Streaming Data Analytics

Administrative items

Emanuele Della Valle

prof @ Politecnico di Milano
emanuele.dellavalle@polimi.it
http://emanueledellavalle.org

About me

Emanuele Della Valle

— **Associate Professor**, DEIB, Politecnico di Milano, Italy
— 20+ years between academia and industry
— Expert in Semantic Technologies and Stream Processing
— Brander of **Stream Reasoning**
— Serial startupper :-P

Lectures' Timetable

— **Official**
  — Wednesday 14:15 - 16:15 in classroom 5.02
  — Thursday 16:15 - 18:15 in classroom 21.S.2
— **Pragmatically**
  — Tuesday 14:30 - 16:00 in classroom 5.02
  — Thursday 16:15 - 17:45 in classroom 21.S.2
— **Important NOTES**
  — there is no clear cut between *theory* and *practice*
  — *bring your laptops* we will often code
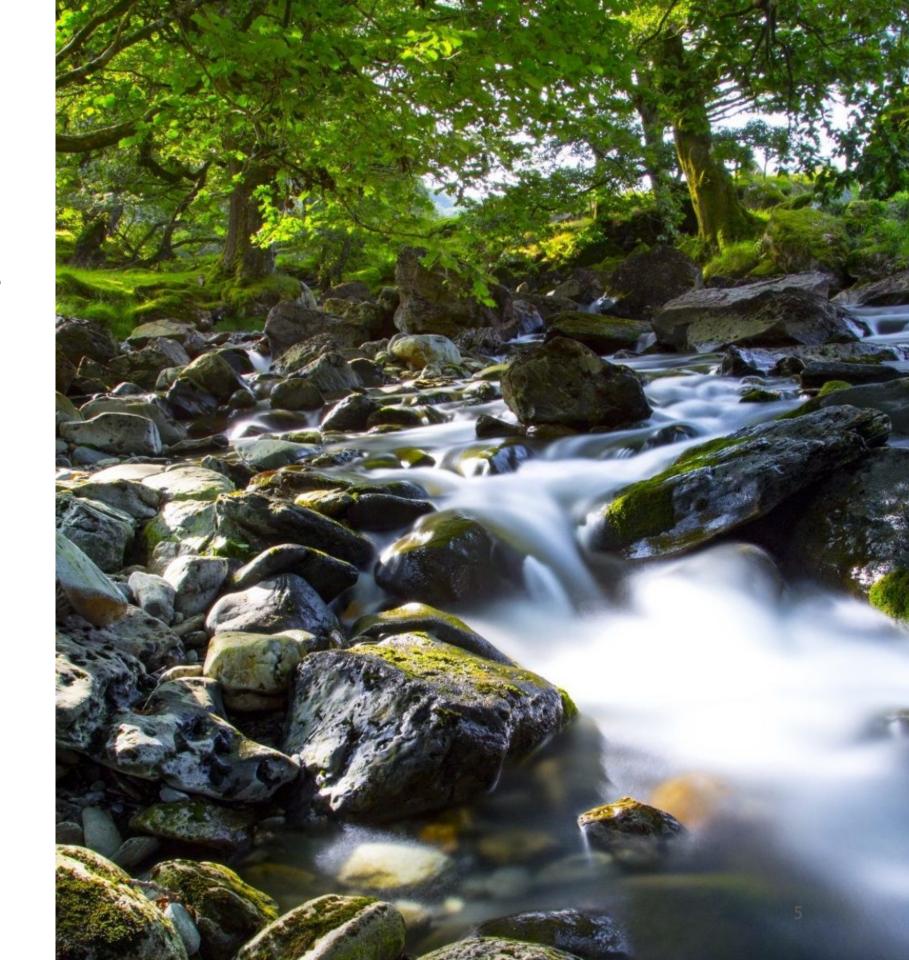 — I **record ALL lectures** but I do not stream them

Lectures' Recording

— I **record** all lectures, and I **make them available the same day**

— I do **no**t **stream** the lectures **outside the classroom** (i.e., I can selectively admit people who are in the room)

Objetives

The course provides the foundational concepts, methods, languages, and systems for **ingesting, processing, and analyzing data that flows** to enable real-time decisions.

The course aims to tame the velocity dimensions of Big Data without forgetting the volume and variety dimensions.

Prerequisites

Students are expected to know the basics of database management and SQL

Topics covered

— Streaming Data Engineering

— Streaming Data Science

IMPORTANT: for tentative scheduling of all lectures, refer to
http://emanueledellavalle.org/teaching/streaming-data-analytics-2022-23/

Topics covered

Streaming Data Engineering

— From the foundations of streaming algorithms to real-world languages and systems

— Languages for Data Stream Management Systems (**DSMS**) and Complex Event Processing (**CEP**) illustrated via EPL

— Horizontally scalable DSMS illustrated via Apache Kafka, Apache Spark, and ksqlDB

— Vertically scalable CEP illustrated via Esper

Streaming Data Science

— From the foundations of **Streaming Machine Learning** to real-world libraries (i.e., MOA and River)

— Learning one sample at a time, prequential evaluation, and concept drift

— Methods for streaming classification

— Ensembles for Streaming Machine Learning

— From the foundations of **Time Series Analytics** to statsmodels-tsa

— Decomposing and detrending time series with and without seasonality

— From stationarity and Time-series forecasting to SARIMA and Recurrent Neural Networks

… more in the next presentation

Expected learning outcomes

— **Knowledge and understanding**

  — Students will learn how to **identify problems** that can be addressed with big data techniques tailored for **velocity** and apply the stream data analysis technologies for **solv**ing real-world problems

— **Applying knowledge and understanding**

  — Given specific **project cases**, students will be able to define and implement a streaming data analysis solution for the problem and apply it to natural data streams from social media and IoT sensors

— **Making judgments**

  — Given specific project cases, students will be able to learn how to **decide** which streaming data analysis solution to apply and how to **evaluate** this decision

Expected learning outcomes

Only those who will do the project

— **Communication**

    — Students will learn to write a report on a project describing and motivating the decisions taken and the results obtained and present their work in front of their colleagues and teachers

— **Lifelong learning skills**

    — Students will learn how to develop a realistic streaming data analysis project in all its phases

Evaluation -

— The exam consist of three parts

    — a **mandatory** written exam

    — an *optional continuous evaluations*

    — an *optional practical project work* with oral presentation

— **Example**:

    — 27 written text

    — + 1 optional continuous evaluation

    — + 3 optional practical project work

    — = 30L

The "mandatory" written exam

— The written exam is composed of

 — a mix of **theoretical questions** regarding any course subjects and

 — **exercises** regarding the technical content and how to apply it in practice.

— Students can get **up to 30/30** in the written test.

The "optional" continuous evaluations

— The optional continuous evaluations are in-presence quizzes proposed **during the lessons**

— Students can get **up to 2 marks** with continuous evaluations

The "optional" practical project

— The optional practical project requires using one or more of the technologies presented in the lectures.

— It **solves a realistic** streaming data analysis **problem** based on actual or practical datasets publicly available or provided by the teachers.

— Students can get **up to three marks** with the practical project

— **Only students, who will get at least 27/30 in the written exam, can opt for it.**

Where to find more information

— The Web page of the course

    — http://emanueledellavalle.org/teaching/streaming-data-analytics-2022-23/

— The Webeep page of the course

    — https://webeep.polimi.it/course/view.php?id=7424

Q / A

Streaming Data Analytics

Administrative items

Emanuele Della Valle

prof @ Politecnico di Milano
emanuele.dellavalle@polimi.it
http://emanueledellavalle.org