Pilot Case Study Report and Recommendations

Utilising the BalticLSC Environment for specific calculations based on real business needs

Version 1.0

Priority 1: Innovation

Warsaw University of Technology, Poland
RISE Research Institutes of Sweden AB, Sweden
Institute of Mathematics and Computer Science, University of Latvia, Latvia
EurA AG, Germany
Municipality of Vejle, Denmark
Lithuanian Innovation Center, Lithuania
Machine Technology Center Turku Ltd., Finland
Tartu Science Park Foundation, Estonia

Pilot Case Study Report and Recommendations

Utilising the BalticLSC Environment for specific calculations based on real business needs

Work package	WP3
Task id	3.5
Document number	O3.5
Document type	Report
Title	Pilot Case Study Report and Recommendations
Subtitle	Utilising the BalticLSC Environment for specific calculations based on real business needs
Author(s)	Henri Hanson
Reviewer(s)	Vitalija Kolisova, Krzysztof Marek
Accepting	Michał Śmiałek
Version	1.0
Status	Final version

Interreg Baltic Sea Region
EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND

History of changes

Date	Ver.	Author(s)	Change description
26.10.2021	0.1	Henri Hanson (TSP)	Document creation
15.11.2021	0.2	Henri Hanson (TSP)	Document creation
06.12.2021	0.3	Krzysztof Marek (WUT)	Review
21.12.2021	0.4	Vitalija Kolisova (LIC)	Review
22.12.2021	1.0	Henri Hanson (TSP)	Final version

Executive summary

Baltic LSC project's activity 3.5 was focusing on pilot case studies utilising the Baltic LSC Environment for specific calculations based on real business needs. Current report provides an overview of the activities regarding approaching the potential partners for pilot cases, their technical needs and feedback on the Baltic LSC Environment and recommendations for the future.

All partners were involved in the activity, issuing public calls for participants, organizing promotional events for case studies, defining and executing case study acceptance procedures and maintaining a case study list. Also, collecting case study results and feedback from participants. Technical partners of the project focused on actively supporting case study participants in defining and executing their computation applications.

Baltic LSC partners gathered basic information on these mini-projects, incl. their domains, calculation needs, calculation application definitions etc.

The technical partners added description how the calculation applications were developed and executed. Together with these descriptions, certain quantitative data will be presented, where special importance will be given to effectiveness (esp. time) of calculation definition and execution. Analysis on how this effectiveness can be compared to the previous approaches to LSC. This should give specific "hard" data for communication activities to promote the system around the Baltic Sea region.

Table of contents

List of figures

List of tables

1. Introduction

1.1 Objectives

Baltic LSC project's activity 3.5 was focusing on pilot case studies utilising the Baltic LSC Platform for specific calculations based on real business needs. Current report provides an overview of the activities regarding approaching the potential partners for pilot cases, their technical needs and feedback on the Baltic LSC Environment and recommendations for the future.

Baltic LSC partners gathered basic information on these mini-projects, incl. their domains, calculation needs, calculation application definitions etc.

The technical partners added description how the calculation applications were developed and executed. Together with these descriptions, certain quantitative data will be presented, where special importance will be given to effectiveness (esp. time) of calculation definition and execution. Analysis on how this effectiveness can be compared to the previous approaches to LSC should give specific "hard" data for communication activities to promote the system around the Baltic Sea region.

1.2 Scope

Overall, more than 100 companies have been approached, including participants at project events, individual meetings with potential service providers/end-users, targeted online marketing campaigns etc. From these, more than 30 companies have had in-depth discussions on using the Baltic LSC services and 8 successfully run cases, where there is a chance for continuous future services.

1.3 Relations to Other Documents

The document is developed based on several Baltic LSC outputs, mainly the Output 2.3: Handbook on Doing Business Using the BalticLSC System

2. Case studies

The BalticLSC environment consists of an online platform developed by the project technical partners which is designed to affect two major groups across the Baltic Sea region, namely:

1. Large **LSC service providers** who can offer their computational power to perform calculations by joining the network. Even start-ups and SMEs can take up BalticLSC specifications and emerge as LSC service providers by creating their own local computing centers. Furthermore, several of such centers can combine to form a network and increase their computing capability which can be distributed across Baltic Sea region.
2. **End-users** such as SMEs, start-ups, research and development centers, design centers and others can utilize the platform to develop and test their own applications and carry out complex computations and calculations thus reducing the time to market in the delivery of innovative products and services. With the help of this environment, the end-users can now make use of LSC services easily which were previously difficult to access.

All partners were involved in the activity of promoting the Baltic LSC environment, issuing public calls, approaching individually potential participants and organizing workshops as well as promotional events for finding case studies, defining and executing case study acceptance procedures and maintaining a case study list. Also, collecting case study results and feedback from participants. Technical partners of the project focused on actively supporting case study participants in defining and executing their computation applications. The project strives to develop a solution that is replicable, transferable, sustainable, and easy to implement.

Pilot Case Study template, available as Appendix A of the current report, was used to ask for feedback from the case study participants.

2.1 Case study participants

Overall, more than 100 companies have been approached, including participants at project events, individual meetings with potential service providers/end-users, targeted online marketing campaigns etc. In the tables below there are descriptions of 30 examples of companies/case studies, that have had in-depth discussions on using the Baltic LSC services and 8 successfully run cases, where there is a chance for continuous future services.

As a lot of companies were not eager to share their names publicly, thus these are not shown. The information about their sector/activity is mentioned.

2.1.1 Successful cases

Cases in which pilots used the BalticLSC Environment and shared their experience and feedback.

Table 1 BalticLSC successful pilot cases

	Sector/Industry description	Status	Reason for cooperation or lack of cooperation
1	University HPC center	Selected	Estonian BalticLSC node run by HPC at University of Tartu, Estonia Case: platform/LSC service provider, with potential for increase availability of resources as needed.
2	Civil Engineering (Pot-hole detection on photos)	Selected	Training of a NN responsible for detecting if the photo contains a pot-hole and how serious it is
3	Tools with AI (E-nose)	Selected	Neural network training. Multiple NN training in parallel (no communication during computation -> ideal case)
4	Education (Curious Intellect)	Selected	NLP-based virtual learning assistant
5	Smart routing for waste collection	Selected	Combinatorial optimization. The task is to build a set of reusable modules that allows building a CAL app that optimizes the sequence of customer visits for the given vehicle fleet and waste fields.
6	University research projects	Selected	Multiple student research project run on BalticLSC. The core example is using BalticLSC computation power to detect moods of people on provided photos. This required detecting faces (module provided by the BalticLSC experts) and an implementation of mood detection on detected faces.
7	Welding	Selected	Welding quality assessment based on ultrasound or X-ray pictures. Case where an expert in AI uses the BalticLSC to quickly develop a neural network dedicated to solving a specific problem. The BalticLSC allows for much quicker training of the network and the trained model can be later used in real time at the manufacturing facility.
8	Old black-and-white photo colouring	Selected	Using a computation intensive AI algorithm to colour black-and-white photos from historical archives of a small local town in Poland (Tarczyn)

2.1.2 Pending cooperation cases

Cases in which the use of BalticLSC Environment was widely discussed but the pilots decided to delay the use of the BalticLSC Environment. Provided reasons included: lack of human resources, change of company priorities (e.g. due to Covid-19 pandemic), currently sufficient own computation resources (but open to using BalticLSC if the need increases), and more.

Such cases allowed to gather different type of useful feedback.

Table 2 BalticLSC pending pilot cases

	Sector/Industry description	Status	Reason for cooperation or lack of cooperation
1	Water/app for shower, WC etc.	Pending	They have been busy inventing a new solution for the Corona situation and couldn't put time a side for our project, but they are interested
2	University	Pending	They are interested in joining with some student projects; however, our contact person has been on paternity leave
3	Software for services	Pending	Good case: download data from many servers -> process data -> upload result to AWS storage. The pilot has resigned due to lack of own resources and a change in priorities.
4	Energy	Pending	Right now they don't see the need to have more computation power than their own; however, they follow the project
5	AI/IoT	Pending	A group of companies working within the sector of AI/IoT - they all resides in the same network/business house and have been informed about the project in a seminar; however, now it is time to follow up on them individually. Corona has made that effort more difficult, but around ten companies relevant to contact and follow-up.
6	Green Technology	Pending	A group of companies working within the green tech sector - they all resides in the same network/business house and have been informed about the project in two seperate network meetings.
7	Food/VR/AR/PR	Pending	Needing an AI expert to help them in order to be able to use the Baltic LSC - They have a meeting with University of Aarhus (DK) Friday 8.1.2021
8	Fintech	Pending	Looking into our possibilities, do not have their own computation power, sensitive data
9	Food/VR/AR/PR	Pending	Needing an AI expert to help them in order to be able to use the Baltic LSC - They have a meeting with University of Aarhus (DK) Friday 8.1.2021
10	Fintech	Pending	Looking into our possibilities, do not have their own computation power, sensitive data
11	Space industry	Pending	Have some data, not too much at the moment computing on their own, maybe in the future
12	AI/IoT	Pending	Working with AI, a lot of data, now no need for more computation power, managing on their own or using google cloud
13	AI/IoT	Pending	working with medical images, testing products on their own computation power, do not see the need to start new project now
14	Laser industry	Pending	Creating different materials small products for clients, running simulations, interested in project maybe in the future
15	Energy	Pending	working with solar and other renewable energy, complex approach to energy supply, interested in the project in the future, now they see no need, there is not so much data
16	IT	Pending	Different deep learning solutions as a service
17	Wildlife/mobile apps	Pending	Start-up company, that sees a great potential in free help and the network opportunities in project countries. Good case: neural network training on large sets of data (image processing)

2.1.3 Cancelled/not suitable cooperation cases

Cases in which the use of BalticLSC Environment was discussed with the organizations but the cooperation wasn't possible at the moment due to the BalticLSC Environment not fitting the organizations' requirements (in many cases the organizations didn't need Large Scale Computing).

Table 3 BalticLSC not selected pilot cases

	Sector/Industry description	Status	Reason for cooperation or lack of cooperation
1	Environment and Health	Not selected	Baltic LSC do not have the storage facilities needed to include this company in our project right now, maybe later
2	Energy	Not relevant	They don't need more computation power for their company and don't have any surplus computation power either
3	Energy/software	Not interested	They have great AI/IoT competences and would rather use their own team as well as own hardware and software than the Baltic LSC.
4	Robotics	Not relevant	They have some data, managing on their own, not ready for LSC
5	Geo data	Not relevant	interested in computation, especially in cheaper way of doing it, but it is too expensive to move the data, case closed
6	Robotics	Not relevant	working with client data, need the information real time, everything calculated in the factory at the client
7	Medicine	Not relevant	creating the robot for surgeries, have some data, no need for LSC
8	Manufacturing	Not relevant	modelling sophisticated products, testing it in simulations, no need for computation power, managing on their own
9	Design/IT	Not relevant	Interactive 3D solutions for E-commerce and other clients, not much computation, working on their own recourses
10	University	Not relevant	Computational simulations of minimum material usage for maximal strength
11	IT	Not interested	Platform risk - the company is not willing to use the platform, because of sensitive business data and algorithms.
12	Engineering	Not selected	Incompatibility issues - the CAD/CAM calculations needed by company cannot be easily transferred to Baltic LSC platform.
13	Geo data	Not selected	Earth Observation data is really large and not easily transferrable to Baltic LSC environment
14	Geo data	Not selected	Platform risk - the company is not willing to use the platform because of the maturity and missing possibility to use high precision object detection from earth observation imagery

2.2 BalticLSC Pilot Case examples in CAL

Not all CAL (Computation Application Language) applications could be shared, because of company-specific trade secrets, but below there are two examples of applications developed by the pilot cases.

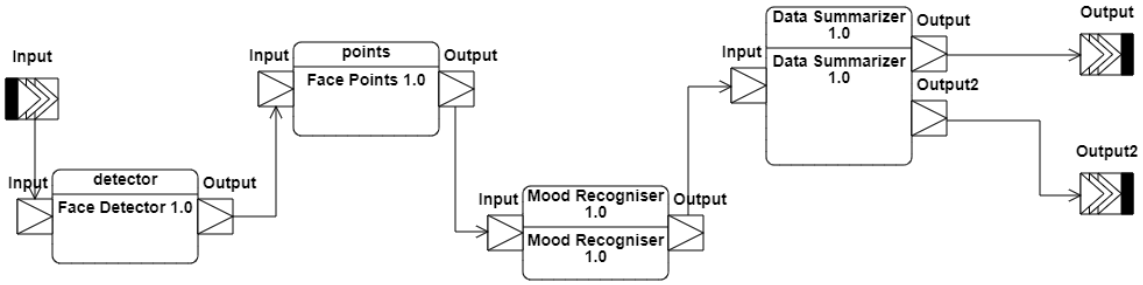

Figure 1 Mood recognizer

Figure 1 shows an example of a simple CAL application developed during a pilot case. It uses a Computation Module already provided by the BalticLSC Developers (Face Detection) and adds additional Computation Modules detecting the mood of people whose faces have been detected on the pictures.

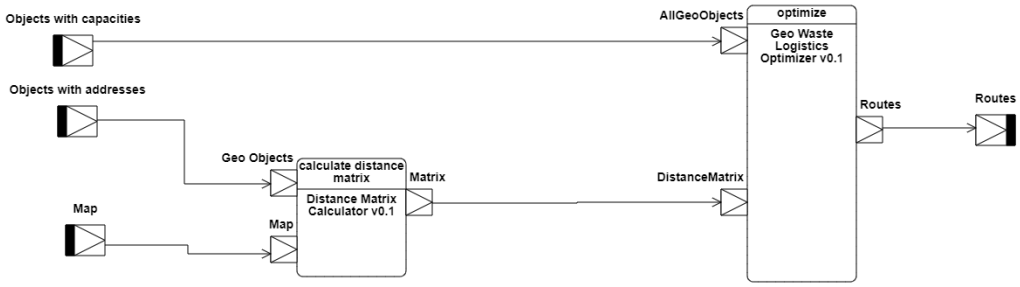

Figure 2 Capacitated Vehicle Routing Problem (CVRP) Solver

In Figure 2 a more advanced Computation Application has been shown. It also uses some already available Computation Modules but requires multiple new Modules and an entire separate Computation Application shown in Figure 3 used by the main application as a Computation Module (this is a test of one of the most advanced features of CAL, using other Computation Applications as Computation Modules)

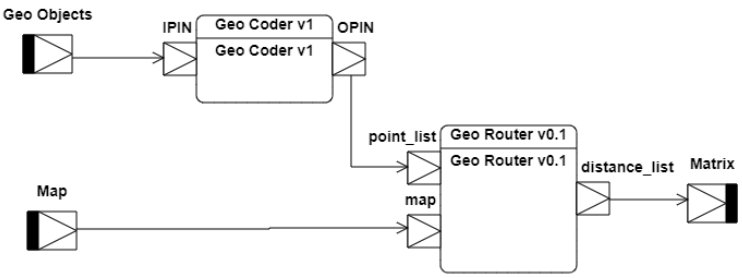

Figure 3 Distance Matrix Calculator (used by the CVRP Solver)

3. Issues and recommendations

Working with the companies providing case studies, a number of issues were mapped, including:

1. **Data quality** - a lot of data collected within business environments is usually with bad quality. To overcome this key issue, one must focus on 3 Cs of data collection: Complete, Clean, Consistent. For business applications, data must be available quickly, processed immediately and automatically, also must be connected to business processes. Thus, there is a need for data specialists to work within the company to produce this kind of data, that might be easier for future utilisation within Baltic LSC or wider HPC context.

2. **Issues with big datasets:**

- Cluster - Filesystem relations are complicated.
- Difficult to understand how much resources a job uses - needs profiling! (Log into node and check process resource consumption with standard linux tools like top, iotop, htop, ps, etc. Use profiling tools like valgrind. With interpreted languages, write parts of code to handle profiling, like psutil in python. Learn with small datasets and generalize to bigger ones. Understand multiprocessing for CPU usage decisions.)
- Not difficult to make a mess. Avoid heavy IO workloads on a cluster/network congestion (e.g. packing things, huge data files). Don't expect resources to run your job instantaneously.
- Parallel, concurrent and multithreaded jobs.
- Module system lacks standards.
- Some specific datasets, e.g. Earth Observation datasets are really large and not easily transferrable to Baltic LSC environment as the platform is not providing data storage service.

3. **Integration with other programs/datasets**, e.g. lack of integration with COMSOL Multiphysic, used as cross-platform finite element analysis, solver and multiphysics simulation software. This might be possible in future versions.

4. **Platform risk for businesses** – from the business perspective, working with a totally new platform run by public service providers, there are huge risks involved, incl. sharing sensitive business data and algorithms.

5. **Data and programming experts needed** to help companies set up their calculations at the Baltic LSC platform. During the project the experts were freely available, however this might be an issue in the future, after the end of the project to help the companies. For example, from one of the case studies - although BalticLSC provides module development templates in C# and

Python, specific modules were developed from scratch using Java and Spring Boot. It took approximately one person-month to develop three working modules. There was adaptation of already existing software (Optimizer, Router) as well as building completely new software (Coder). This might be actually solvable and support university-industry cooperation, if the companies would be willing to pay for such developments and the university would be flexible enough to find specific programmers for each similar case.

4. Discussion

Baltic LSC project's activity 3.5 was focusing on pilot case studies utilising the Baltic LSC Platform for specific calculations based on real business needs including activities to find potential partners for the pilot cases, work together with the case studies regarding their technical needs – available data and needed output parameters.

As a new type of activity for most partners involved, the project was overall a success, opening doors for future cooperation in the LSC/Big Data field. Baltic LSC partners gathered basic information on these mini-projects, incl. their domains, calculation needs, calculation application definitions etc. The technical partners added description how the calculation applications were developed and executed. Analysis on how this effectiveness can be compared to the previous approaches to LSC should give specific "hard" data for communication activities to promote the system around the Baltic Sea region in the future.

All partners were involved in the activity, issuing public calls for participants, organizing promotional events for case studies, defining and executing case study acceptance procedures and maintaining a case study list. Also, collecting case study results and feedback from participants. Technical partners of the project focused on actively supporting case study participants in defining and executing their computation applications.

Overall, more than 100 companies for potential use cases have been approached during the duration of the project, including participants at project events, individual meetings with potential service providers/end-users, targeted online marketing campaigns etc. From these, more than 30 companies have had in-depth discussions on using the Baltic LSC services and 8 successfully run cases, where there is a chance for continuous future services. The small turnaround from companies approached to successfully run cases stems from several factors, including from one side the low level of knowledge in the companies (e.g. no data scientists available or the low quality of data available), but also on the platform risk, as Baltic LSC is run as a cross-border service by publicly funded partners. Additionally, the Baltic LSC platform itself is not yet flexible enough for some specific calculations, needing further development. As a result, although the business plan was developed, additional investments into not only building the platform itself, but also train the wider client base in the possibilities and tools available on the market for the gain of their business. Building Baltic LSC service as a long-term sustainable business in the future will continuously be the challenge in this field.

Appendix A - Pilot Case Study template

Case study: general description of use case, e.g. detection of persons in a video

Company: Company X, Country

User status: end-user, supplier (resource provider), developer. (pick one)

Case study prepared by: partner name

1. Description of the case study and company

The company should be described within 1-2 paragraphs, including country, main business sector(s), size, etc. Ask the company, if ok with public report with company name or prepare anonymous report with general description.

2. Description of use case

General description: Which domain is the use case from, e.g. weather forecasting, chemistry, biotech, space, etc.? What are the main challenges for the company regarding (big) data/LSC/HPC?

Detailed description: What the company is currently using for addressing the issue, what are the main expectations for Baltic LSC platform? How the BalticLSC platform is performing for the company?

Include visuals: pictures, components and processes involved, etc.

3. Technical parameters/details of the use case

What kind of data is available? Amounts and time used of the BalticLSC system, incl. CPU/GPU resource, memory/storage usage, support by technical partners setting up the case study, etc. Potential budget for similar calculations in the future.

Include visuals: e.g. descriptions and diagrams of the calculations

4. Discussion with focus on key issues and potential solutions

Free form input regarding key findings, suggestions for improvements, discussion on viability of this specific kind of service to other similar clients, etc.