

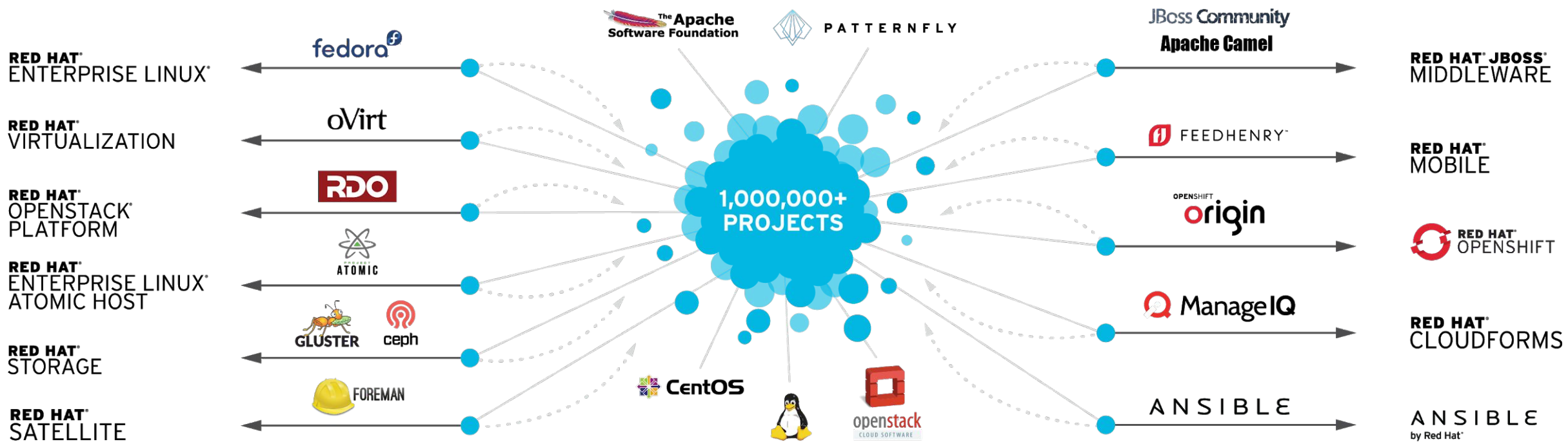
cloudera



AN END-TO-END OPEN SOURCE ARCHITECTURE FOR THE INTERNET OF THINGS

Scott Seighman
Solutions Architect / Red Hat
ss Leighma@redhat.com

FROM COMMUNITIES TO ENTERPRISE



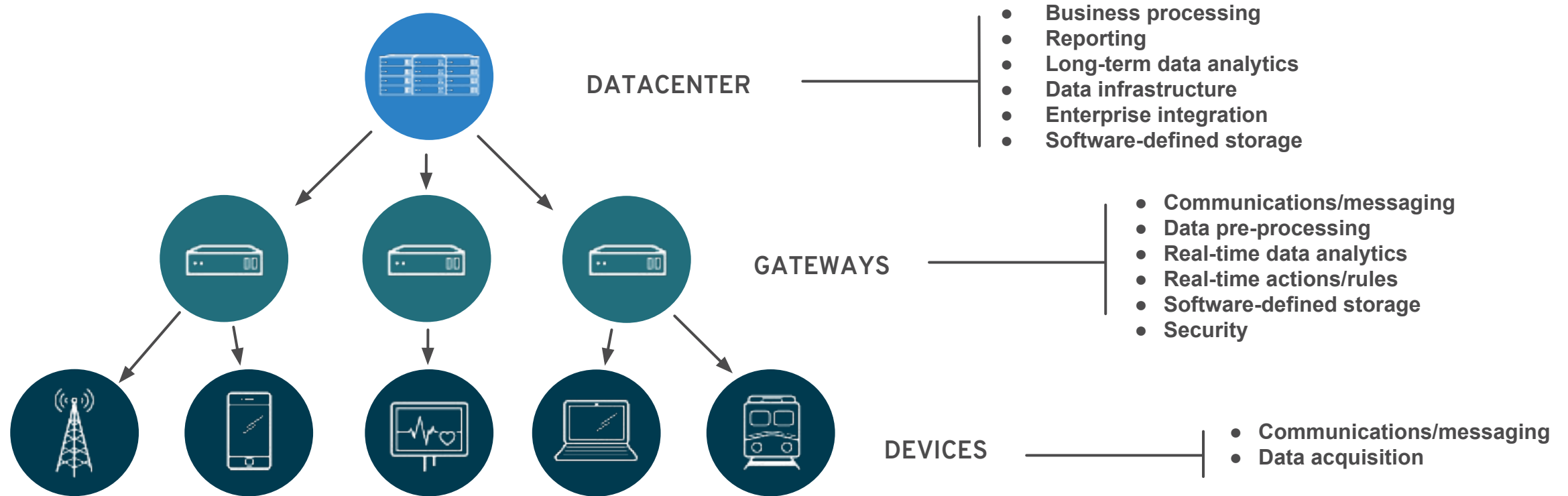
RH0064-3

DATA-DRIVEN IoT

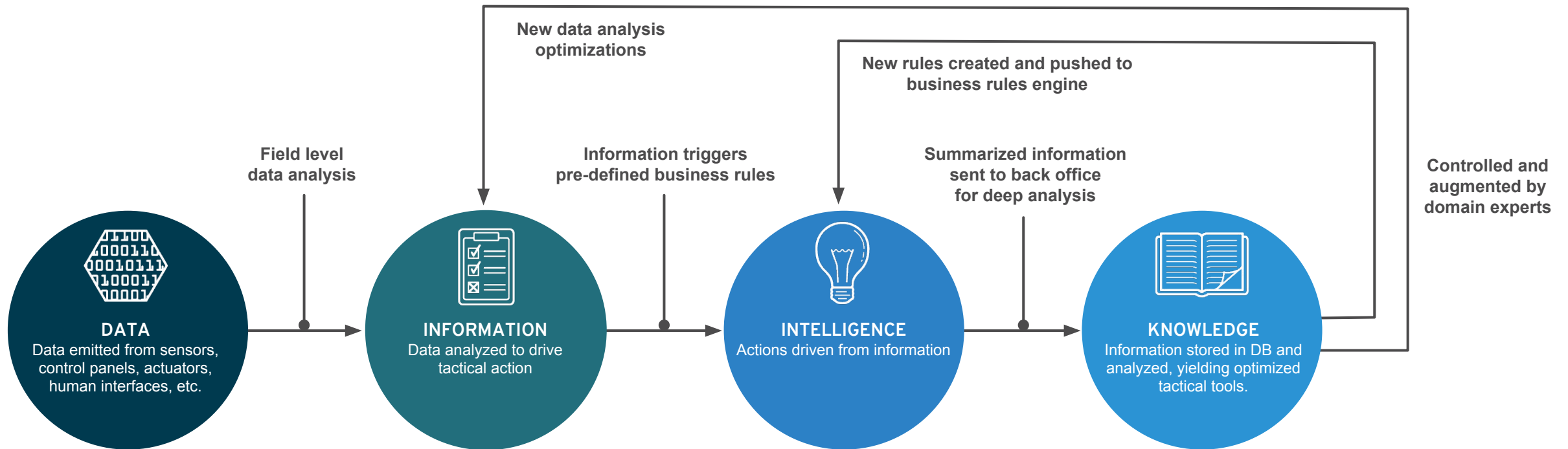


Devices are the eyes and ears of the intelligent system, not its brain.

ENTERPRISE IoT ARCHITECTURE



IoT INFORMATION LIFECYCLE



ADDRESSING ENTERPRISE IOT REQUIREMENTS

Integrating IoT operating technology, data management, analytics, and applications in an open end-to-end IoT architecture

Operational Technology (OT)

- Device Management
- Industrial protocols
- OT Middleware
- Intelligent gateways
- MQTT co-inventors
- OT security



Information Technology (IT)

- Messaging & Integration
- Business Rules & CEP
- Open Hybrid Platform-as-a-Service
- Enterprise Linux Platform
- IT security



Data Management & Analytics

- Enterprise Data Mgmt
- Persistent Data Storage
- Big Data Processing & Analytics
- Real-Time Analytics
- Machine Learning
- Data Security & Compliance



Enterprise IoT open source community



ECLIPSE IoT OVERVIEW

“We believe the best way to support this complex environment is to base our commercial IoT platform, the Bosch IoT Suite, on open source components and open standards. These projects establish a horizontal open technology for IoT and provide the technical breeding grounds for successful business ecosystems.”

- Dr. Stefan Ferber, VP of Engineering,
Bosch Software Innovations



2.4

million
lines of code



30*

projects



250+

developers



130K

monthly
visitors



BOSCH

Invented for life

cloudera



EUROTECH

Imagine. Build. Succeed.



redhat



Google

SIEMENS



SAP



cloudera



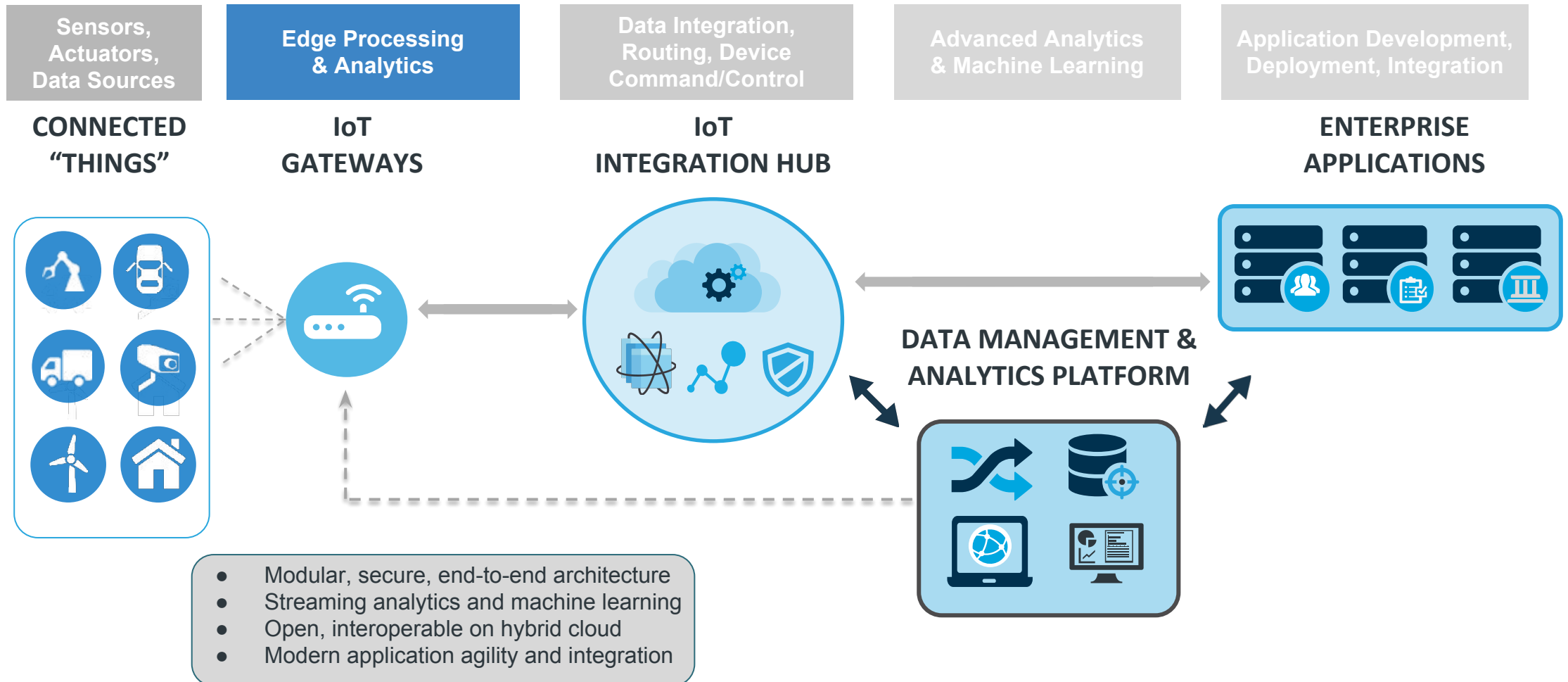
EUROTECH



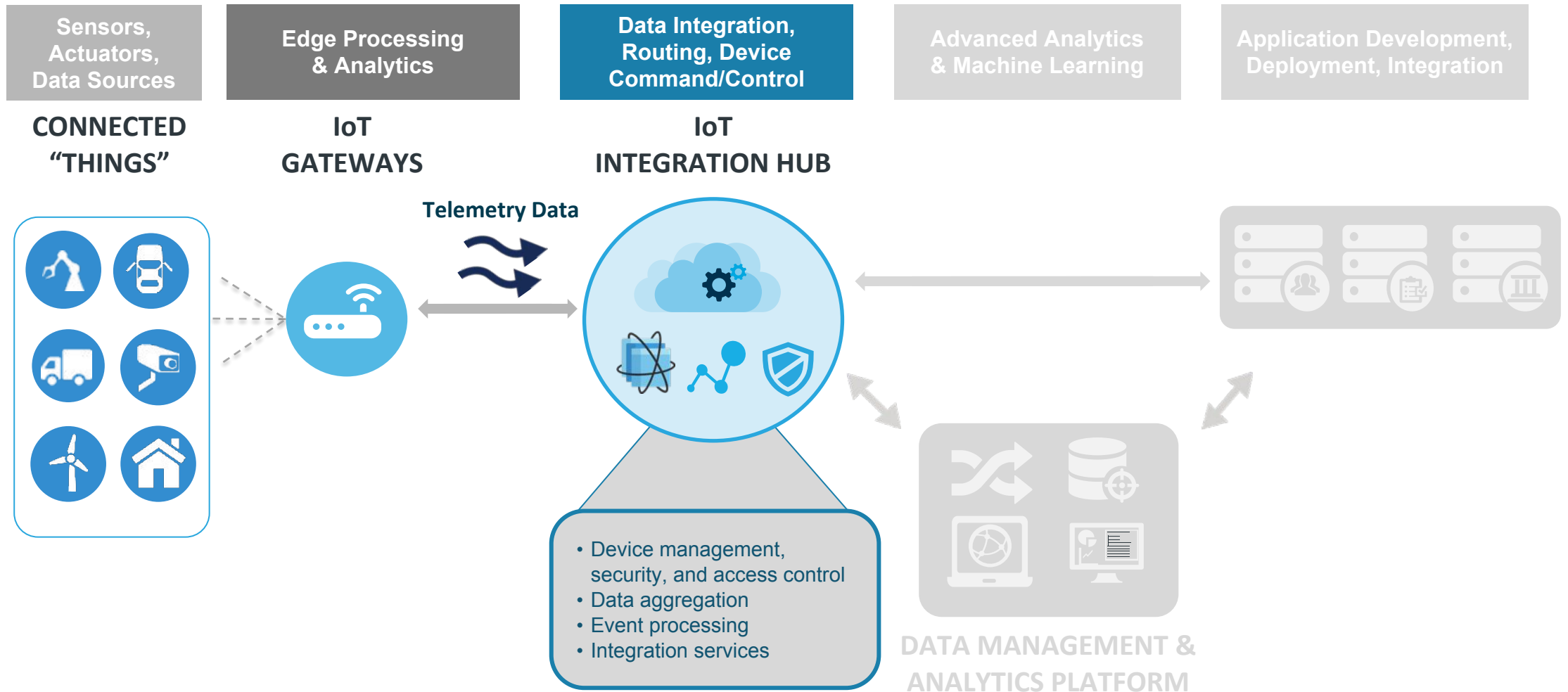
redhat

OPEN END-TO-END IoT ARCHITECTURE

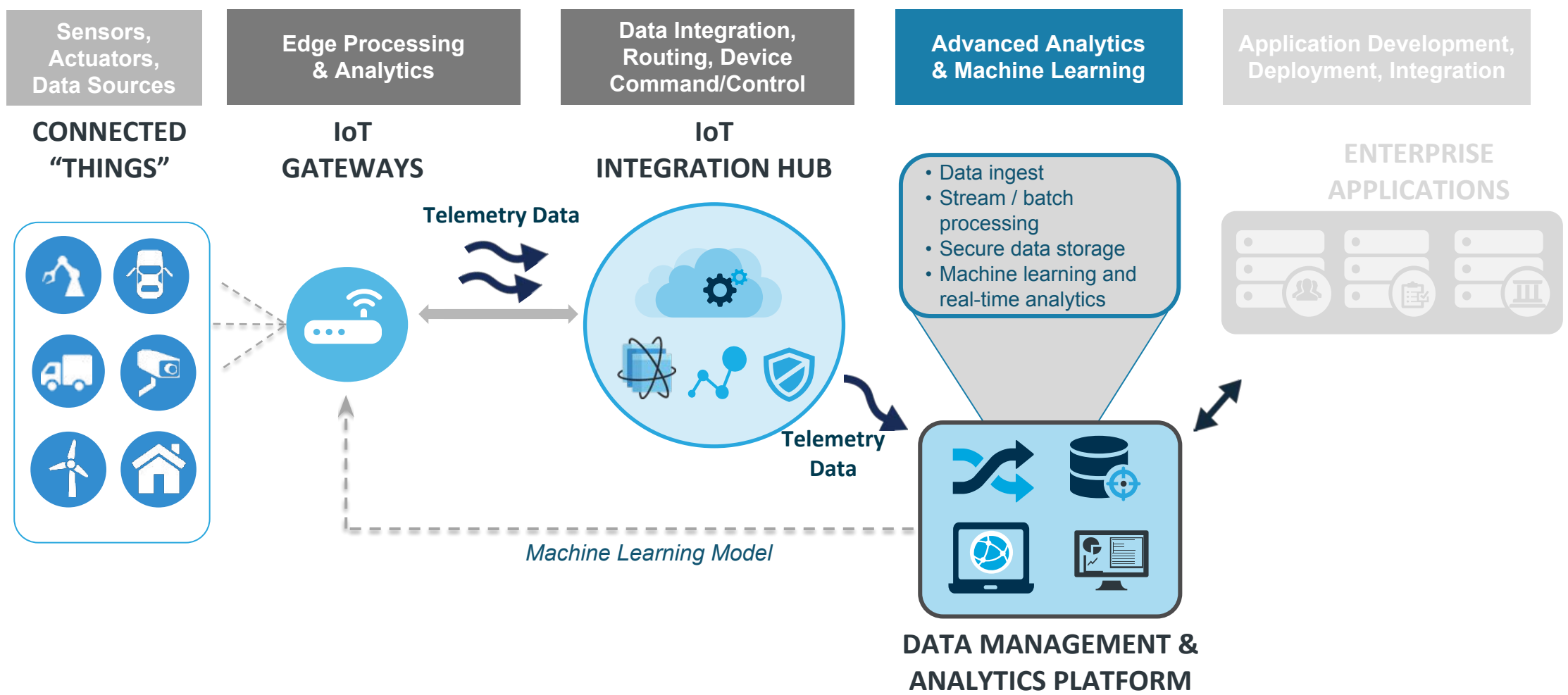
Integrating IoT operating technology, data management, analytics, and applications



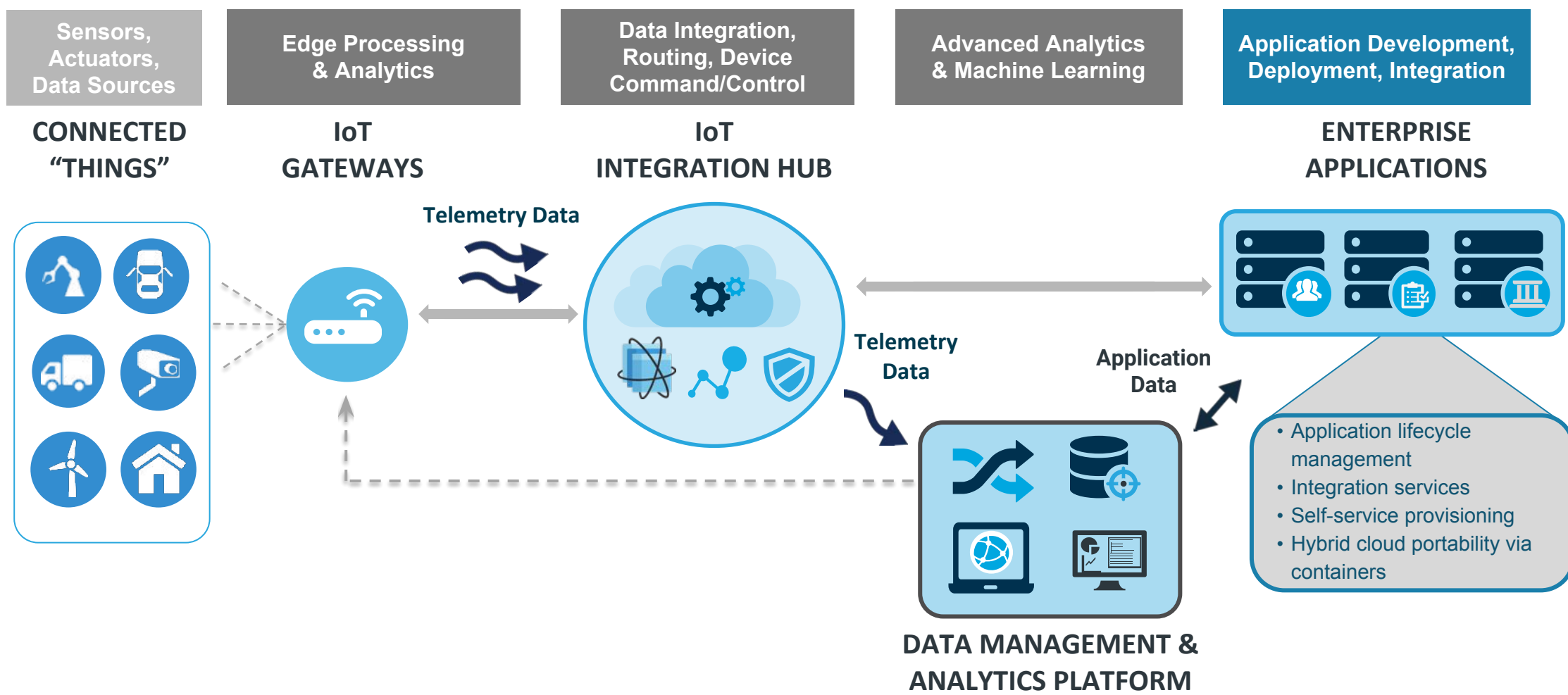
DATA INTEGRATION, ROUTING, DEVICE COMMAND/CONTROL



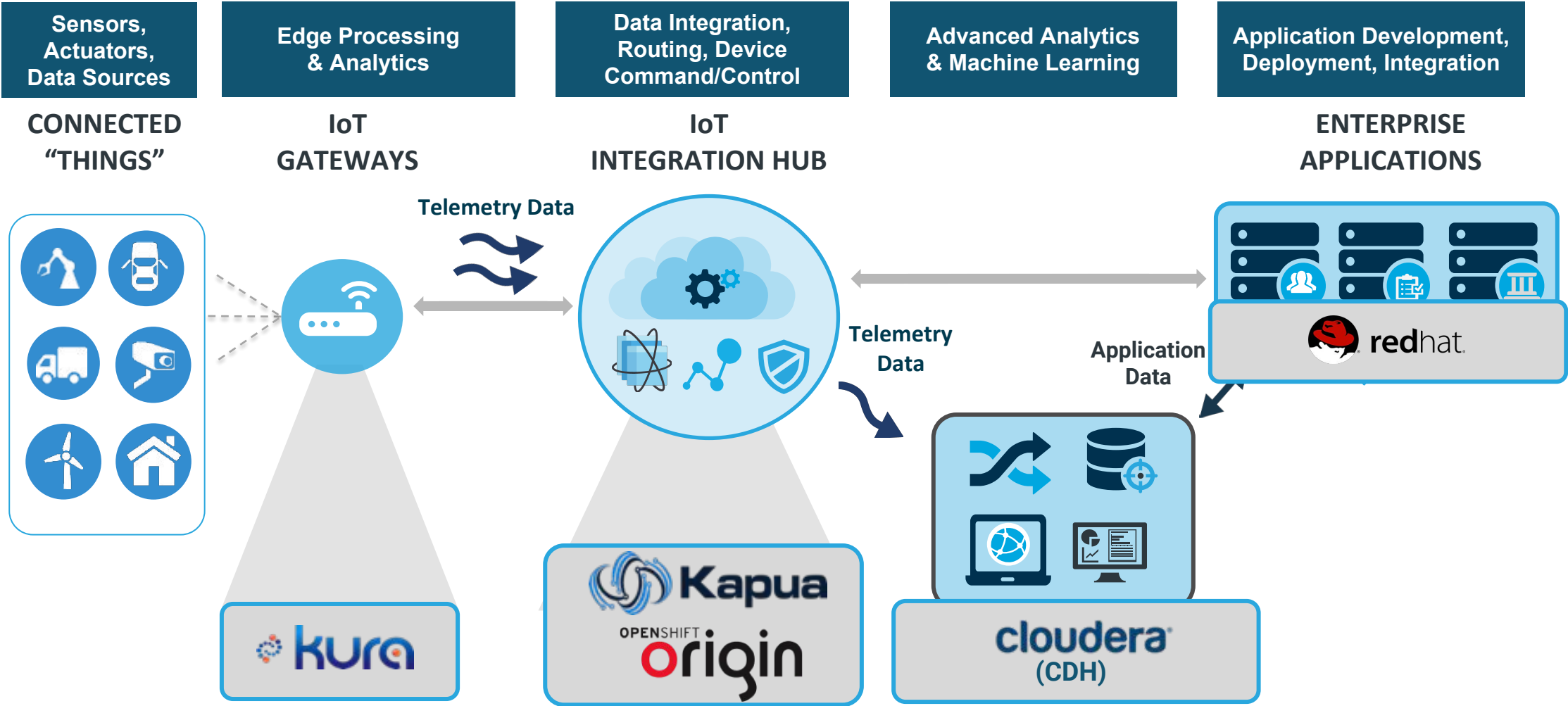
ADVANCED ANALYTICS AND MACHINE LEARNING



APPLICATION DEVELOPMENT, DEPLOYMENT, INTEGRATION



AN END-TO-END OPEN SOURCE ARCHITECTURE FOR IoT



Kura Project (<http://www.eclipse.org/kura/>)

Eclipse Kura™ is an extensible open source IoT Edge Framework based on Java/OSGi.

Kura offers API access to the hardware interfaces of IoT Gateways (serial ports, GPS, watchdog, GPIOs, I2C, etc.). It features ready-to-use field protocols (including Modbus, OPC-UA, S7), an application container, and a web-based visual data flow programming to acquire data from the field, process it at the edge, and publish it to leading IoT Cloud Platforms through MQTT connectivity.



Eclipse Kura offers a platform that can live at the boundary between the private device network and the local network, public Internet or cellular network providing a manageable and intelligent gateway for that boundary capable of running applications that can harvest locally gathered information and deliver it reliably to the cloud.

Kapua Project (<https://www.eclipse.org/kapua/index.php>)

Kapua is a modular IoT cloud platform to manage and integrate devices and their data. Kapua provides a solid integrated foundation of IoT services for any IoT application.

Connect IoT devices to Kapua via MQTT and other protocols.

Manage device applications, configurations, and resources and enable remote administration.

Store and index the data published by IoT devices for quick analysis and visualization into dashboards.

Integrate Kapua services with IT applications through flexible message routing and REST API.



Cloudera CDH (Cloudera Distribution including Apache Hadoop)

CDH is Cloudera's 100% open source platform distribution, including Apache Hadoop and built specifically to meet enterprise demands.

CDH delivers everything you need for enterprise use right out of the box. By integrating Hadoop with more than a dozen other critical open source projects, Cloudera has created a functionally advanced system that helps you perform end-to-end Big Data workflows.



Red Hat OpenShift Origin (<https://www.openshift.org/>)

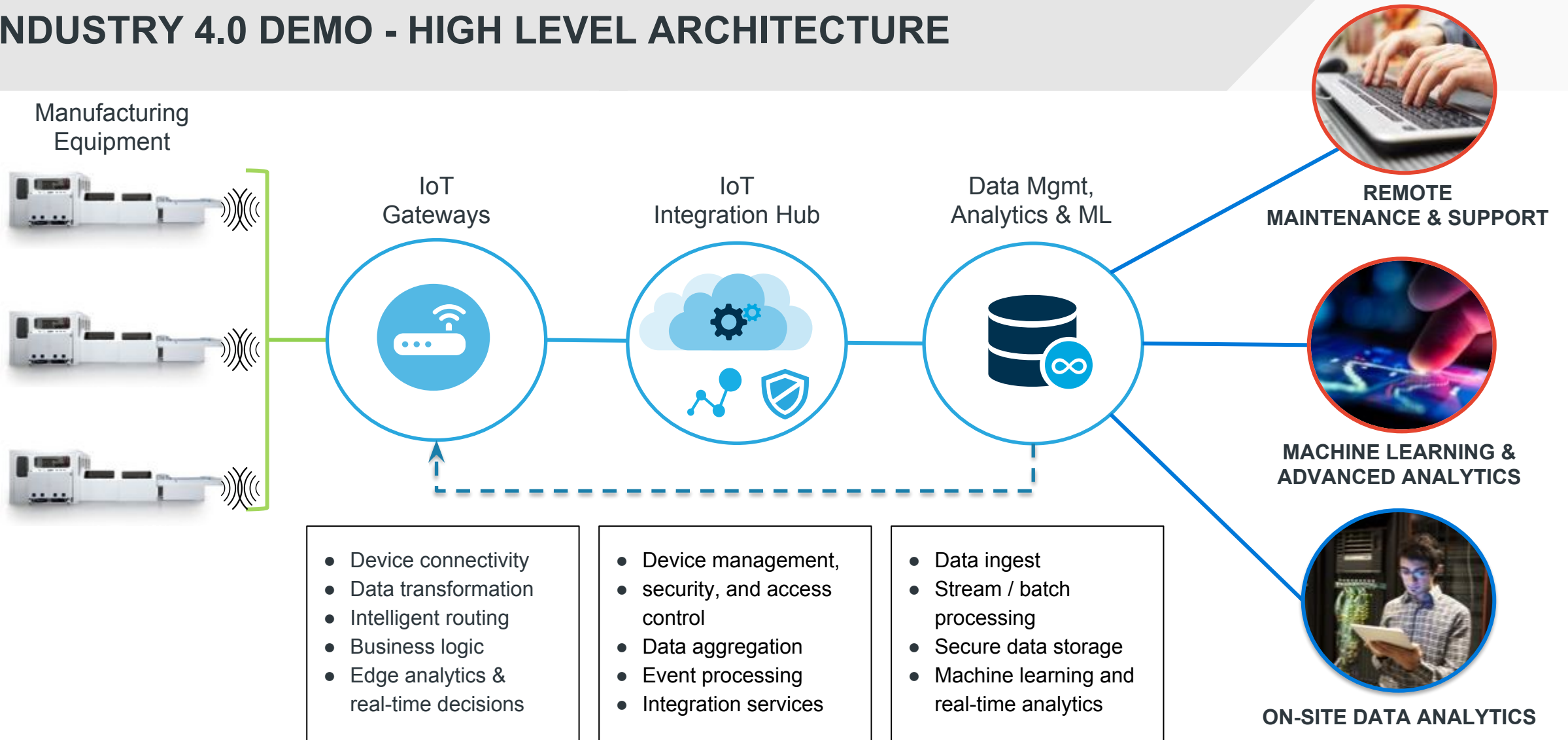
Origin is the upstream community project that powers OpenShift.

Built around a core of container packaging and Kubernetes container cluster management, Origin is also augmented by application lifecycle management functionality and DevOps tooling.

Origin provides a complete open source container application platform.



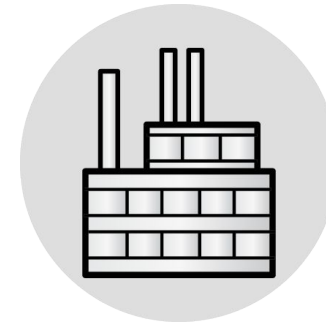
INDUSTRY 4.0 DEMO - HIGH LEVEL ARCHITECTURE



INDUSTRY 4.0 DEMO - BUSINESS CHALLENGES

The scenario

- A manufacturing company is incurring **unplanned machine downtime** leading to increased maintenance and costs.
- Lack of a **real-time view on asset condition** makes it difficult to predict downtime and therefore optimize factory productivity.
- Lack of **business and operations visibility** due to inability to correlate real-time with historic data results.

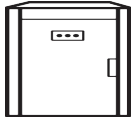


The goal

- Optimize production factory capacity by **scheduling planned maintenance windows**.
- Predict failures on connected machines to **reduce downtimes & increase margins**.

INDUSTRY 4.0 DEMO - USE CASE

Optimize production factory capacity by scheduling planned maintenance windows;
Predict failures on connected machines to reduce downtimes & increase margins



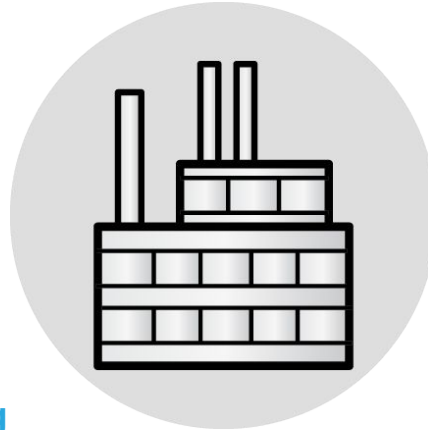
Machine Telemetry

- Current
- Temperature
- Noise
- Speed
- Vibration
- Voltage



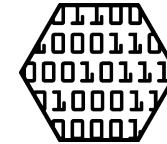
Condition based monitoring

- Predictive & preventative
- Analytics & ML at the edge
- Automatic alerts & decisions



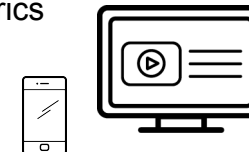
Centralized Data Mgt & Analytics

- Historical archive
- Iterative model training
- Analytics & ML at the core



Real-time dashboards & task lists

- C-level business metrics
- Factory operators
- Field technicians



INDUSTRY 4.0 DEMO - SOLUTION AND BENEFITS

Proactively handle risk through condition based monitoring

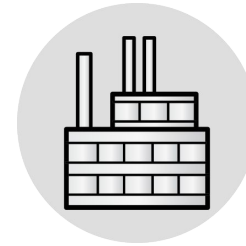
The solution

Place sensors on all critical machines to:

- collect telemetry data
- track conditions in real-time

Run analytics and ML models at the edge with rules to:

- react faster to unpredicted errors with automatic alerts
- shut down machines or entire lines to minimize damage.
- avoid dependency with centralized operations location to reduce latency and optimize communications



The benefits

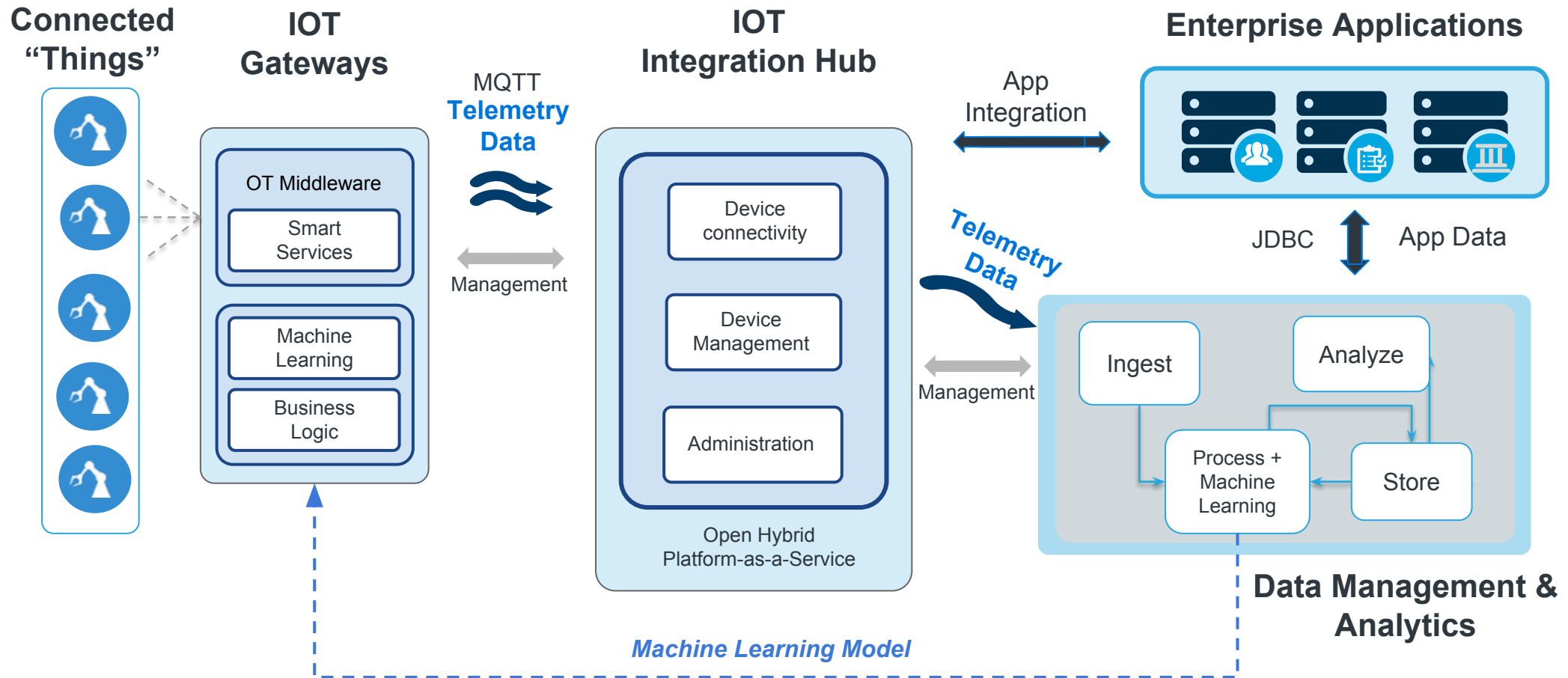
Manage data holistically from a centralized location and **run analytics and ML models** to provide additional insight to field technicians, operators and HQ.

- determine when machines are not performing optimally
- automatically schedule planned maintenance windows
- learn from data gathered over time to improve ability to detect failures before they happen.

Make tracking data available to executives, investors and partners.

- create reporting capabilities to provide visibility to factory utilization, capacity, revenue, margin and costs.

INDUSTRY 4.0 DEMO - ARCHITECTURE: FUNCTIONAL



Data Projects

Apache Spark is a unified analytics engine for large-scale data processing.



Apache Kafka is a distributed streaming platform. A streaming platform has three key capabilities:

- Publish and subscribe to streams of records, similar to a message queue or enterprise messaging system.
- Store streams of records in a fault-tolerant durable way.
- Process streams of records as they occur.



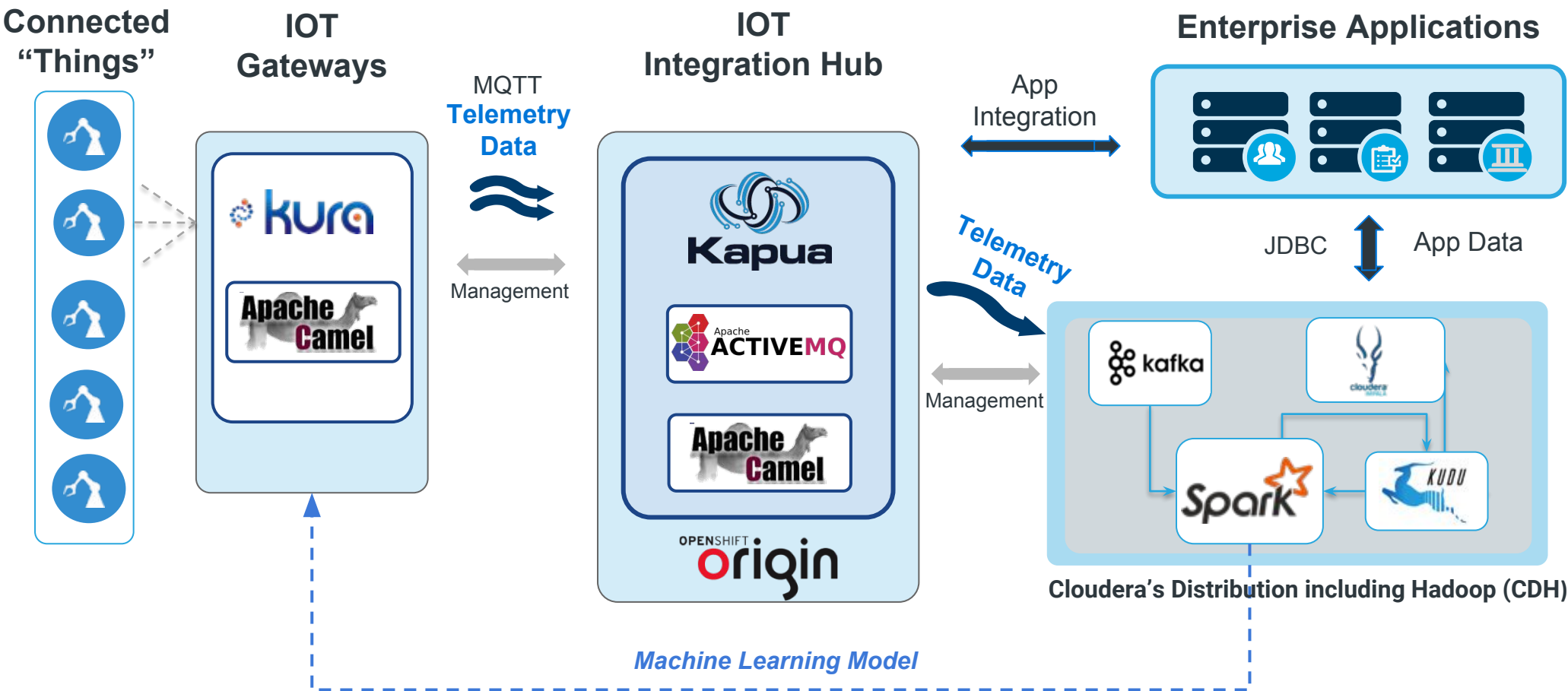
Kudu is storage for fast analytics on fast data providing a combination of fast inserts and updates alongside efficient columnar scans to enable multiple real-time analytic workloads across a single storage layer.



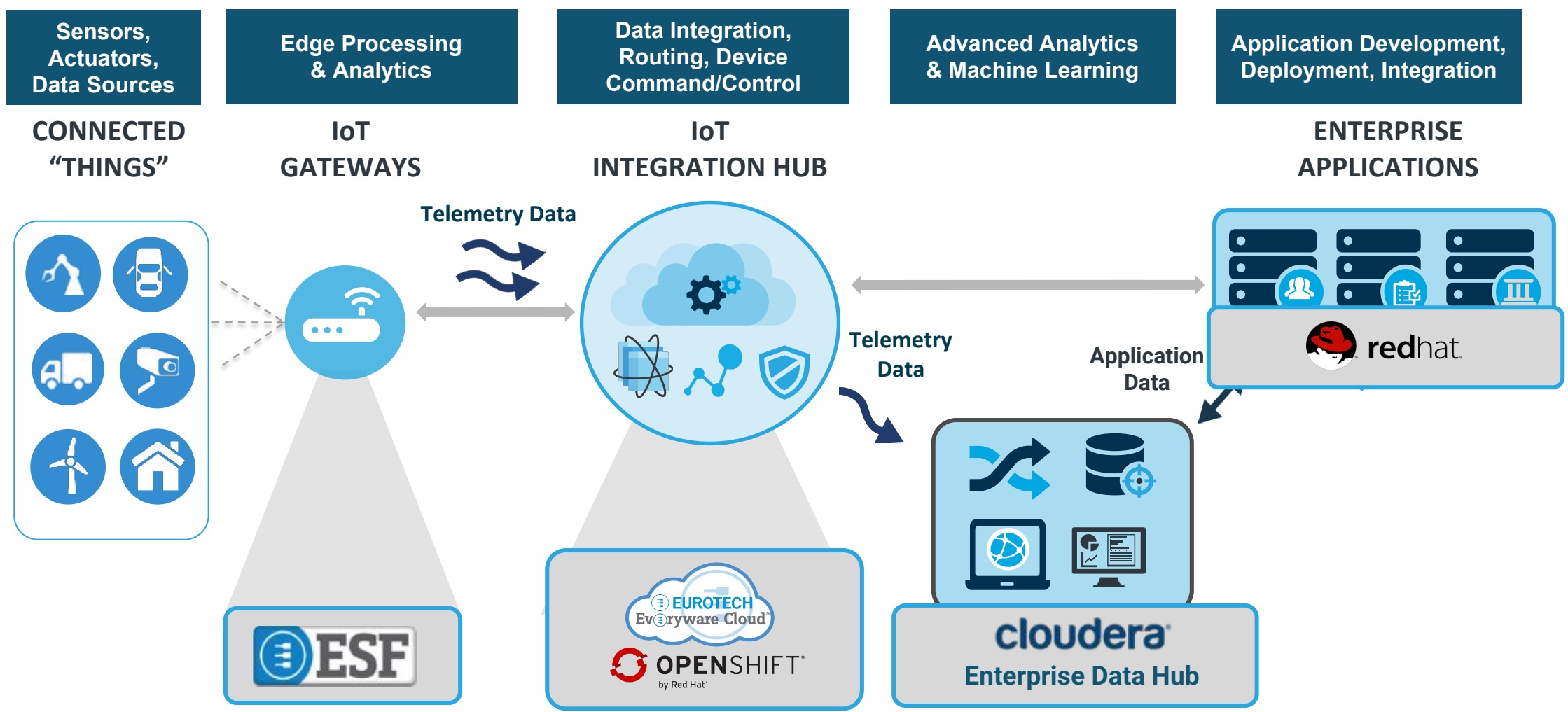
Apache Impala is an open source massively parallel processing (MPP) SQL query engine for data stored in a computer cluster running Apache Hadoop.



INDUSTRY 4.0 DEMO - ARCHITECTURE: PROJECTS



ENTERPRISE GRADE PRODUCTS FOR END-TO-END ARCHITECTURE



LEARN MORE

Integrating IOT Operating Technology, Data Management, Analytics, and Applications

Solution deep-dive

At your request, Red Hat will present a deep dive on the solution and its supporting technologies.

Contact Red Hat at iotquestions@redhat.com.

OR

Visit

<https://engage.redhat.com/iot-live-demo-201709141110>

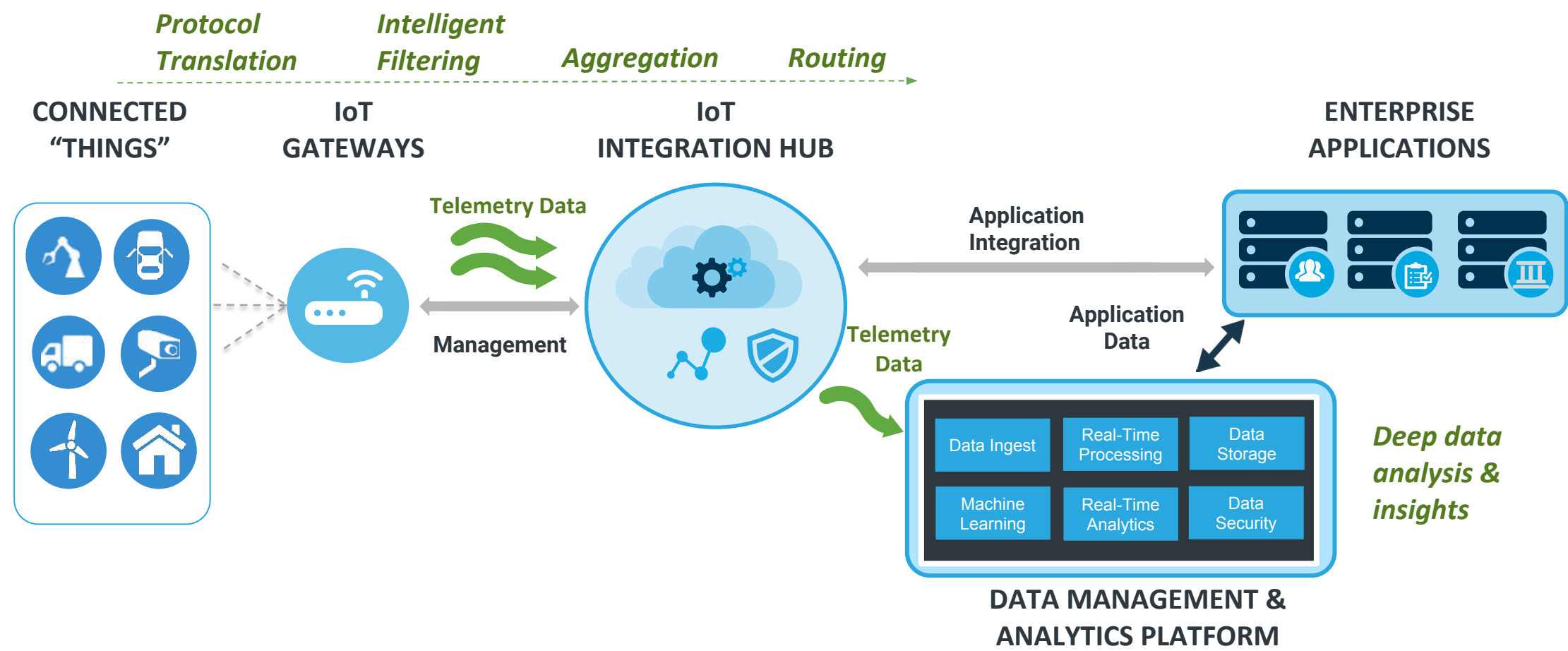
to request a deep dive

THANK YOU

Scott Seighman
Solutions Architect / Red Hat
sseighma@redhat.com

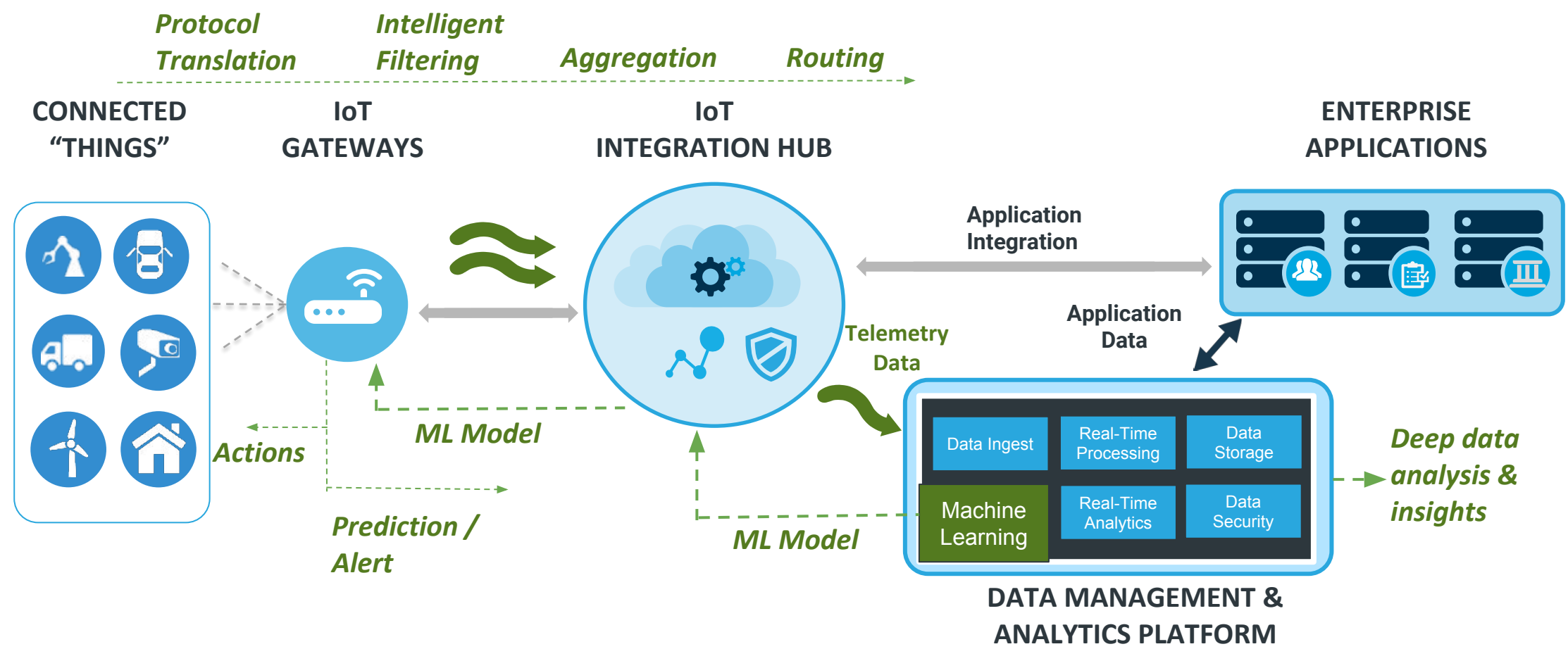
END-TO-END ANALYTICS

Data flow to derive deep business insights and actionable intelligence



END-TO-END ANALYTICS

Data flow to derive deep business insights and actionable intelligence



KUDU Project

Kudu is storage for fast analytics on fast data—providing a combination of fast inserts and updates alongside efficient columnar scans to enable multiple real-time analytic workloads across a single storage layer.

Fast analytics on fast data

Kudu can provide both inserts and updates, in addition to efficient columnar scans, enabling the Apache Hadoop™ ecosystem to tackle new analytic workloads.

Simplified architecture

Kudu fills the gap between HDFS and Apache HBase formerly solved with complex hybrid architectures, easing the burden on both architects and developers.

Ecosystem integration

Kudu was specifically built for the Hadoop ecosystem, allowing Apache Spark™, Apache Impala, and MapReduce to process and analyze data natively. Additionally, Kudu tables can be joined with data in HDFS or HBase.

