launchlabs

Optimizing Accuracy and Reach in *Identity Graphs*

with Ground-Truth Validation from Customer Data Platforms (CDPs)

Table of *Contents*

ABOUT THE AUTHOR

Garrett Roach is the Founder and CEO of Launch Labs, where he leads the charge in helping businesses unlock the full value of their customer data.

In the white paper, Garrett shares a behind-the-scenes look at Launch Labs' identity resolution framework and how it continues to evolve to meet the needs of modern marketers.

Garrett Roach

CEO Launch Labs

Executive Summary

Identity resolution is a critical component of modern marketing and analytics systems, yet most resolution models operate without a clear understanding of their accuracy. These systems typically rely on probabilistic methods to associate anonymous web visitors with known customer identities, often with little to no visibility into whether the resolutions are correct.

At Launch Labs, we've developed a framework to address this gap by using Customer Data Platforms (CDPs) as a source of truth. CDPs uniquely contain both deterministic (known) and anonymous (unknown) visitor data, creating a powerful testing ground for evaluating identity resolution models.

By deploying our resolution system against known visitors, we can directly measure accuracy. Then, by varying key settings — such as data freshness limits and match requirements — we can test how each change impacts both the number of resolutions and their correctness.

To solve the inherent tradeoff between resolution count and accuracy, we built an optimization process that identifies the best possible combinations of settings using Pareto frontier analysis. This gives marketers and data teams the ability to make data-driven decisions about how to balance scale with quality.

This white paper outlines our methodology, the unique role of CDPs in resolution optimization, and a framework that any identity system can adopt to continuously improve both precision and performance.

The Challenge of *Identity Resolution*

THE CHALLENGE OF IDENTITY RESOLUTION

Identity resolution systems are designed to connect fragmented signals — devices, IP addresses, browser behaviors, location data — into a single, coherent understanding of "who" a visitor is. These systems underpin critical functions in modern marketing: personalization, targeting, measurement, and CRM enrichment.

But despite their importance, most identity resolution systems operate in a fundamentally probabilistic space. They make educated guesses about users based on patterns, often with limited input data and no ground truth to verify the match.

This leads to two major challenges:

// 01

Unknown Accuracy

Since anonymous visitors aren't tied to known identities, there's no way to tell whether a resolution was correct. Businesses are forced to trust that their resolution vendor's logic works — without visibility into performance metrics like precision or false positive rates.

// 02

Tradeoffs Between Scale and Precision

Identity systems can be tuned to be more aggressive (resolving more users with looser match criteria) or more conservative (prioritizing precision with stricter rules). But without measurement, it's impossible to know whether those settings are yielding the right tradeoff for the business.

SALES ATTRIBUTION USE CASES

For retrospective use cases like sales attribution, it's often acceptable — even desirable — to prioritize scale over precision. Because the system is ultimately matched against known sales records, some false positives in identity resolution can be tolerated. The key is ensuring that attribution signals are large and broad enough to be useful for reporting and optimization.

CUSTOMER ENGAGEMENT USE CASES

For outbound efforts like retargeting, email follow-up, or SMS messaging, false positives are much riskier. Here, precision must meet a minimum threshold to protect brand experience and avoid misdirected outreach. But scale still matters — overly strict resolution logic can cut engagement volume significantly, reducing campaign impact.
These competing pressures create a need for flexible, measurable identity systems — ones that can be tuned based on context, and evaluated based on their true performance. Without this, businesses are flying blind, risking wasted spend, missed opportunities, or customer experience failures.

THE CDP ADVANTAGE

Ground Truth for *Identity Models*

Most identity resolution systems struggle with evaluation because they operate in environments where the visitor's true identity is unknown. Without a benchmark, there's no clear way to measure accuracy, test different configurations, or understand which signals are driving performance. Customer Data Platforms (CDPs) change that.

CDPs integrate directly with known customer sources — such as CRM systems, transaction logs, and loyalty databases — and track real customer behavior across websites, apps, and messaging channels. This creates a unique advantage: they contain both deterministically identified visitors and anonymously tracked visitors within the same dataset.

Launch Labs leverages this distinction to evaluate and optimize identity resolution models in a way that traditional systems cannot.

Known Visitors

Customers who have been identified deterministically (e.g., by logging in, completing a form, or clicking from a known email) provide a ground-truth record. These are gold standard observations — we know exactly who they are and how they behaved.

Unknown Visitors

These users have not taken any action to explicitly identify themselves, so any resolution must be probabilistic. The system uses behavioral and contextual signals to infer identity.

By applying the identity resolution model to both groups, Launch Labs can compare the model's predicted identity against the known identity for the deterministic cohort. This makes it possible to calculate true accuracy metrics — such as match precision, false positives, and resolution rates — without relying on synthetic data or assumptions.

This approach effectively turns the CDP into a live test lab for identity resolution models. Instead of tuning parameters in the dark, we can:

- Test real combinations of settings
- Measure the tradeoffs between resolution count and accuracy
- Choose configurations that match the needs of specific campaigns or use cases

Building the Optimization *Framework*

With a measurable environment in place, the next challenge is tuning the identity resolution system. At Launch Labs, we identified a set of key parameters within our resolution engine that directly influence both the accuracy and volume of resolved identities.

These parameters form the core of our optimization framework:

Max Data Freshness Days

Controls how recent a match signal must be to be considered valid. Allowing older data increases match volume but can reduce accuracy.

Allow Geolocation Match

Enables matches based on user location history. Helpful for mobile or location-rich behavior, but may introduce ambiguity in dense areas.

Require Match to Store Customer

Restricts matches to individuals who have previously made a purchase from the specific store website the current session occurred on. This increases confidence in the match, but limits scale to known customers of that store.

Require Match to Global Customer File

Expands match eligibility to a broader pool — such as historical buyers across multiple store locations or related brands. This widens the potential match base but can lower precision.

These parameters form the core of our optimization framework:

Require Last Name Match

A strong filter for identity accuracy — ensures that inferred identities share the same last name as the known customer record. Useful for high-integrity campaigns, but can reduce match volume.

By systematically adjusting these parameters and running the resolution system against a ground-truth cohort of known customers, we can directly observe how each configuration impacts system performance. For each combination, we collect:

Accuracy

The percentage of predicted identities that matched the true, known identity

Resolution Count

The total number of visitor sessions that were resolved under that configuration

These results are logged and analyzed, allowing us to rank setting combinations based on campaign goals. For example:

- A sales attribution initiative might tolerate looser rules to capture more matches, knowing that final attribution will be reconciled against actual sales records.
- A customer engagement campaign might require stricter parameters to avoid messaging the wrong individual.

Rather than tuning blindly or defaulting to vendor presets, this framework enables measurable, data-driven optimization based on the specific business context.

Multi-Objective Optimization and the *Pareto Frontier*

In optimizing an identity resolution system, we face two competing goals:

// 01

Maximize the number of resolved visitors (resolution count)

// 02

Maximize the accuracy of those matches

Improving one typically means sacrificing the other. For example, relaxing match requirements may increase volume but allow more incorrect matches. Tightening filters increases precision but can drastically reduce the number of resolved users.

To navigate this tradeoff, Launch Labs applies a multi-objective optimization approach. Instead of aiming for a single "best" setting, we evaluate every viable configuration across both metrics — and then identify which ones represent the best available tradeoffs.

This is where the **Pareto frontier** comes in.

THE PARETO FRONTIER IS THE SET OF ALL SETTING COMBINATIONS WHERE:

No other setting achieves both higher accuracy and higher resolution count

In other words, every configuration on the frontier is non-dominated — you can't improve one metric without hurting the other. These are your optimal tradeoff points.

Visualizing the Pareto frontier allows us to:

- See how changes to settings influence accuracy vs. scale
- Identify diminishing returns (e.g., a big drop in accuracy for only a small gain in volume)
- Provide marketers or data analysts with multiple viable options depending on their campaign goals

For example, one configuration might offer 92% accuracy with 1,500 matches, while another delivers 88% accuracy with 2,200 matches. Depending on whether the use case is attribution or outreach, either may be the "best" choice — but both are clearly better than configurations that underperform on both metrics.

By surfacing this frontier, Launch Labs enables clients to move beyond static resolution models and toward adaptive, context-aware resolution strategies that maximize value without compromising integrity.

Implementation
& Results

After defining the resolution parameters and establishing a validation framework using deterministic CDP data, Launch Labs implemented a full-scale evaluation across dozens of setting combinations. The goal was to observe real-world tradeoffs in identity resolution performance and provide a structured way to optimize based on business objectives.

The system iteratively tested each configuration of the five primary resolution parameters:

- Max Data Freshness Days
- Allow Geolocation Match
- Require Match to Store Customer
- Require Match to Global Customer File
- Require Last Name Match

For each configuration, the identity model was run against a cohort of deterministically identified website visitors — users whose identities were already known through login or CRM matching.

This enabled Launch Labs to calculate:

Accuracy

The percentage of predicted identities that matched the true, known identity

Resolution Count

The total number of visitor sessions that were resolved under that configuration

The results were compiled and analyzed to:

- Identify the top-performing configurations by resolution count while meeting a defined accuracy threshold
- Generate a full dataset for plotting accuracy vs. resolution volume
- Extract the Pareto frontier — the optimal set of tradeoffs for any given use case

SAMPLE OUTPUT METRICS (ILLUSTRATIVE)

Setting Combination	Accuracy	Resolutions
Strict filters (fresh data, all match rules)	0.95	900
Relaxed geolocation, global file match allowed	0.89	1,700
Broader data window, no last name match required	0.85	2,400
Very loose (all constraints off)	0.76	3,200

Launch Labs built tooling to rank combinations dynamically, allowing clients to input their acceptable minimum accuracy threshold and automatically receive the top 5 configurations that maximize reach without dropping below their required precision.

Additionally, all results were exported to CSV for deeper analysis and visualization. Scatter plots of accuracy vs. resolution count made it easy to identify high-performing zones and avoid poor tradeoffs. Internal tools now make this process repeatable — allowing resolution systems to be re-optimized over time as customer behavior and data quality change.

Broader Impact on Marketing Strategy & *Customer Data Management*

Launch Labs' optimization framework doesn't just improve resolution performance — it creates a more agile, transparent, and outcome-aligned approach to customer identity as a whole. By making resolution tunable, measurable, and context-aware, it unlocks new capabilities across multiple areas of marketing and data operations.

Campaign Performance

Identity resolution quality directly impacts campaign results. A more accurate resolution model means more relevant outreach, less waste, and better engagement metrics. By selecting different resolution strategies for different campaign types (e.g. strict for direct mail, looser for audience modeling), teams can tailor tactics to maximize ROI.

Sales Attribution & Media Measurement

Enables matches based on user location history. Helpful for mobile or location-rich behavior, but may introduce ambiguity in dense areas.

CRM Enrichment & Retargeting

Improved resolution accuracy enables smarter enrichment of CRM records — adding browsing behavior, ad engagement, or offline visit data to known customers. It also supports high-integrity retargeting, where marketers can be confident they're speaking to the right person on the right device.

Data Governance & Privacy

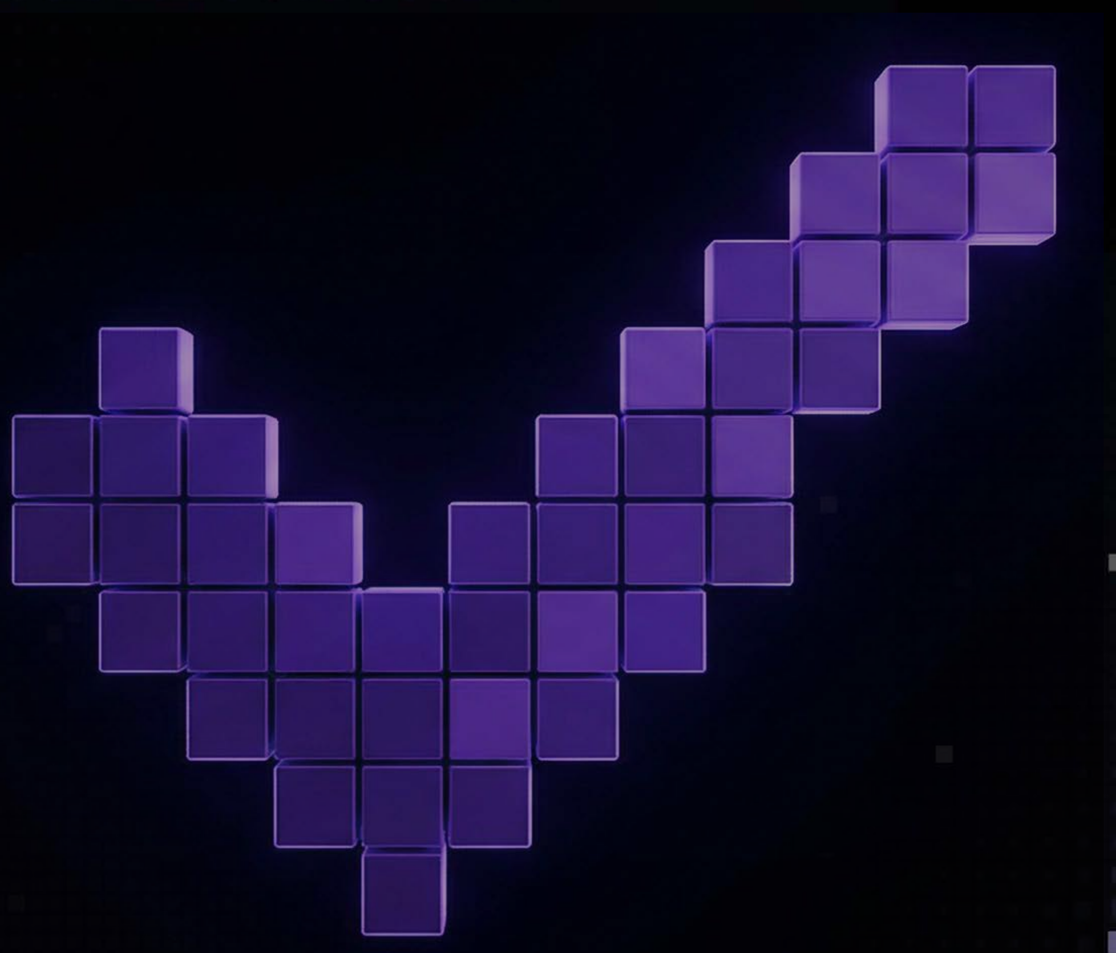
By embedding a precision-aware resolution process, the system builds in natural safeguards against overly aggressive or misdirected identity usage. Marketers can define their acceptable risk tolerance, ensuring outreach campaigns maintain ethical and compliant targeting standards.

Adaptive, Real-Time Tuning

Most importantly, the framework is designed to be repeatable and adaptive. Businesses can re-run the optimization process regularly to reflect:

- Seasonal or regional traffic changes
- Shifts in data quality or signal strength
- Evolving campaign strategies or goals

Conclusion



CONCLUSION

Identity resolution is one of the most powerful — and most opaque — components of the modern marketing stack. Without a way to measure or control it, businesses are left guessing whether their audiences are real, their attribution is valid, or their outreach is reaching the right people.

At Launch Labs, we believe identity resolution shouldn't be a black box. By leveraging the ground-truth data available within Customer Data Platforms (CDPs), we've developed a framework to directly measure and optimize resolution systems in a way that's grounded, repeatable, and adaptable.

Our approach brings:

Visibility into the accuracy and scale tradeoffs of identity resolution

Control through tunable settings that can be adjusted by business use case

Optimization via multi-objective analysis and Pareto frontier modeling

Whether the goal is attribution accuracy, CRM enrichment, or scalable customer engagement, this framework allows businesses to align identity resolution performance with the outcomes that matter most.

As identity continues to evolve in a privacy-conscious, signal-fragmented world, this kind of flexible, data-driven approach will be essential. Resolution isn't just about linking data — it's about making every downstream action more intelligent, more efficient, and more trustworthy.

Final Results
& Visualization

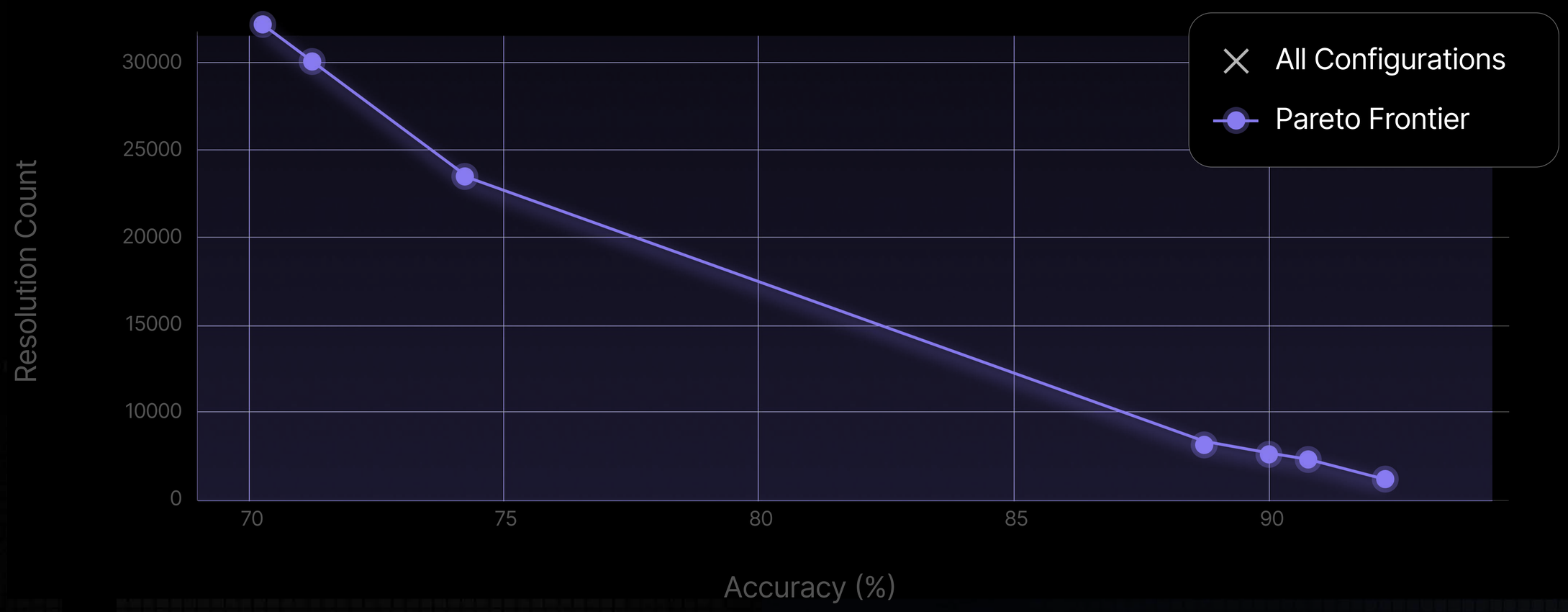
FINAL RESULTS & VISUALIZATION

Using real-world data from a multi-location retail group, Launch Labs ran 64 combinations of identity resolution model settings. Each configuration was evaluated against a cohort of deterministically identified website visitors — allowing us to directly measure both accuracy (percent of correct matches) and resolution count (total number of resolved identities).

The chart below visualizes all tested configurations:

- Each point represents one unique combination of resolution settings.
- The x-axis shows measured accuracy.
- The y-axis shows the number of identities resolved under that configuration.

Accuracy vs. Resolution Count with Pareto Frontier



The red line traces the Pareto frontier — the set of optimal tradeoffs where no setting is strictly better in both accuracy and resolution. Configurations below the line are dominated; they can be outperformed by others in at least one dimension without sacrificing the other.

FINAL RESULTS & VISUALIZATION

This visualization enables stakeholders to:

- Select high-volume configurations for use cases like sales attribution where precision is less critical.
- Choose high-precision settings for engagement campaigns where customer experience and targeting integrity are paramount.

These results reflect the behavior of an actual retail network with multiple stores and a diverse customer base. However, results can vary based on factors like geographic concentration of store locations, customer overlap between stores, and available signal quality (e.g., mobile behavior, past purchase volume, etc.).

Launch Labs' framework is designed to be re-run per client, vertical, or regional footprint — ensuring that each business can optimize identity resolution using data from its own ecosystem.

Ready to optimize your identity resolution performance?

Launch Labs' Enterprise Data Solutions give you the tools to measure, tune, and scale identity resolution with confidence. From accuracy benchmarking to real-time optimization, our platform helps you make smarter decisions grounded in your own data.

Explore how our solutions can bring clarity, control, and impact to your marketing strategy.

LAUNCHLABS.AI/ENTERPRISE-DATA-SOLUTIONS

APPENDIX

Key Terms *& Concepts*

Identity Resolution

The process of associating anonymous or semi-anonymous digital signals (devices, IPs, cookies) with a known individual.

Deterministic Identity

A known identity verified by login, email match, CRM record, or other explicit data source.

Probabilistic Identity

An inferred identity based on behavioral, contextual, or location-based signals.

Resolution Count

The number of visitor sessions successfully matched to a known identity.

Accuracy / Precision

The percentage of resolved identities that were correctly matched to the true customer (as verified by deterministic data).

Pareto Frontier

A set of optimal tradeoffs in multi-objective optimization where no solution can improve one metric without sacrificing another.