LOw Power Heterogeneous Architecture
for NExt Generation of SmaRt Infrastructure and Platforms
in Industrial and Societal Applications

Innovation Potential of OPERA

Platform 2

Co-funded by the Horizon 2020
Framework Programme of the European Union

DELIVERABLE NUMBER	D2.6
DELIVERABLE TITLE	Innovation Potential of OPERA Platform 2
RESPONSIBLE AUTHOR	Dirk Harryvan (Certios)

GRANT AGREEMENT N.	688386
PROJECT REF. NO	H2020- 688386
PROJECT ACRONYM	OPERA
PROJECT FULL NAME	LOw Power Heterogeneous Architecture for Next Generation of SmaRt Infrastructure and Platform in Industrial and Societal Applications
STARTING DATE (DUR.)	01/12/2015
ENDING DATE	30/11/2018
PROJECT WEBSITE	www.operaproject.eu
WORKPACKAGE N. | TITLE	WP2 | Low Power Computing Requirements and Innovation Engineering
WORKPACKAGE LEADER	ISMB
DELIVERABLE N. | TITLE	D2.6 | Innovation Potential of OPERA Platform 2
RESPONSIBLE AUTHOR	Dirk Harryvan (Certios)
DATE OF DELIVERY (CONTRACTUAL)	31/08/2017 (M21)
DATE OF DELIVERY (SUBMITTED)	29/09/2017 (M22)
VERSION | STATUS	V1.0 | Final
NATURE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Dirk Harryvan (Certios); Frank Verhagen (Certios), Joel Nider (IBM); Richard Chamberlain (Nallatech); Idan Yaniv, (Technion); Gallig Renaud (HPE) and Cristian Griua (HPE); Albert Scionti (ISMB); Simone Ciccia (ISMB); Giulio Urlini (ST).

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
v 0.1	Table of Content (TOC)	14/06/2017	Frank Verhagen, Dirk Harryvan (Certios)
v 0.2	Preparation of adding content	02/08/2017	Frank Verhagen (Certios)
V 0.4	Contributions from partners	22/08/2017	Joel Nider (IBM) Richard Chamberlain (Nallatech) Idan Yaniv (Technion) Albert Scionti (ISMB) Simone Ciccia (ISMB)
V 0.5	Contributions from partners	15/09/2017	Gallig Renaud (HPE)
V 0.6	Contributions from partners	18/09/2017	Giulio Urlini (ST)
V 0.7	Contributions from partners	19/09/2017	Richard Chamberlain (Nallatech)
V 0.8	Contributions from partners	20/09/2017	Cristian Gruia (HPE)
V0.9	Finalization for first review	21/09/2017	Frank Verhagen, Dirk Harryvan (Certios)
V0.92	Final version after review by Isere	27/09/2017	Jean-Christophe Maisonobe (Isere), Frank Verhagen, Dirk Harryvan (Certios)
V0.93	Final version after review by CSI	28/09/2017	Luca Scanavino (CSI), Frank Verhagen (Certios)
V 1.0	Final version delivered	29/09/2017	Dirk Harryvan (Certios)

PARTICIPANTS		CONTACT
STMICROELECTRONICS SRL		Giulio Urlini Email: Giulio.urlini@st.com
IBM ISRAEL SCIENCE AND TECHNOLOGY LTD		Joel Nider Email: joeln@il.ibm.com
HEWLETT PACKARD CENTRE DE COMPETENCES (FRANCE)		Gallig Renaud Email: gallig.renaud@hpe.com
NALLATECH LTD		Craig Petrie Email: Richard.Chamberlain@molex.com
ISTITUTO SUPERIORE MARIO BOELLA		Olivier Terzo Email: terzo@ismb.it
TECHNION ISRAEL INSTITUTE OF TECHNOLOGY		Dan Tsafrir Email: dan@cs.technion.ac.il
CSI PIEMONTE		Vittorio Vallero Email: Vittorio.vallero@csi.it
NEAVIA TECHNOLOGIES		Stéphane Gervais Email: s.gervais@lacroix.fr
CERIOS GREEN BV		Frank Verhagen Email: frank.verhagen@certios.nl
TESEO SPA		Stefano Serra Email: s.serra@teseo.clemessy.com
DEPARTEMENT DE L'ISERE		Olivier Latouille Email: olivier.latouille@isere.fr

ACRONYMS LIST

Acronym	Description
ADAS	Advanced Driver Assistance Systems
API	Application Programming Interface
ASID	Address Space Identifier
BSP	Board Support Package
CAPI	Coherent Accelerator Processor Interface
CDNN	CEVA Deep Neural Network
CNN	Convolutional Neural Network
CPS	Cyber-Physical System
CPU	Central Processing Unit
CRIU	Checkpoint/Restore In Userspace
D2.5	Deliverable 2.5
DiaB	Datacenter in a Box, a hardware product of the OPERA project
DSP	Digital signal processing
EC	European Commission
FD-SOI	Fully Depleted Silicon On Insulator
FPGA	Field Programmable Gate Array
GPU	Graphics Processing Unit
GPGPU	General-Purpose Graphics Processing Unit
HPC	High Performance Computing
HW/SW	Hardware/Software
I/O	Input/Output
IoT	Internet of Things
ISA	Instruction Set Architecture
ISP	Image Signal Processing
LLVM	Low Level Virtual Machine
PCIe	Peripheral Component Interconnect Express
PUE	Power Usage Effectiveness
PWC	Page Walk Cache
RDMA	Remote Direct Memory Access
RF	Radio-frequency

RF MEMS	Radio Frequency Microelectromechanical system
SoC	System on Chip
SOTA	State of the Art
SWaP	Size Weight and Power
T2.3.1	Task 2.3.1
TCO	Total Cost of Ownership
TMC	Traffic Management Center
ToC	Table of Contents
TOSCA	Topology and Orchestration Specification for Cloud Applications
TPU	Tensor Processing Unit
ULP	Ultra-Low Power
WP	Work Package

Table 1 Acronyms List

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

1.1 POSITION OF THE DELIVERABLE IN THE WHOLE PROJECT CONTEXT

This project's deliverable D2.6 researches the latest innovation potential of the OPERA project and the products realized in the context of this project and the deliverable D2.5. There will be 4 iterations; this document is the result of the second iteration. The documentation of the innovation potential is part of the WP2. The relationship of this work package to the other WP's of the OPERA project, can be visualized in the following graph:

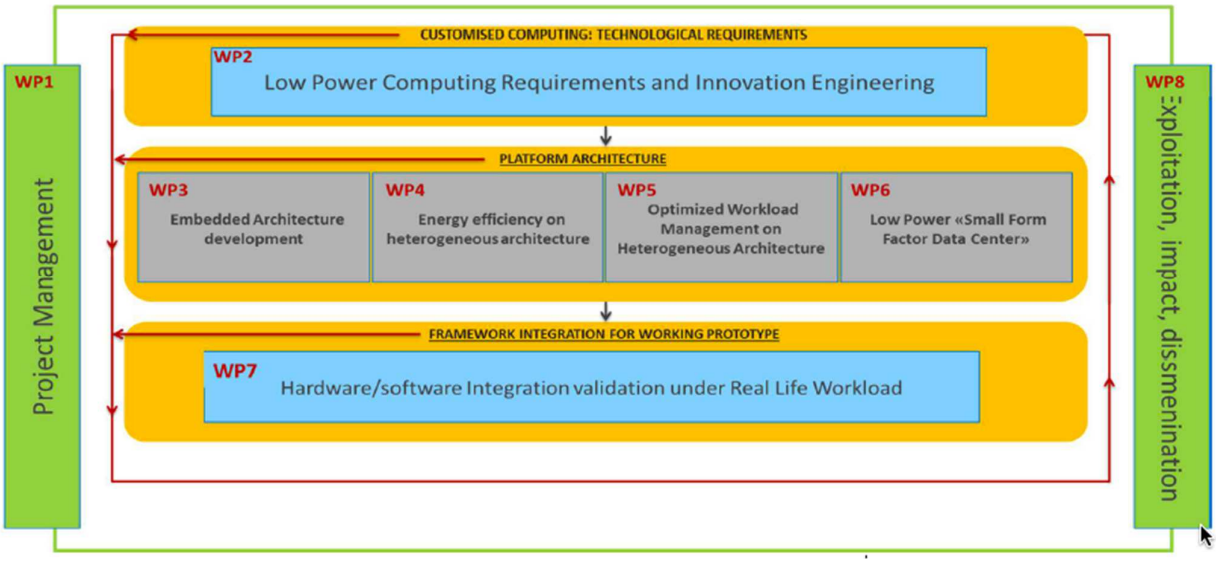

Figure 1 Linked WP's

1.2 DESCRIPTION OF THE DELIVERABLE

This deliverable D2.6 is delivered as an addendum to the D2.5 Innovation Potential deliverable. Presenting D2.6 in this way avoids repeating text of D2.5. This addendum is the second of the total of 4 iterations of the D2.5 deliverable. The four iterations have been spread in time, over the duration of the OPERA project. Where the D2.5 deliverable had the original idea of delivering the OPERA project Potential in M10 of the project, the following iterations have been added after M10, in order to enable addendums (D2.6 and D2.7) to D2.5, to be integrated as a final document on the OPERA Innovation Potential, D2.8 in M28. The decision to deliver 4 (intermediate) reports is consistent with the idea of OPERA as a research and development project; in the progress of the project, more and more is known about the innovative potential of OPERA products.

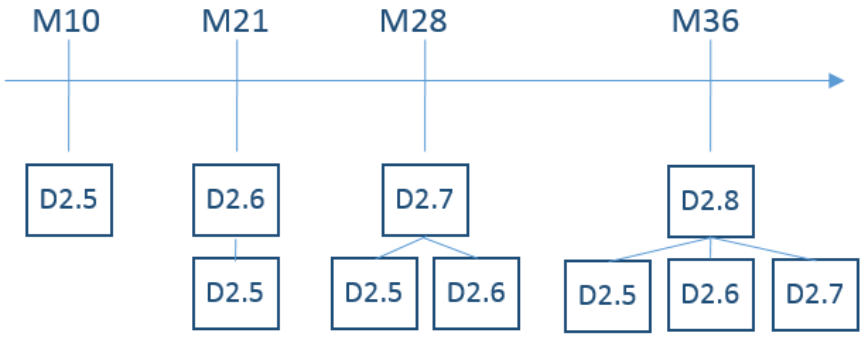

Figure 2 Timeline iterations D2.5, 2.6, 2.7 and 2.8

1.2.1 Positioning the iterations

1.2.1.1 D2.6 (M21)

In D2.6 there will strongly refer to the D2.5 deliverable that was submitted in M10 and updated and resubmitted in M18. D2.6 has a focus on the partners and how each of the partners who are actively innovating, is developing its technology in the context of the OPERA project. In addition to this, the section 4, identifying of Potential Markets and section 5, Market Analysis have been enriched; identifying extra new entrants in the competitive arena for OPERA products.

1.2.1.2 D2.7 (M28)

At the time of the expected delivery of the next D2.7 addendum, there will be more concrete ideas of the market potential for each of the partners involved. Thanks to the progress of the OPERA products, new product market combinations are likely to be identified. By then, the organization's market and product development departments will have been involved. In D2.7 will focus on the way of translating the innovations, the new technology, to the market and creating opportunities for OPERA partners, for each of the innovations and involved partners.

1.2.1.3 D2.8 (M36)

The final document D2.8 will be an integration of the latest insights, the first deliverable D2.5 and the 2 intermediate deliverables, the addendums D2.6 and D2.7. Deliverable D2.8 will be an independent report that can be read independently from the earlier iterations.

1.3 LIST OF ACTIONS AND ROLES

	LIST OF ACTIONS										
ACTIVITIES LIST AND PARTNERS ROLES	**CERTIOS**	CSI	HPE	IBM	**ISMB**	LD38	NALLATECH	NEAVIA	**ST**	TECHNION	TESEO
Summary of the initial requirement	P										
Components to be integrated	P		P	P	P		P		R		
Integration output: first iteration (D2.5)	P		P	P	P		P	R	R		R
Integration output: second iteration (D2.6)	P	R	P	P	P		P		P	I	I
Integration output: third iteration (D2.7)	P	I	P	P	P		P		P	I	P
Integration output: fourth iteration (D2.8)	P	I	P	P	P	R	P	I	P	I	P

Table 2 List of actions and roles

- P = Participating (includes I & R)
- I = Input delivery (Includes R)
- R = review
- **Bold** → assigned in project proposal to contribute to this the task

As shown in Table 2, especially the privately held organizations are expected to participate, give this deliverable input and review the document D2.6 itself. The further the project will progress, the more organizations will be able to contribute in detail of what to expect from OPERA's innovations.

TABLE OF CONTENTS

2 INTRODUCTION

2.1 OBJECTIVE OF THIS DELIVERABLE

The objective of this report is to present the innovation potential of the OPERA products and services. This report is an update of D2.5 Innovation potential of OPERA – Platform 1 which was submitted in M10 and resubmitted in M18. This report, deliverable D2.6, is a next iteration in a sequence of 4 iterations in total:

- D2.5 Innovation potential of OPERA – Platform 1 (M10)
- D2.6 Innovation potential of OPERA – Platform 2 (M21)
- D2.7 Innovation potential of OPERA – Platform 3 (M28)
- D2.8 Innovation potential of OPERA – Platform 4 (M36).

In order to prevent producing a document that is growing 'bigger' with each iteration, and in which it is increasingly difficult for the reader to find the latest input to the document, we have chosen to leave D2.5 as the base document and only to present new insights (new since producing the former iteration) in the newer iterations. The Table of Contents (ToC) in the later iterations will only focus on the 'dynamic' parts of the content, i.e. the contents that are progressing over the duration of the project. The more 'static' parts of the project, like the

- Methodology (D2.5, section 3)
- Innovation Envisaged (D2.5, section 4)
- How to achieve the 9 objectives (D2.5, section 5)
- Full description of the OPERA products and services (D2.5, section 6)
- Description of Other H2020 Projects (D2.5, section 9)

will therefore not be repeated in the iterations D2.6, D2.7.

The objective of this iteration, D2.6, is to give some update of the dynamic parts of the D2.5 and see what the innovation for the different partners is going to be like (Section 3):

- Section 3 Innovation of OPERA Solutions (reference: D2.5, section 6)
- Section 4 Potential Markets (reference: D2.5, section 7)
- Section 5 Market Analysis (reference: D2.5, section 8)
- Section 6 Impact envisaged (reference: D2.5, section 10).

3 INNOVATION OF OPERA SOLUTIONS

3.1 PRODUCTS AND SERVICES

When we look at the objectives of OPERA and what objectives need to be reached in order to make the products of OPERA work (see Figure 3), we expect to develop the following technical solutions as described in detail in Section 6 of D2.5:

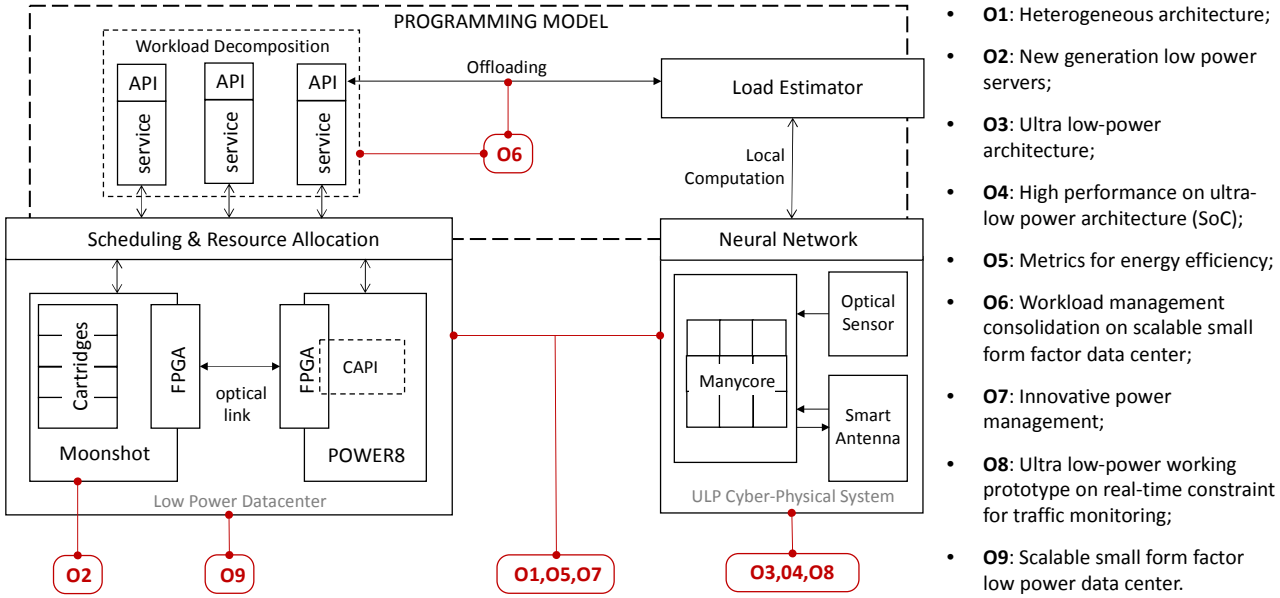

Figure 3 The OPERA main objectives mapped on the reference overall architecture

1. A FPGA card as application accelerator and high-speed interconnect, developed for OPERA also outside of the OPERA platform
2. Low-power heterogeneous data center architecture incorporating both HPE Moonshot and IBM POWER8
3. An ultra-low power autonomous traffic monitoring video sensor.
4. Extension of open source Cross ISA compiler to work with POWER8, enabling application migration over different CPU architectures
5. An energy aware workload management to maximize power efficiency of the heterogeneous architecture under different application loading
6. Smaller mobile datacenter solutions offering superior compute power per Watt in a small form factor for the mobile application
7. Extension of the TOSCA application description TOSCA is an open standard developed with the aim of ease the description and deployment of an application in a cloud environment.
8. Using CAPI (Coherent Accelerator Processor Interface) to attach a low latency interconnect for communication between servers
9. Virtual Memory Model, Developed for OPERA platform and applicable also outside of the OPERA platform
10. Incorporation of the reconfigurable antenna for low power wireless connectivity
11. Implementation of Redfish on the FPGA (PCIe) card
12. Coupling of multiple disparate FPGA boards.

3.2 TECHNOLOGY INVOLVED IN OPERA

To illustrate the innovation potential, it is necessary to illustrate the difference between the state of the art at the time of the start of the OPERA project. For the partners involved, we here go into the technology developments, the innovations, in specific areas of OPERA in the next sections of this paragraph. After indications that the state of the art (SOTA) has been surpassed, we will go into the market opportunities that these innovations may unfold in the next section, in chapter 4.

3.3 WHAT TECHNOLOGY ARE OUR PARTNERS WORKING ON?

The different partners are working both individually as well as in combinations to achieve the products and services listed in paragraph 3.1. The link between the technology and these products and services is given, in square brackets [ ], for each of the mentioned technologies using the numbering schema from paragraph 3.1. This linkage is given in order to better understand the role of the technologies that these partners are working on within the OPERA project.

3.3.1 HPE

HPE is involved with OPERA on the following technologies:

1. Moonshot servers: improving existing workload and density optimized heterogeneous servers [2,6].
2. RedFish interface: working on the addition of new schema for accelerators and power monitoring [2,5].

3.3.2 IBM

IBM is working on 3 main innovations in OPERA:

1. Post-copy migration [5]
2. Low latency interconnect [2,8]
3. Cross-ISA compiler technology [4,5].

3.3.3 ISMB

ISMB carried out applied research on designing and implementing a reconfigurable antenna aimed at pushing the wireless communication beyond its limits. Wireless communication capability is becoming a critical element in modern cyber-physical systems, thus efficient mechanisms to provide such capabilities are of worth. From this perspective, innovation potential of such system is given by:

- Reduced power consumed by the communication subsystem;
- Longer range then omni-directional antennas;
- Robustness against interferers.

This work contributes to the development of the ULP video sensor [3,10]. Next to this work, ISMB is also leading the research into the TOSCA extensions needed for the development of the energy aware workload manager. TOSCA describes the application resource needs and will be extended with “affinity” attributes that link it to the heterogeneity of the OPERA solution [2,5,7].

3.3.4 Nallatech

Nallatech is working on 2 main innovations in OPERA:

- Development of an SoC FPGA accelerator card with high speed serial interconnect [1,2,11,12].
- The optimisation of an Open Source code (MICMAC). This involves innovative approaches to partition code between x86 and FPGA, utilising high precision server power measurements as guidance [2,6]. More can be found in the WP7 deliverable D7.4, where the evolution of the Truck Use Case and how OPERA Project wants to take advantage by partitioning the code has been described.

3.3.5 ST

STMicroelectronics is working in the context of OPERA on two main pillars. The first one is about the design and improvement of hardware architecture for the implementation of heterogeneous SoCs (with general purpose processors placed side by side with accelerators for specific functions) for the improvement of performances of embedded systems in the video computing domain. The REISC5 and 6 processors, with the accelerators for CDNN (CEVA Deep Neural Network) processors are the most important examples.
The second direction is the development of the technology used to manufacture the architecture. The creation of SoCs in FD-SOI 28nm allows ST to realize ULP platform by design and with a ULP technology capable to reduce drastically the current leakage.
The innovation supported by OPERA activities are both in terms of new applications and refinement of the design, based on a strict collaboration with the partners expert in the video computing applications for the traffic monitoring domain, and the validation of the solutions obtained in a real environment.
This work contributes to the development of the ULP video sensor [3,10].

3.3.6 Technion

The Technion conducts a virtual memory research activity under OPERA, focused on hashed page tables. Like other computer-architecture works, the research results are based on simulation and will thus not manifest in actual OPERA systems. Still, the Technion research has some practical implications because hashed page tables are implemented in IBM Power and Intel Itanium platforms. Specifically, Technion studied the Intel Itanium virtual memory design, proposed several ways to optimize it, and showed that our carefully optimized hashed page tables outperform virtual memory implementation with radix, hierarchical page tables in existing x86-64 processors. In future work, Technion hopes to also examine the IBM Power architecture for similar improvements, this future work however will not be part of the OPERA project [9].

3.4 WHAT IS THE CURRENT STATE OF THE ART?

3.4.1 HPE

3.4.1.1 Moonshot

The first generation of Moonshot servers has been released in 2013. The initial goal of Moonshot was to bring optimized servers for defined workloads to the market. Most of the IT-services in large companies are always providing the dual socket system to their internal customers whatever the function is. It has been demonstrated[1] that leveraging optimized servers can dramatically reduce the price, the used space and the power usage over generic systems.

3.4.1.2 Redfish

Redfish is a growing standard created to replace insecure and unscalable management protocols such as IPMI. Redfish standard has been created in September 2014 by leading IT-companies such as Dell, Microsoft, Intel and HPE. This standard is growing fast due to its adoption by major software and hardware vendors, including IBM for the POWER8 server line. The fact that Redfish is using well known technologies (RESTful interface over HTTPS in JSON format based on OData v4) make it very compelling for any cloud deployment.

[1] https://www.hpe.com/h20195/v2/GetPDF.aspx/4AA5-4902ENW.pdf

3.4.2 IBM

3.4.2.1 Post-copy container migration

The current state of the art of post-copy container migration is non-existent. The standard tool used for container migration is CRIU (Checkpoint/Restore In Userspace), which until the start of the OPERA project did not support post-copy.

3.4.2.2 Low latency interconnect

The current state of the art of the low latency interconnect is non-existent. There are some technologies that may be fitting for this purpose, such as RDMA (Remote Direct Memory Access) but to our knowledge, no one has tried to use RDMA for this purpose so far.

3.4.2.3 Cross-ISA compiler

The current state of the art of cross-ISA compiler technology is what has been released by Virginia Tech as the "Popcorn Linux Cross-ISA LLVM (Low Level Virtual Machine) compiler". The current implementation supports ARMv8 and x86_64 instruction sets, but not the Power8 ISA, which is needed for the full OPERA solution.

3.4.3 ISMB

The current state of the art solutions for wireless communications employing reconfigurable antennas use digital beamforming. This kind of solution is very effective against interferers but is not yet efficient for low power communication. In particular, such technology requires a radio front-end coupled to each of the antennas that, by emitting the same signal with a phase difference, together create a directed beam. The addition of these radio front-ends increases the overall consumption.

3.4.4 Nallatech

At the commencement of the OPERA project the most powerful SOC FPGA device on the market, in respect to flops/watt, was the Arria 10 660 device from Altera, now Intel. This device has been superseded by the Intel Stratix 10 device. At the time of writing this document, only Startix 10 development boards are available; these development boards are only available populated with the larger Stratix 10 devices. However, these development boards consume significantly more power than the OPERA FPGA device and are therefore not suitable for the OPERA project.

The other major FPGA vendor, Xilinx, has also released updated FPGA devices in the form of the Ultra scale Plus. As with previous devices, this device does not have native floating-point logic. Given the arguments presented in deliverable 6.1, this does not appear a threat to the OPERA innovation.

FPGA accelerator companies, other than Nallatech are also developing new FPGA accelerator products. This can be seen as a commercial threat to Nallatech, but given the portability of FPGA designs, it does not provide a threat to the innovation of the OPERA as a heterogeneous architecture and the approach being investigated.

In respect to CAPI implementation as part of work package 5, there have since been further products released that incorporate CAPI pre-enabled hardware with FPGA processing have been released since the start of OPERA project. Such a device has been developed by Mellanox (Mellanox Innova™ IPsec 4 Lx EN Adapter Card). This a threat to the FPGA accelerator card in regards of the CAPI implementation, but can be seen as complementary to the heterogeneity of the OPERA system.

3.4.5 ST

The actual state of the art, in line with the research conducted by STM in OPERA and in other parallel activities, is the Orlando R&D System on Chip. It is the first version on a CDNN processor, still in

progress and with foreseen improvements. The use of this SoC in the OPERA context in planned for the final prototype, and if possible also in the next version, if its integration in the global system can be completed in the needed timeframe.

3.4.6 Technion

Intel and AMD implement virtual memory with radix, hierarchical page tables, which organize the translations in a 4-level radix tree. Finding a missing translation in this hierarchy---a.k.a. ``walking" the page tables---thus incurs an overhead of four memory references. Attempting to mitigate the cost of TLB misses, AMD & Intel introduced special page walk caches (PWCs) to accelerate the page table walks (Intel Corporation, 2015) (AMD Inc., 2013). PWCs store partial translations---of prefixes of virtual addresses---thus allowing the hardware to quickly skip over upper levels in the radix tree hierarchy instead of traversing them. In the best-case scenario, when the table walker always hits the PWC, a walk requires only one memory access instead of four. Previous studies (Barr, 2010) have shown that SPEC CPU2006 benchmarks, for example, require 1.13 memory references per walk, on average.

3.5 HOW ARE YOU GOING TO MOVE PAST THIS STATE OF THE ART?

3.5.1 HPE

3.5.1.1 Moonshot

Leveraging System on Chip with more and more integrated functions seem to be a very good option to continue improving system performance efficiency. Intel will be releasing X86 processor for FPGA elements integrated to it which, in itself, is already a proof of the good foundation of the OPERA project. Going further and adding memory onto the package, such as HBM (high Bandwidth Memory) but also PMEM (Persistent MEMory) will dramatically increase the fields of applications for Moonshot and any systems leveraging SoC.

3.5.1.2 Redfish

Redfish is already a pretty robust and well represented standard (Spear, 2017) but it still lacks functionalities due to its relative youth. Redfish enables every vendor to implement their own private schemas following specificities of their hardware (as SNMP today) but a single entry that can make its way to the public area of the schema will be increasing the value and potentially the adoption of this standard (versus IPMI which is still the number one protocol). Having monitoring sensors on standardized PCIe boards is a very good start initiated with OPERA but there is still a lot of work to be conducted on defining new schemas, adding more options for refresh rate and historical data.

3.5.2 IBM

3.5.2.1 post-copy container migration

We are moving past the state of the art by adding support for post-copy migration to CRIU. Our initial implementation is provided by using the userfaultfd feature of the Linux kernel, which handles pages faults in user space. The connection is over a standard TCP/IP link, who’s latency can likely be improved by at least an order of magnitude by using RDMA technology.

3.5.2.2 low latency interconnect

To use RDMA for this purpose of low latency interconnect, requires specialized network equipment, but more and more NICs are starting to support this protocol. To push the envelope even further, we plan to implement a specialized protocol that will lower the latency and power consumption even further. This will be implemented inside the FPGA.

3.5.2.3 Cross-ISA compiler

Cross-ISA compiler being still in its infancy, we are working to move the compiler beyond the state of the art in two ways: first, by adding support for the POWER instruction set, and to make the compiler more stable and general. We aim at being able to compile common, readily available open source projects such

as Redis and Apache without making modifications. If successful, the resulting applications would then be ready for cross-ISA migration.

3.5.3 ISMB

Considering the energy inefficiency of state-of-the-art wireless communication systems (using digital beamforming), we aim at employing reconfigurable antennas in first place to reduce the energy consumption thus avoiding the negative effect of digital beamforming systems. In particular, we employ reconfigurable antenna coupled with single radio front-end.

3.5.4 Nallatech

By utilising a combination of low power servers (HPE Edgeline) and the latest SoC FPGA devices (as of 2016) we aim to achieve significant performance improvement per Watt whilst also reducing compute time versus a state-of-the art high end server. This will in part be achieved through careful management of power at the function level, partitioning code between, x86, ARM and FPGA where appropriate. Where high performance is required above efficiency, work can be offloaded to a Power system via the low latency serial interconnect on the platform agnostic FPGA accelerator.

3.5.5 ST

The results that can be obtained with a traditional approach in the video processing field are limited, and the new approach based on the CDNN could improve them, and create new applications. The actual progress over the SOTA is based on the improvement of the resources available on the SoC given the area and power consumption constraints.

3.5.6 Technion

The most important insight from our work is that the current, widely-available x86-64 platforms are inherently less scalable than hashed designs because they use more and more levels of translation as the Memory size increases. In fact, Intel just announced its plans to extend the virtual memory address spaces supported in x86-64 CPUs ((Intel Corporation, 2017)). The new Intel design will lengthen the page walk length from 4 to 5 memory references, further amplifying the overhead of virtual memory on modern, memory-intensive applications. Our proposed hashed page tables, in contrast, require only 1 memory reference for address translation regardless of the address space size. We believe that our work may guide future computer architects in their designing of new computer systems.

4 POTENTIAL MARKETS

Next to the industries mentioned already in D2.5, thinking about industries, extra markets have been identified by the OPERA consortium:

- CPU design (Intel/IBM)

The work executed by partners of the consortium (Technion) with regard to the memory management will potentially have effect on the way CPU hardware is designed. The hashed page table design uses considerably less transistors when compared with the TLB caches needed in the current implementations. Big CPU producers like Intel and IBM may see the capacity for transistors (buffering, other processes) increase, when the need for memory management transistors on the CPU will decrease.

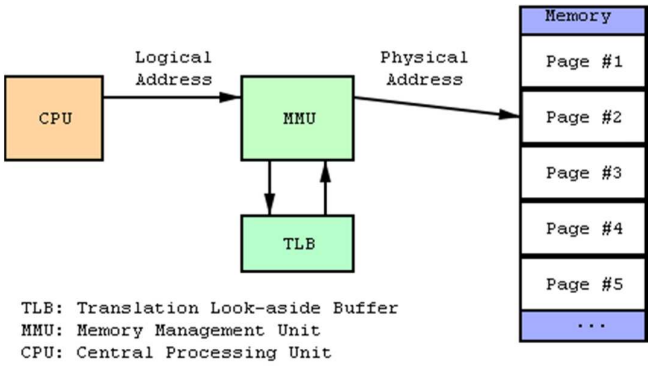

Figure 4 Memory Management Unit

Figure 5 a CPU

- Cloud solution providers

Thanks to the OPERA innovations, solution providers (like CSI), who offer their software solutions in the form of micro services (Open Source), will be able to dramatically increase the number of potential users while reducing the hardware and energy cost of operation.

- Sensor industry and monitoring of public infrastructure

The investigations about use cases as road traffic management, reinforced the potential interest of OPERA technology for the sensor providers: OPERA could provide new opportunities to develop the capability of the sensors, to locally treat in real time and to collect data for further investigation by very high power centralized smartness. Sensor industry could provide new solutions to monitor in particular road network and more generally public infrastructure (road network, electrical grid, gas network, water network). OPERA technologies enable to detect critical event in real time and to massively monitor infrastructure in postponed mode, by making merging big data tools possible.

4.1 ENVIRONMENTAL MANAGEMENT

The preliminary results obtained from the smart camera proves not only the potential for low power sensors, but the need for, and resulting efficiency from, embedded intelligence. The embedding of local processing power and autonomous decision making greatly diminishes the amount of data that is sent and stored. As such the sensor intelligence contributes to energy/equipment and cost savings in both the connecting networks as well as the end point datacenters. The smart camera created within the OPERA project can be used for demonstrating that this intelligence can be added to a sensor, even in a very tight energy envelope.

4.2 COLD STORAGE

Cold storage is a computer system or mode of operation designed for the retention of inactive data (Techtarget.com, sd). Cold storage techniques often use ARM based servers, rather than x86. If such a server were to be used in a hyper-converged data center, you may need to run the same application across various architectures. The FPGA connectivity developed for OPERA can benefit operations by allowing higher performing processors, from a different node, to run the cold storage software in high load conditions while still retaining access to the original physical data store.

4.3 CLOUD

Several large vendors (Google, Baidu, possibly Facebook) are experimenting with alternative chip architectures (ARM, POWER) to prevent vendor lock-in (Intel). If the trend continues/proves itself, it could drastically change the makeup of large data centers. Following the example set by these very large companies, it may be expected that other large data centers will implement heterogeneity within their configurations. The work done by OPERA, characterizing applications in terms of processor affinity under varying load and the experience using heterogeneity in a single virtual system has the potential to accelerate such a shift resulting in a better resource usage and higher energy efficiency for the data centers.

4.4 HAAS

HAAS or ‘HPC as a Service’ offers market potential. Recent discussions during the HPC conference in Frankfurt ISC-2017 (Ena-HPC.org, sd) clarified that the HPC community is already actively engaged in using accelerators of all kinds in the HPC infrastructures. The most pressing issue within this community is lack of skills and time needed to optimize the software running on the HPC infrastructure for the use of the particular infrastructure, including the accelerators (FPGA and others). Consequently, due to these programming difficulties, the performance / performance per Watt is not used effectively whereas it is crucial to these installations. In general, the averaged performance is below 10% of the listed peak performance of these installations (from discussion during the ISC workshop “energy aware HPC” on Thursday 22/6/2017). As such, the OPERA innovations will not be of great impact on HPC since we do not target to facilitate programming for these specific infrastructures.

5 MARKET ANALYSIS

In this section, we will reflect the progressing insights in the matters that were presented in the Market analysis section of D2.5. We have additions to the threat of substitute products, the threat of new entrants and on the way partners start to consider the market approach for OPERA products.

5.1 MICHAEL PORTERS 5P ANALYSIS

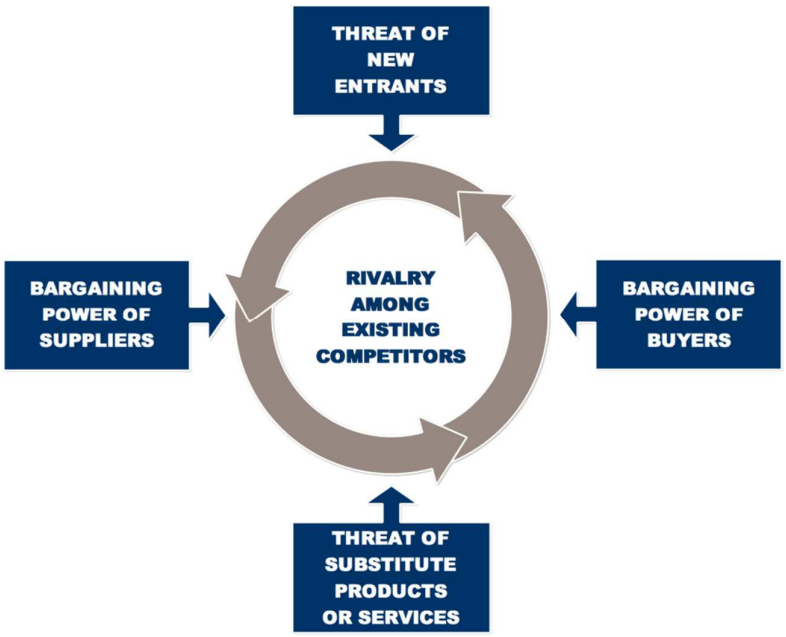

Figure 6 Five forces of Competition

5.1.1 Threat of substitute products

Machine learning, particularly Convolutional Neural Networks (CNN), has seen a boom in new dedicated processors. Below is a list of the current devices available and their relative threat to the OPERA hardware.

- Google TPU: The Google Tensor Processing Unit (TPU) has be designed specifically for machine learning. The threat to OPERA would be for the CNN offload compute. The TPU is only offered as a cloud service and therefore the power consumption would be difficult to verify. It's therefore unlikely that this would be a thread to the potential markets envisioned by OPERA despite its impressive performance capability.
- GPGPU: The General-Purpose Graphics Processing Unit (GPGPU) has consistently increased its share of the HPC market share over the past few years. The GPGPU provider NVidia owns the largest share of the market. Traditionally, GPGPU cards would be considered too power hungry to be considered as a low power technology suitable for OPERA. However, evermore powerful NVidia Quadro chips are appearing in low powered laptop devices. The Quadro P1000 processor available as of Aug 2017, is a single PCIe slot device physically similar to the FPGA device used for OPERA. This has a max power rating of 47 Watts and peak flop rating of ~1.89 Tera Flops. The device is there similar in performance to the FPGA device. As a pure performance offload device, this has to be considered a real threat to the OPERA hardware. However, this device lacks the diversity of the FPGA and would not be suitable for either the CAPI implementation for the VDI use case or as a bridge between Power8 and x86 servers as envisioned in OPERA.

5.1.2 Threat of new entrants

There are new FPGA technologies by new players identified. Deliverable D6.1 explained the reasons behind the choice of device/FPGA to be used for the OPERA FPGA accelerator. This has base on the best FPGA device currently available at the time. Taking into account the long development time of an FPGA accelerator the choice of device has been locked for the duration of the project. Therefore, the Arria10 SOC device was selected being the most versatile and powerful FPGA device available at the time. Since the commencement of the project new FPGA devices have been manufactured by both Xilinx and Altera. The following table compares the OPERA device versus these new emerging technologies.

Device	Floating point Performance GFlops/Sec	Peak Power (Watts)	Watts/Flop	Comments
OPERA 385 FPGA Baseline	1,519 (Intel Corporation)	50[2]	0.033	Single slot NIC, allows up to 4 devices in Moonshot EL4000 server. Devices are fully programmable using OpenCL.
Stratix 10 (Nallatech 520 device) (Nallatech, 2017)	11,520 (Intel Corporation, 2017)	250[3]	0.022	The heat sink required for the device limits the minimum form factor of the card to a GPU form factor. This would be too large of the EL4000 server and requires more than twice the physical space. The power per flop is less but the overall power footprint is significantly increased. It is unlikely a Stratix 10 device would be appropriate for the low power OPERA system.
Xilinx Ultrascale Plus	4,903 (Parker, 2017)	50% improvement versus 7 series, or 30 % versus Ultrascale.		Xilinx devices still do not have hardened floating-point units and therefore the floating-point IP requires extra FPGA logic unlike the Altera devices. This increases the logic requirements and may limit what can be implemented on the FPGA. The device power consumption has been reduced by ~30% from Ultrascale which was also a 25% reduction from Xilinx 7 series devices (see Figure 1). Table 1 lists the power required for floating point units performed on different technologies. An improvement for 30% for Ultrascale plus, still leaves it short of the OPERA Arria10 device in terms of efficiency. Xilinx now have a beta version of an OpenCL compiler available as of now (Aug 2017). However, this is still limited to IP components and does not provide a full system integration solution unlike the Intel offering. Therefore, the programing environment is still inadequate for OPERA's purposes.
Nvidia V100 (PCIe) (Nvidia, 2017)	14,000	250	0.018	The latest Nvidia GPGPU (Aug 2017) has a lower flop to watt ratio than the Stratix 10 device. However, the overall power consumption and GPU form factor make it unsuitable of the low powered OPERA system. GPUs are only suitable for offload, i.e. the "Datacentre in a truck use case".

[2] See D6.1 for details

[3] Approximate power required for 520 card with memories, optical connections, etc.

Device	Floating point Performance GFlops/Sec	Peak Power (Watts)	Watts/Flop	Comments
Quadro P1000	1890	47	0.025	NIC sized GPU device based on latest NVidia silicon (Aug 2017)

Table 3 OPERA device versus these new emerging technologies

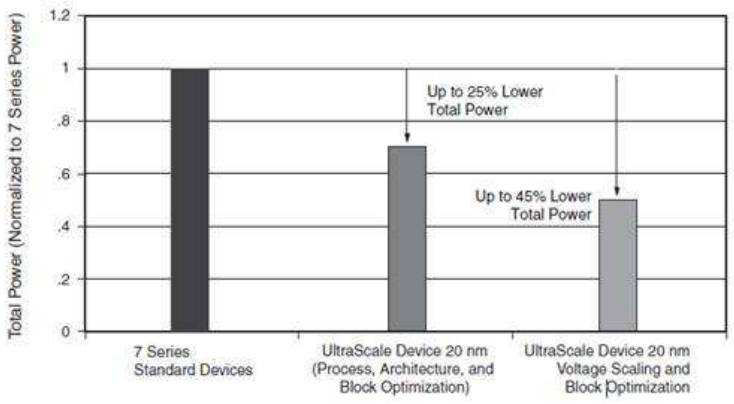

Figure 7 Power improvements from7 series Xilinx FPGAs moving to UltraScale (Kolluri, 2015)

Function	*LUTs*	*FFs*	*Freq (MHz)*	*Toggle %*	*Routing Factor*	*Multipliers*	*Power (Watts)*
Floating Point DSP Addition (Arria 10)	***0***	***0***	***250***	***12.5***	***3***	***2***	***0.0009***
Float Point Multiply Accumulate (Arria 10)	***0***	***0***	***250***	***12.5***	***3***	***2***	***0.0009***
Floating Point DSP Addition (Zynq)	***219***	***309***	***250***	***12.5***	***3***	***2***	***0.003[4]***
Float Point Multiply Accumulate (Zynq)	***777***	***1077***	***250***	***12.5***	***3***	***2***	***0.011***

[4] The Zynq power figures were created using the Xilinx Power Estimator; https://www.xilinx.com/products/technology/power/xpe/license-7series.html

Table 4 DSP power requirements. (From deliverable D6.1)

Graphcore (https://www.graphcore.ai/) is a recently established private company developing a new device dedicated to machine learning. Unlike the TPU it will be available off the cloud. There is currently no performance data available for this device, however it may pose a threat for machine learning and machine inference applications that may have otherwise targeted the OPERA hardware.

5.2 MARKET APPROACH

At this stage of the second intermediate release, there is no definite answer yet to the 'how' these products will be successfully brought to the markets identified. This will be the focus of D2.7. Some of the partners have ideas how to make the next step though. IBM for example, aims to push its technological advances that are being developed as part of the OPERA project internally within the company. "By meeting with leaders of various business units (such as the Cloud business unit and POWER business unit) we can sell our technologies, and help the business units develop the advancements into products. As part of the research division, we cannot directly control the choices of products, inclusion in products, etc. However, one of our primary goals is to see our technologies being adopted inside the company to improve our products and services provided to customers" (Joel Nider, IBM Haifa).

6 IMPACT ENVISAGED

6.1 CONCLUDING REMARKS FOR THIS ITERATION

Next to the statement already delivered in D2.5, the work performed as part of the "OPERA use cases" will provide insight into what compute problems can be distributed over heterogeneous architectures in a way that increases performance and minimises power. It is hoped that this work will encourage the compute industry to look towards heterogeneous systems as platform to significantly reduce power of new and existing compute problems, by illustrating performance improvements and along with improved design approach (RedFish, OpenCL).

The progress of the project in general and of the WP2 in particular, reinforces the innovation potential related to the coupling between the embedded sensor systems and the low power servers.

Considering the research in the traffic monitoring use case, results will have envisaged impact for the management of large public and private infrastructures, like transport and energy infrastructures. By coupling embedded smart sensors system and centralized low power server, OPERA opens new perspectives to detect events, impacting operations in real time, to investigate finer and more systematically the status of these infrastructures.

The demonstration of the traffic monitoring is in progress, coupling real time detection of critical events (like 'congestion' and 'wrong way vehicle') and finer data analysing (as accurate cycle counting). Beyond the SOTA traffic management; new opportunities emerge about the management of the road infrastructure (road, bridge and tunnel status). Beyond the road infrastructure, new opportunities could be considered about smart electrical grid, water distribution system, etc.

Such advancements enable implementation of innovative management of such (public) infrastructures. The capability of an autonomous embedded platform to define what is locally treated in real time and what is treated by HPC in postponed mode, will have crucial impact.

7 REFERENCES

AMD Inc. (2013). AMD64 Architecture Programmer's Manual, Volume 2.

artemis-emc2. (2016, 8). *artemis-emc2*. Retrieved from http://www.artemis-emc2.eu

axiom. (2016, 8). *axiom*. Retrieved from http://www.axiom-project.eu

Barr, T. W. (2010). "translation caching: Skip, don't walk (the page table)". *In ACM/IEEE International symposium on Computer architecture (ISCA), pages 48-59, 2010.*

Consortium, O. (2015). *OPERA_PartB_section1-3_2016_08_05*. OPERA. ST.

Distributed Management Task Force inc. (2017, 03 30). *redfish developers hub homepage*. Retrieved from redfish developers hub: http://redfish.dmtf.org

DRedBox. (2016, 8). *DRedBox*. Retrieved from http://www.dredbox.eu

Ena-HPC.org. (n.d.). Retrieved 2017, from http://www.ena-hpc.org/index.html

exanest. (2016, 8). *exanest*. Retrieved from http://www.exanest.eu

Gartner. (2016, april 27). *G00301050*. Retrieved from Cool Vendors in Cloud Infrastructure.

Gartner. (2016, april 26). *G00301053*. Retrieved from Cool vendors for compute platforms.

Gartner. (2016, 5 13). *G00301565*. Retrieved from Cool vendors in sustainability.

Gartner. (2016, 5 11). *G00302247*. Retrieved from Cool vendors in internet of things analytics.

Gartner. (2016, July). *Gartner Hype Cycle for Data Science.*

Gartner. (2016). *Internet of Things Primer for 2016 doc.nr.G00303518*. Gartner.

Gitlab. (2016, 8). *aparapi-ucores*. Retrieved from https://gitlab.com/mora/aparapi-ucores

HPE. (2017, 03 30). *HPE Moonshot system family guide*. Retrieved from HPE Moonshot system: https://www.hpe.com/h20195/v2/getpdf.aspx/4AA4-6076ENW.pdf

Intel Corporation. (2015). Intel 64 and IA-32 Architectures Software Developer's Manual, Volume 3B.

Intel Corporation. (2017). *5-Level Paging and 5-Level EPT.*

Intel Corporation. (2017). *Stratix 10 Product Table.* Retrieved 2017, from www.altera.com: https://www.altera.com/content/dam/altera-www/global/en_US/pdfs/literature/pt/stratix-10-product-table.pdf

Intel Corporation. (n.d.). *Intel Arria 10 Product Table.* Retrieved from https://www.altera.com/content/dam/altera-www/global/en_US/pdfs/literature/pt/arria-10-product-table.pdf

Intel. (n.d.). *Intel Arria Product Table.* Retrieved from www.altera.com: https://www.altera.com/content/dam/altera-www/global/en_US/pdfs/literature/pt/arria-10-product-table.pdf

Kim, W. C. (2005). Blue Ocean Strategy. Harvard Business review Press.

Kolluri, S. (2015, 10 15). *UltraScale Architecture Low Power Technology Overview*. (Xilinx, Producer) Retrieved 2017, from www.xilinx.com: https://www.xilinx.com/support/documentation/white_papers/wp451-ultrascale-pwr-reduction.pdf

Nallatech. (2017). *Intel Stratix 10 FPGA – Nallatech 520 FPGA Acceleration Card*. Retrieved 2017, from Nallatech.com: http://www.nallatech.com/store/fpga-accelerated-computing/pcie-accelerator-cards/nallatech-520-compute-acceleration-card-stratix-10-fpga/

NVIDEA. (2016, august 16). *correcting some mistakes*. Retrieved from https://blogs.nvidia.com/blog/2016/08/16/correcting-some-mistakes/

Nvidia. (2017). *WELCOME TO THE ERA OF AI.* Retrieved 2017, from www.nvidia.co.uk: https://www.nvidia.co.uk/data-center/tesla-v100/

Parker, M. (2017). *Understanding Peak Floating-Point Performance Claims.* Retrieved from www.altera.com: https://www.altera.com/en_US/pdfs/literature/wp/wp-01222-understanding-peak-floating-point-performance-claims.pdf

Spear, K. (2017, 1 3). *Cisco Supports Redfish Standard: API Enhances UCS Programmability*. Retrieved 2017, from Cisco Blogs: https://blogs.cisco.com/datacenter/cisco-supports-redfish-standard-api-enhances-ucs-programmability

Standard Performance Evaluation Corporation. (n.d.). *www.spec.org*. Retrieved 2017, from https://www.spec.org/cpu2006/Docs/

Techtarget.com. (n.d.). Retrieved 2017, from http://searchstorage.techtarget.com/definition/cold-storage

Villa, N. (2016, august 30). *ICT and sustainability: unleashing innovation*. Retrieved from ICT4S keynotes: http://2016.ict4s.org/program/keynote-speakers/nicola-villa/

Vineyard. (2016, 8). *Vineyard*. Retrieved from http://www.vineyard-h2020.eu

Xilinx Inc. (2016, Feb 19). *Xilinx Corporate and ortfolio Transformation*. Retrieved from youtube.com: https://www.youtube.com/watch?v=9jtvNPySPkc