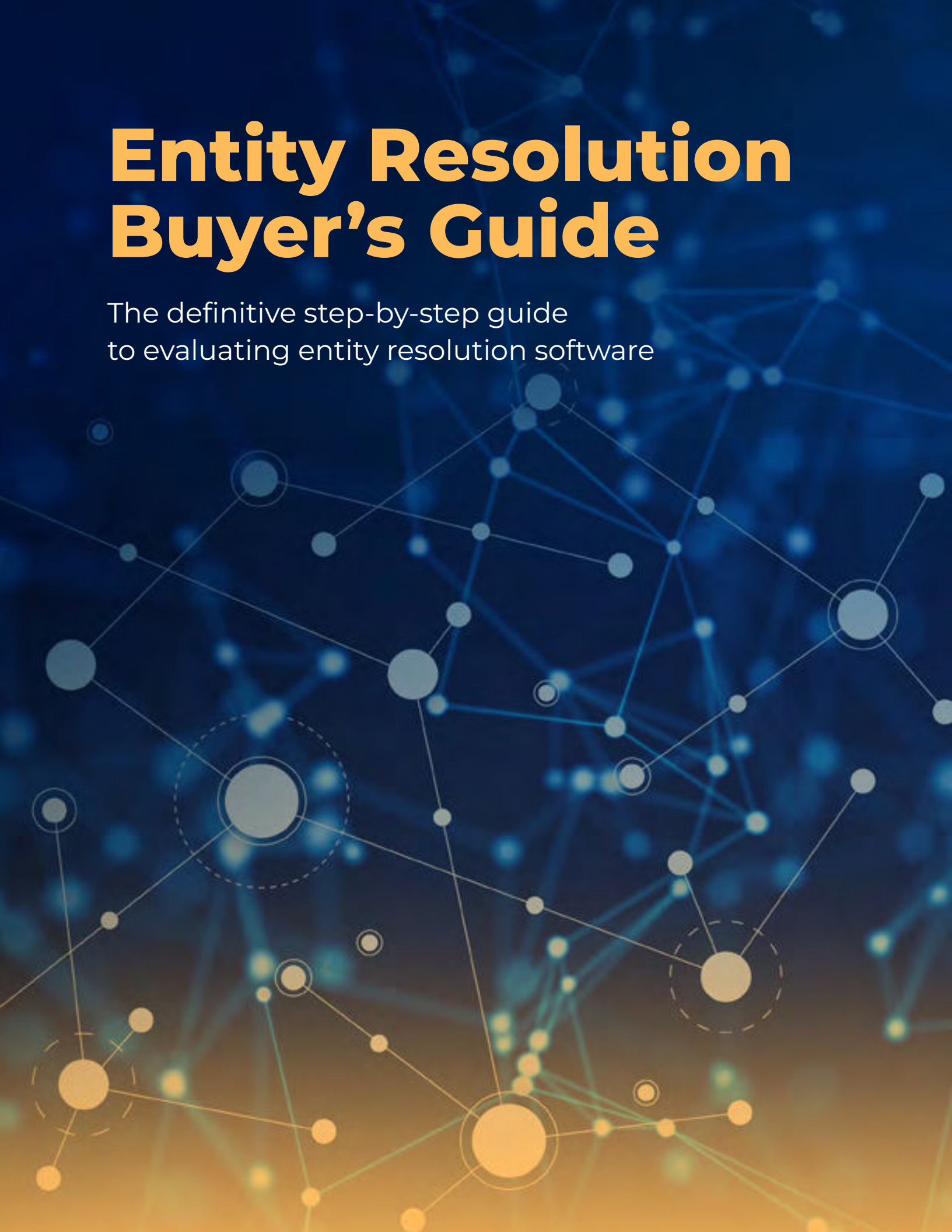
Entity Resolution Buyer's Guide

The definitive step-by-step guide
to evaluating entity resolution software



Purpose of this Buyer's Guide

This guide will cover:

- The value of entity resolution
- Different types of solutions available
- Key capabilities for evaluation
- How to get started with entity resolution

The purpose of this Buyer's Guide is to provide an overview of different types of entity resolution solutions and help you understand what to look for when evaluating and selecting one.

By reading the guide, you will be better equipped to make an informed decision about which type of entity resolution solution is right for your needs today, and into the future.

What's Inside...

Entity Resolution Defined 4

The Value of Entity Resolution 5

How Entity Resolution is Used

- Five Primary Business Use Cases 6
- Four Primary Deployment Methods 7

Technology Choices for Entity Resolution

- Full Stack vs. API 9
- Cloud or On-Prem 10

Evaluation Criteria 11

- Accuracy 12
- Time-to-Value 13
- Total Cost of Ownership 14
- Ease of Use 15
- Relationship Detection 15
- Real-Time vs. Batch 15
- Explainable Results 16
- Scalability 16
- Privacy 16
- Security 16

How to Get Started 17

Entity Resolution

compound noun

en·ti·ty res·o·lu·tion | \ ˈen-tə-tē ˌre-zə-ˈlü-shən

Entity resolution is the process of identifying records that refer to the same unique real-world entity – such as a person or organization.

Entity resolution can help an organization gain a competitive advantage.

Entity resolution matches and links data about people and organizations to create 360-degree views of entities.

ENTITY RESOLUTION POWERS DISCOVERY + INSIGHT

- Improves decision-making
- Bolsters analytics and insights
- Enhances data quality and accuracy

The five primary business use cases for entity resolution today include:

1. **Risk and Fraud Detection**

   Identifying anomalies to proactively find bad actors, mitigate risk and curb fraudulent activities.

2. **Financial and Regulatory Compliance**

   Providing the data necessary to meet government and industry requirements and standards.

3. **Customer Data Management**

   Enhancing customer matching to increase accuracy and insights within and across systems.

4. **Marketing and Customer Analytics**

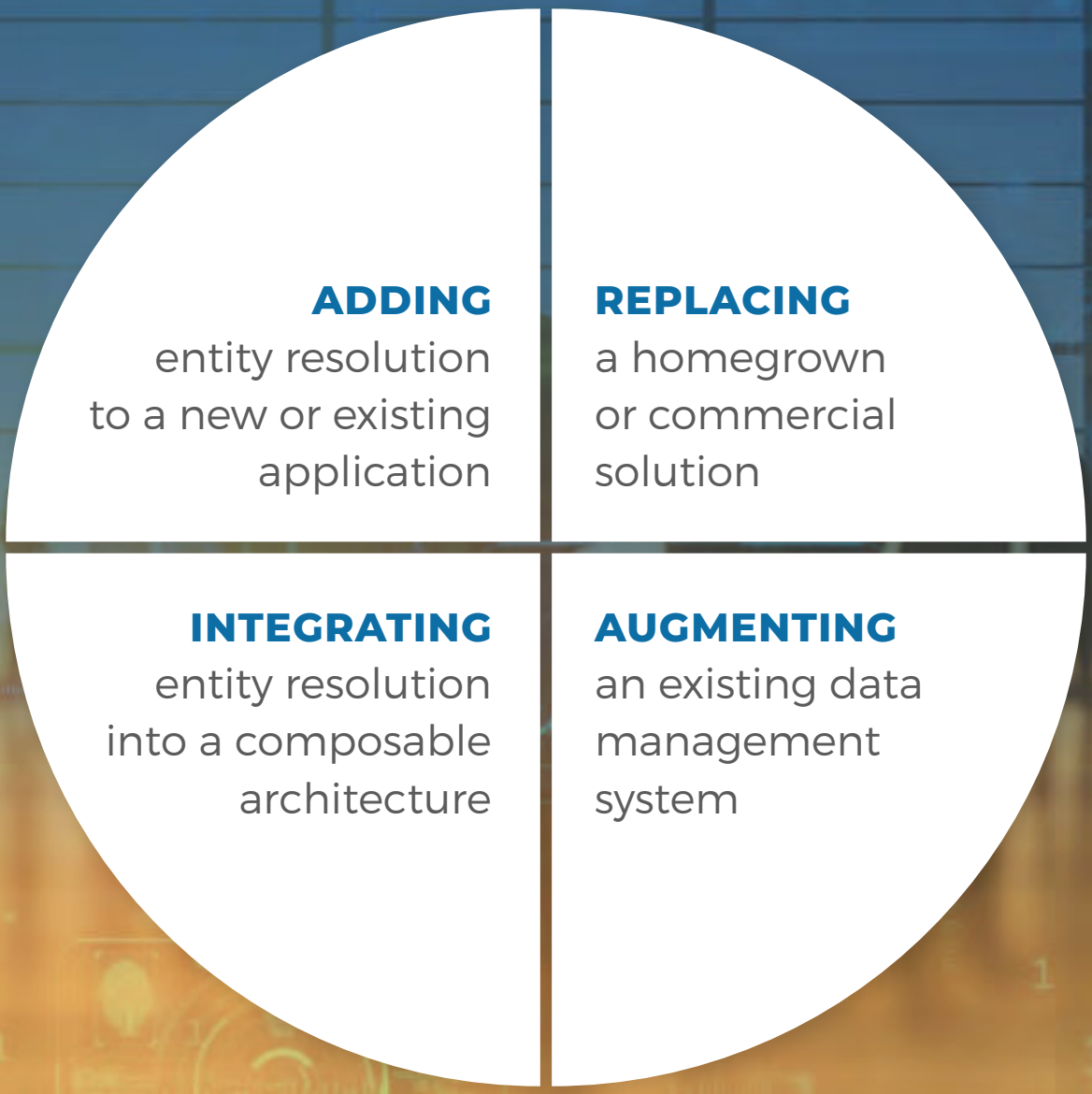
   Gaining complete 360-degree views of customers to improve and refine marketing strategies.

5. **Enhance MDM, CRM, CDI and Other Software**

   Increasing accuracy, enabling new capabilities, and enhancing downstream analytics.

Define your organization's use case for entity resolution before evaluating solutions.

There are four primary ways organizations use entity resolution today:



Clearly understand how you plan to use your entity resolution system.

Four Primary Deployment Methods

Let's take a closer look at the four ways that enterprises are using entity resolution:

1. **Adding to a New or Existing Application**

   Many enterprises and ISVs want to add entity resolution to a new or existing product to improve data matching, eliminate manual processes, or support new deployment models.

2. **Replacing a Homegrown or Commercial Solution**

   The need to replace existing, limited entity resolution capabilities is driving many organizations to look for options that are more accurate, cost effective and enable business expansion.

3. **Integrating Entity Resolution into a Composable Architecture**

   Enterprises building data fabrics and data pipelines, or implementing digital transformation initiatives, often deploy entity resolution as an enterprise-wide service in their architectures.

4. **Augmenting an Existing Data Management System**

   Some organizations want to run entity resolution systems alongside applications that handle large amounts of data about people and organizations to improve data matching and extend capabilities.

Assess Full Stack vs. API Technologies

The first consideration when purchasing an entity resolution solution is to understand the differences between full stack and API approaches.

Full Stack Solutions

Full stack entity resolution solutions are comprehensive offerings that deliver a wide range of functionality but can be costly and complex to integrate and manage. They often provide more capabilities than you need and have specific infrastructure requirements that can impact how you architect and deploy your products or services.

Despite these issues, full stack solutions have been the norm until recently, and are still widely used.

API-Based Solutions

API-based entity resolution solutions are modern, modular components that are usually easy to manage and quick to deploy. They can be added into enterprise or commercial applications and services or deployed as part of an enterprise-wide composable architecture.

API solutions can offer integration and deployment flexibility and other operational advantages. In addition, they can substantially reduce the learning curve to get started versus full stack solutions.

Define and rank your project's requirements to better assess whether a full stack or API approach will meet your needs.

Choose Between Cloud or On-Prem Infrastructure

You'll need to determine where you will initially deploy your entity resolution system. Many organizations start with a cloud-based infrastructure due to lower initial costs but may eventually transition to on-premises. Others deploy on-premises first but plan to move to the cloud over time.

Anticipating the future trajectory of your entity resolution system will enable you to select a solution that provides the flexibility you need to meet evolving needs.

Make sure the solutions you evaluate support all the deployment options you'll need.

How to Evaluate the Capabilities of Different Systems

As you compare different entity resolution options, several criteria should guide your evaluation.

We'll first delve into the top three criteria:

ACCURACY | TIME-TO-VALUE | TOTAL COST OF OWNERSHIP

Then we'll explore seven additional factors to consider in your evaluation.

1. Accuracy

What's involved in time, effort and costs to achieve target accuracy?

Accuracy is one of the top factors to evaluate when considering entity resolution systems. Accuracy is critical because low error rates ensure your organization's data is dependable and credible.

Pre-Configured Accuracy vs. Tuned and Trained Accuracy

It's important to understand the difference between out-of-the-box accuracy and tuned and trained accuracy. Some solutions come pre-tuned and pre-trained and actively learn to deliver highly accurate results more quickly, and can be microtuned as needed for edge cases.

Other systems require substantial time and effort to reach and maintain target accuracy levels. Rules-based entity resolution systems require experts to write rules that improve accuracy and modify or add rules as new data sources are added. Probabilistic systems require thresholds to be adjusted to tune for accuracy based on the attributes within each data source.

Evaluate how much time and effort is needed to achieve and maintain the accuracy levels your project requires.

2. Time-to-Value

How long to deliver value in production?

Time-to-value refers to how long it takes to get a system operational from installation to production. Calculations should include the time required for installation, onboarding data sources and training, tuning and testing the system. Results can vary widely depending on the solution.

Hardware Setup and Software Installation

Full-stack solutions typically demand specialized vendor assistance for installation, turning it into a weeks- or months-long process. API solutions often offer user-friendly setup procedures and installation that takes a day or less.

Data Source Onboarding

The time required to onboard initial data sources is crucial when calculating time-to-value. Some systems can take hours or days to onboard all sources. Others may need a month or longer to add just a single source.

Initial Training and Tuning

Certain systems may require a year or more of training and tuning before yielding results accurate enough for production. Determine how long training and tuning will take (days or months) and how many internal or external resources are required.

Understand everything involved to achieve time-to-value when evaluating entity resolution options.

3. Total Cost of Ownership

How to fully evaluate all costs over time.

The total cost of ownership (TCO) is based on initial deployment costs plus ongoing operation and maintenance costs. Since the solution you choose will be deployed for many years, ongoing costs will be a major part of TCO.

Initial Deployment Costs

Initial deployment costs include all the software you'll need, hardware or cloud infrastructure, and the resources required to onboard data sources and tune and train the system. Some entity resolution solutions require expensive, vendor-specific hardware and software that can make integration challenging.

Ongoing Operating and Maintenance Costs

When projecting ongoing costs, be sure to include both staffing expenses and additional costs associated with software, infrastructure complexity, and increasing data volumes. Adding a new data source can be complex and costly for many solutions, both to onboard the data and tune and train the system. Some systems also require data to be reloaded regularly to maintain accuracy, which increases maintenance and admin costs.

Ensure your TCO calculations focus on both initial deployment and ongoing operations and maintenance.

4. Ease of Use

Systems vary significantly in terms of the user-friendliness of the entire entity resolution process. Make sure you understand how easy a system is to use from installation and data source integration to tuning and training.

5. Relationship Detection

Most entity resolution systems perform record matching to identify people and organizations. Some systems also identify disclosed and discovered relationships between entities for richer context and deeper insights. If your organization would benefit from understanding households or other relationship networks, confirm any system you evaluate supports these capabilities.

6. Real Time vs. Batch

While many organizations rely on batch-based entity resolution, the need for real time is becoming increasingly important. Batch data is easily supported by real-time systems, but transforming a batch system to real time can be impossible or expensive. Even if you don't need it today, we recommend choosing a real-time system to future proof your solution.

If you opt for real time, clearly identify the specific real-time capabilities a solution provides, as different vendors define "real time" differently. Some just mean real-time querying of data that's already in the system. Others mean continuously ingesting, resolving, querying, deleting and self-correcting streaming data as it is received in real time.

7. Explainable Results

Systems vary in terms of the types of explanations they provide about how entity resolution decisions are made. Some offer no information at all. Since it's common for business users, data scientists and auditors to want to know why records matched or didn't, you want a system that provides details about how specific decisions were made.

8. Scalability

Your system may need to scale to support larger data volumes, more data complexity or new features. Since scalability is difficult to retrofit, determine whether a solution can meet the volumes of data and types of performance-intensive capabilities you'll need for current and future requirements.

9. Privacy

It's important to know if anyone outside your organization will have access to your data. Some systems are so complex that you may need to give a vendor access. You should also investigate the details about where your data will flow and what data is or isn't encrypted.

10. Security

The more complex the solution, the more authentication and authorization mechanisms your team will need to learn and audit, and the more systems you'll be required to monitor, patch and maintain. Some solutions require you to add security patches to a variety of technologies and services while others need few if any changes to your existing security methods.

Senzing Entity Resolution API

Senzing software makes it easy and affordable to add advanced entity resolution capabilities to your enterprise systems and commercial applications.

The Senzing API provides highly accurate data matching and relationship detection to improve analytics, insights and outcomes with no entity resolution experts required.

You can be up and running in minutes and deploy into production in weeks.

Learn more about Senzing entity resolution:

Consult with an Expert

Schedule a call with a Senzing entity resolution expert to discuss your requirements.

Try it Yourself

There are three easy ways to take Senzing entity resolution for a test drive – a simple desktop eval tool (for Windows or Mac) and QuickStarts for Linux and Docker. You can install the software, load data and evaluate results in as little as 15 minutes.

Visit Senzing.com