

URBAN CO-CREATION DATA LAB

SUPERCOMPUTING RESOURCES REPORT

DECEMBER 2020

MS 7 - Report on supercomputing resources

Milestone Title	MS 7 - Report on supercomputing resources
Related Activity	Activity 2: Services analytics refinement and HPC preparation
Related Task	HPC and services set-up and deployment
Author's	NOVA IMS
Due Submission Date	30.11.2020
Actual Submission	31.12.2020
Action Number	2018-EU-IA-0099
Agreement Number	INEA/CEF/ICT/A2018/1837945
Instrument	Connecting Europe Facility (CEF) - Telecommunications Sector
Starting Date of Action	01.10.2019
Completion Date	30.09.2021
Main objective	Set-up the necessary HPC platform and deploy analytical services, with connected Lisboa city open data, ready to conduct experiences and validation of use cases

Table of Contents

1	Executive Summary	4
2	Objectives	4
2.1	Strategic Objective	4
2.2	Operational Objective	4
2.3	Task	4
2.4	Outputs:	5
2.5	Milestones and means of verification	5
3	Assessment and use of HPC resources	5
3.1	Supercomputing Resources	8
3.2	Case Study	9
3.2.1	Objective	9
3.2.2	Tasks	9
3.2.3	Case studies	10
3.2.4	Preliminary results and conclusions.....	10
3.2.5	Visualisation of the case study 2 - Model of air pollutants natural gas propagation in the city of Lisbon after a rupture in gas supply sanitation network ...	11
4	Annexes	11

1 Executive Summary

This document is the **Milestone 7 - Services HPC Configuration Report** of the Activity 2: Services analytics refinement and HPC preparation of the Action **Urban Co-creation Data Lab**, funded by the European Commission under the H2020ICT- 28-2017 Connecting Europe Facility (CEF) - Telecommunications Sector under the grant agreement n. INEA/CEF/ICT/A2018/1837945.

The main objective of the activity is to deploy the identified services in the co-creation labs and to connect them with the HPC infrastructure and with, the applications and the city services to be provided to users and stakeholders.

In order to achieve that main objective two tasks have been implemented under this activity, namely:

- Task 2.1: Preparation: data models and analytical services
- Task 2.2: HPC and services set-up and deployment

Due to COVID-19 pandemic crises, the other two cities approached to participate in the project, have not yet provide the necessary response regarding their participation in the project activities. Also, a delay occurred in the availability of Lisbon data, as the municipality is facing huge challenges in terms of allocation of resources in their daily operations due to the pandemic.

Despite these constrains we have been able to identify the case studies where is the need HPC resources, namely, #3 - Parking, #5 - Emergency and for case study #4 - Pollution a simulation as already been made.

A status of the project case studies is presented in annex (please see annexe n.º 2).

2 Objectives

2.1 Strategic Objective

The main objective of this activity was to deploy the identified services in the co-creation labs and to connect them with the HPC infrastructure and with, the applications and the city services to be provided to users and stakeholders.

2.2 Operational Objective

HPC and services set-up and deployment.

2.3 Task

In the last consortium meeting held on November 3 of 2019, the consortium agreed that to guarantee that the analytical models that are being developed in **Task 2.1 - Preparation data models and analytical services** should take into account the impacts of COVID-19 in the city, and should incorporate data for 2020. The data for 2020 that was possible to guarantee was relatively to the waste collection freights for case study #2 - Waste management; occurrences of irregular parking of the Municipality Police for case study #3 - Parking and traffic accidents for case study #5 - Emergency, that is being collected at this moment. In this sense there is a delay in the conclusion of the development of the analytical services, except for case study #1

MS 7 - Report on supercomputing resources

- Micromobility that have already been concluded. In short time the deployment process of the analytical services will start and the setting-up of necessary HPC platform, with connected Lisbon city open data, ready to conduct experiences and validation of use cases in Activity 3.

Regarding the use of HPC resources, a simulation has been already done for case study #4 - Predictive models for the propagation of liquid and atmospheric pollutants in the city of Lisbon”, more concretely for the propagation of natural gas in Alcântara (Lisbon).

2.4 Outputs:

- Data models and analytical services specifications (MS 6),
- Services HPC configuration report,
- Lisbon use cases implementation (on going);
- Validation of datasets per service with the European Data Portal (to be done in 2021).

2.5 Milestones and means of verification

<i>Milestone number</i>	<i>Milestone description</i>	<i>Indicative completion date</i>	<i>Completion date</i>	<i>Means of verification</i>
7	HPC resources have been used in order to process the data as required by the Action	30/11/2020	31/12/2020	Report supercomputing resources (systems/organisations providing the resources and computing hours) that have been used

3 Assessment and use of HPC resources

As already mentioned, and despite the constrains regarding the COVID-19 pandemic, we have been able to identify the case studies where HPC resources are needed, namely:

- #3 - Parking,
- #5 - Emergency, and
- #4 - Pollution.

The HPC resources needed for case studies #3 - Parking and #5 - Emergency are related with the need to find the best parameters when training an Autoregressive integrated moving average (ARIMA) model and computing Dynamic Time Warping (dtw) algorithm for measuring the similarity between two temporal sequences that will be used for clustering. This need for HPC resources for case study #3 - Parking is mainly due to the number of observations to be modelled as the data of occurrences regarding irregular parking is about 28 382 400, as the data was aggregated at street level with a time frequency of 3 hours for two years (2018 and 2019). In Figure 1 are presented the irregular parking occurrences recorded by the Municipal Police aggregated at street level for 2018 and 2019.



Figure 1. Irregular parking occurrences for 2018 and 2019 in Lisbon aggregated at street level.

A first test has been already done with this data using clusters derived from the Getis-Ord G_i^* , in a Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) model (i.e. without seasonal component), using as exogenous factor, a flag for business days. In Figure 2 are presented the results obtained for the train and test set, of a cluster that represents the hotspots occurrences of irregular parking in which was obtained a weighted mean absolute percentage error (WMAPE) of 0.86.

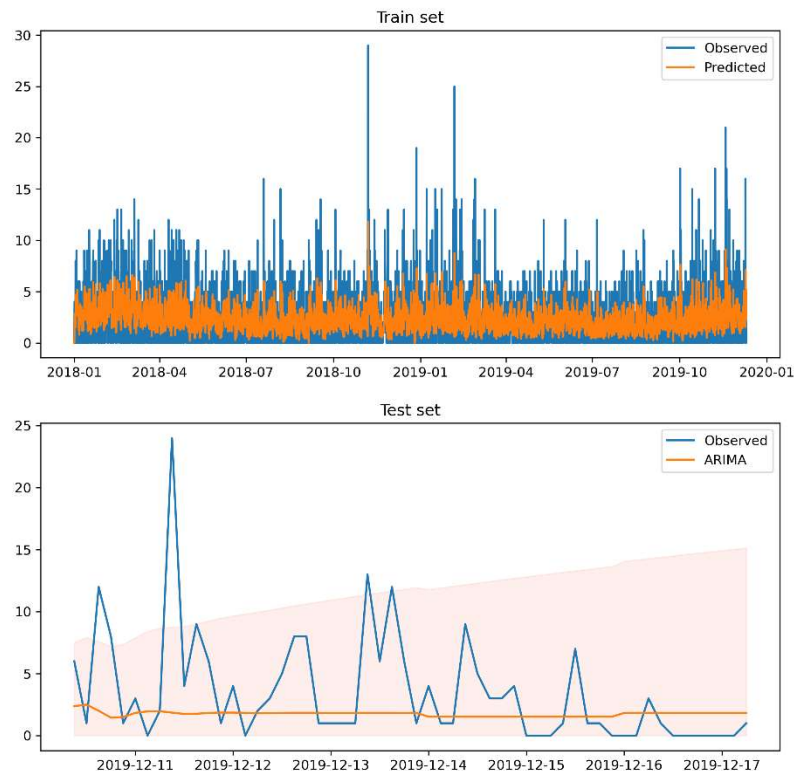


Figure 2. Observed and predicted occurrences of irregular parking for the train and test sets for 2018 and 2019 using an ARIMAX model.

As it can be stated the error is very high, probably due to non-inclusion of the seasonal component and more exogenous variables in the model. For instance we obtained very good results for the case study #1 - Micromobility in which was used a Seasonal Auto regressive Integrated Moving Average with Exogenous factors (SARIMAX) model, using as exogenous factors a flag for weekends, temperature and precipitation to predict the demand for a bike sharing

service, using data for a period of 9 months. With this model was obtained WMAPE of 0.15. In Figure 3 are represented the observed and predicted values for the train and test samples.

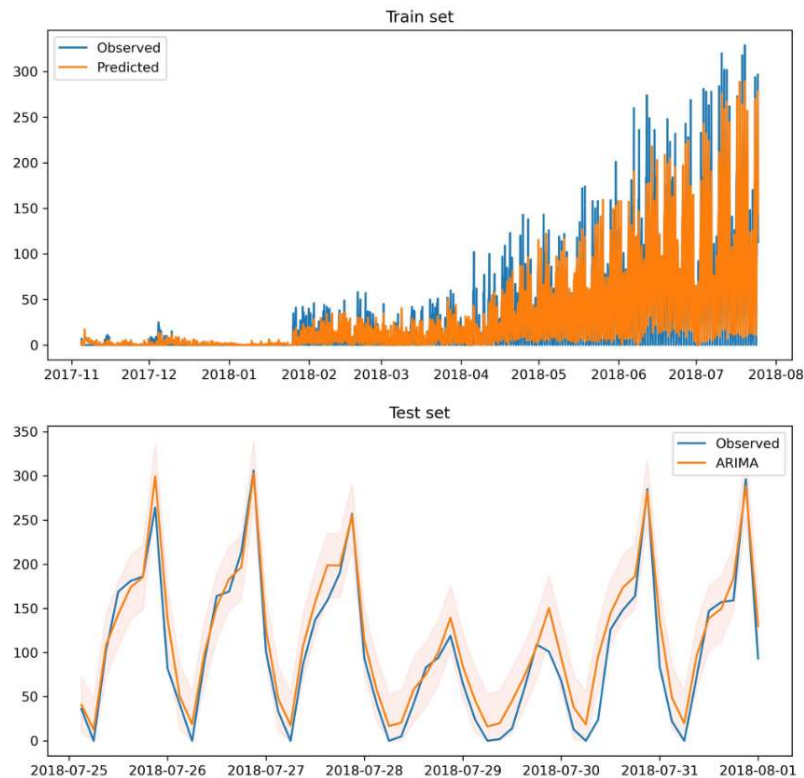


Figure 3. Observed and predicted demand of a bike sharing service for the train and test sets using a SARIMAX model.

As is our intention to include data for 2020 regarding irregular parking to include in the developed models the impacts of COVID-19, along with the integration of the seasonal component and more exogenous variables, is expected that will be necessary to resort to HPC resources.

Regarding the case study for #5-Emergency where is pretended to develop a model for prediction of traffic accidents, we received from the Lisbon Municipality data of traffic accidents recorded by Lisbon firefighters from 2019 and 2020 on the 15th December of 2019. At this moment we are aggregating all the historical data that we had access previously (i.e. from 2013 to 2018) and the data from 2019 and 2020 at street level with a time frequency of one hour. As we pretend to implement a similar modelling strategy as for case study #3 - Parking, we expect that there will be also the need for HPC resources for the development of this case study.

For case study #4 - Pollution a simulation has been already made, in which was simulated the propagation of natural gas due to a rupture in the gas supply network in the area of Alcântara (Lisboa).

The simulation for this case study was made resorting to MareNostrum 4 supercomputer that is managed by the Barcelona Supercomputing Centre (BSC). MareNostrum is managed by the BSC Operations Department, which takes care of its availability, security and performance. An important task of this team is to support scientists in the usage of MareNostrum, as well as to help them to improve their applications to get better research results. They are also in charge of carrying out the updates of the machine.

3.1 Supercomputing Resources

MareNostrum 4 has a disk storage capacity of 14 Petabytes and is connected to the Big Data infrastructures of BSC-CNS, which have a total capacity of 24.6 Petabytes. A high-speed Omnipath network connects all the components in the supercomputer to one another. Like its predecessors, MareNostrum 4 is also connected to the European research centres and universities through the Rediris and Geant networks. is a supercomputer based on Intel Xeon Platinum chips, each with 24 processors, amounting to a total of 165.888 processors and a main memory of 390 Terabytes, Lenovo SD530 Compute Racks, a Linux Operating System and an Intel Omni-Path interconnection, with the following characteristics:

- Peak Performance of 11,15 Petaflops
- 384,75 TB of main memory
- 3.456 nodes:
 - 2x Intel Xeon Platinum 8160 24C at 2,1 GHz
 - 216 nodes with 12x32 GB DDR4-2667 DIMMS (8GB/core)
 - 3240 nodes with 12x8 GB DDR4-2667 DIMMS (2GB/core)
- Interconnection networks:
 - 100Gb Intel Omni-Path Full-Fat Tree
 - 10Gb Ethernet
- Operating System: SUSE Linux Enterprise Server 12 SP2

MareNostrum 4 is able to perform more than eleven thousand trillion operations per second, with an energy consumption of 1.3 MWatt-year.

MareNostrum 4 is formed of clusters of three different emerging technologies that will be added and updated as they become available. These are technologies currently being developed in the US and Japan to accelerate the arrival of the new generation of pre-exascale supercomputers.

- A cluster consists of IBM POWER9 processors and NVIDIA Volta GPUs, which are the same components that IBM and NVIDIA will use for the Summit and Sierra supercomputers that the US Department of Energy has commissioned for the Oak Ridge and Lawrence Livermore National Laboratories. Computing power over 1.5 Petaflop/s;
- A cluster made up of AMD Rome processors and AMD Radeon Instinct MI50. The machine will have a similar processor and accelerator than the Frontier supercomputer that will be installed on 2021 at ORNL. The computing power of the machine will be 0,52 Petaflop/s;
- A cluster formed of 64 bit ARMv8 processors in a prototype machine, using state-of-the-art technologies from the Japanese Post-K supercomputer. Computing power over 0,65 Petaflop/s.

Further information regarding MareNostrum 4 is available at <https://www.bsc.es/marenostrum/marenostrum/technical-information>.

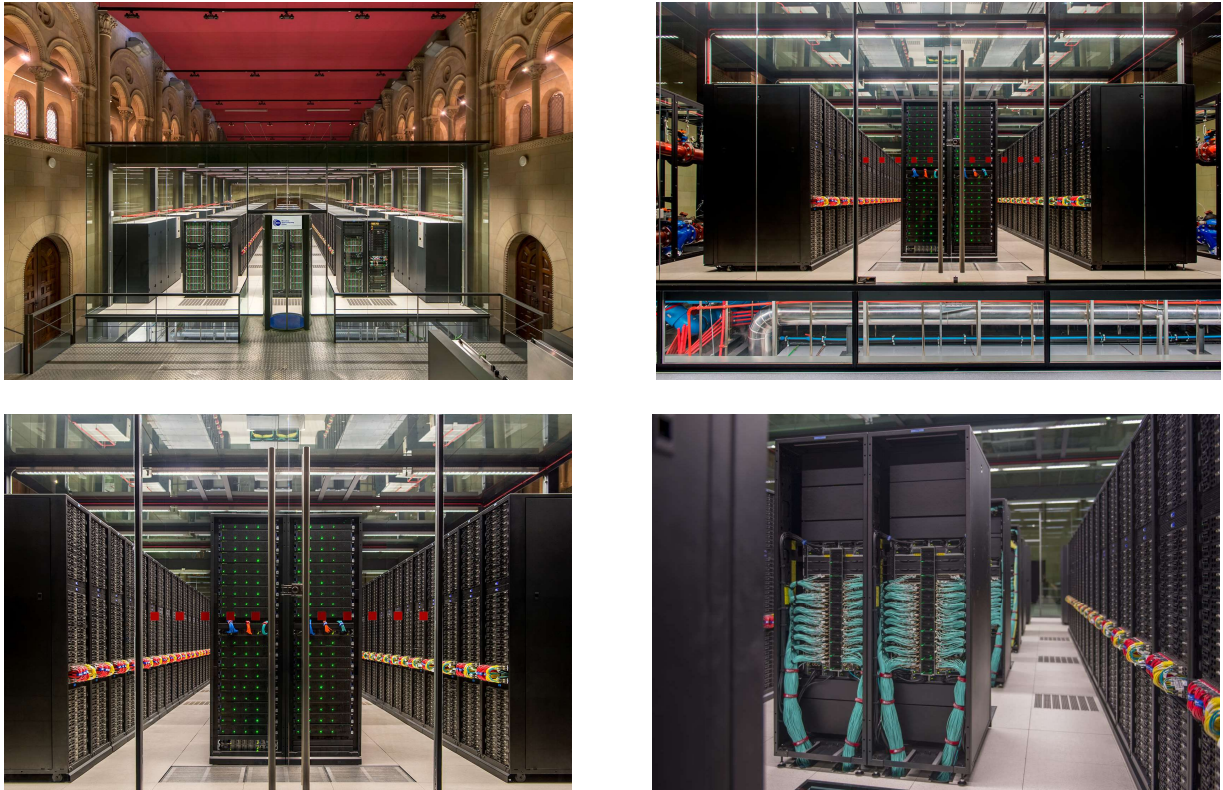


Figure 4. MareNostrum 4.

3.2 Case Study

A simulation, using HPC resources for case study #4 - Pollution was made, more concretely a simulation of the propagation of natural gas in Lisbon (Alcântara) in an area of about 1 km², with the following parameters:

- wind direction: East (1 simulation)
- computational grid: 40 million of elements
- parallel partition: 1.008 cores
- run time: 72h/simulation
- total computational cost: 72.576 cpu-hours/simulation
- simulated real time: 2,31h

Since the computational domain (i.e. the study area) will be expanded for the final simulations (in which will be studied the influence of the surrounding buildings in the simulation), it is expected that the above mentioned numbers will increase significantly.

3.2.1 Objective

The objective is the numerical simulation of a series of test cases proposed by Lisboa City Council.

3.2.2 Tasks

The following tasks have been developed:

1. Test cases definition;
2. Meteorological conditions analysis;

3. Geometry and mesh generation;
4. Running simulations;
5. Conclusions.

3.2.3 Case studies

Within the case study #4 - Pollution, two case studies are being developed, namely:

Case study 1: Model of liquid pollutants in the city of Lisbon after a train accident with hazard liquids.

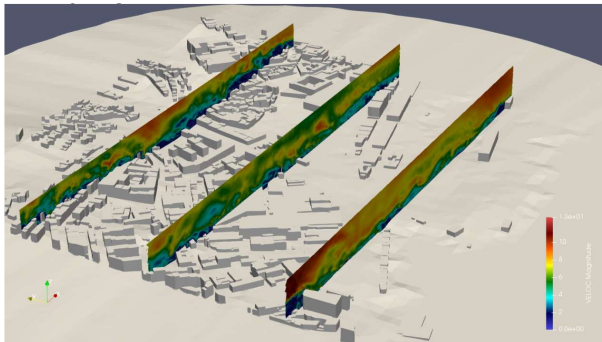
In this case the study is pretended to model what is the propagation of nitrobenzene after a train accident in Roma-Areeiro station with a temporal frequency of 1 hour, in an area of 1 km² from the location of the incident.

Case study 2: Model of natural gas propagation in the city of Lisbon after a rupture in gas supply network

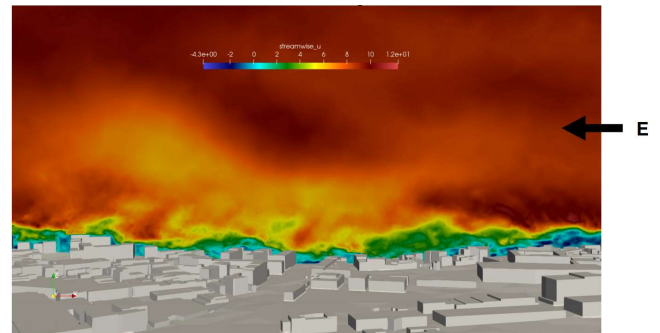
In this case study is pretended to model a leakage of natural gas in Alcantara (with residential, tourist, harbour area and an hospital), with a temporal frequency of one hour, in an area of 1 km² calculated from the location of the incident.

At this moment a simulation has been made for case study 2, while for case study 1 the simulations work already started. Attached to this report please find the complete presentation of the case study regarding the simulation for the model of natural gas propagation. (please see annexe n. 1)

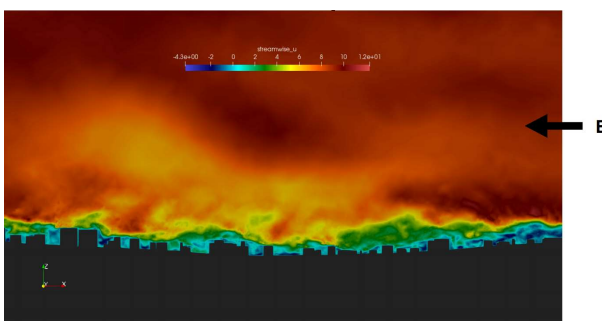
3.2.4 Preliminary results and conclusions



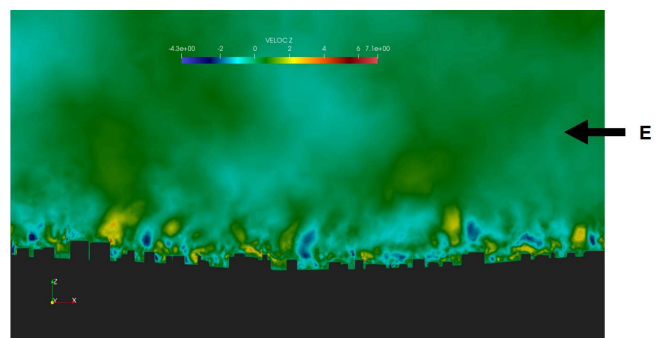
Velocity magnitude sections



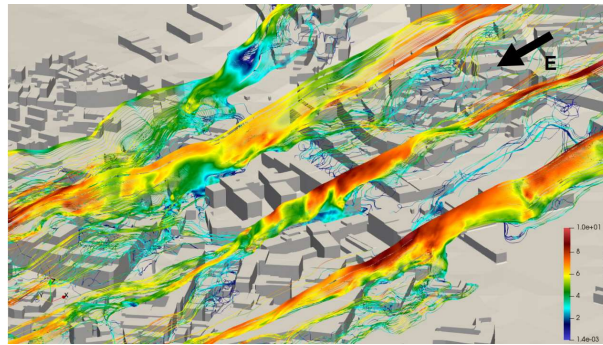
Flow at street level: streamwise velocity



Flow at street level: streamwise velocity



Flow at street level: vertical velocity



Flow streamlines

Figure 5. Case study 2: Model of natural gas propagation in the city of Lisbon after a rupture in gas supply network

Following the preliminary results some conclusion can be highlighted, namely

1. The feasibility of the two proposed test cases has been analysed.
2. The provided meteorological data has been appropriately processed to provide insight about the adequate boundary conditions for the simulations.
3. The provided geometry has been successfully repaired and processed. Subsequently, a high-resolution mesh could be generated.
4. An initial simulation has been performed for the predominant wind direction. Preliminary results have been obtained showing the characteristic flow patterns in urban areas.
5. The following step will be the introduction of the passive scalar transport equation in the model.

3.2.5 Visualisation of the case study 2 - Model of air pollutants natural gas propagation in the city of Lisbon after a rupture in gas supply sanitation network

The videos of the simulation are available in:

- <https://drive.google.com/file/d/1qsbINOxWsybkifKbOU1GAoYoNtnQ7ZlZ/view?usp=sharing> ;
- <https://drive.google.com/file/d/1tY1cVabR0pzclYCIHnLomyLIB-uOL8K3/view?usp=sharing> ;
- https://drive.google.com/file/d/1_XKneVDEvtyJ-QTJHfj5Acd1QV7tXwnX/view?usp=sharing .

4 Annexes

Annexe 1 - BSC Presentation at the project M12

Annexe 2 - NOVA IMS Presentation at the project M12