Poznań University of Economics and Business
OFFIS Institute for Information Technology

MarTERA
HANSA

14th NATO ORA Conference

Application of AI and In-memory Computing for Extracting Vessel Movement Patterns from Historical Data

Presenter: **Krzysztof Węcel**
krzysztof.wecel@ue.poznan.pl

Costa Concordia Disaster

January 13th, 2012

Taken by Roberto Vongher, CC BY-SA 3.0, Wikimedia Commons

> Italian cruise ship Costa Concordia ran aground and overturned after striking an underwater rock off

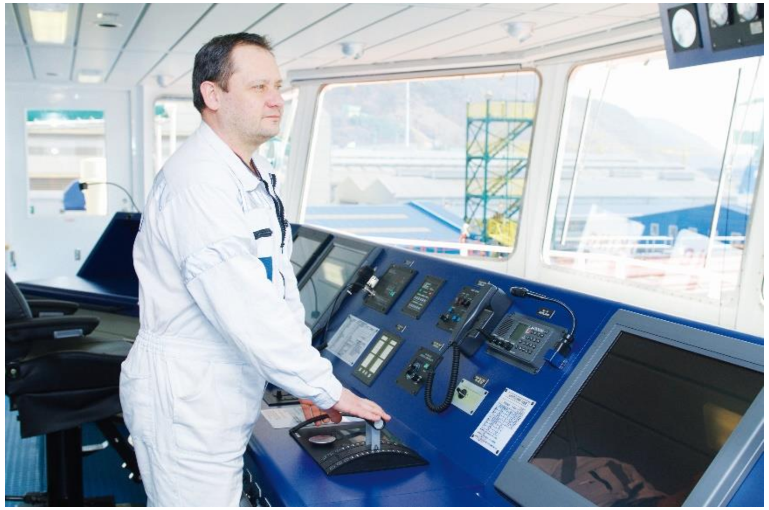
> 4,252 people on board, 32 deaths

> Total cost of the disaster estimated at approx. $2 billion (3x construction cost)

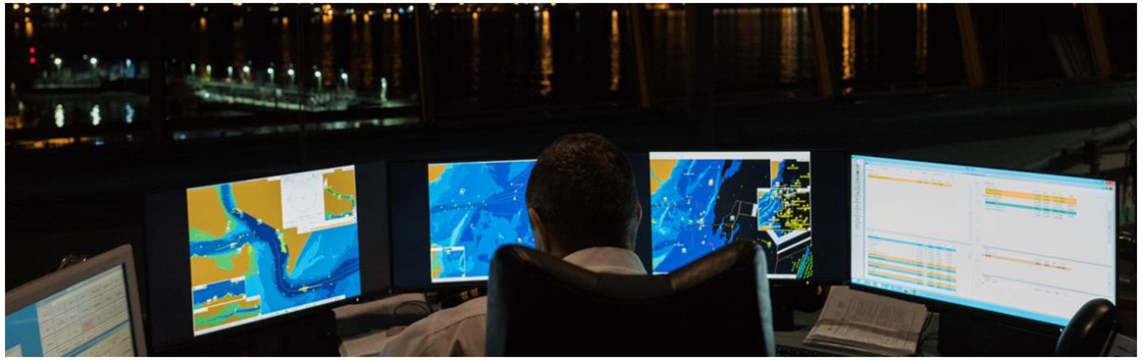
> Navigation by sight and experience

Working questions

On-board



On-shore

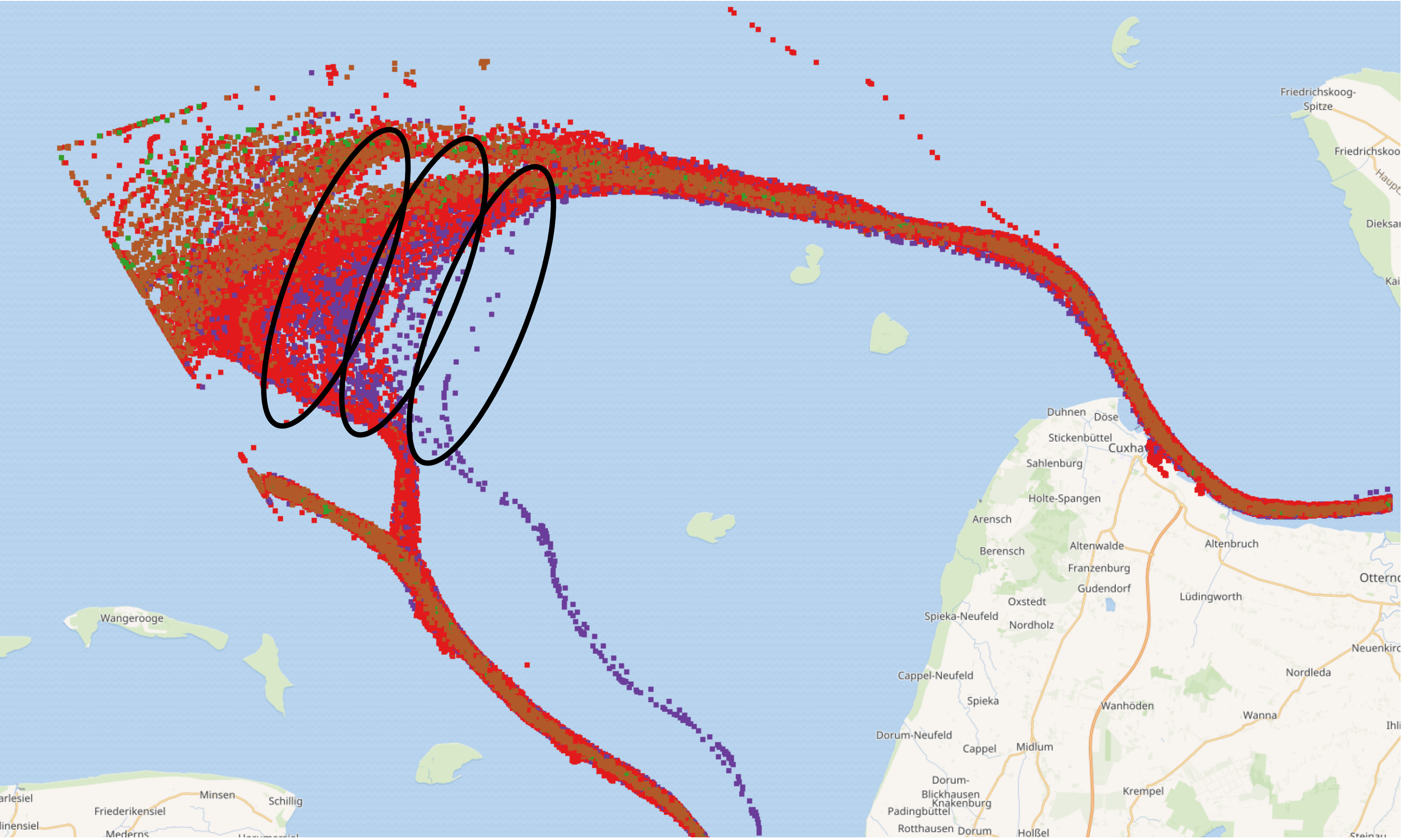


https://hha.co.uk/wp-content/uploads/2017/01/af_8-2400x750.png

> **Route checking functionality**
>
> > How can navigators be supported during route planning and monitoring to avoid unsafe waters?

> **Anomaly Detection**
>
> > How can potentially dangerous vessel behavior be automatically detected on-shore?

Concept of recommended routing corridor

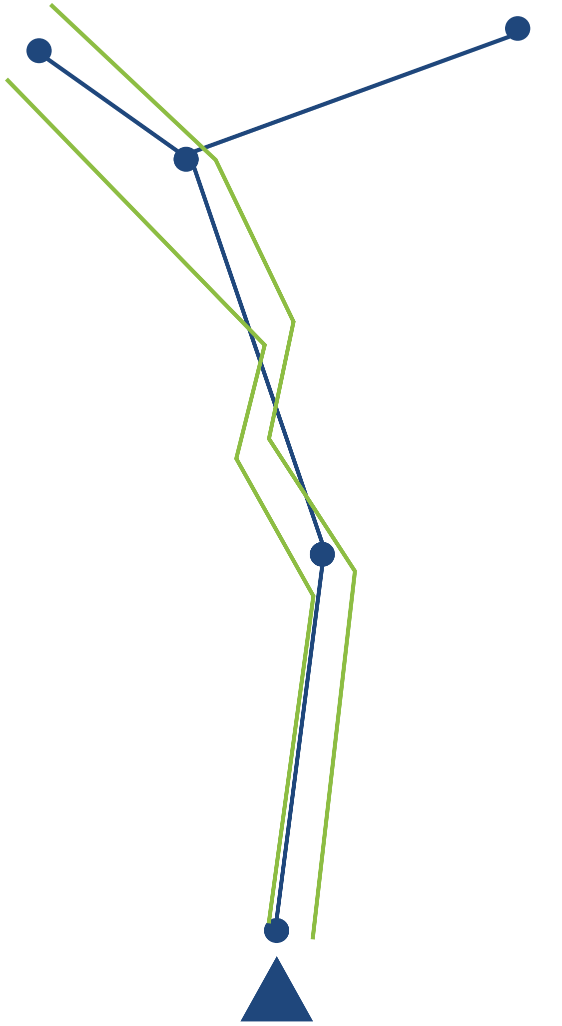


„A recommended corridor (RC) describes a safe route from a port of departure or ship position to destination. This corridor is most commonly used by ships with similar attributes (…) and under similar weather conditions (…).“

Also considers:

> *Direction of travelling/obligation to pass right*

Route Planning



Route checking functionality

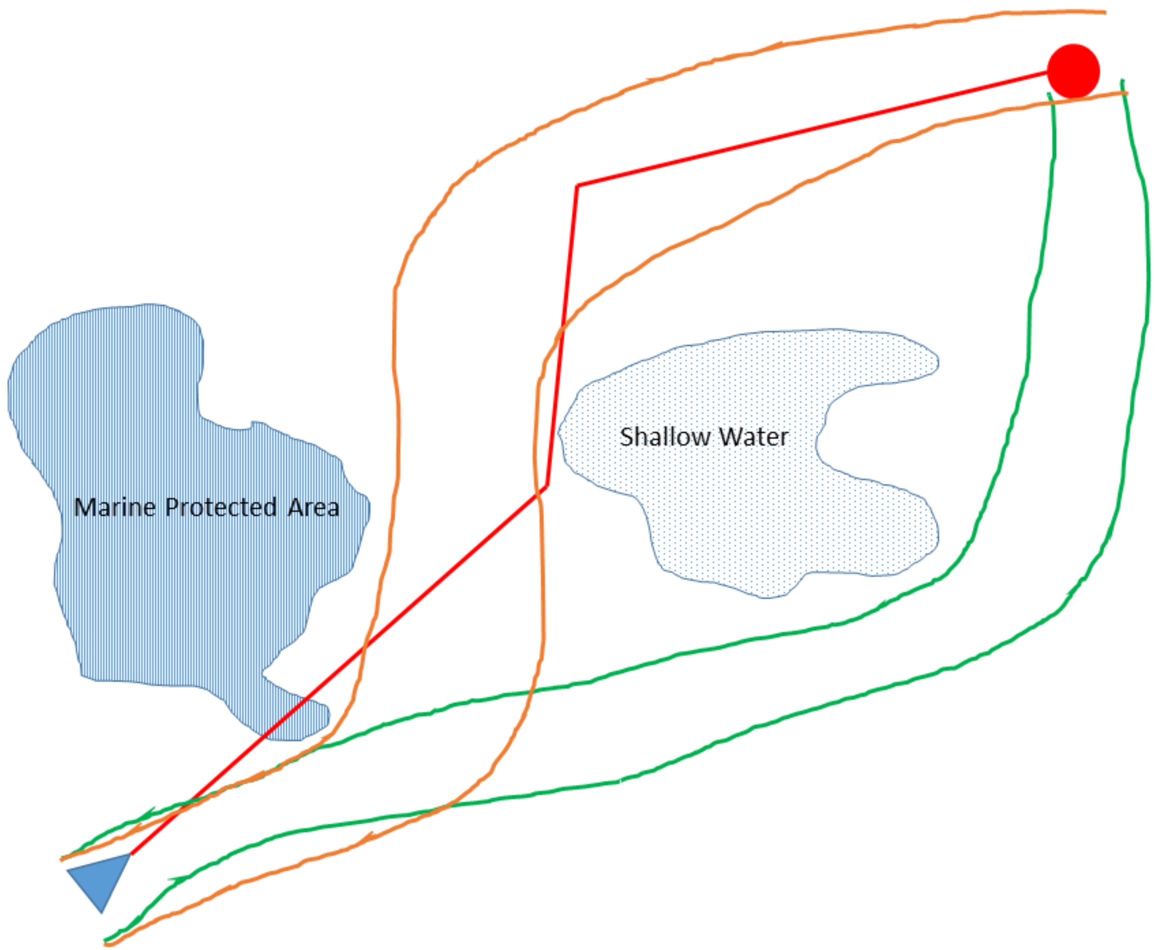
> Checks an intended route against where similar vessels have sailed in the past

> Considering vessel's attributes like type of vessel, draught or length

> Considering environmental influences like current or wind

Recommended Routing Corridor Concept → Data Preparation & Integration → Data Analysis and Anomaly Detection → System Integration and Demonstration → Exploitation

Concept of recommended routing corridor



Key:

Unsafe planned route

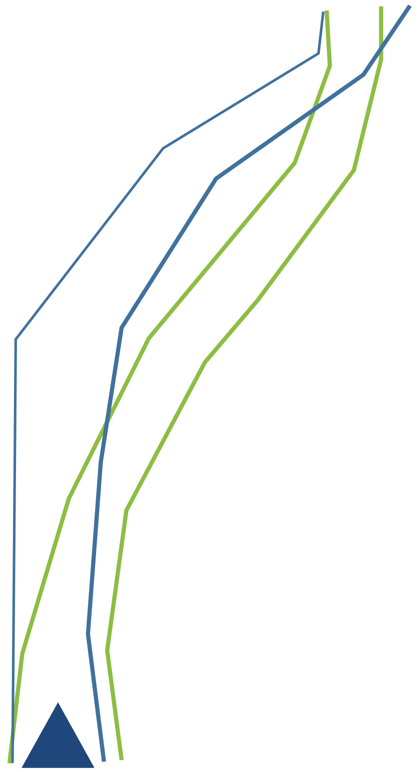
RC alternative 1

RC alternative 2

Anomaly Detection

Bad weather route

Good weather route



> Is the current vessel position inside the RC, recommended for the current weather conditions?

> Is the vessel faster/slower as usual?

Recommended Routing Corridor Concept → Data Preparation & Integration → Data Analysis and Anomaly Detection → System Integration and Demonstration → Exploitation

Why AI/machine learning is need

- huge volume and velocity of data
  - about 1 GB (gigabyte) of data daily
- classical operational research (OR) algorithms do not scale
  - time necessary to compute increases significantly when size of input increases
- more complex scenarios
  - calculations need to be repeated with additional constraints

The method for traffic patterns extraction

- **all possible routes represented as a mesh (a network)**
  - nodes
    - maneuvering points identified by CUSUM
    - then clustered using genetic algorithm (GA) -> waypoints
      - run on a partitioned data (k-d trees)
  - edges
    - nearest waypoints need to be identified for each AIS position
    - reconstruction based on trajectories
    - several filtering and quality methods implemented

* CUSUM – cumulative sum control chart, the method used in statistical quality control for change detection

Edge reconstruction - refinement

> **FILTERAIS**

  > selects a subset of data for a given vessel type (e.g. tanker) or weather conditions (e.g. heavy wind)

  > used to build a mesh from a subset of points, e.g. only important maneuver points as identified by CUSUM

> **FILTERTRAJECTORY**

  > the function is applied to trajectories of a ship

  > selects points fulfilling a condition in trajectories, e.g. only points that are sent within specific time period

> **FILTEREDGES**

  > select edges fulfilling a condition, e.g. can filter out edges that are too long (e.g. distance > 100 Nm) or are very rare (e.g. followed by only a single ship)

Scope of data

Historical vessel movements

> AIS* data from 2017 to 2018
>
> geographical coverage: the German Bight and the Baltic Sea

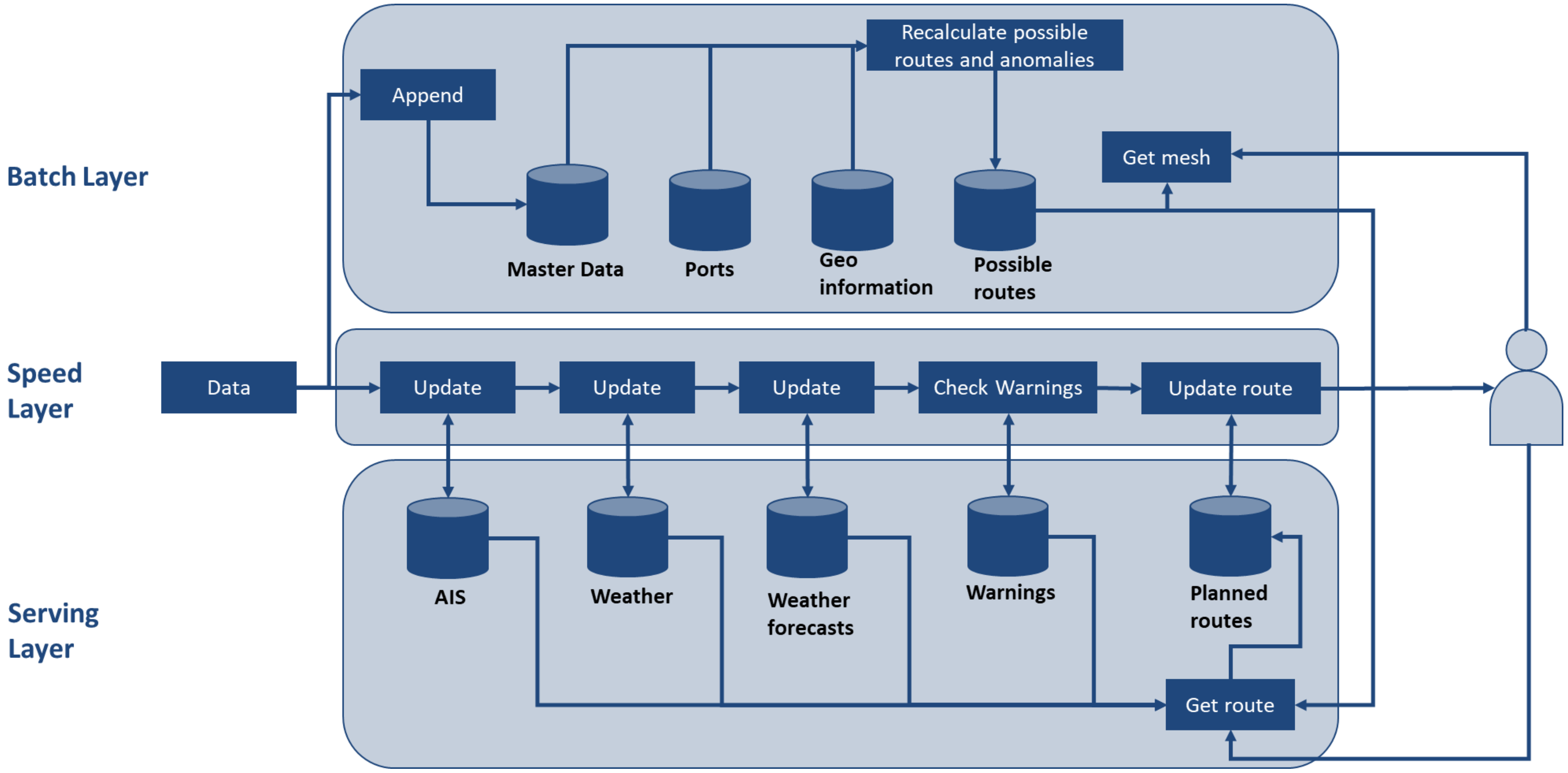
Weather data

> sea weather data obtained from Copernicus

* AIS – Automatic Identification System, used by vessels to report their positions, originally used for collision avoidance

System architecture

- > Pattern extraction is costly
- > Patterns can change as soon as new data is available
- > Request must be answered within seconds
- > Lambda Architecture - a *data processing architecture* to separate slow operations (pattern extraction) from current requests



Implementation

- Apache Spark for calculations
  - Spark is a unified analytics engine for big data processing
- Python as an implementation language
  - PySpark (implementation of Spark in Python)
  - k-nearest neighbour (kNN) from scikit-learn
  - optimization based on user-defined functions (Pandas UDF)
- performance:
  - good implementation decisions can result in performance 100,000x faster than naïve solutions

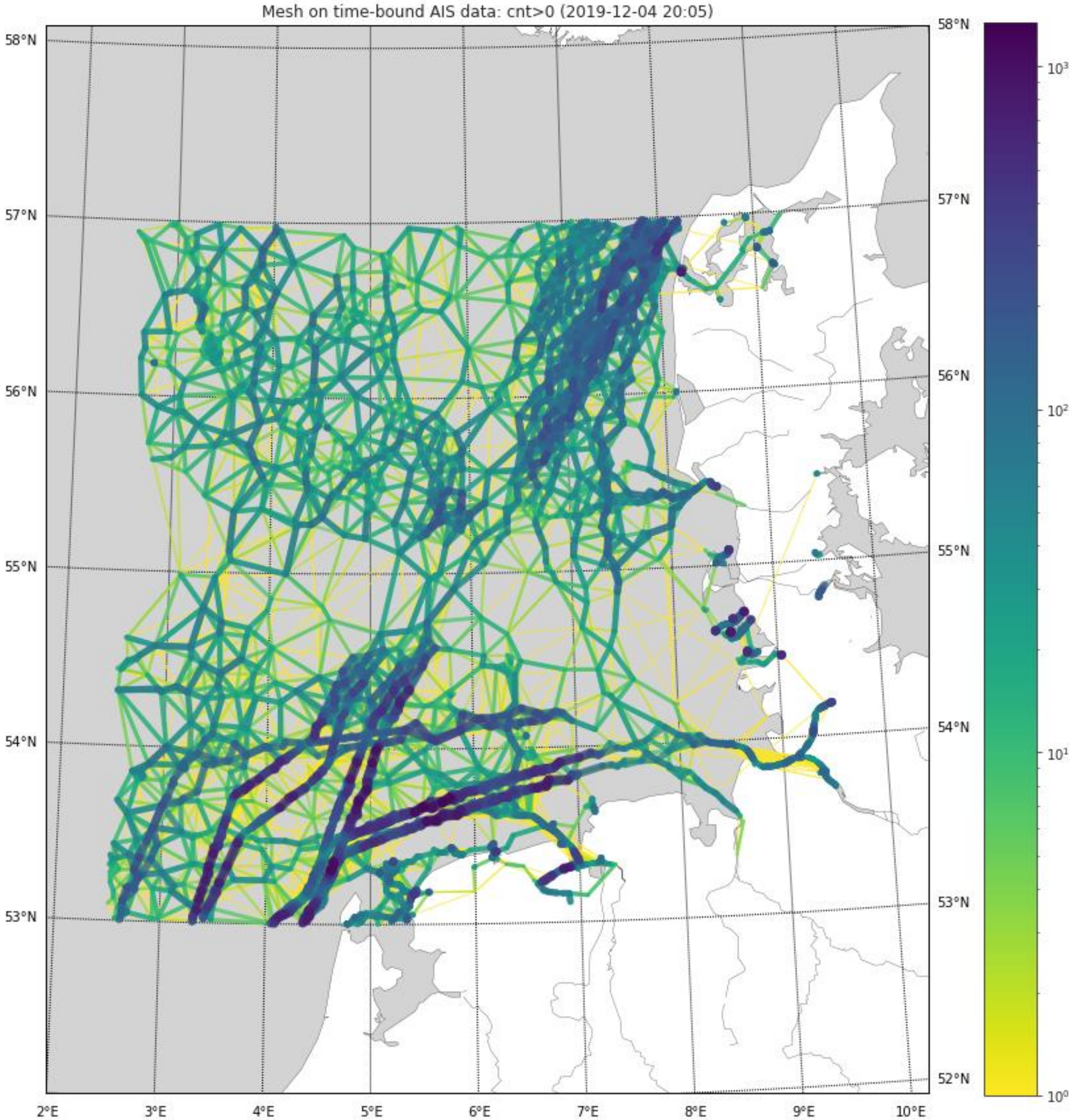
Method	Throughput (rows/s)
kNN, iteration over RDD with flatMap	7.3
kNN, with UDF	399
Brute-force search, with UDF	16,508
kNN Minkowski, with Pandas UDF	1,517,969
kNN haversine, with Pandas UDF	824

Results

> edges generated from about 1.5 million notifications of vessel positions

> the darker the edge, the more vessels travelled between the nodes

> number of edges in this example: 156,103



Summary

> predictive voyage planning as a prerequisite for safe and efficient vessel operations
> recommended corridor (RC) - the method for representing the context-sensitive vessel movement patterns
> mesh – generated on real historical AIS data
> significant work devoted to improving performance
> potential for on-board and on-shore operators

Acknowledgement

> This work has been conducted in the context of the HANSA project, which is funded by the MarTERA partners German Federal Ministry of Economic Affairs and Energy (BMWi), Polish National Centre for Research and Development (NCBR) and Research Council of Norway (RCN) and is co-funded by European Union's Horizon 2020 research and innovation program under the framework of ERA-NET co-fund.