

Advanced duplicate invoice analysis case

Use the power of SAP HANA

Many companies struggle with the new possibilities that open with new technology and combining technological solutions. In the case at hand, we describe how we implemented a duplicate invoice analysis at a major oil company. This article will provide hands-on tips on how to leverage the power of HANA for duplicate invoice analysis and how this can be integrated with SAP Process Control.



Stefan Zuijderwijk MSc RE
is a senior manager at KPMG
Enterprise Solutions.
zuijderwijk.stefan@kpmg.nl



Sylvester van der Giesen MSc
is a manager at KPMG Enterprise
Solutions.
vandergiesen.sylvester@kpmg.nl

The ability to monitor incoming invoices efficiently is becoming more and more critical

WHY YOU SHOULD IMPLEMENT INVOICE ANALYSIS

With hundreds or even thousands of incoming invoices every week, assessing invoices manually is nearly impossible for organizations today. The ability to monitor incoming invoices efficiently is becoming more and more critical. It is important not only for payment processing, but also to stay ahead of the risk around erroneous or fraudulent invoices and mistakes in internal invoice processing.

During the invoice review process, mistakes and errors are easily made. Since this is a common occurrence, most ERP systems contain automatic checks on incoming invoices. These checks are for example the three-way match – comparing the purchase order, goods receipt and invoice amounts and quantities – and a duplicate invoice report. These checks remain relevant, but with new database technologies available, the opportunities to implement advanced analytics are increasing.

In this article we will not only explain the default solutions offered by ERP systems like SAP, but we will also explore the use case of advanced duplicate invoice analytics. We will go through examples of how clients are using advanced analytics and how the solution was implemented.

This specific case contains an explanation of a duplicate invoice analysis; however, the methodology can be used for multiple purposes. The final sponsor in this specific case was the CISO, who wanted to use this analysis in internal audit and in the invoice management process. The analysis itself was implemented early 2017 at a large company, listed in the UK and US.

DUPLICATE INVOICE ANALYSIS

With the development of in-memory database technology, the possibility for more advanced analysis has been introduced. SAP entered the database market with SAP HANA, their in-memory database technology solution which offers the possibility to implement real-time advanced analytics. These real-time analytics have proven to be useful in business analytics. One application of the SAP HANA analytics is to report on duplicate invoices in real time.

Duplicate invoices entail a risk of actual financial loss. When duplicate invoices are not noticed and corrected in time, it will result in a duplicate payment. Duplicate invoice management prevents money from unnecessarily flowing out of your company.

The business case is often quite easy to make: most companies save money by implementing duplicate invoice management. In some cases, the investment is earned back in the first run, either by preventing the payment of duplicate invoices or by reclaiming what has already been paid.

Most ERP systems – SAP being one of them – have reports indicating possible duplicate invoices. This has a huge efficiency gain in (duplicate) invoice management. Using the default reports on duplicate invoices has some prerequisites and limitations:

- the duplicate invoice check must be enabled;
- the parameters on what to check must be enabled;
- the number of parameters that can be enabled is limited;
- it only shows when two invoices have a 100% match on the selected parameters.

As such, a report alone does not support the duplicate invoice management process. Additional follow-up procedures or controls should be considered and implemented to adequately address the risk.

DUPLICATE INVOICE SCENARIOS

As discussed, the standard reporting possibilities in ERP systems does not allow companies to implement an efficient and effective duplicate invoice management process. Besides that, limitations exist in the standard available functionalities of an ERP system. To come to a good definition of the end state analysis, we must investigate the different root causes of duplicate invoices and determine the best course of action.

The standard ERP reports do not identify duplicate invoices posted on two different vendor numbers. A duplicate vendor in the ERP system is a vendor that has two different vendor numbers. Although it might be the procedure to have duplicate vendors in the system within some companies, this is not common practice. To have a solution for this scenario, the analysis should be able to identify duplicate vendors.

Another scenario that is not covered by the default ERP reports are invoices with a minor difference in the amount, say 1 cent. Such a small difference can have different causes, for example due to currency translation. For our analysis, the root cause of this difference is not important; the analysis must identify minor differences. The solution for this scenario is that we can identify small differences and determine the materiality within the analysis.

Other scenarios that are not standard available in the ERP system reports are scenarios where the invoice date, reference documents or description do not match. If there is no exact match, the duplicate invoice reports within ERP systems do not report the possibility of a duplicate.

DEFINITION OF THE ANALYSIS

While the analysis aims to improve the default duplicate invoice analysis report, it is important that the analysis also covers a wide range of invoice processing scenarios. The baseline of the analysis should:

1. analyze all invoices;
2. consider near matches;
3. cover different master data sets;
4. convert data to matching formats.

A simplistic analysis results in a binary outcome: either the invoices are duplicates or not. In this case it means that the output of the analysis is a list of duplicate invoices. Invoices that are not on the list are defined as single entries. The problem with this approach is that duplicate invoices that are entered slightly different in the system are missed as being duplicate.

On the other hand, when the list contains too many possible duplicate invoices, the process becomes inefficient as the number of false positives increase. These false positives are invoices that are not duplicate, but are identified as being duplicates.

To solve the issue of missing duplicate invoices versus too many false positives, a logical likelihood calculation has been introduced. Based on the different parameters, a probability of two invoices being duplicate is calculated. Since the analysis uses input from master data, the probability of duplicate master data is calculated as well.

Because the calculations that must be executed can be quite time consuming, the number of calculations needs to be limited. The earlier it is identified that two sets of vendor master data or invoices are not duplicate, the more efficiently the analysis will run. Since the result of the analysis is not binary (true or false), the analysis itself should not be the basis of the scoping. Therefore, a threshold value is introduced. Since a two-step approach is being used – first determine duplicate master data, second determine duplicate invoices – two threshold values can be used.

SAPHANA has default functionalities that help in developing the analysis. Additionally, SAP HANA integrates very well with the SAP Business Suites and other SAP tooling.

Figure 1. Segments of duplicate invoice probability.

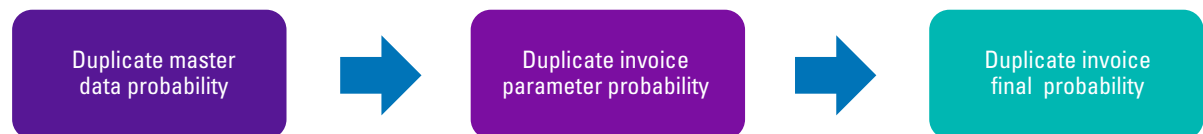


Figure 2. Duplicate invoice probability calculation - example.



RESULT OF THE ANALYSIS

For each combination of two invoices, the analysis determines the probability whether two invoices are duplicate. The result contains a 0% to 100% probability, based on the different parameters, scores and weights. A final score is calculated based on the scores of each parameter.

Depending on the processes and data available at a client, the parameters may deviate. The methodology remains the same, however.

The results are used in different ways. The output of the analysis is first reviewed by an accounting employee. This way, the analysis results in a working package for these accounting employees. The results can be distributed to budget owners, managers and even C-level executives.

As mentioned before, the analysis uses different thresholds for different reporting requirements. A manager is probably only interested in seeing the top scoring duplicate invoices with a specific amount. On the other hand, an (IT) auditor or risk manager might also be interested in the results of an analysis with a score of 80%, and the follow up of these cases.

USING SAP HANA FUNCTIONALITY

SAP HANA contains functionality and algorithms that can be used in advanced analytics. Within the case at hand, the SAP HANA functionality of temporary tables, fuzzy search and currency translations were used.

Fuzzy search

Most databases have a fuzzy search (or fuzzy matching) function, but the functionality is quite limited. For example: some database management systems will only compare the first four characters of the strings. The fuzzy search function in SAP HANA is well defined and powerful. But, it still is a fuzzy search function: it expects one (1) input string and gives an output value for each row: the matching percentage. This means that only one string at a time can be searched. Since the duplicate invoice analyses loop through a dataset with hundreds of thousands or even millions of invoices, the analysis will take a substantial amount of time and resources to run.

A type of fuzzy join is what is really needed. A normal join combines two sets of data using defined columns to link the datasets. For a normal join, the columns that are used in the join should contain the same content. The join that was developed for this analysis, the fuzzy join, should compare the content of the columns based on a specified threshold. This threshold is defined by either the end-user or by predefined scenarios.

Input	Fuzzy search result (%)	Possible duplicate
KPMG	100	Yes
KPGM	88,33	No
KP	58,57	No
GMPK	58,57	No
KPMG NV	96,31	Yes

Table 1. Fuzzy search: fuzzy results compared with 'KPMG', with threshold 95%.

INTEGRATION WITH SAP PROCESS CONTROL

Once the results from the analysis come in, it may overwhelm the end-user and it may be too much to follow up on without proper tracking. To ensure the follow up is performed, a process needs to be in place. However, just creating a new process does not ensure that people will also act accordingly. To overcome this, SAP Process Control (SAP PC) can be used. SAP PC will get the results from the duplicate invoice analysis, split them based on identifiers, such as company code and create a follow-up task for the respective owner. On top of that, a reminder and escalation procedure can be set up in the system to alert the owner, or the owner's supervisor on the outstanding task.

Once the owner receives the task, all processing can be captured in SAP PC. A remediation plan can be created, individual results can be marked as exception (i.e. false positive) or the risk can be accepted and documented. In PC there are reports that can show the status of controls (e.g. the duplicate invoice check), the number of open tasks and the follow-up actions taken. By leveraging SAP PC, the follow-up process is automatically initiated and easier to monitor.

**By leveraging SAP PC,
the follow-up process is
automatically initiated
and easier to monitor**

SAP S/4HANA contains functionality and algorithms that can be used in advanced analytics

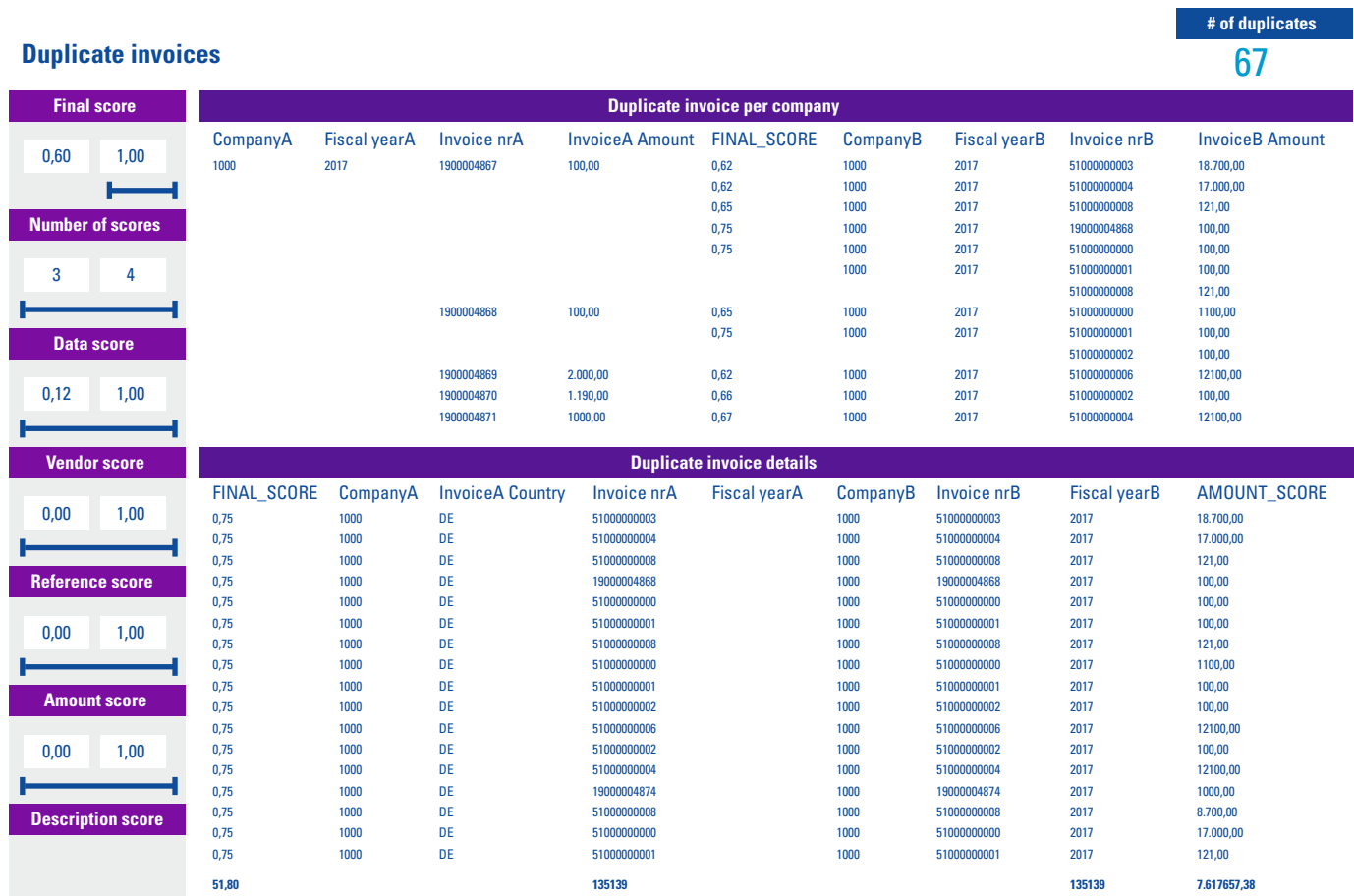
Reporting

In the case discussed, SAP GRC Process Control is being used to further process/distribute the results. We also did a proof of concept using a visualization solution to display the results (in this case MS PowerBI).

Most visualization solutions allow direct connections to SAP HANA. This way, the strength of SAP HANA can be used, and the dashboard is updated in real-time.

As shown in Figure 3, the results are available in a dashboard with details of each likely duplicate invoice. The score sliders on the left-hand side of the dashboard allow filtering on the results in real time. The sliders allow filtering on the minimal average (final) score, the number of scores/parameters to consider executing the analysis and the minimal score values per score.

Figure 3. Duplicate invoice results in MS Power BI.



A simplistic analysis results in a binary outcome: either the invoices are duplicates or not

CONCLUSION: THE BUSINESS CASE

In this article we explained a case of advanced duplicate invoice analytics. Our experience is that many companies are struggling with duplicate invoices. The incoming invoice management process is an inefficient process, because it contains different checks and balances. Since the result of this process is an outgoing payment, the execution of the process is crucial. However, the technical support offered by ERP systems is limited most of the time. Therefore, we propose to use the latest technology to make this process more efficient.

The solution described in this article can be applied to different other fields. Examples are duplicate master data, sales order monitoring, and even supporting the decision-making when ordering materials.

About the authors

Stefan Zuijderwijk MSc RE is a senior manager at KPMG Enterprise Solutions. Stefan works as a SAP risk consultant since 2009 and is specialized in analytics and dashboarding. In the past four years he has implemented several SAP HANA solutions.

Sylvester van der Giesen MSc is a manager at KPMG Advisory. He has been with KPMG since September 2012 and focuses on the intersection between business and IT with a strong focus on SAP products. In this role he has grown as a subject matter expert in SAP GRC Process Control and Risk Management.

SAP Process Control setup requirements

To enable SAP Process Control, there are some prerequisites.

HANA views

HANA analyses are developed in so-called SAP HANA views. The HANA views are stored in a certain location in the HANA database (e.g. Package) and that location has a 'path'. The path, including the view name can only consist of 60 characters, otherwise PC will not be able to find it once the connection is made.

Furthermore, the field types supported by SAP PC are limited. The only supported field types are: Integer, Date, Decimal, and NVarChar. Any other field types used will cause an error in the connection.

HANA connection

To set up the connection, a database connection must be established from SAP GRC to SAP HANA. The connection can only be set up if the HANA client is installed on the SAP GRC application server. Once this is done, the connection can be set up in GRC via transaction DBCO.

The connection user needs to have enough authorizations in SAP HANA to retrieve the results of the views that have been set up for follow-up in PC.

Data source and business rule

Once the connection configured, a data source and business rule need to be created to enable the monitoring in PC. For the data source it is important to select the correct integration scenario, in this case: HANA Integration. When setting up the data source, the correct view needs to be selected. The data source can be tested by running an ad-hoc query to see if any results are returned.

The business rule serves to determine whether exceptions are reported and what rating they would get. In the case of the duplicate invoices, scores below 80% are not considered, scores between 80% and 90% are considered as potential duplicate, scores between 90% and 95% are considered likely duplicate and scores over 95% are considered as highly likely duplicate. These thresholds can be set differently based on requirements. Additionally, filters could be placed to only display values over a certain amount (e.g. 500.000 Euros).