

Cross-Domain Data Exploitation

The impact of being data-driven and a glimpse into the future?

Ir. Steven Beeckman, Attaché, CC V&C / Dept AD&M
steven.beeckman@mil.be

I. INTRODUCTION

Since 2009 the Belgian Defense started enabling structural Cross-Domain Data Exploitation by enabling easier access to corporate data coming from domains such as HR, MR, BudFin, Ops & Trg, ... The presentation will briefly describe the history and current state of affairs of the XDE concept as well as look at new technologies available on the market. These new technologies can enable us to lower the barrier to entry, as well as enable us to detect better insights into the data generated by our Defense.

II. HISTORY OF XDE

Traditional tools follow a centralized data warehousing approach, which requires a significant amount of resources like disk storage, CPU and RAM. A data warehouse also requires a significant number of personnel trained to format the data in the required views and who would have the mandate of all the domains (e.g. HR, MR, BudFin) to model the data correctly. MRC&I-CIS thus decided to opt for a hub & spokes model where data is ingested into a central repository but ventilated into local nodes according to each node's data requirements for reporting and in accordance with each data provider's permissions.

III. THE IMPACT OF XDE

A description will be given of the current tools made available to the personnel of Defense (Datamart, DBNode, Loader and LRF) as well as the required competences. The author will furthermore highlight the impact of XDE on the Belgian Defense's business processes.

IV. A GLIMPSE INTO THE FUTURE

XDE currently uses technology widely available in 2009 but the world did not stand still in the mean time. Open source technologies such as Hadoop [1] and even Spark [2] have become mature allowing users to process more data, not only in batch but also in real-time. New visualization libraries, frameworks and tools have become available on the market, allowing us to condense lots of data in a limited amount of space. Concepts such as the IPython and Spark notebooks [3][4] have emerged, allowing its users to share not only the end result but also the way to achieve that result. But, given that our data center resources have been increased recently, why not look into full-stack centralized data warehouses again?

V. CONCLUSION

Opening and integrating data across domains allows any organization to become more streamlined and more efficient.

REFERENCES

- [1] The Apache Hadoop project. <http://hadoop.apache.org/>
- [2] Apache Spark. <https://spark.apache.org/>
- [3] The IPython Notebook. <http://ipython.org/notebook.html>
- [4] Use Apache Spark straight from the Browser. <https://github.com/andypetrella/spark-notebook>