Large-scale EXecution for Industry & Society

Deliverable D9.5

Market Analysis of Converged HPC, Big Data and Cloud Ecosystems in Europe

Co-funded by the Horizon 2020 Framework Programme of the European Union
Grant Agreement Number 825532
ICT-11-2018-2019 (IA - Innovation Action)

DELIVERABLE ID | TITLE	D9.5 | Market Analysis of Converged HPC, Big Data and Cloud Ecosystems in Europe
RESPONSIBLE AUTHOR	Piyush Harsh (CYC)
WORKPACKAGE ID | TITLE	WP9 | Impacts on Targeted Sectors
WORKPACKAGE LEADER	TESEO
DATE OF DELIVERY (CONTRACTUAL)	30/06/2020 (M18)
DATE OF DELIVERY (SUBMITTED)	30/06/2020 (M18)
VERSION | STATUS	1.1 | Final
TYPE OF DELIVERABLE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	BLABS, LINKS, LRZ, IT4I, ATOS, TESEO, CYC
INTERNAL REVIEW	Natalja Rakowsky (AWI), Katerina Slaninova (IT4I)

Project Coordinator: Dr. Jan Martinovič – IT4Innovations, VSB – Technical University of Ostrava
E-mail: jan.martinovic@vsb.cz, **Phone:** +420 597 329 598, **Web:** https://lexis-project.eu

DOCUMENT VERSION

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Initial ToC and allocation of sections	22/04/2020	Sean Murphy (CYC), Piyush Harsh (CYC)
0.2	Initial contributions from all partner institutions	13/06/2020	Piyush Harsh (CYC), YuanYuan Li (LINKS), Etienne Walter (ATOS), Florin Apopei (TESEO), Marc Derquennes (BLABS), Tomas Karasek (IT4I), Stephan Hachinger (LRZ)
0.3	Section 6 extension	18/06/2020	Marc Derquennes (BLABS)
0.4	Executive Summary, Conclusion	19/06/2020	Piyush Harsh (CYC)
0.9	Review comments integrated, added list of acronyms	25/06/2020	Piyush Harsh (CYC)
1.0	Further fixes and formatting alignment	28/06/2020	Piyush Harsh (CYC)
1.1	Integration of comments from IT4I	29/06/2020	Piyush Harsh (CYC), YuanYuan Li (LINKS), Jan Martinovic (IT4I), Katerina Slaninova (IT4I), Tomas Karasek (IT4I)

GLOSSARY

ACRONYM	DESCRIPTION
HPDA	High performance data analytics
FPGA	Field programmable gate array
HPC	High performance computing
SME	Small to medium enterprise
GFLOPS	1 billion (Giga) floating point operations
RPEAK	Theoretical peak performance of a supercomputer using LINPACK benchmark
IAAS	Infrastructure as a service
PAAS	Platform as a service
SAAS	Software as a service
BPASS	Business process as a service
AWS	Amazon web services
IOT	Internet of things
DC	Data centre
LL	Lessons learned
KPI	Key performance indicator

DDI	Distributed data infrastructure
API	Application programming interface
ISV	Independent software/service vendor
ROI	Return on investment
EU	European Union
PID	Persistent identifier
DOI	Digital object identifier
CAGR	Compound annual growth rate
GDPR	General data protection regulation

TABLE OF PARTNERS

ACRONYM	PARTNER
Avio Aero	GE AVIO SRL
Atos	BULL SAS
AWI	ALFRED WEGENER INSTITUT HELMHOLTZ ZENTRUM FUR POLAR UND MEERESFORSCHUNG
BLABS	BAYNCORE LABS LIMITED
CEA	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES
CIMA	Centro Internazionale in Monitoraggio Ambientale - Fondazione CIMA
CYC	CYCLOPS LABS GMBH
ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
GFZ	HELMHOLTZ ZENTRUM POTSDAM DEUTSCHESGEOFORSCHUNGSZENTRUM GFZ
IT4I	VYSOKA SKOLA BANSKA - TECHNICKA UNIVERZITA OSTRAVA / IT4Innovations National Supercomputing Centre
ITHACA	ASSOCIAZIONE ITHACA
LINKS	FONDAZIONE LINKS / ISTITUTO SUPERIORE MARIO BOELLA ISMB
LRZ	BAYERISCHE AKADEMIE DER WISSENSCHAFTEN / Leibniz Rechenzentrum der BAdW
NUM	NUMTECH
O24	OUTPOST 24 FRANCE
TESEO	TESEO SPA TECNOLOGIE E SISTEMI ELETTRONICI ED OTTICI

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

LEXIS project is creating an innovative solution: to bring cloud and HPC resources into one seamless compute and storage offer towards European SMEs. LEXIS hopes is to enable SMEs innovate faster while keeping the core tenets of a European society, privacy and data safety, into consideration from start.

Position of the deliverable in the whole project context

This deliverable is the first in the series of two investigating the cloud, HPC, and Big Data markets trend globally, as well as the recent European trends specifically. This report is the result of investigation carried out as part of Task 9.3 within WP9 with equal participation from HPC centres, Industrial Partners, and SMEs involved in the LEXIS consortium. This deliverable, D9.5, identifies the market signals and assesses LEXIS innovation in that context. The LEXIS innovations carried out in work packages WP3 - the LEXIS Data System, WP4 - Orchestration and secure HPC/Cloud Services Provisioning, and WP8 - LEXIS Portal for 3rd Party Companies and SMEs, are used to position LEXIS in the correct context with respect to its potential to create disruption in the current market context.

Description of the deliverable

As the title suggests, this deliverable analyses the market key indicators such as market growth trends, application drivers in three marquee IT sectors: HPC, Big Data and Cloud. It presents both the global and European trends. This report presents the recent convergence trends in these sectors, and also presents the emerging trend of heterogeneity in datacentres as a key validator in LEXIS approach. This report places the LEXIS innovation in the overall market context and presents the potential in disruption as well as unique opportunity for European SMEs for easy consumption of converged Cloud/HPC services that will be made available post availability of LEXIS outcomes.

It is important to keep in mind that this deliverable presents an initial assessment and positioning of LEXIS. This positioning is bound to get corrected and more precise in the sequel which will be prepared towards the end of this project.

1 INTRODUCTION

Europe is at a computing crossroad today. Cloud computing with the myriads of offerings and deployment models has been embraced by both large enterprises and SMEs in the same spirit. From academic research to rapid transfer to industry, Europe has delivered a success story when it comes to embracing, nurturing and maturation of the field. The result has been an establishment of an ecosystem of vendors, users and policy-making bodies that has created an environment of sustainability in Europe centring around the core tenets of data security, user privacy and sustainable innovation.

HPC as a theme has been around much longer. It has been characterised by huge amount of capital investments at a few regional centres largely funded by public financing, stifling commercialization frameworks, and public bureaucratic interference (largely due to the nature of financing). Because of such artefacts, exploitation of large HPC capacities has been largely limited to being utilized in work done by other public funded bodies such as universities, meteorological centres, etc. SMEs have been visibly absent. Although this trend is changing lately, there are some key elements missing that would ease the pain of adoption among European SMEs.

LEXIS project’s main goals are to bring on-demand nature of computing from the cloud and scheduled execution batches from HPC world in a seamless interplay, and further make access of such resources transparent and seamless to SMEs in Europe. The purpose of this document is to assess the existing ecosystem of cloud and HPC operations within Europe and elsewhere; identify key drivers of adoption of these compute offerings - mainly from the applications perspective; investigate the market trends in order to justify and benchmark the unique role that the outcomes from LEXIS project will play in improving the adoption of converged cloud + HPC offerings among SMEs and Enterprises in Europe.

1.1 STRUCTURE OF THIS DOCUMENT

The deliverable is structured as follows - Section 2 deals with HPC market analysis where the available compute capacities in HPC centres are analysed, specific compute task requirements as well as challenges of exascale computing with heterogeneity is also outlined. In Section 3, Big Data as a key driver of innovation in both HPC as well as cloud is assessed, including how HPC centres and Big Data application needs are becoming a mutual driving force. Section 4 deals with trends in adoption of cloud technologies, together with market analysis focusing on various deployment models. In section 5, convergence aspects of HPC, Big-data applications and cloud services are analysed. In essence, Sections 2, 3, and 4 validate the market potential of cloud, big-data and large scale HPC solutions and Section 5 highlights the potential of interplay which exists in these sectors as a validation of the direction the LEXIS project is taking. Section 6 deals with the value proposition of LEXIS project results within emerging market trends, analysed in depth with respect to its facilitation as well as disruption potential. Section 7 concludes this document with key findings as well as a synopsis of what can be expected in the final version of this report series at the end of the LEXIS project execution.

2 HPC MARKET ANALYSIS

2.1 ANALYSIS OF CLASSICAL HPC SECTOR IN EUROPE

The Digital Economy is the most important driving force of innovation, competitiveness, and economic growth worldwide. Employing HPC together with AI and the use of Big Data provides huge opportunities for transformation of businesses, public services, and societies towards the digital era. HPC is making possible to address the most challenging issues in areas such as engineering, public health, climate change, and natural disasters. For example, the use of supercomputers allows researchers to solve complex problems such as developing new aircraft engines, or better predicting and mitigating the effects of natural disasters such as floods or tsunamis by using advanced computer simulations.

The potential of HPC is immense but key challenges persist. Main challenge is to close the gap in investment in HPC between Europe and the rest of the world. Table 1, Figure 1, and Figure 2 clearly show that there is a huge disproportion in number of the supercomputers and their power when we compare EU with the Americas and Asia. Data taken from the latest list (November 2019) of top 500 computers in the world[1] shows that this list is dominated by Asia followed by Americas. Europe is outnumbered in all aspects i.e. number of the supercomputers, number of available cores as well as computational power.

CONTINENTS	COUNT	SYSTEM SHARE (%)	Rmax (GFlops/s)	Rpeak (GFlops)	Cores
Asia	274	55	704,084,236	1,398,851,520	36,306,992
Americas	129	26	632,500,077	901,123,552	17,839,200
Europe	94	19	302,570,000	430,707,601	8,790,394

Table 1 Supercomputers listed in TOP500 by continents

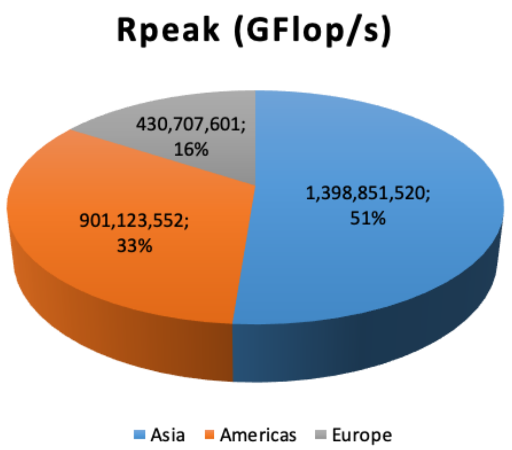

Figure 1 Total number of cores by continents

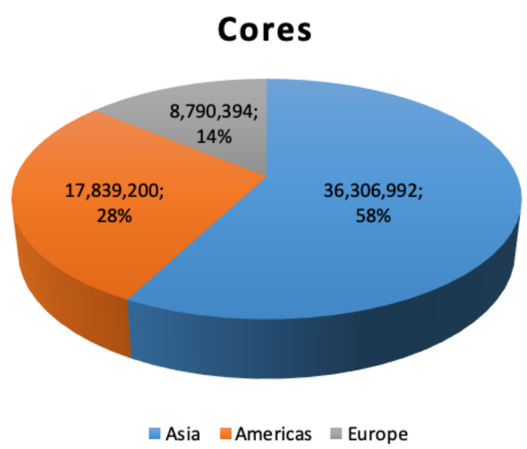

Figure 2 Total Rpeak performance by continents

Comparison of EU countries (see Table 2, Figure 3, and Figure 4) show that only 15 out of the 44 European countries operate supercomputers large enough to make it to the TOP500 list.

[1] TOP 500: www.top500.org

European Countries	COUNT	Rmax (GFlops/s)	Rpeak (GFlops)	Cores
France	18	68,889,020	105,646,033	2,122,784
Germany	16	66,894,410	98,393,728	1,732,022
Netherlands	15	24,736,650	31,795,200	864,000
Ireland	14	23,087,540	29,675,520	806,400
United Kingdom	11	32,143,142	39,454,658	1,189,608
Italy	5	30,098,790	47,843,836	794,032
Russia	3	10,347,350	15,029,555	199,120
Norway	2	3,298,220	4,239,360	115,200
Sweden	2	4,771,700	6,773,346	131,968
Spain	2	7,615,800	11,699,115	171,576
Switzerland	2	23,126,750	29,347,305	453,140
Czech Republic	1	1,457,730	2,011,641	76,896
Poland	1	1,670,090	2,348,640	55,728
Finland	1	1,706,730	2,688,000	40,000
Austria	1	2,726,078	3,761,664	37,920

Table 2 Supercomputers in TOP500 by European countries

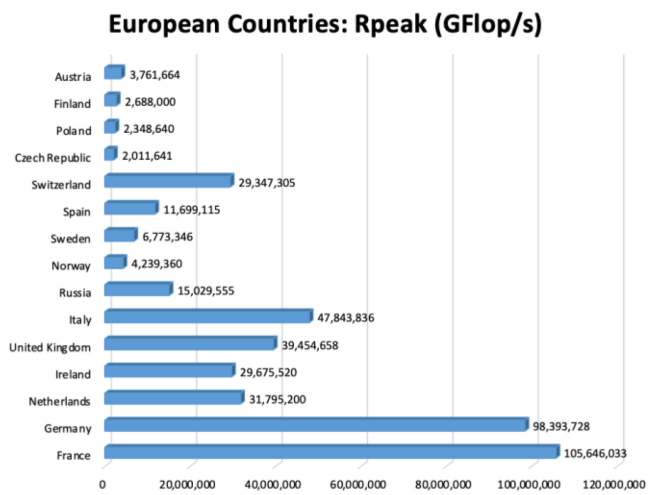

Figure 3 Supercomputers: Total number of cores by countries

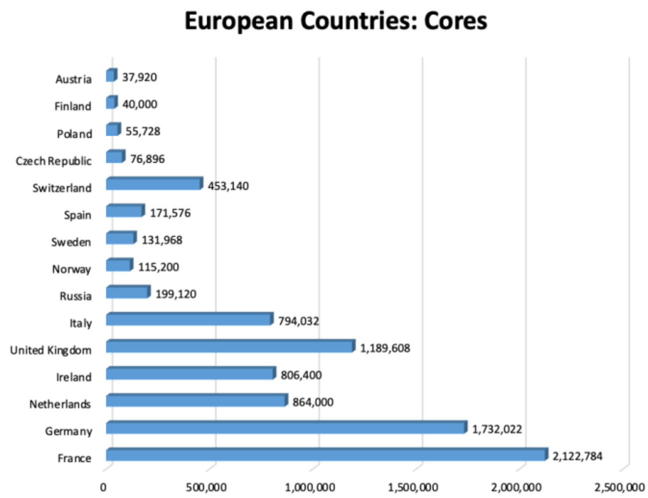

Figure 4 Supercomputers: Total Rpeak performance by countries

Although one third of the global demand for HPC capabilities are by European industries, SMEs and researchers, only 5% of the HPC capabilities are being provided by European HPC centres[2] European industry is aware of the

[2] Financing the future of supercomputing: https://www.eib.org/attachments/pj/financing_the_future_of_supercomputing_en.pdf

benefits of the HPC and their role in digital future which is demonstrated that more than 45% of the supercomputers from TOP500 list located in Europe is owned and operated by private companies, see Table 3.

COUNTRY	TOTAL	ACADEMIC	INDUSTRY	RESEARCH
France	15	3	3	9
Germany	16	6	3	7
Netherlands	15	0	15	0
United Kingdom	11	2	0	9
Ireland	14	0	14	0

Table 3 Top 5 European countries by count of supercomputers

Moore's law still applies in HPC: Although Moore's law states that density of transistors doubles every second year, this interpretation is coming to an end. When analysed with respect to compute power in specialised hardware, the trend is in line with Moore's law. The main trend of HPC in the last decade is constant increase of the performance. Data analysis of top500 list[3] between November 2012 and June 2018 shows that installed capacity and performance of the most powerful supercomputers almost double every two years (exponent of exponential regression 0.345, doubling period 2.0075 years). This is in line with Moors law[4] which predicts doubling in every other year.

Three worlds of HPC: The HPC world could be divided into three categories. "Mainstream" where supercomputers with an output up to 10 PFlop/s belong. This category makes up to 96% (www.top500.org: November 2019[5]) of all supercomputers worldwide. In "Knee" category with power up to 30 PFlop/s, 3% of supercomputers could be found. And last but not the least elite club "Leaders" with 1% of supercomputers with power exceeding 30 PFlop/s. Current No.1 of the supercomputers possess power of 149 PFlop/s (Summit, Oak Ridge National Laboratory, USA). Most powerful European supercomputers are in "Knee" category, led by Piz Daint, CSCS, Switzerland.

Exascale era: Considering the above-mentioned trend in growth of performance the 1,000 PFlop/s i.e. 1 EFlop/s barrier is expected to be broken soon. United States Department of Energy recently announced their plan to install supercomputer Aurora in Argonne National Laboratory which should exceed 1E Flop/s[6]. However, this is not a simple task considering that it means increasing power of existing system by 7 times, while maintaining the same energy consumption. European countries are trying to keep up in this race and three pre-exascale systems (LUMI in Finland, Mare Nostrum 5 in Spain and Leonardo in Italy) should be installed by 2021[7].

Heterogeneity, way forward: Homogeneous supercomputers i.e. supercomputers without any accelerators or specialized processors are represented in TOP500 list (November 2019) mainly in "Mainstream" category (those with power up to 10 PFlop/s) where only 25% of them are accelerated. On the other hand, all supercomputers in "Leaders" category are accelerated. Also, in "Knee" category 83% of supercomputers are accelerated. Currently 70% of the supercomputers in TOP500 are homogeneous ones but in 2013 this number was at 90% which shows a clear trend towards using of accelerators thus heterogeneous systems. The segment of accelerators is dominated by General Purpose Graphics Processing Unit (GPGPU) mainly produced by NVidia. In TOP500 list (November 2019) 94% of accelerated HPC systems are equipped with NVIDIA. Those accelerators itself are heterogeneous. For

[3] TOP500: www.top500.org [accessed: 19/06/2020]
[4] Moore's law: https://en.wikipedia.org/wiki/Moore%27s_law [accessed: 19/06/2020]
[5] The TOP500 supercomputer list gets updated twice a year!
[6] TOP500 News: https://www.top500.org/news/retooled-aurora-supercomputer-will-be-americas-first-exascale-system/ [accessed: 19/06/2020]
[7] Call to acquire world-class supercomputers in Europe: https://ec.europa.eu/digital-single-market/en/news/first-call-acquire-world-class-supercomputers-europe [accessed: 29/06/2020]

example, new NVidia V100 (50% of above mentioned HPC systems in TOP500 from November 2019 are equipped by this accelerator) consists of specialized cores for tensor operations. This shows that heterogeneity is the future trend.

AI, when 64 bit arithmetic's is not needed: In AI applications 16 bit arithmetic is the key of the speed [1]. In those applications using of 64 bit arithmetic will not bring any increase of the results fidelity so less precise but much faster 16 bit arithmetic is preferred. Increasing popularity of the AI also influences HPC technologies. The main change is that calculations in 16 bit precision are supported by hardware which doubles performance comparing to 32 bit precision and four times higher comparing to full 64 bits arithmetic. It also lowers memory demand by same order. Some climate research centres are also moving to reduce the precision (64 bit -> 32 bit) as it is a rather cheap way to increase the performance.

High Bandwidth Memory increases the performance of many HPC applications. The HPC applications could be divided into two categories [2]:

- Applications with relatively low number of operations to data access, irregular memory access, and recursive computations,
- Applications with relatively high number of operations, regular memory access, and large computational blocks.

Applications from the first category are limited by memory throughput, thus increasing memory throughput will directly impact their performance. Since approx. 50% of all HPC applications are of this type introducing HBM memory is a game changer because they provide better performance by up-to 10 times than DDR memory.

Increasing energy efficiency is the only way. The increase in performance makes only sense when constant energy consumption is maintained. Otherwise HPC systems will not be competitive. The energy efficiency of supercomputers is measured in GFlop/s per Watt (GF/W). Currently, most energy efficient supercomputers reach 15GF/W. A further increase in efficiency up to 50GF/W is desirable for Exascale supercomputers. This energy efficiency increase has been and most likely will be achieved by heterogeneous supercomputers with the use of accelerators.

3 BIG DATA MARKET ANALYSIS

The economy and society have been transformed by digital technologies over the last few years. Data is at the centre of the transformation. According to the prediction from IDC[8], globally, the total amount of data created is to increase sharply from 33 zettabyte in 2018 to 175 zettabyte in 2025. Specifically, 51% of the data will be in data centres and the rest will be in the public cloud. The rapid growth is from social media, E- commerce, mobility, and other factors.

The increasing volume of data and adoption of Big Data tools boost revenue growth during the forecast period. As a result, the global Big Data market is forecasted to grow from 42 billion U.S. dollars in 2018 to 103 billion U.S. dollars in 2027 as shown in Figure 5. The fastest growing Big Data market is the Chinese market with a compound annual growth rate (CAGR) of 31.72%. Its revenue is to reach 9 billion U.S. dollars in 2020 from 1.2 billion U.S. dollars in 2014. For European Big Data market, it is expected to expand with an annual growth of 25.7% from 2020 to 2025.

[8] The Digitization of the World: From Edge to Core: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf, [accessed: 27/05/2020]

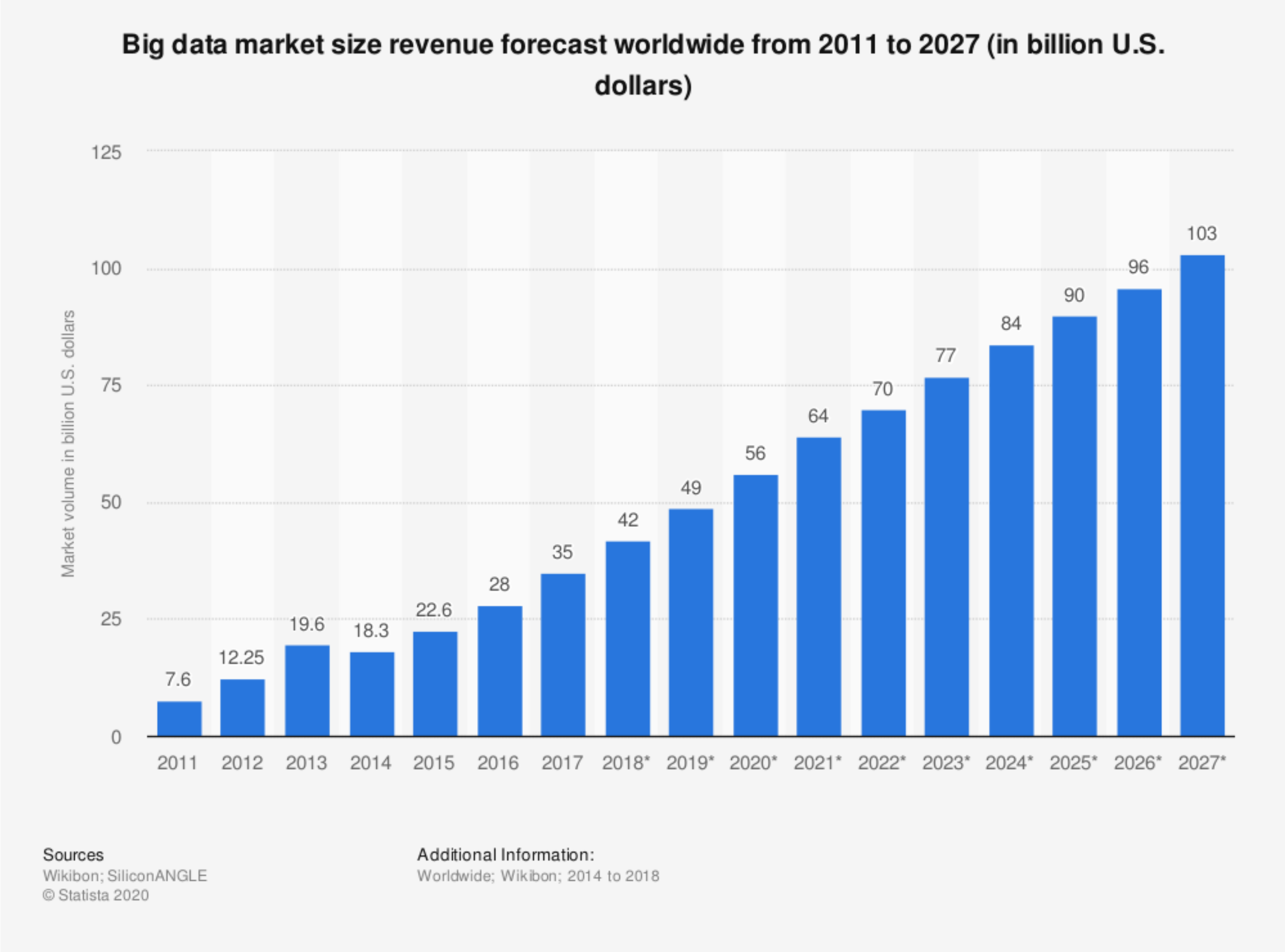

Figure 5 Big Data market size revenue forecast worldwide from 2011 to 2027 (in billion U.S. dollars) [9]

The major growth factors of the Big Data market include the increasing awareness of Internet of Things (IoT) devices, which forms the base for investment in Big Data services, increasing government investments for advancing digital technologies, and increasing availability of data across the organization for gaining more insights to remain competitive[10].

While data is growing dramatically, to get valuable insights from the huge amount of data becomes more critical. In crucial situations, urgent computing often plays a role to provide decision support. It is estimated that by 2025 nearly 30% of all generated data will be real time compared to 15 % in 2017. As a growing amount of data is generated every millisecond, it is essential to understand the data fast enough. From earthquake and tsunami assessments to the tracking of epidemic spreads, a few seconds' delay in processing information could cause the loss of lives. Besides, to keep pace with the storage demands from the data generation, it is predicted that over 22 zettabyte of storage capacity must ship across all media types from 2018 to 2025. Data centre power, space limitation and movement of data etc. all remain as challenges. There are other issues including a lack of the proper skill sets to incorporate and control the Big Data ecosystem, the matter of legal compliance that has been observed in industrial applications (non-personal data sharing), etc.

[9] Statista: Big data market size revenue forecast worldwide from 2011 to 2027: https://www.statista.com/statistics/254266/global-big-data-market-forecast/ [accessed: 28/05/2020]

[10] Big Data Market Forecast: https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html [accessed: 27/05/2020]

To cope with some of the challenges, a set of techniques are prevalent[11]:

- Data analysis automation. The traditional data analytics performed by analysts is too slow. With the addition of artificial intelligence to business intelligence, it increases the possibility to build data analytics directly into business processes.
- Augmented analytics. Through, the use of augmented analytics enabled by technologies such as natural language processing to assist with data preparation, insight generation and results explanation, how people explore and analyse data is augmented.
- Analytics and edge computing. To collect, manage and analyse the huge volumes of data generated by edge devices, commonly real-time streaming data, edge computing is a solution.
- Data as a service. Traditionally data has been locked up in data stores that is accessible by specific applications. Data as a service utilizes cloud technology to provide on-demand access to data regardless of the locations of people or applications.
- Analytics: Beyond data prediction to data prescription. Historically data analytics was used to discover or understand things happened in the past, which is called "descriptive analytics". Now we are seeing "prescriptive" data analysis goes beyond discovering what has happened and providing insights for future actions.
- Increased data management regulation. In May 2018, the European Union's General Data Protection Regulation (GDPR12) took effect. It is predicted by Ventana Research by 2021 one-quarter of organizations will establish new centres of excellence focused on data governance and security to reduce the risk of misuse or threat.
- Increased use of graph databases and analytics. Increasing complexity of data structures brings challenge to analytics. Graph databases, analytics software and visualization tools are designed to face the challenge by showing the relations among people, places, etc.

To keep pace with the Big Data market, many existing applications have invested in information management and get benefits from it. For example, in healthcare, a data-driven analysis that enables personalized medicine and prescriptive analytics is available thanks to the increasingly adoption of mobile devices, wearable technologies, etc. In nanoscience, to accurately describe realistic samples, Big Data analytics provide the solutions to navigate and analyse enormous amount of information. In banking sector, the techniques from Big Data such as clustering help identify new branch locations where the demand is high, and the integration with Machine Learning and Artificial Intelligence helps many banks detect fraudulent activities. In some disciplines real time does not play a huge role. The amount of data grows due to increased resolution of simulations and re-analysis, and due to new measuring techniques, in particular from satellites. In such scenarios, adoption of FAIR[13] data principles are crucial.

Moreover, the rapid increment of data leads to the emerging of application in various fields. Such as intelligent mall shopping, which facilitates with location-based mobile augmented reality applications; network anomaly detection, which efficiently uses network Big Data to perform accurate anomaly detection; autonomous vehicle, which rely entirely on the data they receive through GPS, radar and sensor technology, and information processed through cameras.

[11] CRN: Emerging Trends In Big Data Management And Analytics To Watch In 2020: https://www.crn.com/slide-shows/applications-os/8-emerging-trends-in-big-data-management-and-analytics-to-watch-in-2020/2 [accessed: 16/06/2020]

[12] EUR-Lex, Document 32016R0679: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679 [accessed: 28/06/2020]

[13] FAIR data: https://en.wikipedia.org/wiki/FAIR_data [accessed: 28/06/2020]

4 CLOUD MARKET ANALYSIS

Many organizations are considering the cloud as a primary resource for complex and large compute intense workloads: cloud based solutions offer a number of advantages, chief amongst them (i) the set of sophisticated tools and services offered by cloud services and (ii) the great flexibility to easily scale both storage and compute requirement up and down that is possible.

This chapter provides an overview of cloud-based offerings that are changing for both Enterprise and SME sectors and how it relates to complex analytics and Big Data workloads discussed in Section 3 above.

4.1 PUBLIC-PRIVATE CLOUD FOR STORAGE AND COMPUTE

In this section, we focus on the infrastructural aspects of cloud computing solutions.

4.1.1 Growth in cloud computing

It is well known that Cloud Computing remains a rapidly growing area within IT even today. This growth manifests in multiple dimensions. At the most basic level, the original premise of Cloud Computing - virtualized infrastructure continues to grow at very healthy rates; Table 4 below shows an aggregate growth of 31.3% in revenue for IaaS products offered by the public cloud providers. Amazon remains the market leader, but Microsoft is competing aggressively with deep integration in many enterprises and hence will continue to grow aggressively and likely increase market share.

COMPANY	2018 REVENUE (US$BN)	2018 MARKET SHARE (%)	2017 REVENUE (US$BN)	2017 MARKET SHARE (%)	2018 - 2017 GROWTH (%)
Amazon	15,495	47.8	12,221	49.4	26.8
Microsoft	5,038	15.5	3,130	12.7	60.9
Alibaba	2,499	7.7	1,298	5.3	92.6
Google	1,314	4.0	820	3.3	60.2
IBM	577	1.8	463	1.9	24.7
Other	7519	23.2	6,768	27.4	11.1
Total	**32,441**	**100.0**	**24,699**	**100.0**	**31.3**

Table 4 Worldwide Cloud Computing Infrastructure Revenues - Public Cloud Providers[14]

Forecast growth rates for infrastructure spend continue to be healthy with a projected doubling to over $70bn in 2022 - see Table 5 below.

Table 5 also captures more general cloud based service offerings, highlighting other areas within the cloud service offerings that are poised for growth; clearly Software as a Service (SaaS) services will continue to grow aggressively but these are less relevant within the LEXIS context. The area in which significant growth is perceived which relates to LEXIS is the Application Infrastructure Services which refers to advanced services tied to infrastructure and incorporates data management and analytics.

[14] Gartner Press Release: https://www.gartner.com/en/newsroom/press-releases/2019-07-29-gartner-says-worldwide-iaas-public-cloud-services-market-grew-31point3-percent-in-2018 [accessed: 15/05/2020]

SERVICE TYPES/REVENUE	2018 (US$BN)	2019 (US$BN)	2020 (US$BN)	2021 (US$BN)	2022 (US$BN)
Cloud Business Process Services (BPaaS)	41.7	43.7	46.9	50.2	53.8
Cloud Application Infrastructure Services (PaaS)	26.4	32.2	39.7	48.3	58.0
Cloud Application Services (SaaS)	85.7	99.5	116.0	133.0	151.1
Cloud Management and Security Services	10.5	12.0	13.8	15.7	17.6
Cloud System Infrastructure Services (IaaS)	32.4	40.3	50.0	61.3	74.1
Total Market	**196.7**	**227.8**	**266.4**	**308.5**	**354.6**

Table 5 Worldwide Cloud Computing Revenue Forecast by service type[15]

From the above two tables, it is clear that projected revenues for cloud based services will grow significantly in the coming years.

Growth within cloud offerings is not limited to infrastructure, naturally; cloud providers are offering an increasingly diverse set of services. AWS is a clear leader in terms of diversity of service offerings, ranging from Alexa integrations to AR/VR supports to more general IoT integrations, long term storage etc. The other public cloud providers are also innovating - Google is noteworthy in terms of its hardware innovations with its TPU and Edge TPU offerings which can support more efficient Machine Learning computation than other compute hardware.

However, this is not the only way to view the cloud sector - it is interesting to also consider the number of services offered by the cloud providers. AWS, in this regard, is clearly the leader, offering the most feature rich platform and evolving it rapidly. AWS is evolving its platform in many dimensions, ranging from Alexa integrations to AR/VR supports to more general IoT integrations, long term storage etc. Other cloud providers are also enriching their platforms with, for example, access to new hardware (ARM servers, GPU enabled servers etc) - the Google TPU work and the linkage between the TPU edge devices and cloud hosted data management service is another example.

One area that is receiving much attention is tools to support many different forms of advanced, compute intense data processing - an area which historically was mostly the domain of HPC centres. One example in the Google context is support for Tensorflow processing on custom designed TPU hardware but there are many others, including access to pretrained function specific networks such as object detection or speech analysis. AWS offers a powerful solution for training and management of Machine Learning models via their Sagemaker platform. AWS also offers solutions for large scale data ingestion via their Snowball products for some 10s of Terabytes of data up to many PB.

The diversity of service offerings available from cloud providers in both infrastructure as well as higher level services provides an opportunity for both HPC centres as well as users of HPC centres - the former can experiment in different ways with cloud provider infrastructure without requiring substantial HPC hardware investment; the latter can experiment with different application designs and different application optimizations under different assumptions on a pay as you go model.

4.1.2 Public cloud and private cloud

Although cloud computing is growing at a healthy pace, it must be considered in context. Considering total IT infrastructure spend, cloud computing accounts for just under 50% (see Table 6), even when considering public and

[15] Gartner press release: https://www.gartner.com/en/newsroom/press-releases/2019-11-13-gartner-forecasts-worldwide-public-cloud-revenue-to-grow-17-percent-in-2020 [accessed 17/05/2020]

private cloud spending together. Private cloud spending accounts for just under 1/3 of all cloud infrastructure spending.

	2018 ($BN)	2019 ($BN)	Change (%)
Public Cloud	45.2	45.2	0.1
Private Cloud	20.1	21.5	6.6
All Cloud	66.3	66.8	2.3
Non-Cloud	70.5	67.6	-4.2
All IT	135.9	134.4	-0.1
Cloud Percentage	48.1	49.7	1.6

Table 6 Worldwide infrastructure spend divided into categories[16]

These basic data points highlight that while the cloud is very important and is growing at a healthy pace, it still accounts for a minority of total IT spend. Interestingly, infrastructure expenses spent on private cloud deployments are growing faster than those for public cloud (with the caveat that infrastructure expenses alone do not give a complete perspective).

In the context of scientific HPC centres, on-premises (private) cloud deployments have become increasingly attractive, e.g. for pre- and post-processing tasks related to jobs run on supercomputers. With their competence in building large computing clusters, many centres have made it to offer medium-sized IaaS cloud infrastructures at an affordable cost. For industrial researchers as possible customers, these infrastructures clearly have advantages in terms of privacy. OpenStack clusters and VMWare infrastructures make up for a large - if not the largest - share of the mostly Linux-centric IaaS environments at public HPC centres. OpenNebula is another attractive open-source alternative; in the future, PaaS solutions and on-premises deployments of Kubernetes (e.g. with OpenShift) will probably play a larger role. Both OpenStack and OpenNebula offer a “burst” possibility to public (AWS, Google, Azure, ...) clouds, making hybrid-cloud setups possible (see also below). Clearly, more vendor-specific solutions such as AWS Outpost, Google Anthos, or Azure Stack have convenient on-premises deployment capabilities as well, with seamless connectivity to the respective public cloud services. These solutions, on the other hand, may come with some probability for a vendor lock-in.

4.1.3 Hybrid and multi-cloud developments

The first phase of cloud computing was characterized by a debate of the relative merits of public cloud and on-prem deployments. This has evolved into a much more complex set of possible configurations in which the IT sector understands that the set of possible options is much more diverse, and most organizations will employ combinations of different approaches.

The essential characteristic of hybrid cloud solutions is that it combines infrastructure which sits within a public cloud with on-prem infrastructure. How this is used varies greatly - in simple cases, it can be just used to connect extra more flexible infrastructure to an enterprise DC while in other cases, it can comprise of a complex implementation of distributed storage and compute with different backup and reliability requirements.

One specific variant of hybrid cloud solutions is worth noting: the approach in which systems controlled by cloud providers operate within an enterprise context. Both AWS and Azure offer solutions which fit into this category,

[16] The Next Platform News: https://www.nextplatform.com/2020/04/16/the-outlook-for-infrastructure-is-cloudy-in-a-good-way/ [accessed17/05/2020]

effectively blurring the lines between the resources of the cloud provider and those of the organization. AWS offers its Outpost solution and Azure offers its Azure Stack solution: both of these solutions support integration with, for example, on-premise storage systems such that large or sensitive data sets can be processed using platforms of the cloud providers but using compute entities under control of the cloud providers. This approach adds an extra dimension to the scenario where applications can be tested in the public cloud at smaller scale and real job processing performed on site.

As well as hybrid cloud solutions, there is an increasing interest in multi-cloud solutions in which organizations consume services of more than one cloud provider. There can be multiple reasons for this: very large organizations may not want to be dependent on a single provider for risk reasons, other organizations may see different strengths in different cloud provider offerings and may use, for example advanced hardware of one provider with productivity suite integrations of another.

Although multi-cloud is interesting in principle, at this early stage of Cloud-HPC integration, there are complexities integrating with a single provider: integration with multiple cloud providers to support different capabilities is still some way off. Open on-premises/private cloud solutions such as OpenStack and OpenNebula also offer some capabilities of interaction with different public cloud providers in order to get close to a strongly integrated hybrid-cloud solution.

4.2 TRENDS IN EUROPE

According to survey of Nov 2019 from GMInsights[17] European cloud market is slated to grow at a healthy Compound Annual Growth Rate (CAGR) of >12% and become a >$75Billion market by 2026. Increasingly the Telcos adoption of cloud stack in their 5G rollout across Europe is a major driver of the growth.

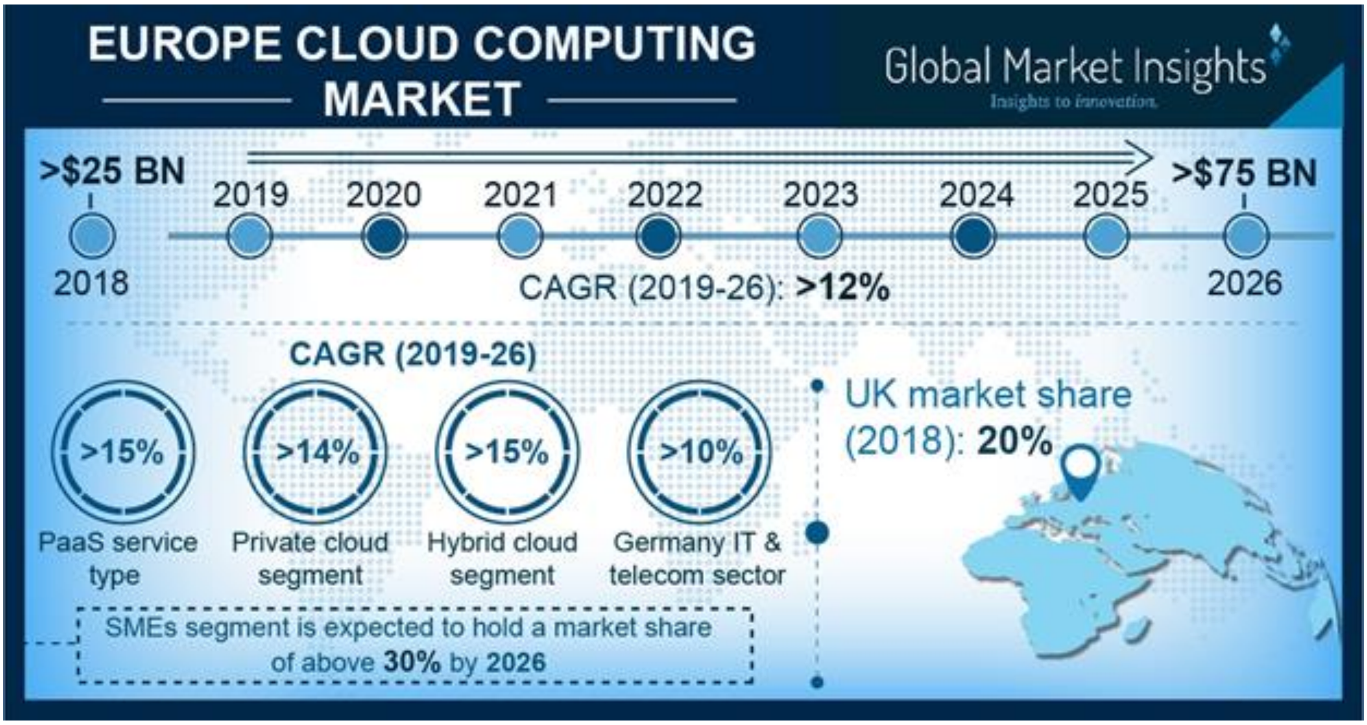

Figure 6 European Cloud Computing Market Trend

Despite a significant growth prospect, European market is still disadvantaged due to the lack of a European mega-provider. American companies such as AWS and Microsoft have been quick on expanding their European footprint.

[17] GMInsights survey: https://www.gminsights.com/industry-analysis/europe-cloud-computing-market?utm_source=globenewswire.com&utm_medium=referral&utm_campaign=Paid_globenewswire [accessed: 19/06/2020]

Lately, Chinese cloud operators such as Alibaba and Tencent have also been expanding aggressively in Europe[18]. There are niche cloud providers appearing, but no one has the scale to compete with the big players.

GDPR has been a significant reason for many companies in EU to embrace the cloud, which may seem counter-intuitive at first. With the onus of data safety and privacy of EU citizen as sole responsibility of the companies, many companies are hesitant to maintain high safety standards in-house, and are relying increasingly on best practices and comfort of mature security practices by large cloud operations by Google, Microsoft, Amazon. Despite growing adoption of cloud solutions, some high-tech SMEs hesitate to use cloud solutions due to fear of industrial espionage.

In order to offset this lack of mega-cloud providers from Europe, initiatives supported by the European Commission are encouraging allied scientific supercomputing centres to pool resources and create pan-European cloud infrastructures. These infrastructures are aimed at supporting science, but increasingly also industrial research and SMEs. In the field of compute-centric IaaS-clouds running virtual machines, the “EGI federated cloud” is probably the largest distributed European infrastructure (when not considering the European branches of the US-based cloud providers mentioned above). This infrastructure[19] was launched in 2014. It is based on OpenStack (and other frameworks) and the hardware infrastructure of about 20 European service providers. Login is based on EGI’s federated AAI, providing the so-called “EGI Check-in” service. On the “storage cloud” side, EUDAT[20] offers large-scale solutions such as B2DROP and B2SAFE (where the latter is rather a federated data-management system than a storage cloud). These are coupled to the B2ACCESS AAI. With EGI, EUDAT and other organisations joining forces in the context of the European Open Science Cloud (EOSC[21]) initiative, Europe’s research sector (prospectively including industry and SMEs) will have access to a rich, independent and open environment with all kinds of IaaS and PaaS services.

5 ANALYSIS OF CONVERGENCE TRENDS

The HPC community observes more and more request for use of the HPC infrastructures for Deep Learning, HPDA using GPUs, FPGAs, and low latency interconnect. The convergence has been speeding due to various factors, some of them are:

- A growing trend is AI for simulation aiming at extracting valuable datasets as inputs for heavy simulations,
- Quantum algorithm simulations,
- Enhanced orchestration for complex workflows in a distributed hybrid infrastructure,
- Data access acceleration,
- Burst buffers - demand for specialised data nodes.

Most of these workflows require additional interactive tasks: 3D remote visualisation, access to Jupiter notebooks, R studio, Databricks, etc.

For lowering the entry barrier to the world of HPC computing without sacrificing execution performance HPC has to implement some recipes from Cloud domain especially about remote user experience in the context of HPC as a Service.

As pointed out in ETP4HC SRA 4, HPC appears more and more as an element of complex workflow (“HPC in the loop” and part of a larger e-infrastructure to realise complex, efficiently managed and orchestrated workflows, including the interfaces of this structure with external devices (distributed and Edge devices), as indicated in Figure 7 below.

[18] CNBC news: https://www.cnbc.com/2019/01/09/cloud-computing-in-europe-salesforce-amazon-are-big-winners---.html [accessed: 19/06/2020]

[19] EGI Federated Cloud docs: https://egi-federated-cloud.readthedocs.io/en/latest/introduction.html [accessed: 24/06/2020]

[20] EUDAT: https://eudat.eu [accessed: 24/06/2020]

[21] EOSC: https://www.eosc-portal.eu/ [accessed: 24/06/2020]

HPC has always advanced science by delivering results only made possible using cutting-edge computer technologies. Throughout the last decade numerical computing has been growing rapidly in many directions: higher fidelity, coupled Multi-physics and multi-scale models; a deluge of observational data from sensors and of simulated data; semi-automatic data analysis and post-processing; uncertainty quantification and newly AI-based models. Combining all these aspects will result in a highly complex application (software) architecture, which is becoming a research topic by itself.

Furthermore, we see an increase in a wide range of new "data driven applications" deploying functionalities such as analysis of very large data sets and machine learning (relying on large data sets for the training phase), which have given rise to a wide range of new applications. And indeed, weather forecast, and seismic events prevention are among this new type of HPC enabled applications.

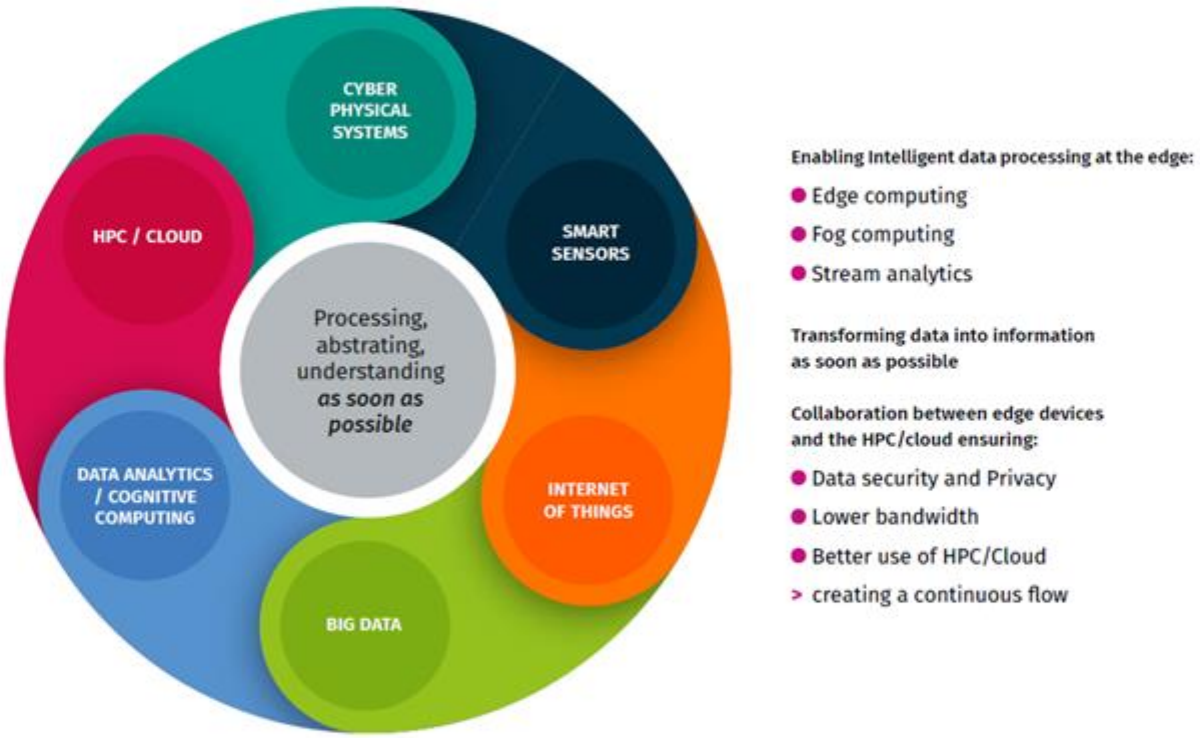

Figure 7 HPC in the loop concept (ETP4HPC SRA 4)

Additional analysis from HPC, Cloud, and Big-data applications perspective shows that there is a clear trend in more and more companies gravitating towards a hybrid approach to the computational resource access. Big Data growth analysis shows that the demand for handling large data volumes in a timely manner will only increase with time. Managing huge volumes of data will be a challenge. Digitalization of basic sustenance tasks including healthcare, agriculture, together with multi-modal data sources, satellites, sensors, smart phone applications, etc. points to a future where innovation that uses a mix of cloud oriented as well as HPC solutions will have competitive advantage, and the commercial sector understands this interplay. This is shown by a clear increase in European HPC centres that are increasingly provisioned and operated by commercial entities as compared to research centres and academic institutions.

Large scale sectors such as agriculture will benefit from heterogeneous resources being deployed at the edge - wherein compute intensive processing can be batch or stream processed in co-located cloud/HPC capabilities in bigger data centres. Large scale data computation using distributed frameworks such as Hadoop and Spark ecosystem of tools is also quite conducive to a harmonious interplay between regional cloud operators as well as European HPC centres.

There will be a need soon to have specialized data nodes (including capabilities offered by inclusion of burst buffers) that can allow better data and computation locality. Increasing market share of interactive data manipulation frameworks: iPython, Databricks, etc. points to a future where cloud like ability in HPC centres (on-demand access to resources, pay-as-you-go finance models) will be increasingly asked for features by enterprise customers. Legal requirements of data safety and protection enshrined in GDPR regulations at European level will add a dimension where European Enterprises will increasingly wish for a solution which is largely European. LEXIS project is clearly aligned in this direction as described in the following chapter next.

6 LEXIS MARKET POSITIONING

When preparing and developing the LEXIS Project, the Consortium had a clear prospective view about the potential impact of the project on the Market. By market, we include both non-commercial and commercial, the Research and Scientific field (the traditional market of HPC since its inception) and the Industry/SME/Start-up field, where Europe clearly lags behind the USA and China in today's landscape, with all consequences in terms of risk on our competitiveness and ability to build our own future. The EU made it a strategic axis of development to increase dramatically the impact of the HPC/Big Data on the global economy and particularly SMEs.

In this framework we will develop in this section our vision for the positioning of the LEXIS Project as a major contributor to the success of the EU policy and our reasons to do so.

This market positioning is still a work-in progress as we are now still in the first half of the LEXIS Project, and will be refined along the coming months to allow us to define a final positioning to prepare the development of the LEXIS platform after the official end of the project.

6.1 SPECIFIC OUTCOMES IN LEXIS & DIFFERENTIATION FACTORS

Innovative at multiple levels, the LEXIS Platform inserts itself in the global trends as explained earlier in this document, but also is built to be future proof by natural agility to scale, to adapt to technical evolutions and end users needs and by being ready for the next steps towards Exascale.

This is based on both the Technical innovations and outcomes in one hand and the implementation of Services provisioning (see Deliverable D4.2 [3]), monitoring and billing in the other hand (see Deliverable D8.1 [4]).

All of this contributes to differentiate LEXIS in this highly competitive and growing global market.

6.1.1 Technical outcomes & innovations

LEXIS is building a powerful computing and data management platform through integrating hardware and software components from HPC, Cloud and Big Data domains, including validation within real-life scenarios (Aeronautics, earthquake and tsunami, weather and climate) and providing easy access to industries, SMEs and Academia through LEXIS Portal [4]. The LEXIS platform lays on a three-layer architecture, being composed of:

- The LEXIS Portal Layer, providing access to the platform for pilots and external users.
- The LEXIS Service Layer, running on top of the infrastructure layer. It includes federated security infrastructure (Authentication & Authorization Infrastructure, AAI), data management (Data Distribution Infrastructure, DDI), LEXIS back-end, and orchestration services (Orchestrator, see Deliverable D4.2 [3]).
- The HPC/Cloud Infrastructure Layer, providing the computing power and data storage space to the upper layers through interactions among HPC and Cloud hardware systems. It is implemented as a federation of multiple HPC providers and data centres.

The well-defined three-layer architecture avoids exposing technical complexity to non-experts, i.e. easy exploitation of the rich computing resources. Meanwhile, the LEXIS platform guarantees the high level security of the federated platform by following "No Trust" principle, which means LEXIS services are all untrusted by default and must be directly registered in the LEXIS AAI or interact with HEAppE (See Deliverable D4.5 [5]).

An innovation that contributes to the future successful building of LEXIS platform is on the co-design activity. While building the LEXIS platform, the co-design process is carried out progressively from investigation, implementation, and integration of technical modules to validation with pilots. In the process, the co-design process demonstrates some good practices, differences between foreseen objectives, and challenges, which are recorded as lessons learned (LL). LL are used for improving implementation of newly proposed approaches, preventing, and minimizing risks and better planning later project phase. A set of LL is available in the Section 3 of Deliverable D2.3 [6]. Through incorporating the feedback from the active co-design process, the LEXIS platform is being improved iteratively and moving towards the LEXIS goals.

The orchestration service is one of the key components of the LEXIS platform. It enables LEXIS users to run their workflows using federated resources. Instead of the commonly adopted static policy, an innovation of the LEXIS orchestration service lies on its dynamic allocation policy (see Deliverable D4.4 [7]). While LEXIS provides distributed infrastructure, the design of the dynamic allocation policy considers the capability of selecting the most appropriate set of resources, i.e., moving data closer to the selected resource to minimize cost. As stated in Section 3 of the current deliverable, urgent computing is demanded nowadays to make decisions fast enough especially in critical situations. LEXIS orchestration service supports urgent computing through strategies such as submitting the task to the least utilized location site. In addition, LEXIS orchestrator service also supports the selection of location sites with special purpose systems including Burst Buffer and FPGA. Thanks to the well-designed LEXIS technical integrations among OpenStack, HEAppE middleware, DDI system and AAI, the LEXIS orchestration service is able to perform the execution of tasks on virtual instances, running jobs on HPC resources and moving data accordingly in a secure environment.

6.1.2 Services provisioning, monitoring, and billing

One of the primary KPIs for the LEXIS project is to support the more rapid access to HPC services generally and more rapid HPC service provisioning in particular.

Key to facilitating this is having a good understanding of the overall resource consumption, as these decisions - pertaining to providing access to new users - is dependent on knowing that there are enough available resources, and in particular, that the total aggregate resource consumption for third party service users does not exceed a specific threshold.

The LEXIS Project has developed a monitoring solution which supports obtaining system usage information from both HPC cloud resources and from HPC scheduling systems. This information is being aggregated to construct a view of resource consumption over different periods of time. The LEXIS monitoring solution also accepts as input, the usage of storage resources on the Distributed Data Infrastructure (DDI) over different time periods and thus this can be monitored. Thinking in regards to the HPC business model and pricing model, the data from the monitoring solution developed could also become the initial baseline information data to be fed into a billing system linked to an Open Source Accounting and Billing system, which is currently under development by Cyclops. The billing solution being used in LEXIS also facilitates detailed credit management, assists in audit reports, and enables billing of services and resources used even if these resources are spread over multiple HPC centres (see Deliverable D8.1 [4]).

6.1.3 How LEXIS project is differentiating itself in the market

At a time where natural trends in the market drive most players providing HPC to fully integrate Cloud services and some added value services it is important to characterise correctly the different approaches and how the LEXIS Project is differentiating itself in this evolving landscape.

Differentiation is managed at multiple levels around the main following axis: technical factors, user experience factors, security and data management, ROI and Costing, Services, ability to evolve and adapt. All these factors will contribute in making the LEXIS Project a significant one impacting the European landscape by dramatically increasing the ability of researchers, industries and specially SMEs to take much more advantage of HPC/Big Data/AI technologies, in better conditions.

MAIN DIFFERENTIATION FACTORS

We can summarize the main differentiation factors as follow:

- **User friendly access**, lowering the barrier to entry, especially for first time / low experienced users,
- Technical multilayer architecture providing a **high level of security and data protection** (compliant with EU standards and strategic goals),
- **Scalability and Flexibility**: LEXIS is designed to extend its capacities by welcoming in its infrastructure additional computing centres and data centres, and provides **readiness for the next Exascale** step,
- Designed to welcome both:
  - Classic computing scenarios,
  - **Event triggered computing scenarios** (also named Urgent Computing).
- Build-in collaborative co-design allowing **continuous and iterative improvement** to the whole platform by using a Lessons Learned approach (continuous exchange with stakeholders and end-users),
- **Data storage, data management and data access** allowing agile, reliable, fast usage in a context where data security is at the centre of the design and Big Data is a main focus of the LEXIS Project,
- Integration of services, software stack and infrastructures,
- Designed to **facilitate a high level of coordination and integration** (when suitable) **with other EU projects and EU infrastructures** to maximise ROI and Impact and to be more efficient in boosting usage by all parties with a specific focus on Industries and SMEs,
- Designed to allow the creation of a **"one-stop" user experience** for existing and future HPC/Big Data/ AI users.

All of this is made possible by the following main technical factors of differentiation:

- Federated data/metadata layer based on iRODS,
- High-performance transfer of Big Data sets, staging mechanisms to HPC machines (EUDAT-B2STAGE, SCP),
- Fine-grained management of rights and data distribution based on iRODS rules,
- Federation with EUDAT Collaborative Data Infrastructure (EUDAT CDI, EUDAT-B2SAFE),
- Transparent, unified view on data at different sites,
- Storage of basic metadata suitable for acquisition of DOIs and PIDs (EUDAT-B2HANDLE),
- Exclusive control of data and rights via HTTP-REST APIs,
- Encapsulation with APIs sanitizes usage patterns and keeps data machine-actionable.

SERVICE DIFFERENTIATION

- LEXIS will differentiate itself in two ways. First by delivering a user experience lowering significantly the entry barrier for new users, and particularly SMEs or teams not experienced in HPC. Secondly by providing users (or application experiments holders) complementary services and software stacks allowing third parties submitting an application experiment to optimise their cost and efforts by finding everything needed in one place, with a dedicated Project Manager to help them coordinate, implement and execute theirs projects, avoiding the usual harassment coming from multiple parties to be involved (this is very important for new users from SMEs).

- The **User Experience differentiation** will be supported mainly by the following factors:
  - Potentially an intuitive and consistent interface to interact with various HPC/Cloud capabilities via the LEXIS Portal.
  - Single point of entry to request HPC and cloud resources - approval system, job submission and transparent manner in how the execution reports, data, invoices are accessed - via a web based user interface instead of plethora of tools that were used before.
  - Ability to provision resources in a dynamic manner - a precursor to cost-controlled HPC job scheduling and move towards a potential prepaid or even pay-as-you-go model.

Differentiation by the portfolio of services available:

- Software stack available from scratch, and that can later evolve by complementing it with additional software stacks from partners and ISVs,
- Expertises available, technical and scientific, from the Pilots to start with and later on from the various partners involved in LEXIS and new ones we intend to propose to join the LEXIS platform, with a perspective to broaden the offer for the development of LEXIS in a second phase, post the official end-of-project date,
- Consulting, Integration and Training services, build-in the LEXIS offer and platform from scratch.

ROI & COST DIFFERENTIATION (Indirect and direct costs, optimisation of resources usage/ROI)

- By simplifying the processes and reducing the number of parties to be involved for helping a project holder to implement and execute its application experiments, LEXIS reduces the direct costs for the user/customer.
- The same approach drives LEXIS to propose solutions reducing the time-to-delivery and the time-to-execution for each project submitted. In doing so industries, SMEs and researcher can get faster the expected results, allowing a significant increase of the ROI by reducing delays.
- A simplification of the legal framework, for the user/client, is also participating to lowering the directs and indirect costs.
- A better optimisation of computing cycles, load balancing and data management is also contributing significantly to reducing the computing direct costs, allowing LEXIS users to take advantage of a more competitive pricing.

SECURITY & DATA PROTECTION

- **BETTER DATA PROTECTION by nature, in line with EU STRATEGIC INTERESTS**: Today's regulations, at international level can guarantee a user to be relatively well protected toward a competitor (public or private) from direct risks. But as the American and Chinese laws demonstrate it, nothing protects users from having their Data being legally spied or used by Intelligence organisations as long as the hosting is directly or indirectly belonging to American interests. LEXIS is differentiating from most players by providing a fully EU based solution keeping its user away from above mentioned risks.
- **SECURITY AT THE HEART OF THE DESIGN OF THE LEXIS PLATFORM**: Complex system have complicated and opaque attack surface. In LEXIS, as resources are spread among participating HPC centres, where the internal organizational security policies are generally divergent among these participating entities. Getting a multi-layer security apparatus operations in a seamless manner is a challenge. LEXIS components' design is being done in a way to facilitate multi-layered approach to access control and security. Within LEXIS security architecture, authentication and role determination happens at the portal ingress, but authorization decisions are taken up at every technical component in the lower stack. This way LEXIS is trying to prevent unauthorized access as every level, from portal to HPC middleware and infrastructure.

6.2 GLOBAL POSITIONING OF THE LEXIS PROJECT

6.2.1 Structure of the market as of today

Today's European market and ecosystem is characterised by the existence of mostly 3 categories of players, in addition to existing or potential HPC users. As described in the report of the European Investment Bank in 2018[22] ("*Financing the future of supercomputing - How to increase investment in high performance computing in Europe*") the different categories of players are presented in the illustration about the HPC Value Chain below (see Figure 8).

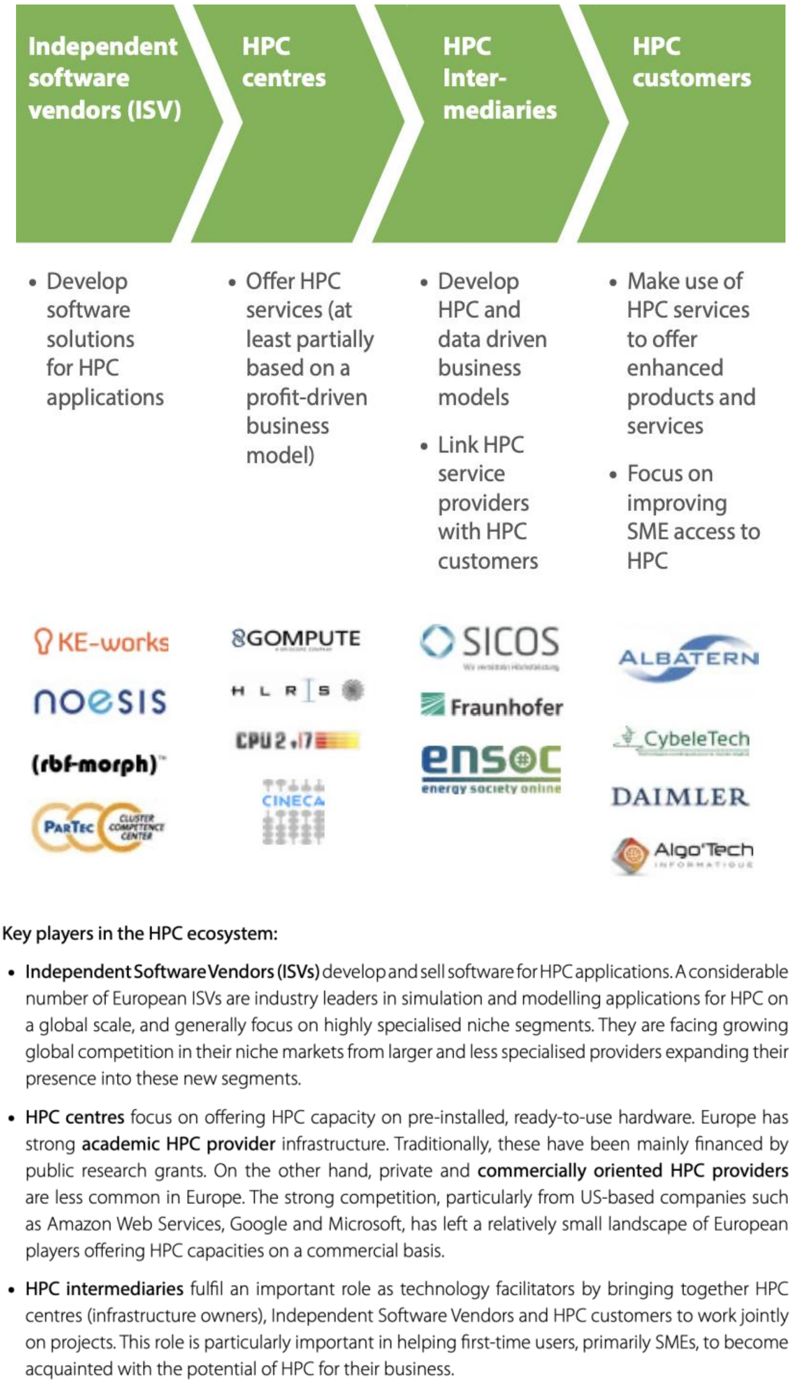

Figure 8 HPC Value Chain and Market structure

[22] Financing the future of supercomputing: https://www.eib.org/attachments/pj/financing_the_future_of_supercomputing_en.pdf [accessed: 25/06/2020]

Most of the offers are focusing only on academic and scientific purposes, with very little room for industries, SMEs.

Most HPC Centres have an average 4 to 5% of their activity with industries and SMEs, with very few exceptions. CINECA in Italy is around 20% but most of it is coming from a single collaboration with the energy giant ENI. HLRS in Germany is around 20 to 25 % but mostly with large companies.

Most players providing directly or indirectly HPC in the EU can be characterised (excluding pure cloud-based players like AWS, Microsoft Azure, Google, or Alibaba) as shown in Table 7:

	Public HPC Centres	**Private HPC Centres**	**HPC Intermediaries**
Main characteristics	• Majority of the HPC capacity is public • Own most powerful computing capacities in the EU • Mainly used for scientific & academic purposes • Technical scalability limited • Heterogeneous technical ecosystem with limited flexibility	• Niche market in EU, few players • Less powerful computing capacities • Focused on niche markets and type of application experiments • Business model based on off-take agreements • Technical scalability limited • Most players have a technical architecture adapted to very limited specialised type of usage	• Facilitators between HPC centres and users/clients • Mainly public entities under the umbrella of HPC centres • Mostly organised by sector, or geographical area • Can access highly skilled professionals but with significant limitations
Capacity available for Industry & SMEs	• <20%, average in the EU closer to 5%	• Up to 100%, but each player is usually limited to a niche approach, with little flexibility	• Supposedly dedicated to industries and SMEs but in reality mostly depending on public funding, project grants and membership fees (mostly from Very large corporations, not from SMEs, i.e. CINECA and ENI in Italy)
Comments	• Lack of viable business model, and legal limitations blocking a serious development toward industries, SMEs, lack of technical flexibility	• Costly, lack of technical flexibility and in lack of value-added services	• Similar to HPC public centres with similar limitations

Table 7 Classification of HPC resource providers in Europe

- **ISVs:** (Independent Software Vendors) their business model and ultra-specialisation makes most of them in Europe, even the most successful ones, unable to grow at large scale, making them fragile. As a consequence, they quickly tend to be acquired by major large players from America or China, making the European ISVs lagging behind non-European ones (not in term of technology but in term of ability to grow and finance their development and technical evolution). Another consequence is that HPC users can access ISVs solutions via the big American (and more and more Chinese like Alibaba) players like AWS, Google, Azure, and less via EU based HPC providers. Because scaling is vital for ISVs, as long as the EU market is in lack of a scalable, flexible cloud based infrastructure for HPC/Big Data potential users will depend on third parties that mostly are non-Europeans and at the expense of the EU competitiveness , strategic interests and EU SMEs.
- **GLOBAL MAJOR PLAYERS: AWS, AZURE, GOOGLE, ALIBABA &Co:** these top world global players, natively cloud based, are taking market shares in the supply of HPC computing cycles, but:
    - Most of the value added services proposed come from approved partners (hence multiple contractual layers for each project, high level of costs) and have very little capacity to provide scientific expertise that will still be to be found mostly within HPC academic & research players.
    - Infrastructure services offered propose a competitive pricing, but the value-added services required by most industries and SMEs are still at high cost and difficult to handle by most potential users keeping high the barriers to entry to most potential users.
    - Data issues: no protection from intelligence from state players particularly the USA, loss of strategic independence, risk on data protection in a world competition supervised today by state policies and strategies.
    - Developing solutions within their ecosystem drives to a situation where the potential move of applications/data to another environment is very difficult and costly: users are de facto very quickly locked in their platforms with real life limited options to change.

Developed by the European Investment Bank and Roland Berger Consulting[23], the above description of the structure of the market in terms of players demonstrates:

- Potential users face the challenge to manage numerous contractual relationships (ISVs, HPC Centres and Intermediaries) to help them develop HPC based solutions for their R&D needs or for exploitation. This is a source of unnecessary complexity, increased timelines (Time-to-compute), and major costs for any project.
- Each category of player tends to develop a vertical offer, helping to bring to the market more sophisticated and specialised solutions, but at the cost of great lack of interoperability, lack of flexibility, higher costs for the end-users. As a consequence the EU players are not in the best position to take advantage of the increased market and needs for HPC/Big Data/AI, leaving the space to mostly American and Chinese solutions and keeping the EU economy and competitiveness more and more lagging behind.
- The global ecosystem is lacking efficiency by lack of coordination and flexibility, and this generates a higher level of costs (or lack of ROI) making it harder to make SMEs and even large industries accessing and using these technologies due to higher costs.
- The lack of scalable, flexible cloud based solution at EU level is blocking the ISVs and specially the EU ones, despite their level of excellence, to broaden their impact toward the industry, SMEs and research organisations by making it very difficult from economical point of view to participate to lowering the barriers to entry and take advantage of a much larger market.

[23] Financing the future of supercomputing: https://www.eib.org/attachments/pj/financing_the_future_of_supercomputing_en.pdf [accessed: 28/06/2020]

As a consequence of this market structure, there is a significant lack of adoption by Industries and SMEs in the EU at the difference of USA and China:

- The diversity of players mixed with their multiplicity as the cause of an entry barrier excluding most industry players. Only public funded institutions can really take advantage of the HPC / Big Data infrastructures in the EU (Difference with USA and China).
- On top of direct costs, the indirect costs coming from the skills gaps are worsening the situation and limit adoption of HPC.
- Complexity of the legal frameworks and multiplicity of players to be involved in projects make the situation hardly manageable by SME's, start-ups and leave the advantage to more reactive competitors (China and USA).

It is to be noted that 2 other factors are massively impacting the lack of adoption by industries, SMEs and indirectly also research:

- Industries/SMEs have a real difficulty to finance their projects in HPC/Big Data, hence a low rate of adoption.
- Investing in HPC projects does not provide enough quality collateral to entice banks (lenders) and equity investors (shareholders, financial investors) to support what is de facto considered as a high risk project. This phenomenon is largely and extensively described in the survey published by the European Investment Bank in partnership with the Roland Berger consulting firm[24].
- HPC/Big Data/AI projects very often have a direct and strong impact on the competitive situation and on the Business Model of companies. It is very difficult, and specially for SMEs to evaluate this impact and to assess its consequences on the business development. This difficulty is mostly linked to the absence of skilled resources on the evaluation and modelisation of disruptions in business models and as a consequence, the cost of specialised consulting is to be added to the cost of the HPC/Big Data/AI project.

6.2.2 LEXIS project market positioning: a transverse approach to change the paradigm for industry and research

From its inception the LEXIS Project has been driven with the perspective to innovate significantly in order to propose a platform of services able to disrupt the traditional approach to HPC as we know it in Europe. The gap between Europe and the USA and now also China is only the visible sign that we needed to change the paradigm about our practice of HPC, if as Europeans we want to stay at the top in the world competition, both for Research and for the Economy and Society.

In the framework designed by the European Institutions, the LEXIS Project is breaching the usual vertical approach typical from HPC Centres, Intermediaries and ISVs in order to capitalise on its existing infrastructures, HPC Ecosystem on one hand and on the second hand in order to use a strong level of innovation to re-think the practice and the development of the HPC within research and Industries.

In doing so LEXIS is building a transversal positioning (by opposition to the vertical approach above mentioned) allowing an unprecedented integration and enhancement of computing capabilities, software stack, orchestration, data management, flexibility looking toward the future and the needed evolutions up to the next to come Exascale era.

From practical point of view LEXIS positioning is toward dramatically lowering the barrier to entry for every new user in HPC, from the research teams to large industries or SMEs and facilitating a very large expansion of the rapid adoption of the HPC services at every level of our society, with the perspective to allow the European society to stay on top of the international competition, preserving and improving our competitiveness, independence and security.

[24] Financing the future of supercomputing: https://www.eib.org/attachments/pj/financing_the_future_of_supercomputing_en.pdf [accessed: 28/06/2020]

As a consequence, the LEXIS Project positioning relies on:

- Innovations proposed by LEXIS as a way to significantly disrupt the practices, the accessibility and the ability to adapt permanently for HPC/Big Data and their existing and potential users,
- The technology innovations as the base of a layer of service innovations able to boost the competitiveness of research, industry in an ultra-competitive environment (society, geopolitical landscape, mid and long-term evolution of the economy),
- Service innovations, based on the technology innovations of LEXIS as the tool for significantly increasing the adoption by the industry players and the SMEs, start-up included,
- A platform leveraging the resources in the EU by allowing a multi-level cooperation and federation for flexibility, scalability and reactivity, and increasing the ROI of existing and future Infrastructures, including at Exascale level (ability to welcome new partners, computing centres, data sources, etc. and ability to partner, and sometime integrate with other projects in the EU),
- Preparing the future rather than repeating the past.

The prospective outcomes of the LEXIS project drive us to propose a positioning based on:

- Ease of access,
- Data and scaling of datasets,
- Processes,
- Scaling of computing, flexibility to welcome new infrastructure including at Exascale levels,
- Accompanying users to remove major hurdles for non HPC professional and focusing on making their application experiments working efficiently at a reduced TCO and Operations Cost,
- Preserving key interests in IP, security, and confidentiality within the global interest of the EU,
- Preparing the community to use the next level (Exascale) in a near future.

As a consequence, the **global positioning of LEXIS** can be visualised as follow (**simplified representation**):

	LEXIS PROJECT	HPC Centres in the EU (Private)	HPC Centres in the EU (Public)	HPC Intermediaries in the EU	ISVs	Global Players (AWS, ALIBABA, GOOGLE, AZURE...)
Data Protection from Non EU players and Intelligence gathering organisations (under American Law)	YES	Partially	YES	Partially	Partially	NO
HPC Computing Services	YES	YES	YES	NO	NO	Partially
Big Data Services & Optimised Data Management	YES	Partially	Partially	NO	NO	YES
Cloud services	YES	Partially	Partially	NO	NO	YES
Software Stack	YES	Partially	Partially	NO	NA	YES
Added Value Services, Consulting & Software developement	YES	Partially	Partially	Partially	Partially	Partially
Large execution Scalability and heterogeneity of architectures	YES	YES	YES	NO	NA	Partially
Optimised orchestration in a distributed environment	YES	NO	Partially	NA	NA	Partially
Linking HPC service providers with HPC customers & users	YES	NO	YES	YES	NO	NA
One Stop Shopping experience & Billing for all components of a computing project	YES	Partially	NO	NO	NO	YES
User Friendliness	YES	NO	NO	Partially	YES	YES
Flexibility	Partially	YES	YES	Partially	Partially	Partially

Table 8 Global positioning of LEXIS project

The above table aims at visualizing the global positioning of LEXIS, simplified, its key differentiations with the other players and the strategic direction chosen by the LEXIS Consortium. The final positioning and its details will be refined for the end of the project in Q4 2021.

The main outcome of the LEXIS Project after the end of the project is to:

- Plan the set-up of a structure (Spin-off) to address the Research and Industry market by capitalising on the LEXIS platform. This spin-off will give the legal framework and business structure to allow a much more efficient interface with researchers, industries, SMEs at a lower cost, with a much more efficient relation with end-users and clients.
- Extend the number of partners, particularly Computing Centres and Data providers to:
    - Increase computing capacities on a diversity of architecture (CPUs/GPUs/FPGAs),
    - Create a diverse and attractive software stack at the benefit of end-users,
    - Allow a broader access to an increasing range of data sets,
    - Improve service levels and efficiency, thus lowering costs and barriers to entry,
    - Manage an efficient coordination and optimisation of resources with other European projects and infrastructures.
- Set LEXIS as a “one-Stop-Shopping” point for any industry and research unit looking at developing HPC /Big Data projects in order to increase the impact in favour of EU research and EU economy via its research labs, industries and SMEs.

This positioning is today a work in progress and will evolve in the next coming months until the end of the project mainly due to 2 factors:

- Being only at month 18 of the project (when LEXIS is ending 31st December 2021), some elements of the positioning will have to be refined due to the final evolutions of current developments,
- It is expected that some significant changes in the HPC ecosystem, under the impulsion of the EU and the competition, may bring the LEXIS consortium to adjust and refine the final positioning.

Nevertheless the strategic direction chosen by the LEXIS Consortium should not be affected and only refinements and adjustments are expected, allowing to finely tune the anticipated follow-up of the LEXIS Project via the preparation of a structure in a spin-off approach for allowing LEXIS to be effectively a key player at the service of Research and Industry.

7 SUMMARY AND CONCLUSIONS

This deliverable is first in a series of two that analyses the market for trends in HPC, Big Data, and Cloud ecosystem in Europe. In this deliverable D9.5, the state of the ecosystem was analysed from several angles: market growth trends, application trends, convergence signs, as well as understanding the European context from a prism of a larger global view. While, in isolation, each of the sectors - one application centric, and two infrastructure centric, shows a healthy CAGR in Europe and global context, what is interesting is the fact that lately there is a sign of convergence among HPC as well as the Cloud actors in terms of services and capabilities they offer. Traditional HPC centres that used to be homogeneous, are seeing demands for heterogeneity even from their traditional user-base. Lack of investment in HPC sector in Europe is clearly visible when compared to Asia and the Americas, but engagement by commercial entities in this space is heartening. Lack of a European mega-cloud operator still remains a concern. GDPR, counter-intuitive for many, has acted as another catalyst for adoption of cloud by enterprises in Europe.

The findings in this deliverable places LEXIS project on a sound footing. LEXIS appears to be favourably placed and appears to have all the making of a disrupter in the market with a potential to speed up the convergence in HPC and Cloud ecosystems. Although, much remains to be seen and depends on how the execution of LEXIS proceeds in the 2nd half of the project.

The next deliverable D9.11 in this series, which is due at the end of the project, will analyse the market trends one more time, and reassess the LEXIS goals in that revised context.

REFERENCES

[1] J. L. Gustafson, The End of Error: Unum Computing, Chapman & Hall/CRC Computational Science. Taylor & Francis., 2015.

[2] S. Williams, A. Waterman and D. Patterson, "Roofline: An Insightful Visual Performance Model for Multicore Architectures," *Communications of the ACM,* vol. 52, no. 4, pp. 65-76, 2009.

[3] LEXIS Deliverable, *D4.2 Design and Implementation of the HPC-Federated Orchestration System - Intermediate.*

[4] LEXIS Deliverable, *D8.1 First Release of LEXIS Portal (will include report).*

[5] LEXIS Deliverable, *D4.5 Definition of Mechanisms for Securing Federated Infrastructures.*

[6] LEXIS Deliverable, *D2.3 Report of LEXIS Technology Deployment - Intermediate Co-Design.*

[7] LEXIS Deliverable, *D4.4 Definition of workload management policies in federated cloud/HPC environments.*