



DIPLOMATERVEZÉSI FELADAT

Klenk Botond (FTNYN1)

Mérnök informatikus hallgató részére

Adatterek minőségbiztosításának támogatása

Az adatmegosztást és -feldolgozást, intelligens algoritmusok ill. MI szolgáltatások felhasználását és moduláris rendszerbe illesztését az elmúlt időszakban az EU által is támogatott adatterek technológiai és rendszerszervezési megközelítésében alkalmazzák számtalan területen, gyártási adatok megosztásától kezdve energetikai alkalmazásokon át személyre szabott képzési ajánlatok elkészítéséig. Ezekben a megközelítésekben közös, hogy az adattér, mint infrastruktúra célja alapvetően az adatot megosztó és feldolgozó felek közti kapcsolat felépítése, ugyanakkor a konkrét adattranszfer a felek közt közvetlenül történik.

Ha az adatmegosztás és az arra épülő alkalmazások üzleti célt szolgálnak, kiemelten fontos az adatmegosztás sikerének biztosítása, az átadott adat és az arra adott válasz minőségének kiértékelése. Ezekre a kérdésekre ugyanakkor a legtöbb elérhető megoldás (pl. az Eclipse Data Connector vagy a különböző referencia architektúra modellek-IDSA, DSSC, Gaia-X, stb.) nem adnak választ, ill. nem kezelik a rendszer kiemelt feladataként.

A hallgató feladatának a következőkre kell kiterjednie:

- Vizsgálja meg és röviden értékelje ki, milyen kérdések merülnek fel egy adat átadásakor annak minőségével kapcsolatban, és hogyan oldható meg ezek kiértékelése adatterek (data spaces) tervezésénél és működtetésénél.
- Mutassa be, hogyan lehet egy ellenőrző komponenst az adatmegosztási folyamatba integrálni.
- Mutassa be, milyen szerepe lehet a blokklánc technológiának az adatmegosztás jellemző tulajdonságainak rögzítésében.
- Készítsen prototípus implementációt és mutassa be annak jellemzőit konkrét adatmegosztási példákon keresztül. Munkája során törekedjen megoldása általánosíthatóságára.

Tanszéki konzulens: Dr. Gönczy László, egyetemi docens

Budapest, 2024. március 17.

.....
Dr. Dabóczi Tamás
tanszékvezető
egyetemi tanár



Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Measurement and Information Systems

Quality assurance support for dataspaces

MASTER'S THESIS

Author
Botond Klenk

Advisor
dr. László Gönczy

April 17, 2024

Contents

| | |
|---|-----------|
| Kivonat | i |
| Abstract | ii |
| 1 Introduction | 1 |
| 1.1 Dataspaces | 1 |
| 1.1.1 IDSA | 1 |
| 1.1.2 Prometheus-X | 1 |
| 1.1.3 Relevance of blockchain | 1 |
| 1.2 Quality assurance | 1 |
| 1.3 Technologies | 2 |
| 1.3.1 Express | 2 |
| 1.3.2 OpenAPI | 2 |
| 1.3.3 Hyperledger Fabric | 2 |
| 2 IDS RAM | 3 |
| 3 Data Veracity | 4 |
| 4 Architecture | 5 |
| 5 Implementation | 6 |
| Acronyms | 7 |
| Bibliography | 8 |

Kivonat

Magyar abstract.

Abstract

English Abstract.

Chapter 1

Introduction

In the ever-evolving realm of data management, a unified framework is essential to provide a structured environment for seamless integration and management of diverse data sources. Dataspaces are a concept that provides this unified view of data sources, enabling organizations to utilize each other's data.

1.1 Dataspaces

In today's data-driven landscape, the importance of creating a framework, where organization can utilize each other's data is significant.

1.1.1 IDSA

The International Data Spaces Association (IDSA) created a Reference Architecture Model (RAM)[1] to provide this structured environment for seamless integration and management of diverse data sources.

1.1.2 Prometheus-X

Prometheus-X is a project that aims to provide a reference implementation of the RAM.

1.1.3 Relevance of blockchain

Blockchain technology is a decentralized, distributed ledger that records the provenance of a digital asset. The relevance of blockchain in the context of dataspaces is that it can provide a secure and transparent way to record the data transactions between organizations.

1.2 Quality assurance

The goal of this thesis is to provide a solution that ensures the veracity of data in the context of dataspaces. The participants must be able to trust the data that is being exchanged between

them, and the process that checks the veracity of it.

Checking the veracity of data in a chaincode could be a solution to this problem. But with large amounts of data, it would be an inefficient way to do it. A better solution would be to check the veracity of data in a separate service.

1.3 Technologies

The technologies used in this thesis for creating a veracity checking service are:

1.3.1 Express

Express is a minimal and flexible Node.js web application framework that provides a robust set of features for web and mobile applications.

1.3.2 OpenAPI

OpenAPI is a specification for building APIs. It provides a standard way to define the structure of an API.

1.3.3 Hyperledger Fabric

Hyperledger Fabric is a private, permissioned blockchain platform that provides a modular architecture with a delineation of roles between the nodes in the infrastructure.

Chapter 2

IDS RAM

Chapter 3

Data Veracity

Data veracity can be measured by several factors, such as accuracy, consistency, and trustworthiness.

Chapter 4

Architecture

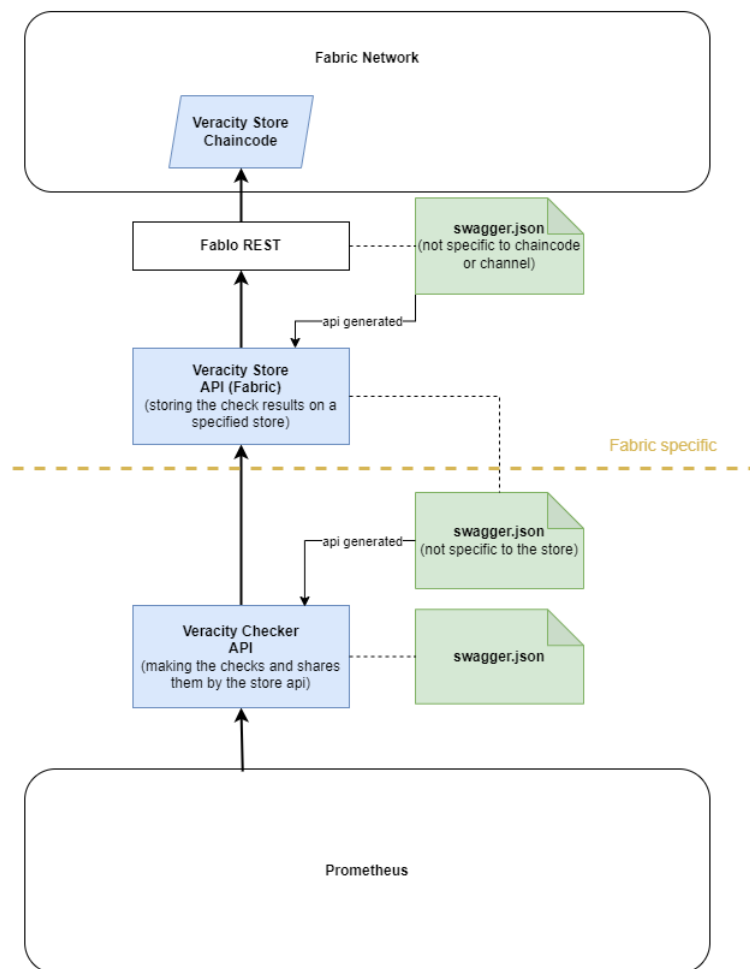


Figure 4.1: Architecture of the system.

Chapter 5

Implementation

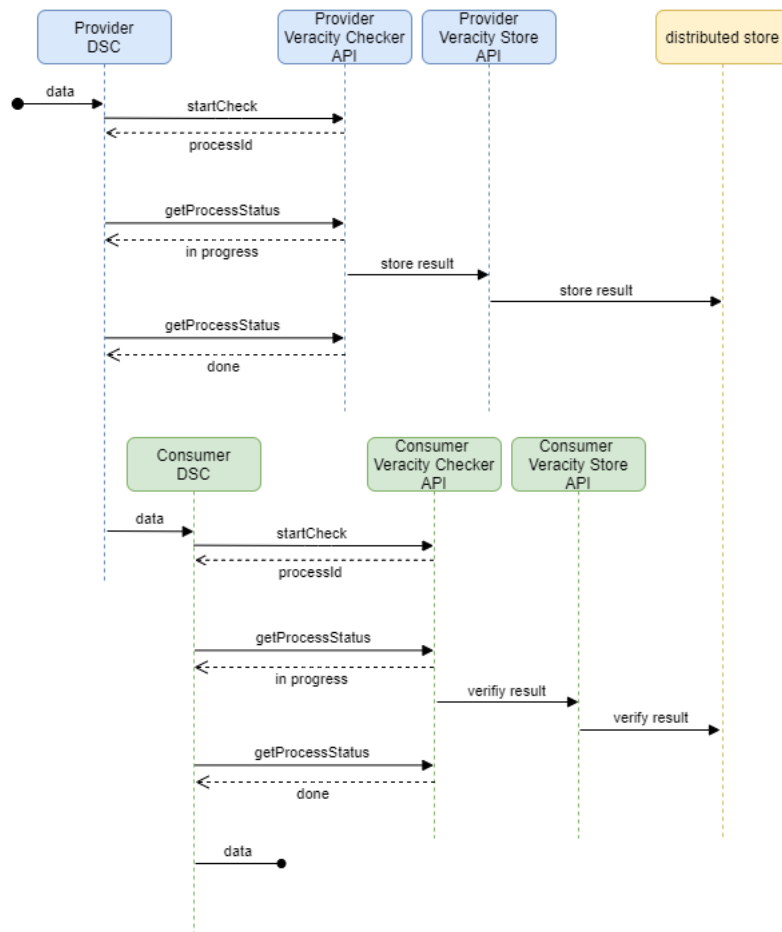


Figure 5.1: Check with verification call flow.

Acronyms

| | | |
|------------|---|---|
| IDS | International Data Spaces Association | 1 |
| RAM | Reference Architecture Model | 1 |

Bibliography

- [1] International Data Spaces Association. Ids ram 4, 2023. URL <https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4>.