Report on the consultation with industry representatives on research challenges in cloud computing for H2020 Work Programme 2018-2020

Workshop date: 4 February 2016

Final version

12 February 2016

Rapporteur: David Griffin

SECTION 1. EXECUTIVE SUMMARY

The *Software & Services, Cloud* unit (DG CONNECT E.2) invited representatives of industrial stakeholders in the cloud computing sector to assist in the identification of the main research and innovation challenges in the sector in the context of the H2020 ICT Work Programmes 2018-2020. A meeting was held on 4 February 2016 where the contributions were presented and discussed.

Research and business-related challenges were identified in the following areas:

- Cloud management platforms linking compute, storage and networking
- Fog and edge computing linked with the Internet of Things
- Data-centric cloud computing
- Cloud federation, hybrid clouds and cloud scalability
- Enhanced and automated cloud deployment methodologies and tools
- Novel cloud-native applications
- Migration of applications and services to the cloud
- Cloud and data centre infrastructure
- Standards and openness in cloud computing platforms and data
- Privacy, security, trust and accountability in cloud computing

In addition comments and suggestions on the H2020 Work Programme and funding instruments were raised by the participants.

TABLE OF CONTENTS

SECTION 2. MEETING REPORT

2.1 Introduction

The *Software & Services, Cloud* unit (DG CONNECT E.2) invited representatives of industrial stakeholders in the cloud computing sector, including cloud service providers, SMEs, research centres and equipment manufacturers to assist in the identification of the main research and innovation challenges in the sector which would require funding at EU-level in the context of the forthcoming H2020 ICT Work Programme 2018-2020.

Eighteen organisations were involved in the consultation. Thirteen written contributions were received prior to a meeting held on 4 February 2016 where the contributions were presented and the identified research priorities were discussed. Sixteen presentations were made during the meeting. Appendix A lists the written contributions and meeting presentations.

The remainder of this report summarises the research challenges and remarks on the H2020 work programme as raised by the written contributions and as discussed in the meeting.

2.2 Cloud management platforms linking compute, storage and networking[1]

- Use of software-defined networks (SDN) and software-defined infrastructures (SDI) to integrate the management of networks, including overlay networks, with compute and storage capabilities and to federate management across data-centres, core networks and multiple cloud providers. Allocation and control of bandwidth and QoS on demand for distributed cloud services and applications. Use of higher level abstractions of underlying resources to simplify and improve the effectiveness of management. Automated platforms and dashboards for bandwidth and QoS management.
- A unified operating environment for coherent end-to-end management and orchestration of cloud-based infrastructure and platforms across centralised, distributed and hybrid clouds, the networks interconnecting providers and out to sensors and cyber-physical systems in the Internet of Things.
- Management of mash-ups of virtualised telecommunications functions (network functions virtualisation, NFV) and application/service functions. Applications linked with operator services for monitoring, OSSs and BSSs.
- Management platforms for low latency applications to improve performance and dynamic adaptation of applications.
- Management platforms for automated configuration and deployment, selection of cloud providers and locations, optimised placement of services across geographically distributed cloud locations to assure performance and QoS, balance load over clouds and networks, improving energy efficiency. Management of the interconnection of DCs, exchange traffic, close to users to reduce latency and optimise traffic delivery between data centres including micro data centres out to edge computing platforms and to end users. Support for reconfiguration and migration between cloud providers, without loss of SLAs or security.
- Cost prediction and optimisation managing resource consumption/reservation.

[1] Some of these research topics have a potential overlap with the activities of Unit CNECT E1.

2.3 Fog and edge computing linked with the Internet of Things[2]

- Deployment of applications and services across the spectrum of centralised and distributed clouds to point of presence (PoP)-based micro-clouds including edge/fog computing, ambient connected devices, cyber-physical systems and the Internet of Things. Linking with ad-hoc clouds and mobile devices, considering admission control, security and low-power issues. Models that integrate cloud resources across this spectrum are likely to be low cost and scalable at the centre yet increase flexibility, reduce latency and save bandwidth at the edge.
- Research into scalable and vendor-independent platforms for the deployment of cloud-based IoT applications and services that deal with sensor volatility, reliability and heterogeneity. Software development environments covering a wide range of computing devices from mega- to micro-watt. Addressing security requirements and privacy of data collected from connected objects. Vendor-independent standards for the integration of cloud computing with mobile devices, CPS and IoT.
- Deployment of edge-processing functionality to reduce the need to transmit large amounts of raw data back to centralised cloud locations and reduce latency. Real time data analytics at the edge, dealing with the scalability required by billions of connected objects. Support for low latency, dynamic, services at the edge.
- Admission control and software placement issues on mobile resources optimised for their availability and performance.

2.4 Data-centric cloud computing[3]

Research challenges

- Platforms for data-intensive and data-centric computation moving away from compute-centric models, providing tighter control of where data is processed and stored. Using databases as the basis for PaaS delivery with intelligent interaction with underlying IaaS tools for managing data requirements.
- Support for cloud-based big data applications:
    - Techniques for undertaking data-analysis close to the data rather than moving large-scale data to distant processing nodes. Flexible and dynamic pre-processing of data. Support for analytics of real-time data at large scale, such as that generated by earth observation satellites.
    - New models and standards for open data access and the federation of data between data-centres, across distributed clouds and between cloud providers. Methods for community data updates and sharing of data at peta-scale. Management of data deletion in large-scale or shared datasets.
    - Approaches for efficient access to large datasets including the development and use of cross-domain search metadata. Support for integration of heterogeneous mixes of structured and unstructured datasets with incompatible meta-data.

[2] Some of these research topics have a potential overlap with the activities of Unit CNECT E1.

[3] Some of these research topics have a potential overlap with the activities of Unit CNECT G3.

    - Coordination of data replication, considering the type of data, the frequency of access, etc. Auto scaling and self-management of compute and storage resources for big data applications.
    - Business models for data analytics on open data.
- Software techniques, platforms and tools to supporting memory-driven computing and memory-centric processing, making use of storage-class memory. Techniques for the efficient implementation, operation and federation of in-memory databases.

Experimentation challenges

- Trials across multiple IaaS providers for experimentation and prototyping of federation of memory-centric data.

2.5 Cloud federation, hybrid clouds and cloud scalability

Research challenges

- Federation between cloud providers moving towards extremely large scale cloud infrastructures, incorporating hyperscale computing in large data centres with high density compute, storage, memory and networking resources.
- Hybrid clouds blurring the distinction between local, private clouds and distributed, public clouds. Techniques for offloading of processing and storage from local to cloud resources, including the dynamic selection of offload target, considering also mobile cloud users.
- Management of security and quality of service in hybrid and federated clouds.
- Cross-provider orchestration techniques and tools.

Business-related challenges

- Interworking between, and federation of, clouds is not a new research topic. It has been identified as a research priority in several previous work programmes. Many of the barriers to seamless federation may not be wholly technological, but the gaps are as yet difficult to identify due to a lack of real-world deployments of applications across multiple cloud providers.
- Deployment models moving above IaaS to PaaS may encourage more interworking and federation as common platforms are deployed over heterogeneous clouds.
- New business models and open standards for federated and hybrid clouds are required, including business models for SMEs innovating in the cloud. Cloud marketplaces that address and harmonise quality, compliance and security between cloud providers.
- Non-European, large cloud providers threaten small to medium sized cloud service providers. There is an opportunity for the development of European PaaS and SaaS products to address the dominance of non-EU providers in the IaaS marketplace.

2.6 Enhanced and automated cloud deployment methodologies and tools

- Automated software deployment and management of applications in the cloud including orchestration and placement of functions across multiple clouds, exploiting the features and capabilities of low-cost and scalable centralised cloud infrastructures at one end of the spectrum and the increased flexibility, reduced latency and bandwidth of edge/fog computing at the other.

    - Tailored to application requirements with workflow-driven processing and workload management at large scale.
    - Adaptation of deployment over time based on data analytics of service behaviour and performance metrics.
    - Improved performance/QoS.
    - Cost prediction.
- Increased use of PaaS based on simplified platforms and APIs:
    - In general and for edge/fog computing integrating IoT and CPS in particular.
    - Users define aspects such as what, when and at what cost tasks should be performed rather than managing detailed interactions with low-level infrastructure resources.
    - Cloud-native development – open platforms and models; components and micro-service integration in larger services.
- Open software development models that move away from monolithic service deployments to predictable compositions of micro-services, based on agile DevOps, continuous operations and application evolution, using PaaS.
- Intuitive APIs for developers for deployment on abstract and virtualised cloud resources which hide multi-cloud heterogeneity; and simplified interfaces for the control and management of applications by end users without detailed technical knowledge of the underlying infrastructure. APIs enabling richer interactions with the cloud infrastructure provider – for example to match resources to the type of data being managed, differing, for example, in access frequency or structured vs unstructured.
- Tools to enable migration of active, live databases between cloud locations using techniques at a higher level than raw VM migration.
- Storage and data management tools at scale for big data and IoT.
- API-driven provisioning of expensive resources such as super and quantum computing.

2.7 Novel cloud-native applications

- Support for mission-critical and demanding applications.
- Data analytics for cloud applications.
- Quantum cloud.
- New AI-based applications.
- Specialised clouds built on public cloud providers for vertical applications, making use of specialised hardware.
- Cloud-only terminals/devices.
- Seamless and universal cloud accessibility including through high capacity and low latency access networks, such as 5G.

2.8 Migration of applications and services to the cloud

- Broader adoption and increased take-up of cloud computing:
    - Tools and techniques for the smooth adaptation of confidential, safety-, mission- and business-critical services.
    - Large organisations/enterprises are often unwilling to put data into the public cloud.
    - Further research is required into the social implications of cloud take-up.
    - There is a lack of tools and business models for migration of existing applications to the cloud, especially for public sector organisations.

- Migration of legacy applications to the cloud (see also challenges related to simplified tools and software deployment platforms in section 2.6).

2.9 Cloud and data centre infrastructure

- Greater exploitation of specialized hardware in data centres, including GPUs, FPGAs, NPUs, quantum computing and energy-efficient hardware.
- Further work is required on the use of next generation virtualization, including containers and unikernels to improve efficiency of deployment and operations and security and multi-tenancy isolation.
- Secure elastic compartments in the cloud offering equivalent privacy levels as on-premises cloud infrastructure.
- Development from the ground up of a native open source cloud and data centre operating system to address the deficiencies of Linux due to its development as a general purpose operating system rather than being tailored for cloud computing.

2.10 Standards and openness in cloud computing platforms and data

- Migration tools, further standards and open APIs to avoid vendor lock-in to cloud providers and to facilitate interoperability between providers in multi-cloud environments, including techniques for content and application portability, substitution and reusability.
- More openness and provider-neutrality moving away from dependency on the few incumbent large cloud providers.
- Establishment of common metrics for SLAs, performance and trust to define and quantify pre-accredited capabilities of cloud providers.
- Standards for data exchange and access/processing policies.
- Cross-border and free movement of data between providers and geographical locations which complies with EU legislation.

2.11 Privacy, security, trust and accountability in cloud computing

Research challenges

- Consistent service capability/metric description between cloud providers including scientific assessments of privacy and security risks. Enforcement of SLAs related to security.
- Increased customer/user control of cloud deployments and of data management, including the assured deletion of data. Especially in federated and distributed clouds in multiple locations under different jurisdictions.
- Management of sovereignty of data. Techniques against unauthorized analysis or linkage of data. Techniques to enforce appropriate accountability and audit trails of data access, linkage and processing.
- Modelling of security risks, such as those posed by the analysis of meta-data and information inferred by analysing side-channels (e.g. monitoring of CPU, memory or network resources in multi-tenant environments to infer cryptographic activity).
- Identity management across federated clouds and techniques to link data access to identity.
- Compartmentalisation and isolation for multi-tenancy data centres and clouds.

- Increased security of cloud applications and data through the further research into the development and use of homomorphic encryption, post-quantum cryptography, quantum key distribution and block-chain in real cloud deployments.
- Security in post-quantum computing world: techniques against breaking of encrypted data using quantum computing.
- Development of trust models based on transparency and privacy governance.
- Research into mitigation and prevention of new malware attack patterns for cloud-based applications and data.
- Establishment of trusted software environments for the compute and storage capabilities of telecommunications equipment and devices that are integrated as part of decentralised cloud infrastructures.

Business-related challenges

- Challenges related to privacy and trust are not purely scientific or technological topics and the involvement of legal experts and regulators to collaborate with engineering and development personnel in common research projects will assist in meeting these challenges.
- Regulations for privacy are not the same in all jurisdictions within the EU and world-wide. Cloud-based systems need to take this into account to comply with local legal requirements, but also the expectation of privacy and security of the users of cloud computing in hybrid, federated and multi-cloud deployments across multiple regions.

2.12 Remarks on H2020 research and innovation projects

- A more focussed and commercially realistic approach is suggested for H2020 WP 2018-2020. More dialogue between research and industry is required to ensure research problems are appropriately addressed.
- Future projects should build on previous results and not revisit the same problems or rehash existing work.
- Some research challenges identified are not purely cloud computing issues: security, support for IoT, big data analytics and the management of networking technologies, for example. Cloud-specific research challenges need to be identified and appropriate collaboration with researchers in related areas should be established for cross-topic themes.
- Further dissemination and exploitation of research project results should be encouraged to facilitate project results becoming market realities. A broader evaluation of project success is required. Commercialisation can take 12-15 months to realise after the end of a project, which is not necessarily visible in project final reviews. The use of technology readiness levels (TRLs) in H2020 projects does not guarantee exploitation and commercial adoption.
- Medium-large pilots would encourage realisation of project results on commercial platforms. Provision of platforms for SME deployment of applications. More venture capital-like funding is requested for focussed research and technology transfer.

SECTION 3. ACRONYMS AND ABBREVIATIONS

5G	Fifth Generation (mobile networks, or wireless systems)
API	Application Programming Interface
BSS	Business Support Systems
CPS	Cyber-physical Systems
CPU	Central Processing Unit
DC	Data Centre
DevOps	Development and Operations
DG CONNECT	European Commission Directorate General for Communications Networks, Content and Technology
EC	European Commission
EU	European Union
FI	Future Internet
FPGA	Field-programmable Gate Array
GPU	Graphics Processing Unit
H2020	Horizon 2020
IaaS	Infrastructure as a Service
ICT	Information and Communications Technology
IoT	Internet of Things
ISP	Internet Service Provider
IT	Information Technology
NaaS	Network as a Service
NFV	Network Function Virtualisation
NPU	Network Processing Unit
OSS	Operations Support Systems
PaaS	Platform as a Service
PoP	Point of Presence
QoE	Quality of Experience
QoS	Quality of Service
R&D	Research and Development
SaaS	Software as a Service
SDI	Software-defined Infrastructure
SDN	Software-defined Networking
SLA	Service Level Agreement
SME	Small and Medium Sized Enterprise
TRL	Technology Readiness Level
UI	User Interface
VM	Virtual Machine
WP	Work Programme

APPENDIX A. LIST OF CONTRIBUTORS

Industry representative	Written contribution	Presentation
ATOS	X	X
BT	X	X
CloudSigma	X	X
DLR & CISPA	X	X
Ericsson	X	X
ESA	-	X
HP Enterprise	X	X
IBM Research	X	X
INDRA	X	X
Interxion	-	X
Nokia	X	X
Orange Labs	X	X
SAP	X	X
SixSq	-	X
Telefonica	X	X
T-Labs	-	X
Software AG	X	-