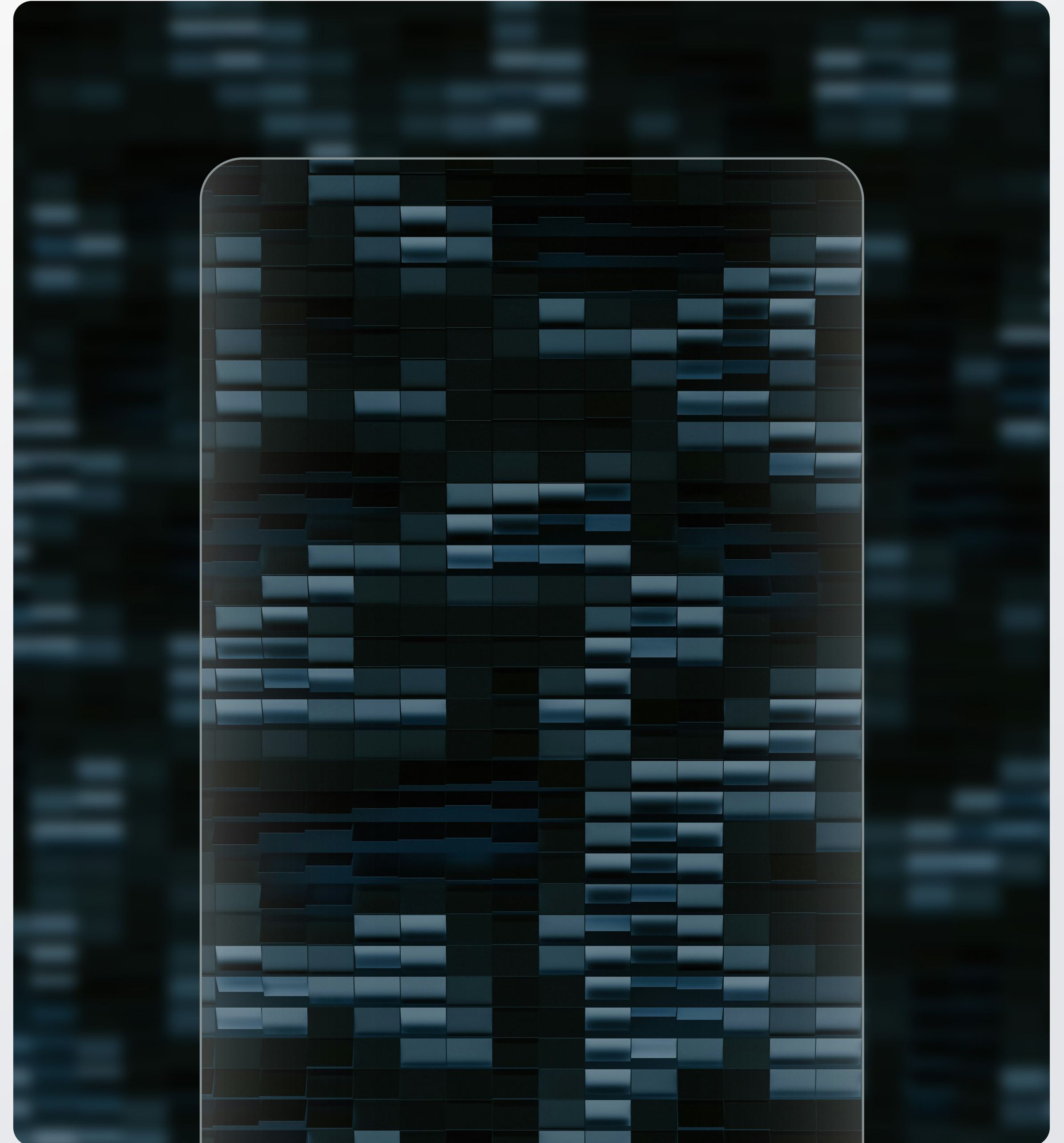# kernel

Kernel's Cleaning & Hierarchies Guide

How AI Cleaning & Hierarchies work →

Cleaning action

The cleaning action is a singular field that summarizes the main action to take for each account

Cleaning action	Definition
None	No action
Delete	The account is dead or dormant and it is safe to delete (e.g. no active opportunities, actively being worked on, etc.)
Merge	The account has a duplicate and is safe to merge according to the safety rules (e.g. no active opportunities, actively being worked on, etc.)
Associate	The account has a parent in the CRM and should be linked to the Parentid
Associate (missing parent)	The account has a parent that isn't in the CRM, which should be added as a new account to complete the family tree
Review	Account identity issues or name/URL wrong; Kernel provides auto-recommendation

Examples of using the cleaning action

— Routinely mass-delete all accounts with the cleaning action "Delete"

— Routinely mass-merge all accounts with the cleaning action "Merge"

   — You can also merge accounts without this cleaning action, but the "Merge" action indicates that the account are safe to merge

— Run Kernel's automatic suggestions on accounts with the cleaning action "Review"

— Link all accounts to their parent in the ParentId field that have the cleaning action "Associate"

   — By default, Kernel will add the identified parent ID in a separate custom field

Duplicate types

Kernel categorizes all duplicates by the following types. The purpose of categorizing duplicates into different types is to provide you with the flexibility to define what a "duplicate" means for your business. This allows for a more nuanced approach than a simple "match" or "no match" system. For example, you can configure rules to automatically merge Exact duplicates while excluding Regional duplicates, allowing you to keep international offices as separate accounts if desired.

Type	Definition
Primary	In a group of duplicates, this is marked as the surviving account
Exact	Exact match found after 'cleaning' and standardizing the URLs; merging logic depending on internal data such as activity, creation date, tasks, leads, contacts, and opportunities, completed account fields, and custom fields specific to you.
Subdomain	Similar to exact match, e.g. shop.ccs.com vs ccs.com
Regional	fr.amazon.com, amazon.fr, amazon.com are all regional duplicates, but apollo.de and apollo.com are not.
Potential	This is a catch-all category for all the hard cases that require expensive work, e.g. inside-sephora.com vs. sephora.com.

How Duplicates are Identified

To identify duplicates, Kernel performs a comprehensive analysis that reviews an account's complete identity as it exists within your CRM. This includes analyzing the account name, associated notes, and opportunity data. This internal context is combined with an analysis of the company's website content and a visual comparison against all other potential duplicate candidates in your CRM. The goal is to understand the relationship between accounts as a sales representative would.

Safeguards for Merging

These classifications allow you to build a tailored deduplication strategy. It is important to note that an account will only be assigned the "Merge" cleaning action if it is identified as a duplicate and it has a low Quality Score. This safety feature ensures that important accounts with active sales cycles or significant history are protected from being merged or deleted.

Family tree types

An account's position within a corporate family tree is defined by its type and sub-type. These classifications determine how accounts are related to one another within your CRM.

Type	Definition	Examples
PARENT	Any entity that is not owned by another operating entity (excluding holding companies)	Berkshire Hathaway, Kraft-Heinz
CHILD	Any entity that is owned by another operating entity	Heinz, Jell-O, Oscar Mayer

Type	Subtype	Definition	Examples
PARENT	OPERATING	Any parent entity that is not a holding company	Kraft-Heinz
PARENT	HOLDING	Any parent entity that is a holding company	Berkshire Hathaway
CHILD	STANDALONE	Any subsidiary that is not absorbed into its parent company	Jell-O, Oscar Mayer
CHILD	ABSORBED	Any subsidiary that is absorbed into its parent company	Heinz
CHILD	REGIONAL	Any subsidiary that is a regional subsidiary of its parent company	Jell-O UK

Building and Viewing Hierarchies

Kernel first builds the complete corporate hierarchy externally. This structure is then provided to you for review as a flat file (such as a CSV or spreadsheet). This file clearly maps out the proposed parent-child relationships before any changes are made in your CRM.

Upon your approval, Kernel adds all missing accounts to your CRM via the API and implements the hierarchy by populating the standard Parent Account field in Salesforce. Each child account is linked to its parent using their unique and permanent 18-digit Salesforce Account IDs, creating a clean and reliable structure directly within your CRM.

Kernel's Approach vs. the D-U-N-S Number

While a unique company identifier is essential, traditional systems like the Dun & Bradstreet D-U-N-S number often create more problems than they solve. The core issue is that numbers are assigned without first resolving underlying data chaos.

For instance, the D-U-N-S system frequently assigns multiple numbers to the same core business for different reasons—a new legal entity, a separate branch, or a slight name change. The result in a CRM is predictable: duplicate accounts that disrupt territory management, skew reporting, and create operational friction.

Kernel's approach is fundamentally different. We believe in cleansing the data first. Our methodology focuses on resolving identity conflicts and eliminating duplicates to establish a pristine and accurate company hierarchy.

Upon this clean foundation, we will introduce our own unique Kernel ID, guaranteeing it serves as a trustworthy identifier that avoids the pitfalls of rigid, legacy systems.

Safeguards

Kernel uses an adjustable algorithm, using internal data and outputs of the website analysis, to (a) safeguard accounts that are actively being worked, preventing cleaning from disrupting existing business activities, and (b) identify the primary account in a group of duplicates.

The output is a quality score, ranging from 0 to 100, which is used to determine the cleaning action and duplicate type.

The following data points are used in creating the quality score:

Component	Definition	Weight
Recent activity	Favours accounts that have been recently worked on	Medium
Creation Date	Favours older accounts over new ones	Low
Opportunity	Accounts with many opportunities - open and closed - are heavily restricted from updates	High
Contacts	Favour accounts with many contacts associated	Medium
Tasks	Favour accounts with many tasks associated (emails, phone logs)	Low
Ownership status	Favours active sales owners over non-sales owners, inactive owners, and integration users	Medium
Core account fields	Favours accounts that have more core (non-custom) fields filled in, like address, description, and type.	Low
Custom logic	Custom logic based on custom account fields or custom logic, e.g. fields indicating account lifecycle stage or status	High
TLD	Certain top-level domains (TLDs), such as .com, are favoured over others	High
Redirects	Websites that redirect to another website are penalized	Medium

kernel

Kernel's Cleaning & Hierarchies Guide

Build AI customization on clean master data with Kernel.

Book a demo