

this year project





DEBS 2024

18TH ACM INTERNATIONAL CONFERENCE ON
DISTRIBUTED AND EVENT-BASED SYSTEMS

June 25th–28th 2024

Lyon, France



Venue Information

The 2024 DEBS Conference will be held in the facilities of INSA Lyon situated on the LyonTech-la Doua campus.

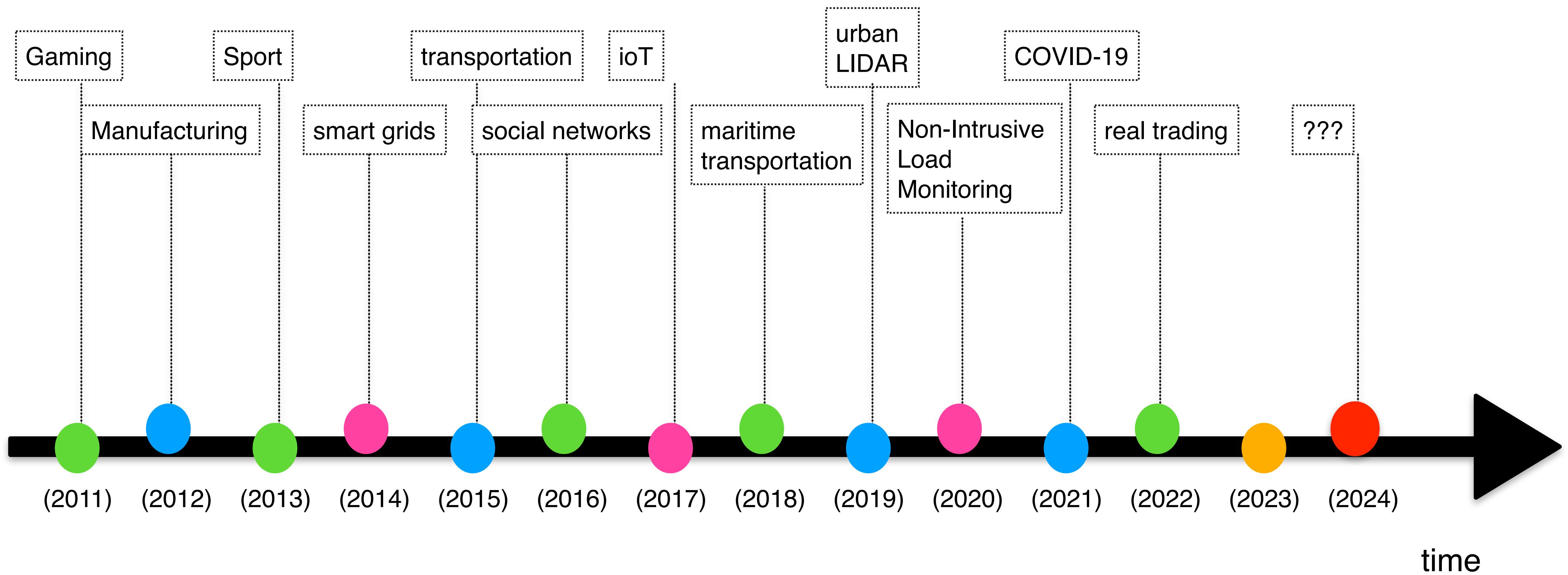
Check out details about the venue in the [venue information](#) page.

Objective

Important Dates

Events	Dates (AoE)
Research Papers	
Abstract Submission	February 16th, 2024
Paper Submission	February 22nd.

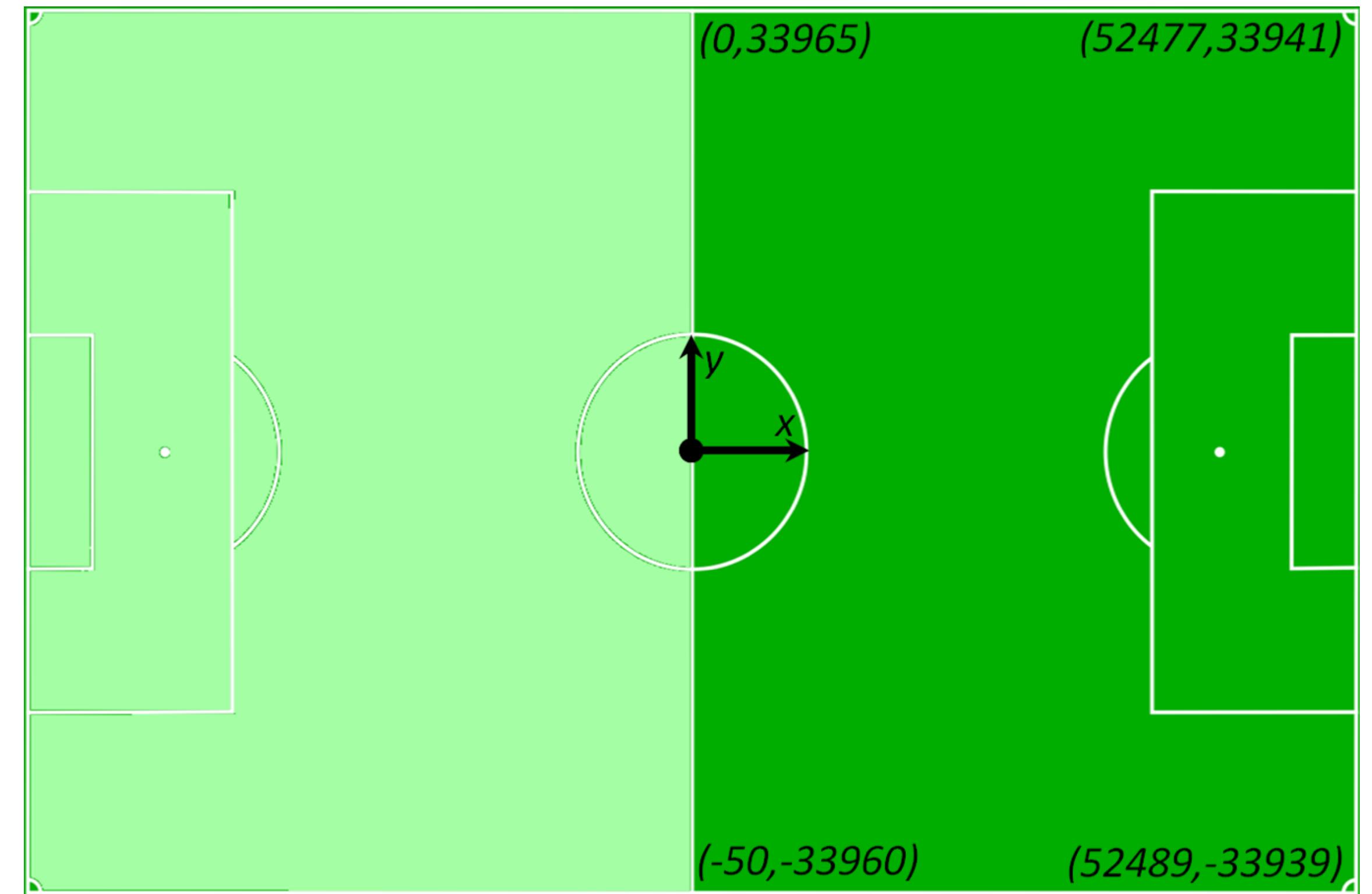
DEBS Challenges



Example 1

In 2013

- The 2013 edition of the Grand Challenge focuses on real-time, event-based sports analytics.
- data set was collected during a football match carried out at a Nuremberg Stadium in Germany



Example 2

In 2015

- The focus of the DEBS 2015 Grand Challenge is on processing of data streams originating from the New York City Taxi and Limousine Commission.
- The data is made available under the Freedom of Information Law and provides information pickup, drop off, and payments made in New York City medallion taxis.



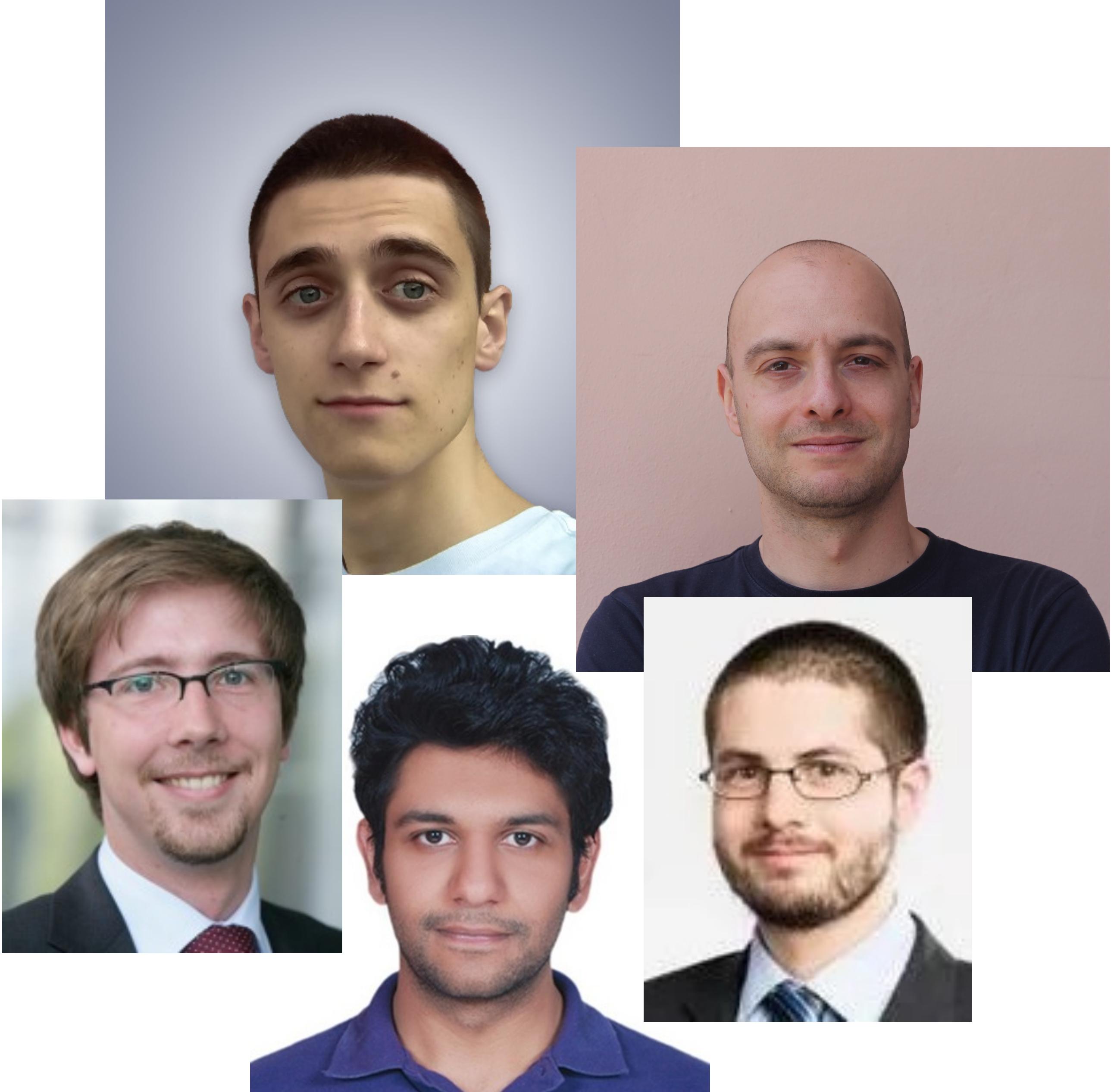
2024 DEBS Challenge

Telemetry data for hard predictive drive failure & maintenance

Who's Behind the Challenge

This Year Chairs

- Luca De Martini, Politecnico di Milano
- Alessandro Margara, Politecnico di Milano
- Sebastian Frischbier, Allianz Global Investors
- Jawad Tahir, TU Munich
- Christoph Doblander, TU Munich



Participation

Consists of three steps:

- (1) registration,
- (2) iterative solution submission, and
- (3) paper submission.

Registration

Submit your registration abstract to Easychair <https://easychair.org/my/conference?conf=debs24> in the "Grand Challenge Track"

Send an email to one of the Grand Challenge Co-Chairs (see <https://2024.debs.org/organizing-committee/> to obtain access to the platform landing page).

Solutions to the challenge, once developed, must be submitted to the evaluation platform (<https://challenge2024.debs.org/>) to get benchmarked in the challenge.

Reporting

The last step is to upload a short paper (minimum 2 pages, maximum 6 pages) describing the final solution via the central conference management tool Easychair.

The DEBS Grand Challenge Committee will review all papers to assess the merit and originality of submitted solutions.

All solutions of sufficient quality will be presented during the poster session at the DEBS 2024 conference.

You are very welcome but not required to submit the paper to the conference, but to us!

2024 DEBS Challenge

Telemetry data for hard predictive drive failure & maintenance

The Actual Challenges

This year's DEBS Grand Challenge requires you to implement two queries

- Query 1: Count of the recent number of failures detected for each vault (group by storage servers) (Continuous Querying)
- Query 2: Use this number to continuously compute a cluster of the drives (Streaming K-Means)
- Query 3 (optional): Creativity Challenge: what else can I do with this data?
- Additional clarifications to queries will be provided here: <https://challenge2024.debs.org/documentation/>

Input

Input data consists of batches and each batch contains

- SMART readings for a list of drives
- vault_ids: a list of vault identifiers of interest for this batch (used in Q1, see below)
- cluster_ids: a list of cluster identifiers of interest for this batch (used in Q2, see below)
- day_end: a flag that marks the end of one day of readings

Query 1

- For every vault v , count the number of failures NF_v in a sliding window W
 - Size = 30 days
 - Slide = 1 day
- For a given day i , NF_v^i is the count of failures in the window that starts at $i-31$ (included) and ends at $i-1$ (included)
- The first window closes at day 0: you can assume 0 failures for every day $i \leq 0$
- Each batch of input data will contain the identifiers of 5 vaults: you are to return the current value of NF_v for those vaults

The Actual Challenges

- For each reading at day i regarding a drive d belonging to vault v , add NF_v^i to the reading
 - Normalize data
 - Rescale the smart values within the provided ranges
- The ranges can be downloaded from here
- Compute dynamic K-means clustering
- Assign incoming readings to the nearest centroid
- At the end of one day, which is marked by a ‘day_end’ flag in the batch, update the centroids positions with the average coordinates of all the readings currently associated to that centroid
- The initial coordinates of centroids are provided, and can be downloaded from here
- Each batch of input data will contain a list of cluster identifiers (clusters are identified by a sequence number): you are to return the number of drives associated to each of those clusters

The Data

The data provided for this DEBS Grand Challenge is based on S.M.A.R.T (https://en.wikipedia.org/wiki/Self-Monitoring,_Analysis_and_Reporting_Technology) data and additional attributes captured by Backblaze on a daily basis for six months (starting Q2 2023).

The data set contains events covering over 200k hard disks. Each data point resembles the S.M.A.R.T. status of a dedicated disk on a specific day.

The full data set will be used for benchmarking the submitted solutions. The smaller test data set will be provided upfront via our eval platform for testing purposes.

The test data set contains event notifications for a small number of drives. Consequently, participants must not pay attention to filtering out event notifications that do not contain attributes relevant to the Grand Challenge.

Selection Process

- Participants of the challenge compete for two awards:
 - The performance award determined through the automated evaluation platform (see above), according to precise evaluation criteria that measure the speed and correctness of submitted solutions
 - The audience award winner will be determined amongst the finalists who present in the Grand Challenge session of the DEBS conference (YOU need to submit)

Awards

- There are two ways for teams to become finalists and get a presentation slot in the Grand Challenge session during the DEBS Conference (YOU NEED TO SUBMIT):
 - up to two teams with the best performance (according to the final evaluation) will be nominated;
 - the Grand Challenge Program Committee will review submitted papers for each solution and nominate up to two teams with the most novel concepts.
- All submissions of sufficient quality that do not make it to the finals will get a chance to be presented at the DEBS conference as posters.

Criteria

Every submission will be evaluated for how it addresses our non-functional requirements in its design and implementation while assuming that all submitted solutions strive for maximum performance. In this regard, we distinguish between hard and soft non-functional requirements.

- **Hard criteria** are must-haves that have to be addressed in the design and must be implemented, e.g., Configurability, Scalability (horizontal scalability is preferred), operational reliability/resilience, accessibility of the solution's source code, integration with standard (tools/protocols), documentation.
- **Soft criteria** do not necessarily have to be implemented but must be covered in the submission's description, e.g., Security measures implemented/addressed, deployment support, portability/maintainability, and support of special hardware (e.g., FPGAs, GPUs, SDNs,...).

Dates

Release of the challenge, initial data set: November 6th, 2023

Test platform opens: December, 2023

Evaluation platform opens: February, 2024

Deadline for uploading final solution to the evaluation platform: May 12th, 2024

Evaluation of the solution: May 19th, 2024

Deadline for short paper submission: June 2nd, 2024

Notification of acceptance: June 9th, 2024

Camera ready submission: June 16th, 2024

Conference June 25th – 28th June, 2024 here at the INSA Library!!!