# Integrating supercomputing clusters into education: a case study in biotechnology

Álvaro Fernández[1,2] · Camino Fernández[3] · José-Ángel Miguel-Dávila[1] ·
Miguel Á. Conde[3]

## Abstract

The integration of a Supercomputer in the educational process improves student's technological skills. The aim of the paper is to study the interaction between science, technology, engineering, and mathematics (STEM) and non-STEM subjects for developing a course of study related to Supercomputing training. We propose a flowchart of the process to improve the performance of students attending courses related to Supercomputing. As a final result, this study highlights the analysis of the information obtained by the use of HPC infrastructures in courses implemented in higher education through a questionnaire that provides useful information about their attitudes, beliefs and evaluations. The results help us to understand how the collaboration between institutions enhances outcomes in the education context. The conclusion provides a description of the resources needed for the improvement of Supercomputing Education (SE), proposing future research directions.

**Keywords** Supercomputing training · Education · Biotechnology

✉ Álvaro Fernández
aferg@unileon.es

Camino Fernández
camino.fernandez@unileon.es

José-Ángel Miguel-Dávila
jam.davila@unileon.es

Miguel Á. Conde
miguel.conde@unileon.es

[1] Business Department, Universidad de León, León, Spain

[2] Fundación Centro de Supercomputación de Castilla y León (SCAYLE), Leon, Spain

[3] Mechanical, Computer and Aerospace Engineering Department, Universidad de Leon, Leon, Spain

# 1 Introduction

High-performance technologies are an increasingly prominent theme in higher education. The fast introduction of Supercomputing in education is a challenge that provides a broader range of opportunities. In this sense, the possibility of managing a high volume of data and information has prompted educators and researchers to take a pedagogical view to promote the use of these infrastructures in teaching. Computing Education (CE), as a popular research field, faces new challenges that require students' motivation, which is the foundation in systems as the SAIL—a System for Adaptive Interest-based Learning [3]. New tasks and challenges arise for the Higher Education System (HIES) which require extensive and complete training. This training has to face the limitations of the integration of technologies in teaching [44, 47] and at the same time be included in the standard curriculum as a constraint related with Supercomputing [26]. Currently, the use of Supercomputers is becoming a key element in the improvement of certain disciplines of higher education, in spite of the lack of experience and basic skills in this field [1] and low availability of Supercomputing facilities due to the high costs [19].

Supercomputing Education (SE), as one of the sub-fields of CE, is considered important for innovation, modernization and technological development, contributing to create a new generation of professionals, companies and organizations related to Science and Technology [11]. This area is also in charge of specific training, mainly for students of computational science and engineering, and also for students of various fields of knowledge, that use computational, mathematical and engineering tools for problem solving [45]. Currently, many educational institutions worldwide are seriously introducing Computer Science (CS) concepts into the basic training of students [2], which raises the question of how important the training in Supercomputing for the professional future of the students is.

The introduction of new information and communication technologies (ICT) in the training sessions of students of all levels has been considered one of the main tools in the modernization of all educational systems. It has contributed to evolution from the "society of the information" to the "society of knowledge" [33], allowing forming students, in an appropriate way, for the exercise of a great variety of professions. This consideration means that CS could be suggested for the core area of many postgraduate disciplines, where the use of a Supercomputer will be essential for solving complex data problems. The fast developments in computer and electronics technologies provide father-reaching developments especially in the areas of biomedical and health informatics, which are very important for improving solutions [25] for the needs of training sessions of these fields.

For companies and organizations in general, technological qualification is ensuring corporate competitiveness by improving productivity and reducing the consumption of human/physical resources through the change of strategic management environments using ICT technology [31]. In this, supercomputing training represents a specific learning activity. It provides an integrated and experimental approach in the

use of technologies for higher education students in advanced contents. These contents are designed to provide experiences in the analysis of real cases to be used as pedagogical resources. At the same time, it increases learner engagement, making students feel involved in the course content, improving the level of technical expertise and expectations regarding technology. The level of training provided in these courses covers a wide range of areas that depend on the target category of the trainees. They are supported by specialists in many fields of knowledge that help students in solving complex computation-intensive applied problems. SE is based on a wide variety of courses and tools, under different forms of educational programs that permit the development of these skills. In general, courses linked to Supercomputing differ from those of advanced training for IT specialists, and can be classified into three groups [24]:

- System architecture: the goal is to teach how to design, implement and deploy a supercomputing platform.
- Applications in Computing Science/Computing Education: based on the implementation of high-performance programs on a supercomputing platform.
- Specific-domain courses: uses the computing power of a Supercomputer to solve specific problems of a field, such as computational physics [32], computational mathematics [27], and computational biology [8].

Regarding higher education students, the advantages of Supercomputing training are the following:

- Student explores real problems through hands-on experience [7, 10, 20, 21, 26, 41, 43], learning the most advanced functionalities of computers, linking the theory with a practical problem solving.
- It is not always necessary, previously to the course, to have skills in the use of a Supercomputer [38].
- Supercomputing training prepares students in a continuous process to get a level of knowledge and proficiency adequate to deal with the largest computing systems available [7].
- Supercomputing training helps in the use of mathematical models [9, 42, 50] which allows describing a problem in a specific research field, for extracting conclusions and evaluating performance on the proposed problems.
- Simulations and data design in a Supercomputer prepare students for the role of leaders in research [50] by the use of very large database systems [1], covering, at the same time, a great variety of applications [39].
- Supercomputing contributes to inventing new tools and new philosophies of work, which enhance the learning about computing [20, 29].

Bloom's Taxonomy of Educational Objectives [12] represents a convenient way to describe the degree to which students understand and use concepts, demonstrate particular skills, and have their values, attitudes, and interests affected. According to this taxonomy, supercomputing training enhances learner engagement by providing experiences that allow students to practice application skills.

This article presents first an analysis of Supercomputing in education through a review of relevant literature and secondly, a case study that illustrates the benefits and challenges of organizing a course of high technology for higher education students. The main goal is measuring the quality of training sessions through the use of HPC tools by analyzing how the integration of technology, in the teaching and learning process, engages students, providing motivation, attitudes, achievement and interactions [48]. In order to achieve this goal and based on both the literature review and the experience of the collaborating entities of this study, a survey questionnaire was developed to evaluate the opinion of the students in relation to the training received in both courses. The feedback obtained was used as a base both to validate the flowchart proposed and to improve future sessions and, at the same time, to look for better performance [17, 18].

The paper's main objective is the description of a flowchart of the educational process related to Supercomputing. The secondary objectives are: (1) to study the main application fields identified in the review of the literature for both STEM and non-STEM or non-computing students in High Education Institutions (HEI) where Supercomputers are used; (2) to study the types of computational courses, related to Supercomputing, identified through the review of the literature; (3) to provide a relational study juxtapositioning the field or sector that uses Supercomputing and the type of specific computational training and (4) to analyze how the use of a Supercomputer in a practical course, organized jointly a University and a Supercomputing Center, permits the enhancement of training in higher education environments.

The conclusions highlight the strengths of the experiential approach taken through the case study. It is argued that technology, in general, is a powerful way to engage students using innovative practices, thereby improving their performance.

The outline of the paper is as follows: in Sect. 2, we briefly discuss the material and methodology followed; Sect. 3 presents, in greater detail, a general overview of the results obtained, with data analysis related to the use of Supercomputers in education. In Sect. 4, a case study of training is described. An overall discussion of the results is outlined in Sect. 5. Finally, in Sect. 6, the conclusions are presented.

## 2 Materials and methodology

The scope of the study, both with regard to its technical and educational content, is suited for a combination of methods. In this sense, the design of the targeted process combines theory with the findings of the literature review and, at the same time, is heavily based on the case study method.

Firstly, the base of the population studied was the one used in [22], which follows guides related to literature reviews [23], with special attention to systematic reviews of the literature and based on a taxonomy about bibliographic reviews in education [15]. It is considered as true hypothesis-based research, in which the studies are selected and combined through the use of a predefined protocol that reduces subjectivity and the possibility of bias of the researcher [6, 30]. The items of the study were filtered based on a review of relevant journals, conferences and books, focusing on the analysis of terms related to training, Supercomputing and technological subjects
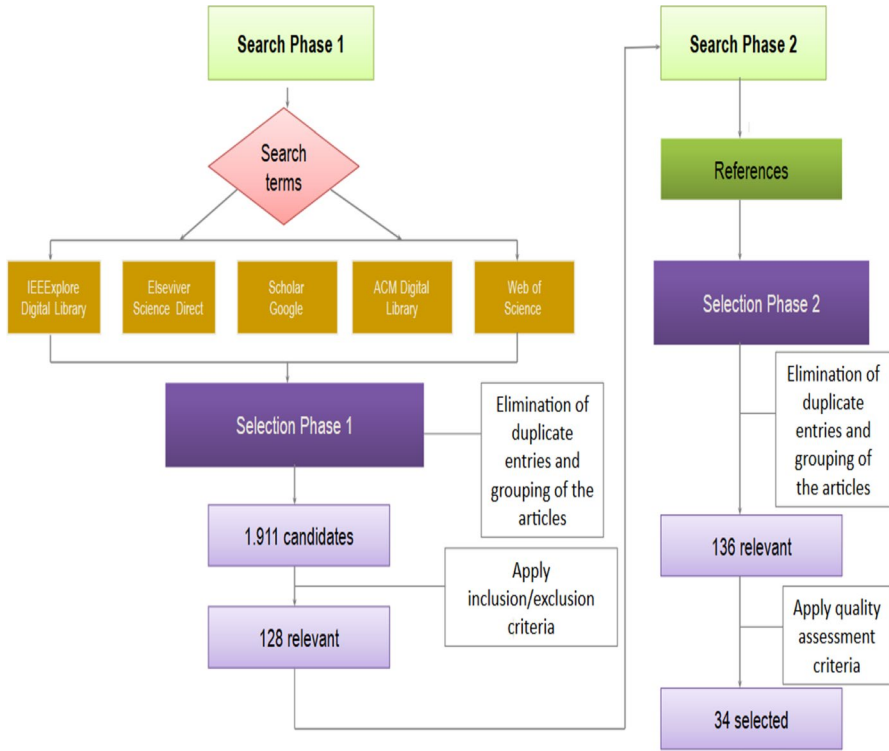
**Fig. 1** Search and selection process

in various sectors. The databases consulted considered to be the main repositories for searches on computer science are IEEExplore Digital Library, ACM Digital Library, Elsevier ScienceDirect, Google Scholar and Science Web. As a result of the first phase of the search, 1911 works were extracted. Once some inclusion/exclusion criteria were applied, 128 items were considered as relevant. In the second phase, 8 new works were considered, bringing the items up to 136. At the end of this phase 2, quality assessment criteria were applied, and the final number of items selected was 34. Figure 1 describes the process.

The methodological framework of the case study is based on an approach related to Supercomputing courses carried out in collaboration between a University (University of Leon, Spain) and a Supercomputing Center (SCAYLE). The use of a case study design is used to describe and explain what experiences based on field research are like, in order to establish their relationship with their real-life context [51]. In general, the case study results offer insights and expand the reader's experiences [34] in their construction of knowledge [40].

The courses included in the case study enhance computer competency and improve teaching and learning processes, introducing students to the use of

several technological tools and technology integration practices, to effectively use supercomputers.

In order to provide a better quality, the implementation of both courses follows the National Educational Technology Standards for Teachers (NETS) from the International Society for Technology in Education (ISTE). The ISTE (2008) standards were categorized as follows:

(a)  facilitate and inspire student learning and creativity,
(b)  design and develop digital age learning experiences and assessments,
(c)  model digital age work and learning,
(d)  promote and model digital citizenship and responsibility, and
(e)  engage professional growth and leadership.

In this article, the interest lies in how experiences provided by supercomputing training improves the performance of higher education students, instead of analyzing deeper technical aspects in relation of HPC.

## 3 Results

As mentioned above, the main objective of this paper is to present the description of a flowchart of the educational process related to Supercomputing, whose definition was completed once some partial objectives of this first study were developed: (1) the study of the main application fields in which students use Supercomputers; (2) the identification of the type of computational courses related to Supercomputing in the literature review and (3) the relational study of the field or sector that uses Supercomputing and the type of specific computational training. In the next section, the fourth secondary objective, the analysis based on the case study of a practical course, is described.

Figure 2 describes a flowchart in which the development of a Supercomputing course is analyzed, from the previous analysis of the students' profile to its closure. The process is based on the description of the tasks carried out by the entities that have collaborated in the case study presented in this work, which is closely linked to the findings of the literature review. The first step of the process, once the student has been registered, is to check their level of prior knowledge, especially in the use of Unix-like operating systems such as Linux. In the event that the student does not possess the minimum level of necessary knowledge, he will be prohibited from completing of the course. Once the course begins, special attention will be paid to the practical cases, since students seek to know how to apply Supercomputing in the solution of their tasks. In this process, feedback on students' performance is provided in order to improve course development, and a digital repository stores the cases to be used as examples in future training sessions. Before the end of the course, the assessment of students on the different topics discussed is analyzed and, if necessary, aspects that have not been clarified are repeated. When the teachers
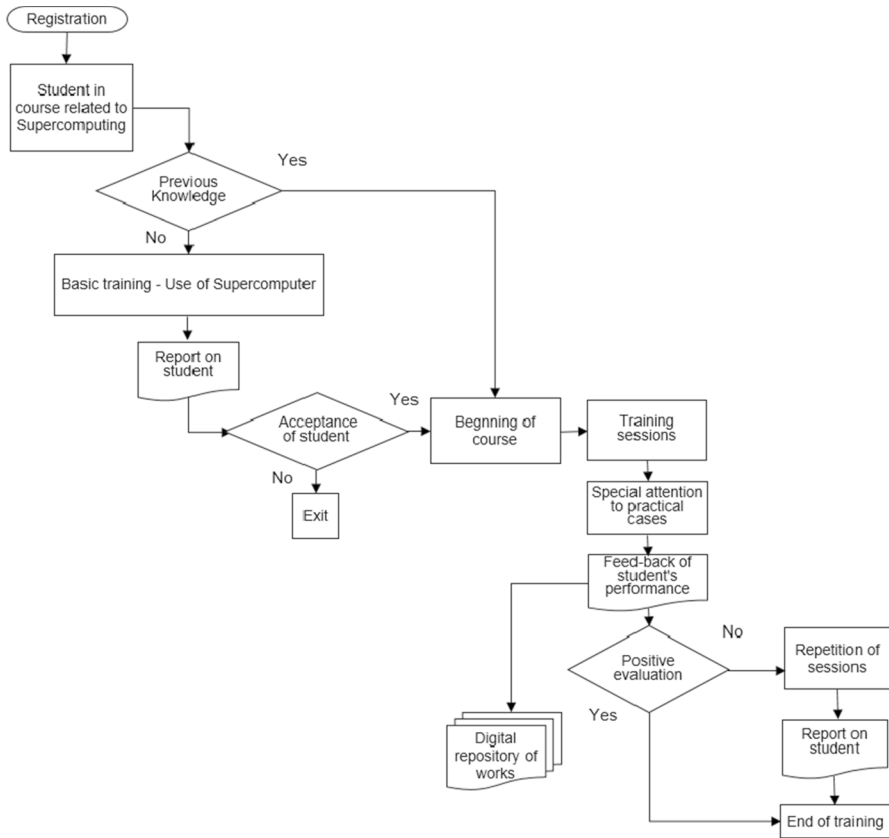
**Fig. 2** Flowchart of a training process related to supercomputing

consider that the students have understood all theoretical and practical concepts, the course ends, offering the possibility of repeating some sessions.

The next sections will describe the solution proposed for every partial objective, described previously:

## 3.1 Application fields

In order to better understand the details of the studies of each field analyzed in the review, the International Nomenclature of Unesco was used to group the most relevant fields of knowledge in the use of Supercomputers, according to each article where the field is referenced.

Table 1 shows the fields that use Supercomputers found in the literature review. The column "Category" shows a total of 52 references found related to fields of knowledge. In the column "Specialty", a more specific reference of the type of activity inside every category is included.

**Table 1** Fields of knowledge using Supercomputers

| Category | Specialty | Number of articles |
|---|---|---|
| Mathematics (4) | Mathematics and numerical analysis | 1 |
| | Control problems | 1 |
| | Automatic control | 1 |
| | Design automation | 1 |
| Astronomy and astrophysics (3) | Astrophysics | 1 |
| | Exploration of space | 1 |
| | Aerospace engineering | 1 |
| Physical (8) | Physical phenomena | 1 |
| | Physical processes | 3 |
| | Thermal and fluid dynamics | 4 |
| Chemistry (6) | Computational chemistry, chemical engineering and structural chemistry | 6 |
| Life sciences (8) | Biology | 3 |
| | Bio-informatics | 2 |
| | Genomics and proteomics | 2 |
| | Molecular simulation | 1 |
| Earth and space sciences (9) | Meteorology and weather forecasting | 5 |
| | Material science | 3 |
| | Prediction for forest re-propagation | 1 |
| Science of technology (12) | Image processing | 2 |
| | Drug design | 2 |
| | Construction of vehicles (land/amphibious/aerial): structural design, engine design, financial modeling and crash tests | 1 |
| | Engineering | 2 |
| | Microelectronics | 2 |
| | Sensor technology | 1 |
| | Nanotechnology | 1 |
| | Media | 1 |
| Economic sciences (1) | Economics and finance | 1 |
| Linguistics (1) | Linguistics | 1 |

## 3.2 Type of computational courses related to Supercomputing

In the scope of the same review mentioned before, an analysis of the computational courses related to Supercomputing was conducted.

As seen in Table 2, the column "Category" shows a total of 97 references detected through the review of the literature, related to the type of computational course identified. In the column "Specialty", a more specific reference to the content of the course is described.

**Table 2** Computational courses related to Supercomputing

| Category | Specialty | Number of articles |
|---|---|---|
| Programming (16) | Parallel programming | 10 |
| | Multiple-device programming | 1 |
| | Programming techniques for supercomputers | 1 |
| | Parallel and distributed programming | 2 |
| | Multicore programming and performance evaluation | 1 |
| | Code optimization | 1 |
| Computing (17) | Cluster computing | 3 |
| | Cloud computing | 2 |
| | Multicore, multicore computing and multicore clustering | 1 |
| | Parallel computing | 3 |
| | High-performance parallel computing | 2 |
| | Parallel and distributed computing | 2 |
| | HPC technology (hardware) | 1 |
| | Scalable computing | 1 |
| | Matrix calculation | 1 |
| | Heterogeneous computing | 1 |
| Software (6) | Applications | 1 |
| | User developer | 1 |
| | Software engineering development tools | 1 |
| | Software engineering | 1 |
| | Software design for complex and parallel systems | 1 |
| | Software parallelism | 1 |
| Systems and architectures (17) | Architecture of parallel systems | 8 |
| | Distributed and parallel system | 4 |
| | System administration | 1 |
| | Parallel systems | 1 |
| | High-performance computing systems | 1 |
| | Scientific computer systems | 1 |
| | Operating systems | 1 |
| Data management (14) | Visualization of scientific data and mathematical tools | 4 |
| | Distributed and parallel database | 1 |
| | Workflow management | 1 |
| | Data analysis/post-processing | 1 |
| | Data partitioning | 1 |
| | Data extraction | 1 |
| | Distributed and parallel databases | 2 |
| | Big data | 2 |

**Table 2** (continued)

| Category | Specialty | Number of articles |
|---|---|---|
| Others (28) | Parallel algorithms | 6 |
| | Optimization | 1 |
| | Cluster and grid middleware | 1 |
| | Network technologies | 1 |
| | Modeling and simulation | 6 |
| | Operational issues | 1 |
| | Performance analysis | 1 |
| | Verification and validation | 1 |
| | Domains | 1 |
| | HPC application packages | 1 |
| | Scientific resources | 1 |
| | Parallel libraries | 1 |
| | Scientific applications of parallelism | 1 |
| | Load balancing | 1 |
| | Checkpoints | 1 |
| | Introduction to accelerators | 1 |
| | Network infrastructure | 1 |
| | Parallel processing | 1 |

## 3.3 Relational study comparing the field of knowledge that uses Supercomputing and the type of specific computational training

The relational study was based on the identification of 12 articles that define together, a field of knowledge and a type of computer course. The initial number of papers selected was reduced, because not all articles mentioned data of both variables at the same time. To assess the relationship between the variable "field of knowledge" and the variable "type of training", The UNESCO criteria have been followed to define two dimensions: Sciences (including all the science-related subject and those related to economics and linguistics) and Engineering (includes articles related to the sub-discipline of Mathematics, Astronomy and Astrophysics, Chemistry and Physics). In order to achieve the proposed objectives, the construction of two multiple regression models has been carried out, with the dependent fields being the Sciences and Engineering fields mentioned above. In all the cases, the independent variables were the different types of computational courses developed, described in Table 2.

It is necessary to emphasize that in the relationships sought for the development of the models, the number of references, or frequency in which the variables are mentioned in the final selection of articles detailed in the index, is analyzed as reflected in Tables 1 and 2. The final number of articles to be used in the study was 12, due to the coincidences of references, both in the field of knowledge, as in the type of course mentioned.

The results obtained in relation to the parameters of each model, with a level of significance of 95%, are shown in Table 3.

In the case of training in the field of Sciences (Model 1), the parameter of training in Programming is the one that is positively related in greater proportion to explain the increase in training of computational courses. In Model 2, the parameter that best explains the increase in the number of courses in the Engineering is that related to Others (referring to algorithms, load balance, simulations, etc.).

In order to evaluate the validity and the degree of correlation that exists in the models, as well as their statistical significance, the analysis of variances (ANOVA) is shown in Table 4. The correlation coefficients between the dependent and independent variables of each model in Table 5 show the percentage for which the variables explain each other.

From the data presented above, it can be observed that only Model 1 satisfies the condition of critical value F being less than 0.05 and, therefore, this would be the only model of the two proposed in which the null hypothesis would be rejected. Therefore, Model 1 can be globally considered a valid model.

Table 5 analyzes the adjusted R2 coefficient to compare which of the proposed models best explains the correlation among variables. In Model 1, where the dependent variable is the "Sciences" field, it has a higher correlation percentage, which explains the degree of variation of the dependent variable as a function of the

**Table 3** Models created according to the criterion of dependent variable "field of knowledge"

| Model | Criteria | Predictors | Coefficients (β) | Typical error | Statistic $t$ |
|---|---|---|---|---|---|
| Model 1 | Sciences | Interception | 0 | N/A | N/A |
| | | Programming | 2.34 | 0.94 | 2.49 |
| | | Computing | −0.10 | 0.75 | −0.14 |
| | | Software | −0.24 | 2.56 | −0.09 |
| | | Systems and architectures | −0.68 | 0.77 | −0.89 |
| | | Data management | −1.31 | 1.34 | 0.98 |
| | | Others | 0.16 | 0.51 | 0.31 |
| Model 2 | Engineering | Interception | 0 | N/A | N/A |
| | | Programming | 0.01 | 1.09 | 0.02 |
| | | Computing | 0.32 | 0.87 | 0.36 |
| | | Software | −0.33 | 2.97 | −0.11 |
| | | Systems and Architectures | 0.04 | 0.89 | 0.05 |
| | | Data management | 0.55 | 1.54 | 0.35 |
| | | Others | 0.79 | 0.59 | 1.32 |

**Table 4** Details of ANOVA results for each model

|  | Degrees of freedom | Sum of squares | Average of squares | $F$ | Critical value of $F$ |
|---|---|---|---|---|---|
| Model 1 |  |  |  |  |  |
| Regression | 6 | 60.09 | 10.02 | 6.75 | 0.0268 |
| Waste | 6 | 8.91 | 1.48 |  |  |
| Total | 12 | 69 |  |  |  |
| Model 2 |  |  |  |  |  |
| Regression | 6 | 21.08 | 3.51 | 1.76 | 0.274 |
| Waste | 6 | 11.92 | 1.99 |  |  |
| Total | 12 | 33 |  |  |  |

**Table 5** Details of the correlation of the models

| Regression statistics | Model 1 Sciences | Model 2 Engineering |
|---|---|---|
| Multiple correlation coefficient | 0.93 | 0.80 |
| Coefficient of determination R2 | 0.87 | 0.64 |
| R2 adjusted | 0.60 | 0.17 |
| Typical error | 1.21 | 1.41 |
| Observations | 12 | 12 |

independent variables selected. The corresponding value of Model 2 is below 0.5 in the lower scale of the parameters that measure this coefficient, and therefore does not adequately explain the correlation as it happens in Model 1.

## 4 Case study

This section describes the case study based on the courses completed by the students related to Supercomputing. Courses analyzed were organized by the University of Leon and the Supercomputing Center of Castile and Leon (SCAYLE), and held during 2016. The objectives in both cases were to provide fundamental and advanced concepts by using very large database systems in a Supercomputer. Courses can be classified in the "Specific-domain courses" category. The analyzed courses are in the field of biotechnology, due to the fact that of those related to Supercomputing the courses from this field, developed by the entities involved, held the majority.

In the description of the case study, a prior analysis is of great importance, since, as described in this study, practically all of the findings of the literature review are applicable to the development of the training actions described.

The names of the courses were: ''Practical Course of Introduction to the use of Supercomputing applied to the analysis of RNA-Seq data—2nd Edition'' (C1) and "Practical Course on the use of Supercomputing applied to metagenomics and comparative genomics" (C2). The most relevant characteristics of the courses are summarized in Table 6.

**Table 6** Characterization of courses (significant/meaningful data)

| | C1 | C2 |
|---|---|---|
| Organizer | University of Leon | Supercomputing Center of Castile and Leon |
| Date | July 2016 | October 2016 |
| Number of students | 16 | 14 |
| Course duration | 28 h | |
| Previous Seminar of Linux for better use of Super-computing resources | Yes | |
| Number of students attending the previous seminar | 16 (100%) | 12 (86%) |
| Participants of the course | Researchers and/or higher education students interested in genomic studies, Computer Science professionals, Biologists and/or Biotechnologists related to genetic diagnosis and postgraduate University Students (with technical qualifications in the experimental field) and, in general, any person related to the topic of research, innovation and development | |
| Objectives | Managing and interpretation of data from global gene expression of Next Generation Sequencing using a Super-computer | |

Training programs in Supercomputing are normally conducted by highly qualified specialists with broad expertise in HPC systems and applications. They usually also have skills for the use of Supercomputers for a wide range of computer science issues. These specialists teach how to optimize the software used to work efficiently with HPC technologies. During the development of the courses, new teaching strategies were important in order to adopt technology based on Supercomputing. The new strategies have to enhance the pedagogical processes and ensure that the educational approach has been correctly applied while allowing personalization of the learning plans. They also have to be adapted, depending on previous knowledge and individual learning goals, described in a system in the literature where students come from a variety of backgrounds, and specific weaknesses [4]. For better organization and quality, courses were designed, developed and tested using the infrastructure of a real supercomputer cluster and a software appropriately programmed and optimized to generate quick results to solve complex problems.
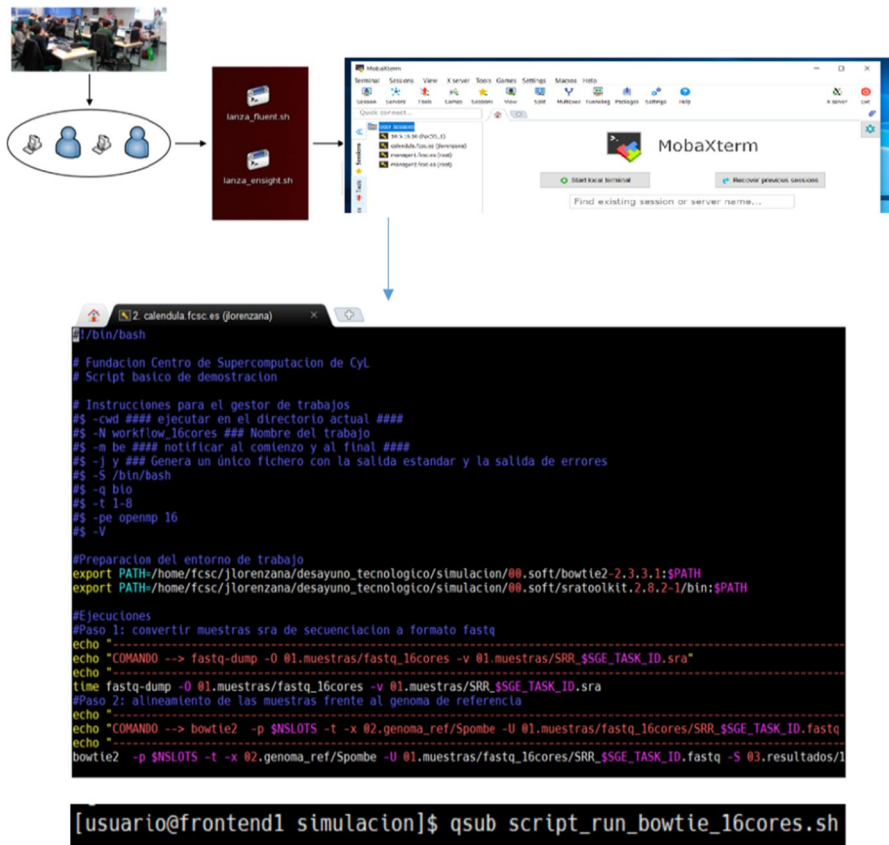


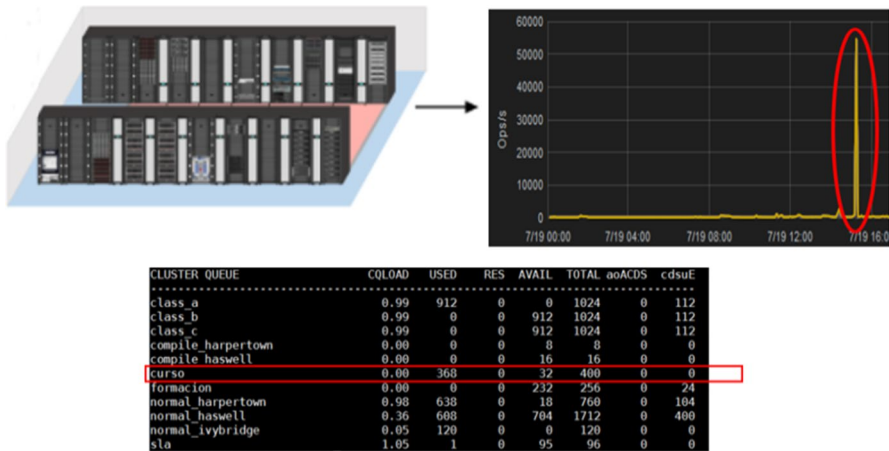**Fig. 3** Overview of students' work in Supercomputing courses

| CLUSTER QUEUE | CQLOAD | USED | RES | AVAIL | TOTAL | aoACDS | cdsuE |
|---|---|---|---|---|---|---|---|
| class_a | 0.99 | 912 | 0 | 0 | 1024 | 0 | 112 |
| class_b | 0.99 | 0 | 0 | 912 | 1024 | 0 | 112 |
| class_c | 0.99 | 0 | 0 | 912 | 1024 | 0 | 112 |
| compile_harpertown | 0.00 | 0 | 0 | 8 | 8 | 0 | 0 |
| compile_haswell | 0.00 | 0 | 0 | 16 | 16 | 0 | 0 |
| curso | 0.00 | 368 | 0 | 32 | 400 | 0 | 0 |
| formacion | 0.00 | 0 | 0 | 232 | 256 | 0 | 24 |
| normal_harpertown | 0.98 | 638 | 0 | 18 | 760 | 0 | 104 |
| normal_haswell | 0.36 | 608 | 0 | 704 | 1712 | 0 | 400 |
| normal_ivybridge | 0.05 | 120 | 0 | 0 | 120 | 0 | 0 |
| sla | 1.05 | 1 | 0 | 95 | 96 | 0 | 0 |

**Fig. 4** Details of graphic tools during the development of a course

The analysis of the graphics shown in Figs. 3 and 4 provides information on how the supercomputing infrastructure was used during the course. In Fig. 3, we can see the diagram of how the Supercomputer was used in C1 and C2 sessions. Once students complete the theoretical phase, they begin to test practical cases, related to the specialization of the course, sending jobs to the Supercomputer. They have an immediate feedback of the simulation they run, and they can directly see directly the quality and efficiency of the use of these infrastructures in the optimization of the performance of a process, enhancing the learning.

As a basic part of the courses, students run computational simulations using real data. A graphic tool is in charge of controlling the load of the infrastructure in order to provide the best efficiency for the users (Fig. 4). With the support of the Supercomputer, the graph shows a peak marked with a red circle, which represents the operations per second in the storage system. The highlighted detail corresponds to the exact moment when this data was processed in the practical sessions, and shows the peak of data access when students access data stored on disk to begin calculations. It is made from SCAYLE's monitoring systems (Grafana, see web https://grafana.com). The screen image at the bottom shows a complete list of the jobs that were running on the Supercomputer at the same moment, highlighting the servers that were assigned to the course in the Cluster Queue manager. This is obtained from the SGE job manager qstat command, which displayed all created queues and their status (available cores, total cores, used cores, average load, etc.).

## 4.1 Participants and data collection

Both courses were designed for students with diverse backgrounds, and are usually made up of 12 to 20 participants at the most, in order to enable a more personalized

**Table 7** Students' profile

| Student profile | C1 | C2 |
|---|---|---|
| Doctorate with experience | 2 | 2 |
| Post-doctorate | 2 | 2 |
| Researcher | 11 | 10 |
| Master student | 1 | |

education for students. The number and characteristics of the students are shown in Table 7 below.

A key step in every research is to be clear about its purpose and scope. The implementation of methods for data collection and analysis are essential for all types of evaluations. Data collection is an iterative process based on the search strategies of the study. In the case of this paper, in order to complete the quantitative analysis for obtaining information to analyze the results, a questionnaire consisting of ten items which mixed Likert-type, closed- and open-ended questions, was provided to students at the end of the sessions in both courses.

The questionnaire reliability was tested previously in SCAYLE, where this tool is used as the standard for evaluating the quality of the training provided to the students in their own training programs. This questionnaire provides, through the use of specific key evaluation questions, the evidence needed to make appropriate judgements about the quality of the courses. The consistency of the items of the questionnaire was evaluated by experts who also contributed to the design and adaptation of data collection instruments.

### 4.2 Survey results

The evaluation questionnaire includes a first part that evaluates the introductory aspects of the course, a second part requiring an evaluation of the teachers and, in the last part, a final consideration of the whole course. Table 8 includes the full text of questions that were asked in the survey, and displays the results of the students who answered the survey. Items can be classified into three groups: (i) introductory aspects (Q1 to Q5); (ii) assessment of the teachers (Q6 and Q7); (iii) overall satisfaction and final considerations (Q8 to Q10).

The analysis of the responses of every group mentioned above is the following:

– *Introductory aspects* The introductory seminar was considered very useful by all the students in both courses, so it is essential to remedy the lack of knowledge regarding aspects needed to use a Supercomputer. In this connection, the mark of both biotechnologist and computer engineer profile teachers of the introductory seminars was very high, with marks close to the maximum. Finally, most of the students consider that the other introductory aspects, such as the material provided in the course and the dissemination, were adequate.

**Table 8** Results of the survey

| Questions | | C1–N = 15 | C2–N = 13 |
|---|---|---|---|
| Q1 | Was the introductory seminar to Linux useful for you? | Yes: 15 | Yes: 13 |
| Q2 | Average assessment of the introduction seminar teacher, biotechnologist profile (1–5) | 4.62 | 4.92 |
| Q3 | Average assessment of the introductory seminar teacher, computer engineer profile (1–5) | 4.60 | 4.40 |
| Q4 | Do you consider the teaching material as adequate? | Yes: 12 | Yes: 11 |
| | | Unanswered: 3 | Unanswered: 2 |
| Q5 | Has the course been adequately disseminated? | Yes: 14 | Yes: 13 |
| | | No: 1 | No: 0 |
| Q6 | Would you change anything about the seminar? | No: 5 | No: 8 |
| | | Unanswered: 5 | Unanswered: 4 |
| | Reduce the technical part of the startup seminar | 1 | 1 |
| | Include more practical aspects in the sessions | 3 | |
| | Take into account that students have low computer skills | 1 | |
| | Lower the level of the course | 1 | |
| | Go further and deeper in course content | 3 | |
| Q7 | Average evaluation of sessions of research professors (1–5) | 3.59 | 4.46 |
| Q8 | Did the course cover your expectations? | Yes: 12 | Yes: 10 |
| | | Unanswered: 3 | Unanswered 3 |
| Q9 | Has the course methodology been attractive? | Yes: 12 | Yes: 10 |
| | | Unanswered: 3 | Unanswered: 3 |
| Q10 | Average of overall assessment of the course (1–10) | 8.83 | 9.69 |

– *Assessment of the teachers* In order to ascertain the student's perception of the courses, we asked question Q6, relating to possible changes in the conduct of the course. The number of unanswered questions was high, but in any case, a more practical content and further extension and depth were proposed. In the case of the evaluation of the specialized teachers (Q7), we can see that teachers of course C2 were evaluated better, probably because C1 was a very intensive course, with the same number of sessions and double the number of teachers, making the teaching task more difficult.

– *Overall satisfaction and final considerations* Questions Q8 to Q10 provide the final assessment of the course. Most of the students consider that the course met their expectations (Q8) and that the method of teaching was attractive (Q9). The final assessment of both courses was very high: 8.83 for C1 and 9.69 for C2.

# 5 Discussion

This study first analyzes the results obtained in the literature review and, later, those of the case study. At the end of this section, information regarding the relationship of both studies will be provided.

## 5.1 Supercomputing in the educational process

The increase in supercomputing infrastructures has paralleled the complexity of research and the problems facing today's society and has been developed to obtain better results in various fields of knowledge and sectors of activity. Currently, numerical simulation for obtaining results of research is widely practiced, both in the academic arena and in industry in general, to solve problems in a short period of time. For this reason, training in these technologies is essential for an adequate performance of professionals working in diverse types of fields and activities.

For the consolidation of Supercomputing in the educational process, it is essential to invest in the training of competent personnel in the proper use of Supercomputers [16]. Supercomputing training improves the results of the STEM and non-STEM, or non-computing students, in a Higher Education Institution (HEI), among others. With the help of many tools such as new and easy-to-use software computer simulation is made accessible to a wider group of people.

For implementing Supercomputing, the need of cooperation between professionals of multiple skills and of different backgrounds—an important added value—should be based on two interrelated aspects: (1) different profiles of a variety of fields of knowledge in different competences in STEM and non-STEM, non-STEM students; (2) different experiences in the integration of technology in the Higher Education system.

Our findings suggest, through the results seen in the literature review, that including Supercomputers in the learning processes of a great variety of fields has provided many opportunities for learners.

### 5.1.1 Outcomes

Many factors, as constraints in the curriculum, such as the rapid changes in the field of Supercomputing in general, or the sophistication of the mathematical models, influence the use of supercomputing facilities.

This study provides three main outcomes related to state-of-the-art Supercomputing training, which respond to the objectives raised at the beginning of this study and which are summarized below.

- Objective 1: through the analysis of the 52 articles of dealing with different fields related to the use of Supercomputers, an important number of fields that are considered non-STEM or non-computing were detected, such as a wide number of specialties, that are the basis of many industrial processes.

- Objective 2: it was found that many computational courses are described in the literature, which influence activities related to Supercomputing. These courses clearly enhance the performance of the users of these infrastructures, by improving the processes on which research is based, allowing a better interpretation of the results.
- Objective 3: according to the relational study, Model 1, linked to the items related to the field of Science, show that programming is the variable close related to this field of knowledge.

### 5.1.2 Limitations

During the course of the literature review, the greatest difficulty encountered was the heterogeneity in the type of articles, and also in the content related with the typology of training. Another limitation of the study was the fact that, due to the nature of the information, it did not focus on analyzing training related to Supercomputing as an element of the teaching–learning process, so it would be necessary to complement the study with others related to the use of high-performance computers as new didactic tools.

Finally, another limitation is related to the methodology used in the regression models, which is based on a correlational approach. The results achieved thereby allow us to speak of a certain relationship between the field of knowledge and the type of specialized computational course, but which does not necessarily mean a causality between these variables. For further development of this research, studies based on experimental designs will be necessary to confirm the effect of independent variables on the development of computational training in the fields of knowledge.

Our hope is also that the present study will encourage researchers and institutions in general to contribute to the development of agreements which address the challenges of using a Supercomputer as a fundamental part of the research of many fields, through the use of a real Supercomputing infrastructure and the inclusion in the academic curriculum of both STEM and non-STEM students from subjects related to this issue [35].

### 5.2 Case study

### 5.2.1 Outcomes

This study of cases gives an opportunity to examine how the procedures of training in a technological environment provide and facilitate the engagement of the students in the improvement of their performance [14]. The issues raised in this part of the paper should be relevant in order to provide new tools for higher education students, thereby complementing their skills in the use of technology. We highlight that, in spite of some cases described in which students participate in curriculum design [13], currently, contents in HPC are practically testimonial in the academic curriculum of higher education [26, 37, 41].

Supercomputing training appeared to support student's performance and improve learning activities by providing a platform that helps in the pursuit of real cases in many fields. In the future, the design of technologically advanced courses should be conducted with the use of a Supercomputer, thus moving forward with regard to the training offered and following new pedagogical methods.

However, although these courses have, in general, received favorable comments and positive assessment, some students expressed the need for some aspects to be improved in the future, such as the lack of more time for practical sessions and deeper explanations. Additionally, the high level of the course in terms of specific computational skills presumed requires attention, because some students declared having a low-level computational background using a Supercomputer in an optimal manner. The most important questions that can be considered for future courses are the following:

– Students consider the support of the introductory seminar essential. In some open questions they even express the belief that they would need more training in some areas, prior to the beginning of the main content of the course in order to optimize their performance.
– Increasing the support of the technicians of Supercomputing Centers, in order to improve and expand the capacity for using the infrastructures to test the practical aspects of the courses—something highly demanded by students.
– Enhancing planned didactic aspects of the classes, to ensure the best performance possible from teachers.
– Implementing the design of an easy-to-use interface needs to be carried out, to facilitate human–computer interaction with an intuitive experience during the use of a Supercomputer.

### 5.2.2 Limitations

This study has certain limitations that should be considered when interpreting its results and conclusions. Firstly, a larger number of participants would be more suitable, because, due to the relatively small number of participants of both courses (30, of which 28 responded to the survey), results should be interpreted with caution. Secondly, the courses considered in the study were taught by different instructors, in specific areas, and the profiles of the students were also different, which complicates the study of a homogeneous population. In order to minimize the effects of these limitations, similar levels of education and experience of the teachers, and a close monitoring of the implementation of the content in both courses, would be essential.

In the development of the work, it has been possible to verify how most of the aspects analyzed in the literature review have been observed in the courses analyzed in the case study. Thus, aspects such as lack of prior knowledge in the use of a Supercomputer, the notable improvement in the qualification of the students, or the effective availability of supercomputing infrastructures to carry out practical cases on simulating reality, have made it possible to carry out actions to improve the training offered.

## 6 Conclusions

The paper shows how the implementation of an innovative pedagogy for teaching STEM-related subjects for non-STEM and/or non-computing students is very important in today's society. There is a strong potential of expansion for interdisciplinary studies and addressing other major questions all through the networks and connections to be created by highly qualified personnel involved. A step forward in the integration of Supercomputing in STEM and non-STEM fields of knowledge through a change in policies might lead to a wider innovation in the traditional way of the educational process: improving teaching and learning.

Currently, the use of software tools for the deployment of theoretical methods and data analysis used for mathematical modeling, and techniques of computational simulation, are wide extended for the study of fields such as biological systems, through high-performance computing [46]. This will create new learning environments which allow the best use possible of the available technologies, showing a way for the improvement in the offer of practical sessions in the programs in higher education.

In the courses related to supercomputing, teachers work in a technology-oriented manner based on authentic situations and problems that illustrate the potential of HPC technologies in learning practices. In order to reach the goal proposed, the implementation of the courses should follow these principles: (1) organization of an introductory seminar that contributes to a better performance for the rest of the course; (2) support of specialized technicians and on-site Supercomputing infrastructures, providing an optimal use of the computational resources for running simulations as well as testing the capacity for solving practical problems, and (3) instructions provided by highly qualified experts.

We can conclude that HPC technologies are good pedagogical tools, which allow students to learn in different ways, in the most diverse fields of knowledge, considering that training in Computer and Information Sciences (CIS) is largely influenced by the way big data scientists are trained through the use of machine learning and artificial intelligence techniques [49]. Currently, deep learning is also described as a new philosophy of training based on technology [5]. By changing the way Computer Science is advertised and taught, a more inclusive Computer Science is possible [2].

Related to the support provided by the Supercomputing Center in the design, implementation and monitoring of the training offered to students, we can conclude that it contributes to the training sessions in the following positive aspects:

- By using the Supercomputing Center infrastructures for training, the work of the students is constantly monitored. This way, the use of the resources can be optimized, providing a higher computational capacity if necessary.
- Continuous support by technicians, experts in HPC, covering several areas of Supercomputer use, solving problems and providing a high level of security is assured.

– Software involved in computational simulations is being tested permanently in order to offer more reliable results.
– Students carry out their software assignments with the help of specialized super-computing technicians, resulting in better performance.

The use of algorithms and flowcharts for the identification of students' needs is extended, allowing the use of diagnosis decision trees through the use of a distributed architecture that combines shared-data and client–server styles [28]. Currently, algorithms that provide an optimized environment for solving large scale optimization problems, helping researchers in the application of evolutionary algorithms for a wide range of fields [36], are commonly used. In this light, the flowchart described in this paper shows a successfully tested way of conducting training sessions related to Supercomputing based on the experience of the cases studied.

Finally, future work should be focused, comprehensively and simultaneously, on analyzing new methods and tools to evaluate the quality of Supercomputing training in order to improve it, constantly adapted to the needs of students in every moment. Accordingly, it is important to explore the requirements of all the fields the use of Supercomputers could be helpful, trying to find proper ways of promoting such training in students of higher education.

# References

1. Abuzaghleh O, Goldsmit K, Elleithy Y, Lee J (2013) Implementing an affordable high-performance computing for teaching-oriented computer science curriculum. ACM Trans Comput Educ TOCE 13(1):1–14
2. Aguar K, Arabnia HR, Gutierrez JB, Potter WD, Taha TR (2016) Making cs inclusive: an overview of efforts to expand and diversify cs education. In: 2016 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, New York. pp 321–326
3. Aguar K, Safaei S, Arabnia HR, Gutierrez JB, Potter WD, Taha TR (2017a) Reviving computer science education through adaptive, interest-based learning. In: 2017 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, New York. pp 1161–1166
4. Aguar K, Sanchez CC, Beltran DB, Safaei S, Asefi M, Arnold J et al (2017b) Considerations on interdisciplinary instruction and design influenced by adaptive learning. A case study involving biology, computer science, mathematics, and statistics. arXiv preprint arXiv:1703.06010

5.  Amirian S, Rasheed K, Taha TR, Arabnia HR (2019) A short review on image caption generation with deep learning. In: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV) (pp 10–18). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp)

6.  Anderson LM, Petticrew M, Rehfuess E, Armstrong R, Ueffing E, Baker P, Francis D, Tugwell P (2011) Using logic models to capture Complexity in systematic reviews. Res Synth Methods 2:33–42

7.  Armosky B, Brown S, Drumond T, Ferguson J, Gerber R, Hacker TJ, Traxler K (2007) HPC University. In: TG08-TeraGrid Conference, Madison, WI

8.  Bader DA (2004) Computational biology and high-performance computing. Commun ACM 47(11):34–41

9.  Bergeron K, Cummings R, Decker CR, Freeman MJ, Hoke CC, Seidel J, Mc Daniel DM (2008) An undergraduate computational aerodynamics curriculum. In: DoD HPCMP Users Group Conference, 2008. DOD HPCMP UGC. IEEE, New York. pp 18–24

10. Bernabé G, Cuenca J, García LP, Giménez D, Rivas-Gómez S (2014) A high performance computing course guided by the LU factorization. Procedia Comput Sci 29:1446–1457

11. Bethel EW, Van Rosendale J, Southard D, Gaither K, Childs H, Brugger E, Ahern S (2010) Visualization at supercomputing centers: the tale of little big iron and the three skinny guys. IEEE Comput Graph abd Appl 31(1):90–95

12. Bloom B (1994) Reflections on the development and use of the taxonomy. In: Anderson L, Sosniak L (eds) Bloom's taxonomy: a forty-year retrospective. The National Society for the Study of Education, Chicago, pp 1–8

13. Carey P (2013) Student as co-producer in a marketised higher education system: a case study of students' experience of participation in curriculum design. Innov Educ Teach Int 50(3):250–260

14. Coates H (2005) The value of student engagement for higher education quality assurance. Qual High Educ 11(1):25–36

15. Cooper HM (1988) Organizing knowledge syntheses: a taxonomy of literature reviews. Knowl Soc 1(1):104

16. Coveney P, McGuire S, Parchment O, Kenway R, Parsons M (2012) UK participation in high performance computing (HPC) at the European level—publications. https://www.gov.uk/government/publications/uk-participation-in-high-performance-computing-hpc-at-the-european-level

17. Crouch CH, Mazur E (2001) Peer instruction: ten years of experience and results. Am J Phys 69(9):970–977

18. Denny P, Hamer J, Luxton-Reilly A, Purchase H (2008) PeerWise. In Proceedings of the 8th International Conference on Computing Education Research—Koli'08. ACM Press, New York. pp 109–112. http://portal.acm.org/citation.cfm?doid=1595356.1595378

19. Fabricius U, Freundl C, Köstler H, Rüde U (2005) High performance computing education for students in computational engineering. Lecture Notes Comput Sci 3515(II):27–35

20. Farian H, Anne KM, Haas M (2008) Teaching high-performance computing in the undergraduate college CS curriculum. J Comput Sci Coll 23(3):135–142

21. Fernández Slezak D, Turjanski PG, Montaldo D, Mocskos EE (2010) Hands-on experience in HPC with secondary school students. IEEE Trans Educ 53(1):128–135

22. Fernández Á, Fernández C, Miguel-Dávila JÁ, Conde MÁ, Matellán V (2019) Supercomputers to improve the performance in higher education: a review of the literature. Comput Educ 128:353–364

23. Fink A (2010) Conducting research literature reviews: from the internet to paper, 3rd edn. Sage Publications, New York

24. Hacker T (2010) Hands-on high performance computing developing a cluster computing course for real world supercomputing. In: Proceedings of the National ASEE Conference. 3

25. Hemanth JD, Kose U, Deperlioglu O, De Albuquerque VHC (2012) An augmented reality-supported mobile application for diagnosis of heart diseases. J Supercomput 76(2):1242–1267

26. Holmes V, Kureshi I (2015) Developing high performance computing resources for teaching cluster and grid computing courses. Procedia Comput Sci 51:1714–1723

27. Joiner DA, Gray P, Murphy T, Peck C (2006) Teaching parallel computing to science faculty: Best practices and common pitfalls. In: Proceedings of the 11th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP'06). ACM, New York, pp 239–246

28. Joshua R, Scuse DH (2005) Modeling beliefs and solution strategies in a distributed learning system. J Supercomput 34(1):27–39

29. Kim K, Moon N (2019) Activity index model for self-regulated learning with learning analysis in a TEL environment. J Supercomput 75(4):1971–1989
30. Kitchenham B (2004) Procedures for performing systematic reviews, vol 33. Keele University, Keele, pp 1–26
31. Lee CM, Chang H (2014) A study on security strategy in ICT convergence environment. J Supercomput 70(1):211–223
32. Lu F, Song J, Cao X, Zhu X (2012) CPU/GPU computing for long-wave radiation physics on large GPU clusters. Comput Geosci 41:47–55
33. Martín AG (2007) Integración curricular de las TIC y educación para los medios en lA sociedad del conocimiento. Revista Iberoamericana de educación 45:141–156
34. Merriam SB (1998) Qualitative research and case study applications in education. San Francisco: Jossey-Bass Publishers, 350 Sansome St, San Francisco, CA 94104
35. Minaie A, Sanati-Mehrizy R (2009) Incorporating parallel computing in the undergraduate computer science curriculum. In: Proceedings of the National ASEE Conference
36. Mohammadi FG, Amini MH, Arabnia HR (2019) Evolutionary computation, optimization and learning algorithms for data science. arXiv preprint arXiv:1908.08006
37. Munshi R, Coalson RD, Ermentrout GB, Madura JD, Meirovitch H, Stiles JR, Bahar I (2006) An introduction to simulation and visualization of biological systems at multiple scales: a summer training program for interdisciplinary research. Biotechnol Prog 22(1):179–185
38. Ryabinin K, Chuprina S (2015) Using scientific visualization tools to bridge the talent gap. Procedia Comput Sci 51:1734–1741
39. Shiflet AB, Shiflet GW (2010) Simulating the formation of biofilms in an undergraduate modeling course. Procedia Comput Sci 1(1):895–901
40. Stake RE (1994) Case studies. In: Denzin NK, Lincoln YS (eds) Handbook of qualitative research. Sage, Thousand Oaks, pp 236–247
41. Smith J, Wolffe G (2007) Introducing AP computer science students to high-performance computing. J Comput Sci Coll 23(1):70–76
42. Stainsby H, Muresano R, Fialho L, González JC, Rexachs D, Luque E (2009) Teaching model for computational science and engineering programme. In: Computational Science–ICCS 2009. Springer, Berlin, pp 34–43
43. Stojanovic N, Milanovic E (2015) Teaching introductory parallel computing course with hands-on experience. Int J Eng Educ 31(5):1343–1351
44. Suarez TL (2007) Access Grid technology in classroom and research environments. J Supercomput 41(2):133–145
45. Sung NS, Gordon JI, Rose GD, Getzoff ED, Kron SJ, Mumford D, Onuchic JN, Scherer NF, Summers DL, Kopell NJ (2003) Science education. Educating future scientists. Science 301:1485
46. Tran QN, Arabnia H (2015) Emerging trends in computational biology, bioinformatics and system biology. Moran Kauffmann publishers, Elsevier, Amsterdam
47. Wyld M (1996) Technology refusa: rationalising the failure of student and beginning teachers to use computers. Br J Edu Technol 27:134–143
48. Yang CH, Tzuo PW, Komara S (2011) WebQuest and collaborative learning in teacher preparation. Singapore study. Educ Media Int 48(3):209–220
49. Yang MQ, Yu S, Yang W, Milanova M, Zhao W, Yang JY, Arabnia HR (2017) Developing secure privacy preserving and causal genetic alteration research in building an innovative systematic pedagogy for integrated research and education: the INSPIRE model. In: 2017 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, New York, pp 1149–1154
50. Yasar O, Maliekal J, Little LJ, Jones D (2006) A computational technology approach to education. Comput Sci Eng 8(3):76–81
51. Yin RK (2009) Case study research: design and methods, 4th edn. Sage, London