Face-2-face Spoke 2 Centro Nazionale “HPC, Big Data e Quantum Computing”

WP2.4 - planning iniziale e considerazioni
(piu’ che altro … appunti e qualche considerazione per discutere insieme)

Alexis Pompili (UNIBA & INFN-Bari)

18.07.2022 - Bologna & remote - https://agenda.infn.it/event/32118/

La narrativa del WP2.4 - I

WP2.4: **Boosting the computational performance of Theoretical and Experimental Physics algorithms** (INFN, INAF, UNIBA, UNIMIB, UNINA, UNIBO, UNIFI, UNICT, UNITS, UNIPD, UNIFE): T4.1 Tools and guidelines for developing and porting heterogenous codes and algorithms on modern architectures; T4.2 Competence and training centre for heterogeneous computing.

2 tasks principali

Main Activities: porting of applications to GPUs and heterogeneous architectures (e.g., scalability of scientific codes and applications on GPU/CPU many-cores clusters, local and remote offloading, mission-critical algorithms on FPGAs, ...). The solutions and tools implemented during the project will be easily extendable to other scientific domains of the Centre and to the industrial partners in the Spoke; moreover, the personnel trained within the Centre will help to spread and boost the application of HPC methodologies to Italian academic and industrial fields, for a comprehensive advancement of the Italian system.

Milestones: M9-M15: report on best practices for heterogeneous computing; M22-M26: first training opportunity; testbeds ready for users (initial handshake); user support in place; M25-36: results from testbed and benchmarking activities; final report on technologies, training and support system; white paper for use cases external to the CN

La narrativa del WP2.4 - II

4 Phases:

The three "technical Work Packages" WP 2.4, 2.5 and 2.6 follow a similar approach:

1. planning and identification: landscape recognition for best solutions for the realization of heterogeneous and portable code (e.g. software frameworks, compilers, programming models, …), for the integration of services into a data-lake infrastructure; cross domain software and services will be identified if appropriate. Moreover, solutions for handling user support, user fora, and training opportunities will be identified;
2. a realization phase, in which the services and the support systems are put into place, at least in alpha/beta phase. These include the testbeds to be used for benchmarking of scientific and industrial solutions, the user support system, the training opportunities.
3. a validation phase, in which experience on the supported services and codes are reported, to be used as a touch base before the end of the project.
4. a wrap-up phase, in which results are reported for executed activities, and are disseminated via white papers for future and external use cases.

Nelle slide seguenti vi entreremo con qualche dettaglio e considerazione.

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1

WP2.4: Boosting the computational performance of Theoretical and Experimental Physics algorithms (INFN, INAF, UNIBA, UNIMIB, UNINA, UNIBO, UNIFI, UNICT, UNITS, UNIPD, UNIFE):

Task-1

Identify tools & deliver guidelines } for { developing & porting } *heterogeneous codes and algorithms* … on modern architectures

D: Quali sono questi … ***heterogeneous codes and algorithms*** ?

R: Sicuramente gli …

… ***innovative algorithms*** ed anche ***AI inspired techniques*** di cui ai WP2.x (x=1,2,3) (“WP scientifici”)

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-1

In altre parole…

il Task-1 consiste nel raccogliere/mettere a punto un insieme di conoscenze/competenze che supportino le attivita' dei WP "scientifici" nella transizione dei loro codici ed algoritmi (sia di ricostruzione che di selezione e simulazione degli eventi) …**ad architetture computazionali eterogenee**,

… coerentemente con quanto qui specificato:

Main Activities: porting of applications to GPUs and heterogeneous architectures (e.g., scalability of scientific codes and applications on GPU/CPU many-cores clusters, local and remote offloading, mission-critical algorithms on FPGAs, ...).

… e con la **Fase-1 / Planning&Identification :**

planning and identification: landscape recognition for best solutions for the realization of heterogeneous and portable code (e.g. software frameworks, compilers, programming models, …), for the integration of services into a data-lake infrastructure; cross domain software and services will be identified if appropriate. ~~Moreover, solutions for handling user support, user fora, and training opportunities will be identified;~~ (ne parlo dopo)

… e con la **milestone** **M9-M15: report on best practises for heterogeneous computing**

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-1

Bisogna quindi individuare preliminarmente (iniziando dall'ambito WLCG) quali metodi/soluzioni esistono per…
… **portare e mantenere del codice eterogeneo da eseguire su piattaforme HPC (CPU+GPU).**

I piu' importanti requisiti da garantire sono (`arXiv:1602.08477`**):**

1) **Heterogeneity** : ***write once, execute everywhere***
2) **Testability** : ***validate once, get correct results everywhere***
3) **Sustainability** : ***porting must imply minimal code changes***
4) **Optimizability** : ***fine-tuning for good performance at minimum coding effort***
5) **Openness** : ***open source and open standards***

CUDA non fornisce (1)(2),(3) e naturalmente (5).

Seguono alcune slide giusto per richiamare la consapevolezza che vi sono diverse soluzioni "sul mercato"
(ne richiamo un paio fra quelle citate in `arXiv:1602.08477`):

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-1

alpaka

Abstraction Library for Parallel Kernel Acceleration

https://www.casus.science/research/software-repository/

The alpaka library is a header-only C++14 abstraction library for accelerator development.

Its aim is to provide performance portability across accelerators through the abstraction (not hiding!) of the underlying levels of parallelism.

It is platform independent and supports the concurrent and cooperative use of multiple devices such as the hosts CPU as well as attached accelerators as for instance CUDA GPUs and Xeon Phis (currently native execution only). A multitude of accelerator back-end variants using CUDA, OpenMP (2.0/4.0), Boost.Fiber, std::thread and also serial execution is provided and can be selected depending on the device. Only one implementation of the user kernel is required by representing them as function objects with a special interface. There is no need to write special CUDA, OpenMP or custom threading code. Accelerator back-ends can be mixed within a device queue. The decision which accelerator back-end executes which kernel can be made at runtime.

The abstraction used is very similar to the CUDA grid-blocks-threads division strategy. Algorithms that should be parallelized have to be divided into a multi-dimensional grid consisting of small uniform work items. These functions are called kernels and are executed in parallel threads. The threads in the grid are organized in blocks. All threads in a block are executed in parallel and can interact via fast shared memory. Blocks are executed independently and can not interact in any way. The block execution order is unspecified and depends on the accelerator in use. By using this abstraction the execution can be optimally adapted to the available hardware.

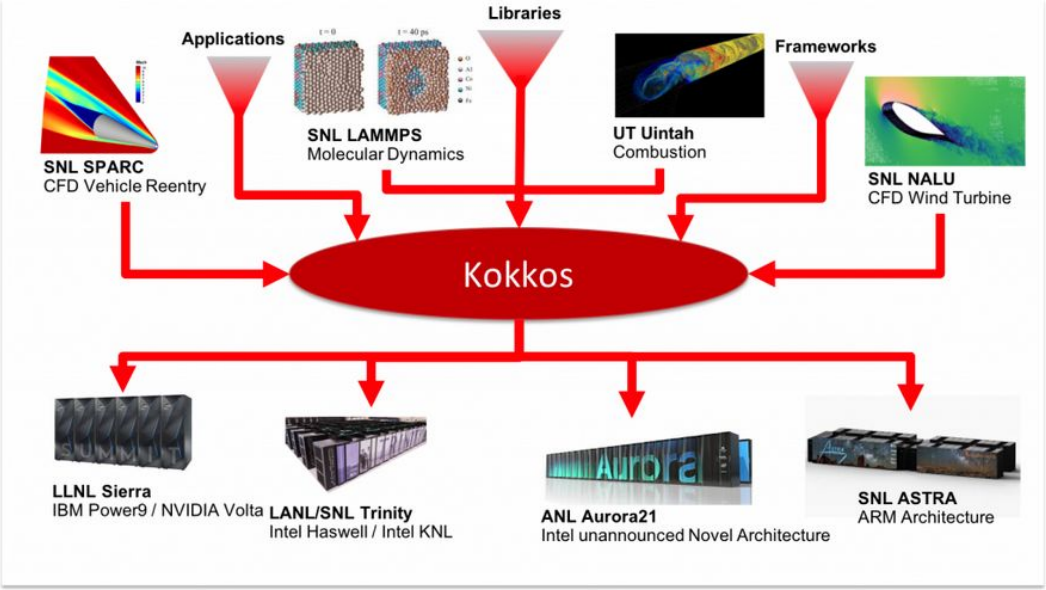
DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-1

kokkos

HOME ABOUT REPOSITORIES DOCUMENTATION BLOG OUTREACH COMMUNITY

About

The Kokkos C++ Performance Portability EcoSystem is a production level solution for writing modern C++ applications in a hardware agnostic way. It is part of the US Department of Energies Exascale Project – the leading effort in the US to prepare the HPC community for the next generation of super computing platforms. The EcoSystem consists of multiple libraries addressing the primary concerns for developing and maintaining applications in a portable way. The three main components are the Kokkos Core Programming Model, the Kokkos Kernels Math Libraries and the Kokkos Profiling and Debugging Tools.



https://kokkos.org/about/

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-1

https://www.khronos.org/api/index_2017/sycl

SPERIMENTAZIONE di diverse soluzioni IN CORSO - I

E' sufficiente dare un'occhiata agli abstracts sottomessi ad ACAT2022, rimanendo solo in ambito CMS, per capire che c'e' vivace esplorazione e sperimentazione in corso…

Adoption of the alpaka performance portability library in the CMS software

Andrea Bocci (talk, abstract on CINCO)

To achieve better computational efficiency and exploit a wider range of computing resources, the CMS software framework (CMSSW) has been extended to offload part of the physics reconstruction to NVIDIA GPUs, while the support for AMD and Intel GPUs is under development. To avoid the need to write, validate and maintain a separate implementation of the reconstruction algorithms for each back-end, CMS decided to adopt a performance portability framework. After evaluating different alternative, it was decided to adopt Alpaka as the solution for Run-3.

Alpaka (Abstraction Library for Parallel Kernel Acceleration) is a header-only C++ library that provides performance portability across different back-ends, abstracting the underlying levels of parallelism. It supports serial and parallel execution on CPUs, and extremely parallel execution on GPUs.

This contribution will show how Alpaka is used inside CMSSW to write a single code base; to use different toolchains to build the code for each supported back-end, and link them into a single application; and to select the best back-end at runtime. It will highlight how the alpaka-based implementation achieves near-native performance, and will conclude discussing the plans to support additional back-ends.

SPERIMENTAZIONE di diverse soluzioni IN CORSO - II

E’ sufficiente dare un’occhiata agli abstracts sottomessi ad ACAT2022, rimanendo solo in ambito CMS, per capire che c’e’ vivace esplorazione e sperimentazione in corso…

Experience in SYCL/oneAPI for event recostruction at the CMS experiment

Felice Pantaleo, Andrea Bocci, Aurora Perego, Tony di Pilato, Wahid Redjeb (poster, abstract on CINCO)

The CMS software framework (CMSSW) has been recently extended to perform part of the physics reconstruction with NVIDIA GPUs. To avoid writing a different implementations of the code for each backend the decision was to use a performance portability library and so Alpaka has been chosen as the solution for Run-3.
In the meantime different studies have been performed to test the track recostruction and clustering algorithms on different back-ends like CUDA and Alpaka.

With the idea of exploring new solutions INTEL GPUs have been considered as a new possible back-end and their implementation is currently under development.
This is achieved using SYCL, that is a cross-platform abstraction C++ programming model for heterogeneous computing. It allows developers to reuse code across different hardwares and also perform custom tuning for a specific accelerator. The SYCL implementation used is the Data Parallel C++ library (DPC++) in the Intel oneAPI Toolkit.

In this work, we will present the performance of physics reconstruction algorithms on different hardware. Strengths and weaknesses of this heterogenous programming model will also be presented.

SPERIMENTAZIONE di diverse soluzioni IN CORSO - III

… a proposito di ***local & remote offloading*** …

Data transfer to remote GPUs over high performance networks

Ali Marafi, Andrea Bocci (poster, abstract on CINCO)

In the past years the CMS software framework (CMSSW) has been extended to offload part of the physics reconstruction to NVIDIA GPUs. This can achieve a higher computational efficiency, but it adds extra complexity to the design of dedicated data centres and the use of opportunistic resources, like HPC centres. A possible solution to increase the flexibility of heterogeneous clusters is to offload part of the computations to GPUs installed in external, dedicated nodes.

Our studies on this topic have been able to achieve high-throughput, low-latency data transfers to and from a remote NVIDIA GPU across Mellanox NICs, using the Remote Direct Memory Access (RDMA) technology to access the GPU memory without involving either nodes' operating system.

In this work we present our approach based on the Open MPI framework, and compare the performance of data transfers of local and remote GPUs from different generations, using different communication libraries and network protocols.

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-2

Task-1 Alla fine della **Fase-1** vanno poi identificate ***solutions for handling user support*** & ***user fora*** da implementare nella **Fase-2** che e' la **fase realizzativa** :

> a realization phase, in which the services and the support systems are put into place, at least in alpha/beta phase. These include the testbeds to be used for benchmarking of scientific and industrial solutions, the user support system, ~~the training opportunities~~. (ne parlo dopo)

… con la **milestone** **M22-M26: testbeds ready for users (initial handshake); user support in place**

Dopodiche' **i test devono essere eseguiti** …diciamo da M22 ad M30 (testbed predisposti fra M15 ed M22)
… se le Fasi 3 (validazione) e 4 (wrap-up & disseminazione) occupano gli ultimi ~ 6 mesi

A tal fine serve aver individuato, alla fine della Fase-1, **anche** degli ***use cases*** !

Il *test* per ogni *use case* dovrebbe consistere nell'attuare il porting su piattaforma eterogenea e confrontare le perfomance rispetto all'applicazione iniziale (con soluzione "non HPC"), cioe'...
accelerators vs standard CPUs (ma anche **ML vs standard algorithms**)
[chiamiamolo ***performances gain assessment***]

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-2

Dunque … serve aver individuato alla fine della prima fase (entro il M15) degli ***use cases*** !

D: ma quali?
R: ragionevolmente ci si immagina un **mix di casi di test** (difficile pensare a dei numeri adesso…):

a) **casi "core" scientifici tipici degli ambiti scientifici coinvolti in WP2.1,WP2.2 e WP2.3 derivanti dalle relative fasi realizzative**
[almeno due casi per ambito sarebbe ragionevole avere…]

b) **casi innovativi accademico/scientifici in ambiti scientifici diversi** (p.es. Spoke3-5)?
[un paio di casi? o almeno uno …] ("cross-spoke"… difficile ma interessante !?)

c) **casi innovativi applicativi in ambito industriale**
(da individuare fra i 15 partners aziendali; possibile coinvolgere anche network partners?)
[un paio di casi? o almeno uno …] (questo sarebbe il campo di azione degli *innovation grants*!)
- le aziende dovrebbero offrire le loro infrastrutture e piattaforme (per test consistenti: *data are data*)

Commento generale: qui servirebbe una forte sinergia con WP2.5 (*scale-up*) anche per supportare gli use cases dei ***nearby scientific domains*** e delle ***academic and industrial realities*** (partners in primis), ed anche con WP2.6 con particolare riferimento ai dati dalla costellazione dei satelliti Mirror Copernicus.

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-2

Quanto delineato sembra coerente con quanto si legge sempre nel documento principale (Xc):

In addition, the Spoke activities aim at demonstrating that the same set of solutions are of value in domains outside the bounds of basic science and of the specific research use cases. We plan to address at least two specific situations:

- the handling of data and processing in the context of the Space Economy Italian Strategy, including the handling and processing of data from the Mirror Copernicus program, by fostering the conditions to enable radically innovative services;
- the seeding of similar solutions in the productive context, with industry-research shared testbeds and proofs of concept; this will cover the productive domains in which data and processing resources need to be geographically dispersed, need a secure and granular solution for the data access, and experience computational problems with an unsustainable predicted future scaling (due to cost, efficiency, or performance).

Sulla dichiarata facilita' dell'operazione che si ha in mente faro' dei commenti fra qualche slide:

The solutions and tools implemented during the project will be easily extendable to other scientific domains of the Centre and to the industrial partners in the Spoke;

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-2

D: Chi sono gli attori principali/***dedicati*** di questa Fase-2 (e della successiva fase di validazione) ?

R: Ci sono le "open calls" e gli "innovation grants" [anche i co-funded PhD grants ?] :

The Spoke 2 activities need a number of support and ancillary services, like a web portal, a ticketing / support system, and help in organizing activities like benchmarking and training. We intend to use a part of the "Open Calls" to this purpose, selecting professionalities from companies or academic institutions with experience on the subjects from previous projects.

The largest fraction of "Open Calls", still, is expected to be dedicated directly to academic institutions. This will allow it to extend the project beyond the direct involvement of partners and will allow it to operate in co-design and co-development mode. The current plan is to devote 11 open calls (~200kEur each), or almost 4 per scientific WP.

A summary of the Open Calls we aim to open, very early in the project (within month 6) is:

- 11 open calls for co-design and co-development, explicitly directed to academic partners; almost 4 per each of the scientific Work Packages 1, 2, 3; qualcuno per i test degli *use cases*?
- 2 open calls for engineering and development of domain scientific services;
- 1 open call for Spoke services, like a web portal, ticketing and support system;
- 1 open call for helping organizing training activities;
- 2 open calls for support to benchmarking and testing activities.

⇨ per WP 2.4-2.6 bastano?

The last 6 open calls are tentatively directed towards private parties, but we expect a possible collaboration also with academic partners with a proven experience.

The testbed and benchmarking ("validation") phase, in all WPs, will be executed partially via the "innovation grants" available via the project, in which the solutions developed in alpha/beta level will be tested together with industrial partners, on hardware either provided by the CN, or acquired via the same grants.

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-1/Fase-3&4

Task-1 La **Fase-3** prevede la validazione (benchmarking) delle soluzioni sotto test, mentre nella **Fase-4** i risultati vengono raccolti e riportati in un White Paper che serve per disseminare i risultati e la loro documentazione:

- a validation phase, in which experience on the supported services and codes are reported, to be used as a touch base before the end of the project.
- a wrap-up phase, in which results are reported for executed activities, and are disseminated via white papers for future and external use cases.

… con la **milestone** **M25-M36: results from testbed and benchmarking activities; final report on technologies and support system; white paper for use cases external to the CN**

DESCRIZIONE del WP2.4 (una delle possibili letture …) - Task-2 / accross phases

Task-2 **Competence** and **training** center for heterogeneous computing

Un sistema di competenze (necessariamente distribuito) messo a sistema

s'intende quanto detto prima:
heterogeneous code and algorithms on modern architectures

l'addestramento e la condivisione e trasmissione delle conoscenze e competenze e' presente in tutte le fasi
(… e' la *mission* delle universita' ed in parte degli EPR) :

moreover, the personnel trained within the Centre will help to spread and boost the application of HPC methodologies to Italian academic and industrial fields, for a comprehensive advancement of the Italian system.

… con le **milestone** **M22-M26:** ***first training opportunity*** (da anticipare a prima di M22?)

M25-M36 include: ***final report on training***

Nota: disponibile una "open call" dedicata per l'organizzazione delle attivita'

D: si potrebbe pensare ad una specie di "master di livello I e/o II" *distribuito*? (formalizzazione?, target?,...)

… EASILY EXTENDABLE TO OTHER SCIENTIFIC DOMAIN … : EASILY ?????

Main Activities: porting of applications to GPUs and heterogeneous architectures (e.g., scalability of scientific codes and applications on GPU/CPU many-cores clusters, local and remote offloading, mission-critical algorithms on FPGAs, ...). The solutions and tools implemented during the project will be easily extendable to other scientific domains of the Centre and to the industrial partners in the Spoke; moreover, the personnel trained within the Centre will help to spread and boost the application of HPC methodologies to Italian academic and industrial fields, for a comprehensive advancement of the Italian system.

Preoccupazione/criticita': ***easily*** e' una parola **MOLTO** ottimistica

⇨ Vi tediero' per qualche minuto - se c'e' tempo - con le prossime slide per riportarvi la mia esperienza in piccolo (che potremmo considerare un "caso di scuola pre-PNRR" *in miniatura*).

Ma prima qualche appunto generale per la discussione di questo aspetto (next slide)

… *EASILY EXTENDABLE TO OTHER SCIENTIFIC DOMAIN* … : **EASILY ?????**

Appunto generale:

- l'estensione ad altri domini applicativi di interesse industriale e' probabilmente la parte piu' delicata

  (cioe' il resto e' piu' nelle corde e nell'esperienza dei partecipanti/proponenti accademici)

- dipendera' molto dalla capacita' e disponibilita' aziendale

  (p.es. expertise interna ed infrastruttura da coinvolgere sinergicamente nell'ambito del progetto)

  - un'azienda che pensa semplicemente di attingere al "parco" di personale addestrato durante il progetto seguira' un approccio piu' comodo ma non ottimale

- la parte di *training* (nel *competence and training centre for heterogeneous computing*) sara' cruciale:

  il personale addestrato (a partire da M18 ?) **opererebbe da "facilitatore" (*spread* & *boost*)**

  interloquendo sinergisticamente con le 15 aziende previste aderire all'hub del CN:

  - D: il personale previsto sara' numericamente sufficiente allo scopo?
  - Non e' scontato o facile raggiungere risultati concreti (*comprehensive advancement*) al M36.

    D: Cosa intendiamo per risultati concreti? TRL5? TRL6? Tech Readiness Level da capire…

CASO DI SCUOLA PRE-PNRR - I

A Sett.2017 presento all'Anvur (selezionatore) un progetto per un …
… ***Dottorato Innovativo a Caratterizzazione*** Industriale, finanziato dal MIUR
… su fondi PON R&I 2014-20 per le Regioni dell'Obiettivo Convergenza Italia
Titolo: ***Deep Neural Networks or the analysis of big data from EO with GPU onboard satellite***

Partner estero: **CERN (CMG group)**

- nello specifico coloro che sviluppano la ricostruzione per l'HGCAL di CMS (Phase-2)

Partner Industriale regionale: ***Planetek Italia***

- azienda di dimensioni medio-piccole (60 staff + altrettanti collaboratori),
  leader regionale nel settore dell'EO e servizi connessi (*),
  membro del DTA-Puglia (Distretto Tecnologico Aereospaziale Pugliese)

L'idea in 2 parole: il dottorando si forma al CERN da M6 a M18 dove impara ad usare le reti neurali ed in particolare quelle CNN-based per l'analisi di immagini (in sviluppo per HGCAL) e poi passa in azienda (da M19 a M30) dove "porta" le conoscenze acquisite e le riadatta ad un'applicazione di interesse industriale.

(*) https://www.planetek.it/azienda/chi_siamo/profilo

Project overview:

- Management of huge amounts of data acquired from Earth Observation missions with new technologies and strategies
  - **Heterogeneous architectures (GPU + CPU)**
  - **Deep learning algorithms**
- Activity at CERN (12 months): CMS-HGCAL software reconstruction
  - **GPU-friendly clustering algorithm**
  - **Deep Neural Network for particle identification and energy regression**
- Activity at Planetek Italia (12 months): Change Detection with Sentinel-2 images
  - **Segmentation and classification of urban changes**
  - **Benchmark measurements on a low power GPU**

HW platform:

Deep learning algorithms are ***extremely costly*** in terms of time, especially for the training phase. Therefore, hybrid architectures are largely exploited for GPU-friendly deep learning models (i.e. CNNs).

GPUs used at ReCaS-Bari HPC*:

- **NVIDIA Tesla K40**
- **NVIDIA Tesla P100** (thanks to INFN e MAECI PGR00970/2018)

(DATA CENTER GPUs / “at ground station”)

GPU provided by Planetek Italia:

- **NVIDIA Jetson Xavier** (LOW POWER CONSUMPTION GPU / “onboard satellite”)

HW platform:

Deep learning algorithms are ***extremely costly*** in terms of time, especially for the training phase. Therefore, hybrid architectures are largely exploited for GPU-friendly deep learning models (i.e. CNNs).

GPUs used at ReCaS-Bari HPC*:

- **NVIDIA Tesla K40**
- **NVIDIA Tesla P100** (thanks to INFN e MAECI PGR00970/2018)

(DATA CENTER GPUs / “at ground station”)

GPU provided by Planetek Italia:

- **NVIDIA Jetson Xavier** (LOW POWER CONSUMPTION GPU / “onboard satellite”)

CASO DI SCUOLA PRE-PNRR - V

Pre-changes

Post-changes

Data acquisition

Satellite

Data transmission

Ground station

Segmentation

Output

Binary change map

Classification

Output

Decision making process

Discard

Download request

CASO DI SCUOLA PRE-PNRR - VI

Prodotti del progetto: risultati interessanti esposti in

1) Tesi di Dottorato, 2) Pubblicazione scientifica

https://www.mdpi.com/2072-4292/13/20/4083

remote sensing

MDPI

Article

Deep Learning Approaches to Earth Observation Change Detection

Antonio Di Pilato [1,*,†], Nicolò Taggio [2], Alexis Pompili [1], Michele Iacobellis [2], Adriano Di Florio [3], Davide Passarelli [2] and Sergio Samarelli [2]

[1] Dipartimento Interateneo di Fisica, Università Degli Studi di Bari Aldo Moro, 70126 Bari, Italy; alexis.pompili@uniba.it
[2] Planetek Italia s.r.l, 70132 Bari, Italy; taggio@planetek.it (N.T.); iacobellis@planetek.it (M.I.); passarelli@planetek.it (D.P.); samarelli@planetek.it (S.S.)
[3] Dipartimento Interateneo di Fisica, Politecnico di Bari, 70126 Bari, Italy; adriano.diflorio@poliba.it
* Correspondence: antonio.dipilato@uniba.it
† Current affiliation: Center for Advanced Systems Understanding (CASUS), 02826 Görlitz, Germany.

Abstract: The interest in change detection in the field of remote sensing has increased in the last few years. Searching for changes in satellite images has many useful applications, ranging from land cover and land use analysis to anomaly detection. In particular, urban change detection provides an efficient tool to study urban spread and growth through several years of observation. At the same time, change detection is often a computationally challenging and time-consuming task; therefore, a standard approach with manual detection of the elements of interest by experts in the domain of Earth Observation needs to be replaced by innovative methods that can guarantee optimal results with unquestionable value and within reasonable time. In this paper, we present two different approaches to change detection (semantic segmentation and classification) that both exploit convolutional neural networks to address these particular needs, which can be further refined and used in post-processing workflows for a large variety of applications.

Keywords: change detection; convolutional neural network; earth observation; deep learning; Sentinel-2

check for updates

Citation: Di Pilato, A.; Taggio, N.; Pompili, A.; Iacobellis, M.; Di Florio, A.; Passarelli, D.; Samarelli, S. Deep Learning Approaches to Earth Observation Change Detection. *Remote Sens.* **2021**, *13*, 4083. https://doi.org/10.3390/rs13204083

Tutto questo *prima* … del progetto C-SCALE H2020 (2021-23) di cui ReCas-Bari e' partner,

mettendo a disposizione … non solo risorse (computing & storage)

… ma anche know-how e servizi evoluti ad-hoc da istanziare su risorse cloud distribuite ed eterogenee

C-SCALE WILL PUBLISH THE FOLLOWING SERVICES IN THE EOSC PORTAL:

Access

Access to a large C-SCALE EO data archive.

Integration

C-SCALE compute services integrated with the EO Data archive.

Analytic tools

Set of analytics platforms and tools that can be deployed on top of the C-SCALE EO data archive and compute services.

https://c-scale.eu/

UN CASO di ECOSISTEMA di utenza di Data Center

Planetek Italia fa parte dell'ecosistema di utenti attuali del ReCas-Bari Data Center:

UN CASO di ECOSISTEMA TEMATICO (settore aereospaziale pugliese) fuori dal contesto PNRR

RIPARTI (*)
assegni di RIcerca per riPARTire con le Imprese

(*) vedi anche https://riparti.regione.puglia.it/

Sei un laureato e vorresti collaborare nel settore aerospaziale pugliese? Abbiamo 13 opportunità per te.

Planetek Italia partecipa al Bando "**RIPARTI-assegni di RIcerca per riPARTire con le Imprese**", l'iniziativa della Regione Puglia con cui si finanziano **assegni di ricerca professionalizzanti per nuovi ricercatori e il loro inserimento nel sistema produttivo regionale** per rispondere ai fabbisogni di innovazione delle imprese.

Il bando mira a rilanciare il ruolo che la ricerca riveste per l'innovazione e lo sviluppo del tessuto socioeconomico e industriale, coerentemente con la Smart Specialization Strategy della Regione Puglia.

Obiettivo: Unire Università ed Enti di ricerca pugliesi, giovani risorse umane di eccellenza e un tessuto di imprese dinamiche. L'obiettivo è promuovere la ricerca sviluppata dagli assegnisti e applicata all'industria, negli ambiti considerati prioritari dall'Unione Europea e al servizio delle filiere produttive regionali: su questi temi, le Università pubbliche e private aventi sede legale in Puglia e degli Enti Pubblici di Ricerca, in collaborazione con un'impresa privata, sono state chiamate a candidare i propri progetti di ricerca.

UN CASO di ECOSISTEMA TEMATICO REGIONALE fuori dal contesto PNRR

13 progetti di ricerca con Planetek Italia (*)

Planetek Italia è assegnataria di n.13 progetti di ricerca in vari ambiti che saranno via via pubblicati dai singoli enti proponenti, ai quali i ricercatori potranno aderire facendo domanda. Gli enti sono: CNR (Consiglio Nazionale delle Ricerche) (n.3 Borse), INdAM (Istituto Nazionale di Alta Matematica "Francesco Severi") (n.1 Borsa), Università degli Studi "Aldo Moro" di Bari (n.2 Borse), INFN (Istituto Nazionale di Fisica Nucleare) (n.2 Borse), Politecnico di Bari (n.2 Borse), Università del Salento (n.2 Borse), Libera Università Mediterranea "Giuseppe Degennaro" (n.1 Borsa).

Elenco dei 13 progetti ammessi e scadenze:

- **UNIBA:** Sistema INntegrato di Tecnologia UAV, GNSS e cloud computing per il Monitoraggio del Rischio Idrogeologico (SIMRI)
- **UNIBA:** SPIKEO Spiking Neural Networks for Satellite OnBoard Processing
- **POLIBA:** Non-Fungible Token for Earth Observations (NFT4EO) **(Scad. 22 giugno pagina info e modulistica POLIBA - Scarica Descrizione Progetto)**
- **POLIBA:** Un sistema automatizzato per il monitoraggio della sicurezza di ponti esistenti basato sull'uso integrato di dati satellitari e informazioni geo-spaziali e visive **(Scad. 22 giugno pagina info e modulistica POLIBA - Scarica Descrizione Progetto)**
- **UNISALENTO:** MITIGA - MonIToraggio e mitIGazione delle isole di calore urbAne **(Scad. 24 giugno - Codice GEO/12 scarica bando e pagina info UNISALENTO - Scarica Descrizione Progetto)**
- **UNISALENTO:** Machine Learning per Space Weather **(Scad. 24 giugno - Codice ING-INF/05 pagina info e bando UNISALENTO - Scarica Descrizione Progetto)**
- **LUM:** Data-driven Marketing for Apulian Aerospace SMEs - DaMaS **(Scad. 21 giugno - informazioni e modulo adesione. - Scarica Descrizione Progetto)**
- **CNR:** Utilizzo di intelligenza artificiale e dati satellitari per il monitoraggio dell'instabilità del territorio
- **CNR:** Earth Observation for Carbon Cycle Studies (EO4CCS)
- **CNR:** Irrigation Management Tool + (IRMAT+)
- **INDAM:** MATHEO: Definizione del modello ottimo di proiezione geografica di immagini satellitari **(Scad. 28 giugno - scarica bando, allegato bando e modulo adesione - Scarica Descrizione Progetto)**
- **INFN:** Quantum-enhanced hyperspectral imaging for Earth Observation
- **INFN:** Analisi Innovative per il monitoraggio dello stress idrico e gestione della risorsa idrica in agricoltura

Potremmo/dovremmo (?) - nello spoke 2- trovare sinergie anche con iniziative simili (tematiche e/o regionali "sparse sul territorio")

(*) https://www.planetek.it/news_eventi/archivio_news/2022/06/bando_riparti_assegni_di_ricerca_in_partenariato_con_le_imprese

To-do-list (si parte dal censimento/ricognizione)

Step-0: "Censimento" da fare (tramite ***call interna centralizzata***; inizialmente tramite gli *institute contacts*):

- identificare una comunita' di interesse tecnologico/scientifico
- … partendo dai ricercatori coinvolti ufficialmente ("esposizione ministeriale/cofinanziamento")
- … ma soprattutto ricercatori e tecnologi interessati scientificamente
  (a partire dai Data Center esistenti in Italia)

Step-1: "Censimento" da fare:

- di interessi ed expertise in ambiti aziendali

Step-2: Individuare persone di contatto all'interno del pool di expertise/interest identificato (step 0-1)

Nota: Step 0-2: entro inizi settembre

Step-3: Kick-off meeting del WP2.4 (ragionevolmente entro fine settembre)

Backup

Definizione di TRL (un indice da contestualizzare)

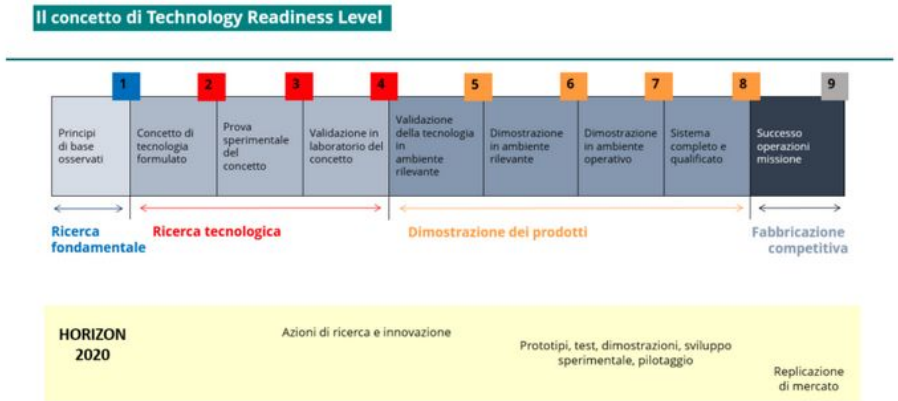
IL *Technology Readiness Level – TRL* è un indice che permette di misurare lo stato di **maturità delle tecnologie**.

Identifica le fasi che partono dalla concettualizzazione di una nuova tecnologia sino alla sua possibile introduzione nel mercato.

Il TRL identifica le fasi di ricerca e sviluppo, di riproduzione in laboratorio, di *testing*, di prototipizzazione e di *piloting* sino al monitoraggio di fasi sperimentali produttive operative.

Il suo valore varia in **una scala da 1 a 9.**

Ai fini del **Programma Horizon 2020** sono stati in particolare individuati 9 Livelli:
TRL 1 = osservazione dei principi fondamentali
TRL 2 = formulazione di un concept tecnologico
TRL 3 = proof of concept sperimentale
TRL 4 = validazione tecnologica in ambiente di laboratorio
TRL 5 = validazione tecnologica in ambito industriale
TRL 6 = dimostrazione della tecnologia in ambito industriale
TRL 7 = dimostrazione del prototipo in ambiente operativo reale
TRL 8 = definizione e qualificazione completa del sistema
TRL 9 = dimostrazione completa del sistema in ambiente operativo reale (prova funzionale con tecnologie abilitanti ed applicazione al settore industriale specifico



Chapter 3 illustrates the study of the performance of Convolutional Neural Networks, using TensorFlow, for image segmentation and classification applied on Sentinel-2 [17] data. More specifically, the change detection task on urban scenes is studied, exploiting an existing training data set generated from Sentinel-2 images. A CNN with a single final score is designed for scene classification, for preliminary analyses, to check whether a significant change occurred in the area, while an encoder-decoder structured CNN is used for segmentation, mapping changes pixel by pixel for a precise localization. Afterwards, the trained encoder-decoder model is also tested on Rome-Fiumicino airport area as a practical application of the novel method, together with additional post processing algorithms.

Finally, in Chapter 4 performance studies are carried out on a low power portable GPU under test in comparison with GPUs in a typical computing data center used as benchmark; change detection algorithms has been developed and studied exploiting the resources of the High Performance Computing (HPC) cluster hosted at the ReCaS Data Center in Bari [18]. The aim of these studies is to verify that such low power device can fit the energy constraints typical of a satellite scenario while preserving enough high computing performance. In addition, the efficiency of the inference accelerator NVIDIA TensorRT [19] on satellite multi-spectral data is also proven in a preliminary test, thus showing very promising performance in terms of speed and throughput.

Esempio di Servizi di un Data Center da mettere in rete

Software-as-a-Service (SaaS) e' un servizio di cloud computing che offre agli utenti finali - tramite un browser web - un'applicazione cloud, munita di piattaforme e dell'infrastruttura IT che la supportano.