Generation of Business Process Indicators
Development of a Data Warehouse for the edoclink process engine

João Miguel Gandum Ferreira

Instituto Superior Técnico, Lisboa, Portugal

January 2021

Abstract

Indicators are essential for any system where constant monitoring and evaluation is important, giving us information on performance, achievement and responsibility. In the context of a company named edocLink, some indicators were designed to allow an analysis of the data generated by its product (edoc, a document management tool), useful for, for example, automating the task assignment process or to simply see the evolution of the workload over the time. We devoleped a solution that allows extracting and transforming data from an edoc generated database and then loading it into a Data Warehouse. This Data Warehouse is then connected with Power BI where the analysis will be made through dashboards, fulfilling the objectives defined jointly with edocLink. In the end, the solution and the results obtained are evaluated in 4 different ways - extensibility, production, scalability and Dashboards' usability. It is concluded that the solution is extensible and productive, but not scalable. In addition to that, the dashboards created were concluded to be simple, interactive and of immediate understanding, even for a user who does not have a vast knowledge of the edoc concepts.

Keywords: Data Warehouse, Business Process engine, Business Process KPI, Database, Star schema, Power BI , ETL

1. Introduction

1.1. Context

This project was developed in the context of a company named edocLink Enterprise, more specifically, in the context of edocLink - a document management tool. We thought it would be interesting to automate the process of allocating tasks to staff. For that, it is first necessary to define how the allocation will be made - what is the criteria used for it. Should a task be allocated to whoever executed the least tasks in the past month ? Or to whoever is fast performing tasks of that specific type? Or to the person who has the least pending tasks to execute?

This project aims to develop indicators that allow us to carry out this type of analysis which will then serve as a basis for the automation of these allocations. In addition to these types of indicators, we also developed other indicators that allow for a more general analysis of the company, such as the evolution of the work volume over a period of time, etc. With this in mind, it was fundamental that we first understood the company's environment, including its concepts, functioning and working method. After that, we will use the data generated from executing their business processes, transform and analyze in order to reach our goals.

1.2. Objectives

The project's objective is to develop a system capable of answering a set of indicators defines directly with the help of edocLink. The indicators are:

- **I. Pending *Etapas***
- **II. Delivered *Etapas***
- **III. Average Time to Accept**
- **IV. Average total Time per *Distribuição* Type**
- **V. Average Total Time per *Etapa***
- **VI. Average Total Time per *Fase***
- **VII. *Distribuição* volume**

1.3. Methodology

For the solution to be able to answer the indicators defined, the following steps were taken:

- Analyze the edoc database and decide which tables would be relevant for the project;

- Create a Data Warehouse, where we defined the dimenson and fact tables as needed;
- Implement the ETL process using Pentaho Data Integration, which extracted the data from edoc's DB, transformed them loaded into the created DW;
- Develop dashboards that allow for a simple and direct visualization of the data needed to answer the indicators

1.4. Results

The obtained solution was a system with 4 different components - edoc's DB, a Data Warehouse, Pentaho Data Integration for the ETL process and Power BI for the data visualization. It was concluded that the solution is extensible, productable interactive and easy to visualize what is intended. However, the solution is not scalable due to the difference between execution times when executed with a big data volume.

1.5. Document Structure

The document has 5 chapters. In Chapter 1 an introduction to the theme is made, explaining the context in which the developed project is inserted and what are the objectives that are intended to be achieved with the developed solution.
Chapter 2 presents the State of the Art - refers to several definitions of important concepts for the theme, exposes different types of data models and introduces several related tools.
The implemented solution is then detailed in Chapter 3, which shows how all the components of the solution were developed. The architecture and technology used in the solution are presented and some development steps are explained
In Chapter 4, the solution is evaluated in order to verify that the objectives defined in Chapter 1 have been met - both in terms of performance and functional.
Finally, Chapter 5 concludes the document, making a brief conclusion about the entire project and suggesting some proposals for future work.

2. State of the Art

Indicators are an essential component of any effective monitoring and evaluation system, providing crucial information about performance, realization and responsibility. For example, it is based on them that decisions are made to optimize a company's performance. It is also based on indicators that the government decices on the strategy to follow in order to respond to a pandemic.
The challenge is not in developing indicators, but in ensuring its quality and integrity, in order to guarantee that it provides us with concise and valuable information.
An indicator is therefore useful because it is a normalized measure to make comparisons over a long period of time, which gives us the ability to assess and interpret the evolution of a determined variable.

2.1. Data

In order to be able to apply an indicator, it is first necessary to have a collection of data that will serve as a basis. This data is typically generated through the execution of an organization's / company's business processes. The purchase of an item or the delivery of an order are both an example of events that occur during a business process and produce crucial information for a further analysis.
The information systems are responsible for managing the processes' execution and saving the data generated by each step of the process. Generally, they process one event at a time, generating a transaction record and identifying the current status. After that, another event may alter the initial state, which leads to the system updating the register. Therefore, the data models of these information systems are optimized to write and update the data fast.
However, this data collection cannot be used to analyze and reach to conclusions directly. The data generated by the business processes is highly standardized to minimize space occupancy and maximize the speed with which these are stored, making it more difficult to users to understand its structure. Despite some data sources allowing for a superficial analysis, through dashboards, for example, it is not flexible enough for an analyst. This is where analytical databases, known as Data Warehouses (DW), arise. By Kimball's definition, a DW is "a copy of transaction data specifically structured for query and analysis". DW are able to integrate data from multiple sources. Another important characterisctic is the fact that all the data is temporarily tagges, allowing for a temporal analysis. The process of extracting the data from the data sources and storing it on the DW is practically automatic, reducing the cost and ensuring the availability and consistency of the information. Also, it is crucial to note that before the storing process, the data usually suffers some kind of transformation.

2.2. Dimensions and Facts

Dimensions are an essential concept for the DW architecture. These allow you to analyze the information from a specific perspective. They are used to select and group the data according to the desired level of detail.
The dimension tables contain descriptive attributes, used to filter queries. Each table contains a unique primary key that corresponds to one part of the composite key of the associated fact table.

The fact tables are another important component, as they store the metrics to be analyzed. Each fact is an object that represents something to be analyzed. Both tables are related, since in multidimensional models the facts are implicitly defined by the combination of the dimensions. Fact tables have a primary key with 2 or more foreign keys and have a one-to-many relationship with dimension tables.

2.2.1 Star Schema

The star scheme consists of a central table (table of facts) surrounded by several dimension tables. The great advantage of this organization is the response's speed, regardless of the data volume. This behavior is possible since the maximum number of join operations is equal to the number of dimension tables linked to the fact table. In this scheme, the dimension tables are highly denormalized and, as such, contain a lot of redundant information (in exchange for better performance).

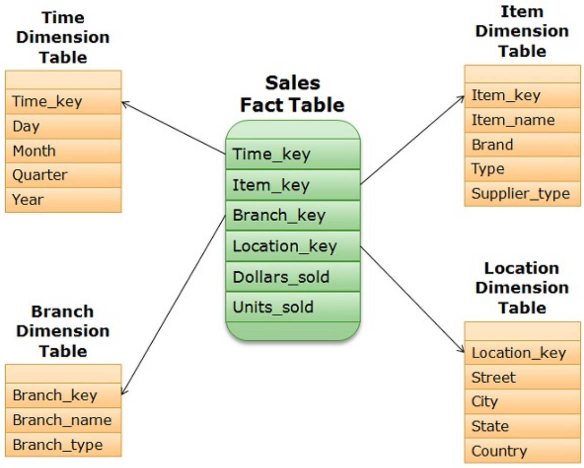


Figure 1: Star Schema architecture

2.2.2 Snowflake Schema

This scheme is similar to the star scheme in the sense that it also has a central fact table that is connected to several dimension tables. However, dimension tables are normalized into several tables (subdimensions), dividing the data to avoid redundancy. This clearly increases the number of tables needed, increasing the number of join operations for a given query, but decreasing the space needed to store the information.

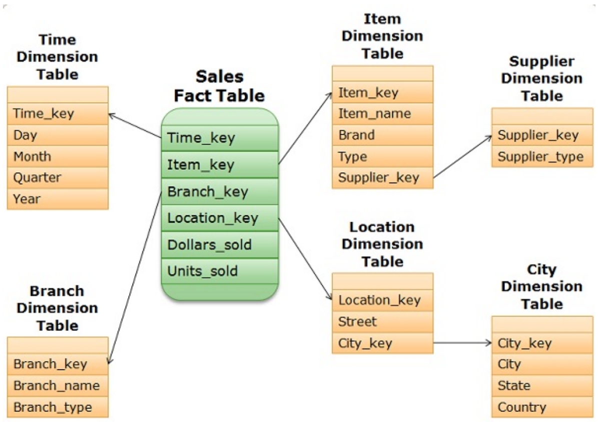


Figure 2: Snowflake Schema architecture

2.3. ETL

Extract, Transform, Load is the process of moving data from an original source to a destination source. In this process, the data is extracted from the original source, transformed as needed and loaded into the destination source. Firstly, the extraction depends on the source type. If it is a database, it is a trivial process - just connecting to the database and then copying the data. However, it is also possible that the data is in XML format or web services, for example. In any case, it is only when the extraction is complete that the process proceeds to the next step - transformation. It is in Transformation that a set of rules and conditions are applied to the data previously copied, in order to change its structure or content - integrating data from several sources into one or just improving its quality. Finally, the transformed data is loaded into the DW, allowing queries to be made about it - the main purpose of a DW.

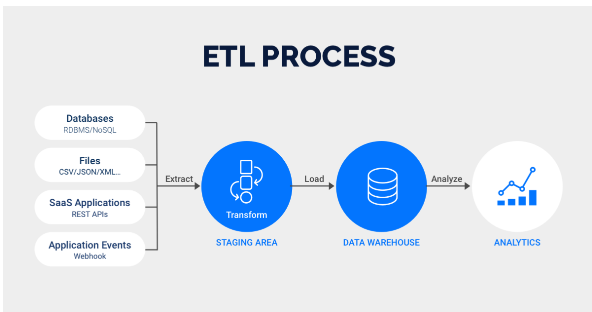


Figure 3: ETL process

In order to facilitate the Data Warehouse development process, several tools have emerged that allow the implementation of the ETL process without having to write any line of code - making it easier and more intuitive for non-programmers to understand. Another advantage of these programs is the fact that they have graphical interfaces that help to speed up the process of mapping tables and columns between the source and destination sources. Pentaho Data Integration is an example of an ETL tool.

2.4. Reporting Tools

After the ETL process is complete, it is important to visualize the data. For this purpose, there are various reporting tools that translate the information present in Data Warehouses into graphs, tables or other forms of visualization, allowing users to browse, filter and sort according to the his needs. These reports allow organizations for an informed analysis and decision making. Power BI is an example of a reporting tool.

2.5. Link Consulting and edoclink

The solution was developed for edocLink Enterprise. Its product, edoclink, is a document management and workflow tool that allows the dematerialization of administrative tasks and decision-making processes. It is already used in several sectors, both public and private, such as Health, Education, Insurance and Public Administration. This solution supports the latest Microsoft platforms, including a series of features capable of satisfying the needs of all organizations (Requirements analysis, organizational consulting, post-production support, etc.).

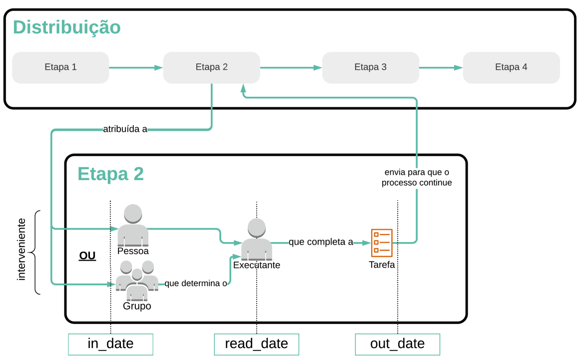


Figure 4: Resumed Edoc Cycle

3. Solution

The implemented solution aims to generate indicators that help to decide to whom a certain *etapa* in edoc should be allocated. For example, should *etapa* A, associated with *Interveniente* X, be allocated to *executante* 1 (who has less pending tasks) or *executante* 2 (who performed more tasks in the last 30 days)? The indicators will serve as a basis for making these types of decisions later.

3.1. Architecture and Technology Used

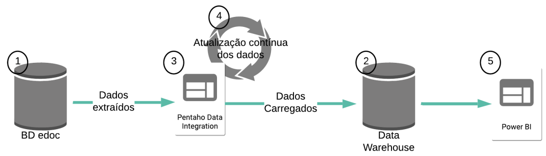


Figure 5: Proposed solution's architecture

3.1.1 1. Edoc Database

Unfortunately,due to data protection laws, it was not possible to access a customer's database. The database used was then a copy of a database used for tests provided by edoclink, so it is expected that the results obtained are not representative of the reality.

The name of the database is EDOCDEMO3_to_azure and is a SQLServer database. It has hundreds of tables and, as such, it was first necessary to analyze the tables and understand which ones would be important for the project. It was decided that the event to be analyzed in the fact table would be each *etapa* of a distribution, so the main table was DISTRIBUTION_STAGES which contains most of the information related to each stage. In addition to that, a few more tables were used in order to obtain all the information necessary to respond to the defined indicators.

3.1.2 2. Data Warehouse

The Data Warehouse created follows a star schema structure. It presents 6 dimensions and a fact table with the measures *tempoAceitacao*, *tempoExecucao* and *tempoEtapa* that will help to answer the objectives defined initially.

- **dim_profile -** Dimension that contains information about the *intervenientes* and *executantes*.
- **dim_distribuicao -** Dimension to filter by *distribuição*.
- **dim_typeDistribution -** Dimension to filter by *distribuição* type.
- **dim_TypeStep -** Dimension that contains the information needed to define an *etapa* type.
- **dim_etapa -** Dimension that contains all the information related to each *etapa*.
- **dim_tempo -** Dimension that will allow filtering by time, more specifically, by month, year, etc.

- **fact_eventos -** Fact table that connects all tables. The 3 measures it presents will be important to understand how long each *etapa* takes to be accepted / executed. As mentioned in the previous section, each entry in this table refers to a *etapa* in a *distribuição*.

3.1.3 3. Pentaho Data Integration

The PDI was the ETL tool chosen for the project. Not only did it have all the necessary functionalities, but it had also been previously used in the course of Analysis and Data Integration of the Master in Informatics and Computer Engineering. This component is the core of the project. This is where the data is extracted from the edoc database, transformed and later loaded into the developed DW. A transformation was developed for each dimension. Also, a job was also implemented to perform all transformations sequentially and daily.

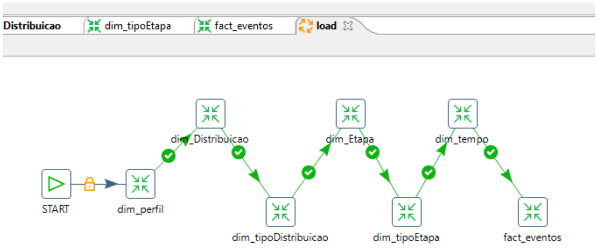


Figure 6: Job to execute all transformations sequentially

3.1.4 4. Continuous data updating

The continuous data updating is crucial because it is what allows us to make an accurate analysis over time. Pentaho Data Integration allows to schedule the job's execution.

The scheduling was set to run the job every day at 12:00. This execution is done locally, but it can also be defined to run on a server (from PDI or from edoclink, for example). In this case, given that the DB is static and is not updated with new data, the daily execution will not load / update new information to the DW. However, the implementation was made as proof of concept.

3.1.5 5. Power BI

We decided to use Power BI since it is well recognized in the business world, has a user-friendly interface and a great variety of data visualization and more detailed documentation. It is through Power BI that the indicators will be designed, therefore it has a direct relationship with the objectives defined at the beginning of the document (one Power BI file per indicator).

3.1.6 I. Pending *Etapas*

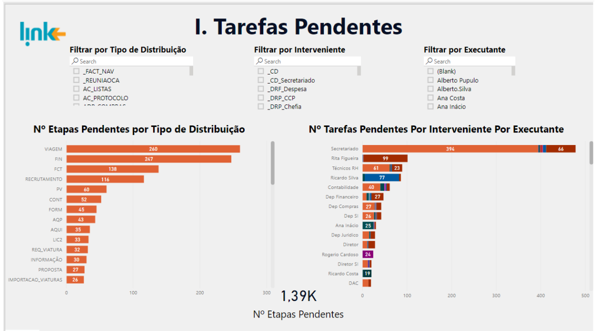


Figure 7: Indicator I

This Dashboard allows to answer questions such as:

- Who has more pending *etapas*?
- Which is the *Distribuição* type with the least pending *etapas*?

3.1.7 II. Delivered *Etapas*

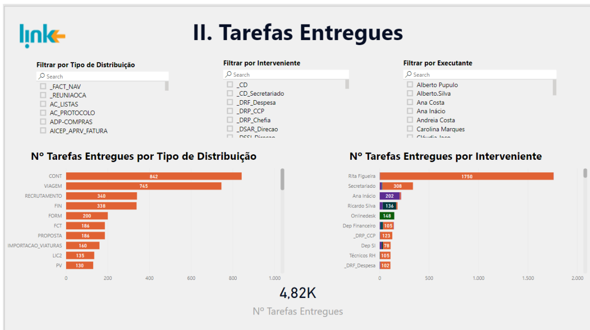


Figure 8: Indicator II

This Dashboard allows to answer questions such as:

- Who delivered more *etapas*?
- Which is the *Distribuição* type with the most delivered *etapas*?

3.1.8 III. Average Time to Accept

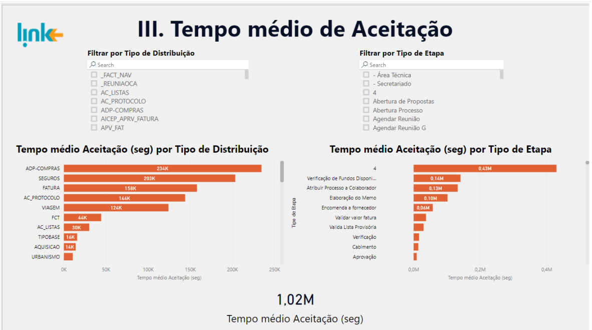


Figure 9: Indicator III

- Which *Distribuição* type usually takes longer to accept?

- Which *Distribuições/etapas* are accepted faster?

3.1.9 IV. Average total Time per *Distribuição* Type

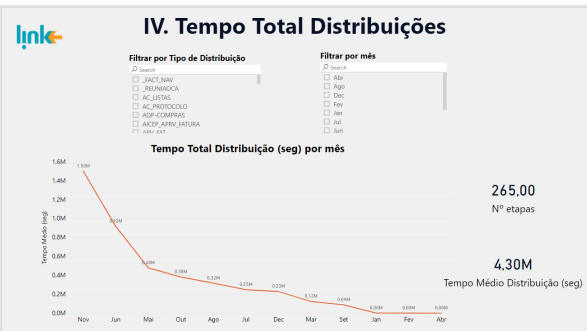


Figure 10: Indicator IV

- Which *Distribuição* type usually takes the least time to execute?

- What was the evolution over time of *Distribuição* X ?

3.1.10 V. Average Total Time per *Etapa*

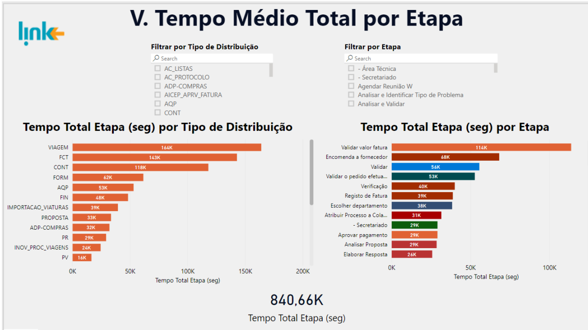


Figure 11: Indicator V

- Which *Etapas* usually take the most to execute?
- What *Distribuição* type has *etapas* that are executed the fastest?

3.1.11 VI. Average Total Time per *Fase*

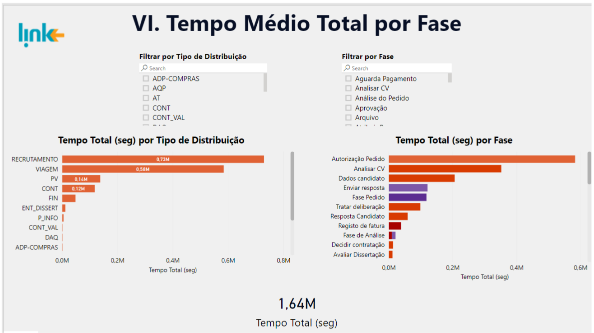


Figure 12: Indicator VI

- Which *Fases* usually take the most to execute?
- What *Distribuição* type has *fases* that are executed the fastest?

3.1.12 VII. *Distribuição* volume

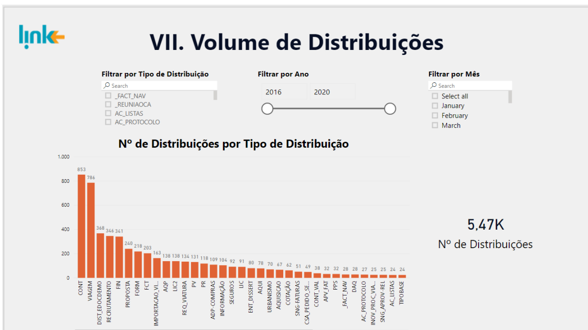


Figure 13: Indicator VII

- What period has the highest *Distribuições* volume?
- How is the evolution of the overall volume over the years?

3.2. Challenges

3.2.1 Indicators

The biggest challenge was to concretly define the indicators. Despite always having an idea of the type of questions that the solution would have to answer, it was only towards the end that the final version of the indicators presented in the document was reached. This decision was taken jointly with edoclink employees, who gave their opinion on what would be important to analyze. Before the final version, there were other versions which were much more widespread, making it more difficult to develop the solution (especially the dashboards in Power BI). The measures in the fact table could also only be thought after the indicators were fully defined.

3.2.2 Edoc environment

Another difficulty that arose at the beginning of the project was to understand the edoc environment. Before the solution development, it was first necessary to learn all the concepts, operation and organization of the company / tool. edoclink provided access to the edoc, together with a training that allowed to interact and visualize most of the concepts, facilitating the initial experience. However, it was the continuous interaction with company employees and with the supplied database that allowed the understanding of the entire system.

3.2.3 Repeated *Etapas*

During the course of the project, a problem emerged - how were we going to handle *etapas* that had to be executed again (due to problems in the first execution, for example)? In the initial BD, there was no way to identify 2 *etapas* as equal, since they had different IDs, so we had to find a way to solve this. This is important since a *etapa* that is executed several times should have as its execution time the total time of all the executions and not just the first one, for example.
It was then studied how we could implement this in the project and the solution was changing the *etapa* dimension to a Slowly Changing Dimension, where the different executions (through versions) associated with the same **etapaUnique_key** would be kept when the criteria defined to identify if two tasks are the same was verified. These versions are also associated with different dates. This solution achieved what was intended, since the total time of all executions of the same *etapa* was obtained by grouping by **etapaUnique_key**.

3.2.4 Power BI

Power BI was another component that initially generated some difficulty. Since there was no previous experience with the program, nor practical examples of dashboards that could be made, the graphs created in an initial phase were very weak compared to those presented in the previous section. The user was unable to have an instant understanding of what the graph represented, the filters were not user-friendly and the data specific to a testing environment was not filtered. It was with the direct help of edoclink that we moved from the initial graphics to the presented dashboards. From examples made for customers to opinions on what to add / remove, the company was crucial in reaching the final versions.

4. Evaluation

4.1. Extensibility

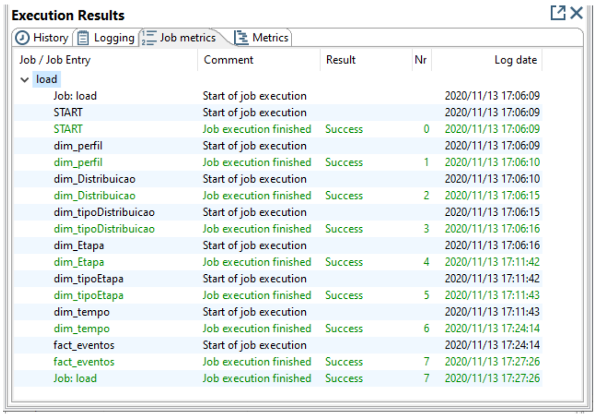
The solution was developed with the previously defined indicators in mind. However, how easy would it be to add another indicator indicator?

- **1. Relationship between current model and new dimension -** Analyze how the current model is related to what is pretended to add and choose which attributes are important.
- **2. Create new dimension -** Create new dimension by adding MySQL code to the existing script that creates the whole DW.
- **3. Create PDI transformation and add to job -** Create a transformation for the new dimension which will extract the data needed, transform it and load it into the DW. After that, add the transformation developed to the job.
- **4. Analyze with PDI -** Create a new dashboard which will give answer to the desired indicator.

Since the solution allows new indicators to be added, we can conclude that it is **extensible**.

4.2. Scalability/Performance

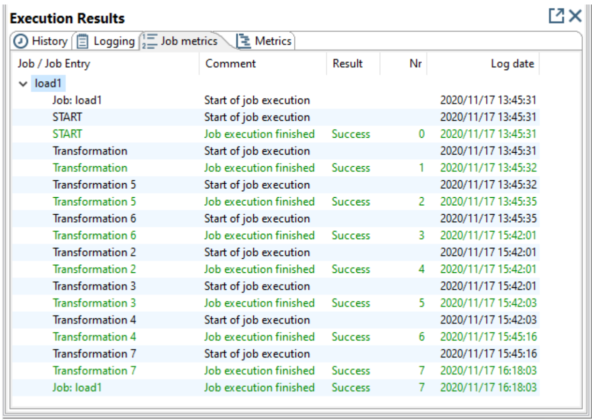
Regarding the performance, the ETL process duration was evaluated. This shows how long it takes for the data to be fully extracted, transformed and loaded into the DW. Below, a log of the job's execution is presented.



Job / Job Entry	Comment	Result	Nr	Log date
load				
Job: load	Start of job execution			2020/11/13 17:06:09
START	Start of job execution			2020/11/13 17:06:09
START	Job execution finished	Success	0	2020/11/13 17:06:09
dim_perfil	Start of job execution			2020/11/13 17:06:09
dim_perfil	Job execution finished	Success	1	2020/11/13 17:06:10
dim_Distribuicao	Start of job execution			2020/11/13 17:06:10
dim_Distribuicao	Job execution finished	Success	2	2020/11/13 17:06:15
dim_tipoDistribuicao	Start of job execution			2020/11/13 17:06:15
dim_tipoDistribuicao	Job execution finished	Success	3	2020/11/13 17:06:16
dim_Etapa	Start of job execution			2020/11/13 17:06:16
dim_Etapa	Job execution finished	Success	4	2020/11/13 17:11:42
dim_tipoEtapa	Start of job execution			2020/11/13 17:11:42
dim_tipoEtapa	Job execution finished	Success	5	2020/11/13 17:11:43
dim_tempo	Start of job execution			2020/11/13 17:11:43
dim_tempo	Job execution finished	Success	6	2020/11/13 17:24:14
fact_eventos	Start of job execution			2020/11/13 17:24:14
fact_eventos	Job execution finished	Success	7	2020/11/13 17:27:26
Job: load	Job execution finished	Success	7	2020/11/13 17:27:26

Figure 14: Job's execution log

The duration was around 20 minutes. Despite being an acceptable time, the DB used is a tests DB and static. Therefore, if a client's DB was used (with real data and constant inserts/updates) the duration is expected to increase to unacceptable times. To test that case, we created a new table with 10 times the data volume of current main table.



Job / Job Entry	Comment	Result	Nr	Log date
load1				
Job: load1	Start of job execution			2020/11/17 13:45:31
START	Start of job execution			2020/11/17 13:45:31
START	Job execution finished	Success	0	2020/11/17 13:45:31
Transformation	Start of job execution			2020/11/17 13:45:31
Transformation	Job execution finished	Success	1	2020/11/17 13:45:32
Transformation 5	Start of job execution			2020/11/17 13:45:32
Transformation 5	Job execution finished	Success	2	2020/11/17 13:45:35
Transformation 6	Start of job execution			2020/11/17 13:45:35
Transformation 6	Job execution finished	Success	3	2020/11/17 15:42:01
Transformation 2	Start of job execution			2020/11/17 15:42:01
Transformation 2	Job execution finished	Success	4	2020/11/17 15:42:01
Transformation 3	Start of job execution			2020/11/17 15:42:01
Transformation 3	Job execution finished	Success	5	2020/11/17 15:42:03
Transformation 4	Start of job execution			2020/11/17 15:42:03
Transformation 4	Job execution finished	Success	6	2020/11/17 15:45:16
Transformation 7	Start of job execution			2020/11/17 15:45:16
Transformation 7	Job execution finished	Success	7	2020/11/17 16:18:03
Job: load1	Job execution finished	Success	7	2020/11/17 16:18:03

Figure 15: Job's execution log with 10 times more data volume

As expected, the duration was much higher (20 minutes vs 2 hours and 30 minutes), demonstrating that the solution is **not scalable** as the data volume increases. This could be solved if, with every job execution, only the new data that was changed or added since the last execution was inserted/updated into the DW. One possible way to implement it would be to create a Control table that stored the date of the last execution and the time interval desired (24 hours if we want to execute the job daily). We could then filter the data coming from the original DB by those attributes, making sure that only the new information was inserted/updated into the DW.

4.3. Production

In order to analyze whether the solution is productable, it is necessary to analyze the architecture and technology used and understand what changes would have to be made to the components of the current version for a client version.

- **Source DB -** Instead of using the test DB of the current version, the customer's DB would be used with the data generated by their business processes. The only necessary change would be the configuration of the connection established between the PDI and the DB.
- **Data Warehouse -** The DW of the current version would not have to undergo any changes, since changing the source DB would not affect the format of the data generated.
- **Pentaho Data Integration -** As mentioned, the only changes needed would be in the configuration of the connection between each transformation of the PDI and the customer's database. The measures could also be changed in the transformation of the fact table to present the time in minutes (instead of seconds), since the data would be more realistic than in the test database.
- **Data update -** When scheduling the job, there would also be no problems whatsoever. The client could decide if he wanted to run it locally or on a server - easily configured directly in the PDI.
- **Power BI -** Finally, in Power BI the only change needed would be the connection between the program and the DW, depending on the host of the DW. If it was locally hosted, no changes would be necessary. The dashboards would automatically present the customer's DW data, which consequently contained data coming from the database generated through the execution of the customer's processes.

Therefore, we can conclude that the solution is in fact **productable**, requiring only a few simple changes, mainly at the level of the connection between the DB of each customer and each component of the solution.

4.4. Dashboards' Usability

The Dashboards fulfill the initially defined objective. Each indicator corresponds to a dashboard that presents the data in an interactive, direct and user-friendly way. The most important thing is the fact that even a customer (who may not have a vast knowledge of the edoclink's environment, concepts

and operation) is able to open a dashboard and immediately understand what is being presented.
However, despite not compromising the analysis of the data, it was unanimous that an improvement to be made would be to have all the dashboards in the same Power BI file so that the access to them was easier. This could have been implemented, if we developed a redundant model directly in Power BI and established the relationships between all the tables, allowing for all indicators to be viewed in the same file.

5. Conclusions

Throughout the document, we presented the theme and concepts related to the project, detailed the proposed solution and discussed the results and its evaluation. The project's objectives were reached, having developed a solution capable of, through the extraction and transformation of data from the edoc database, giving an answer to the indicators defined through simple and interactive dashboards. In addition to that, it was concluded that the solution is extensible and productable, allowing the addition of new indicators / attributes and the production of customers' versions. However, and for future work as well, it would be important to make the solution scalable, since in the current version, it is inefficient and the job execution time will increase as the data volume increases. As suggested in the previous chapter, a possible improvement would be to create a control table with the date of the last job iteration and the time interval between each execution, which would then filter the values to be analyzed in each job iteration. Another aspect that can also be explored is to use the obtained results to automate the task allocation process directly in edoc. For example, if the criteria is to assign the "executante" with the least pending tasks, when a "etapa" is assigned to an "interveniente", then it would automatically assign to that "executante", not needing to be decided and accepted manually. This is something that is already being put into practice by a colleague who started her dissertation at Link.

In conclusion, the project was indeed interesting, the objectives were fulfilled and relevant results were obtained. It allowed me to have my first experience at a business level and interact with various tools and applications unknown to me until the beginning of the project. I would like to thank Link Consulting and, more specifically, prof. Pedro Sousa, Carolina Marques and João Guilherme for all the support and help they gave me throughout the duration of the project and the way they received me.

Throught the document, we introduced the theme and concepts related, presented and detailed a solution that allows to answer the defined objectives and showed and discussed the results that were generated by the solution. For future work, it would be interesting to fix some of the problems we did not solve with our solution - mainly turning it scalable - and crete a script that would atomate the task allocation process based on the indicators we developed. All in all, we are pretty satisfied with what we accomplished.

This project allowed us to:

- **Understand** the theme and the inherent technology.
- **Interact** with an existing business solution and understand its operation and crucial concepts.
- **Implement** a solution that would allow us to make an assessment of the data generated by the company.
- **Learn** new tools used in a business environment.
- **Structure** results in a user-friendly and interactive way that allows the user to understand immediately.
- **Evaluate** the results obtained and determine how they could be improved.
- **Think** about how to extend the solution to meet even more goals.

References

[1] J. Alves. Sistema de business intelligence no projeto educativo de guimarães. 2015.

[2] K. Bartley. *https://rivery.io/etl-vs-elt-whats-the-difference/.*

[3] A. Berti, D. Zang, and M. Lang. *An Open-Source Integration of Process Mining Features into the Camunda Workflow Engine: Data Extraction and Challenges.* 09 2020.

[4] edoclink. edoclink white paper. 2016.

[5] edoclink Enterprise. *https://www.linkconsulting.com/what-we-do/products/edoclink1/.*

[6] O. O. for Economic Co-operation and Development. *Glossário de termos-chave na avaliação.* 2002.

[7] M. Z. Karmani, G. Mustafa, N. Sarwar, S. Wajid, J. Qureshi, and S. Siddque. *A Review of Star Schema and Snowflakes Schema*, pages 129–140. 05 2020.

[8] R. Kimball, L. Reeves, W. Thornthwaite, M. Ross, and W. Thornwaite. *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses with CD Rom.* 1998.

[9] Pentaho. *https://help.pentaho.com/Documentation/7.1/0J0/0C0/020.*

[10] M. PENČIKOV. *Techniques and Methods for Effective Collection and Transformation of Process Metrics Data [online].* 2014.

[11] E. Santos. Indicadores de desempenho de bases de dados. 2017.

[12] I. Solutions. *http://www.innoventsolutions.com/pentaho-review.html.*

[13] F. Taveira. Análise de risco com base em indicadores. 2014.

[14] UNAIDS. *An Introduction to Indicators.*