Contribution ID: **79** Type: **Oral Presentation**

Blockchain-Enabled Secure Management of Scientific Data in Permissioned Networks: a Metadata-Centric Approach (Remote Presentation)

Thursday, 28 March 2024 14:20 (20 minutes)

In recent years, blockchain has emerged as a promising new technology to manage trusted information, making it easier for companies to access and use critical data while maintaining the security of this information.
Permissioned blockchains, unlike permissionless ones, restrict access to a select group of certified entities. They ensure a controlled and secure environment where only authorized participants can join the network and perform operations, a peculiar aspect in sectors where data sensitivity, confidentiality, and limited access are crucial.
In this work, we present the implementation of a permissioned blockchain system aimed at ensuring data immutability, operations traceability, and the ability to reproduce workflows.
Tracking operations performed on the data and guaranteeing reproducibility of research through workflow reconstruction upon data processing become very important in different sectors ranging from scientific communities to private companies and health.
In such regards, we work with Hyperledger Fabric, an enterprise-grade permissioned distributed ledger platform that offers modularity and versatility for a broad set of industry use cases.
Among its features, Fabric supports the generation of digital certificates and the possibility to import externally provided X.509 format to ensure participant identity; smart contract development in different programming languages (such as Go, Javascript or Java); provides a REST API for communication with off-chain service; its modular architecture allows various components and features to be replaced or extended based on the specific needs of a blockchain network.
In our implementation of the private and permissioned blockchain network, we decided to operate with the hash of the data and the related metadata instead of the whole data. In such a way, storage space usage can be optimised, providing the possibility to delete data. In a blockchain framework, in fact, data are not removable. By storing only the metadata, we store essential information on the blockchain, and leave the possibility to manage and delete specific data entries stored off-chain efficiently.
Moreover, the support of a data lake to store data enables the possibility to trigger events when the data is modified.
Furthermore, we take into consideration also aspects such as data tampering and how blockchain can help in capturing and storing the modification, revealing the owner and timestamp of the operation. As already mentioned, including hashes of the data allows a faithful reconstruction and reproduction of the entire workflow, making the data processing management trustable.
In the present work, the blockchain implementation is explained, with some examples, and possible improvements are presented, including the design of a cloud-enabled blockchain as a service on the INFN Cloud infrastructure.

Primary author: RANIERI, Domingo

Co-authors: Mr COSTANTINI, Alessandro (INFN CNAF); Mrs MARTELLI, Barbara (INFN CNAF)

Presenter: RANIERI, Domingo

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data