Python for IoT Data Analytics

PhD Course Intro

EIT4SEMM PHD Program, July 2023

Prof. Marco Di Felice

Department of Computer Science and Engineering, University of Bologna

marco.difelice3@unibo.it

Learning Outcomes

- Overview of the **building blocks** of a data pipeline process, from data acquisition to knowledge extraction via AI/Machine Learning techniques.
- Overview of **Python** data structures and tools for data analytics.
- Focus on **IoT sensor data** (time-series data)
- **Methodology**: theory + practice (live demo + coding exercises)
- **Exam**: data analysis on IoT dataset

(write Python code → collect results → submit a report)

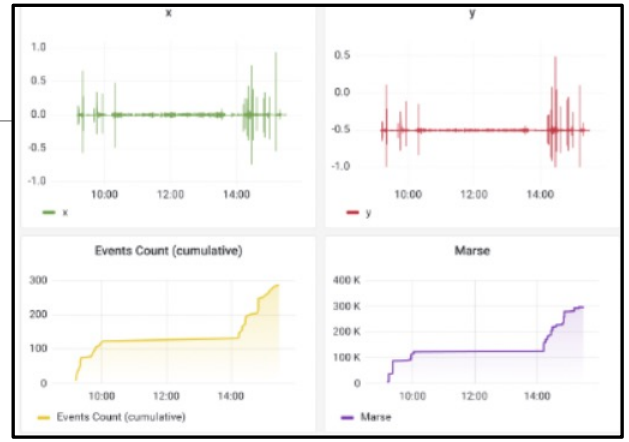
About the lecturer

- **Prof. Marco Di Felice (marco.difelice3@unibo.it)**
  - Full Professor in Computer Science
  - UNIBO, Department of Computer Science and Engineering
  - https://www.cs.unibo.it/difelice

Teaching activities (2022-2023)

- *Database systems* (III year, First Laurea Degree in Business Informatics)
- ***Data Analytics*** (II year, Second Laurea Degree in Computer Science)
- ***Internet of Things*** (II year, Second Laurea Degree in Computer Science)
- *Context-aware systems* (II year, Second Laurea Degree in Computer Science)
- *Things that Compute and Interact* (Bologna Business School)
- *Sport Analytics* (Master in Sport Marketing and Management)

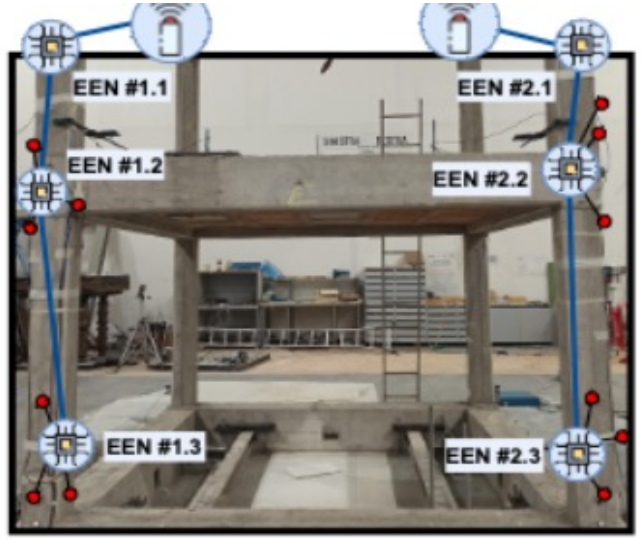
About the lecturer



- **Prof. Marco Di Felice (marco.difelice3@unibo.it)**
  - Head of the PRISM IoT Research Laboratory, DISI, UNIBO
  - Member of the advisory board of the EIT4SEMM PhD Programme

Research activities

- The Internet of Things (IOT)
- IoT interoperability and edge/cloud architectures
- Sensor data management and analytics
- Context (location/activity) aware software systems



Data science: how

DATA ACQUISITION

- LESSON 1: July 6, 2023, 9am-12pm
- **Course Introduction**
- **IoT Introduction**
- **Data loading and acquisition**

Data science: how

DATA ACQUISITION → **DATA VISUALIZATION**

- LESSON 2: July 12, 2023, 3pm-6pm
- **Time-series pre-processing (cleaning/downsampling/…)**
- **Time-series visualization**

Data science: how

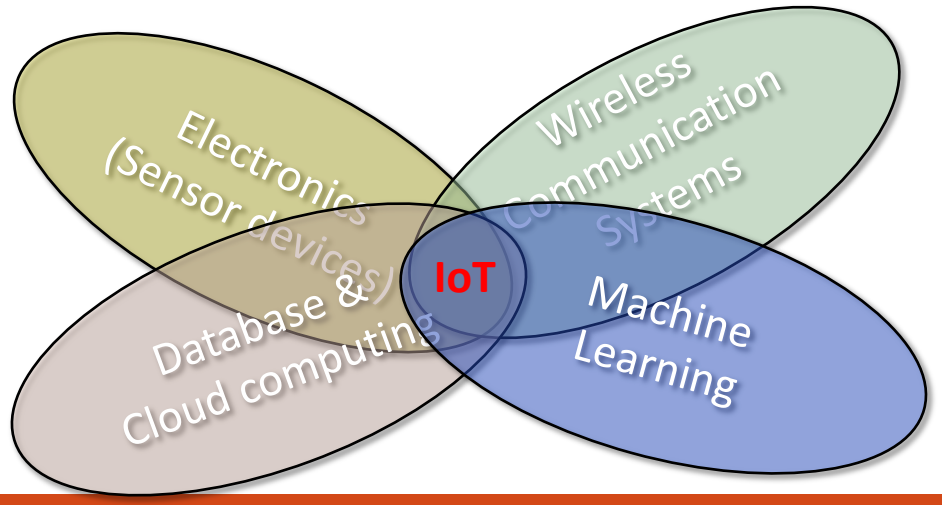
- LESSON 3: July 20, 2023, 3pm-6pm
- **Time-series basics (stationarity, components, etc)**
- **Time-series forecasting via statistical (auto-regressive) methods**

Data science: how

- LESSON 4: July 26, 2023, 3pm-6pm
- **Introduction to Machine Learning techniques**
- **Time-series forecasting via tree-based methods**
- **Time-series forecasting via LSTM architecture**

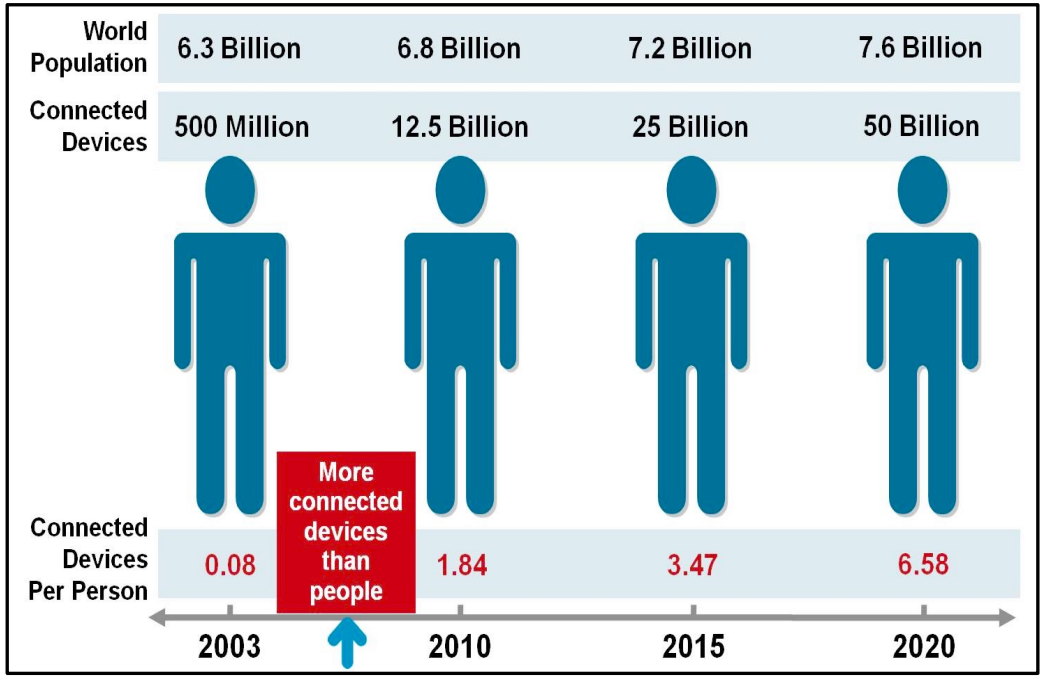
Internet of Things: Intro

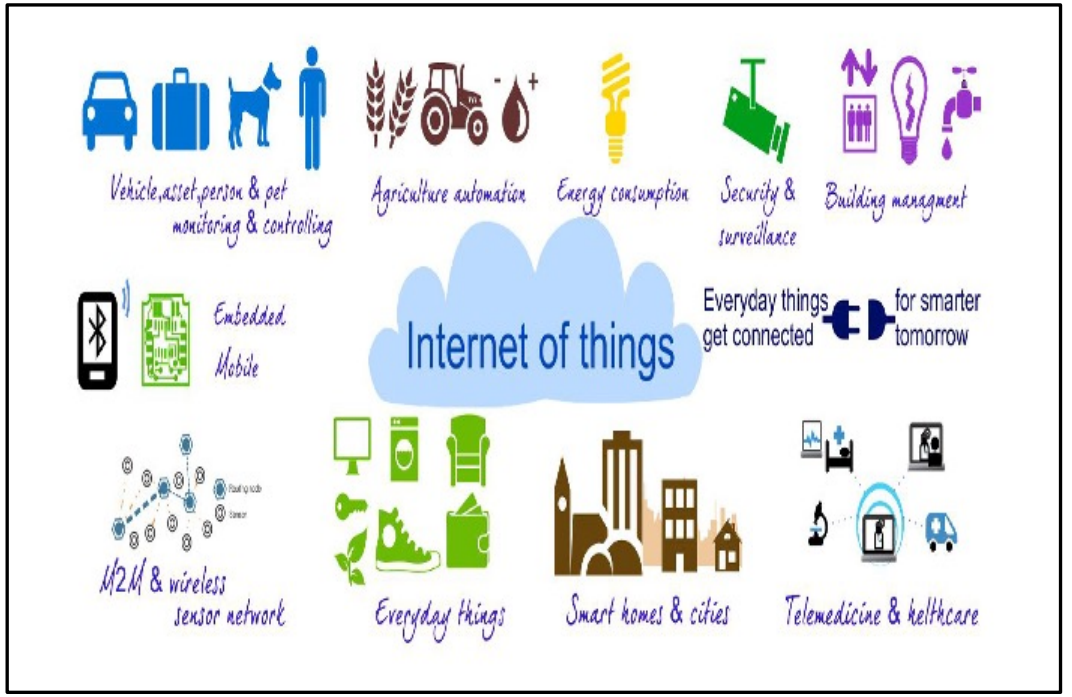
- Pervasive Information and Communication Technology (ICT)
- **GOAL**: Extend the Internet connectivity to **Things**
- Things → Physical (smart) objects, augmented with capabilities to produce data (through sensors), communicate and process data



Internet of Things: Intro

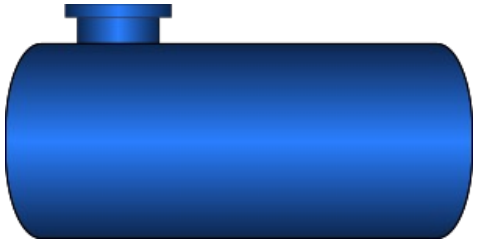
✧ From **Internet of Humans** to **Internet of Things**



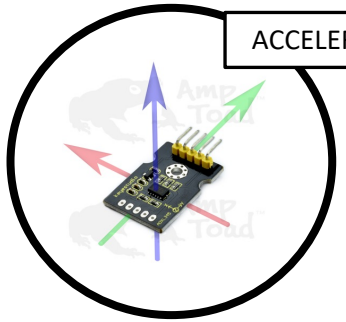


✧ **One single paradigm, many use-cases …**

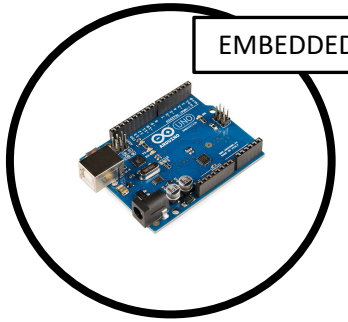
Internet of Things: The full pipeline



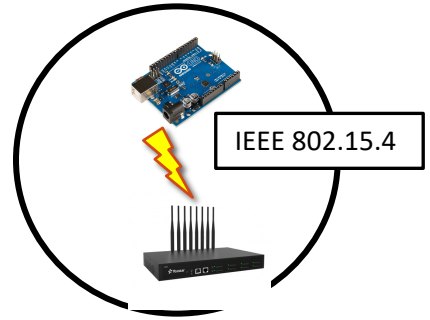
Use case: IoT Monitoring



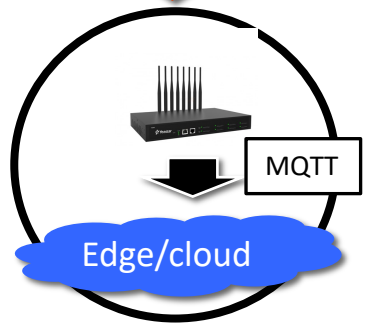
1. **DATA** generation



2. **DATA** acqusition

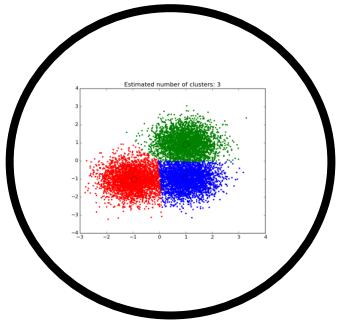


3. **DATA** communication
(sensor → gateway)



4. **DATA communication**
(gateway → cloud)

5. **DATA** storage

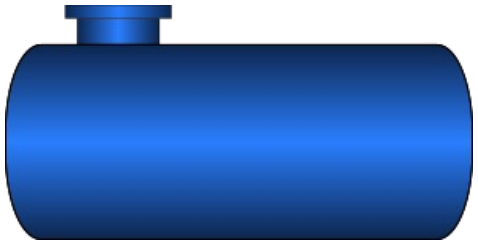


6. **DATA** analytics

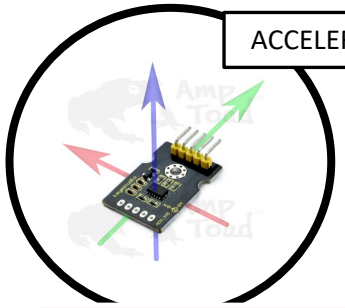
7. **SERVICE** deployment

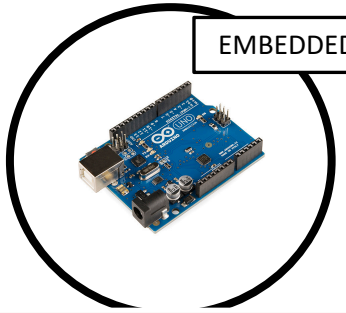
IoT PIPELINE

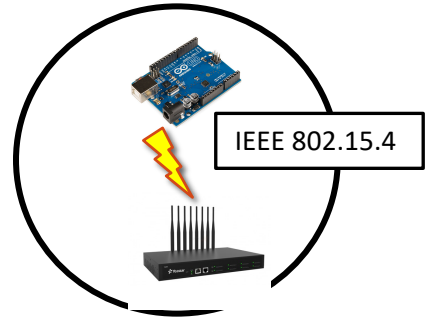
Internet of Things: The full pipeline



Use case: IoT Monitoring





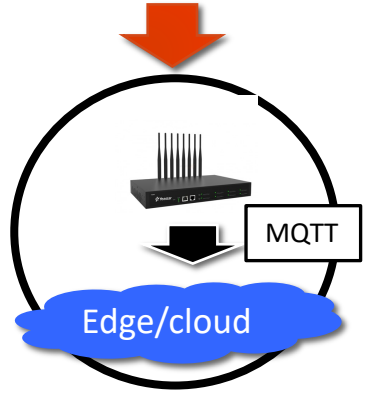


3. **DATA** communication
(sensor → gateway)

1. D

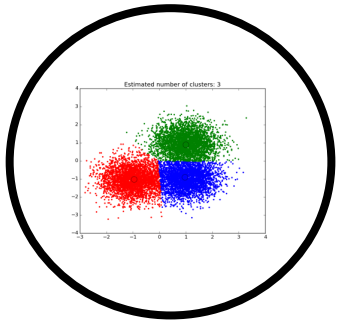
This is our entry point …

IoT PIPELINE



4. **DATA communication**
(gateway → cloud)

5. **DATA** storage



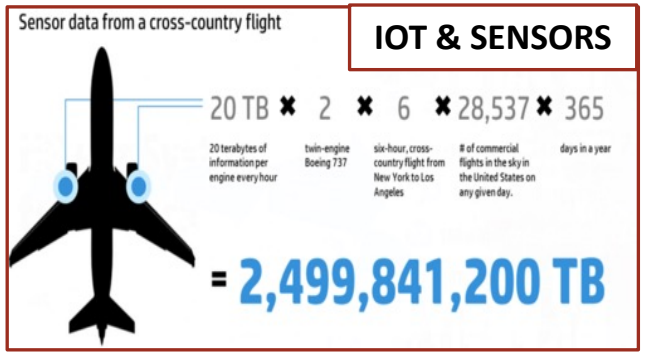
6. **DATA** analytics

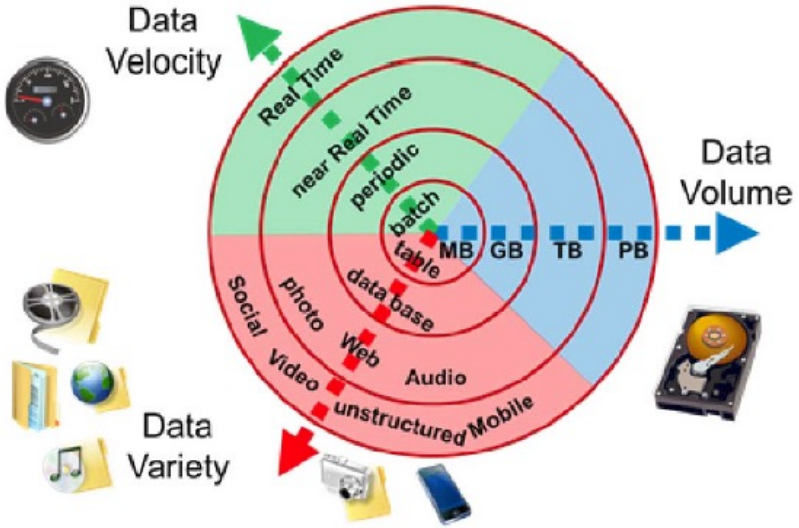
7. **SERVICE** deployment

Internet of Things & big data

- **Big Data**: extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations (definition by the Ofxord Languages)

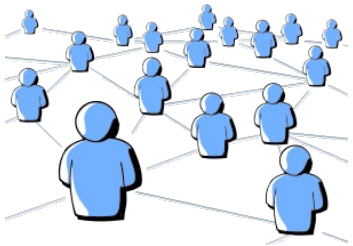
- Volume
- Velocity
- Variety
- Value





By Ender005 - CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=49888192

SOCIAL MEDIA
Twitter: 5700 tweets/seconds



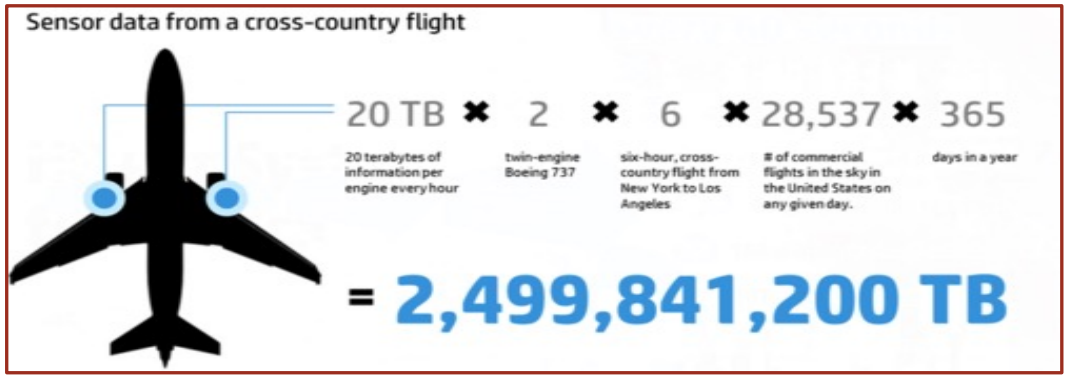
CERN, Large Hadron Collider (LHC)
600 mln of collisions/seconds→
Around 1 PB/s

Internet of Things & big data

✧**IoT** & **Big-data** are two sides of the same coin!

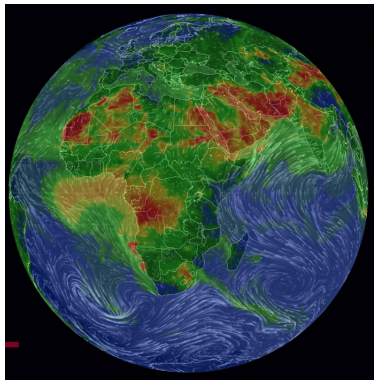
- ❑ Large-scale IoT deployments can produce huge amounts of data
- ❑ Big IoT data can be produced in two ways

FEW SENSORS + HIGH-FREQUENCY SAMPLING RATE



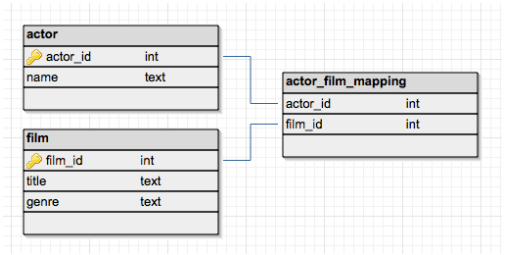
MANY SENSORS + LOW SAMPLING RATE+ LONG-TERM MONITORING

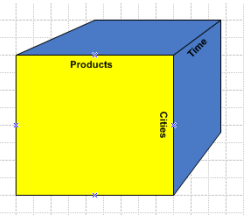
https://www.iqair.com/earth?nav=

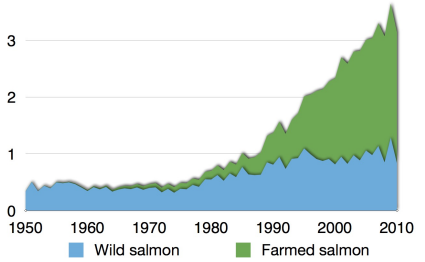


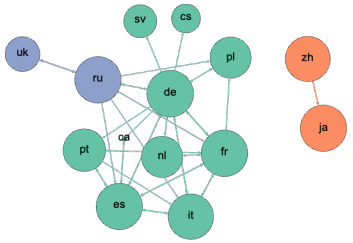
Data types

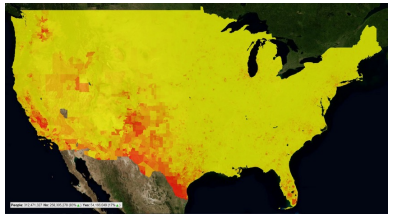
- **Heterogeneous data types**
  - Structured data
    - Transactional data
    - Aggregated data (e.g. warehouse)
  - Unstructured data
    - Textual documents
    - Images
    - Audio & Video
  - Spatial Data
  - Time series
  - Graph
    - Social networks
    - Molecular structures





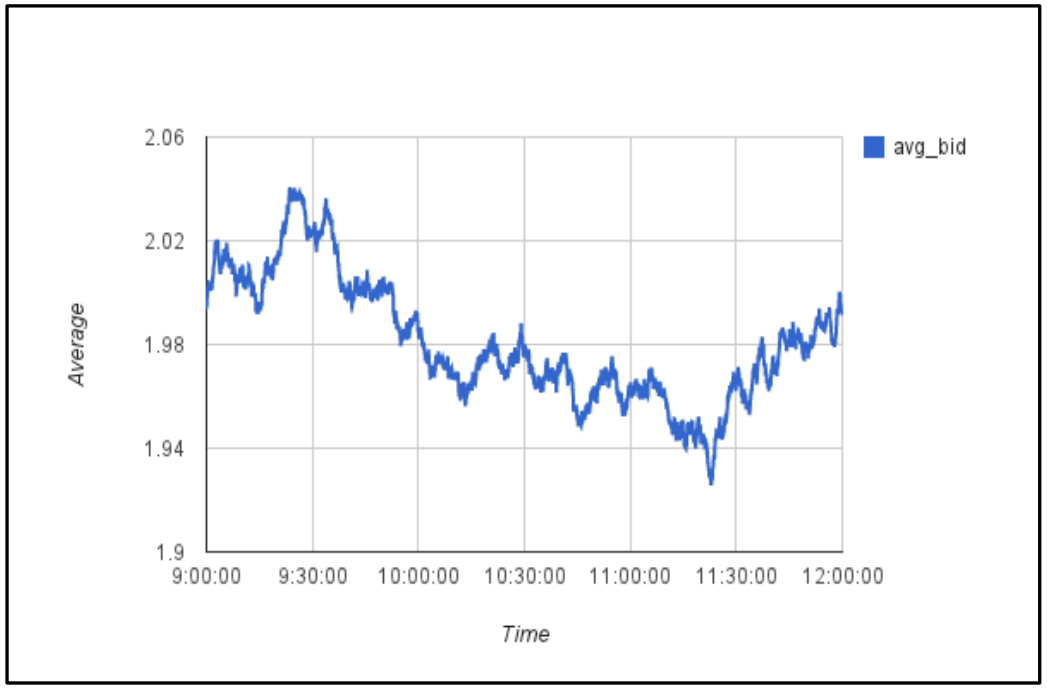


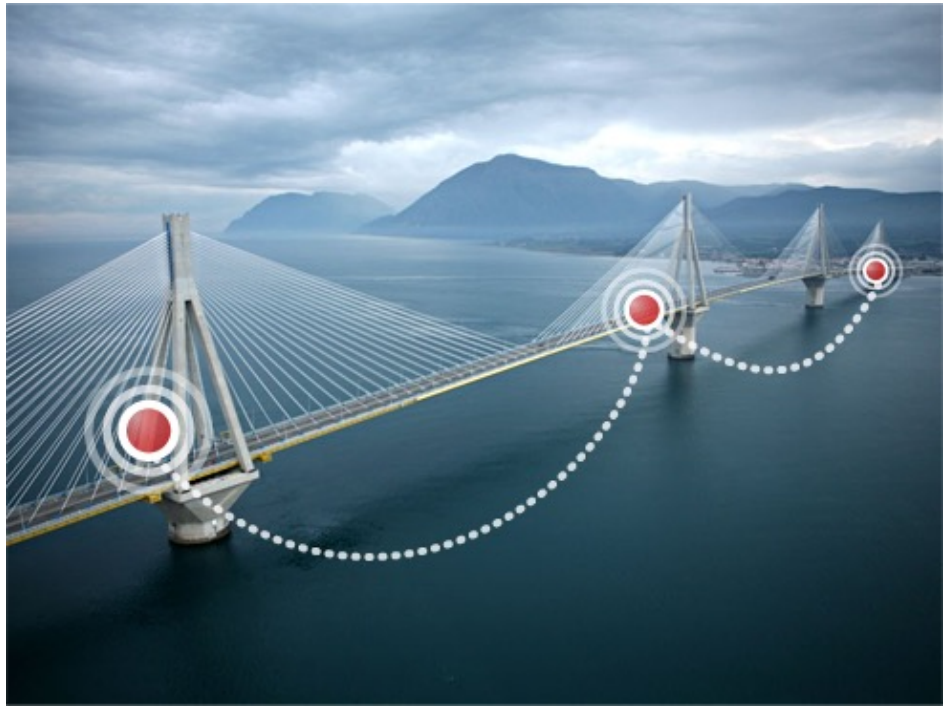




IoT Timeseries

✧Time-series → Sequence of timestamps plus values





IoT Timeseries

✧**Time-series → Sequence of timestamp plus values**

- ❑ Data are **immutable**.
- ❑ Writing in **append**.
- ❑ Reading **contiguous sequence** of samples data.
- ❑ Highly **compressible data**.
- ❑ **High precision** for short period of time.
- ❑ Single value is not **so important**

IoT Timeseries

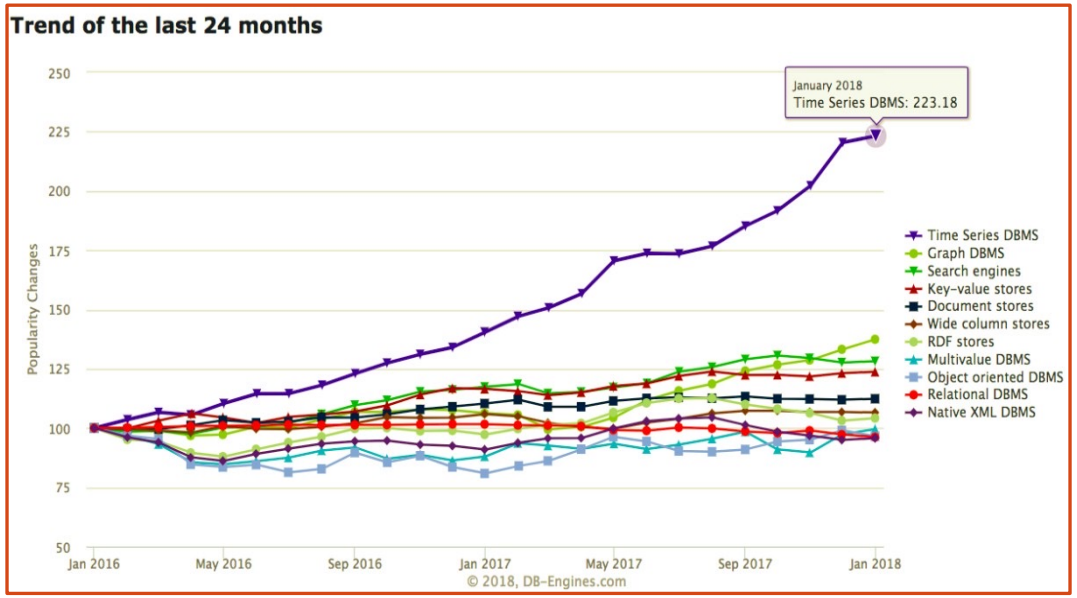
1) Working with time-series: STORAGE

NOT COVERED IN THIS COURSE

TIME-SERIES DATABASE MANAGEMENT SYSTEMS

Enhanced data storage

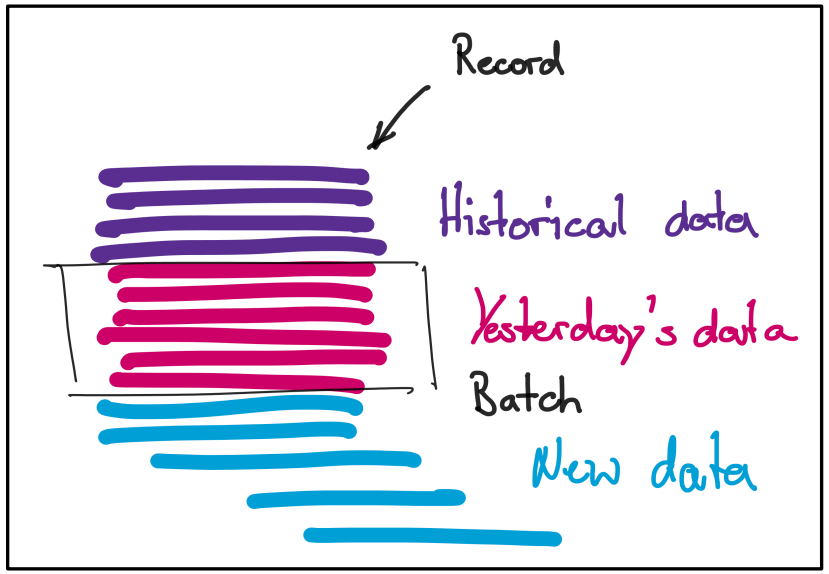
- Operational support (e.g. range-based queries)
- Time-granularity management
- Time-series processing



IoT Timeseries

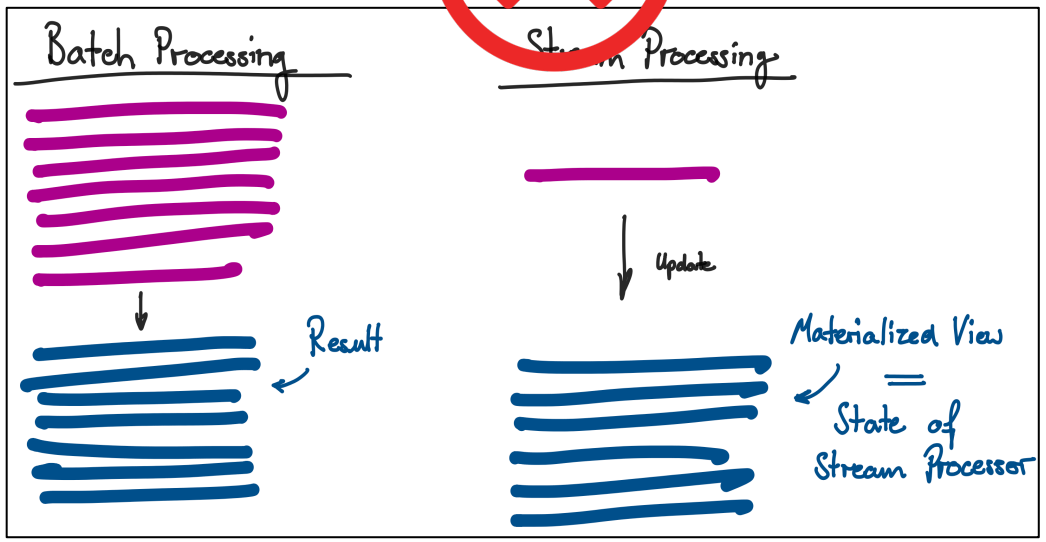
2) Working with time-series: PROCESSING

BATCH PROCESSING



Source: https://blog.philipp-brunenberg.de/entries/batch-stream/

REAL TIME PROCESSING

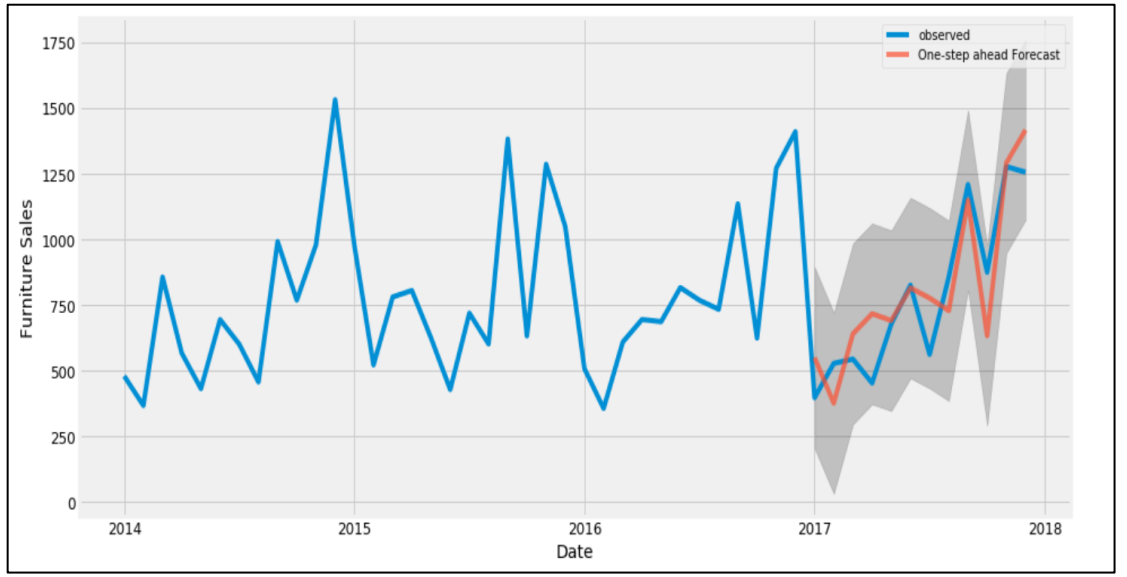


Source: https://blog.philipp-brunenberg.de/entries/batch-stream/

IoT Timeseries

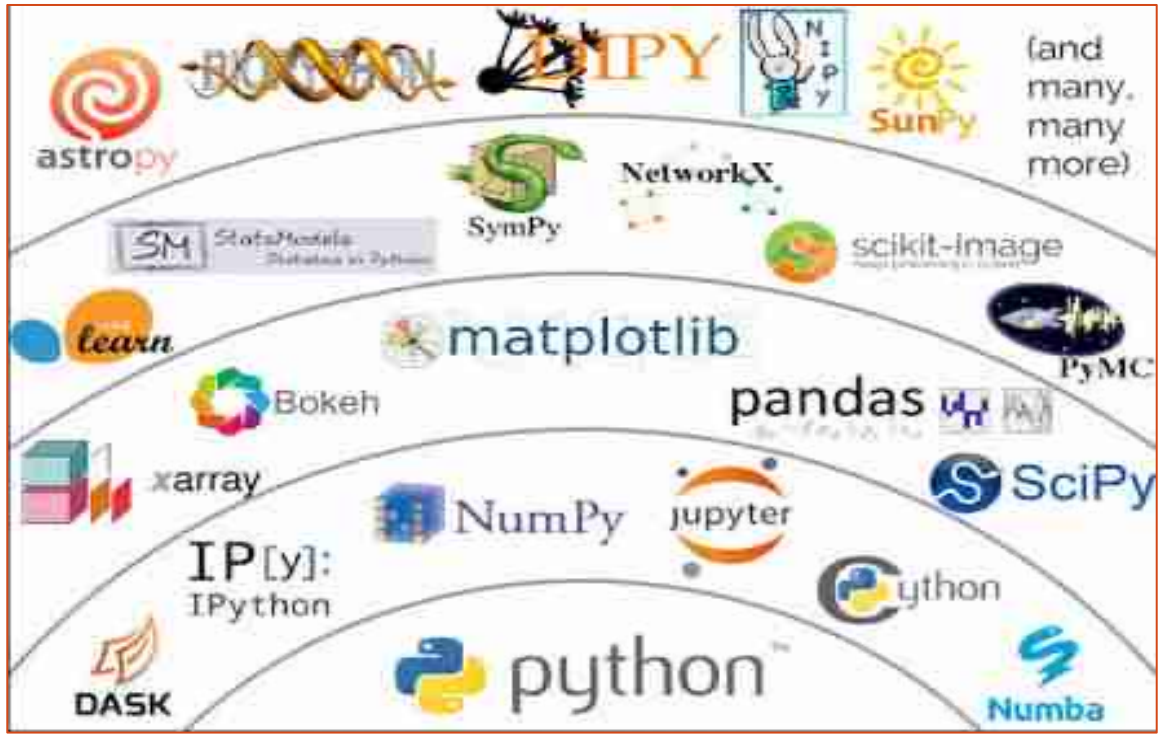
3) Working with time-series: ANALYTICS

TIME-SERIES CLASSIFICATION



TIME-SERIES FORECASTING

Data analytics & Python



Source: https://mc.ai/how-to-master-machine-learning-with-python/