

# Introduction to Customer Data Platforms

**Raab**  
associates, inc.  
730 Yale Avenue  
Swarthmore, PA 19081  
[www.raabassociatesinc.com](http://www.raabassociatesinc.com)  
[info@raabassociatesinc.com](mailto:info@raabassociatesinc.com)

 **TREASURE  
DATA**  
2565 Leghorn St.  
Mountain View, CA 94043  
<http://www.treasuredata.com/>  
[info@treasuredata.com](mailto:info@treasuredata.com)

# Introduction to Customer Data Platforms

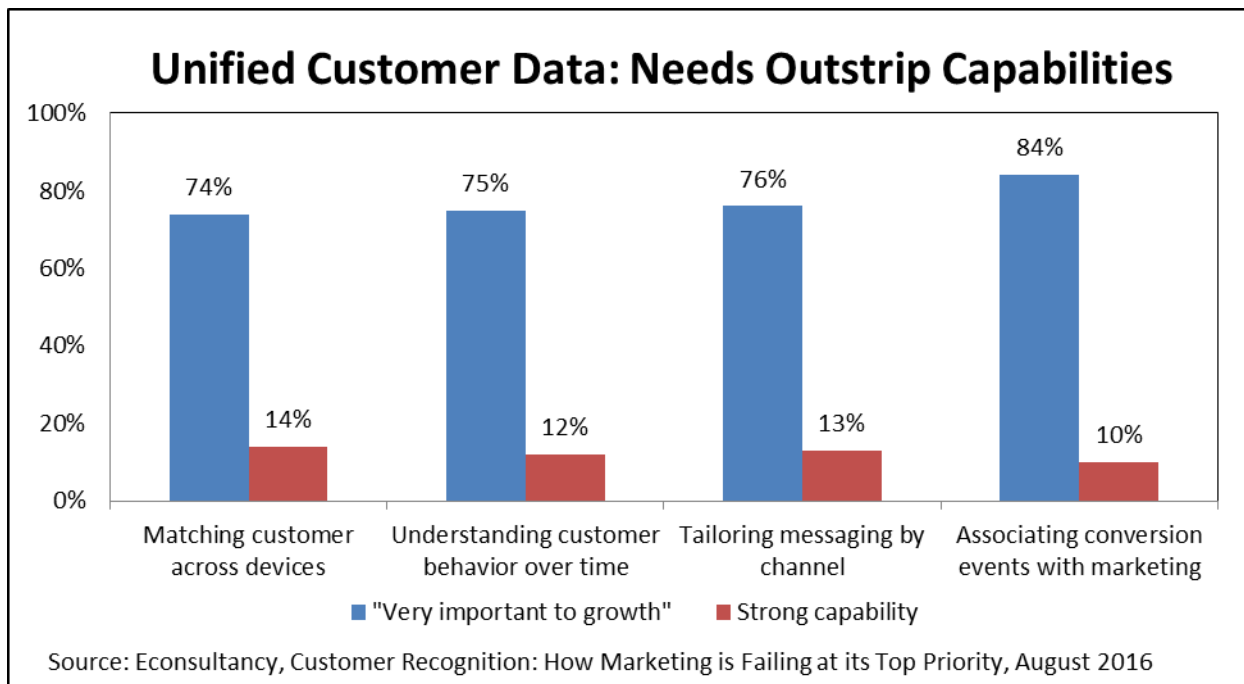
---

## Overview

Many marketers are struggling to assemble the unified customer data they need for success. The Customer Data Platform is a new class of system that makes solving the problem considerably easier than older alternatives. This paper describes the requirements needed to provide unified customer data, shortcomings of conventional approaches, and characteristics that make Customer Data Platforms a superior solution.

## Introduction

Here's the problem in a nutshell: marketers know they need unified customer data but haven't been able to get it. A recent survey illustrates this clearly: while three-quarters or more marketers considered each item very important to growth, just 10 to 14 percent had a strong capability.



Let's assume that you, too, are already convinced that unified customer data is important but don't have it available in your organization. We can then move on to how you can solve the problem.

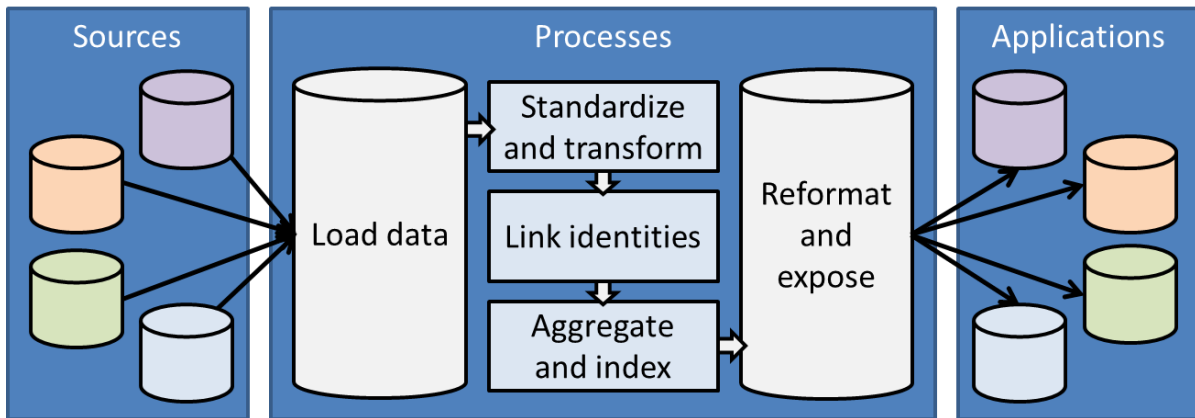
## What Is Unified Customer Data, Anyway?

Let's start by defining what a solution looks like. Unified customer data, also known as a "360 degree customer view", refers to assembling all of a company's information about each customer in one place and making it available for marketing and other uses. This means a solution must meet several criteria:

# Introduction to Customer Data Platforms

- **Gather data from all systems with customer information.** A lot of systems have customer information, including your Web site, email, sales, customer service, order processing, and probably others. But it's still not all your systems: for example, there's probably no customer data in your human resources, manufacturing, purchasing, or logistics systems. So the scope of a unified customer data project is broad but not universal.

## Process Flow for Unified Customer Data



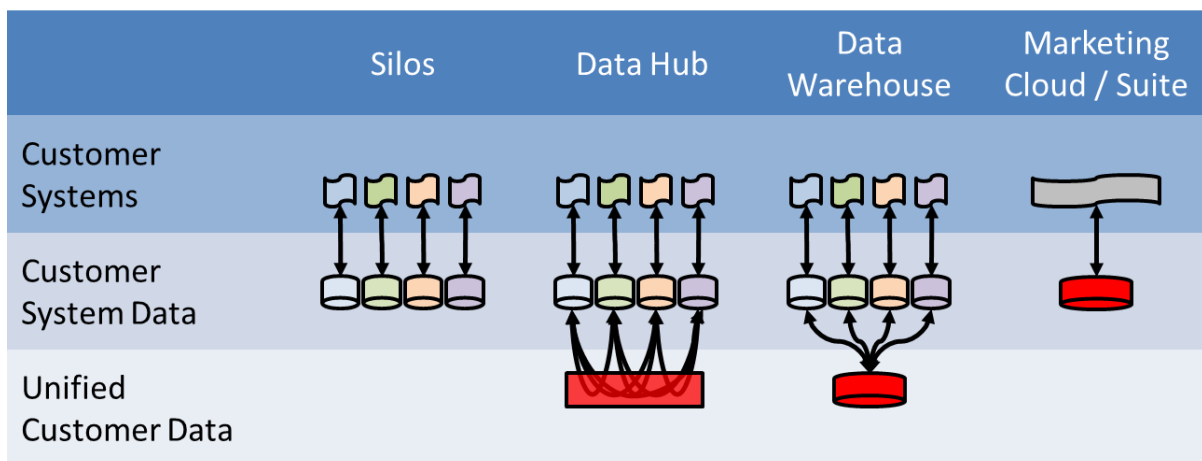
- **Gather all customer-related data from those systems.** Even your customer-facing systems will contain large amounts of data that are not tied to individual customers. For example, details of your Web site operations are usually not relevant (unless slow response is impacting the customer experience). Similarly, it doesn't matter how your sales teams are organized or how your warehouse fulfills customer orders (again, unless those affect how customers are treated). This further reduces how much data must be assembled.
- **Handle many types of data.** Even after excluding the irrelevant systems and data elements, you'll have plenty of data to wrangle. Customer data is no longer just structured information like purchase transactions with each element (customer name, address, product, date, price, etc.) neatly organized and documented. A complete customer view today includes email and chat transcripts, audio and video recordings of service interactions, logs of Web site visits and mobile app sessions, text of social media comments, and much more. Many of these require special processing to extract concise, usable information from massive streams of semi-structured or wholly unstructured inputs.
- **Identify customers across systems.** Different systems often use different identifiers for the same person. Common identifiers include customer or account ID, name and address, telephone number, email address, and device ID. The unified customer view requires linking identifiers that relate to the same individual, business or household. Sometimes the linkage is based on clear relationships: for example, if a customer logs into their account using a specific device, that device can be confidently associated with the customer on subsequent interactions, even if the customer does not log in again. This is called "deterministic" matching. Other linkages are based on "fuzzy" or "probabilistic" matching, such as similarities between two different name and address records ("Bob Jones, 123 Main St." vs. "Robert Jones, 123 Main Street") or two devices frequently being used at the same locations.

# Introduction to Customer Data Platforms

- **Expose data to other systems.** Database experts sometimes joke about “write only” databases – inherently useless systems that import data but never let anyone see it again. To avoid this fate, the unified customer view must organize its inputs in ways that let other systems access them efficiently and effectively. The specific requirements depend on the applications but they usually include extracting complete information about a single customer, often in real time; finding groups of customers that share specified characteristics; identifying customers whose data has changed; and extracting data about selected customer segments. Requirements also include methods to access the data: typical options include direct connections through an Application Program Interface (API); standard database queries such as SQL statements; and bulk files with large volumes of data.

## What Are The Options?

### Architectures for Unified Customer Data



Marketers have several options for creating a unified customer view. The most common include:

- **Enterprise data warehouse.** This can be broadly defined as a central repository for data from corporate systems, used primarily for analysis. Data warehouses have long been IT departments’ standard approach to assembling data from multiple systems. They are designed to serve the entire business, not just to give marketing a unified customer view. This larger goal results in a bigger, more complicated project, made still more challenging by the ever-growing complexity and frequent changes of today’s marketing systems. In fact, data warehouses are notoriously time-consuming and expensive, nearly always involving a multi-year effort and frequently failing to reach completion.

The analytical focus of standard data warehouse design also means that most warehouses are updated periodically – usually weekly or monthly – to let time-consuming processes such as aggregation and indexing run efficiently. Working with not-quite-current information is adequate for most business analyses but not for guiding real-time customer interactions. For those, the latest information is needed to treat each customer correctly. Nor are most data warehouses designed to quickly retrieve information about individual customers: again, this is not needed for most business analyses. For these and similar reasons, most marketers find a data warehouse project does not meet their needs for a unified customer view.

# Introduction to Customer Data Platforms

---

- **Data exchange hub.** This shuttles data between systems without storing it centrally. Hub systems are used primarily for processes that run across multiple systems, such as having an email system send a follow-up message after the order processing system records a transaction, or having a marketing automation system send a list of prospects to an advertising server. While hubs can be effective for such purposes, they have no native features to match different customer identifiers, so they cannot themselves assemble a unified customer view. Lacking a central data store, they require moving any shared data into the system that uses it, which may result in considerable data duplication – assuming the receiving system is even able to store the external information, which is not always the case. Lack of a central data store also means hubs can't keep historical information which is often lost or inaccessible in the source systems. A hub solution can overcome some of these problems by sending data to a permanent central database. But that database would still need features such as identity matching and access mechanisms to meet the full requirements of a usable unified customer view.
- **Marketing clouds and suites.** These offer all major customer-facing systems as part of one solution sharing a single customer database. This inherently provides a unified customer view because there are no separate sources of customer data to combine. Unfortunately, the major marketing suites are not as unified as they seem: all were built through acquisition of separate systems and have usually retained their original, separate databases. Details vary by vendor but there is often little more than a set of linked identifiers to unify the customer data; details are stored in the original systems and must be extracted to assemble a complete customer view. This has a number of drawbacks including how much data can be assembled, how current it is, and how groups of customers can be analyzed. Sometimes even the linked identifiers are not available. Moreover, most marketers use products from multiple vendors, not all products in one marketing cloud or suite. Integrating other systems with the core marketing suite is always some work, and in cases can be difficult or impossible.

## A Better Alternative

Marketers have recently been offered a new approach to building the unified customer view: the Customer Data Platform (CDP). The CDP is defined as *“a marketer-controlled system that builds a unified, persistent customer database which is accessible to external systems.”* Each part of this definition distinguishes the CDP from other alternatives:

- *“marketer-controlled system”* differentiates the CDP from enterprise data warehouses, which are built and managed by corporate IT departments. Marketer control means the marketing department manages the CDP project, using its own staff and external resources to do the work. As a result, the CDP is focused on marketing needs and can more rapidly be adjusted as those needs change. It also implies that less technical skill is needed to build a CDP. This is because the narrower scope of CDPs makes it easier for vendors to predict, and therefore build in, features needed to operate the system. It also means the system must meet a smaller set of requirements, allowing a simpler design. All these factors contribute to a faster, cheaper, lower-risk deployment than a traditional data warehouse.
- *“builds a unified, persistent customer database”* highlights the specific requirements the CDP must meet. It is “unified” in the sense of loading data from multiple systems and linking it with customer

# Introduction to Customer Data Platforms

---

identifiers to create unified customer view. It is “persistent” in storing the customer data permanently, rather than continuously reading it from the original source system or replicating it into multiple customer-facing systems. This improves performance, ensures consistency, avoids redundant data storage, and permits historical analyses that would be impossible if only current data were available. It is a “customer database” in focusing on customer information, thereby limiting the scope compared with a broader enterprise data warehouse.

- “*accessible to external systems*” clarifies that CDP data is intended to support operations of other systems. This distinguishes the CDP from marketing clouds or suites that are closed to external products and from systems that create the database to help deliver applications such as predictive modeling or message personalization. Some CDPs also include such applications – in fact, most provide some practical function, which gives marketers an immediate benefit from buying them. So long as the CDP makes its data easily available to other systems, including integrated applications does not violate the definition.

Many CDP systems share other characteristics. These are not specifically required by the formal definition, but make it easier to fulfill core user needs. They include:

- **Cloud deployment.** The system is run on the vendor’s own servers, rather than installed on the client’s in-house computers. This removes much of the technical work required to manage the system, making it much easier for marketers to operate without support from corporate IT. Cloud deployments also often cost less than on-premise systems, helping to make the systems more affordable than conventional data warehouses.
- **Software as a Service (SaaS) pricing.** This means that buyers rent use of the software rather than purchasing it outright. SaaS systems are almost always cloud-based, although cloud deployments are not necessarily SaaS. Like other cloud deployments, SaaS systems reduce the technical workload of the buyers. SaaS pricing also makes systems easier for marketers to purchase because the monthly fee can often be charged as an operating expense rather than a software investment that will require for approval by corporate finance and IT departments.
- **NoSQL data stores.** Many CDPs use “NoSQL” data stores such as Hadoop, Cassandra, or MongoDB instead of conventional relational databases like Oracle or SQL Server and newer SQL-based systems like Amazon RedShift and Google BigQuery. SQL databases require precise advance definitions of the data they will receive and the types of queries they will return. They must be designed properly for adequate performance, especially when large volumes are involved. NoSQL stores are vastly more flexible, making it much easier to incorporate new data sources in general and unstructured or semi-structured data in particular. This flexibility further reduces the time and technical work needed to construct customer database and to adapt it as new needs arise.
- **Real time access.** Many applications of the unified customer view require real time response. Common examples include Web site personalization and interactive product recommendations. Many CDPs apply special techniques, such as indexes on top of NoSQL data stores, to make this possible. Many also provide APIs for real time access. APIs are an especially flexible method for reading data, making it easy for external systems to issue precise information requests and immediately apply the results to an on-going process.

# Introduction to Customer Data Platforms

---

- **Supplemental applications.** As previously noted, many CDPs supplement their core database with functions that use the data for tasks such as segmentation, predictive modeling, data exploration, customer success management, or even campaign management. Even CDPs that were not originally designed to deliver such applications have tended to add them over time in response to client requests. This clearly adds value for many users, without requiring anyone to use the features if they prefer an alternative solution.

## Reality Check: CDPs Don't Solve Everything

Customer Data Platforms have many advantages over other approaches to creating a unified customer view. But they are not magic. Here are some realities to keep in mind:

- **All CDPS are not the same.** Many systems meet the conditions included in the CDP definition. Quite a few share the additional features as well. But there are still many differences between CDPs. You'll need to look for fit with you mix of inputs, scale of data, rate of change in your marketing systems, preferred deployment model, budget, technical resources, and other factors. As with any other marketing system selection, start with a careful assessment of your current and expected requirements and then factor in a considerable margin for unanticipated future needs.
- **Some technical skills are still needed.** CDPs are designed to deploy easily and are vastly simpler to install than conventional alternatives. But it still takes technical expertise to understand your existing systems, determine which data to send the CDP, select and tune identity matching methods, monitor data quality, and share data with external systems. If you don't have the right technical resources, most CDP vendors can provide the necessary support through their own services teams or partners. Even then, you'll probably still need some cooperation from your corporate IT team to gain access to data sources and to let other systems use the CDP data.
- **Your existing infrastructure could be a problem.** Not all systems are designed to integrate easily with others. Both extracting data to feed the CDP and reading CDP data for customer management and analytics may be difficult or impossible for some of your existing systems. In addition, the quality of source data is often an issue, especially if no one has tried to use that data for customer management before. You'll want to assess these issues before making a CDP investment, both because some CDPs may deal with them better than others, and because – in the worst case – they may make it impossible to use any CDP effectively.

## Summary

Marketers' need for easily accessible, unified customer data is well known. So are the drawbacks of conventional solutions. Customer Data Platforms offer a less-familiar alternative that gives marketers a superior solution at a lower cost. CDPs are not perfect and effort is still required for success. But a carefully chosen CDP can be a solid foundation for an effective marketing technology strategy and long-term marketing success.

# Introduction to Customer Data Platforms

---

## About Treasure Data

Treasure Data is the Live Data Management Platform that makes all data connected, current and easily accessible to the people and algorithms that drive business success. The platform provides turnkey connectors for 300+ data sources, flexible APIs to unify and analyze raw event-by-event data, and the mechanism to define workflows and drive business processes. Common use cases include data silo unification, omnichannel analytics, paid and content marketing optimization and cross-device attribution analysis.

Founded in 2011, Treasure Data has raised more than fifty million dollars in funding and is headquartered in Silicon Valley with offices in Tokyo and Seoul. Its global customer base includes data driven Fortune 500 enterprises like Subaru and GE Healthcare as well as fast-growing startups like Wish, Quip and Pebble.

**Contact:**

Treasure Data  
2565 Leghorn St.  
Mountain View, CA 94043  
<http://www.treasuredata.com/>  
[info@treasuredata.com](mailto:info@treasuredata.com)

## About Raab Associates Inc.

Raab Associates Inc. is a consultancy specializing in marketing technology and analytics. Typical engagements include business needs assessment, technology audits, vendor selection, results analysis, and dashboard development. The company also consults with industry vendors on products and marketing strategy. It publishes the B2B Marketing Automation Vendor Selection Tool (VEST), the industry's most comprehensive independent guide to B2B marketing automation systems.

**Contact:**

Raab Associates Inc.  
730 Yale Avenue  
Swarthmore, PA 19081  
[www.raabassociatesinc.com](http://www.raabassociatesinc.com)  
[info@raabassociatesinc.com](mailto:info@raabassociatesinc.com)