REMARKS:

- All projects are about designing and implementing big data systems.
- In all projects we want you to work with real-world data, to understand the challenges of what happens 'in the wild' and to make the problem more concrete.
- We expect you to use big data technologies that you learned throughout the course, even if the actual dataset you are working with is not really 'big data'.
- Always ask yourself 'What if…?'. What if the data was 100x bigger? What if the data would come in as real-time feed through APIs? What if you had to serve 1M concurrent requests? Etc.etc.
- For some projects, we have added an indication of some possible data sources. This is just meant to help you in the initial phase. You are not required to use the data we suggest, so feel free to look for alternatives. Generally speaking, we welcome (and reward) the fact that you work with additional datasets.
- We welcome (and reward) projects that use creative, non-trivial approaches to the problems at hand.
- Every project should start with a data exploration phase (if you don't know the data well, you can't design/build a good system for managing/processing it). Data exploration and design is something you can start working on right away. For the implementation part, take it easy for a while as there are relevant technologies that you still have to see in class.
- In some cases data is available as real-time feed only, and you may want/need to build some historical data, so start monitoring and downloading asap.
- Remember that you can swap projects among teams - just get an agreement no later than 15/4 EOB and inform us by email (with all team members involved in CC).
- We will provide you details about the report (5 pages) and the code to be shared later on - no need to worry (yet) about them

1. [health] Design and implement a big data system for predicting in near-time the expected waiting time in the emergency rooms of hospitals in Trentino and the likely outcome (hospitalization, transfer to another structure etc.). Consider using data from Open Data Trentino (you may need to build your own dataset for training so start collecting data soon).
2. [health] Design and implement a big data system for short-term predictions (daily, up to 7 days) of COVID-relevant metrics (new cases, hospitalization, ICUs, deaths etc.) in Trentino. Consider using open data from FBK covid-19 dashboard.
3. [sport] Design and implement a big data system for classifying how much a given territory and its inhabitants are sportive. Consider using data from OpenStreetMap (sport structure, running routes, cycling paths etc.) and - if available for a given area - public data about sport activities by individual users/groups.
4. [media] Design and implement a big data system for predicting the (short- and long-term) popularity of a recently uploaded YouTube video. Consider using YouTube8M for training purposes (https://research.google.com/youtube8m/index.html).
5. [finance] Design and implement a big data system for supporting the decisions of a real-estate investor who is considering to buy some properties in Puglia to rent them out on short-term contracts. The system should provide RoI data and suggest locations where to focus the investment. Consider using data from InsideAIRB&B (http://insideairbnb.com/get-the-data.html) and OMI/AdE (https://www.agenziaentrate.gov.it/portale/schede/fabbricatiterreni/omi/banche-dati/quotazioni-immobiliari).

6. [finance] Design and implement a big data system for predicting, based on the dynamics of financial indexes (DJ, S&P500, NASDAQ etc.etc.) the short-term dynamics of Bitcoin (BTC).
7. [finance] Design and implement a big data system for detecting and predicting sudden variations (> X% on a daily basis) for a set of cryptocurrencies.
8. [society] Design and implement a big data system for estimating census data of single individuals based on their socio-demographic data (age, gender, address). Consider using ISTAT or similar data.
9. [government] Design and implement a big data system for predicting, based on historical data, the likely winner(s) for a tender published by a public administration body. Focus on 'significant' tenders (>40k€ budget). Consider using data from ANAC.
10. [transportation] Design and implement a big data system for estimating the usage of multi-modal transportation in London and how different statistics are connected (bikes, car parks, metro etc.). Consider using open data from TFL.
11. [finance] Design and implement a big data system for optimising an investment portfolio. Starting from a portfolio of, let's say, shares quoted on the FTSE-MIB, analyse their historical ability to generate returns, and suggest replacements from the same industry to enhance the performance (e.g., replace share of XX with shares of YY - same industry - 'cause this will give you a return higher by ZZ% Y/Y etc.).
12. [government] Design and implement a big data system for suggesting a company, based on historical data, how much discount should be offered for having good chances of winning a tender published by a public administration body. Consider using data from ANAC.
13. [finance] Design and implement a big data system for predicting near-time strong variations on financial assets (shares, bonds, commodities, crypto) based on tweets by relevant personalities (e.g., Elon Musk).
14. [science] Design and implement a big data system for predicting the long-term impact (number of citations or similar metrics) of a recently published scientific work. Consider using data from Google Scholar (but beware of the lack of public APIs).
15. [society] Design and implement a big data system for monitoring and estimating, based on tweets, the birth rate in various parts of Europe and its dynamics. Consider using Twitter APIs + data from Eurostat.
16. [energy] Design and implement a big data system for predicting short- and mid-term production by renewable energy sources and potential imbalances in generation/load. Consider using data from TERNA (https://www.terna.it/it/sistema-elettrico/transparency-report/download-center).
17. [environment] Design and implement a big data system for rating the performance (accuracy) of a weather service. Consider using Meteotrentino as an example and data from Open Dati Trentino as ground truth (you may need to build your own dataset for training so start collecting data soon). (Of course you can also consider other services, fully up to you.)
18. [society] Design and implement a big data system for automatically building a leaderboard of Italian cities in terms of quality of life. Consider using data from open data portals, OpenStreetMap, authoritative sources, statistical institutes (e.g., ISTAT) and news.
19. [transportation] Design and implement a big data system able to predict, for a given day and time slot (duration 1 hour), the availability of parking slots in the Bolzano area. Consider using the data available from Open Data Hub Südtirol.
20. [energy] Design and implement a big data system able to predict the usage of e-charging stations and plugs in South Tyrol and their status for a given day and time slot (duration 1 hour). Consider using the data available from Open Data Hub Südtirol.

21. [transportation] Design and implement a big data system able to predict in near-time the traffic associated with specific and crucial points of interest monitored by Bluetooth sensors in the Bolzano area. Consider using the data available from Open Data Hub Südtirol.
22. [environment] Design and implement a big data system for predicting the traffic in the Trento area by leveraging the historical dataset available in the Trentino Data Hub. Consider using weather data as an input.
23. [environment] Design and implement a big data system capable of predicting on different time horizons (3 hours, 12 hours, day, week) the status of rivers in the Bolzano area (e.g. water level, flow rate, water temperature) based on historical measurements. Consider using the data available from Open Data Hub Südtirol.