

# Data Minimization Challenges

Data minimization is a critical component to cloud migrations, privacy and protection regulation compliance, data quality initiatives, security considerations, and more. Given the sheer volume of data that enterprises store, data management dynamics have dramatically shifted: duplicate and redundant data amplifies the risk of data breaches, complicates (and compromises) cloud migration strategies, and often falls under specific privacy and protection regulations.

Organizations can address data minimization challenges with a systematic and automated approach to reduce security, privacy compliance and governance risks: removing and consolidating duplicate (or similar) data.

## Data Minimization: The Solution

The first step to data minimization is to identify duplicate and similar data and determine what data should - and can - be minimized. This can be a daunting task after literally years of data sprawl, which has exacerbated the scalability and coverage limitations of traditional tools to scan, discover and classify data.

Data privacy regulations increasingly emphasize that enterprises collect only what they need under the principle of data minimization, and every file that is redundant or a duplicate unnecessarily expands the scope of data that must be secured - especially if it is sensitive or personal data.

Enterprises need to be able to find duplicate and similar data at scale for data consolidation across files and databases to address these interlocking needs.

### Identify Duplicate and Similar Data

Identify duplicate and near-duplicate files containing sensitive and personal data at scale using machine learning across large volumes of unstructured data - whether on premise or in the cloud.

### Accelerate Time To Privacy Compliance

Reduce time and improve accuracy in identifying duplicate files and data copies that potentially violate minimization policies. Apply retention policies to original files.

### Automate Accurate Classification

Automatically classify and apply labels to files containing similar attributes and common key words, minimizing manual steps and enabling scale.

### Reduce Operating Costs

Minimize the number of duplicate and near-duplicate files to reduce storage overhead, identify low value data and enable lower operating costs & minimize on-prem footprint.

# Data Minimization Solutions

BigID helps enterprises execute a multi-pronged data minimization strategy through the ability to classify and label data, files and documents at scale, and automatically identify duplicate and similar data. Using advanced machine learning, BigID brings clarity and visibility to large volumes of unstructured data to help enterprises understand what data to protect, migrate, govern or minimize.

With BigID, enterprises can automatically identify duplicate and similar data, gain accuracy in classification, and improve time to visibility to reduce risk and manage compliance for data minimization.

## How BigID Helps with Data Minimization



### Reduce Data Security Risk

Automatically identify unused, duplicate, similar or redundant data within large volumes of data to reduce unnecessary data exposure (and secure vulnerable data) by leveraging BigID's ML-patented discovery-in-depth technology.



### Operationalize Privacy Compliance

Automatically classify regulated and sensitive data via machine learning, drill down into classification results to identify and remediate duplicate & near-duplicate copies of files containing PI and PII, and apply retention policies to the original files.



### Enable Data Utilization Audits

Automate discovery & classification of high-risk, low-value use of personal and sensitive data for data minimization and identify legal or regulatory exposure from duplicate data assets. Propagate manually defined labels across all discovered files.



### Reduce Data Footprint

Gain automated insight at scale into unstructured data to minimize storage footprint and identify duplicate data before migration for cloud migration initiatives. Use fuzzy hash matching to evaluate whether similar files should be retained.