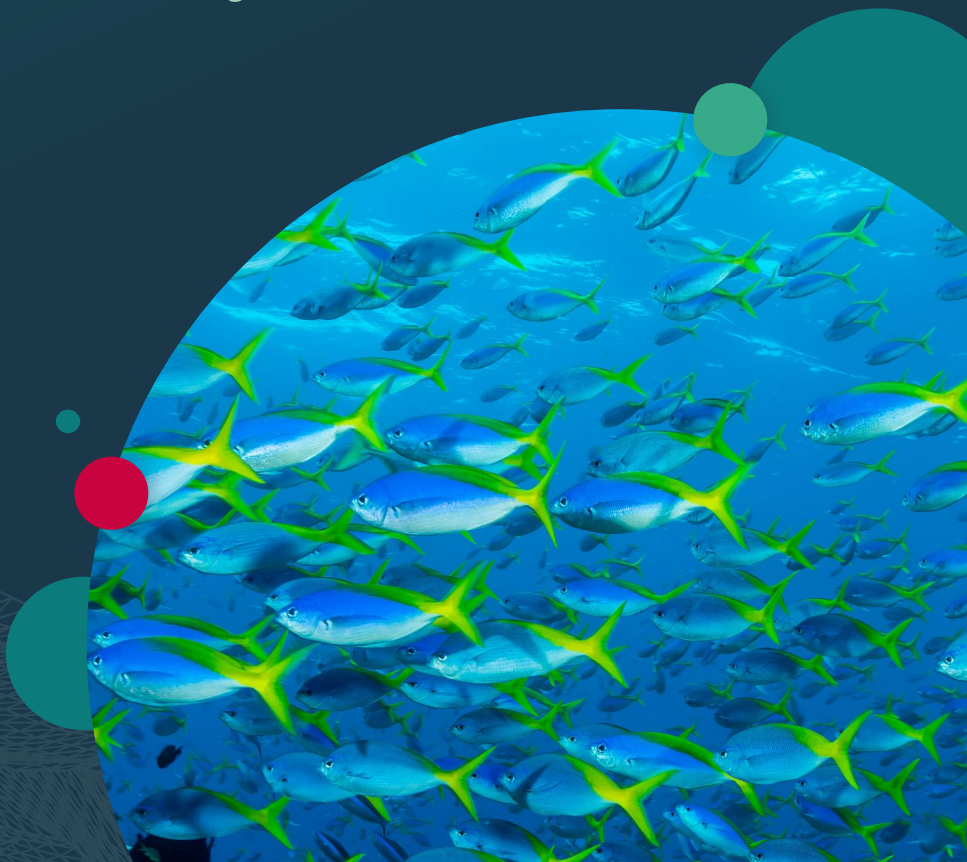
Migrating to SAP S/4HANA with duplicate-free master data

> Discover the essence of mastering data accuracy for a seamless shift to SAP S/4HANA and how deduplication process can optimize the transition.

In this paper, you will learn:

- best practices for achieving precise data for seamless S/4HANA migration
- expert recommendations for cost-efficient deduplication
- transformative impact of enhanced data quality on operations and decision-making
- how CDQ's deduplication works and sets the stage for a successful migration and future data integration



Introduction

In this whitepaper, we are focusing on business partner data, meaning master data for suppliers/vendors and customers.

The shocking truth is that countless companies are pouring in millions of Euros and often dedicating as much as 5 years to S/4HANA migrations. But here's the catch: if your data isn't set for the transformation, you might as well be burning your money!

You expect SAP S/4HANA to be lightning-fast compared to R/3, but speed without precision is simply fast in the wrong direction...

Today, financial controllers are caught up in an endless loop: initiating a laborious SAP transaction, generating a report, downloading it, wrestling with manipulations to fix the erroneous data, and finally, presenting it to the CFO. But in S/4HANA, it's all built-in by design – numbers are ready at your fingertips, and your CFO can access real-time data through the Fiori dashboard.

However, the reports are still inaccurate because the input data is flawed. And if your data isn't clean from duplicates, all the game-changing potential of SAP S/4HANA goes to waste.

FACTBOX

On average, 16% of business partner records in any organization are unintended duplicates.
Considering process costs, data acquisition expenses, and the impact on strategy decisions we have estimated, together with the CDQ Data Sharing Community member companies, that the cost of a single duplicate can be as much as 1,000 Euros!

Duplicate and inconsistent master data lead to operational inefficiencies, increased costs, and hinder decision-making processes.

This whitepaper outlines the process of deduplicating customer and vendor master data with CDQ Duplicate Matching capabilities before migrating to SAP S/4HANA.

Scope

Maintaining accurate and up-to-date master data is crucial for organizations of all sizes. Master data, such as business partner information, is often collected and stored in various systems across an organization.

A systematic approach to identify and manage duplicate records within a dataset is particularly critical during significant system changes, updates, or mergers, such as the mandatory migration to SAP S/4HANA by 2027.

> **Industry view: Raising the data quality in various SAP implementations, including migration to S/4HANA, is a topic for many organizations at the moment.**

Software implementation, configuration, testing, rollouts, and training are all challenges, but imports of master and transaction data are perhaps the most challenging of all. The legacy system's master data must be checked, validated, and prepared for the new system.

Business partner master data deduplication is a foundational step in preparing for S/4HANA migration. It ensures data quality, system performance, cost efficiency, streamlined processes, informed decision-making, compliance, user satisfaction, and the long-term success of the migration initiative.

Deduplicating master data records for SAP S/4HANA migration

In the context of S/4HANA migration, these typical challenges should flag a need for master data deduplication efforts:

1. **Legacy Data:** the presence of extensive legacy data adds complexity during the migration process. Data managers struggle to determine whether two records are in fact duplicate records, and the need for expert review from the business becomes an additional layer of complexity and pressure.
2. **Lack of duplicate prevention in ERP systems:** absence or weak configuration of built-in duplicate prevention mechanisms in the current ERP systems intensifies the risk of data redundancy during migration.
3. **Absence of data archiving activities:** lack of historical data archiving practices, particularly concerning master data, complicates the migration.
4. **Ongoing storage capacity and software license purchases:** the need to continually invest in additional storage capacity due to unchecked data growth has financial implications during migration.
5. **No standardized process for customer and vendor creation:** the absence of standardized procedures for business partner data creation hinders data consistency and accuracy during migration.

While many companies spend months defining the business partner concept, our experts from CDQ Business Consulting can help to establish a clear scope of the project and support your teams at all stages of implementation..

Many companies are moving towards centralized IT systems, reducing customization, and consolidating ERP instances. Legacy companies often use various systems (CRM, ERP, WMS, TMS) per business unit, country, or process, leading to extensive customer and vendor data. To streamline operations, data from hundreds of systems needs to be consolidated into a few dozen.

In the migration journey, organizations typically go through a „*stress curve.*" It begins with simple discovery workshops but becomes more demanding over time, especially for business users. At its peak, the need for data cleansing becomes critical, yet often gets sidelined due to time constraints.
This highlights two key priorities:

1. tackling data cleansing before migration, which is frequently neglected, and
2. maximizing automation where feasible.

Addressing these challenges prior to the S/4HANA migration is imperative to ensure a successful transition with clean, reliable, and optimized data.

CDQ recommended approach

1. **Preparing systems for S/4HANA migration**: including establishing of change management process
2. **Deduplication**: clean up outdated legacy data for the active ERP systems.
3. **Automation**: strive for automated duplicate identification and cleansing.
4. **Maintain quality**: ensure uniformity of data records through quality rules in the ERP systems.

CDQ supports **every step** of the deduplication process for your S/4HANA migration:

1 Data scope definition

Gain clarity from the start. Collaborate with CDQ to clearly define the data scope. This ensures that only the relevant data is targeted for deduplication, preventing wasted effort.

2 Precise duplicate identification

Streamlined, error-free data. CDQ provides advanced data analysis tools, enabling thorough examination and precise identification of duplicate records. This guarantees data accuracy and integrity.

3 Effortless proposal for deletion flags

Clear, structured recommendations. CDQ helps create a structured proposal for marking duplicate records with deletion flags. This ensures a clear plan, reducing confusion and uncertainty.

4 Client-focused review process

Efficient, targeted review. In the review phase, we emphasize efficiency by using representative samples, reducing the workload for the client. This targeted approach ensures alignment with your expectations.

5 Configuring automation with CDQ

Time and resource savings. We collaborate with you to configure automation tailored to your requirements. This not only saves time and resources but also enables automated deduplication based on your business definition of duplicates.

In the context of preparing for S/4HANA migration, data cleanup initiative underscores the business value of data quality – once a thorough examination reveals the presence of duplicate entries, it becomes evident that their presence is undesirable within the new system. With advanced tools and custom rules, the organizations enhance data accuracy, leading to significant cost savings and improved customer service capabilities post-migration.

SAP One Domain Model as baseline

Whether companies choose to handle their future master data processes in S/4HANA or SAP MDG, they will create and maintain business partner data in alignment with the standardized SAP One-Domain-Model (ODM). CDQ delivers business partners with the out-of-the-box mapping already in the SAP ODM, so that data records can be imported directly into S/4HANA or SAP MDG. It's a strategic move towards seamless transition and more efficient business operations.

Industry sneak peek: In CDQ customer's master data system, some regions were burdened with up to 30% duplication rate. Focusing on the top 29 countries, which constituted 80% of the dataset, revealed 740,000 potential duplicates. A successful resolution of 660,000 duplicates represented 17% of the customer records in scope. Additionally, the project not only eliminated redundant data but also established an enduring cleansing strategy. This encompassed global duplicate definitions, a cross-company review process, and an innovative review tool.

OUTCOMES

1. **Duplicates elimination**: duplicate and legacy records are removed, streamlining data across the organization and all divisions.
2. **Cost avoidance**: deduplication will result in lower costs for the migration to S/4HANA, reducing the number of records and avoiding additional licensing and storage fees.
3. **Cost reduction**: additional savings can be anticipated in terms of migration project duration and associated labor costs. Accurate data in the new system enhances its efficiency and lowers the risk of operational hiccups that can be costly to resolve.
4. **Sustainability**: focus on long-term sustainability of clean data, ensuring that data quality remains high after the cleanup.
5. **Streamlined business processes**: clean data will facilitate smoother and more efficient business processes within the new environment and provide support to future carve-in and carve-out activities.

Customer key highlights

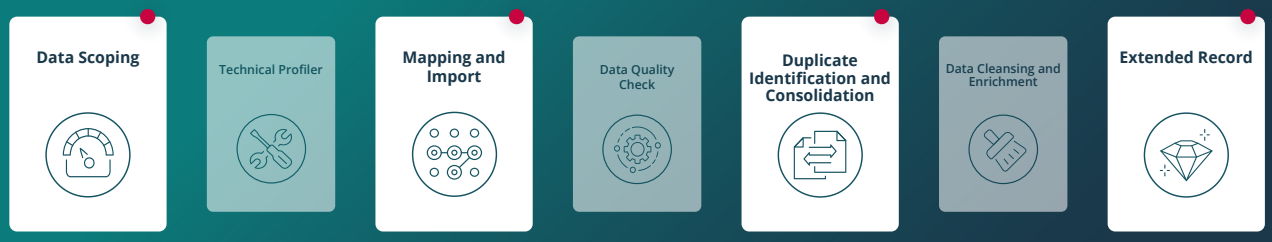
- CDQ guided a leading global integrated technology group through the cleanup of 290,000 business partner master data records within a timeframe of 18 months.
- With data records from 83 different companies spanning 15 distinct systems, an average of 61% duplicate and legacy entries were removed per system (with the highest being 79%).
- Automation was increased through CDQ's out-of-the-box data validation tools, which include a robust set of over 2,300 data quality rules and a well-developed duplicate detection algorithm. This contributed to making the data cleansing processes more efficient and streamlined.
- Duplicate clean-up resulted in cost avoidance of €2,695,000 for the S/4HANA migration. Approximately 60% of duplicates were identified and removed.

Transitioning to S/4HANA involves adopting the business partner object for master data management, replacing the customers and vendors objects. Many companies implement SAP MDG for business partner creation, maintenance, and governance, leveraging CDQ to enhance data quality.

Combining all business partner data into a single table would yield numerous duplicates, posing challenges for identification. CDQ helps resolve these issues by distinguishing between full names and abbreviations, sorting ship-to and bill-to addresses, and reconciling variations in data attributes across systems. CDQ imports data from relevant systems into a Data Mirror, standardizing it according to the SAP One Domain Model Business Partner for consistency and comparability. This ensures business partner data is uniform and standardized.

CDQ holistic data cleansing approach

CDQ's approach to deduplication is embedded in a **holistic 7-step** Get Clean process that is designed to improve data quality for various use cases, including the migration to S/4HANA.



Note: Click on the steps to navigate through the document.

● Deduplication Services provided by CDQ

Deduplication services powered by CDQ

Within a CDQ-powered Get Clean project the following steps show how deduplication service is performed with the complex intelligence built into the algorithm-based matching capabilities.

DATA SCOPING

Performed activities:

1. Data scope: reserve enough time for identifying the relevant data scope for the deduplication. E.g., selecting the relevant customer or vendor data of a specific country, a specific source system, active records of the last two years, etc.
2. Field scope: identify attributes that should be included in an extract, e.g., name, address, identifier, account groups, etc. Although the core data fields are not too many and typically similar in deduplication initiatives, additional company-specific fields may be necessary e.g., to clearly define a duplicate.
3. Project Scope: clarify the question "who will be involved?". Who is able and allowed to define the scope? Who owns a company-wide definition of a duplicate? Who decides on the desired data quality level and the thresholds? Who owns the data itself and can decide on record level?

CDQ Recommendation

- A common pitfall is to discover later in the process (e.g., when duplicates are being reviewed and merged) that the data scope is incomplete. This can be avoided by having all relevant stakeholders decide and approve the scope.
- When creating the field scope, consider all fields needed to identify a duplicate, but also include fields that are needed to evaluate duplicates in the review (making decision if duplicate is indeed a duplicate and how to merge them), e.g., number of transactions.
- As the GDPR regulations enforce specific treatment of personal data, natural persons must be excluded from the scope. CDQ may help in this context as well.

MAPPING & IMPORT

Performed activities:

1. Generate a data extract, based on the data and fields scope defined in previous step. The data can be provided manually in a CSV or text file. The data can also be transferred via API in batch mode or on a record level.
2. Based on your input, data fields of the data extract are mapped to CDQ's data model. This ensures a compatibility to the various data models of the external data sources and allows the successful application of CDQ data quality rules.

3. CDQ sets up the so-called data mirror in the CDQ Cloud and uploads the data from the data extract. The data in the data mirror is the basis for all subsequent activities.

CDQ Recommendation

- When extracting the data manually, stick to a CSV file in UTF8 Format. This ensures that the data content and format is not altered, e.g., by Excel auto format functions.
- Data mirror upload can be automated by implementing an API for data mirror sync. This way the data can be updated regularly to ensure that scope is not obsolete. This will save manual effort particularly when the duplicate check is performed periodically.
- An initial thorough mapping approach will save time in the long run and ensure accuracy. All relevant fields are now mapped, particularly country specific identifiers. Extracts from different systems can be uploaded to the data mirror. The mapping ensures that all data has a common denominator, the CDQ data model. Once mapped to the CDQ data model, all CDQ services, including the duplicate matching, can be applied to the uploaded data.

DUPLICATE IDENTIFICATION & CONSOLIDATION

Performed activities:

1. Define what attributes must match (or show similarity) to be considered a duplicate. Typically, this definition starts with a recommendation from CDQ, based on best practices of the Data Sharing Community. It is an iterative approach, that is supported by examples from the data in the data mirror.
2. Set up a duplicate **matching configuration** (algorithm) based on duplicate definition. The algorithm which is used for identifying duplicate candidates (i.e., the „matching") can be configured according to individual needs.

Matching configuration

Matching configuration lies at the core of CDQ deduplication process.

Matching configuration defines which attributes are to be compared (which comparator should be used for which attributes), which impact identical, similar, or different values of these attributes have on the matching score (confidence scores, how values should be temporarily transformed (which cleaners should be used), which threshold the matching score of potential matches should be exceeded to be considered a duplicate.

CDQ offers a fully flexible configuration:

- Any set of attributes can be used for identifying duplicates.
- Cleaners and comparators can be individually configured.
- Thresholds can be individually configured.
- Configurable consolidation rules for identified duplicates allow for different Golden Record generation strategies based on decision trees.

You can use standard configurations and optimize them iteratively for any specific use case.

Configuration building blocks:

A Cleaners for each attribute

By cleaners, you can remove characters which are not needed (e.g., non-digits, whitespace) or normalize your data (e.g., all lower case, no accents). Note: Order of cleaners matters.

Other available cleaners:

- Attribute in Attribute Cleaner
- **Country Cleaner**
- **Legal Form Cleaner**
- Digits-only Cleaner
- Lower-case Normalize Cleaner
- Non-character Cleaner
- **Phone Number Cleaner**
- Punctuation cleaner
- Replace Cleaner
- Strip Non-text Characters Cleaner
- Trim Cleaner

CDQ special cleaners

ATTRIBUTE	CLEANERS*
Country	1. Lower Case Normalize Cleaner
Name	1. Legal Form Cleaner 2. Lower Case Normalize Cleaner 3. Punctuation Cleaner 4. Care of Cleaner 5. Cleaner Word Freq Name
Street	1. Lower Case Normalize Cleaner 2. Punctuation Cleaner 3. Cleaner Word Freq Name
Postal Code	1. Punctuation Cleaner
City	1. Lower Case Normalize Cleaner 2. Punctuation Cleaner
STRCD1	

**Order of Cleaners matter*

Figure 1 Example cleaners per attributes

B Comparators for each attribute

By comparators, you can define the algorithms to use for data comparison per attribute (e.g., exact comparison or several fuzzy algorithms).

Other available comparators:

- Exact Comparator
- Geoposition Comparator
- Jaro Winkler Comparator
- Levenshtein Comparator
- Metaphone Comparator
- Soundex Comparator
- Weighted Levenshtein Comparator
- q-Gram Comparator
- Longest Common Substring Comparator

ATTRIBUTE	COMPARATOR
Country	Exact Comparator
Name	Name Comparator
Street	Levenshtein
Postal Code	Exact Comparator
City	Levenshtein
STRCD1	Exact Comparator

Figure 2 Example comparators per attribute

C Setting thresholds

The threshold tag within the matching schema defines the threshold for the similarity score to consider two records a match after comparison. Records with a match score above this value are compiled to matching groups, remaining records are listed as singles. The match score is calculated from the fields' comparison results with the Bayes' theorem.

Matching Attribute	Property Name	Lookup	Low	High	Comparator	Cleaners
addresses[0].country.shortName	Country	required	0	0,5	Exact Comparator	1. Lower-case Normalize Cleaner
names[0].value	Name	true	0,3	0,8	NameComparator	1. Lower-case Normalize Cleaner, 2. punctuationCleaner, 3. careOfCleaner, 4. cleanerWordFreqName
addresses[0]postCodes[0].value	Postal code	true	0,3	0,6	Exact Comparator	1. Digits-only Cleaner
addresses[0].localities[0].value	City	false	0,2	0,6	Levenshtein Comp...	1. Lower-case Normalize Cleaner, 2. punctuationCleaner
addresses[0].thoroughfares[0].value	Street	true	0,3	0,7	Levenshtein Comp...	1. Lower-case Normalize Cleaner, 2. punctuationCleaner, 3. cleanerWordFreqStreet
Umsatzsteuer-Id.Nr	STCD1	false	0,2	0,7	Exact Comparator	Select...

Figure 3 Example threshold configuration

The configurator app in the CDQ Web Apps includes: decision on which fields should be compared, adding cleaners for each relevant field, setting an overall threshold and what impact each field should have to identify a duplicate. at the moment.

Run duplicate matching job

The duplicate matching configuration is applied to the data scope. Business partner data will be sorted into matching groups based on similarity measure across multiple fields.

Run the job for the algorithm to solve the typical challenges when matching business partner names or addresses, such as:

- The representation of the name is not standardized. While for some names abbreviations are used (e.g., “Inst. of Information Management”), others include legal forms (“ZF Friedrichshafen AG”) or are written in uppercase or lowercase letters only (“BAYER AG”).
- The name or address is represented in different characters (e.g., Chinese, Cyrillic, and/or Latin characters).
- The name or address includes misplaced information. For example, c/o information is added to the name instead of the respective data attribute, or the house number is sometimes included in the street name and sometimes placed separately.
- In some cases, acronyms are used, while in other cases the full name is used (e.g., “BMW AG” vs. “Bayerische Motorenwerke AG”). Similarly, abbreviations are used inconsistently (e.g., “Lindenstr.” vs. “Lindenstrasse”).
- Misspellings may occur in various forms: characters added (e.g., “Bayern AG” instead of “Bayer AG”), characters omitted (e.g., “Byer AG”), characters replaced (e.g., “Baier AG”), or characters transposed (e.g., “Bayre AG”).
- The order of name components is inconsistent (e.g., “Lindner Hotel Hamburg” vs. “Hotel Lindner in Hamburg”).
- The name is represented differently (e.g., only one attribute “business partner name” in one data model vs. “name 1-5” in SAP’s data model).
- Missing attributes (e.g., one system includes building information, another does not, or data was just not maintained).
- Original names of cities vs. international names (e.g., “München” vs. “Munich”, “Mailand” vs. “Milano”).
- Semantic ambiguities for certain fields (e.g., different post codes available in Ireland: Eircode vs. GeoDirectory vs. Loc8 Code).

Special feature: the duplicate matching tool is independent of data model. This means any field can be uploaded and used in the duplicate identification process. Furthermore, any language/character set can be included in the data scope (comparisons are only possible between the same character set).

Matching Score	External ID	Name	Country	Postal code	City	Street
1	6100042	上海赫能实业有限公司	CN		上海	上海市松江区江田东路185号9栋
0,96	6100487	上海赫能实业有限公司	CN		上海	上海市松江区江田东路185号9栋
0,941176471	100281	上海赫能实业有限公司	CN			上海市松江区江田东路185号9栋
1	100452	太原市畅吉物流有限公司	CN			清徐县徐沟镇北关村北关大街13-1号
0,941176471	200528	太原市畅吉物流有限公司	CN			清徐县徐沟镇北关村北关大街13-1号
0,941176471	300464	太原市畅吉物流有限公司	CN		太原	清徐县徐沟镇北关村北关大街13-1号

Create duplicate report

Duplicate reports as a standard Excel file are also available as a dashboard duplicate overview (PowerPoint). There are two types of reports:

Extended Record

Duplicate Matching Report

The report is the result of the duplicate matching comparing all records of a given set of custom databases to each other, that identifies similar records, and groups „best matches“ in matching groups.

Matching Group	Matching Score	External ID	NAME	COUNTRY	POSTAL_CODE	CITY	STREET	ID_1	ID_2	Matching Explanation
1062dfba-29ca	1	0066	Vily Stav AR	CZ	741 01	Novy Jicin, Loucka	Jicinska	20390885		Pattern record
1062dfba-29ca	0,97288259	0103	Vily Stav	CZ	74101	Novy Jicin	Jicinska	20390885		overall: [CITY: 0.5830123456790123; overall:
e6e6eb62-32d	1	DE-03519	Hecker Service GmbH	DE	44287	Dortmund	Schleefstr.	DE812727397		Pattern record
e6e6eb62-32d	0,84	DE-03159	Hecker Glaskeramik GmbH & Co. KG	DE	44287	Dortmund	Schleefstr.	DE812727397		overall: [CITY: 0.6; overall: 0.6; COUNTRY: 0.
06e96a4c-a0af	1	FR-00710	DUCLOS BOIS ET MATERIAUX SAS	FR	17400	POURSAY GARNAU	2 ROUTE DE POITIERS	FR74349564773	349564773	Pattern record
06e96a4c-a0af	0,866024	FR-00291	DUCLOS BOIS ET MATERIAUX SAS	FR	17180	PERIGNY		FR74349564773	349564773	overall: [CITY: 0.5428089795918367; overall:
06e96a4c-a0af	0,86263117	FR-00293	DUCLOS BOIS ET MATERIAUX SAS	FR	17430	TONNAY CHARENTE		FR74349564773	349564773	overall: [CITY: 0.5356200554295792; overall:
06e96a4c-a0af	0,86218229	FR-00292	DUCLOS BOIS ET MATERIAUX SAS	FR	17600	SAUJON		FR74349564773	349564773	overall: [CITY: 0.534679012345679; overall: (
06e96a4c-a0af	0,84482759	FR-00289	DUCLOS BOIS ET MATERIAUX SAS	FR	17260	GEMOZAC		FR74349564773	349564773	overall: [CITY: 0.5; overall: 0.5; COUNTRY: 0.
06e96a4c-a0af	0,84482759	FR-00290	DUCLOS BOIS ET MATERIAUX SAS	FR	17290	LE THOU		FR74349564773	349564773	overall: [CITY: 0.5; overall: 0.5; COUNTRY: 0.

Duplicate Consolidation Report

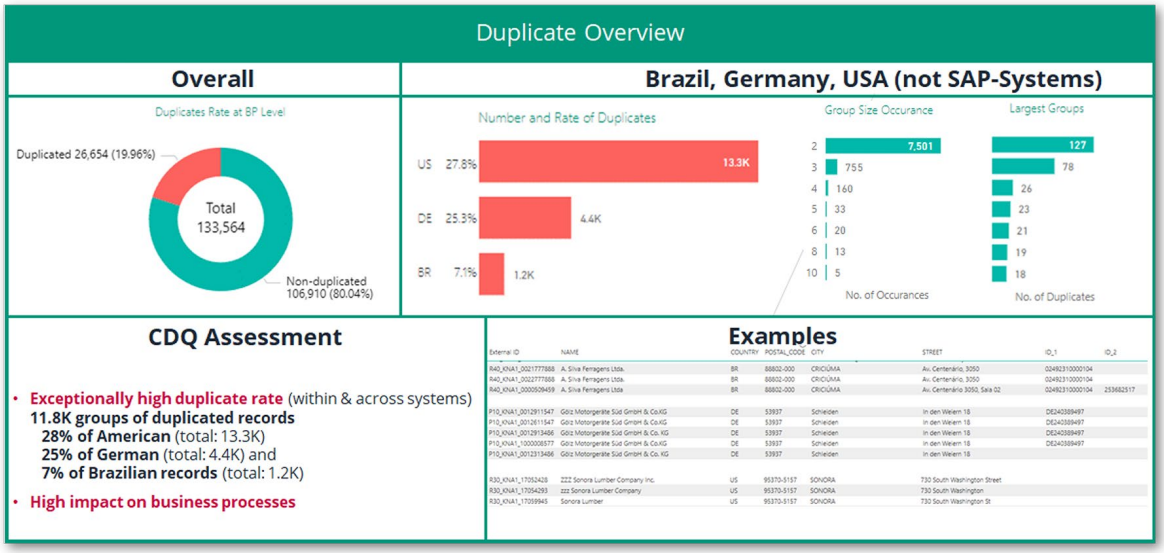
The report is the result of the duplicate matching within one dataset employing standard or custom configurations, that consolidates duplicates into Golden Records, based on golden record configurations. Such an Extended Record is a consolidated record from duplicate matching report with enriched data from data cleaning and enrichment, for a single and most comprehensive view on a given business partner.

Matching Group	Matching Score	External ID	NAME	COUNTRY	POSTAL_CODE	CITY	STREET	ID_1	ID_2	Matching Explanation
1062dfba-29c	*1.0*	*0103*	*Vily Stav AR*	*CZ*	*741 01*	*Novy Jicin, Loucka*	*Jicinska*	*20390885*		*Golden Record*
1062dfba-29c	1	0066	Vily Stav AR	CZ	741 01	Novy Jicin, Loucka	Jicinska	20390885		Pattern record
1062dfba-29c	0,9728826	0103	Vily Stav	CZ	74101	Novy Jicin	Jicinska	20390885		overall: [CITY: 0.5830123456790123; overall: 0.5
e6e6eb62-32	*1.0*	*DE-03159*	*Hecker Service GmbH*	*DE*	*44287*	*Dortmund*	*Schleefstr.*	*DE812727397*		*Golden Record*
e6e6eb62-32	1	DE-03519	Hecker Service GmbH	DE	44287	Dortmund	Schleefstr.	DE812727397		Pattern record
e6e6eb62-32	0,84	DE-03159	Hecker Glaskeramik GmbH & Co. KG	DE	44287	Dortmund	Schleefstr.	DE812727397		overall: [CITY: 0.6; overall: 0.6; COUNTRY: 0.5; o
06e96a4c-a0	*1.0*	*FR-00710*	*DUCLOS BOIS ET MATERIAUX SAS*	*FR*	*17400*	*PERIGNY*	*2 ROUTE DE POITIERS*	*FR74349564773*	*349564773*	*Golden Record*
06e96a4c-a0	1	FR-00710	DUCLOS BOIS ET MATERIAUX SAS	FR	17400	POURSAY GARNAU	2 ROUTE DE POITIERS	FR74349564773	349564773	Pattern record
06e96a4c-a0	0,866024	FR-00291	DUCLOS BOIS ET MATERIAUX SAS	FR	17180	PERIGNY		FR74349564773	349564773	overall: [CITY: 0.5428089795918367; overall: 0.5
06e96a4c-a0	0,8626312	FR-00293	DUCLOS BOIS ET MATERIAUX SAS	FR	17430	TONNAY CHARENTE		FR74349564773	349564773	overall: [CITY: 0.5356200554295792; overall: 0.5
06e96a4c-a0	0,8621823	FR-00292	DUCLOS BOIS ET MATERIAUX SAS	FR	17600	SAUJON		FR74349564773	349564773	overall: [CITY: 0.534679012345679; overall: 0.53
06e96a4c-a0	0,8448276	FR-00289	DUCLOS BOIS ET MATERIAUX SAS	FR	17260	GEMOZAC		FR74349564773	349564773	overall: [CITY: 0.5; overall: 0.5; COUNTRY: 0.5; o
06e96a4c-a0	0,8448276	FR-00290	DUCLOS BOIS ET MATERIAUX SAS	FR	17290	LE THOU		FR74349564773	349564773	overall: [CITY: 0.5; overall: 0.5; COUNTRY: 0.5; o

What happens next?

The accuracy of the reports depends on the matching configuration and inbound data quality. The results can be more or less accurate depending on the configuration, so the report needs to be reviewed before using the results.

→ **Analyze/review duplicate report and finetune duplicate matching configuration if necessary.**



Graphical representation of existing duplicates

Business review

Review the duplicate report to make a final decision which duplicates can be eliminated and/or merged. This decision can be automated based on matching scores. Furthermore, a master record can be selected. The master record will be the leading record when consolidating duplicates. The decisions from this review can be uploaded in the next duplicate matching job (decision log).

CDQ Recommendation

- Use a lower threshold for the initial duplicate job. The threshold can then be finetuned after the first review.
- Perform a sample review and decide which matching score level is accurate enough to automate the review decision.
- Include identifiers in the scope. Identifiers such as EU VAT, business registration numbers, etc. can improve the accuracy of the results significantly.

Closing remarks

The tasks mentioned above serve not only to streamline, refine, and improve business partner data but also play a vital role in integrating historical data into the appropriate data model. Equally crucial is the ongoing effort to uphold data accuracy and ensure the accurate and consistent creation of new business partners.
It's important to note that a well-maintained business partner database will only stay up to date for a short period if no further actions are taken. Sustaining data cleanliness requires the continuous utilization of suitable tools and processes with a high degree of automation, as 21% of master data outdate within 12 months from initial creation.

When considering future acquisitions, the necessity of cleansing and eliminating duplicates from master data will persist as new companies and datasets are integrated. As a result, projects aimed at achieving data cleanliness are foundational but must be supported by robust processes.

CDQ simplifies maintenance efforts by facilitating updates for business partners from various sources, including official business registers and the CDQ Data Sharing Community, and seamlessly integrates them into S/4HANA.

Additionally, CDQ ensures the creation of accurate and error-free business partners through CDQ First Time Right, which is also available in SAP store as an SAP-endorsed app. Integrated into the business partner creation process, this solution utilizes intelligence derived from over 2,300 data quality rules and data sourced from over 70 official business registers at a click of a button.

Migrate to S/4HANA duplicate-free!

CDQ accompanied more than 300 data quality projects of leading Forbes 2000 companies and has been actively shaping a more sustainable approach to master data management for 18 years.

Are you struggling with duplicates?

Let's tackle the challenge together with the data quality approach!

Oliver Reh
Head of Customer Success

E-Mail
oliver.reh@cdq.com

LinkedIn
/oliver-reh/

CDQ AG
Lukasstrasse 4
CH-9008 St. Gallen

Follow us on social media:

LinkedIn YouTube

Visit CDQ website:

www.cdq.com