

## Data Harmonisation in an Industry 4.0 Environment

By Alex McDonald, Developer at ICE

### Some questions for you

- How many different types of data can you find in your company?
- Does your data have complex schemas? Would it benefit semantic analysis?
- How much time do you spend on data cleaning and harmonisation?

### Data Processes in 4.0 Systems

“Industry 4.0” can be considered a nebulous term – a science fiction-like concept that refers to the increase of smart technologies share in industry. These technologies are expected to communicate with other smart technologies to increase automation with the less specific concept of the Internet of Things (IoT). [1]

This communication of smart technologies means that different systems are communicating, from precision cutting tools such as electro-spindles on the factory floor, to the off-site management systems that monitor the machines in multiple different factories. In a perfect world, these systems can all communicate seamlessly, but even the largest Fortune 500 companies have struggled to create a standardised architecture that processes and collects data. [2]

To turn this future vision of an Industry 4.0 connected via Internet of Things you need to harmonise.

### Transformation? Centralisation? Harmonisation?

If projects are examined that have a general focus on data there are various terms that in theory, could be used interchangeably. In some cases they are! (But shouldn't!)

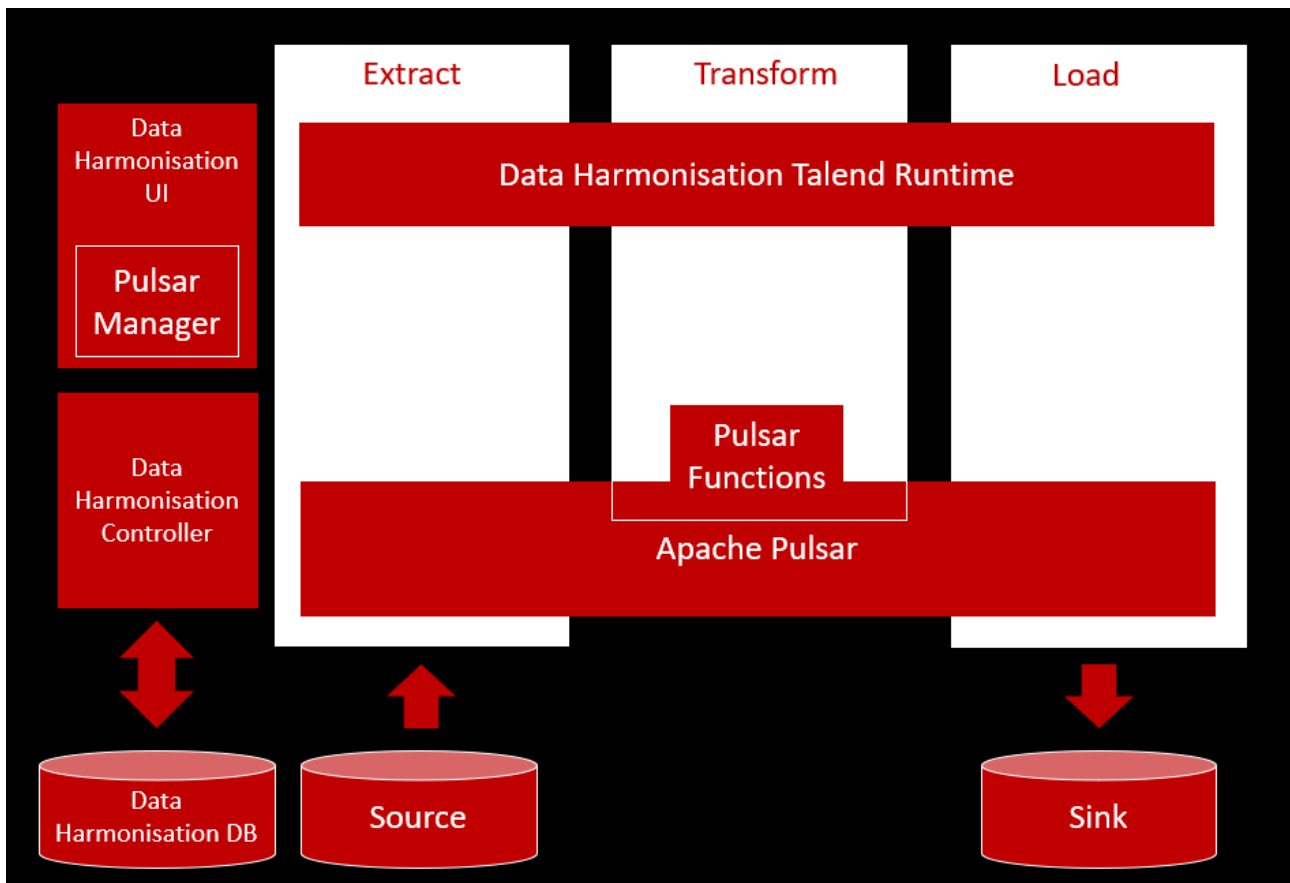
“Data harmonisation” is the process of organising different data into one homogenous type of data. The data inputs to this process can be varied, from different file types, different file structures to simply, different naming conventions. The input data going through a data harmonisation process can then be transformed into a cohesive data set.

How does that differ from “data transformation”? Whilst the process of transforming data can happen during a data harmonisation process, it does not tell the full story. Data harmonisation can include data being transformed in various ways, but it does not have to. The data harmonisation process can occur without any data transformation! For example, organising separate files into a single location, combining different databases, or connecting to multiple interfaces might not involve any transformation.

Is that not be considered data centralisation? Centralising data is an example of one of the many processes that can be undertaken during a data harmonisation process. The concept of data harmonisation is an umbrella process that includes various smaller tasks that affect data, transforming data, centralising data, understanding the semantic analysis of the data, and many steps between.

But what does “semantic analysis” mean? At its simplest level, semantic analysis is the ability to understand meaning from text. It is the process that humans or computers go through to understand words or sentences by analysing the grammar, and understanding what the words mean in different contexts. Using it, companies can extract meaningful data automatically, such as support tickets or feedback. [3]

Together in harmony, all these tasks can eat up a lot of man hours. If companies fail to consider standardisation efforts, differing data formats and data harmonisation they will find that they will lose out to companies making use of the latest advancements in semantic analysis and supervised machine learning techniques.



## What will ZDMP achieve

The Data Harmonisation aspect of ZDMP is made up of two separate components that will work in tandem.

The Data Harmonisation Design-time component helps divorce the complexities of data harmonisation with programming interfaces and gives users from a less technical background the ability to create complex data pipelines that can achieve a successful data harmonisation process with ease. With a built-in semantic reasoner, it can also offer the option of mappings that are done via semantic analysis.

The Data Harmonisation Runtime component is where the technical aspects come into play, where all the hard work that the users of the design-time component accomplish is connected to the wider ecosystem of ZDMP. This component exemplifies the accessibility of the data itself, allowing the data harmonisation to connect to virtualised services (Docker and Kubernetes) to a message bus such as Apache Pulsar.

When working together, these two components will offer a data harmonisation solution to the multifaceted data problems of an Industry 4.0 and Internet of Things lifecycle.

## ZDMP Links

|                             |  |
|-----------------------------|--|
| • Architecture Component(s) | Data Harmonisation Design-Time<br>Data Harmonisation Runtime |
| • Work Package              | WP5 – Core Services and Middleware                           |
| • Tasks                     | T5.3 – Data Harmonisation                                    |

## References

- [1] Control Station, 28th May 2019. 'What are the differences between Industry 4.0, Big Data and the IIoT?' [Available on: <https://controlstation.com/what-are-the-differences-between-industry-4-0-big-data-and-the-iiot/>]
- [2] Machine Design, 24th May 2019, 'Design a Data-Centric Industry 4.0 System' [Available on: <https://www.machinedesign.com/automation-iiot/article/21837825/designing-a-datacentric-industry-4-0-system#:~:text=An%20effective%20Industry%204.0%20architecture,process%20information%20and%20valuable%20insights.>]
- [3] MonkeyLearn, 12th August 2020. 'Semantic Analysis: What Is It & How Does It Work?' [Available on: <https://monkeylearn.com/blog/semantic-analysis/>]