HPCS 2019 Demo

Data Harbours: Computing archetypes for digital marketplaces

Reginald Cushing, Lu Zhang, Yuri Demchenko, Cees de Laat, Paola Grosso
Systems and Networking Lab (SNE)
University of Amsterdam
Amsterdam, Netherlands
{r.s.cushing, l.zhang2, y.demchenko, delaat, p.grosso}@uva.nl

DEMO ABSTRACT

The project of Data Logistics for Logistics Data (DL4LD) aims to help the Dutch logistics sectors with IT tools that promotes digital business processes, with particular support for the trustworthy sharing of sensitive data [1]. Specifically, DL4LD shows how to establish, digitally, sufficient trust to execute a data-transaction between two ad-hoc logistic partners with the concept of Digital Marketplaces (DMP) [2][3]. This includes the digital negotiation of legal contracts for data sharing and data operations. DL4LD also shows how digital contracts are input for automatized setting up of the required digital infrastructure. Two concrete use cases in DL4LD are privacy-preserved healthcare data sharing and collaborations among normally competing airline companies for aircraft maintenance prediction.

Our demo demonstrates the prototype of trust-worthy and policy-driven data sharing among distributed parties with the concept of DMP. More specifically, we show a concrete real-world use case of airline companies in project DL4LD. Three distributed data harbours, KLM, Air France and University of Amsterdam (UvA), are collaborating of federated computing for a common goal, e.g. more accurate algorithm result. *How can these competing parties, such as airline companies, to facilitate policy-driven data sharing with concept of DMP*?

To answer this question, we investigate a federated computing platform with container technology. The platform is molded into different archetypes based on trust relationships between organizations. For each archetype, there is a corresponding contract which contains rules about how the data and/or compute objects can be transferred, accessed and used within the compute infrastructure. Our demonstration currently supports four different archetypes. Also, we have five different applications, each of which is a distributed pipeline. Only the operations complying to the selected archetype are permitted.

KEYWORDS - Open Data Market, Data Marketplace, Trusted Data Market, Industrial Data Space, Data Economics, STREAM Data Properties

OUTLINE

1. Use case

The demo demonstrates the policy-driven data sharing with airline use case. Airline companies, e.g. KLM and Air France, would like to predict the need for aircraft maintenance by operating AI/ML algorithms on the aircraft data. It is commonly known that more reliable prediction result is achieved by better availability of training data. It is beneficial for those companies to gather the data of same aircraft type for collaborative computing. But these companies are competing with each other and normally have a preferred data sharing policy for privacy and confidentiality consideration.

2. Components of digital marketplace architecture

The architecture of a digital marketplace is an active research field and has many components, which is illustrated in Figure 1. Here we explain each of components in detail in the following.

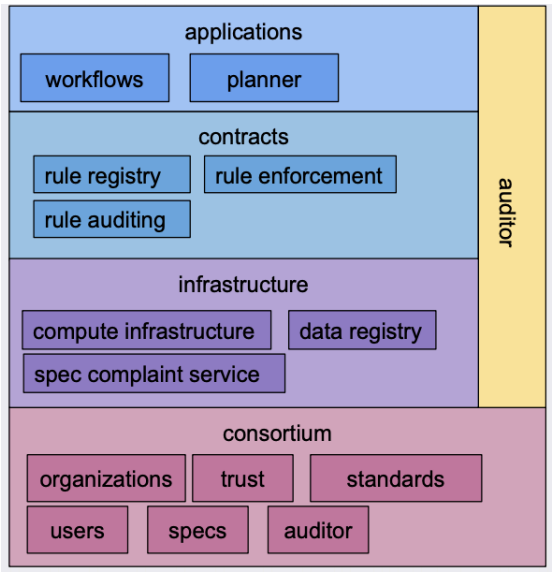

Figure 1. Components of a digital marketplace.

- **Consortium**: is an initial document which brings together organizations that wish to collaborate. It defines static information such as keys to identify parties.
- **Infrastructure**: A single domain organization infrastructure that securely hosts data, compute containers and, optionally, compute infrastructure. We dub this infrastructure a data harbor. A harbor implements a set of protocols that allows it to interact with other harbours.
- **Contracts**: Are a set of rules that are shared amongst participating harbours which describe how objects (data, compute) can be traded between harbours and who can process data. In its simplest form is a 7-tuple which *binds a user, data object, compute container, contract, consortium, harbor, and expiry date*.
- **An application:** Is a distributed pipeline which can make use of several contracts. The combination of application and contract defines the archetype of the computation i.e. how data and compute are moved to effect computation.
- **Auditor**: A trusted entity that collects audit trails for use in litigation of policy violations.

3. Demonstration in action

1) Archetype Selection:

The demo supports four different archetypes to allow users to choose from and each of them has a set of concrete rules for governing the asset sharing. All the transactions within the platform must comply to the rules of the selected archetype, otherwise the transaction would not be permitted and cannot be finished successfully. It's also possible to select multiple archetypes simultaneously to combine the rules together.

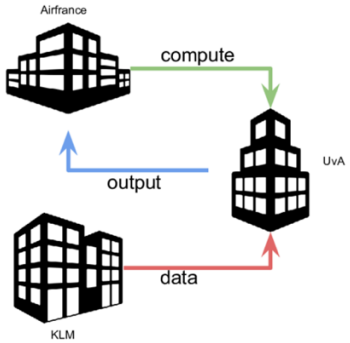

Figure 2. An example archetype of digital marketplaces supported by the demo.

Figure 2 illustrates one of the archetypes. Air France would like to perform its algorithm on the aircraft data from KLM. But they do not trust each other, so they employ a trusted third party UvA and send their compute and data to it. UvA executes Air France's algorithm on KLM's data and sends the result back to Air France.

2) Scenarios Simulation:

After selecting an archetype, the different harbours conduct several transactions, e.g. send data or compute from one location to another, to simulate a concrete application scenario. There are five different applications available in our demo. Figure 3 shows the screenshot of the demo when operating transactions with one of the application. For the scenario to succeed, all those transactions must follow the contractual rules determined by the selected archetype in the previous step.

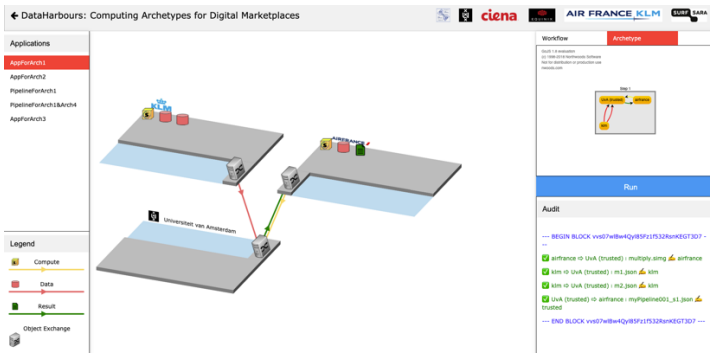

Figure 3. The demo simulates a scenario of federated computing. Data and compute meet in a 3rd party and transfers the result back to algorithm provider.

3) Transaction Protocols:

The transaction protocol involves member authentication and a data exchange. It first identifies both parties are who they say they are through pub/priv key changes. Secondly, that at least a contract rule is matched to allow the transaction.

REFERENCES

[1] T. D. consortium. (2018) Data logistics for logistics data. [Online]. Available: https://www.dl4ld.net

[2] S. Cisneros-Cabrera, A. Ramzan, P. Sampaio, and N. Mehandjiev, "Digital marketplaces for industry 4.0: a survey and gap analysis," in Working Conference on Virtual Enterprises. Springer, 2017, pp. 18–27.

[3] A. Zerdick, K. Schrape, A. Artope, K. Goldhammer, U. T. Lange, E. Vierkant, E. Lopez-Escobar, and R. Silverstone, E-conomics: Strategies for the Digital Marketplace. Springer Science & Business Media, 2013.

REQUIREMENTS AND TARGET AUDIENCE

All attendees familiar with or interested in secure data sharing, collaboration computing, container networks and the concept of digital marketplaces.

DEMO DURATION

The demo will be presented in a 30-minute session (unless otherwise noted).

EXPECTED DEMO FORMAT

Talks + Demo

A/V AND EQUIPEMNT

Projector to be connected to the laptop. Internet connection.

INSTRUCTORS' BIOGRAPHY

Dr. Paola Grosso is associate professor in Systems and Networking Lab (SNE) at the University of Amsterdam. She is the coordinator and lead researcher of all the group activities in the field of multi-scale networks and systems. Her research interests lie in the creation of sustainable e-Infrastructures, relying on the provisioning and design of programmable networks. She currently participates in several national projects, such as SARNET, DL4LD, EPI and SecConNet and in EU H2020-funded projects such as FED4FIRE+, GN4. and ENVRIPLUS. See: https://staff.fnwi.uva.nl/p.grosso/

Prof. de Laat chairs the System and Network Engineering (SNE) laboratory at University of Amsterdam. The SNE lab conducts research on leading-edge computer systems of all scales, ranging from global-scale systems and networks to embedded devices. His group is/was part of a.o. EU projects GN4-2, SWITCH, CYCLONE, ENVRIplus and ENVRI, Geysers, NOVI, NEXTGRID, EGEE, and national projects DL4LD, SARNET, COMMIT, GIGAport and VL-e. See:http://delaat.net/

Dr. Reggie Cushing is a postdoctoral researcher in Systems and Networking Lab (SNE) at the University of Amsterdam. His research interests are distributed computing, computing paradigms, programmable infrastructures and alternative computing. He is currently involved in PROCESS and DL4LD projects.

Lu Zhang is currently a PhD student in Systems and Networking Lab (SNE) at University of Amsterdam. She received her B.SC. and M.Sc from Shandong University, China and RWTH Aachen University, Germany. Her research interests include information security, container networks and novel networking infrastructure

Yuri Demchenko is a Senior Researcher at the System and Network Engineering of the University of Amsterdam. His main research areas include Big Data and Data Intensive Technologies and Infrastructure, Cloud and Intercloud architecture, general security architectures and distributed access control infrastructure for cloud based services and data centric applications. More information about his professional activity, projects, publications, philosophy and hobbies can be found at his homepage http://www.uazone.org/demch/.

Acknowledgment

This demo builds upon the work done within the Dutch NWO Research project 'Data Logistics for Logistics Data' (DL4LD, www.dl4ld.net), supported by the Dutch Top consortia for Knowledge and Innovation 'Institute for Advanced Logistics' (TKI Dinalog, www.dinalog.nl) of the Ministry of Economy and Environment in The Netherlands and the Dutch Commit-to-Data initiative (https://commit2data.nl/)