eBook

The Essential Guide to Customer Data Platform Success

matillion

Executive Summary

In this comprehensive whitepaper, we delve into the critical role of a Customer Data Platform in driving business success by increasing revenues and profitability while effectively managing operational costs. At the core of a Customer Data Platform lies its ability to centralize and unify customer data, offering a complete perspective on customer identities and transaction histories. Beyond conventional metrics, a Customer Data Platform empowers businesses to delve into leading indicators such as customer satisfaction and value, enabling proactive responses to evolving customer behaviors.

Efficient data organization is key, and the star schema model comes into play by reducing data duplication. Cloud data platforms, such as Snowflake and Databricks, provide the ideal infrastructure due to their scalability and seamless compatibility. We explore two fundamental data management approaches: the centralized and federated models. The centralized model places data control in a single team, ensuring uniformity. In contrast, the federated model delegates data ownership to domain-specific teams, fostering agility.

Effectively harnessing data means tapping into diverse sources, including CRM systems, email campaigns, and website analytics. Turning raw data into actionable insights involves safeguarding PII data privacy, creating unified customer profiles, and rigorous data cleansing.

To create analytics-ready data sets, we delve into techniques such as key structuring and data manipulation. These methods enhance our understanding of customer demographics and behaviors. Data aggregation is pivotal for streamlined data analysis and is a cornerstone of a productive data warehouse. We explore the role of database functions and efficient data processing. Syncing enriched customer data back to business applications, known as reverse ETL, enhances accessibility and integration.

A well-implemented Customer Data Platform empowers businesses to cultivate customer relationships, fine-tune marketing strategies, and enhance product development, ultimately leading to increased revenue and fortified customer loyalty. Whether choosing a centralized or federated data management model, along with leveraging Matillion's capabilities, profoundly influences the success of a Customer Data Platform implementation.

Contents

01. Data organization in a Customer Data Platform

02. Best practices for transforming customer data

03. Optimizing your customer data for business analytics

06. Next Steps

01

Data organization in a Customer Data Platform

The bottom line for every business is that they want to make money by increasing revenues and profits. And the only way to do that is to have more customers spending more money on your products and services while keeping your costs down.

A Customer Data Platform is essential to achieving these results. Because it collects all customer information, such as customer identity and transaction history, in one place, a Customer Data Platform helps you measure your success in acquiring and keeping customers. You can analyze a wide variety of metrics, such as customer acquisition costs (CAC), net dollar retention (NDR), the percentage of revenue retained from existing customers, and the overall customer life cycle (CLC). While it's valuable to understand these metrics, they are ultimately lagging indicators that show what has happened in the past. An excellent way to lose customers is to only look at past behavior and treat them like walking wallets, assuming their behavior or preferences will remain the same.

The greater value of a Customer Data Platform is the ability to collect and analyze information obtained from leading indicators that measure a customer's satisfaction and how much value the customer is getting. These customer-focused metrics analyze behavior and preferences through things like web browsing, social media, demographics (industry vertical, location, size, etc.), and interactions with the business (support, chatbot, community activity, etc.)

With this information, a Customer Data Platform now provides a more objective way to measure the value obtained by each customer versus spend. This allows you to become more proactive when dealing with customers and react better to changing customer behaviors.

- Make sure customers are gaining all the value they can
- Fine-tune marketing and advertising campaigns and create more targeted recommendations
- Improve product development and offering strategies

Ultimately, a Customer Data Platform can help businesses achieve their primary goal of increasing revenues through increased customer loyalty and improved customer service.

Streamlining Data for Better Insights: The Star Schema Model

A common way to organize and store your data efficiently in a Customer Data Platform is to deploy a star schema model. In a star schema, different data dimensions – i.e., products, customers, time, etc. – are organized and stored in a manner that reduces duplication. This makes it easier to filter data in a data platform like a data warehouse, a database, or a data mart and makes it more efficient to query larger data sets. The figure below shows an example of the star schema model in a Customer Data Platform and how it can be used to analyze customer opinions of different products.

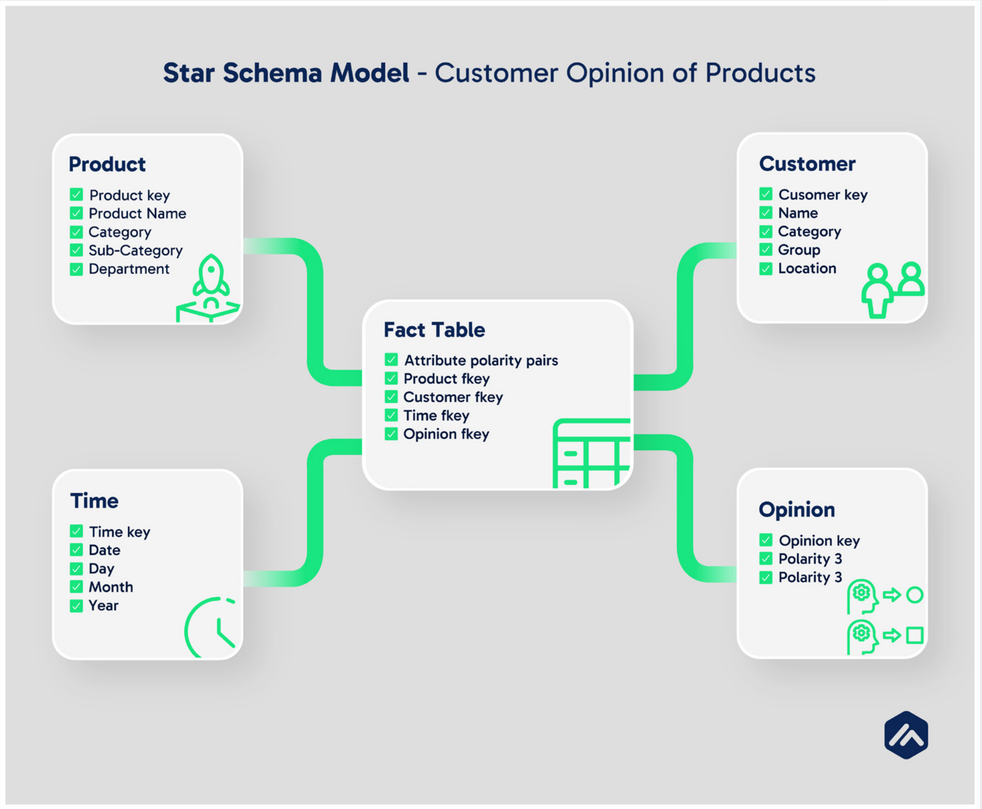


Figure 1. Star schema to measure customer opinions of particular products.

Centralized vs. Federated Data Management Approaches

Today, most modern enterprises rely on cloud data platforms, like Snowflake, Databricks, or Amazon Redshift, as the storage layer and the analytics backbone for their customer data platforms. These cloud-native platforms provide greater scalability, performance, and compatibility with BI, analytics, and governance solutions while also reducing infrastructure costs and maintenance. A critical decision in creating your Customer Data Platform is to determine how you are going to manage all of the data coming into your cloud data platform of choice. This is usually done using one of two main models.

1. **A centralized model**
2. **A centralized model**

1

Centralized model - one team to rule them all

In the centralized model, a single team is responsible for co-locating and integrating all the relevant sources of data into a central location – i.e., your chosen cloud data platform. The team is also responsible for ensuring that the correct customer identity is used in all cases and that the resulting data is easily consumable.

To make the Customer Data Platform actionable, this team will likely need to push the enriched output data somewhere else – most likely into a contact center database or back to source business applications – where the data can be viewed and analyzed by business users. Accomplishing this might require sync-back capabilities for your data environment.

A centralized model requires all the expertise and knowledge to exist in this one team. For example, ownership of the definitions of “customer,” “industry vertical,” and “products.” The same team also owns the technology stack and enforces governance.

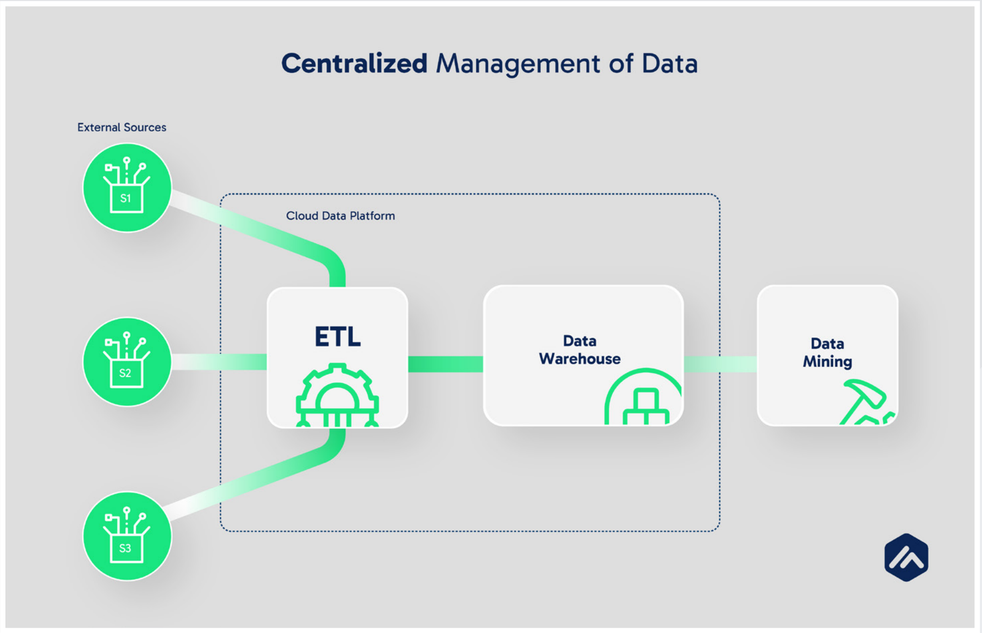


Figure 2. Centralized model with one team managing all data pipelines for the customer data platform.

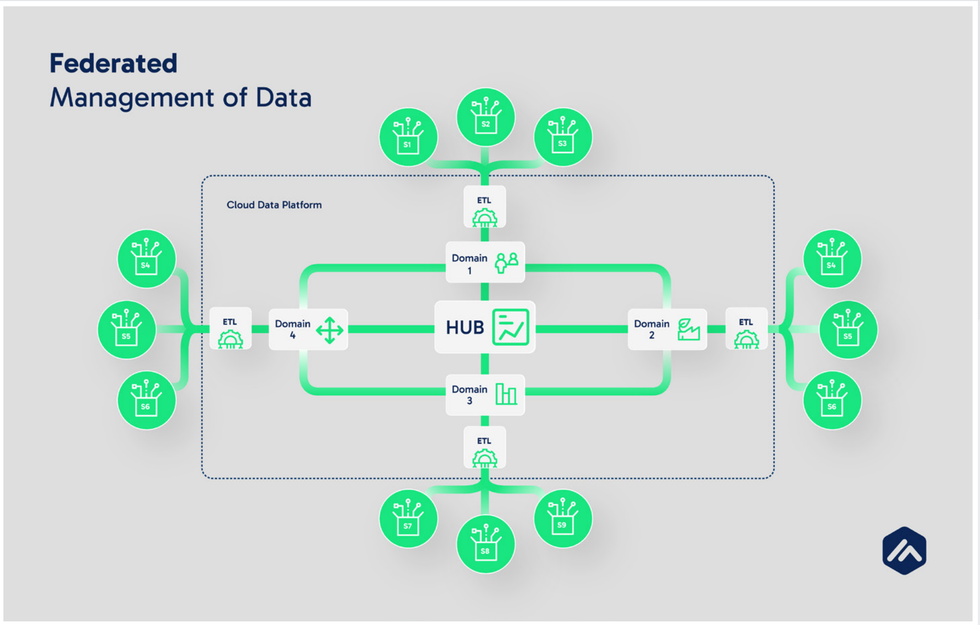
2

Federated model - every team, responsible, at once

The federated model assigns responsibility for managing the data to different teams with domain expertise. Instead of centralized data ownership and governance, each domain is assigned responsibility for a particular set of data based on knowledge and proficiency. So, while the customer data in a federated model, similar to a centralized approach, ultimately resides in a central cloud data platform, the responsibility for ownership and governance of different types of data resides with different teams.

In practice, this means that the marketing team might be responsible for customer data gathered from a customer's web browsing history or social media activity, while the sales team might be responsible for customer transaction data, and the support team is responsible for customer interaction data. A data mesh architecture, in which each domain is responsible for the data pipelines for the data they own, can be used to support a federated model.

An example of how the federated model works within a Customer Data Platform can be seen below:



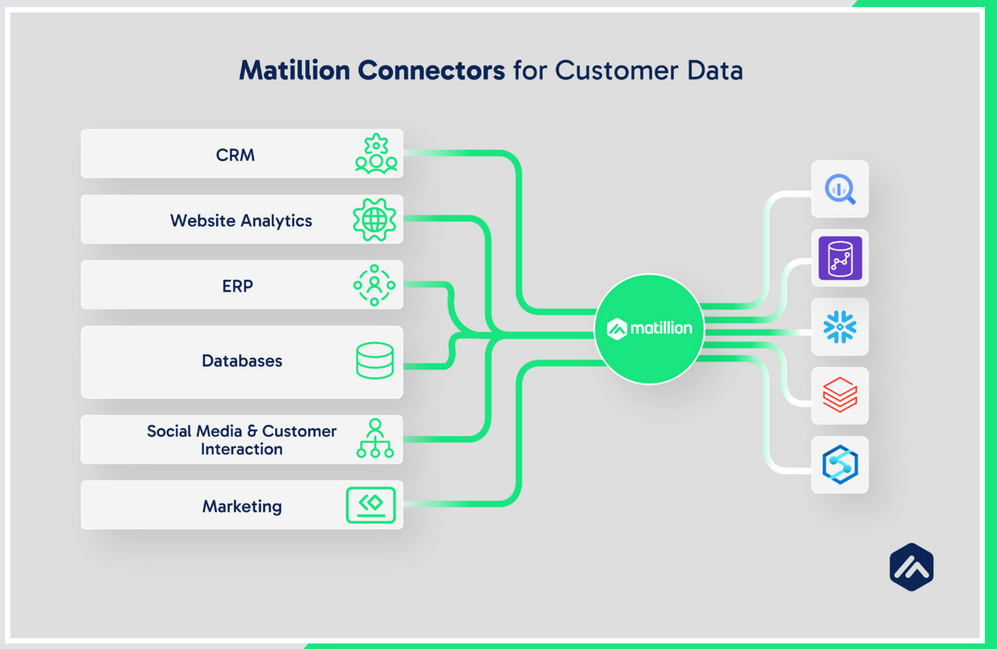
A central hub team still has a vital role in the federated model – to provide a tech stack and enforce governance. Gold/Silver/ Bronze is a sound governance methodology for this approach. Domain teams are encouraged and empowered to publish material. But if they want it to be treated as a "gold" standard, they must adhere to specific requirements.

Getting Data Into Your Customer Data Platform

No matter which approach you choose, the first step to better understanding your customers and making the right decisions to improve customer loyalty and engagement is ensuring you have all the right data. Any successful Customer Data Platform must have access to data from a wide variety of sources to enable a business to acquire a comprehensive view of its customers and their behaviors. The most commonly used data sources for customer data are customer relationship management (CRM) systems, email campaign solutions, website analytics tools, social media platforms, online surveys, loyalty programs, and third-party data providers.

The Matillion Data Productivity Cloud makes it easy to extract and load data from all your customer data sources into the chosen cloud data platform for your Customer Data Platform – whether it is a centralized approach with a single team loading the data or a federated model with various domain owners handling the task.

Matillion has built-in connectors to the most commonly used data sources and applications for customer data. And, with Matillion's Low-Code / No-Code interface, you can connect to a data source in just a few seconds. Matillion also offers Change Data Capture data loading to capture incremental changes to the data and the ability to connect to any REST API with the Create Your Own Connector framework.



02

Best practices for transforming customer data

- private, unified, and clean

The next step, likely the most critical component of any customer data strategy, is to transform your customer data – coming from many sources – into a unified format for reporting and analysis. You must consider three critical aspects: ensuring personally identifiable information (PII) privacy, creating a unified customer profile, and cleansing the data to ensure accuracy and relevance. This will, in turn, enable you to utilize customer information better and get value from it — understand how your business is performing and make the right decisions to improve customer engagement, experience, and returns.

Protect your data - Personal customer data is private

Personally Identifiable Information (PII) is data used to identify a person, such as a name, date of birth, social security number, driver's license number, credit card information, or other similar information. Organizations that collect PII data must keep the information private and limit access to that data. A large chunk of customer data falls under PII, and your Customer Data Platform must protect such data adequately by restricting access and masking.

Limit access - The fewer, the better

Access to PII should be limited, with full access granted only to those who need it. In addition, all persons with access to PII data should attend compliance training and receive applicable certifications – HIPAA, PCI, etc., depending on the type of business. Finally, an organization should have proper documentation to ramp up new people accessing PII.

Matillion helps organizations get proper access control to PII data. You can set up granular access to data based on virtually anything a user can do with Matillion, including permissions based on projects, data pipelines, or tables. Matillion can also mimic the data permissions of the target, i.e., your cloud data platform.

Hide your PII Data - masking or tokenization

In addition to access control, you must mask or hide PII data within your Customer Data Platform in cases where limited access is needed. The simplest option, applicable when a person needs anonymized customer information – for example, to analyze buying habits of particular customer demographics, is to run a transformation job that strips (PII) data from a table stored in your cloud data platform. You can efficiently hide your PII data with Matillion, as shown below.

When greater access to PII is needed, you can add a quarantine schema to the data in your cloud data platform. A quarantine schema can restrict access or mask the data based on rules in the schema — by data type, data user role, or any other rules you set up — while the PII data is at rest. Limiting access to the quarantine schema is especially important as it is for data at rest.

Another option is to leave the tables containing PII data unchanged but mask the data in transit. Under this option, the original tables with the PII data are not masked. When the data is queried by a user who should not have access to PII data, the columns with PII data are masked in transit so the user is only given access to non-PII data on their end.

Under either circumstance, the data can be partially hidden so that PII is not revealed — but the user can still do their job. For example, all but the last four digits of a customer's social security number or credit card info are on file and masked — enabling a customer service rep to verify identity or payment without having access to PII data. Matillion supports both quarantine schemas and dynamic data masking.

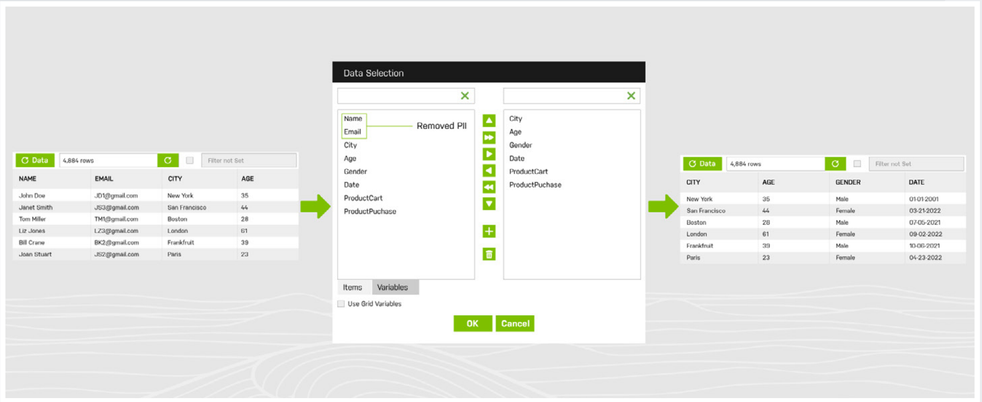


Figure 1 - Removing columns with PII information from a table with Matillion.

Connect your data - Keys can open more than doors

Once you have ensured proper protection of PII data in your Customer Data Platform, the next step is to create a unified customer profile by consolidating customer data from multiple sources into a single master customer record. Now, you can understand your customer base better, identify customer segments, track customer journeys, and improve customer experiences and returns.

Key structuring is an effective method of connecting your data.
For example, keys join data from multiple tables in relational databases like the cloud data platform where your Customer Data Platform data is stored. Combining customer data from different sources creates a single master customer record for each customer.

A key that joins columns from multiple tables can be in the form of a single column that you can create yourself by adding a column to a table and assigning values to that new column for each row of the table – usually sequentially. Alternatively, you can rely on natural keys, a combination of columns of existing data, for example, a customer's name, address, and phone number, that uniquely identify a record or row when combined.

Matillion can help you create and map keys across different tables — making joining data from those tables easy. Then, you can transform that data to create unique customer profiles that can be consumed by the business – e.g., in a customer service portal or your business intelligence and analytics applications. Once you create a unified profile for each customer, use the profiles to improve your understanding of past customer behavior, needs, and preferences — delivering increased customer loyalty and satisfaction.

Customers

Customer ID	Name	Last Name	Email Address
1	Max	Rayfield	maxrayfield@360world.com
2	Ray	Stiffen	ray_stiffen@360world.com
3	Sami	Ruth	sami4ruth@360world.com
4	Steph	Jones	steph.jones@360world.com

Sales

Purchase Number	Date of Purchase	Customer ID	Item Code
1	9/14/2018	1	A_1
2	8/13/2017	2	C_2
3	4/08/2017	3	D_1
4	2/05/2017	1	B_2
5	8/13/2018	4	B_2
6	12/21/2018	2	B_2
7	7/08/2017	4	A_2
8	5/03/2018	3	C_1
9	1/13/2017	1	C_1
10	6/28/2018	2	B_1

Products

Item Code	Item	Price (USD)
A_1	Lamp	$20
A_2	Desk	$250
B_1	Lamp	$30
B_2	Desk	$350
C_1	Chair	$150
D_1	Loudspeakers	$400

Figure 3 - Example of how keys join customer info with purchases of particular products

Cleanse your data – Good data leads to good results

Data cleansing is crucial to maintaining quality data in a customer data platform. We've all heard the saying "garbage in, garbage out." Your Customer Data Platform will have minimal value if the data going into it is not good. Proper cleansing will make your Customer Data Platform accurate, consistent, and complete.

Checking integrity is a crucial step in ensuring that data aligns with its intended purpose. For example, in healthcare, it involves verifying that procedure codes for patient treatments conform to a predefined list of acceptable codes; any records with invalid codes are rejected. Similarly, in retail, validating product codes against the company's inventory ensures accurate purchases.

Filtering mechanisms also play a vital role in eliminating undesired data, such as duplicates or entries with incomplete information. Additionally, data formatting comes into play, allowing for the standardization of dates, numbers, and text, guaranteeing uniformity across data collected from various sources. These practices collectively contribute to maintaining data quality and reliability for informed decision-making.

Matillion automates the process

Matillion makes it easy to cleanse your data and quickly move from bronze (raw data) to silver (cleansed data) to gold (business and analytics-ready data). With Matillion's Dynamic Data Standardization capabilities, you can quickly set up rules to automatically standardize data formats and identify and eliminate any outliers – missing or incomplete data and values significantly different from the majority of values in a dataset. You can also set up business rules to sift out data irrelevant to a particular analysis, and Matillion has Assert components where you can confirm the characteristics or validity of the data before moving the data along the pipeline.

03

Optimizing your customer data for business analytics

The next step is to get your customer data analytics ready so that your business users, who are not experts in the manipulation of data, can get the most value from it. In essence, you have to change the data format in your cloud data platform, making the data easier to understand and interact with.

The Role of Schemas in Modern Data Integration

A database schema defines how data is organized within a database. A transformation schema is used to prepare data for analysis. In contrast, an analytics schema is used to explore the data while analyzing it, making it easier to garner more meaningful information and creating a more efficient reporting and analysis process. Ultimately, getting the data into an analytics-ready state ensures that the data is properly organized and structured to support analytics functions, business intelligence, and analytics tools. You need both types of schemas as part of the data integration process within your modern Customer Data Platform because the transformation schema is where the datasets are built to publish into an analytics schema.

More and more organizations are moving to cloud-native data integration platforms like the Matillion Data Productivity Cloud because legacy solutions with on-premises processing could not handle the scale of transformation processes needed to get the increasing volumes of data into an analytics-ready state. As a result, some BI tools started to add functionality where they would perform the necessary modifications on the data before reporting on it. This not only slowed down the reporting function but also led to potential gaps in business knowledge because the business users did not have the expertise to ensure that the BI tools were accessing and digesting all the necessary data.

Getting your customer data analytics-ready

The key to getting your customer data analytics ready at scale is simplifying the structure of the data. You need to model the data, understand how it is structured and stored, and put together rules to ensure that the incoming data is stored in a manner that complies with that model, making it much easier to comprehend the information in your customer data platform. This, in turn, enables your developers to group the data more effectively into various layers when designing data pipelines within the cloud data platform where your customer data is stored.

Using customer demographics data as an example, let's examine how this works in practice. To create a successful data model, it is important to understand the different types of data, the relationships between various entities, and the attributes of each entity. For customer demographics, the data will typically come from Salesforce, your sales ordering system, and a few additional sources, such as customer survey data. Each source will contain some types of customer demographic data needed for the analytics layer to respond to a business query. Once you understand how your data is modeled, you can then begin to do additional manipulations on the data to facilitate analysis.

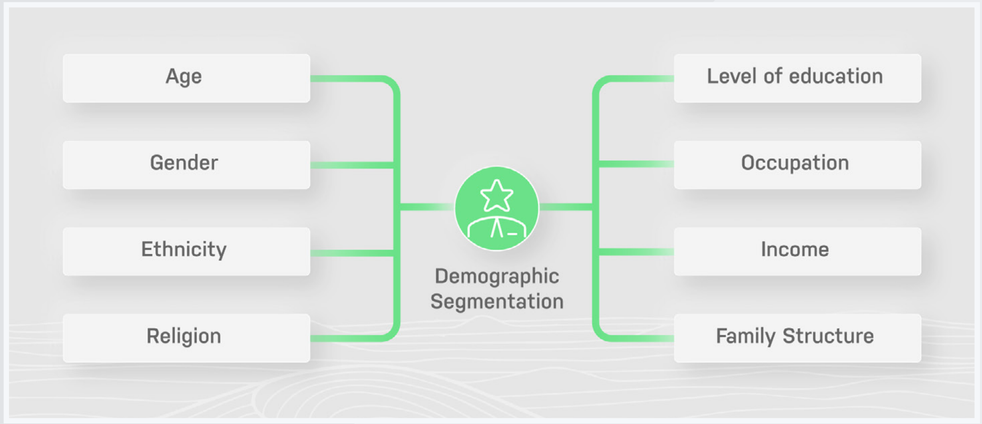


Figure 2. Typical segmentation of customer demographics data.

Filtering - only analyze the data that matters

You can start modeling your customer demographics data with a few simple rules to segment the raw data coming in based on when the data was loaded or updated. You can then use this segmentation to ensure that only the new or updated customer demographics data is fed into your transformation schema, where the data is normalized to your corporate standards. The standardized customer demographics data can then be filtered to ensure that relevant and up-to-date data is sent to your business users and digested by your BI/analytics tools.

Aggregation - group data to simplify the analysis

Aggregating data is the key to a productive data warehouse. This is all about storing data at the correct granular level so it meets our business rules. In practice, this entails grouping data together into conformed data sets to simplify our data structure.

Here, we have a typical enterprise with account information from one source and customer emails in an unstructured format from a different source. We use the Calculate component to figure out the web domain for our accounts. At the same time, we flatten the unstructured email data and use another Calculate component to figure out the customers' email domains. We then join the two datasets to create a new customer accounts data set that includes both account and email information.

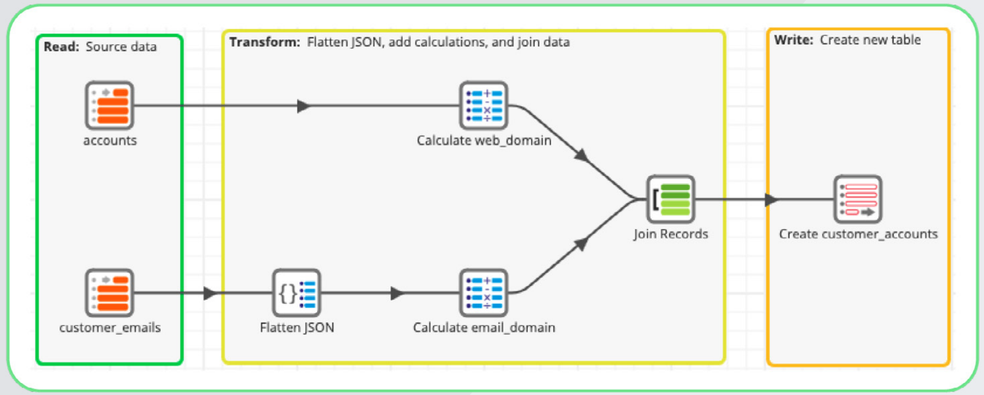


Figure 4. Joining account and email information.

Functions - Do the analysis legwork upfront

A data integration platform like Matillion, with its extensive integration into the database functionality of your cloud data platform, can help you dig deeper into your transformed data. In turn, this helps you gain a greater understanding of your data and develop more refined analytics-ready data sets that can be used both for reporting and for further analysis by your BI/analytics solutions.

Synchronize your data back to business applications

To maximize the utilization of your enriched, analytics-ready customer data and extend its accessibility, as well as to guarantee precise targeting of customers with tailored offers, a proven approach is to integrate this enriched customer data back into the familiar business applications utilized by your sales and marketing teams.

This ability to sync enriched data back to your business applications is called reverse ETL. Simply put, reverse ETL is taking data loaded into a cloud data platform, transforming it to meet certain business standards and analytics requirements, and then pushing it back into the source systems for all users to take advantage of in their operational systems.

Matillion provides a number of options for managing tables via reverse ETL. You can simply create a new table on the fly if there is no corresponding table in the source business applications. You just have to ensure that the new table fits within the table standards of the intended destination – i.e. the original source business application like Salesforce. However, in many instances, you will want to sync your data back to a table that already exists in the destination. In that case, Matillion makes it easy to configure your Matillion table to match before outputting the data back.

Synchronizing and loading data back to the source business applications can be done at any point within a data pipeline and ensures that your business users have access to the most complete and up-to-date information in the applications they are used to accessing on a regular basis.

Customer data metrics that matter

Because it collects all customer information, such as customer identity and transaction history, in one place, a Customer Data Platform helps you measure your success in acquiring and keeping customers. You can analyze a wide variety of metrics, such as customer acquisition costs (CAC), net dollar retention (NDR), the percentage of revenue retained from existing customers, and the overall customer life cycle (CLC).

These insights empower businesses to:

- Ensure customers realize their full value potential
- Refine marketing and advertising campaigns for targeted recommendations
- Enhance product development and offerings
- Increase revenue by fostering customer loyalty and improving service

While it's valuable to understand these metrics, they are ultimately lagging indicators that show what has happened in the past. An excellent way to lose customers is to only look at past behavior and treat them like walking wallets, assuming their behavior or preferences will remain the same.
The greater value of a Customer Data Platform is the ability to collect and analyze information obtained from leading indicators that measure a customer's satisfaction and how much value the customer is getting. These customer-focused metrics analyze behavior and preferences through things like web browsing, social media, demographics (industry vertical, location, size, etc.), and interactions with the business (support, chatbot, community activity, etc.)

With this information, a Customer Data Platform now provides a more objective way to measure the value obtained by each customer versus spend. This allows you to become more proactive when dealing with customers and react better to changing customer behaviors.

- Make sure customers are gaining all the value they can
- Fine-tune marketing and advertising campaigns and create more targeted recommendations
- Improve product development and offering strategies

Ultimately, a Customer Data Platform can help businesses achieve their primary goal of increasing revenues through increased customer loyalty and improved customer service.

04

Next Steps

Implementing a successful Customer Data Platform strategy is a complex yet essential endeavor for modern businesses aiming to thrive in a data-driven era. In this ever-evolving landscape, where customer data serves as the lifeblood of informed decision-making, the critical first steps toward Customer Data Platform success are paramount.

1

Craft your Customer Data Platform implementation strategy:
Evaluate your existing data infrastructure, including data sources, storage solutions, and data management processes. Decide between a centralized or federated data management model based on your organization's needs, resources, and data structure. Research and evaluate Customer Data Platform solutions available in the market to find the one that best fits your business needs and budget. Implement data connectors or integration tools, such as Matillion, to extract and load data into your chosen cloud data platform.

2

Enhance customer data transformation and cleansing:
Ensure PII data privacy and security by restricting access and implementing masking or tokenization where necessary. Create a unified customer profile by consolidating data from different sources. Cleanse the data to remove duplicates and incomplete records and ensure data accuracy.

3

Optimize data for analytics & track key customer metrics:
Prepare your customer data to be analytics-ready by simplifying its structure and ensuring it complies with analytics schemas. Continuously monitor key customer data metrics, including customer acquisition costs, net dollar retention, and customer satisfaction indicators. Use these metrics to make informed decisions, refine marketing strategies, and improve product development to leverage insights from your Customer Data Platform to enhance customer engagement, tailor marketing campaigns, and drive customer loyalty.

4

Prioritize data governance and compliance:
Establish robust data governance practices to maintain data quality, security, and compliance with relevant regulations (e.g., GDPR, HIPAA). Regularly audit your data management processes to ensure adherence to data privacy and security standards.

5

Stay informed and adapt:
Stay informed about industry trends and emerging technologies in customer data management. Be prepared to adapt your Customer Data Platform strategy and implementation to stay competitive and meet evolving customer expectations.

Match the Speed and Scale of Your Data Demands

Matillion is the productivity platform for data teams, making data work more productive by empowering the entire data team - coders and non-coders alike - to move, transform, and orchestrate data pipelines, faster. Matillion simplifies and automates data movement, bridges the skills gap for data transformation, and handles the scale and complexity of pipeline orchestration with ease. Underpinned by a unified platform that enables unlimited scale, users and projects; with setup in minutes and transparent pricing for only what you use.

Sign up for a free trial at matillion.com today.

Trusted by the world's best data teams

TUI AUTODESK DocuSign cisco Western Union slack