ELASTIC

A Software Architecture for Extreme-ScaLe Big-Data AnalyticS in Fog CompuTIng Ecosystems

Distributed data platform for extreme analytics
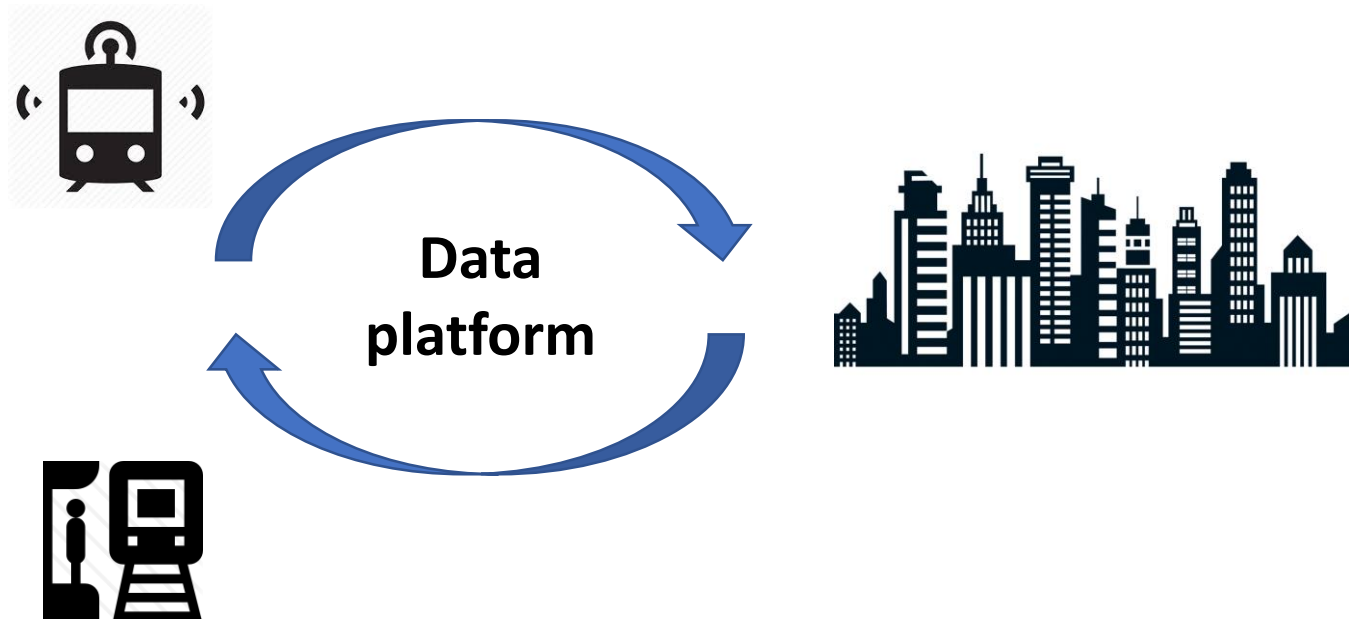
Anna Queralt – Senior Researcher
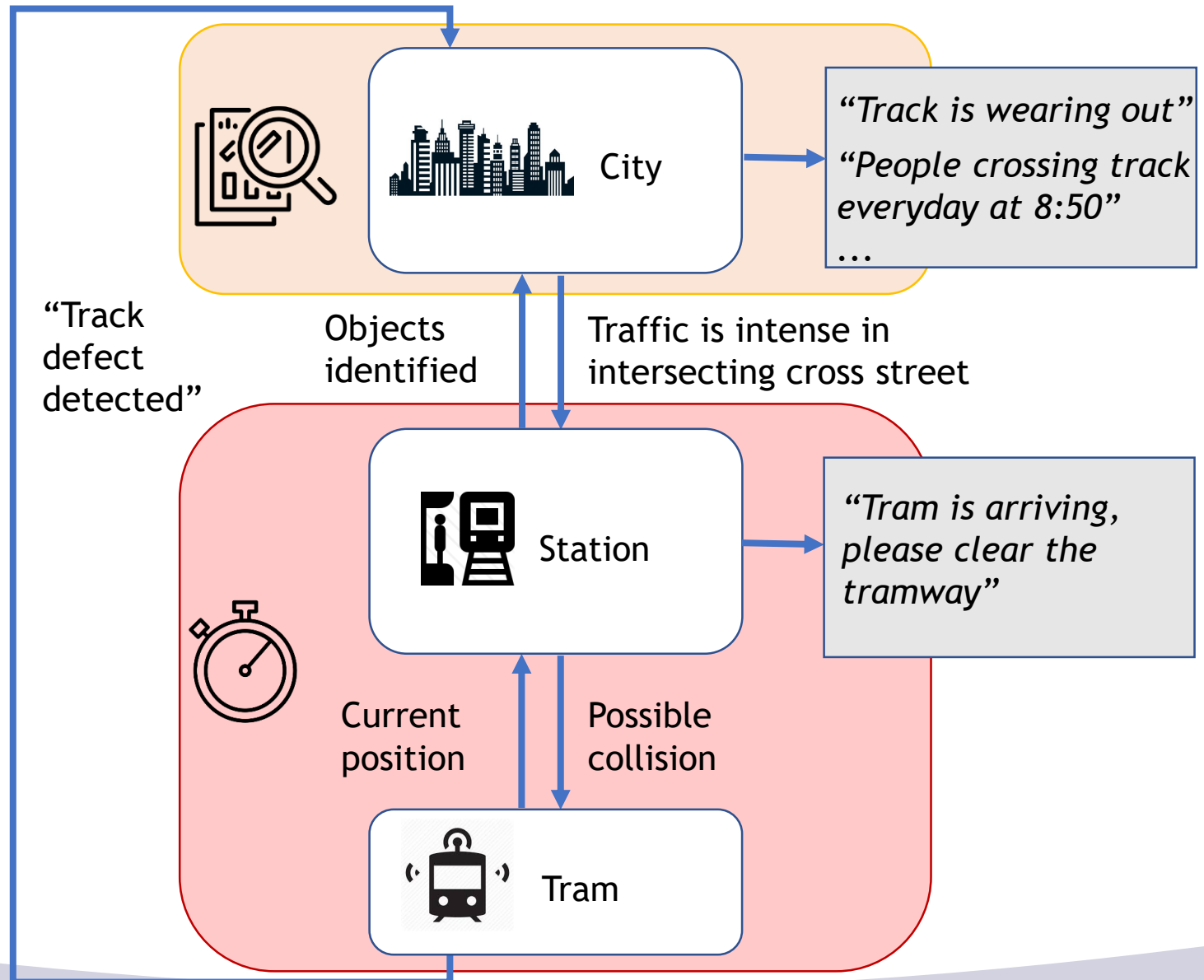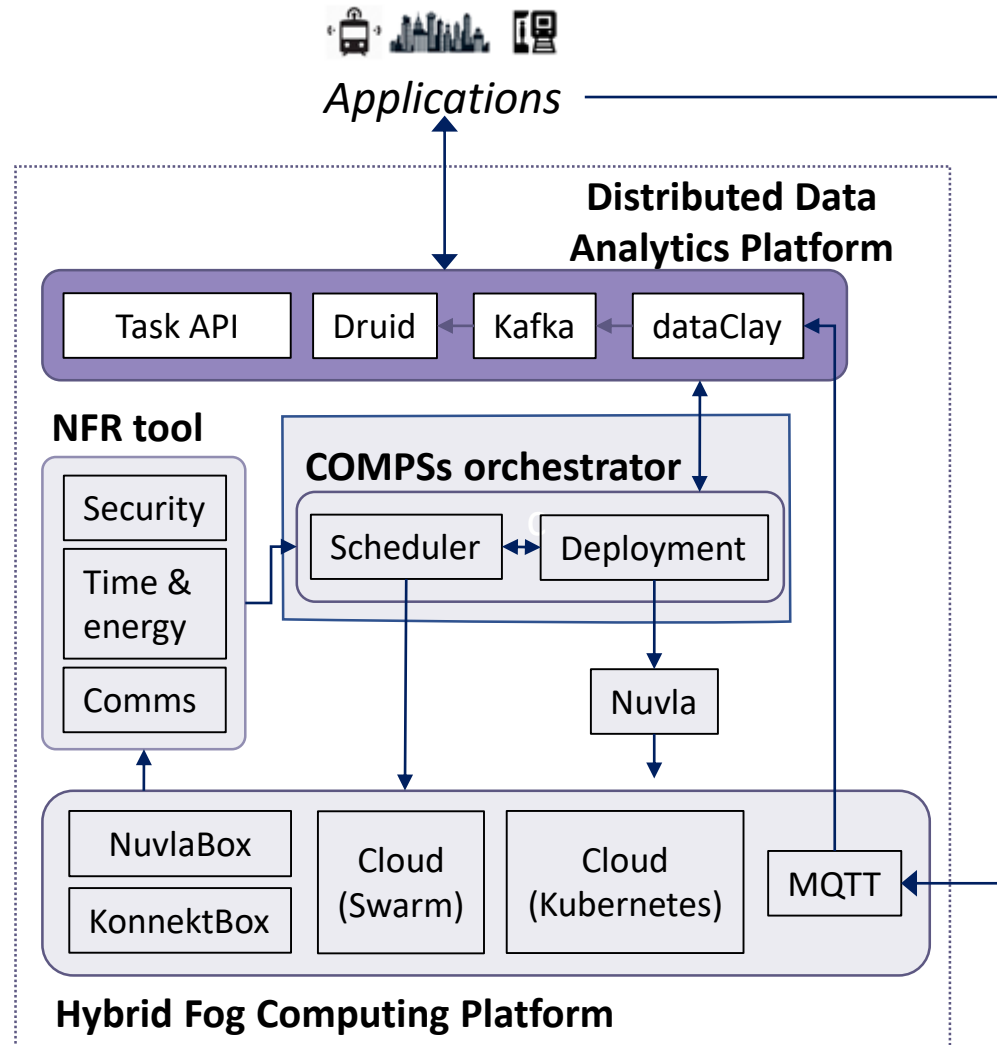
Barcelona Supercomputing Center (BSC-CNS)

04/11/2020

To provide a **distributed data management** platform that ensures data accessibility **across the compute continuum**, to support data analytics applications both on **data-in-motion** and **data-at-rest**.

Data platform

ELASTIC

ELASTIC Software Architecture diagram showing Applications connecting to the Distributed Data Analytics Platform (Task API, Druid, Kafka, dataClay), NFR tool (Security, Time & energy, Comms), COMPSs orchestrator (Scheduler, Deployment), Nuvla, and Hybrid Fog Computing Platform (NuvlaBox, KonnektBox, Cloud (Swarm), Cloud (Kubernetes), MQTT).

- For real-time processing at the edge
  - Lightweight
  - Arbitrary data structures
  - Efficient access and processing on objects of interest

- For analytics in the cloud
  - Fault-tolerance
  - Arbitrary exploration of event data
  - Fast queries on the whole dataset

dataClay

- dataClay is a **distributed active object store** for high performance, big data, and edge-to-cloud applications
  - Developed at the BSC, open source

- Brings **computation to the data**
  - Reduces communication costs
  - Enables in situ processing

- Runs on **heterogeneous infrastructures**
  - A single solution from small devices at the edge to datacenters
  - Simplifies application development and operation

- Creates a **shared data space** between independent applications/devices
  - Provides seamless access to data regardless of its location
  - Devices can dynamically join or leave the infrastructure

- Can be used standalone or integrated with the COMPSs/PyCOMPSs parallel execution runtime

- Successfully used in other edge-to-cloud projects
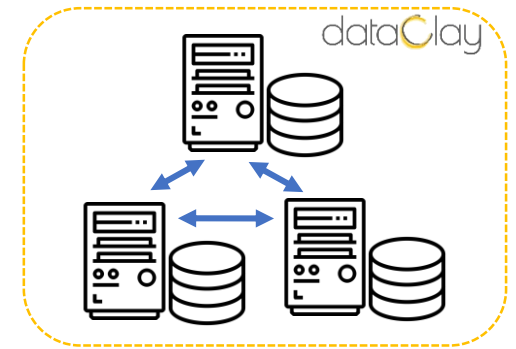
druid

- Druid is a **column-oriented, distributed data store**
  - Part of the Apache ecosystem, open source.

- Quickly ingests massive quantities of **event data**
  - Stream ingestion
  - Batch ingestion

- **Fault-tolerant** architecture
  - Data redundancy, no single point of failure
  - Permanent data backup in deep storage

- Data is **partitioned** and stored on different nodes in the cluster
  - Queries are processed in parallel, results are merged to return the final result

- Used in **business intelligence** applications
  - Arbitrary slice-and-dice (OLAP) data exploration on large datasets
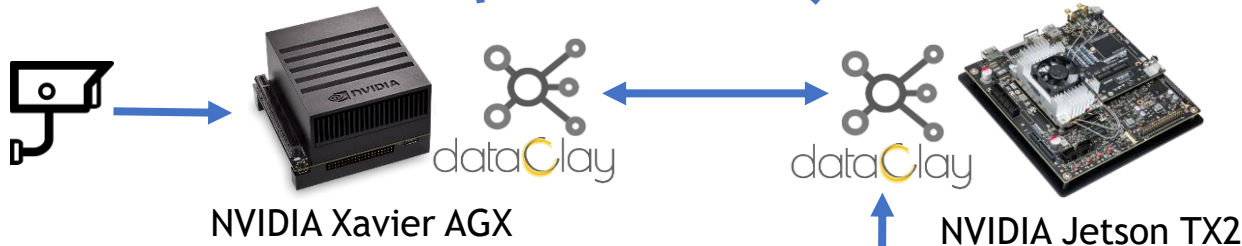 - Fast analytics queries
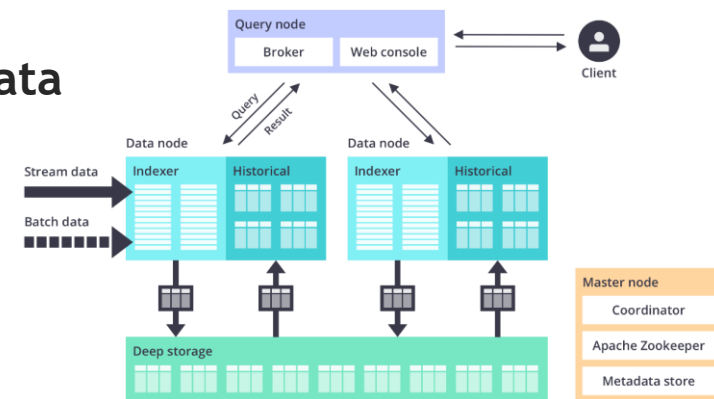
Data analytics platform deployment

- ✓ Data available from the edge to the cloud
- ✓ Supports real-time processing and historical analytics
- ✓ High performance
- ✓ Interoperability

ELASTIC

A Software Architecture for Extreme-ScaLe
Big-Data AnalyticS in Fog CompuTIng Ecosystems

www.elastic-project.eu

Stay Tuned!

anna.queralt@bsc.es

@elastic_EU   www.linkedin.com/company/elastic-project