Manuals+ — User Manuals Simplified.

blueconic G2 Treasure Data Customer Data Platform User Guide

blueconic G2 Treasure Data Customer Data Platform User Guide

Unifying data to create a 'single customer view' is the foundational, and arguable the most critical, capability of a customer data platform (CDP) like BlueConic.

With access to persistent, individual-level profiles that update in real time, business technology users in marketing, ecommerce, analytics, and other teams can activate the data when and where they need to improve how they engage with customers, conduct modeling and analytics, build segments, and more.

Creating this comprehensive single customer view hinges on great identity resolution. If you can't recognize and reconcile data for the same individual across multiple devices, channels, systems, and platforms, then you can't deliver tailored messages and experiences that meet their unique needs and preferences.

However, the growing complexities of identifier fragmentation, combined with browser changes that restrict the use of third-party cookies, present significant barriers to making the construct of identity the core of any customer engagement model. It also makes compliance with GDPR, CCPA, and other consumer privacy laws exponentially more difficult.

At BlueConic, we have helped hundreds of companies collect, consolidate, and normalize customer data from online and offline sources into unified, persistent profiles for a comprehensive and actionable single customer view. This document explains BlueConic's unique approach to identity resolution and why it's faster and more accurate than other approaches.

Contents

Creating Your Complete and Dynamic Customer View

To engage with individuals effectively, companies need to understand all the dimensions of identity that relate to an individual. The problem is these bits of information live in siloed systems that each have a unique way of storing data and recognizing customers.

To create a comprehensive, single customer view, BlueConic aggregates these fragmented identifiers —along with associated identity-related attributes (e.g., browsing behavior, consent status, transaction history, geolocation) — and stores that data in persistent, individual-level profiles that update in real time.

Unlike those of other CDPs, BlueConic profiles are unique in three key ways:

1. Profiles can be created for both known and anonymous individuals, then merged based on any unique identifiers or any combination of them.
2. Profiles exist across any channel and persist as long as you require.
3. Profiles are designed to pull from any other tech in your stack (e.g., CRM, email system) and share customer data back to your technology stack.

Unified profiles are where the value in BlueConic starts. With access to unified, actionable data, growth teams can decide in what ways they want to activate the data to support their use cases and drive business growth.

BlueConic's Profile Merging Process

Profile merging is the underlying mechanism BlueConic uses to ensure you can recognize the same person across multiple channels and devices and engage them in a consistent way. To resolve identities, BlueConic leverages two profile merging methods:

1. Deterministic matching: This is where identities are resolved based on one or more unique identifiers you know to be true, such as a login name, email address, customer ID, or phone number. Deterministic matches are more precise but require a higher threshold for data.
2. Probabilistic matching: This is where identities are resolved based on the probability that two customer records represent the same individual. Also known as "fuzzy" matching, probabilistic matching can be used on top of our deterministic matching capabilities for scenarios where the value of a profile property doesn't match exactly (e.g., due to a typing mistake) Each approach has its purpose but should be considered in the context of how you will use the data. Let's take a closer look at both matching techniques and their intended use cases.

## BlueConic's Deterministic Matching Capabilities

Most BlueConic customers use deterministic profile merging methods, meaning there must be at least one verifiable data point of information shared by two profiles to merge them. Examples of common deterministic merge rules include:

- Email address is the same
- Customer ID is the same
- Phone number is the same AND name is the same AND address is the same
- Date of birth is the same AND name is the same AND address is the same

While any identifier or combination of identifiers can be used for profile merging, for best results, we recommend targeting profile properties that are most guaranteed to be unique, such as an email address and customer ID.

For instance, while two profiles containing the same first and last name likely belong to the same person, it is possible that you are dealing with two different individuals who have the same first and last name.

Therefore, it is important the identifiers you use are those most likely to uniquely identify visitor profiles belonging to the same person.
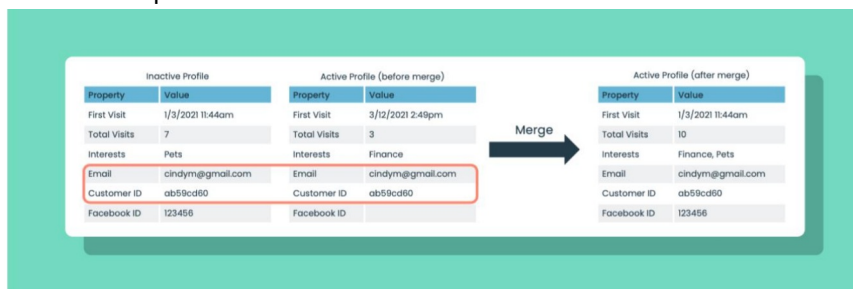

Profile Merging in BlueConic

How It Works

BlueConic enables you to easily define and create your own profile properties using the UI and mark these properties as a "unique identifier" so they can be used to determine when two or more profiles are merged.

In addition to identifying profile properties that serve as unique identifiers, you can also customize the data points that are unified and stored for each profile within your BlueConic database. There is no limit to the number of profile properties you can add, which might include total number of visits and page views, IP address, behavioral scores, time zone, and more based on your business needs.

The logic behind these custom "merge rules" can easily be configured in the BlueConic user interface as well. This sets BlueConic apart from many other customer data platforms that limit you to a developer-only or black-box-only approach when deploying a merge strategy.

You can also establish custom merge rules that specify what to do with the value of a single profile property when multiple profiles merge. For example, when merging a list of customer interests, you'll want to add those lists together, while you sum values together for purchases, and store only one original customer acquisition date. This customization increases the utility of the unified profile and focuses your view of the customer to include the most important data for your business. Moreover, merges occur as soon as the criteria for a match are met so the system always has the latest insight about a person as they interact with your brand.
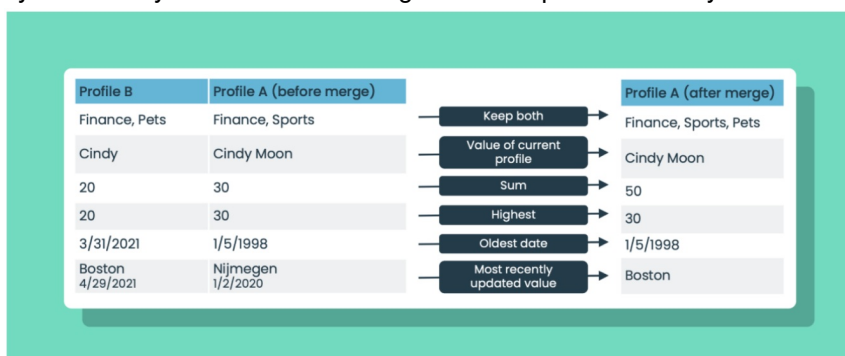
Merge Strategy Examples

When two profiles merge, both profiles are kept (normally, the newest profile becomes the active profile; the older profile becomes inactive). The values of the profile properties shared by both profiles are merged one by one, according to the merge strategy assigned to that property, and any profile properties and their values that did not previously exist are added to the new profile. Watch our video to learn more about profile merging in BlueConic.

Data Normalization and Cleansing

BlueConic also leverages data normalization techniques (e.g., for email addresses, phone numbers, names, street addresses, etc.) to help improve the quality and match rate of our deterministic matching capabilities.

These data processors are pluggable components that apply common data hygiene and transformational processes to ensure profile data is properly validated, corrected, normalized, and cleansed.

For instance, the email-cleaning data processor can be configured to correct many common typing mistakes in email addresses, while the name-normalization data processor can be configured to create a normalized full name.

These data processes can be configured to both pre-process data before it is imported into BlueConic and post-process data that is already stored in BlueConic.

Learn more about BlueConic data processors here.

BlueConic's Probabilistic (Fuzzy) Matching Capabilities

BlueConic also offers probabilistic matching capabilities that use specialized Python code in AI Workbench to determine when slightly different profiles in fact belong to the same user. A simple, but common example of a probabilistic match is where two profiles have identical first and last names, but their phone numbers differ by one digit, possibly because of a typo. This functionality can be useful for customers where a lot of data entry is done manually (e.g., by store clerks or callcenter operators), since this tends to lead to more data-entry errors compared to customers entering their information themselves.

How It Works

Our probabilistic models measure the similarity between values based on the Damerau Levenshtein measure of edit distance, which is the number of operations required to change one word or number into the other. This measure increases by 1 to account for deletions, insertions, substitutions of a single character, or transpositions of two adjacent characters.

$$d_{a,b}(i,j) = \min \begin{cases} 0 & \text{if } i = j = 0 \\ d_{a,b}(i-1,j) + 1 & \text{if } i > 0 \\ d_{a,b}(i,j-1) + 1 & \text{if } j > 0 \\ d_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} & \text{if } i,j > 0 \\ d_{a,b}(i-2,j-2) + 1 & \text{if } i,j > 1 \text{ and } a[i] = b[j-1] \text{ and } a[i-1] = b[j] \end{cases}$$

For example, "Jennifer" and "Jenifer" have an edit distance of 1 (deletion of an "n"), as does "Michael" and "Micheal" (swapping of the adjacent characters "ae" for "ea"). When two values are exactly the same, the edit distance is 0. (Note that capitalizations, hyphens, and special characters do not contribute to the edit distance.) Using the symspell algorithm, BlueConic's probabilistic model can rapidly find values from the dataset that are within the allowed edit distance from another value. BlueConic customers can also easily change the functionality of the model or add additional functionality on top of our probabilistic matching logic via the Notebook editor in the BlueConic UI. For example, you can:
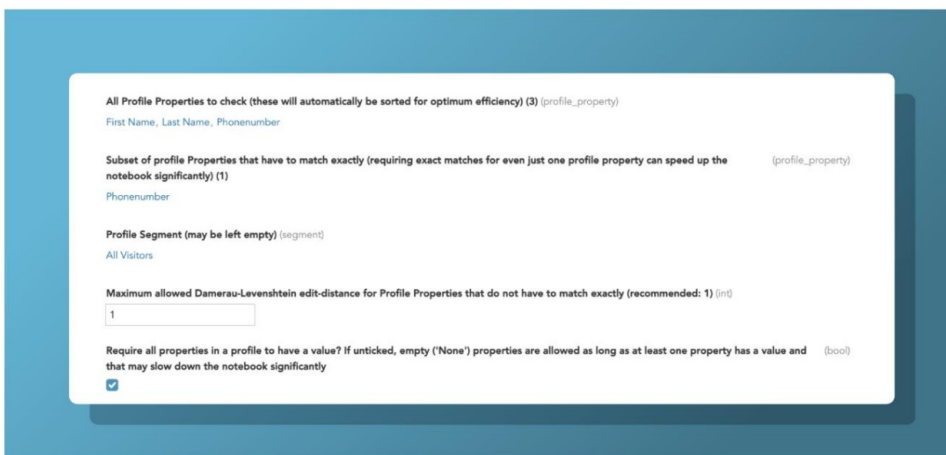
- Generate a score based on the number of profile property values that overlap between two profiles, only merge the profiles if the score exceeds a certain threshold
- Add additional normalization rules for names or addresses
- Use census data to check how common a given name is and take this into account as well before merging two profiles
- Use census data to check how many people live at a certain ZIP code and take this into account before merging two profiles

Weeding Out False Matches

To increase accuracy and reduce the risk of false matches when employing probabilistic matching, we help customers specify the right matching parameters based on their goals and constraints.

These parameters can be set up in the Python code or via the BlueConic UI. We typically recommend customers:

1. **Limit the edit distance to 1:** Common profile property values are often only a few edits away from each other, so the number of false matches (as well as the notebook's runtime) increase exponentially with higher allowed edit distances.
2. **Ensure at least one property matches directly:** Configuring one or more profile properties to match exactly can further increase accuracy. For example:
   1. **First name is within 1 edit distance;** last name and phone number match exactly
   2. **Last name is within 1 edit distance;** first name and phone number match exactly
3. **Include at least one highly diverse property:** We suggest including at least one highly diverse property that has many different values in your dataset (i.e., phone number, email address). For example, let's say there are 40,000 John Smiths in the U.S., and among them, there are hundreds of John Smiths who are 48 years old. However, when we find two John Smiths, both aged 48, with a phone number that differs by one digit, we are likely looking at the same person who mistyped his phone number once, and we have likely found an accurate, probabilistic match.


Notebook Parameters

BlueConic also employs its own set of probabilistic match rules to remove duplicate matches and reduce the risk of false matches. These rules include:

- **Profiles with matches are ranked by the number of matches they have. The profile with the greatest number of matches comes first.** When A matches B and C, while B and C both only match A, then B and C disappear, and A is enriched with their data. This is because A has two matches, while B and C only have one match each.
- **Conflicting matches are removed**. Building on the previous example, if D matches B and C (but not A), it would conflict with the merger of B and C into A. This is prevented by removing B and C from D's list of matches (assuming A ended up higher in the ranking from the first rule).
- **The previous rule ensures duplicate matches that remain after the application of the first rule are always removed.** If E matches F, then F must match E. One of these matches is always removed.

- **These rules always ensure exact matches are merged, as well as many probabilistic matches** . Only probabilistic matches that threaten to create a chain and have few matches themselves (D, in the example above) are not merged.

Probabilistic models can be run manually or via scheduling, and the output can be tied directly to BlueConic's profile merge functionality using your custom merge rules and associated sub-rules previously set up via the UI.

Manual runs allow for more detailed monitoring of the model's execution and timings. With scheduling, you can set a repeating schedule for the notebook to run.
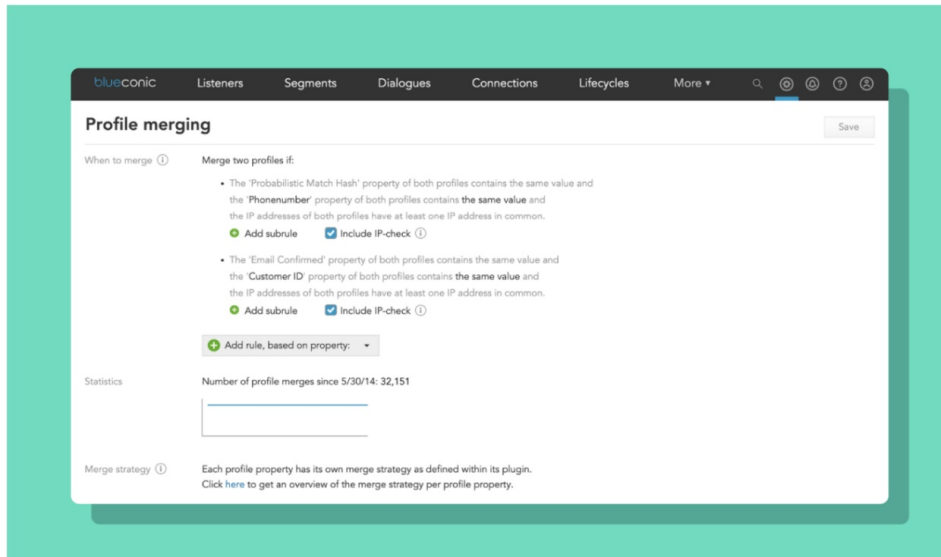
Merge Rules in Combination with Probabilistic Matching Output

For both manual and scheduled runs, the number of profiles, exact matches, and deterministic matches are displayed in the output and log. Matches for profiles that have not been merged are also displayed, along with an associated match score that indicates the likelihood that the records belong to the same individual.
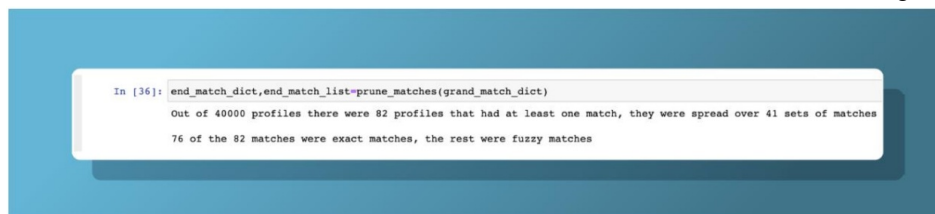
Probabilistic Model Log

Learn more about BlueConic's probabilistic matching capabilities here.

Conclusion

BlueConic's robust and sophisticated approach to identity resolution ensures our customers have access to persistent, unified customer profiles that are comprehensive, accurate, and always up-todate. Unlike other customer data platforms, our identity resolution capabilities are unmatched in terms of our:

- **Flexibility:** With BlueConic, you don't have to choose between deterministic or probabilistic matching. Start with a deterministic approach to identify certain matches, and then augment with probabilistic techniques for those instances where profile property values slightly differ.
- **Configurability:** BlueConic's extensive configuration options give you fine-tuned control over how your identities are managed. Create your own unique identifiers, merge rules, sub-rules, parameters, and even notebook code based on your unique business and data needs — all from our easy-to-use UI.

- **Transparency:** Unlike hard-coded, "black-box" approaches that can only process your data in a certain way, BlueConic's "white-box" approach ensures you have full transparency into how matches are made, and can easily make adjustments when needed to resolve issues or adapt to the needs of the business.

For more information about BlueConic and our identity resolution capabilities, please contact us at **info@blueconic.com.**

Thank you

BlueConic Inc.

+1 (888) 440-2583

info@blueconic.com

www.blueconic.com

Documents / Resources

blueconic G2 Treasure Data Customer Data Platform [pdf] User Guide
G2 Treasure Data Customer Data Platform, G2, Treasure Data Customer Data Platform, Data Customer Data Platform, Customer Data Platform, Data Platform, Platform

References

- **Home | BlueConic**
- **Home | BlueConic**
- **Running probabilistic matching models with AI Workbench – BlueConic Help Center**
- **Overview: Profile merging in BlueConic – BlueConic Help Center**
- **Using data processors with BlueConic connections – BlueConic Help Center**
- **Home | BlueConic**

- [**User Manual**](#)