An Open-Source Streaming Machine Learning and Real-Time Analytics Architecture

Using an IoT example

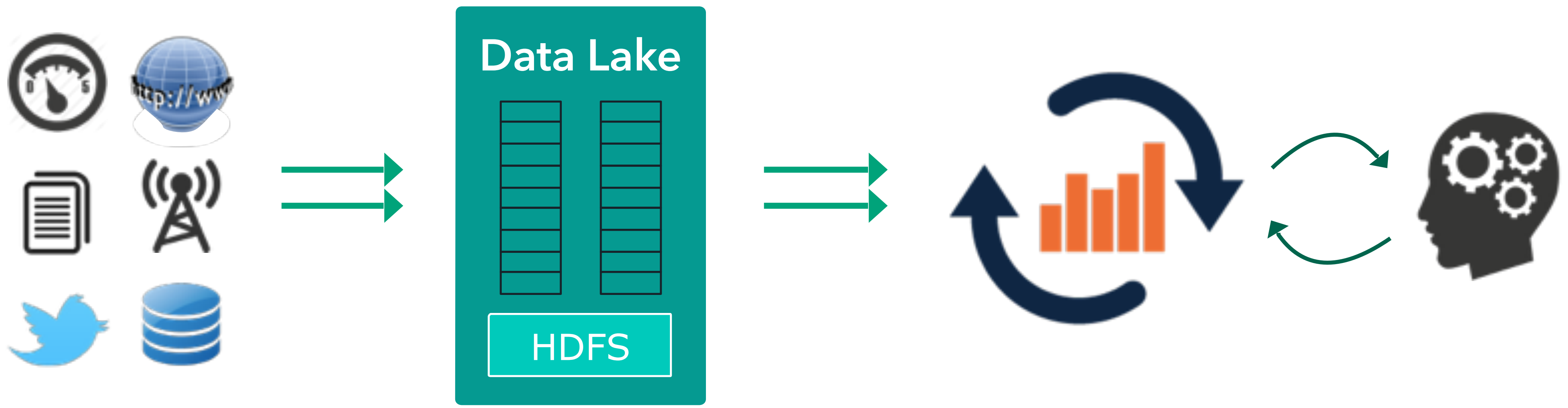
(incubating)

(incubating)

Fred Melo
@fredmelo_br

William Markito
@william_markito

Traditional Data Analytics - Limitations



No real-time information

ETL based

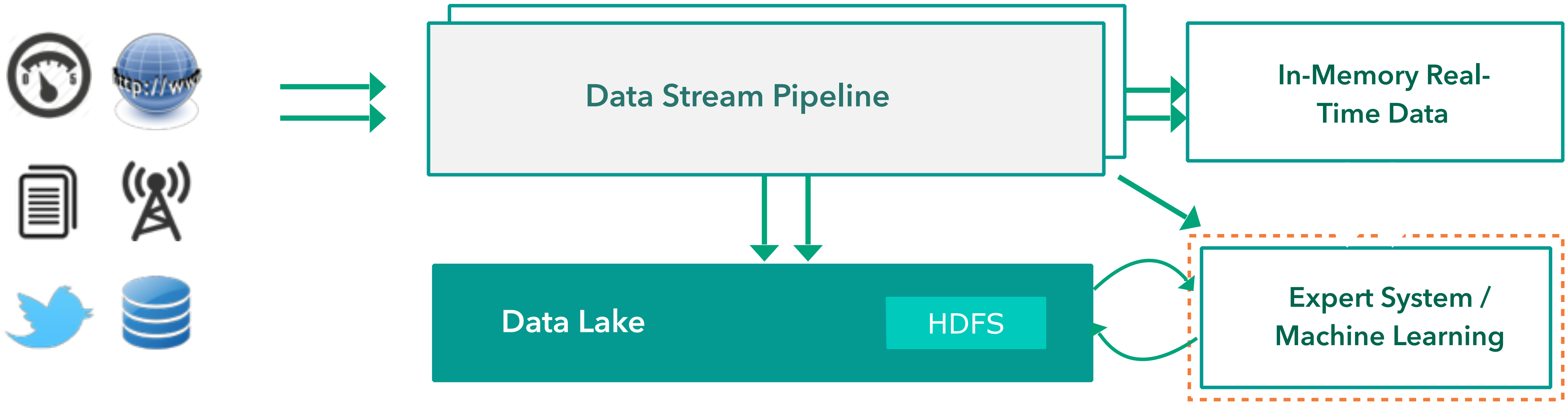
Data-source specific

Hard to change

Labor intensive

Inefficient

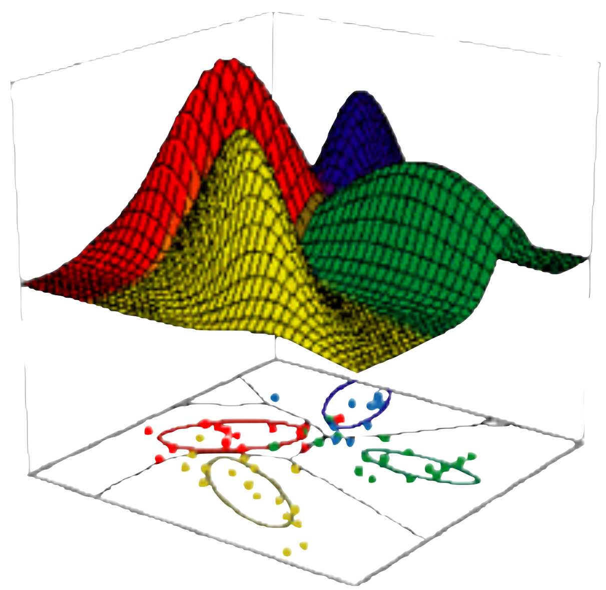
Stream-based, Real-Time Closed-Loop Analytics



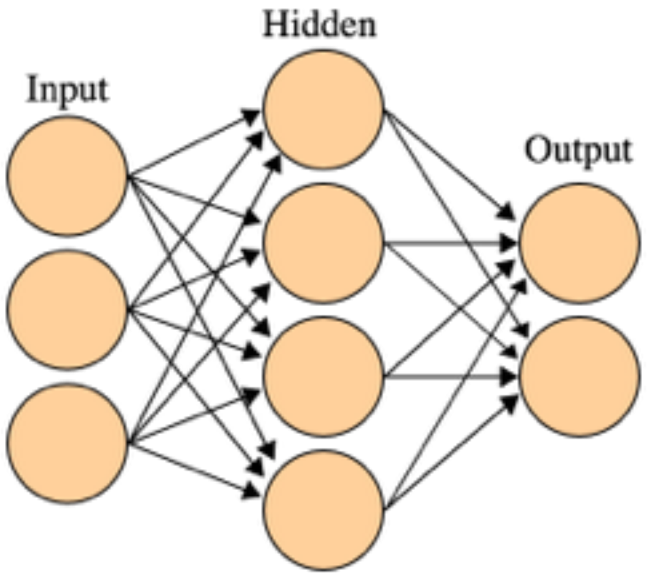
Multiple Data Sources
Real-Time Processing
Store Everything

Continuous Learning
Continuous Improvement
Continuous Adapting

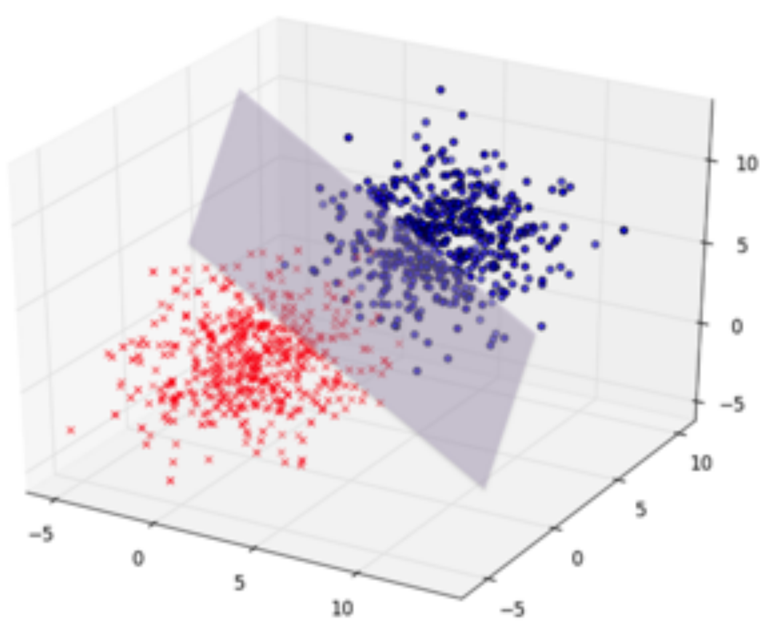
Machine Learning and Smart Systems



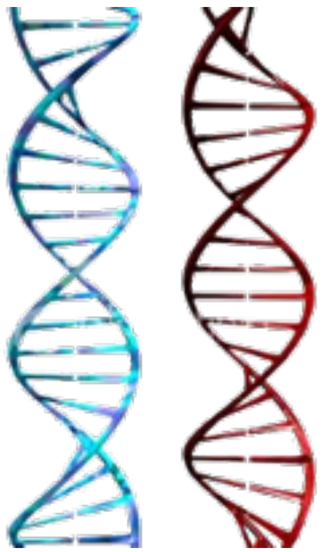
Bayesian Methods



Neural Networks



Classifiers



Genetic Algorithms

A Streaming Machine Learning for IoT Example

Predictive Maintenance Scenario

Sensor Data

Real-Time

Evaluates LIVE DATA

“According to historical trends, there’s an 80% chance this equipment would fail in the next 12 hours"

Live data becomes historical over time

Smart System

Historical

Learns with HISTORICAL TRENDS

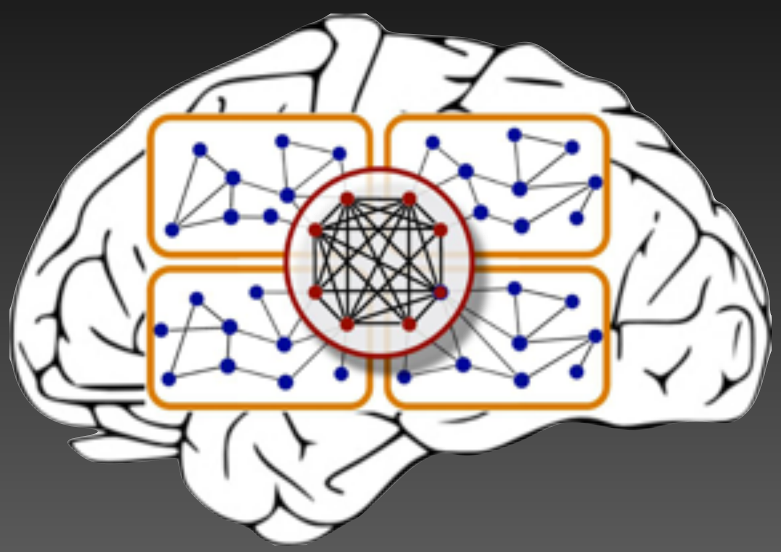
"How were the temperature and vibration sensors reading when the latest failures happened? "

Streaming Machine Learning

Info

Analysis

Machine Learning

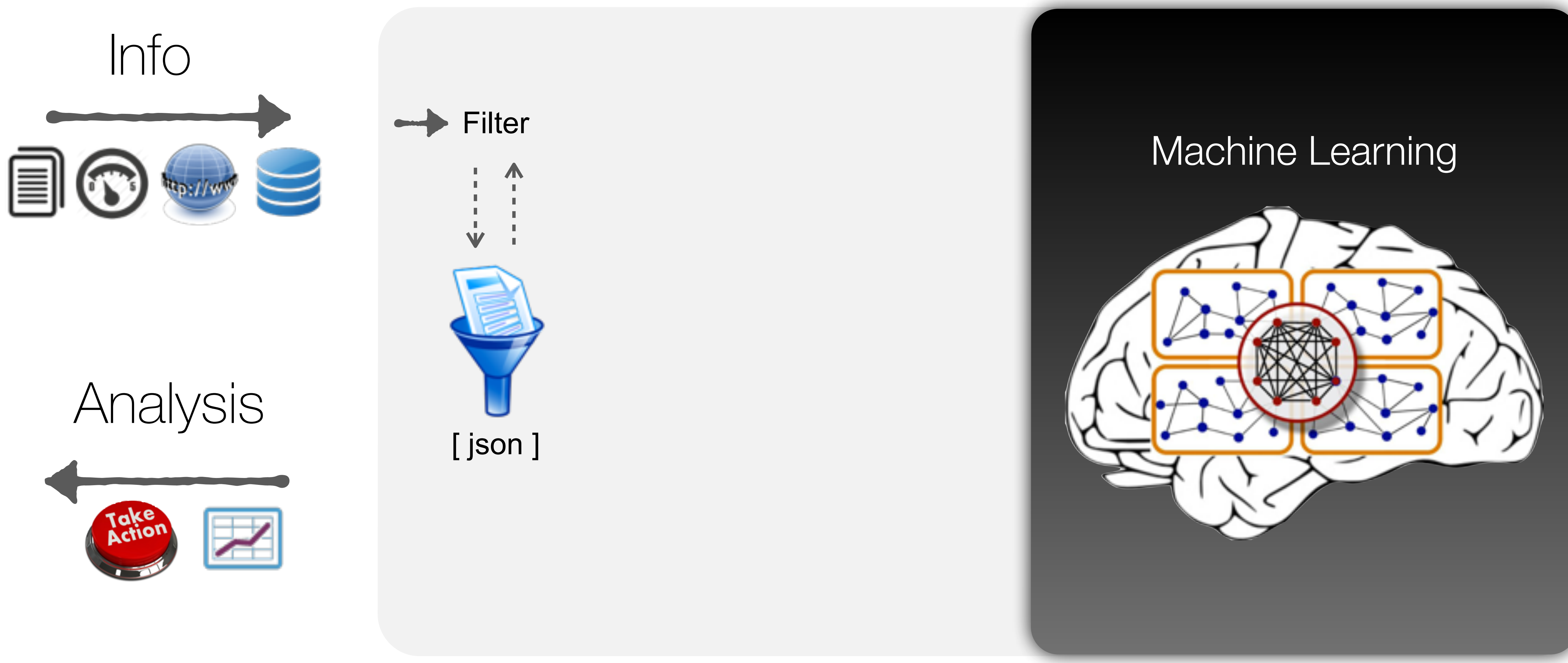


Look at past trends
(for similar input)

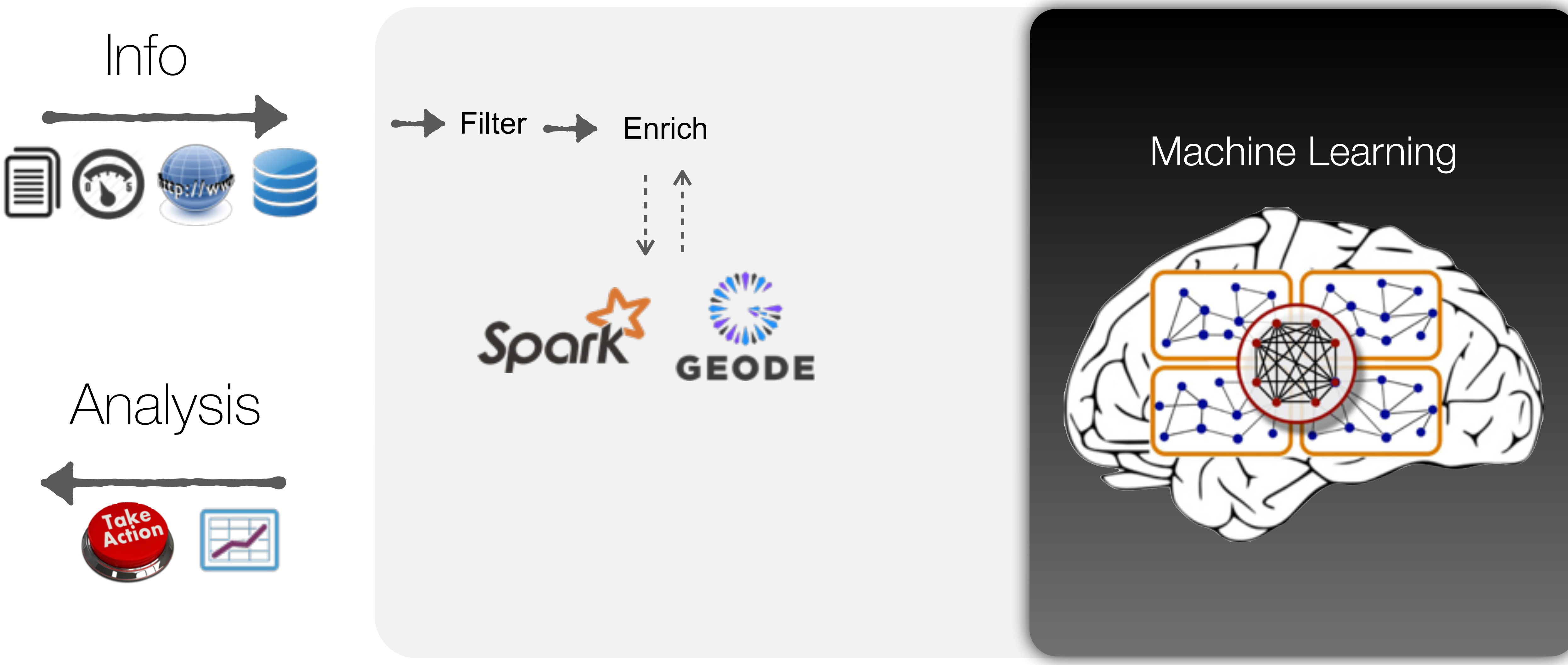
Evaluate current input

Score / Predict

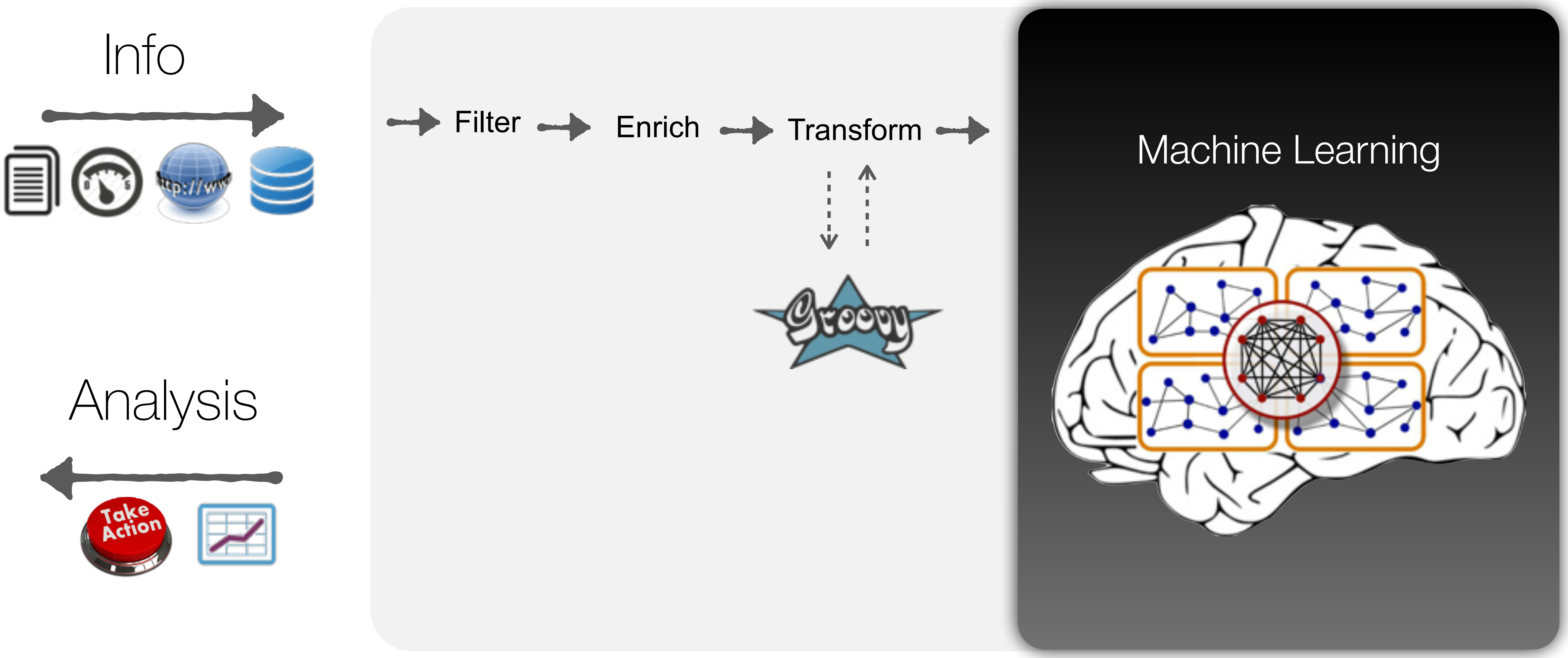
Streaming Machine Learning



Streaming Machine Learning



Streaming Machine Learning



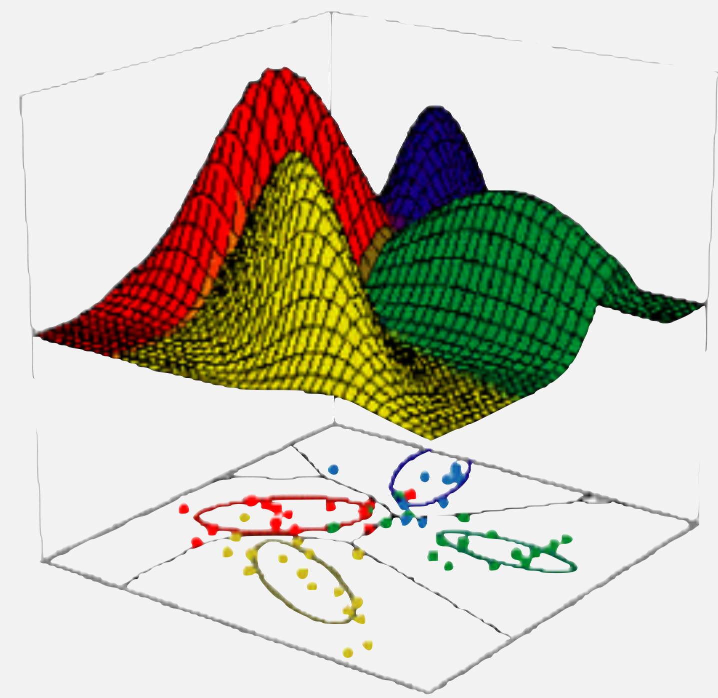
Streaming Machine Learning

Info

Analysis

Filter → Enrich → Transform →

ML Model

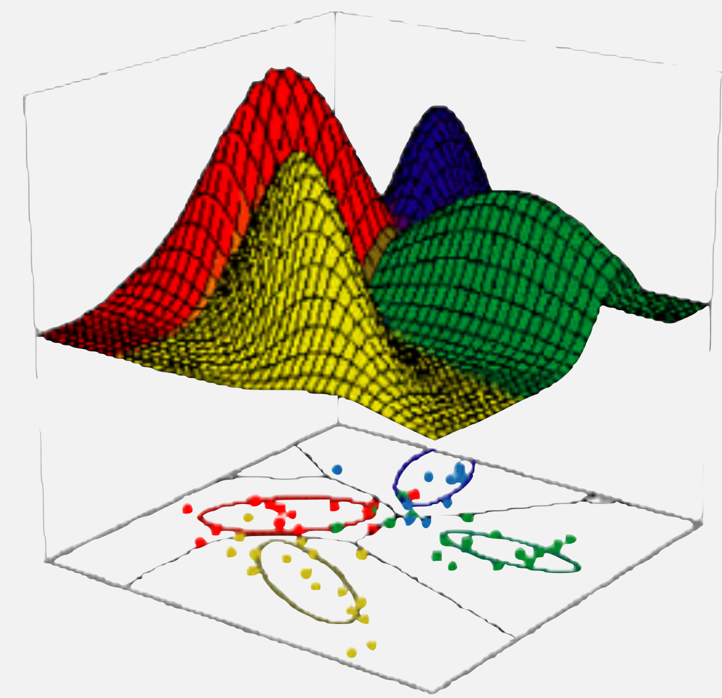


Streaming Machine Learning

Info

Filter → Enrich → Transform →

ML Model

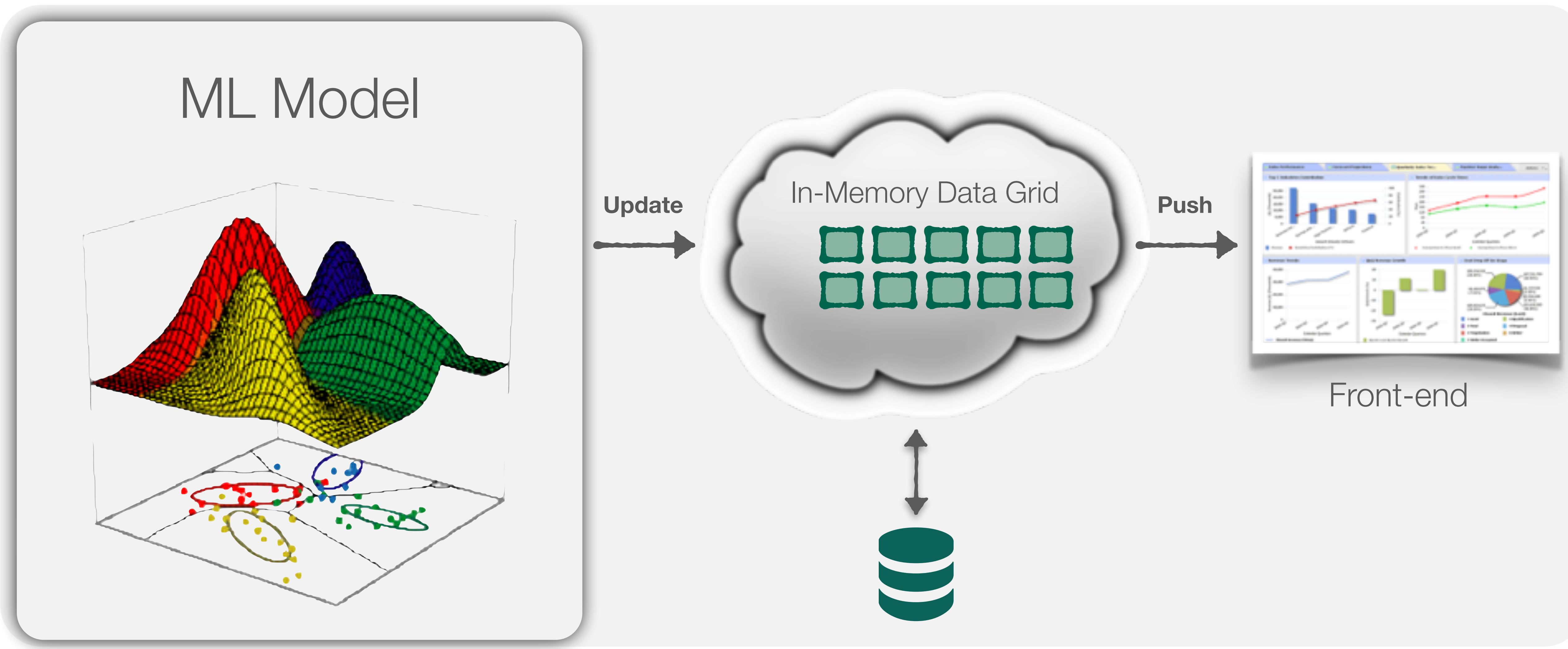


Analysis

← Transform ←

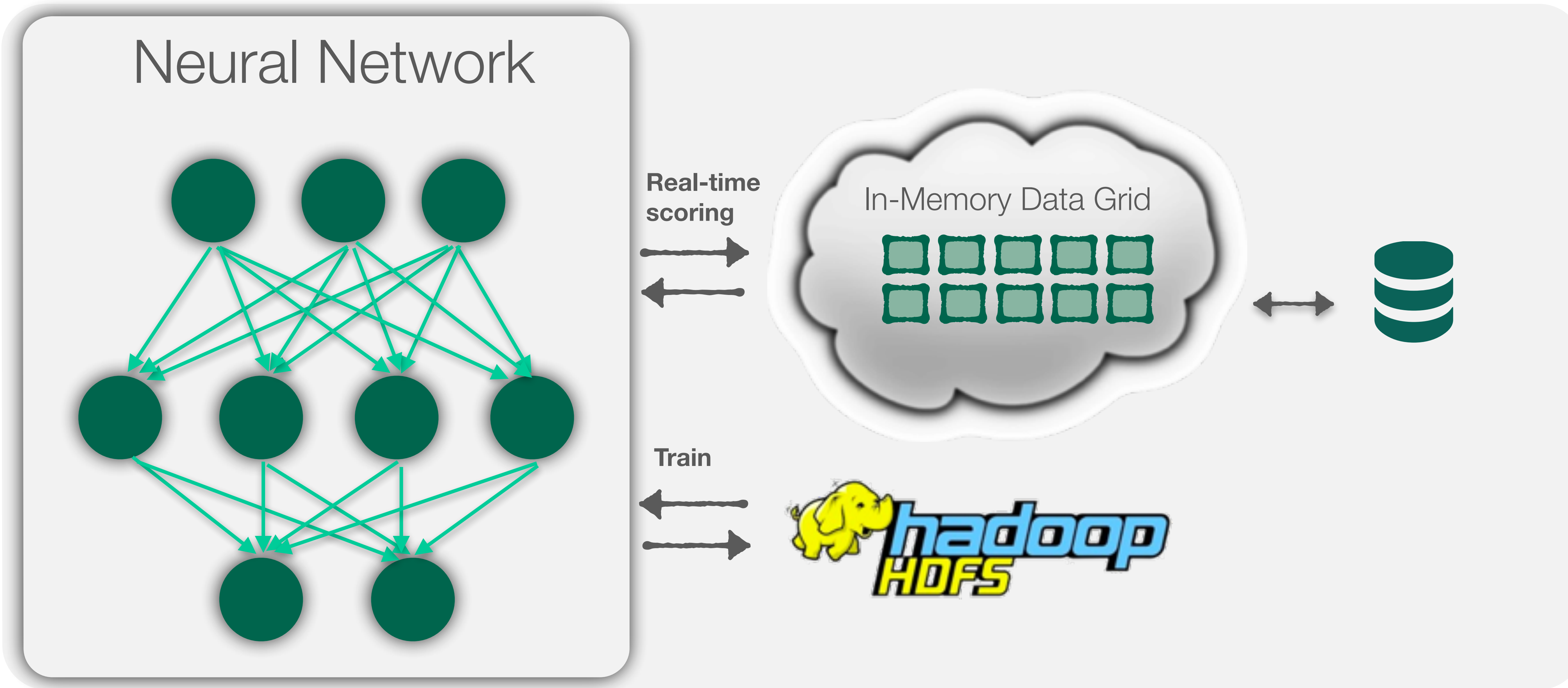
APACHE GEODE

Streaming Machine Learning

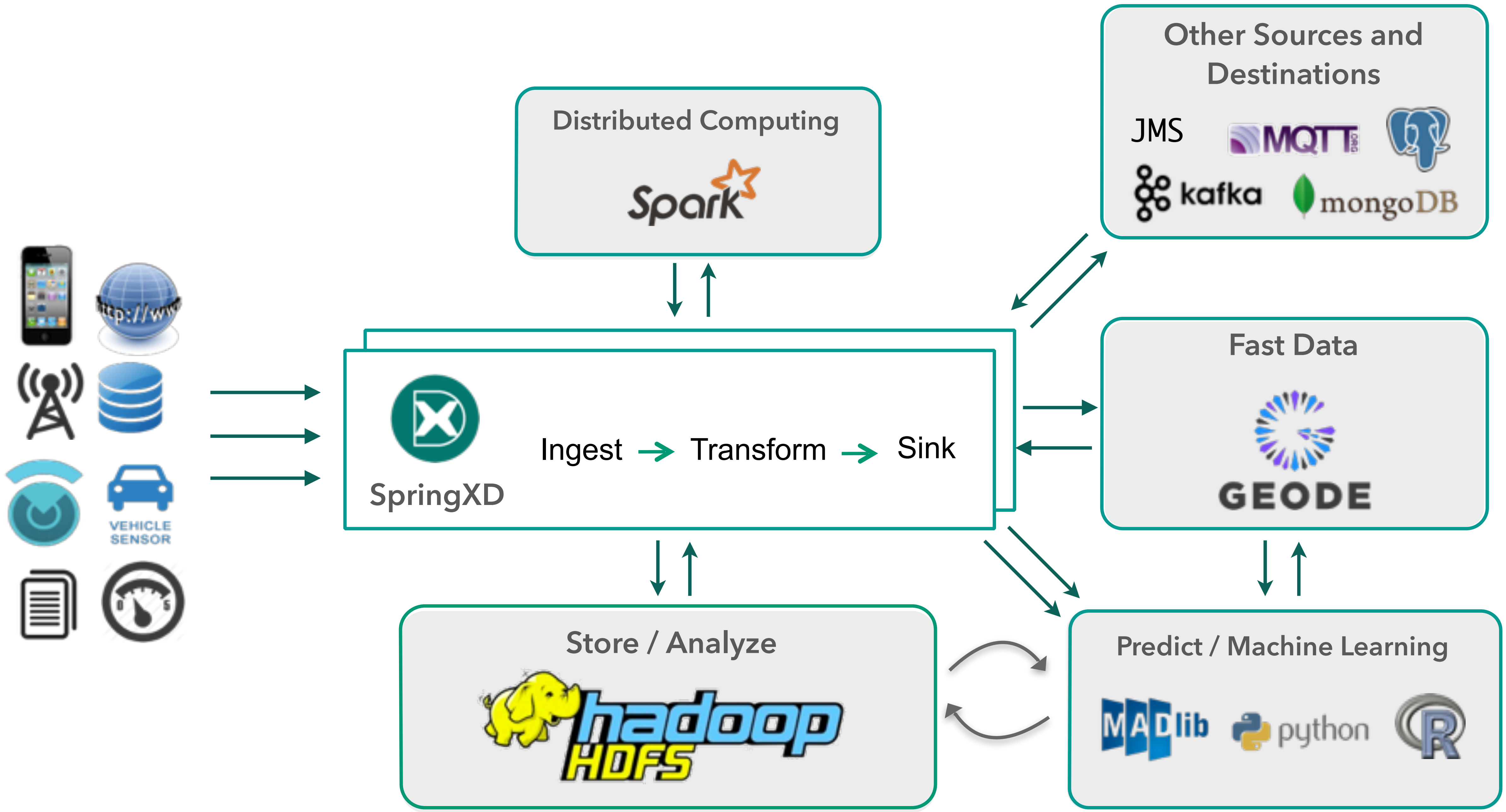


Streaming Machine Learning

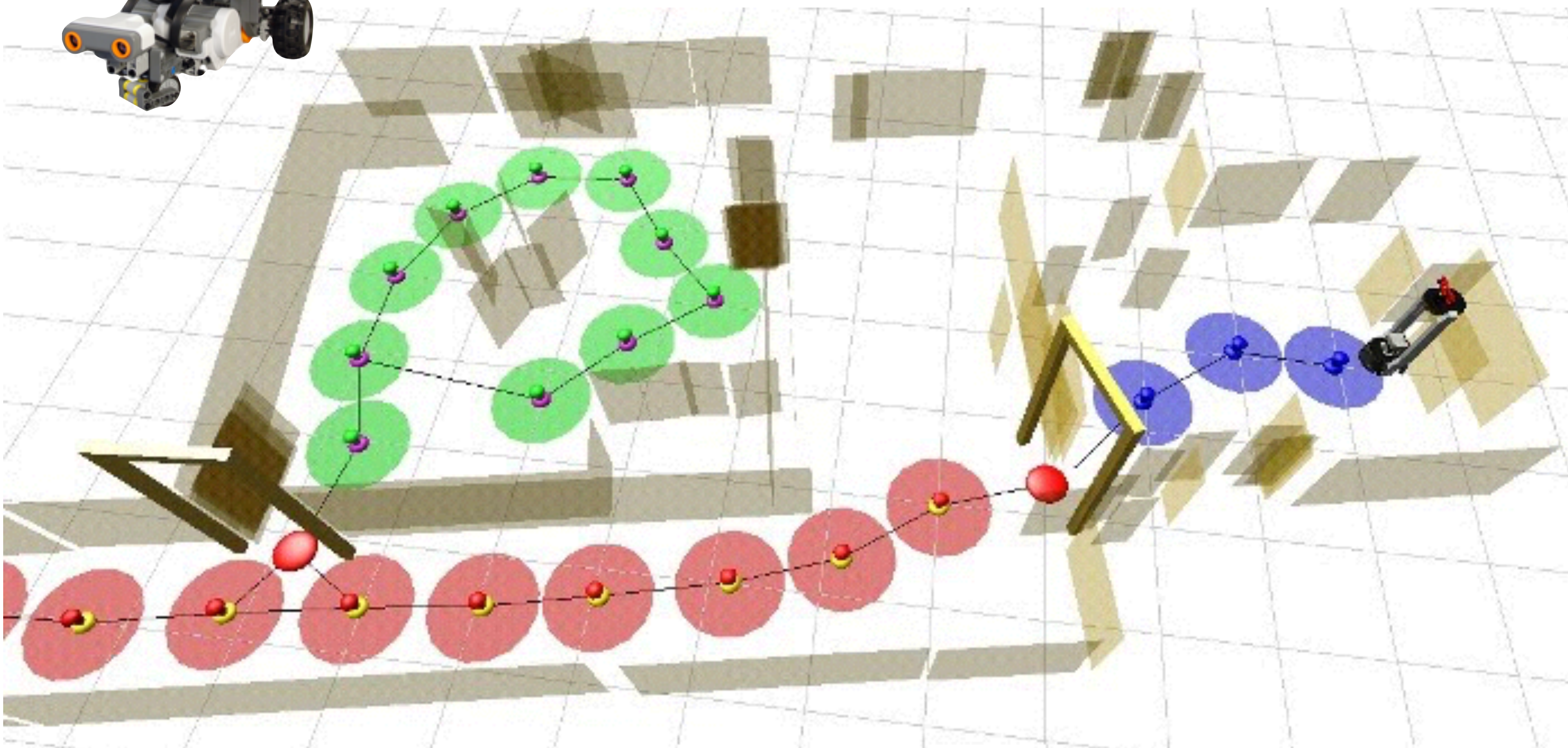
Supervised Learning Example



A Streaming Machine Learning Reference Architecture



Indoors Localization - Applied Example

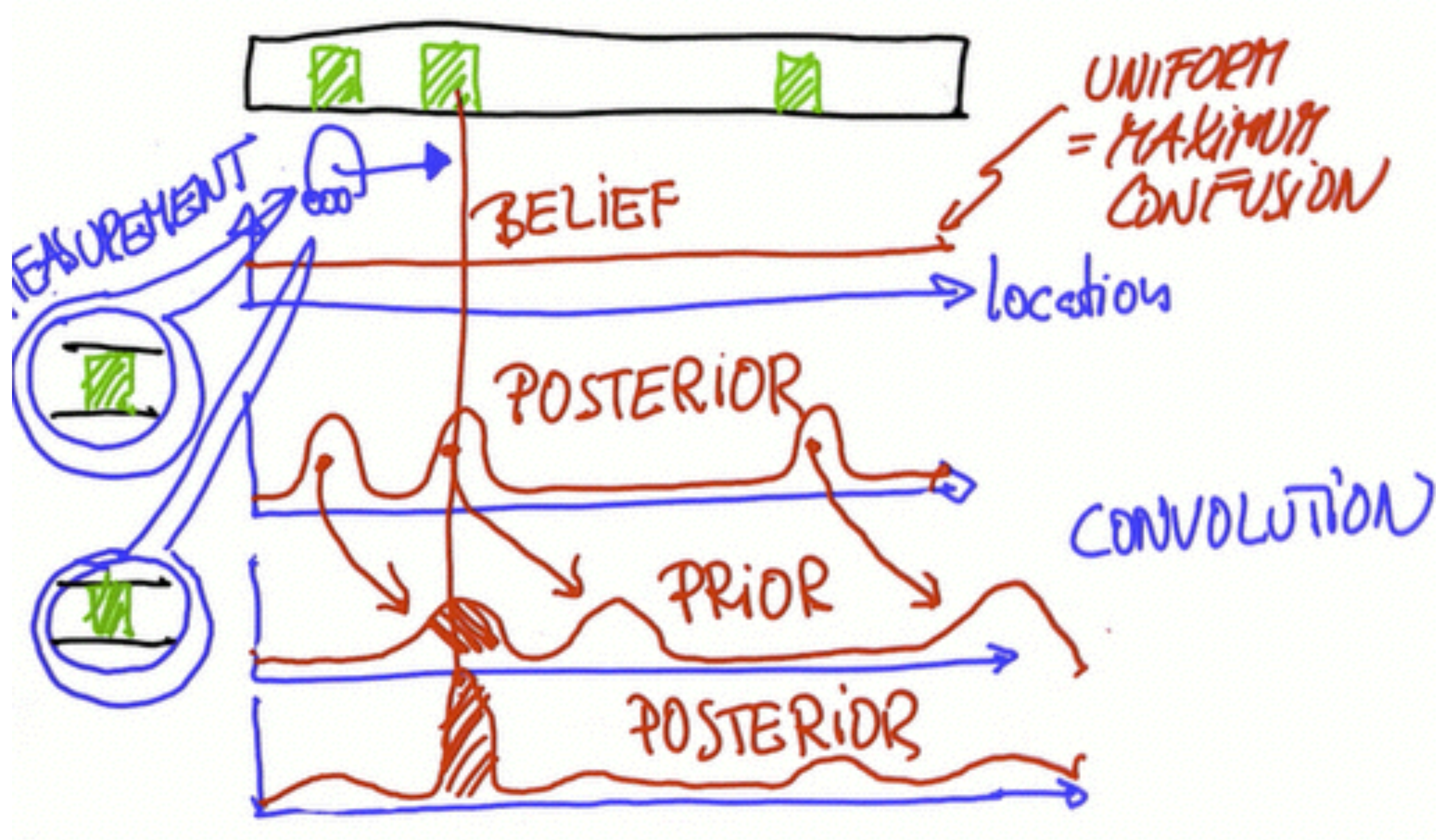


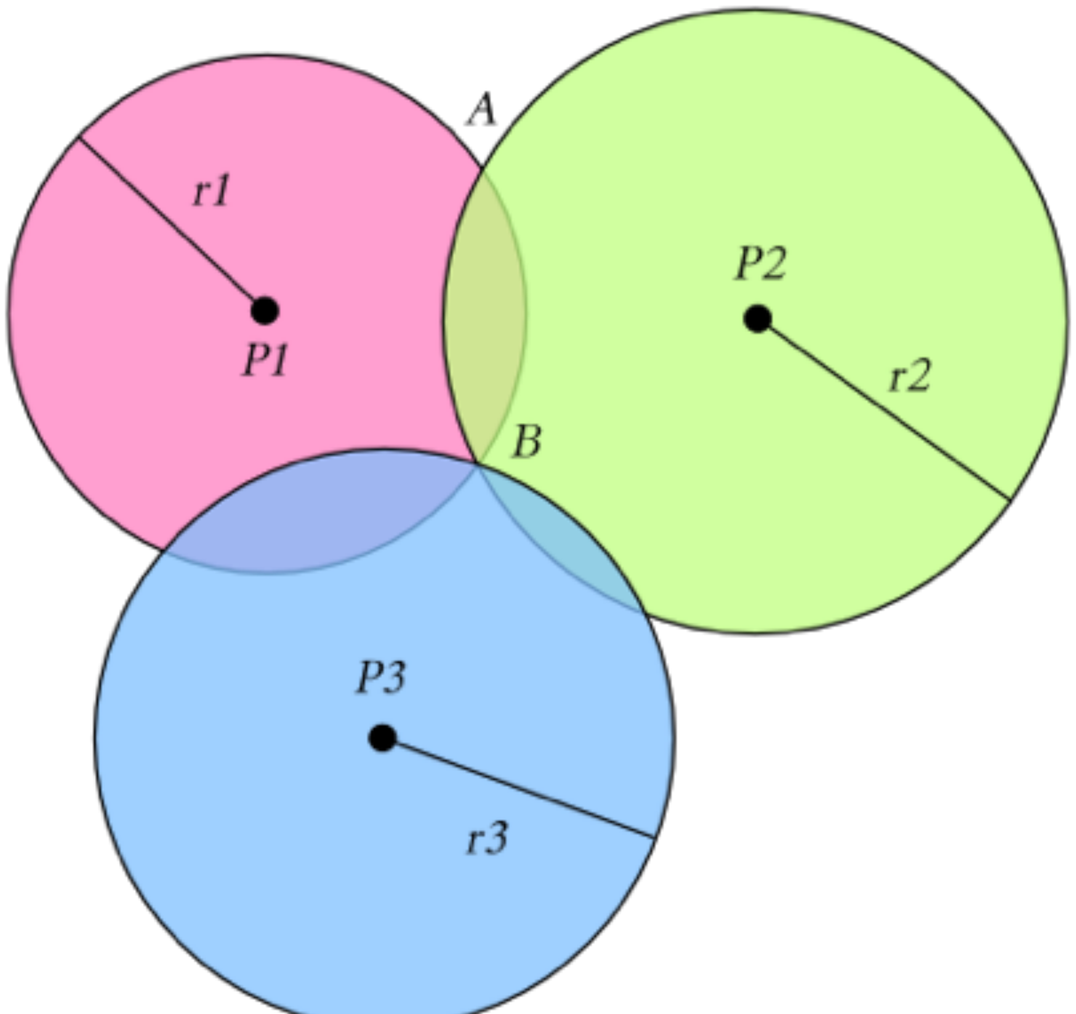
Trilateration and its limitations

Noisy Data

Physical Barriers

Large Overlap Areas



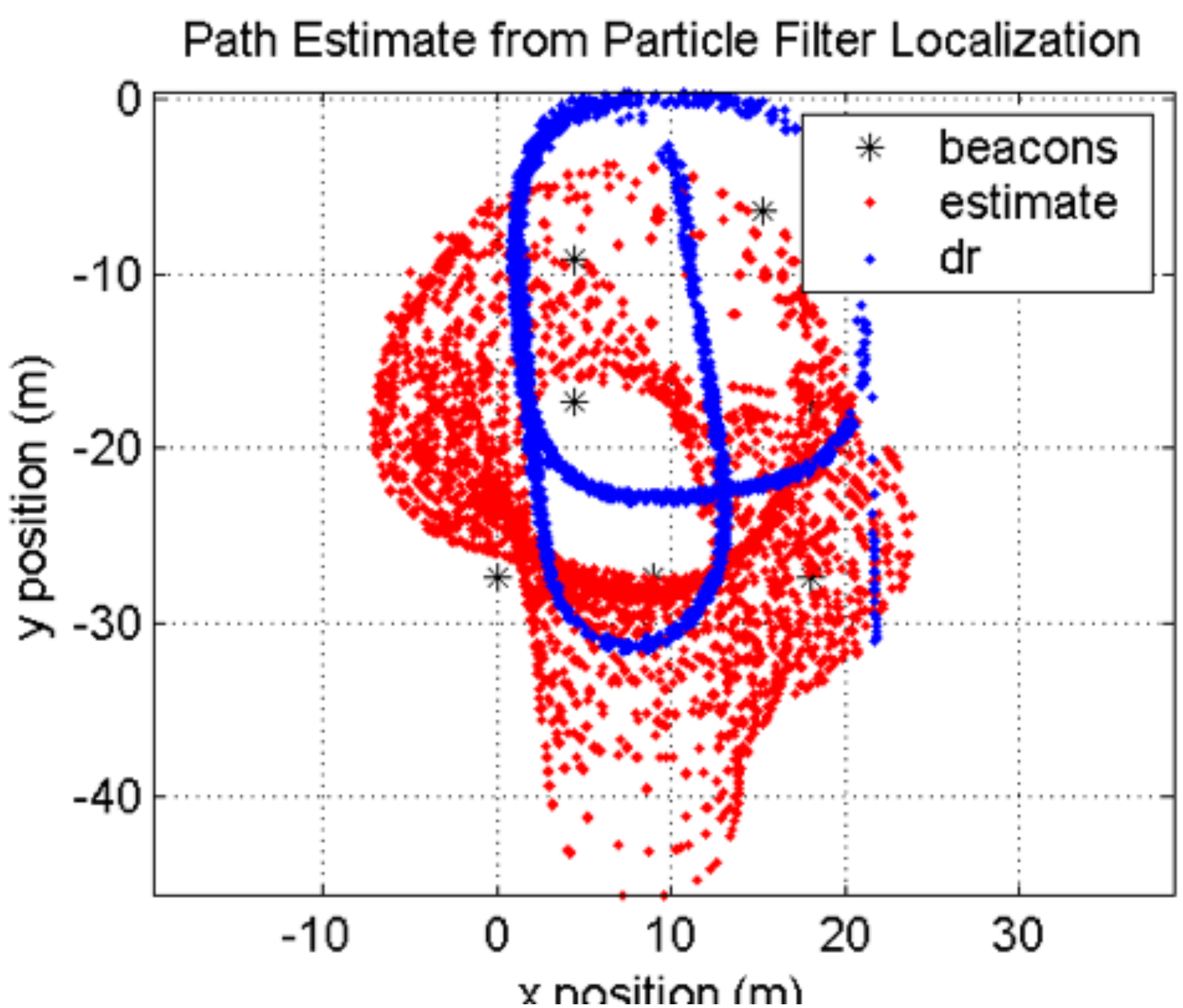


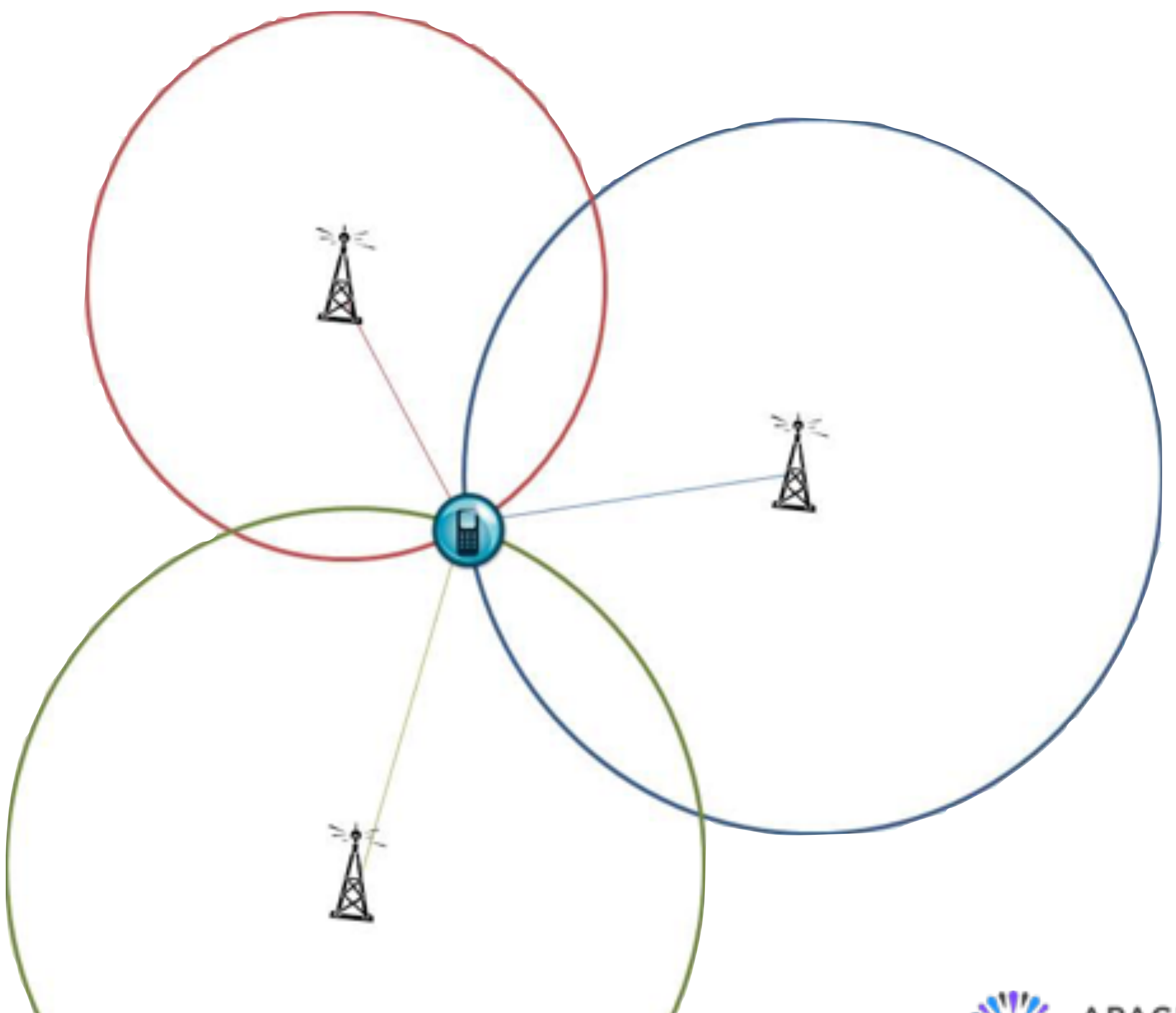
Moving Targets

Innacuracy

Large Overlap Areas

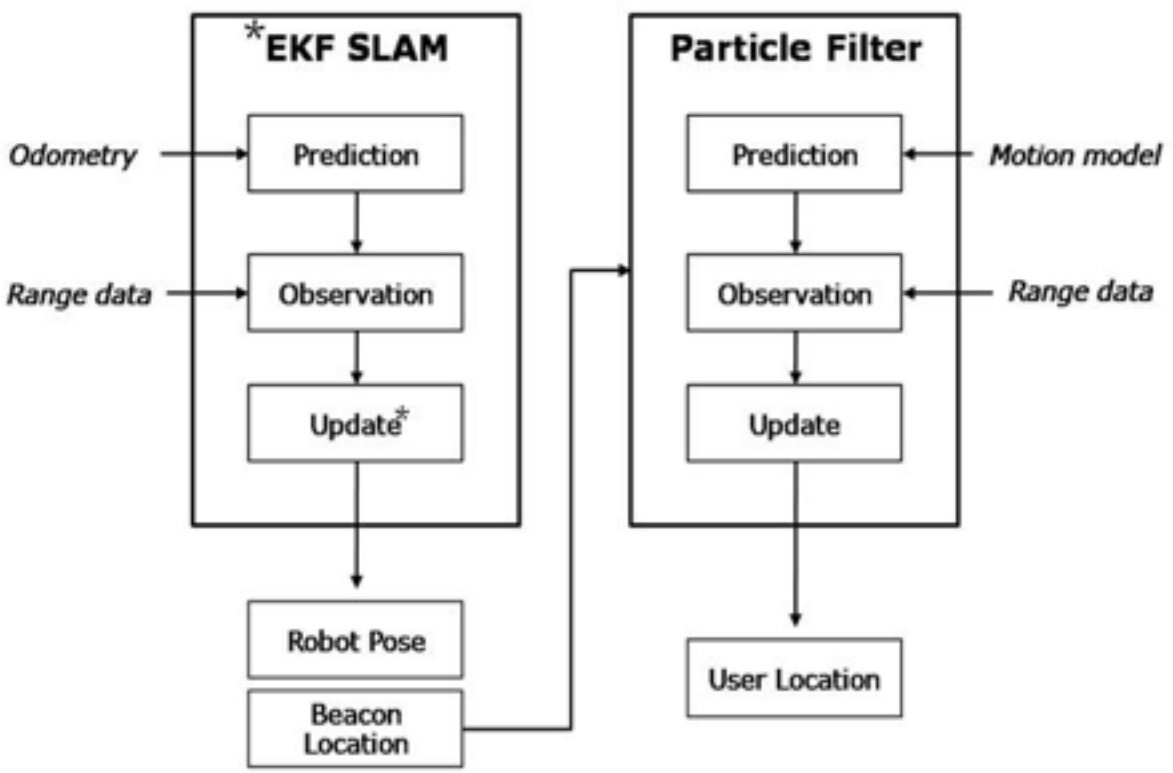
Particle Filters - Calculating the optimum solution





Particle Filters - Calculating the optimum solution

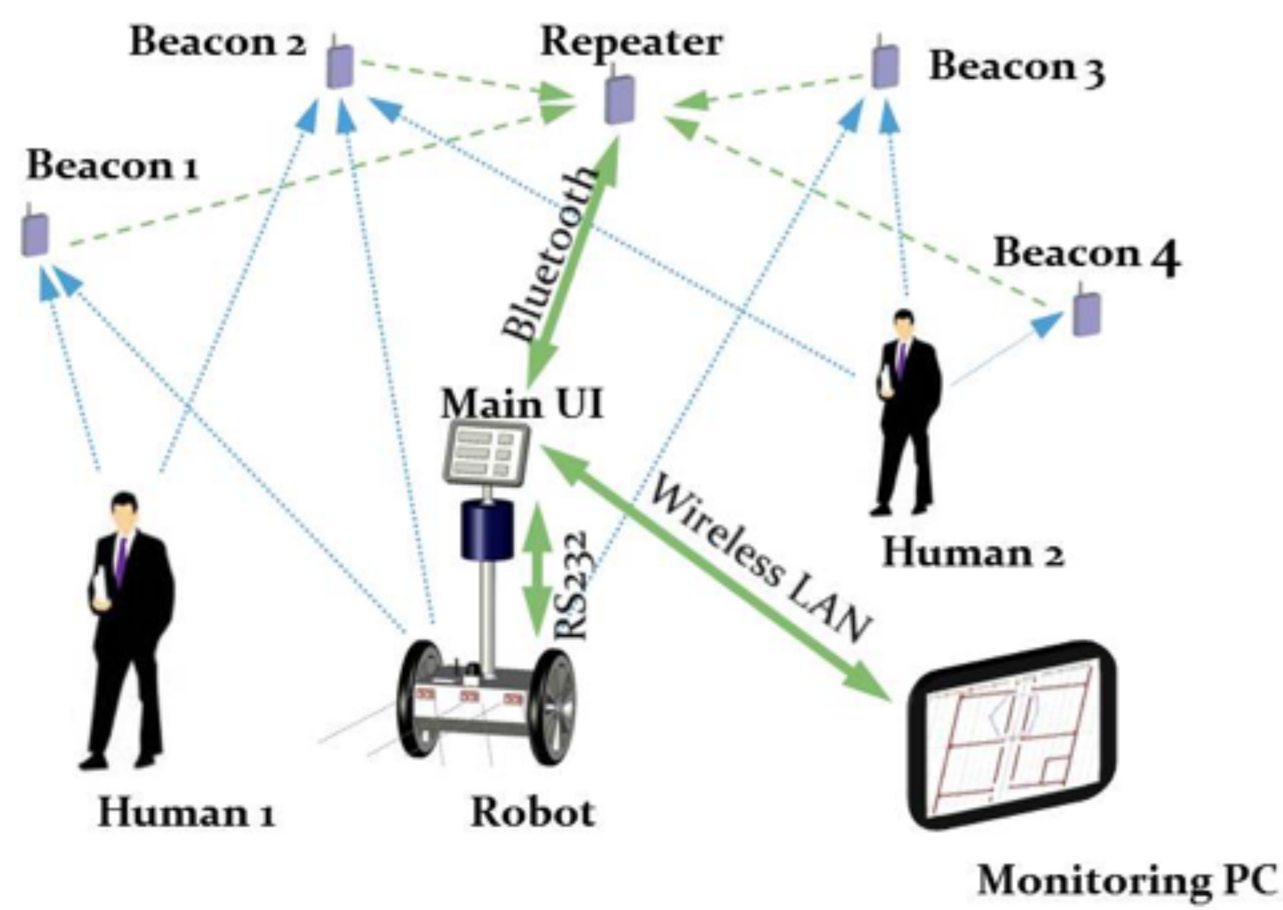
Localization Algorithm



User localization based on the localization of robots and beacons

* Extended Kalman Filter Simultaneous Localization And Mapping

Prototype System

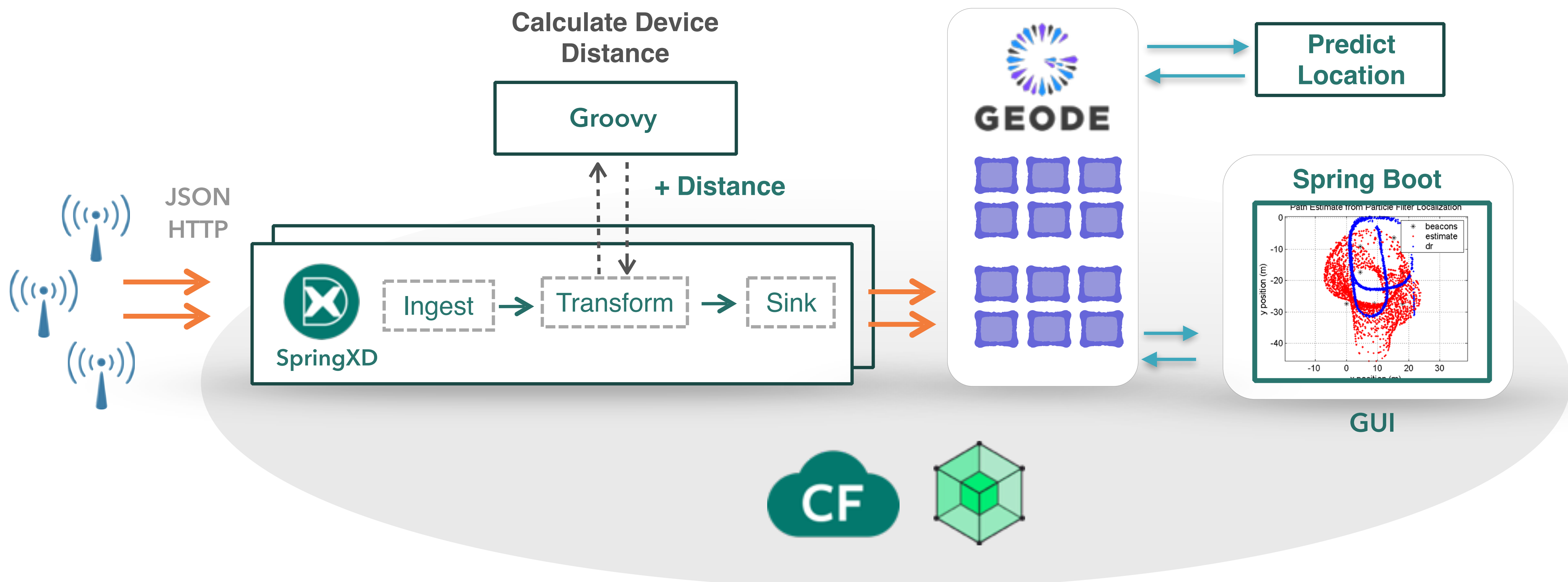


Autonomous Navigation

The Solution

1. Capture signal strength
2. Calculate distance from antenna
3. Trilaterate different sensors to predict location in real-time
4. Show on a map with live updates

Architecture Overview



Application Platform

CLOUD FOUNDRY

APACHE GEODE

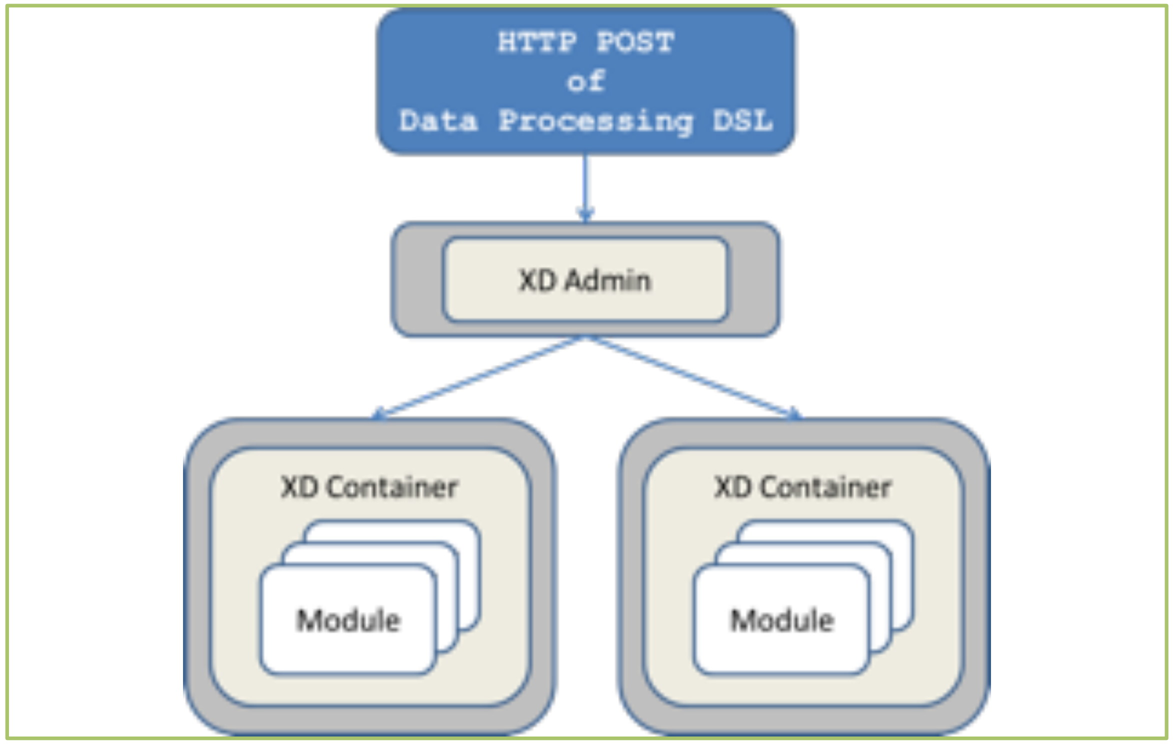
Geode Basic Concepts

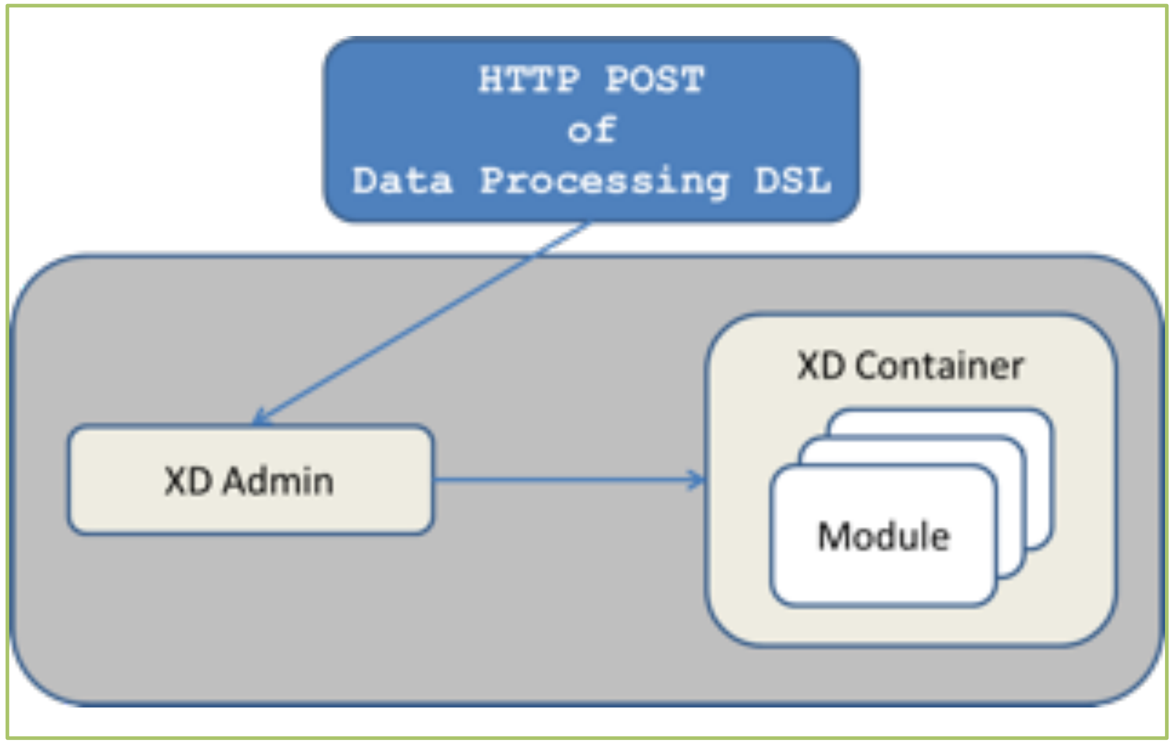
- Cache
  - Configurable through XML, spring ,Java
- Region
  - Distributed j.u.Map on steroids
  - Highly available, redundant
- Member
  - Locator, Server, Client
- Callbacks
  - Listener, Writer, AsyncEventListener, Parallel/Serial

Introduction to SpringXD

spring by Pivotal

Runs as a distributed application or as a single node

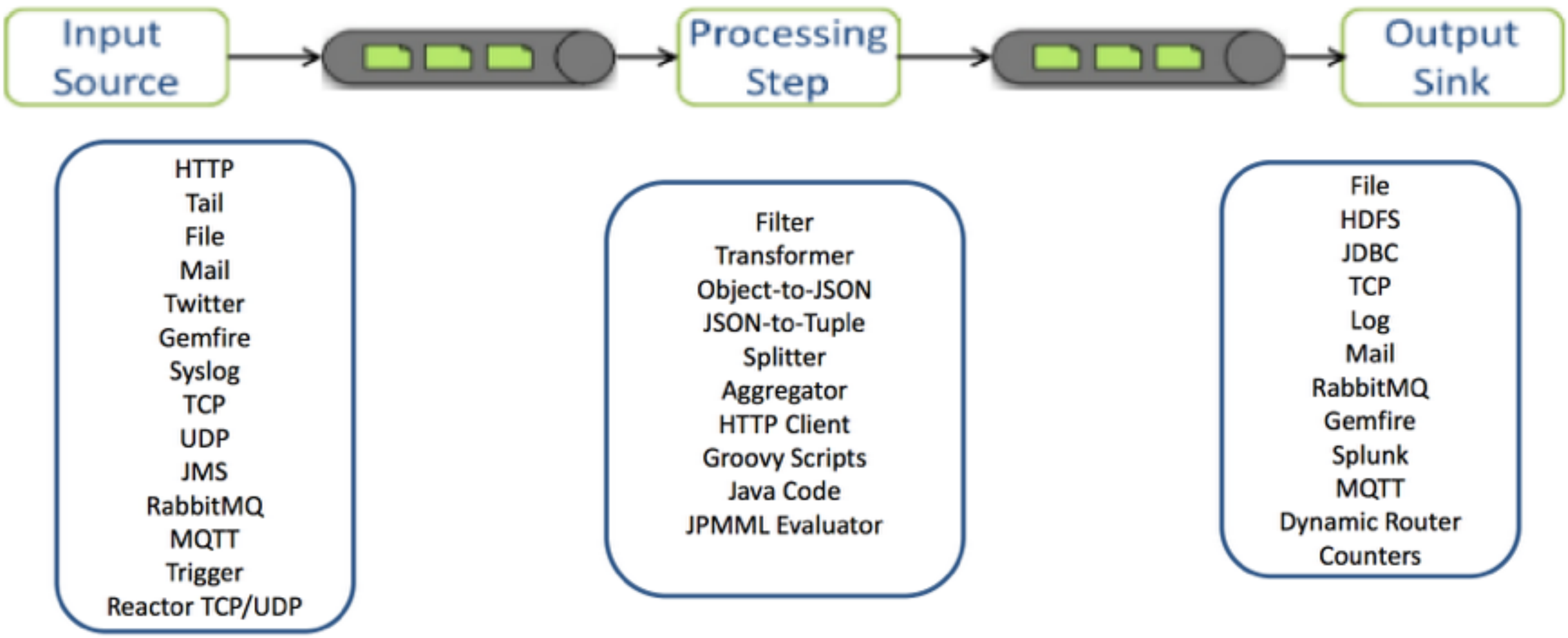




APACHE GEODE

Spring XD

A stream is composed from *modules*. Each module is deployed to a *container* and its channels are bound to the *transport*.



APACHE GEODE

Demo

Why have we selected those projects

Apache Zeppelin

- Iterative & Exploratory model
- Web based REPL
- Multiple Interpreters
    - Apache Geode
    - Apache Spark
    - Markdown
    - Flink
    - Python…

spring by Pivotal

- Productivity
- Built-in connectors
- Cloud Agnostic
- Highly Scalable
- Easy to setup
- Streams without coding

GEODE

- In-memory & Persistent
- Highly Consistent
- Extreme transaction processing
- Thousands of concurrent clients
- Reliable event model

APACHE GEODE

Source code and detailed instructions available at:

https://github.com/Pivotal-Open-Source-Hub/WifiAnalyticsIoT

Tomorrow:

Implementing a Highly Scalable In-Memory Stock Prediction System with Apache Geode (incubating), Spark MLib and Spring XD

Room: Tohotom - 14:30, Sep 30

Fred Melo, Pivotal, William Markito, Pivotal

Follow us on GitHub!

Fred Melo

@fredmelo_br

William Markito

@william_markito