

# Discover, search, and catalog your data



**Duplicate Discovery**

## Automatically discover and identify data duplicates

### The Challenge

Technology is evolving faster than ever before, and data continue to grow and change at an exponential rate. Enterprise data, therefore, become increasingly fragmented, and finding the right information becomes increasingly difficult. Over time, data management practices that solve these fragmentation and discovery challenges become absolutely essential - for enabling business analytics, ensuring regulatory compliance, and most importantly, continuing the successful operations of the enterprise. Traditional, and often manual approaches to discovering and cataloging data are simply too slow to meet the aspirations of the modern enterprise. To capture, manage and govern data, organizations must first discover and understand both the data, as well as the relationships among data across an inherently heterogeneous and complex enterprise landscape.

### The Solution

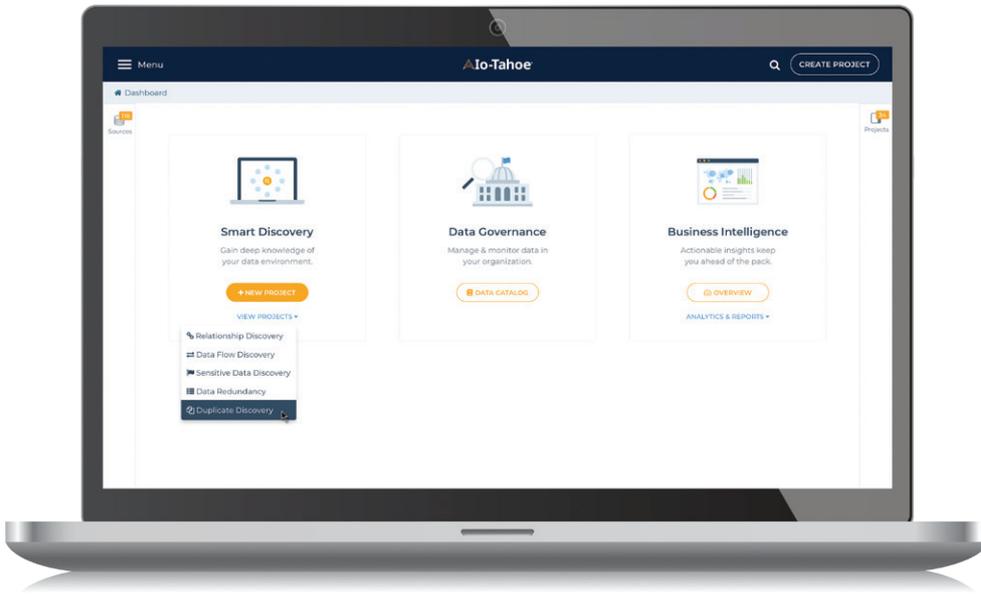
Where most data management products leverage only metadata, Io-Tahoe's discovery goes well beyond traditional metadata - leveraging patent pending artificial intelligence (AI) and machine learning (ML) technology to automatically discover and identify data duplicates across multiple data sources, based on the relationships between various records, attributes and clusters. Through the combination of human and machine intelligence, Io-Tahoe users are able to achieve optimal de-duplication outcome. Additionally, Io-Tahoe's duplicate discovery also provides continuous re-learning capability that becomes smarter on subsequent re-runs.

We provide a **Smart Data Discovery** and **AI-Driven Data Catalog** product that enables **data management**, **data governance**, and **regulatory compliance**.

“*This capability is a significant enhancement to our Smart Data Discovery platform. Organizations have multiple systems, many of which, will contain copies of the same or similar documents, files and even numerous records of data that could be associated with one single customer. With Io-Tahoe's duplicate discovery, organizations are now able to discover and identify data duplicates. The findings of this discovery enable the organization to ultimately achieve a single unique instance of a data record when there were previously two or more similar records.*”

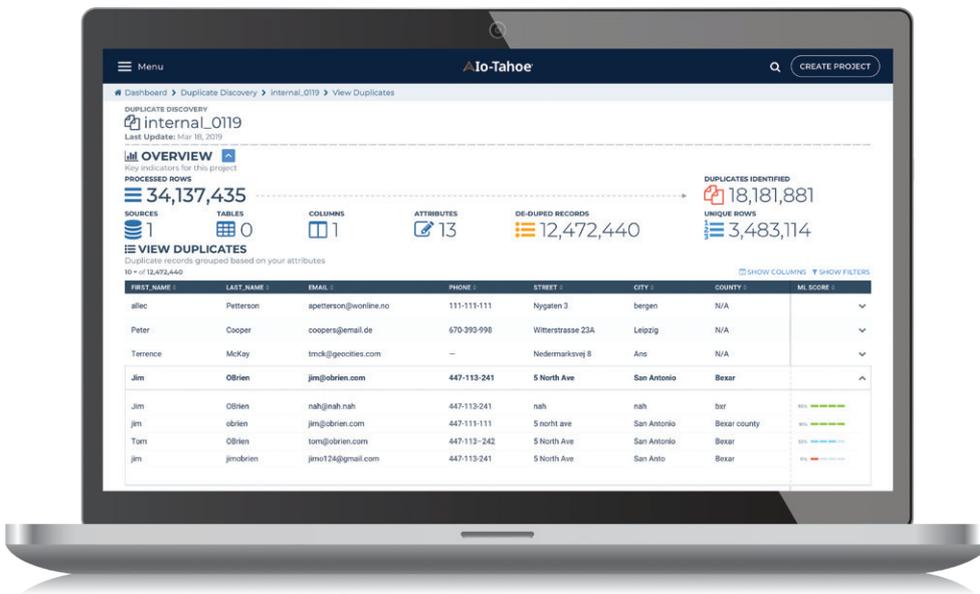
**Rohit Mahajan,**  
Co-Founder, Chief Technology  
and Product Officer

## Duplicate Discovery



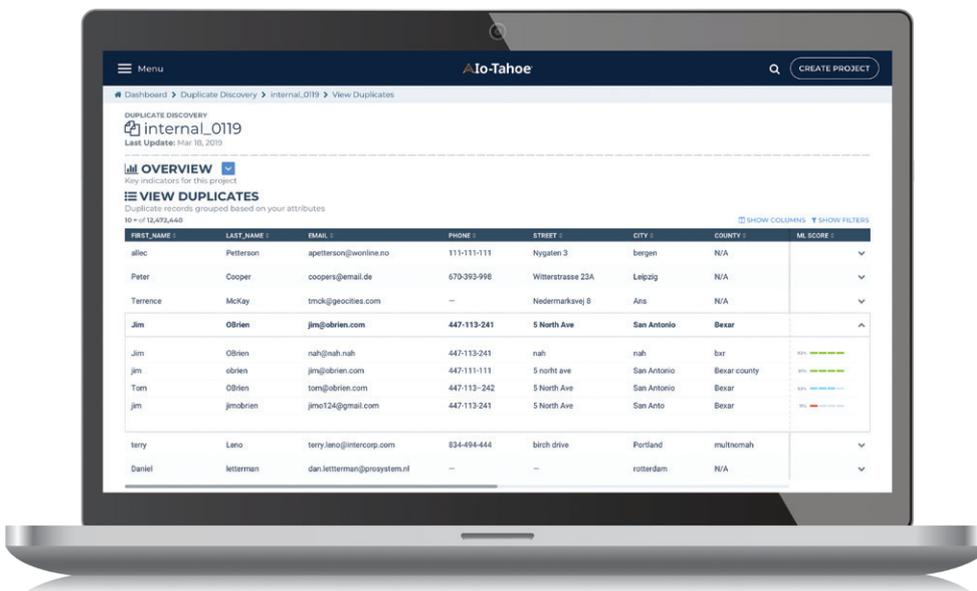
The user selects Duplicate Discovery, and the attributes to feed the algorithm to start the de-duplication process. Automatically identify duplicates across rows.

## Discover Data Duplicates



The user must select up to 20 attributes or columns. A group of duplicates is expanded to show multiple duplicates of the same individual, and best version.

## Create a Single Customer Record



The summary shows an overview of the number of duplicates and clusters identified by Io-Tahoe.