



**Barcelona  
Supercomputing  
Center**  
*Centro Nacional de Supercomputación*



# AI between the Edge and Cloud

## Leveraging IoT and Analytics for Smart Cities

### Cloud Computing

Master en Enginyeria Informàtica

24<sup>th</sup> February, 2020

February/2020

Josep Ll. Berral

# Josep Lluís Berral-García

## Me and my research

Josep Lluís Berral García, from Barcelona and BSC

- Education at UPC:

- *2007 Informatics Engineer*
- *2008 Master on Computer Architecture*
- *2013 PhD on Computer Sciences → ML + HPC*



- Away from academia:

- *Attempt to create a Digital Goods Distribution Platform → Didn't go far ☹*
- *Work in Industry, as QA on Medical Machinery*

- Back to research:

- *BSC-UPC with David Carrera (2014)*
- *Activity Leader for "Applied Machine Learning"*
- *Teaching and Advising PhD + Master students*

- Data Mining/Machine Learning
- Operating Systems
- Computer Engineering

# Introduction

## 1. Deploying the Smart City

- User and Citizen-centric services
- Internet of Things → Computation in local & personal devices

## 2. Data Analytics

- Data from users/systems → Analytics and AI models
- ... for systems optimization/research/business

# Introduction

## 3. Edge Computing

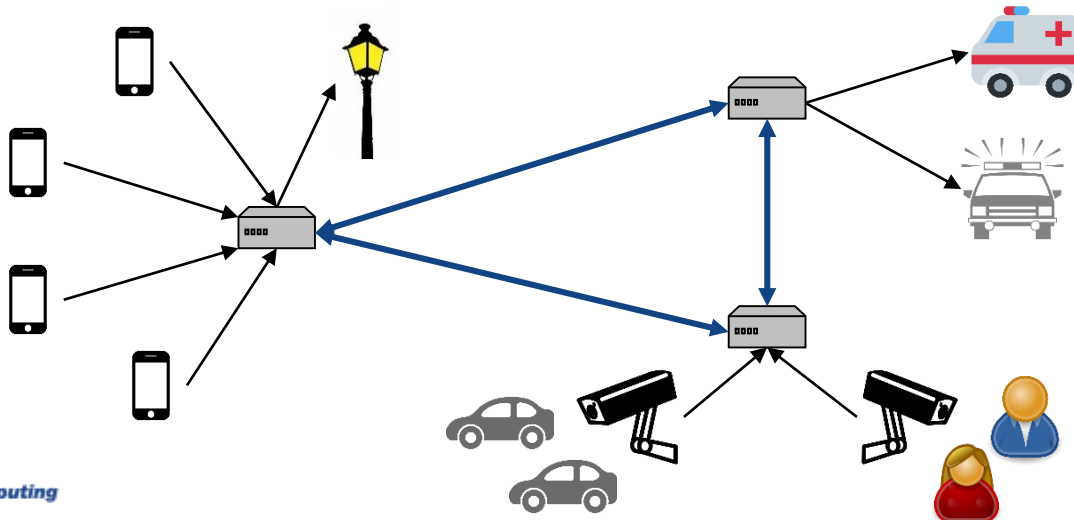
- Move the data processing and analytics to devices close to data source
- Avoid transmission of sensitive data
- Amortize computing power in local devices
- Provide faster services

## 4. Distributed Analytics

- Federated Machine Learning paradigms

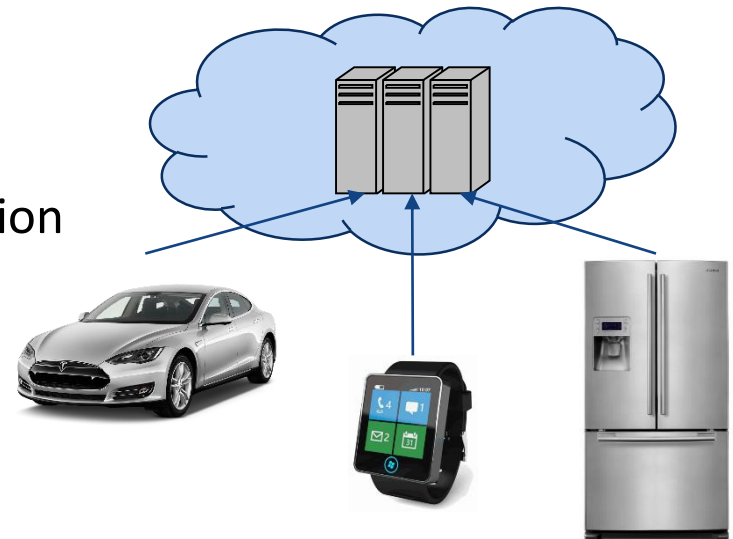
# Smart Cities

- Smart Cities
  - Use of technology to improve citizen services
    - From reducing consumption of street lights...
    - ... to provide emergency assistance to private citizens in the street
  - Pervasive Computing
    - Everything has a computer
    - ... and sensors also actuators
    - Also users → mobile devices
  - Services collect data from users to optimize city services



# Internet of Things

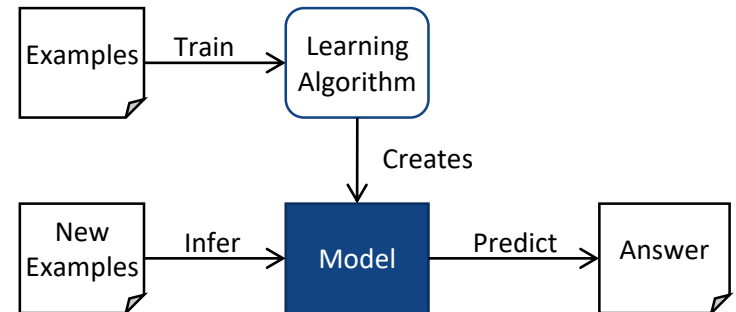
- Everything has a Computer
  - Edge Devices → Devices collecting data, or close to data
- Edge Devices
  - Low power → Computing and Consumption
  - Unreliable → May disconnect, private devices, ...
- IoT vs the Cloud
  - Devices connect to send information and receive instructions
  - Change of Paradigm: Edge Computing



# Analytics, AI and Machine Learning

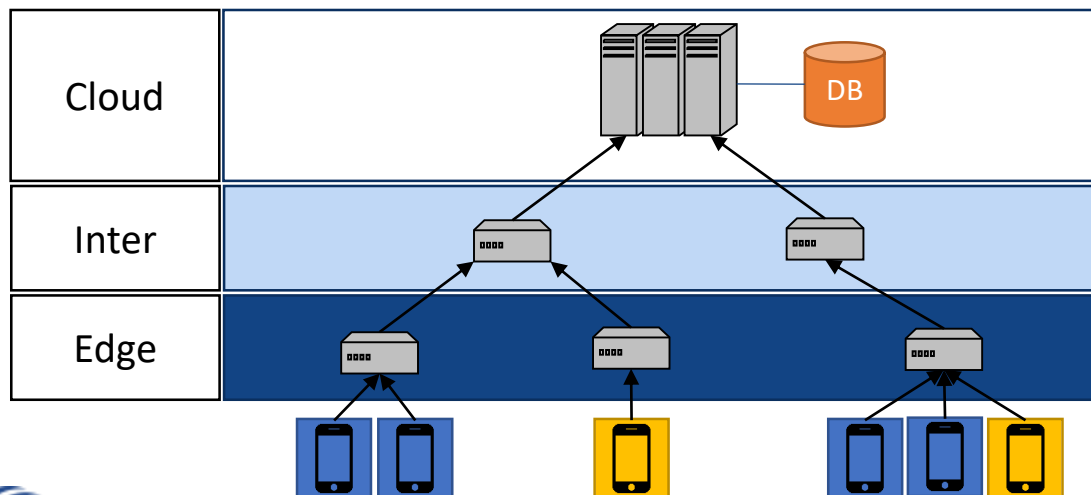
“Collecting data from devices for what?”

- Data Analytics
  - Extract information
  - Learn the behavior of the system/users/vehicles/...
    - Then improve the system/make profit
- Machine Learning
  - Automatic modelling
    - Learn from examples
    - Create a prediction model
  - Knowledge discovery & Data mining
    - Extract patterns from data
- Artificial Intelligence
  - Automatization of decisions and management
  - ... including Analytics and Learning services



# Edge-Cloud Architecture

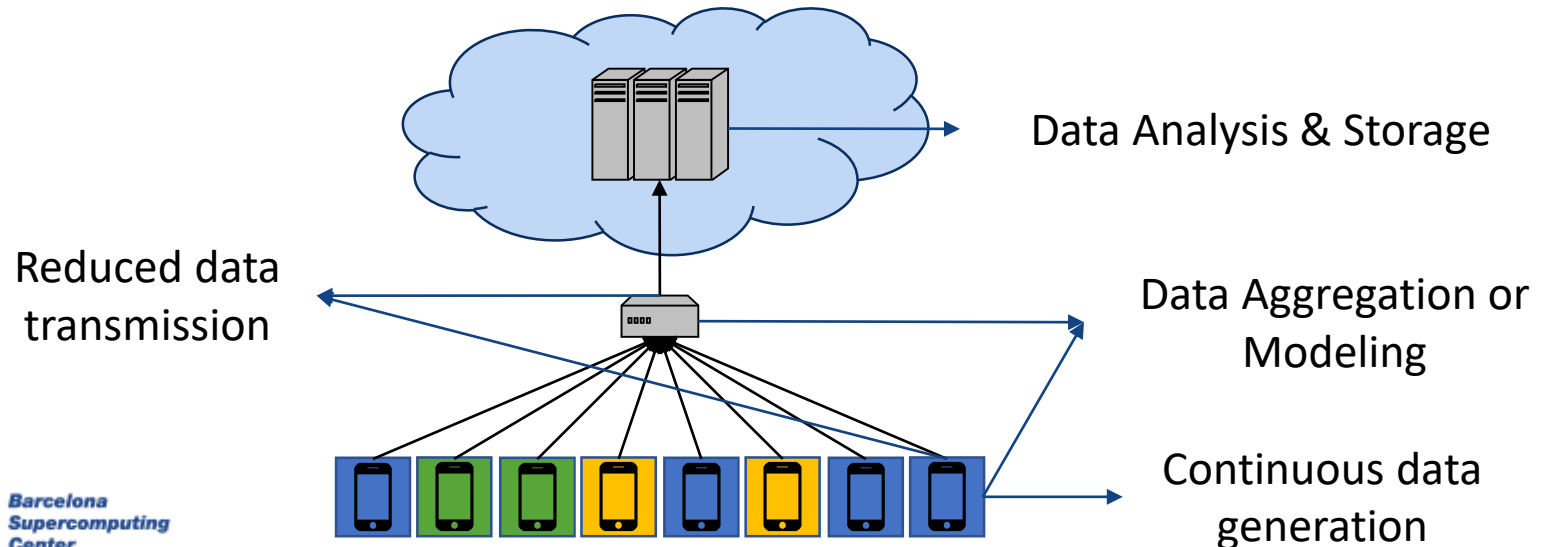
- The Cloud:
  - High performance nodes
  - Concentrate data computing in “far-away” data-centers
- The Edge:
  - Low power nodes
    - ... or personal devices and appliances
  - It is where all data is collected





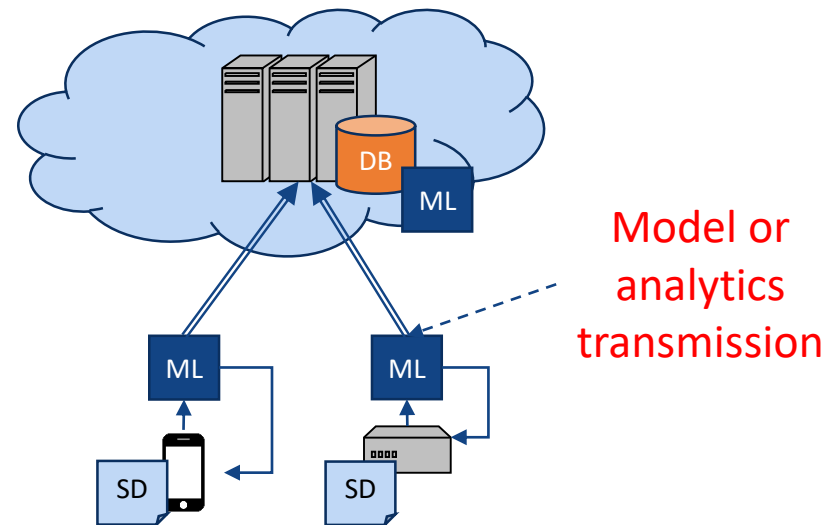
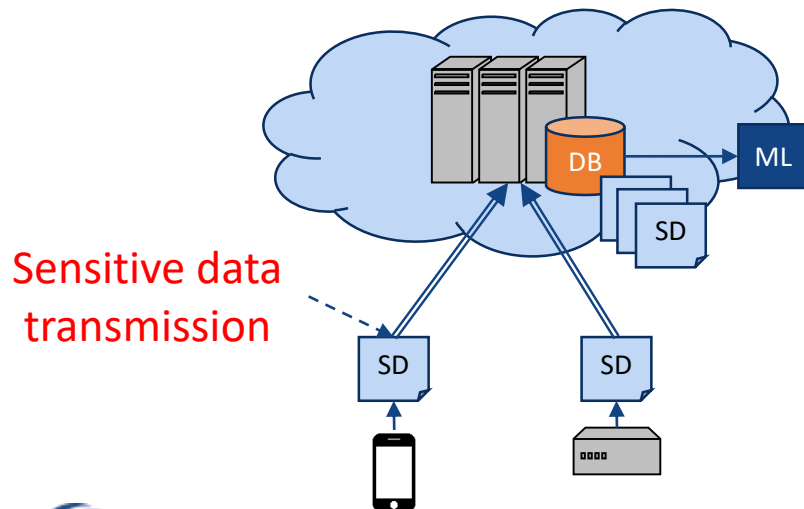
# Analytics in the Edge

- Efficiency
  - Keeping important data
    - Edge devices aggregate local data
    - Only transmit aggregated data ← usable for analytics
    - Raw data ← can be discarded from origin
  - Service to users
    - User data and analytics/aggregation remain local
- Energy
  - Edge devices are “always” on-line and low-power. Most of the time, power is amortised



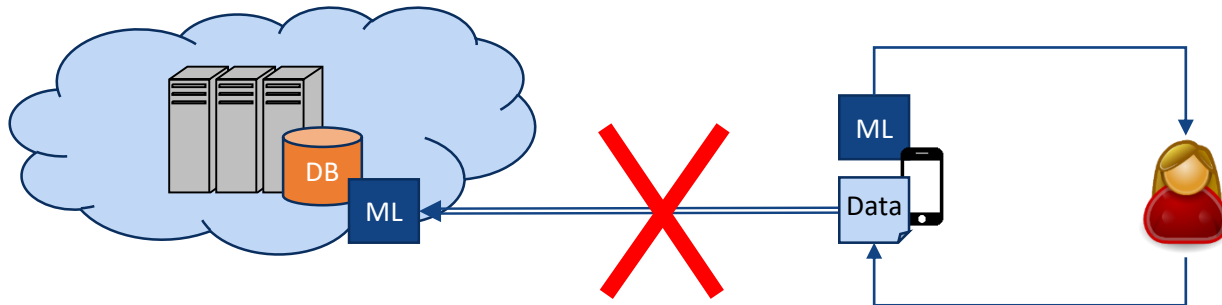
# Analytics in the Edge

- Privacy
  - Aggregation:
    - Allows anonymization
    - Done in Edge devices or personal devices
  - Sensitive data:
    - Analytics and models are created without transmitting it
    - Companies and government can get information about users without it
    - User decides which data is shared, without altering the service



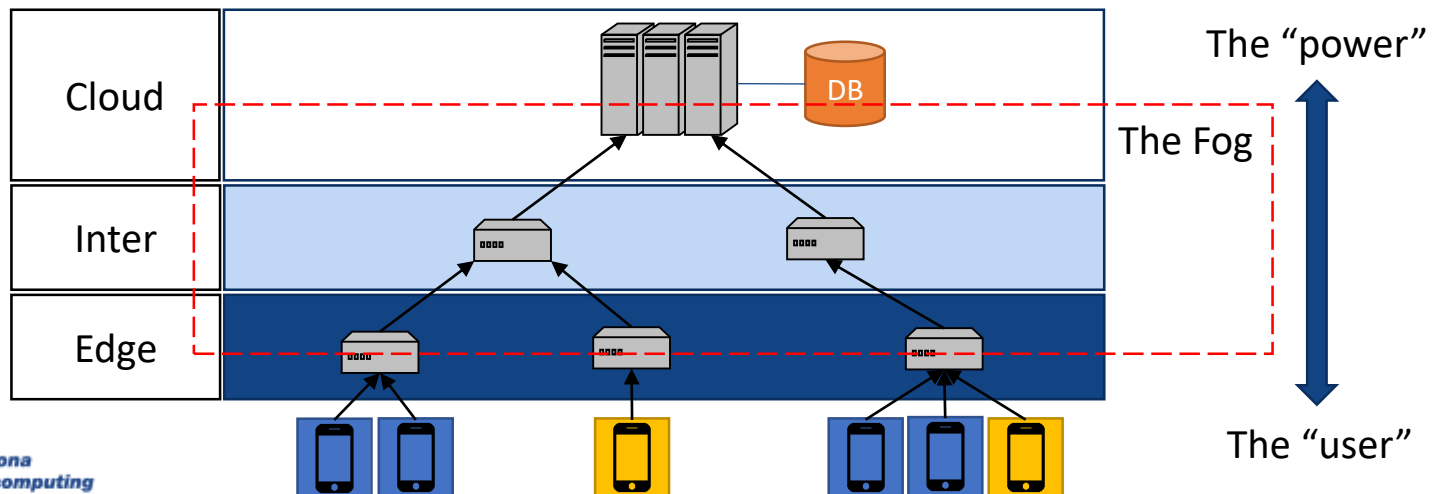
# Analytics in the Edge

- Autonomy
  - Services for User
    - ... can work in local devices
    - The user does not need to be constantly connected
  - Network disconnection
    - Local devices keep aggregated data and local analytics
    - Local services can keep working
  - Transmission of aggregated data
    - ... is resumed when connection restablishes



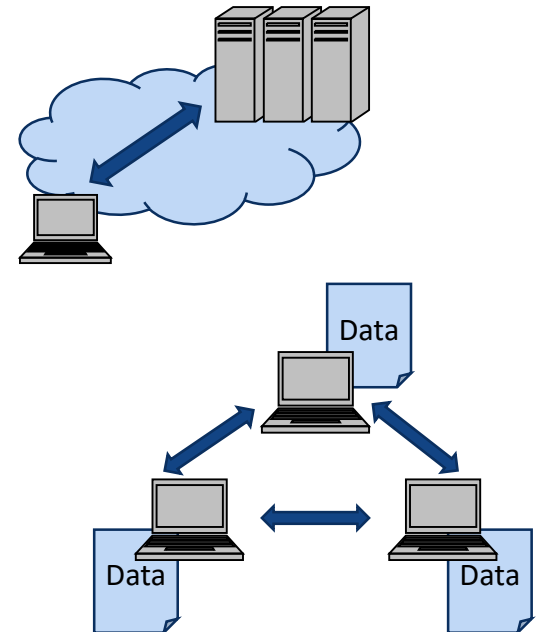
# Leveraging Edge-Cloud

- Upward: more power required
  - Lower nodes = Low Power
  - Higher nodes = High Performance
- Lateral: off-loading
  - Unused sibling nodes ← send over-load
- Downward: network/space reduction
  - Higher nodes = More communication / Data-warehousing
  - Lower nodes = Less Network Usage & Lower Latency (Edge Services)



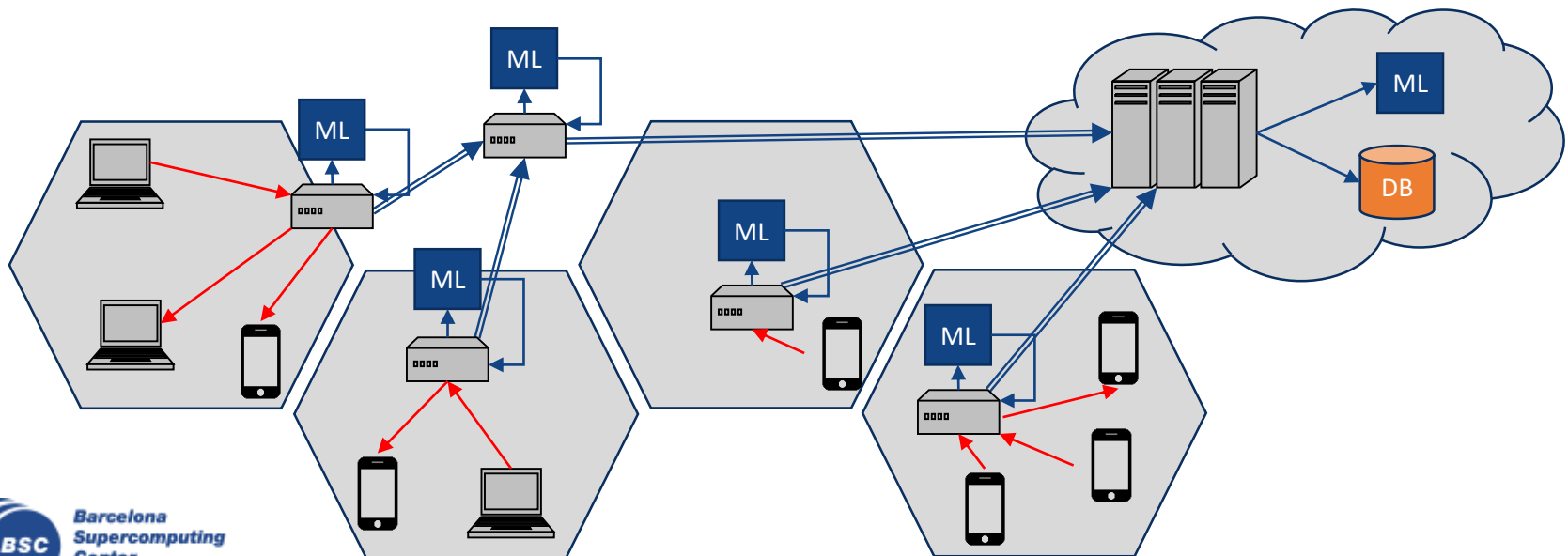
# Distributed Analytics

- Analytics:
  - Either general analytics
  - ... or Machine Learning and AI
- Distribution of ML
  - “Where to run AI”
    - Centralized HPC centers
    - Reunion of low-power computing devices
  - Federated Learning
    - “How to share data & efforts”
    - Beware privacy!
    - Beware scarce computing resources!



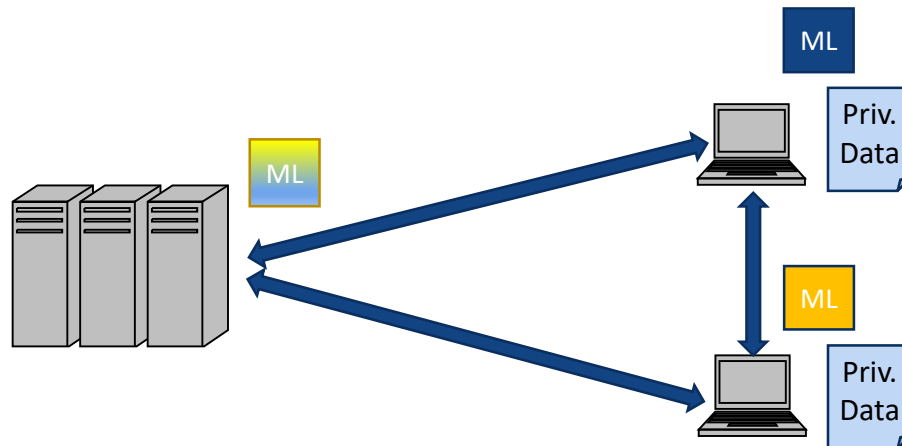
# Federated Learning

- Until now:
  - Centralization (Cloud) collects data and models it
  - Models are passed down (if needed)
- Federated Learning
  - Each local node models its data
  - Transmit and query data/models only when necessary
  - Local autonomous nodes



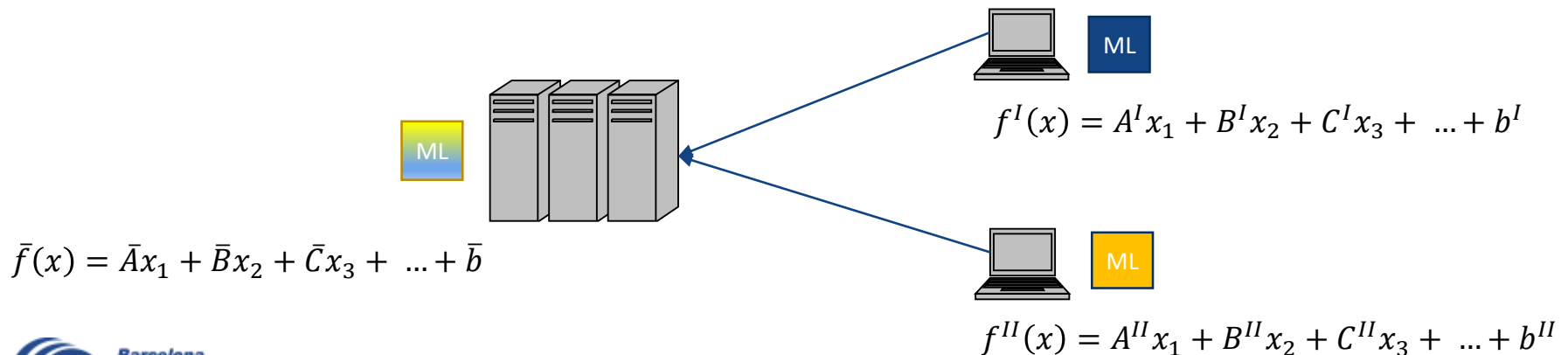
# Federated Learning

- Distribution of ML load
  - Share analytics/modelling efforts
    - Create local models
    - Share the local models
    - Keep data private in local environments
- Model aggregation
  - Instead of aggregating data on the Cloud
  - ... aggregate models on the Cloud
  - ... or in neighbor environments (e.g. P2P systems)



# Model Aggregation

- Machine Learning models
  - Statistical learning: functions and data representations
  - Neural Networks: matrices and vectors
- Aggregate models
  - Find the average function
  - Find the average matrices and vectors
  - Ensemble models and decide which to keep





# New Technologies for Edge + AI

## Low-Power Devices

Machine Learning on Limited Environments, and Enhanced Edge Devices



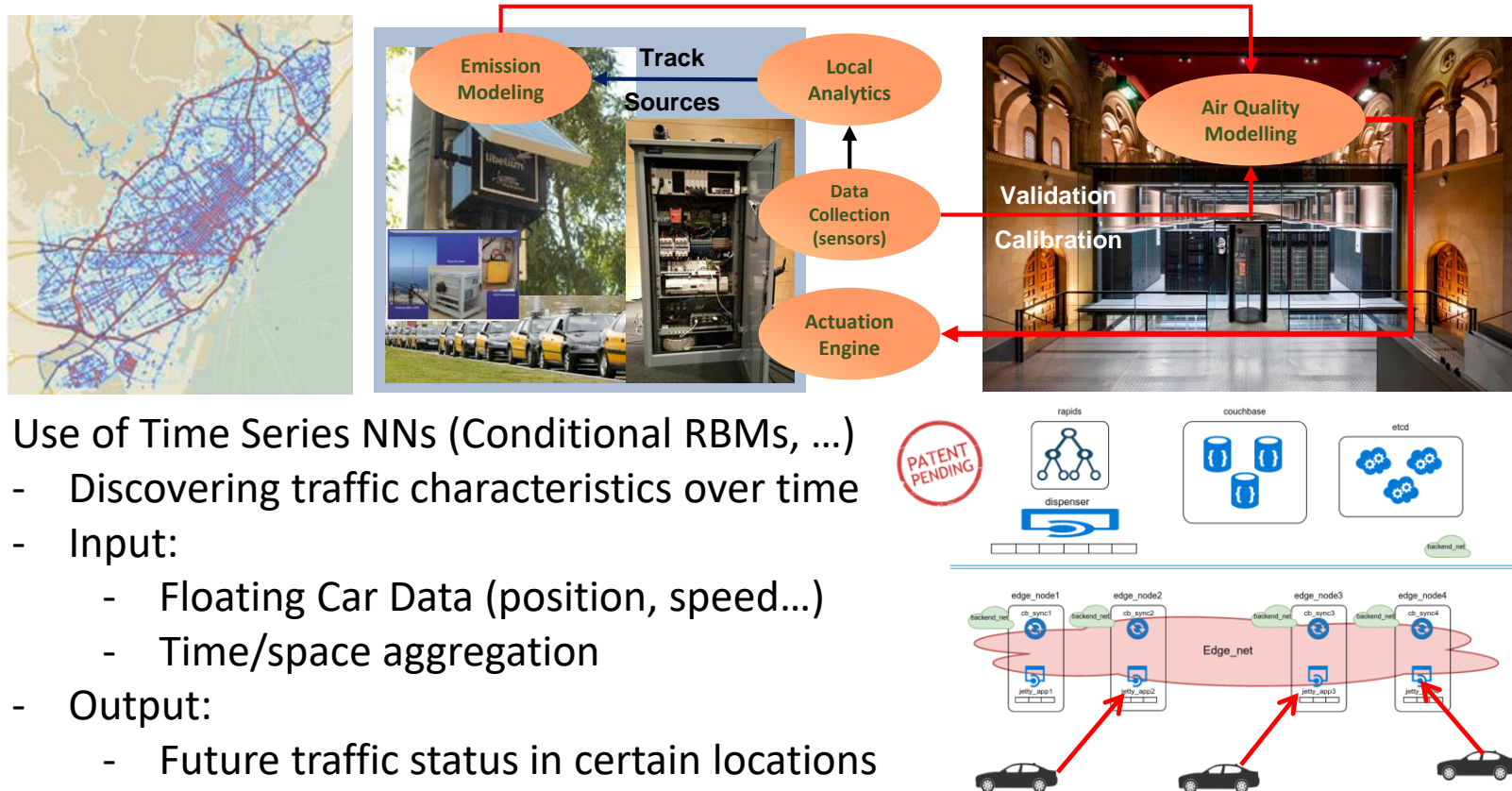
Research:

- Discovering the real limitations vs. advantages of such devices
- Trade-offs between modeling Near-Data vs Near-Power

# Use case: Traffic and Mobility

## Modelling of road traffic towards emissions using Deep Learning

Emission modelling and forecasting based on deep learning



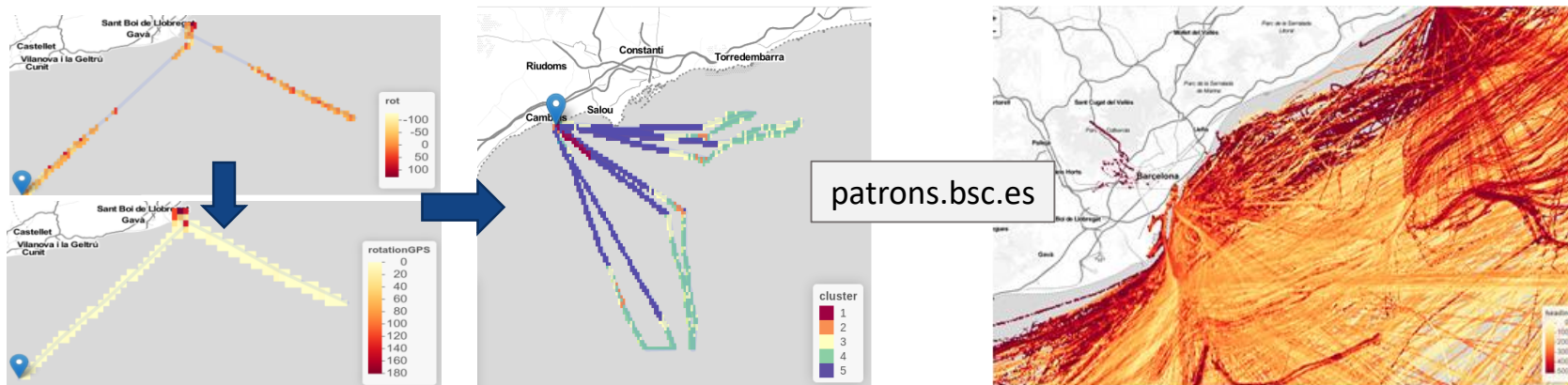
Use of Time Series NNs (Conditional RBMs, ...)

- Discovering traffic characteristics over time
- Input:
  - Floating Car Data (position, speed...)
  - Time/space aggregation
- Output:
  - Future traffic status in certain locations

# Use case: Maritime Traffic Modeling

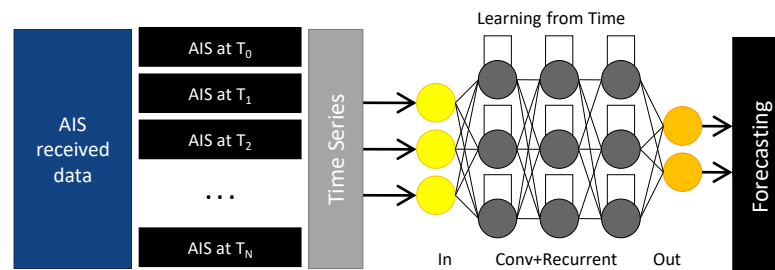
## Modelling of Maritime emissions using Deep Learning

Emission modelling and forecasting based on deep learning



Use of CNNs (Convolutional-LSTMs and 3D convolutional NN)

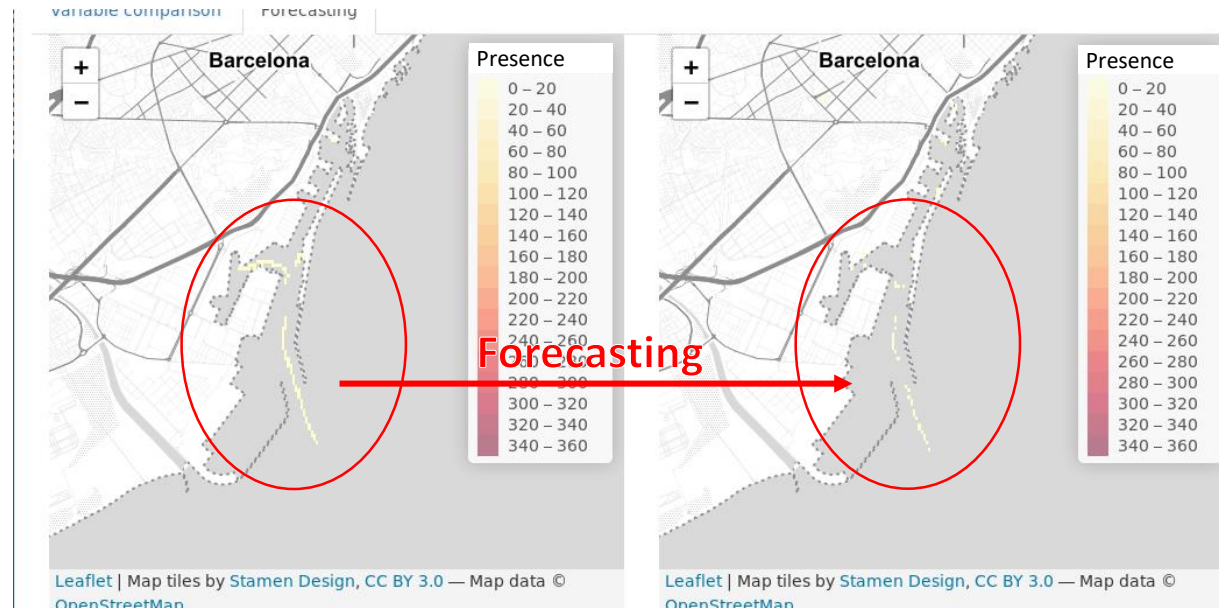
- Learning patterns and forecasting
- Input:
  - AIS traces (position, operation...)
  - Emission modeling (Jalkanen)
  - Time/space aggregation
- Output:
  - Future position and emissions



# Use case: Port Forecasting

## Visualization & Forecasting Tools

Emission modelling and forecasting from AIS & Fog-Computing



Work in Progress:

- Forecasting of occupation and emissions in port vicinity
- Applicability of these techniques to road traffic
  - More complex/chaotic environment

# Challenges on Analytics in the Edge

- AI in Exa-Scale & Extreme-Scale analytics
  - Large amounts of data
  - Large amounts of models to run
  - Continuous arrival of new data to process
- Distribution of Analytics
  - Advances of Edge vs Cloud computing
  - Newly arrived Federation of learning models

## Preserve Privacy:

- Legal and Ethical restrictions on data and models
- Frameworks & methods to share results without sharing data and models

## Maintain Performance:

- Technical restrictions on AI processes, also low resource availability
- Frameworks & methods to share load from experiments



# Summary

- Smart Cities and IoT
  - Devices everywhere, collecting data , to manage everything
- Edge Computing
  - Performing analytics, aggregations, models, processing... on near-data devices
- Federated Learning
  - Distributing machine learning processes and load
  - Avoid transmission of sensitive data
- Some examples
  - Analysis of road and maritime traffic
  - Analysis of air quality
  - De-centralised and everywhere-available citizen services