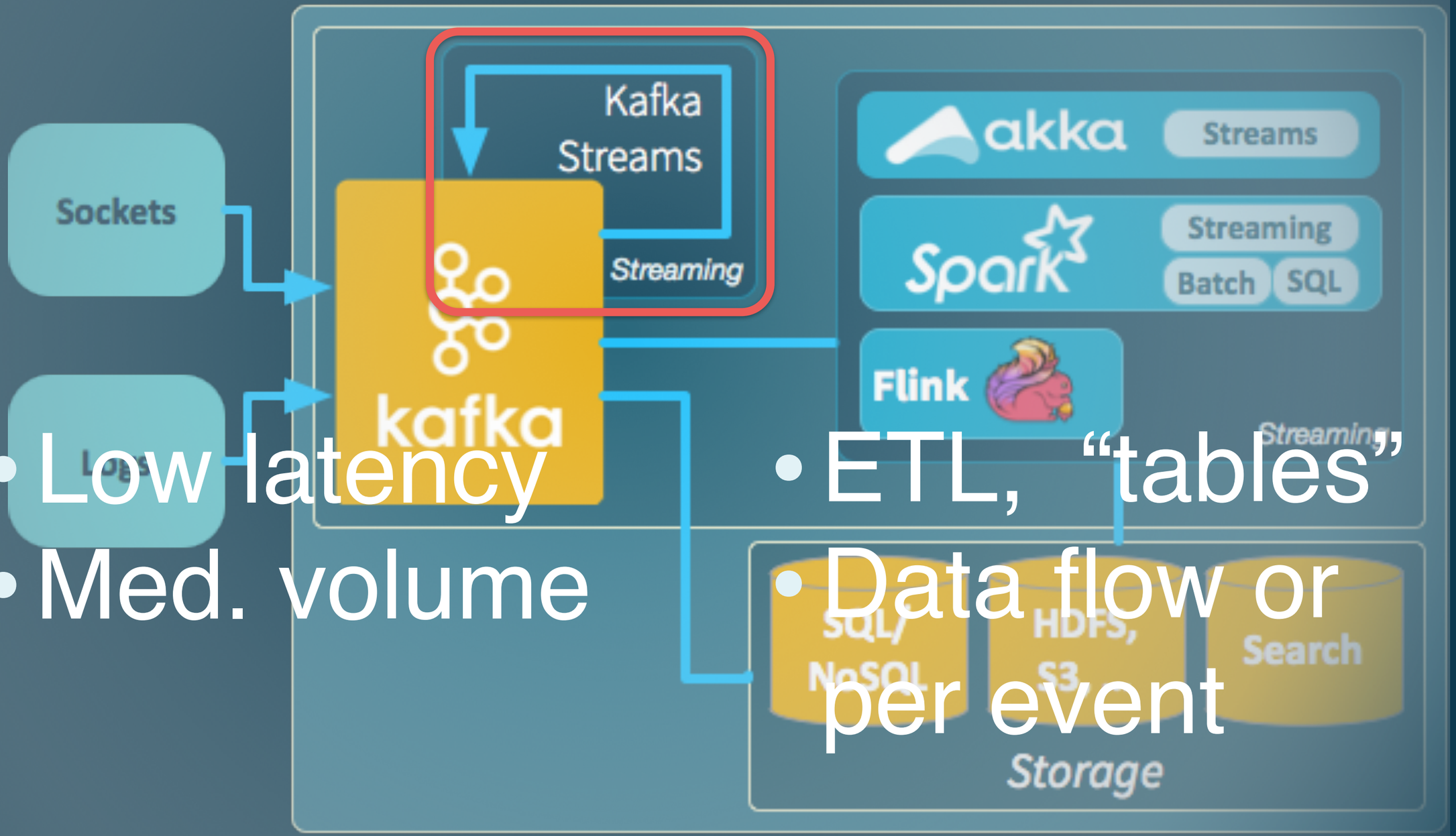


- Med. latency
- High volume
- Data flows, SQL
- *En masse* processing



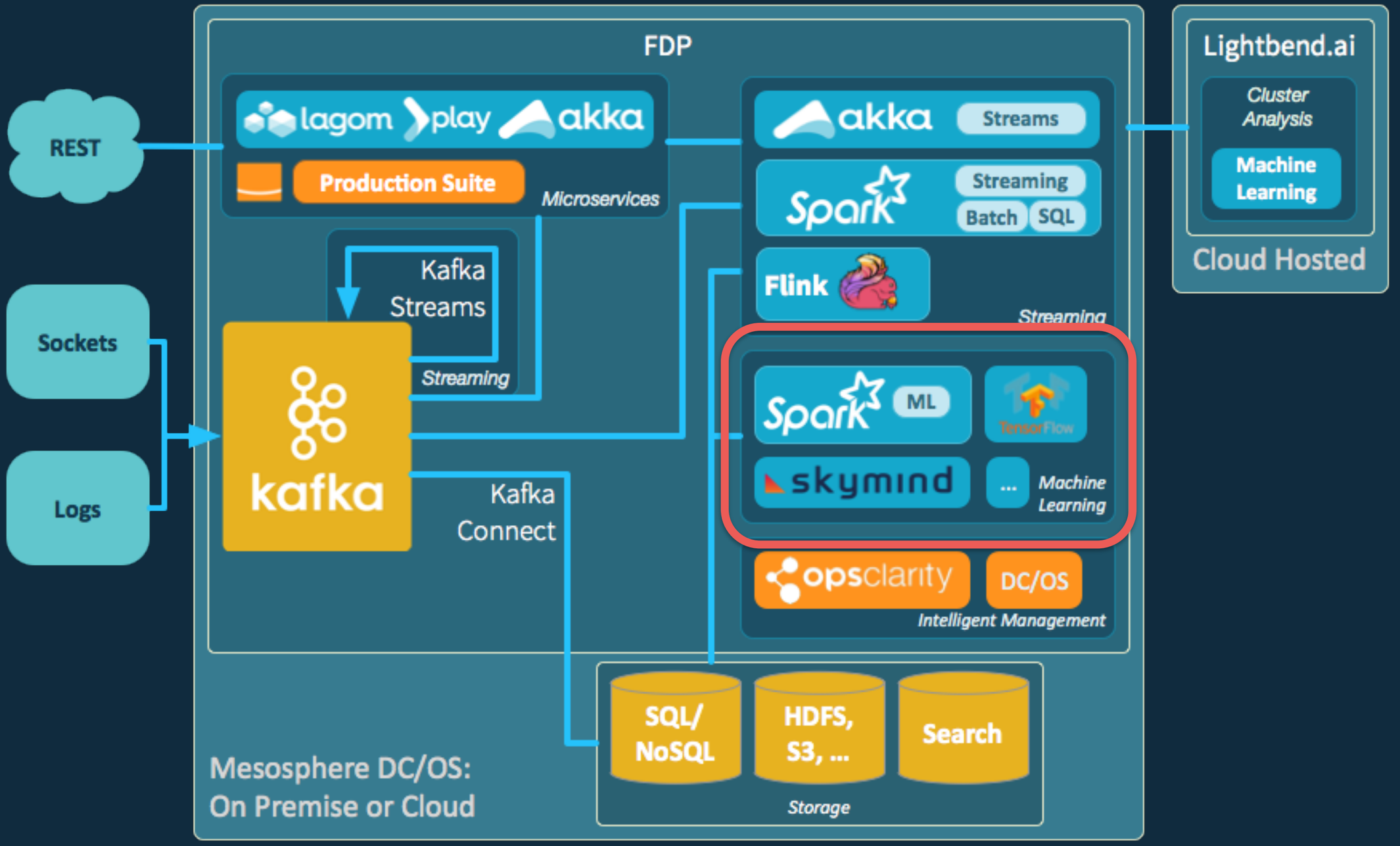
- Low latency
- High volume
- Data flows, correctness
- *En masse* processing



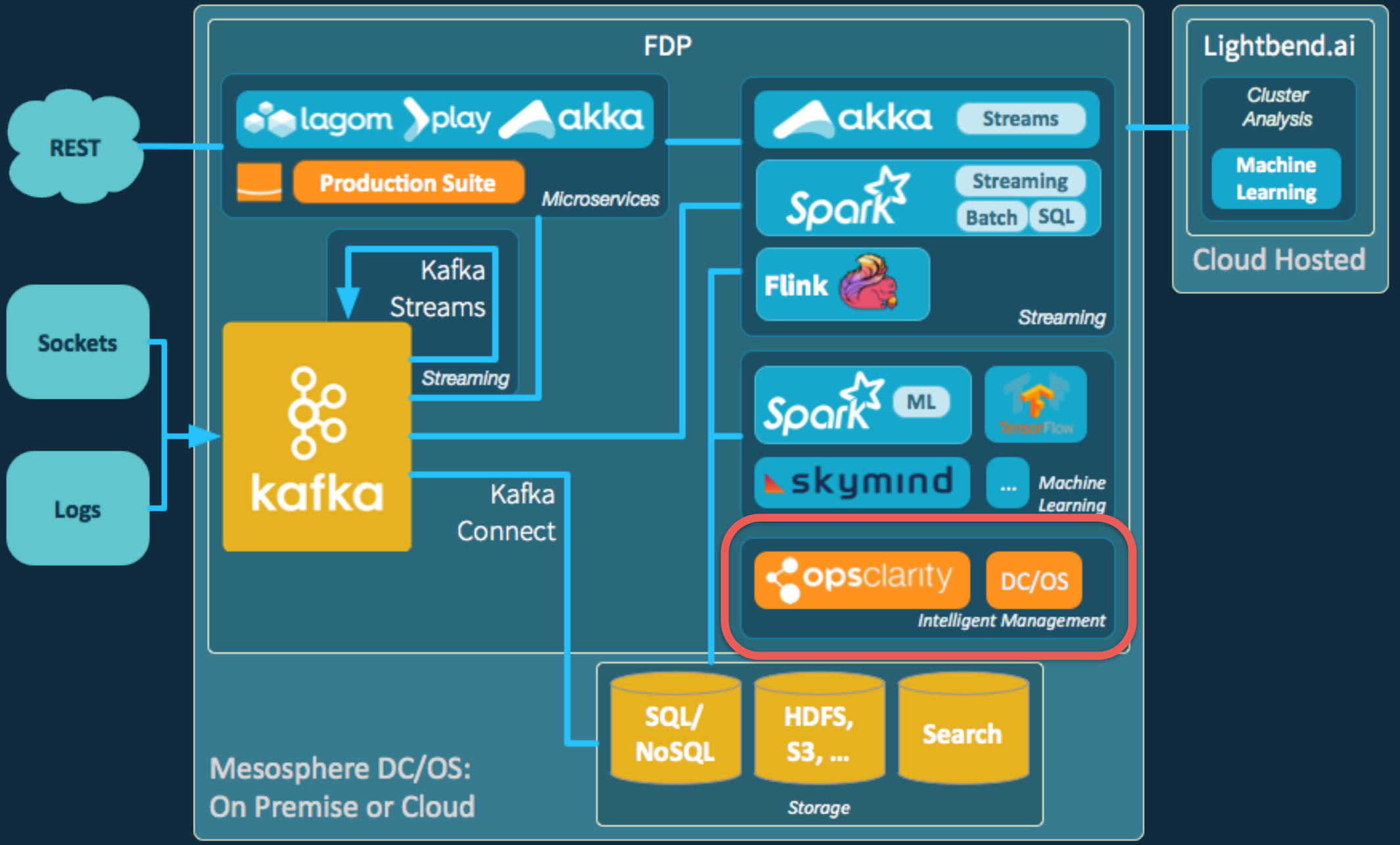


- Low latency
- Med. volume
- ETL, “tables”
- Data flow or per event

Lightbend Fast Data Platform V1.0



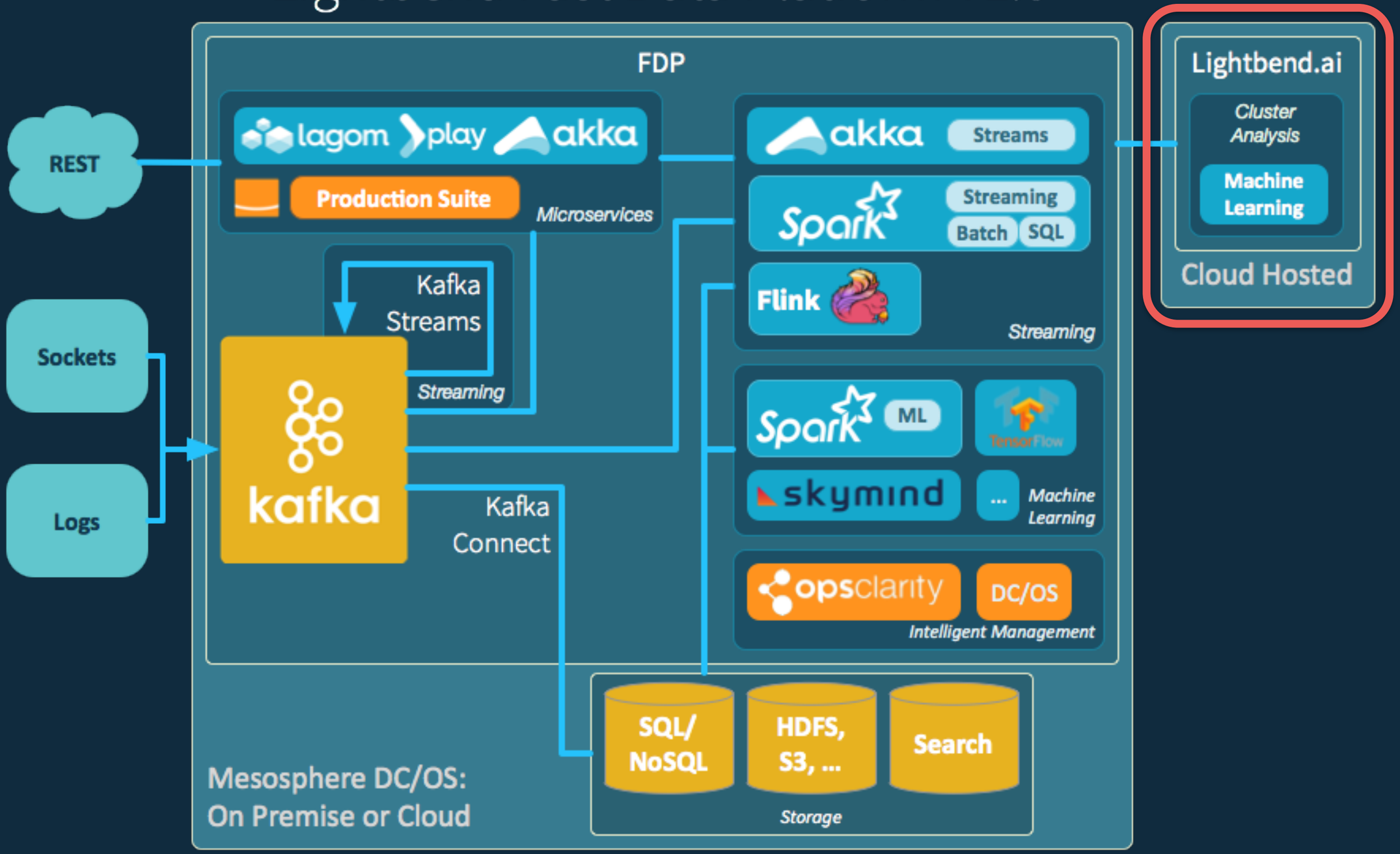
Lightbend Fast Data Platform V1.0



Kafka & Spark Metrics

Correlated Troubleshooting

Lightbend Fast Data Platform V1.0



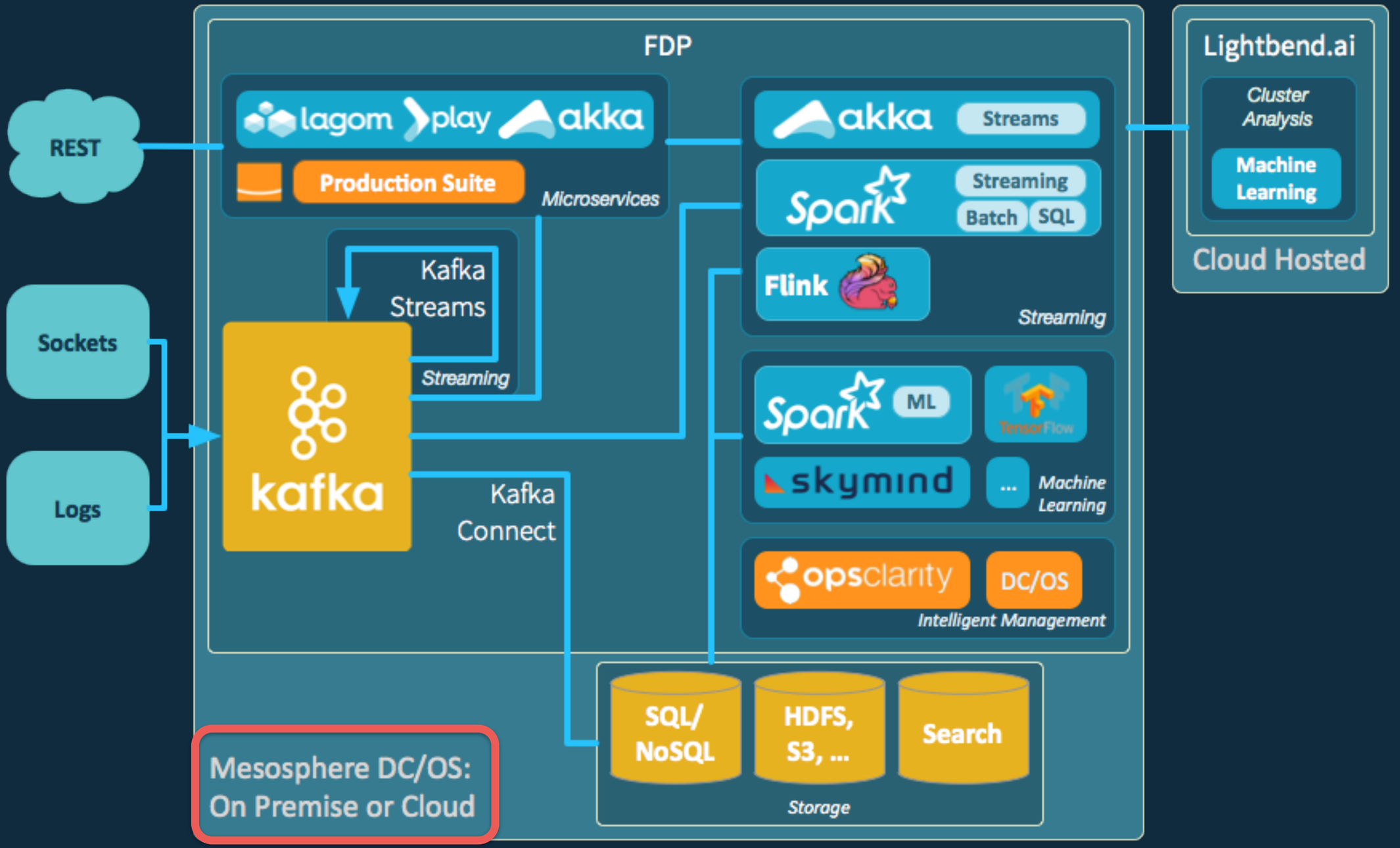
Why Only Hosted Services?

- Only metrics are uploaded, not sensitive domain data.
- One less service for you to manage.
- …

Why Only Hosted Services?

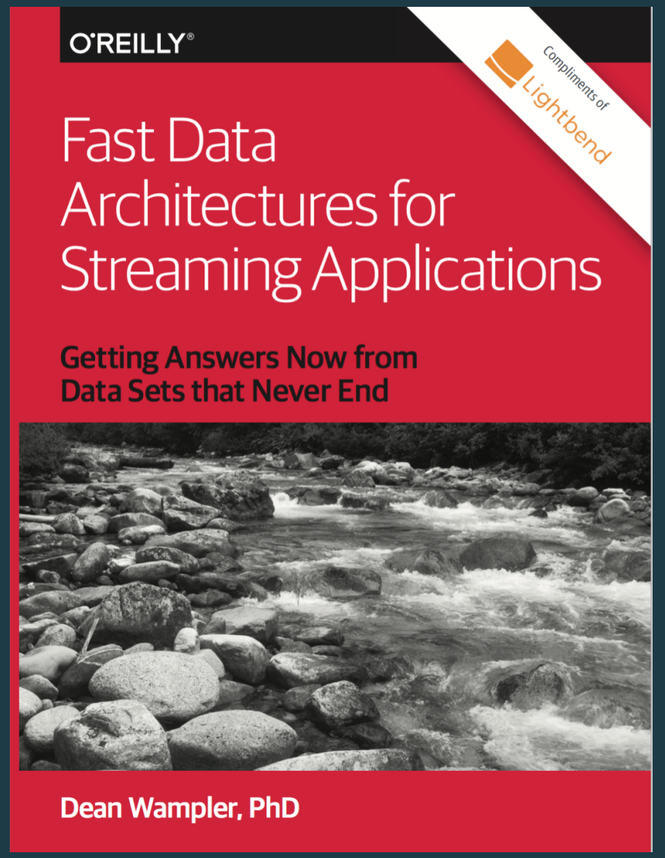
- …
- We can rapidly evolve this service. without impacting your environment.
- You benefit from aggregated knowledge from all FDP clusters.

Lightbend Fast Data Platform V1.0

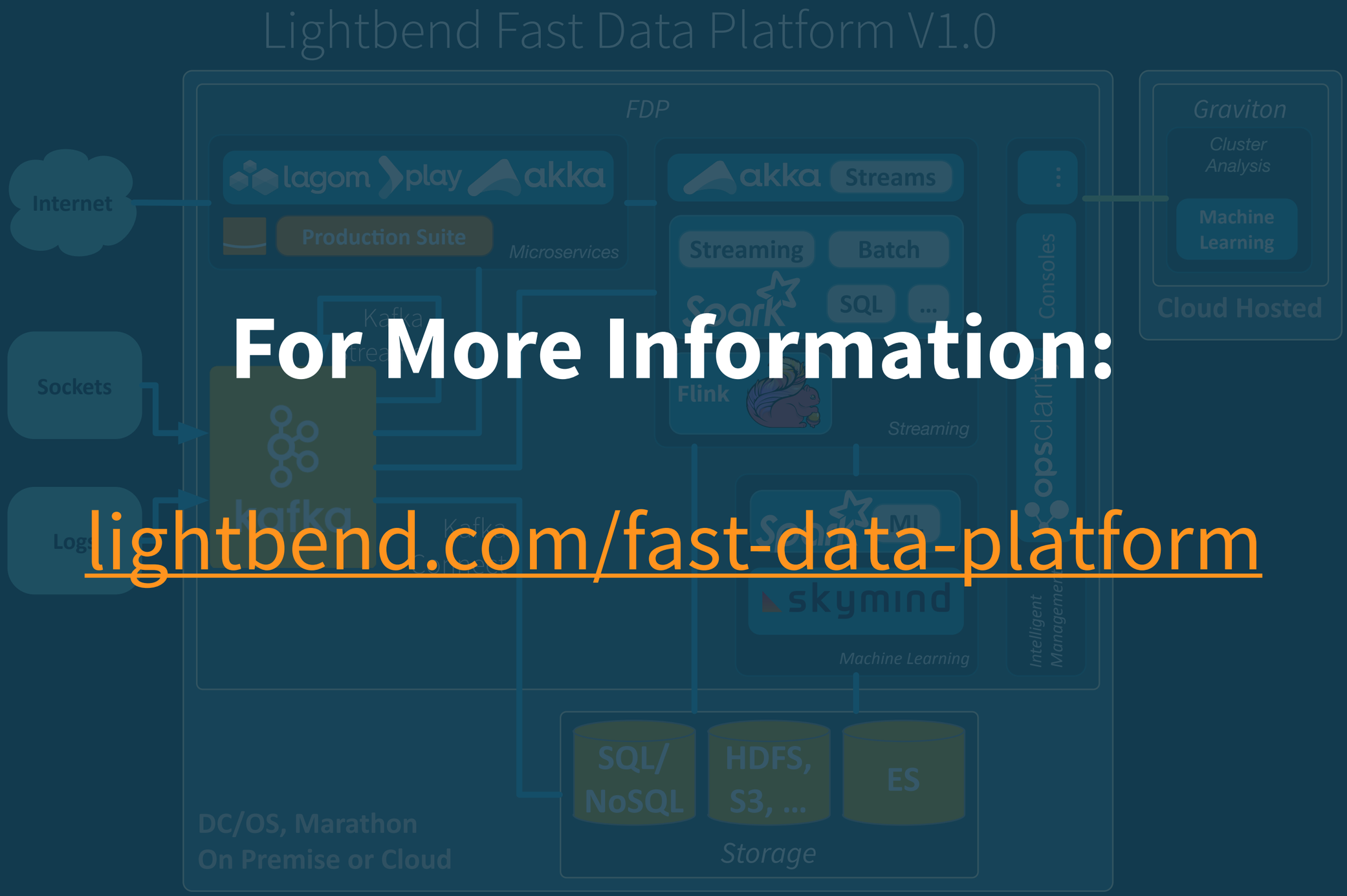


Upgrade your grey matter!

Get the free O'Reilly book by Dr. Dean Wampler,
VP of Fast Data Engineering at Lightbend



bit.ly/lightbend-fast-data



For More Information:

lightbend.com/fast-data-platform