



Clean Core: Data Dimension



por Mayko Silva

Understanding Clean Data in SAP

When we talk about clean data in the context of a clean core strategy, we're referring to keeping your data lean and well-managed. This isn't just about having accurate information – it's about having the right data, in the right place, at the right time, without unnecessary bloat.

Think of your SAP system like your smartphone. If you never delete old photos, messages, and unused apps, eventually it starts running slower and taking up valuable storage space. The same principle applies to your SAP environment.

Right Data

Ensuring your system contains only necessary and relevant information

Right Place

Organizing data logically within your SAP architecture

Right Time

Making data available when needed and removing it when obsolete



Key Goals for Clean Data

Let's talk about what we're trying to achieve with clean data. There are several important outcomes we're aiming for:

1

System Performance

Clean data ensures your SAP system runs efficiently without unnecessary processing load

2

User Experience

Well-organized data makes it easier for users to find and work with information

3

Business Decisions

Quality data leads to more accurate reporting and better strategic choices

Accuracy and Completeness

First, your data should be **accurate** - reflecting the real-world truth of your business. It should be **complete** - containing all necessary information for business processes.

Accuracy

Accurate data correctly represents real-world facts and conditions. For example, customer addresses must match their actual locations, and inventory counts must reflect physical reality.

Completeness

Complete data includes all required fields and information needed for business processes to function properly. Missing data points can cause process failures or incorrect decisions.

Consistency and Timeliness

Your data needs to be **consistent** - the same information should match across all parts of the system. It must be **timely** - available and accessible when needed.



Consistency

When customer information appears in multiple places, it should show the same details everywhere. Inconsistencies create confusion and process errors.



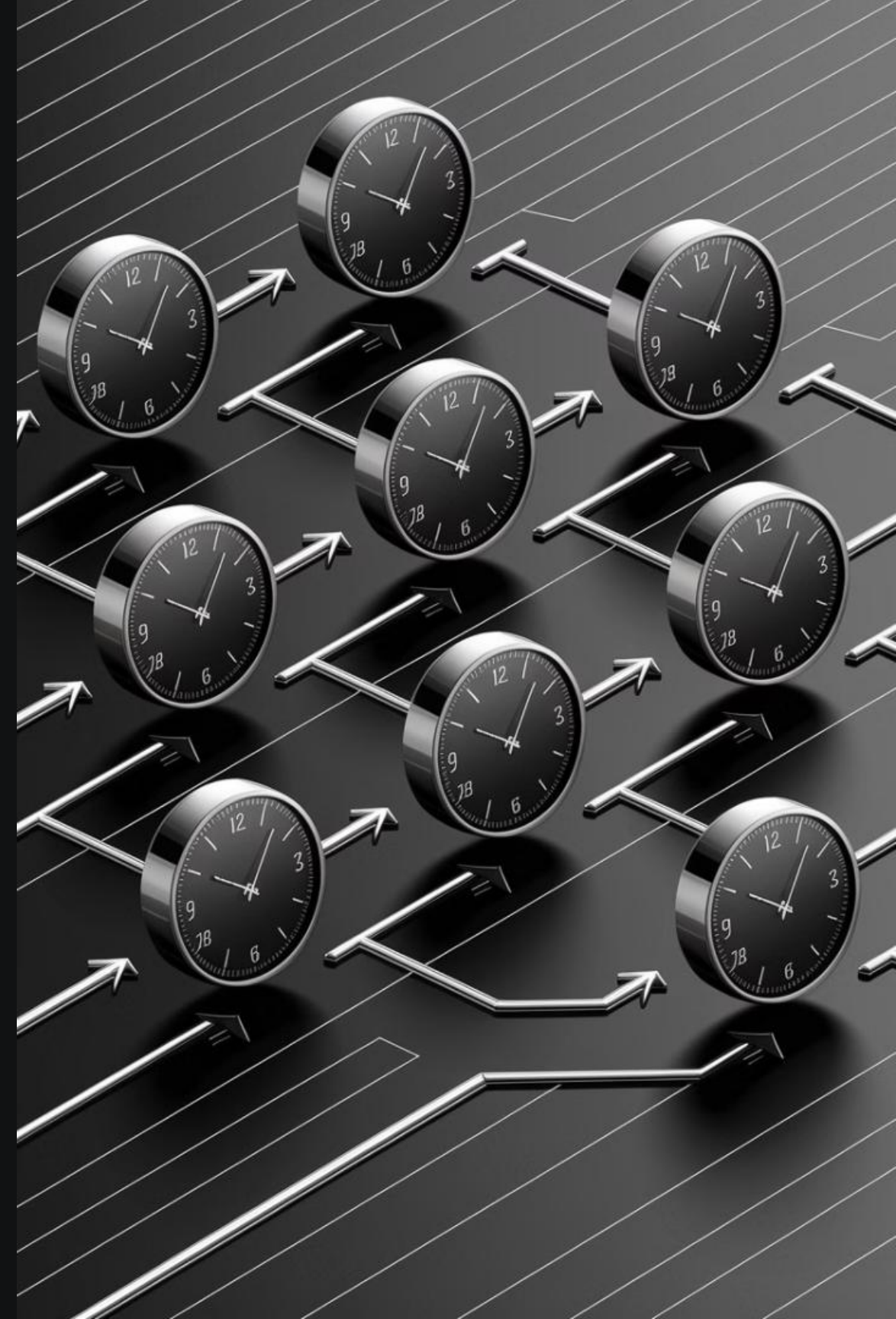
Timeliness

Data must be available when business processes require it. Delayed information can lead to missed opportunities or incorrect decisions.



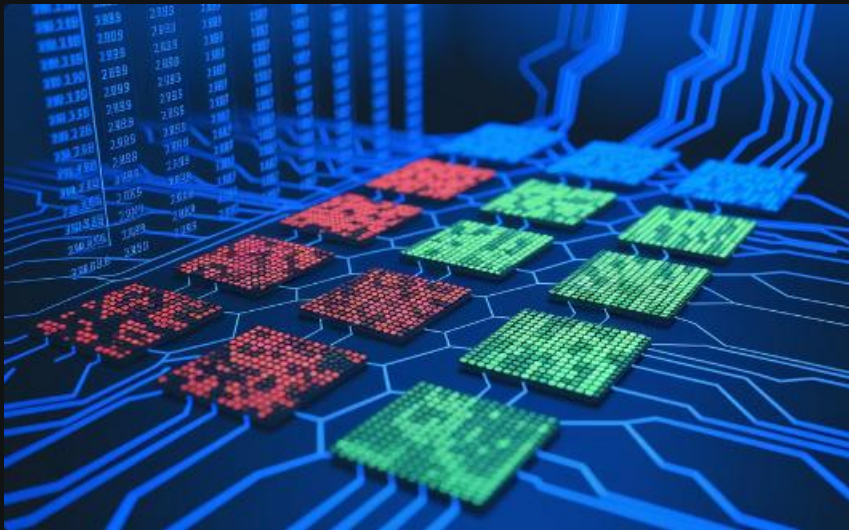
System-Wide Alignment

All modules and components should work with synchronized data to ensure smooth operations across the enterprise.



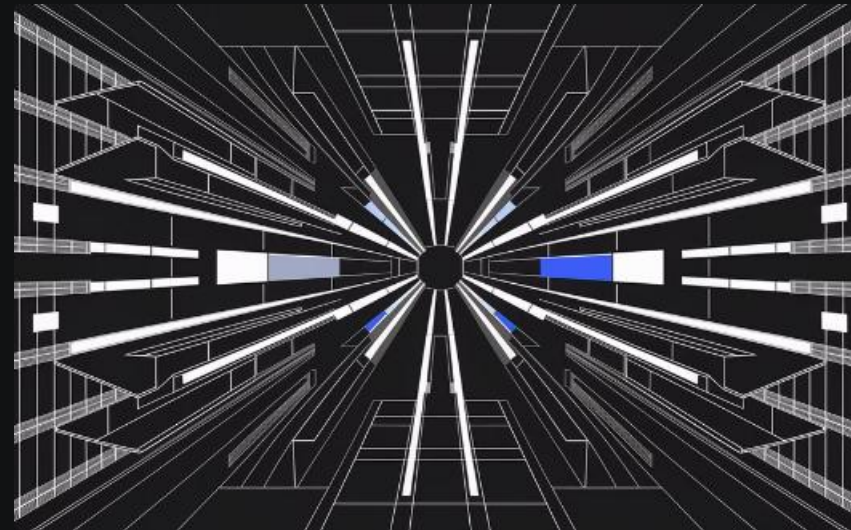
Data Uniqueness

Each piece of data should be **unique** - avoiding duplication that causes confusion.



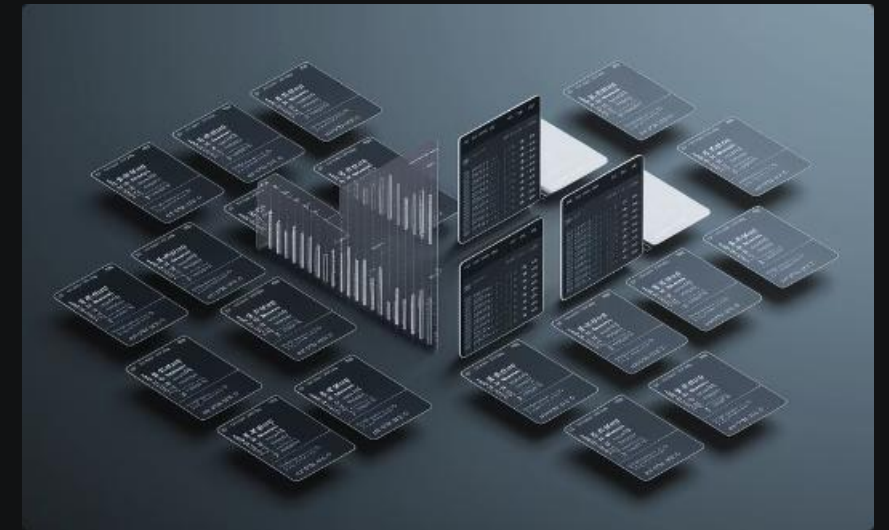
Duplicate Detection

Advanced algorithms identify potential duplicate records based on similarity thresholds and matching rules.



Unique Identifiers

Properly implemented primary keys and unique constraints prevent duplicate entries from being created.



Record Consolidation

When duplicates are found, merging processes combine them into a single "golden record" containing the most accurate information.

Data Validity

It must be **valid** - conforming to defined business rules and formats.

1

Input Validation

Enforcing proper formats and value ranges when data is first entered into the system prevents invalid information from contaminating your database.

2

Business Rule Compliance

Valid data adheres to established business rules and relationships, ensuring logical consistency within your processes and operations.

3

Format Standardization

Consistent formatting for addresses, phone numbers, dates, and other common data elements improves usability and integration capabilities.





Data Relevance

Your data should be **relevant** - regularly used and serving a business purpose.

1 Business Purpose

Every data element should have a clear connection to business processes or decision-making. Data without purpose creates unnecessary complexity.

2 Usage Monitoring

Tracking which data is actually being accessed helps identify unused information that might be candidates for archiving or removal.

3 Regular Review

Periodically evaluating data relevance ensures your system doesn't accumulate information that no longer serves current business needs.

4 Process Alignment

Relevant data directly supports your core business processes and strategic objectives rather than existing as isolated information.

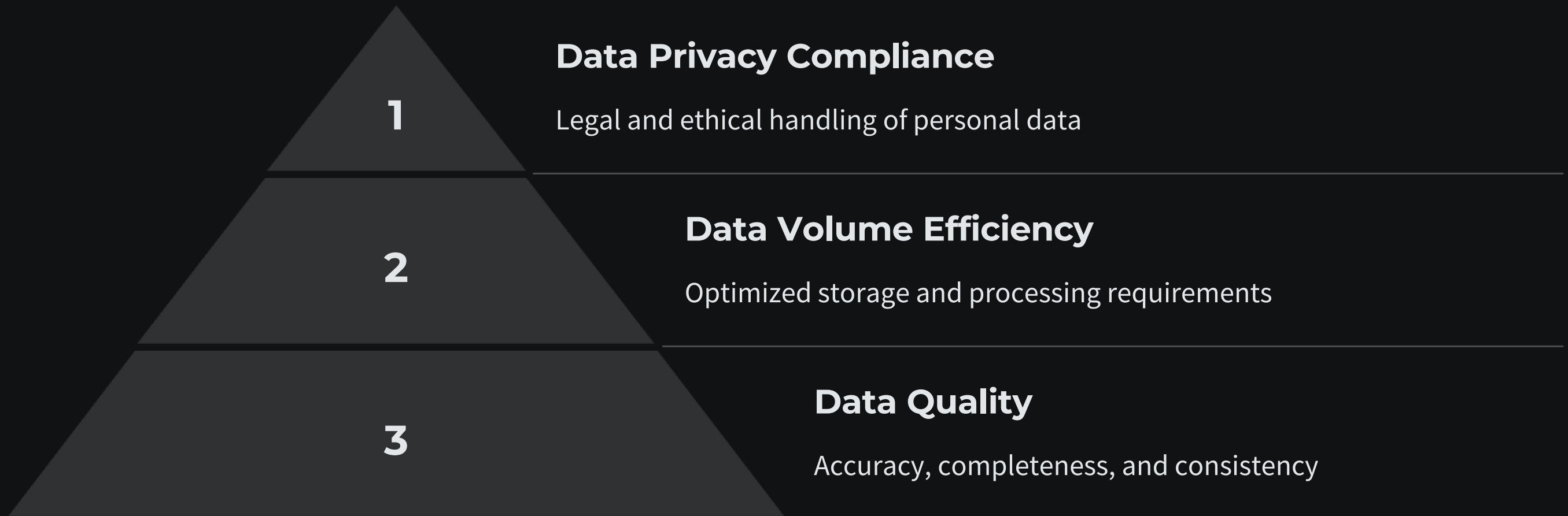
Optimized Storage and Lifecycle

And finally, you want **optimized disk consumption** and proper **data lifecycle management**. These aren't just nice-to-have qualities - they directly impact your system performance, user experience, and ultimately, your business decisions.



The Main Aspects of Clean Data

There are three main aspects we need to focus on when it comes to clean data:



Data Quality

This covers your configuration data, master data, and transactional data. Having high-quality data means your business can make better decisions, operate more efficiently, and provide better customer service.

For example, if your customer master data has duplicate entries or outdated contact information, you might send multiple mailings to the same customer or fail to reach them entirely. This wastes resources and damages customer relationships.

99.9%

Accuracy Target

The goal for critical master data elements in high-performing organizations

40%

Cost Reduction

Potential operational savings from improved data quality

28%

Revenue Impact

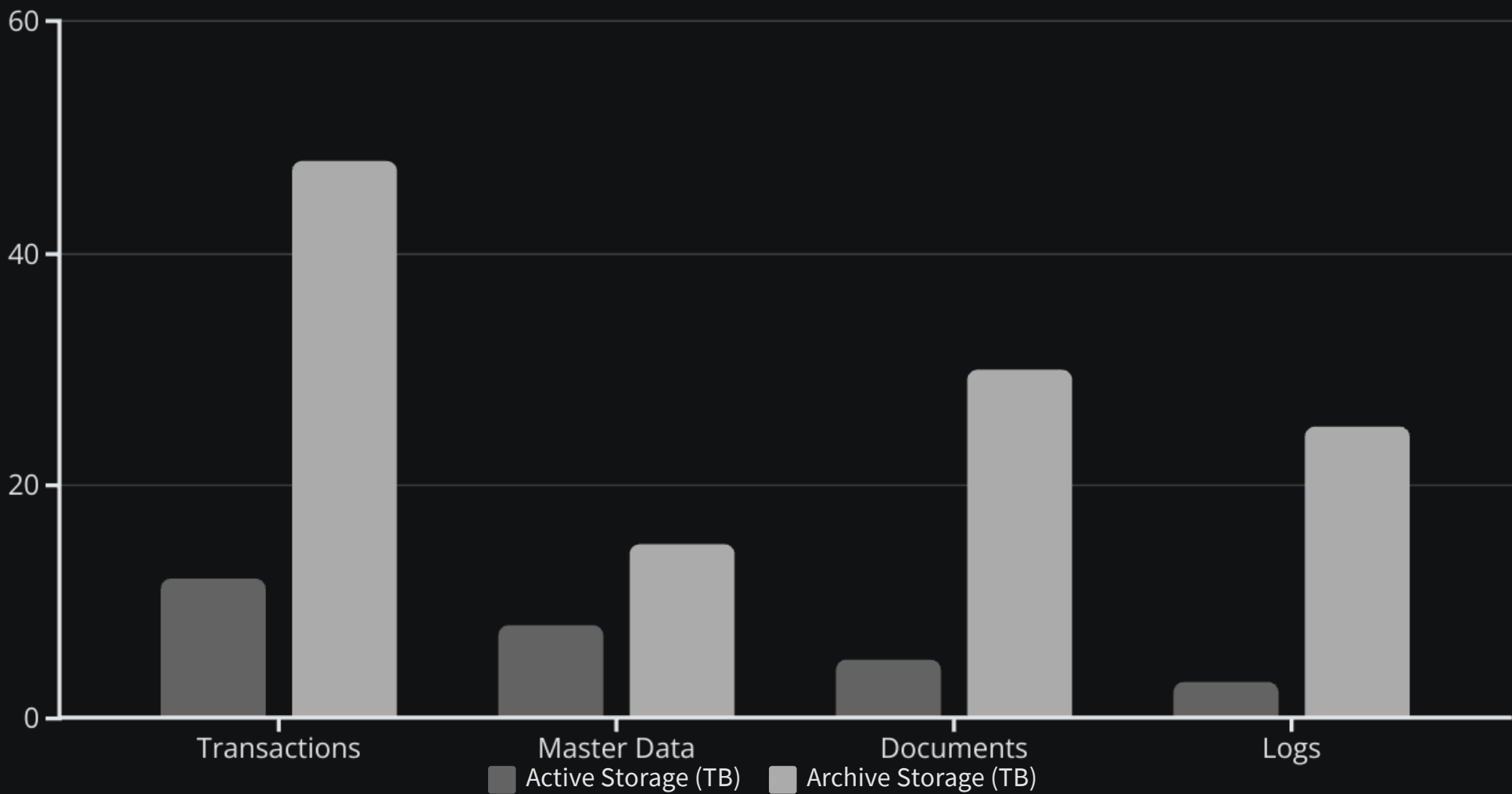
Average revenue increase from better customer data quality



Data Volume Efficiency

This specifically relates to your master and transactional data. Efficient data volume means your system runs faster, costs less to maintain, and is easier to manage.

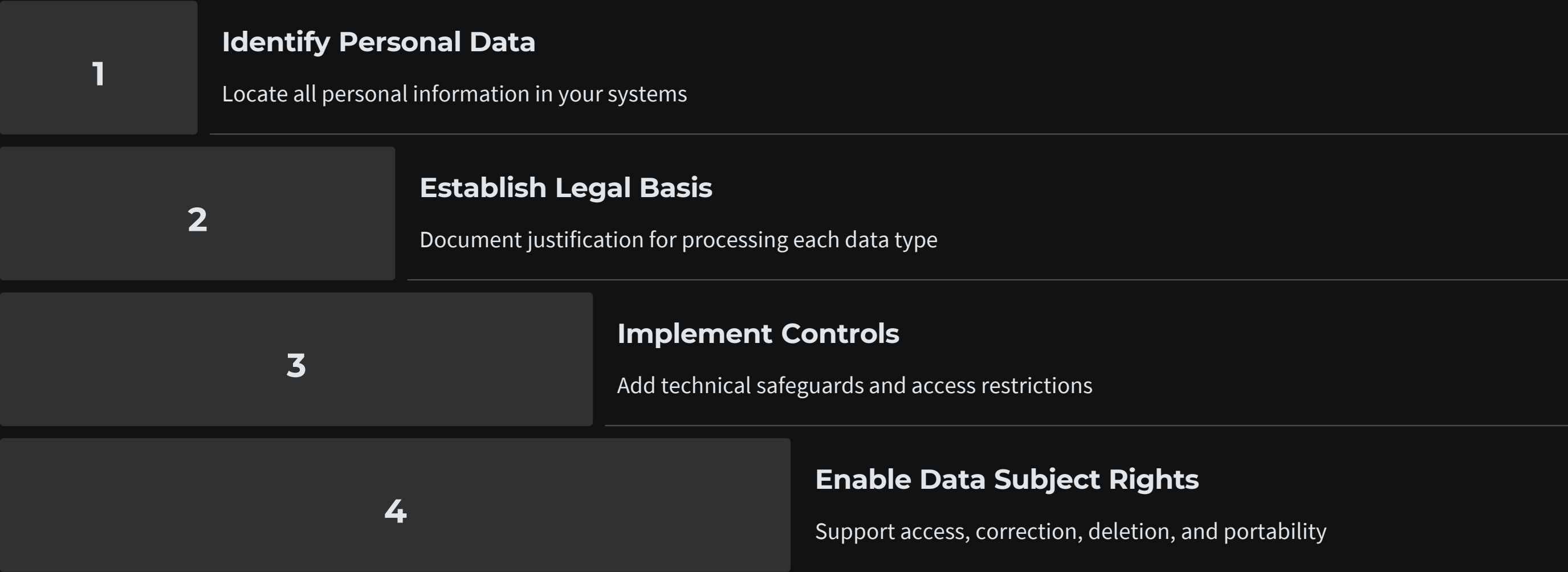
Imagine you're keeping years of detailed transaction logs that nobody ever accesses. This data consumes storage, slows down system backups, and can impact performance. Being efficient about what you keep and for how long makes a huge difference.



Data Privacy Compliance

This focuses on how you store and process personal master data. With regulations like GDPR, CCPA, and others around the world, properly managing personal data isn't optional – it's mandatory.

If you're storing personal data without a justifiable business purpose, you're not only violating clean core principles but potentially breaking the law.

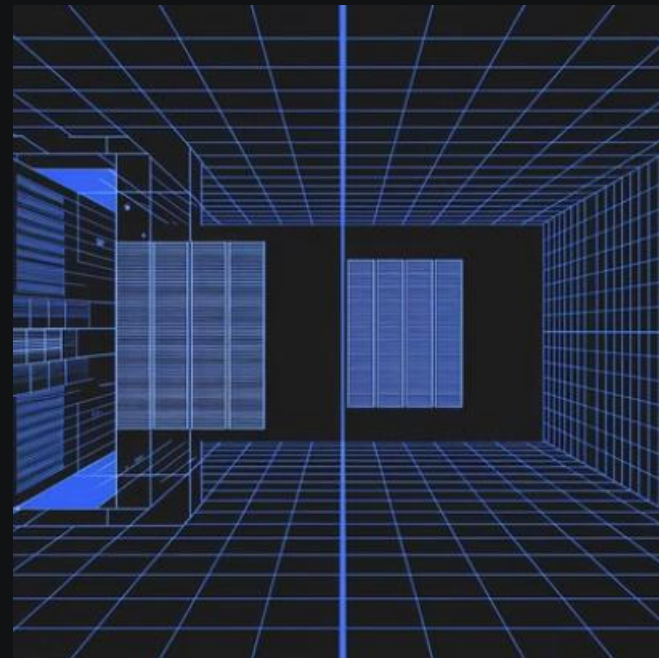


Data Quality Assurance Approaches

First, you need to **analyze and define measures**. This means identifying your critical data objects and establishing comprehensive quality measures that make sense for your organization.

Second, **use the right tools and expertise**. SAP provides tools like SAP Master Data Governance (SAP MDG) and SAP Information Steward. You can also work with third-party vendors like CDQ for specialized guidance in data cleansing.

Third, **implement cleansing processes**. This includes deduplication and generating best records, ideally supported by your tools. Finally, set up **continuous monitoring**. Don't just clean your data once and forget about it.





Data Volume Efficiency Approaches

First, establish **continuous analysis and monitoring** of your database utilization. This helps you identify outdated, unused, or redundant data that could be archived or deleted.

Next, define clear policies for **archiving and data tiering**. Solutions like SAP Information Lifecycle Management (SAP ILM) can help you move less frequently accessed data to cheaper storage tiers while maintaining accessibility when needed.

Analyze Usage Patterns

Identify which data is frequently accessed and which remains dormant

Define Retention Policies

Establish how long different data types should remain in active storage

Implement Archiving

Move historical data to cost-effective storage based on policies

Monitor and Optimize

Continuously review storage utilization and adjust policies as needed

Data Privacy Compliance Approaches

First, conduct thorough **analysis of data usage** to understand why you're collecting and processing personal master data. Every piece of personal information should have a clear business justification.

Second, **establish lifecycle policies** for personal data. Define how long you'll keep different types of personal information and when it should be deleted or anonymized. Again, SAP ILM can help enforce these policies.



Real-World Application: Data Quality

Imagine a retail company that had been operating for over two decades. Their enterprise system contained millions of customer records, many of which were duplicated or inactive for more than ten years. To address this, the company adopted a structured data management approach, focusing on three key areas:

Data Quality: They implemented an advanced master data governance solution to identify and merge duplicate customer records, standardize address formats, and validate contact information.



Duplicate Detection

Advanced algorithms identified over 200,000 duplicate customer records based on name, address, and contact information similarities.



Address Standardization

Automated tools corrected and standardized address formats across all customer records, ensuring consistent formatting and accurate delivery.



Contact Validation

Email and phone verification processes identified and corrected thousands of outdated or incorrect contact points, improving customer communication.

Real-World Application: Volume Efficiency

Data Volume Efficiency: An archiving strategy was introduced, moving inactive customer data and completed transactions older than seven years to archive storage, significantly reducing the active database size.

30%

Database Reduction

Overall decrease in active database size after implementing archiving strategy

7 years

Retention Period

Standard timeframe before moving completed transactions to archive storage

40%

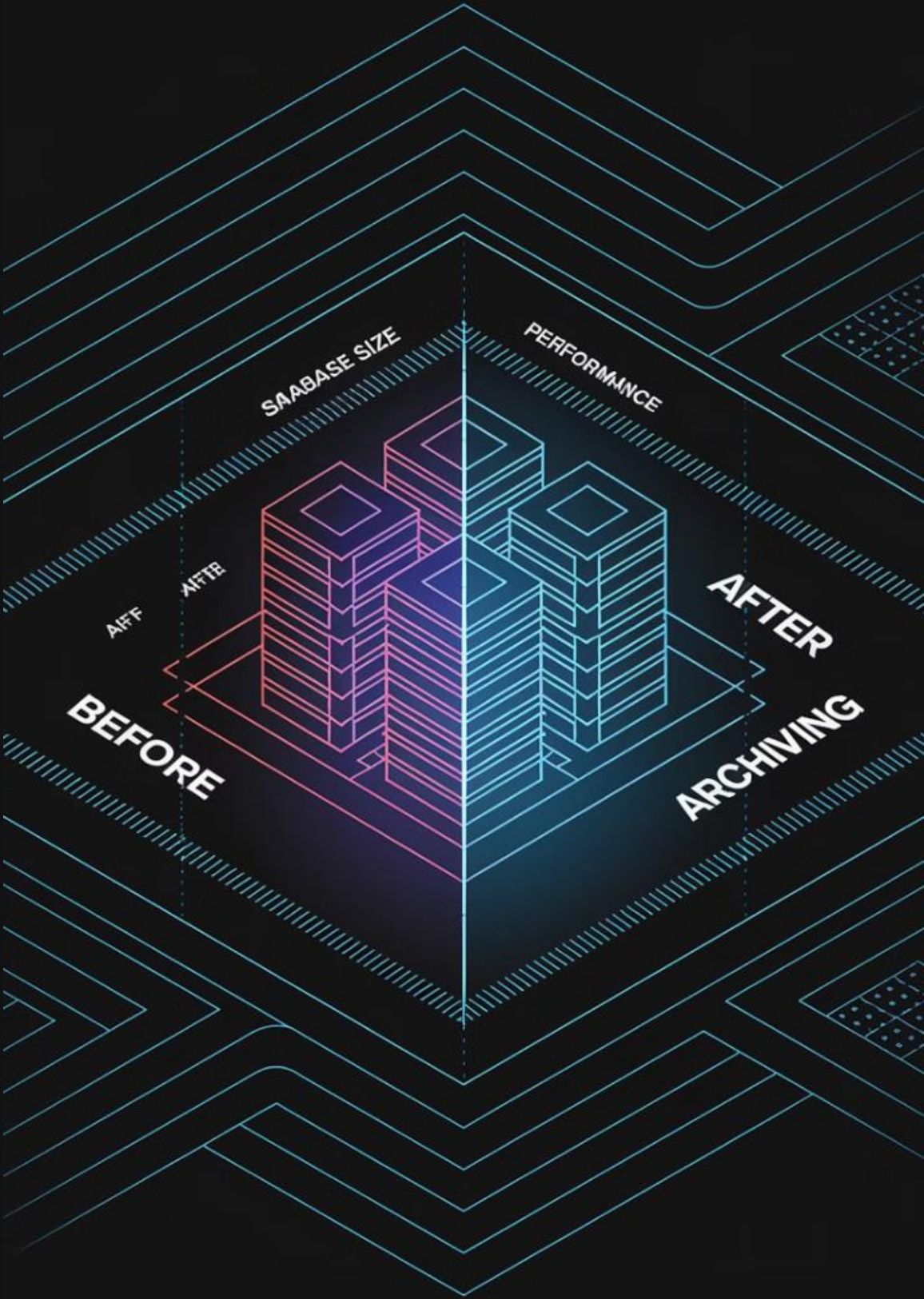
Performance Gain

Improvement in system performance after reducing active database size

60%

Storage Cost Savings

Reduction in storage costs by moving data to lower-cost archive tiers



Real-World Application: Privacy Compliance

Data Privacy Compliance: A full audit of personal data was conducted, retention periods were established based on legal and business requirements, and automated processes were put in place to anonymize customer data after the retention period expired.

The impact was significant: A 30% reduction in database size, a 40% improvement in system performance, full compliance with data privacy regulations, and more accurate reporting and analytics.

This structured approach not only optimized their data management but also enhanced overall business efficiency.

1

Personal Data Audit

Comprehensive inventory of all personal information across systems, identifying what was being collected and why.

2

Retention Policy Development

Creation of data-specific retention schedules based on legal requirements and business needs.

3

Automated Anonymization

Implementation of processes to automatically remove or obscure personal identifiers after retention periods expired.

4

Compliance Verification

Regular audits and reporting to ensure ongoing adherence to privacy regulations.

Key Takeaways for Clean Data

Remember that achieving clean data isn't a one-time project – it's an ongoing commitment. Here are the key points to remember:

Clean data is accurate, complete, consistent, timely, unique, valid, and relevant. Focus on all three aspects: data quality, volume efficiency, and privacy compliance. Use the right tools like SAP MDG, SAP Information Steward, and SAP ILM.

Establish ongoing monitoring and governance to maintain clean data over time. Proper data management directly impacts system performance, compliance, and business decision-making.



Ongoing Process

Data cleanliness requires continuous attention and maintenance, not just one-time projects.



Right Tools

Leverage specialized solutions like SAP MDG, Information Steward, and ILM to support your clean data initiatives.



Business Impact

Clean data directly influences system performance, regulatory compliance, and the quality of business decisions.



Governance Framework

Establish clear policies, ownership, and processes to maintain data quality over time.