

KEEDIO
Data Stack

BIG DATA PLATFORM

KDS Platform

Big Data platform: framework where use cases are developed

It houses every element required for advanced analytics:

- Data ingestion with/out message brokers
- Processing engines: Batch, Real Time and/or Streaming
- Storage: raw (HDFS) and unstructured (NoSQL DBs)
- Rule Engines for complex event processing
- Machine Learning tools
- Data Governance and security
- Connectors with external tools

KDS Tools

Integrated technologies:



KDS Advantages



No Vendor Lock-in

- Only stable widely-used tools
- Our patches and plugins are released as open source



Big Data Tools

- Comprising a selection of Big Data tools to build a modular architecture and an analytical platform



Apache Spark Distributions

- Integration with third-party libraries required for specific scenarios
- Bug-fixes and customizations



Module integration

- Tight modular integration even if independent deployments of modules are possible



Advanced analytics in real time

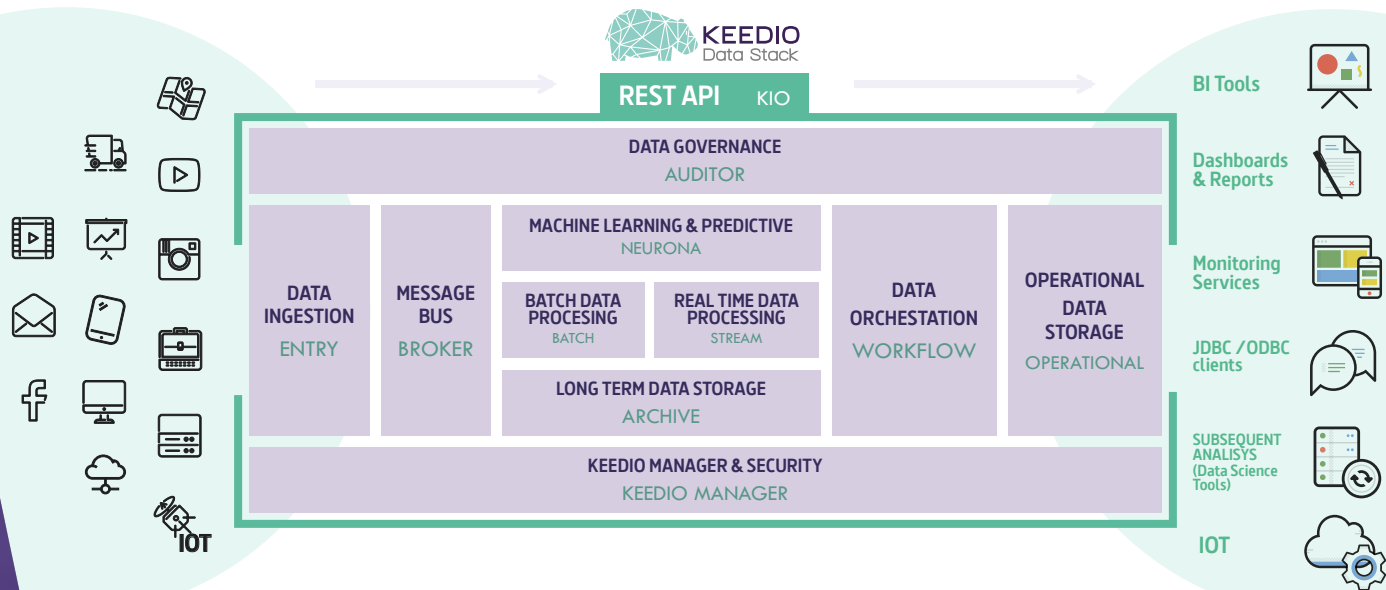
- Batch, real time, streaming and predictive.



Tailored distributions for every customer

- KDS is not monolithic, but adapts to every use case and customer's needs
- Compact platforms are easier to maintain

KDS Modules



KEEDIO MANAGER

KEEDIO Manager based on Apache Ambari

Native integration with Keberos and FreeIPA.
Native support for full HA.
Supported by the community and evolution global principles.

Improvements by KEEDIO

Standardization of service management
Adaptations for easy management of new Big Data tools
Integration with tools unavailable in the Community Version

ENTRY

Unified Ingestion Architecture based on Apache Flume

Versatility and power of dozens of components supported by the community.

Wide range of components (20+ cont'd by KEEDIO)

Full HA if required

Integration with Apache NiFi

Distributed ingestion engine

User interface for configurations of ingestion and transformation pipelines

Open sourced components contributed by KEEDIO.

BROKER

Based on Apache Kafka

Temporal data persistence.

Multichannel high-performance bus

Decouples ingestion and consolidation/processing layers

ARCHIVE

Based on Apache Hadoop HDFS

Guarantees information is stored redundantly with high availability

Compatible with every module in KDS, highly integrated with third-party tools

Fine-grained control access

BATCH & STREAM

Batch

Jobs & tasks executed at some specific time (scheduled or not)

Stream

Processing regular data flow with bounded delay bearable for humans

NEURONA

Analytics: Statistics, Math & Machine Learning

Distributed computation

Batch or stream: offline or online analytics





www.keedio.com



info@keedio.com



[@keedio](https://twitter.com/keedio)



[keedio](https://www.linkedin.com/company/keedio)

Calle Virgilio 25
Edificio Ayessa I, Bajo D
Pozuelo de Alarcón
28223 Madrid