

Assess Before **You Invest**

6 IT Investments you shouldn't make until you
understand your data

A practical guide to using data intelligence to right-size investments, reduce project risk, and avoid carrying unnecessary data into your next enterprise technology initiative.

Before you invest around your data, understand your data.

Your current data footprint should not automatically become the basis for your next technology investment. The better starting question is: What data actually deserves to be supported, and how?

THE DATA BLIND SPOT IN IT INVESTMENT

Your current data footprint shouldn't automatically become the basis for your next IT investment.

Organizations are investing in infrastructure modernization, cloud, AI, cybersecurity, resilience, and governance. But these initiatives share the same underlying dependency: data.

Without a clear view of how much data exists, how it is being used, where sensitive information lives, what must be retained, and what still has business value, IT can end up sizing and designing new technology around an inefficient or risky data estate.

THE BETTER STARTING QUESTION

What data actually deserves to be supported in your project – and how?

What a data assessment should help answer

1 How much data is active vs. inactive?	2 What is redundant, obsolete, trivial, stale, or otherwise low value?
3 What can be deleted, archived, or moved to a lower-cost tier?	4 Where does sensitive or regulated information reside?
5 Who has access to critical data?	6 What data is appropriate for AI, migration, backup, or modernization?

Where understanding the data first can change the investment

1

Storage Refresh

Right-size new capacity instead of recreating today's footprint.

2

Cloud Migration

Decide what deserves to move before paying to migrate and store it.

3

AI & GENAI

Identify data that is useful, safe, relevant, and appropriate for AI to consume.

4

Backup & Disaster Recovery

Prioritize what truly requires protection, replication, and recovery.

5

Security & Governance

Find the risk before deciding where to add controls and remediation.

6

Infrastructure Modernization

Design the future environment around the footprint you need - not the one you inherited.

01 | STORAGE REFRESH

Before you buy more storage, determine whether you actually need it.

- Active vs. inactive data
- Data that has not been accessed in years
- Candidates for deletion, archive, or lower-cost storage
- Sensitive data sitting on expensive production infrastructure
- Capacity that could potentially be reclaimed

Instead of asking	How much additional capacity do we need?
Ask first	How much capacity do we actually need <u>after optimizing the data we already have?</u>

Potential impact

Avoid unnecessary capacity purchases, right-size refreshes, extend existing infrastructure, and reduce ongoing storage costs.

- Reduce or defer capital expenditure by eliminating data that does not need to be carried into the new environment.
- Right-size premium NAS and high-performance tiers so expensive capacity is reserved for data that actually requires it.
- Extend the useful life of existing infrastructure by reclaiming capacity before adding hardware.
- Lower downstream backup, replication, DR, administration, power, and data-center costs because less unnecessary data remains in production.
- Create a defensible, data-driven basis for procurement and capacity planning rather than relying primarily on historical growth rates.

Cost & risk of not assessing first

Without assessing first, IT can size a refresh around an inflated footprint and lock unnecessary data into another multi-year hardware cycle. That can mean purchasing more Tier 1 capacity than the business actually needs, increasing backup and DR requirements, and continuing to pay to manage inactive or obsolete data. It also misses a rare opportunity to clean up the estate before data is copied into the next platform.

02 | CLOUD MIGRATION

Don't pay to migrate data that shouldn't make the trip.

- What is actively used
- What is inactive, obsolete, or low-value
- What should remain on-premises
- What can be archived or deleted
- What sensitive or regulated data requires additional controls
- How data should be prioritized in the destination

Instead of asking	How much will it cost to move all of our data?
Ask first	How much of that data should <u>actually move</u> ?

Potential impact

Reduce migration scope and cloud storage costs while avoiding the transfer of unnecessary risk into the new environment.

- Reduce the volume of data that must be migrated, shortening migration windows and lowering project complexity.
- Lower ongoing cloud storage, backup, egress, and management costs by preventing low-value data from becoming a recurring cloud expense.
- Determine which data belongs in high-performance cloud tiers versus archive, lower-cost tiers, or on-premises environments.
- Identify sensitive or regulated data before movement so security, residency, retention, and access requirements can be designed into the migration.
- Give migration teams a prioritized, actionable dataset instead of treating every file as equally valuable.

Cost & risk of not assessing first

A lift-and-shift approach can turn years of data sprawl into a recurring cloud bill. Unnecessary data may be migrated, replicated, indexed, scanned, and protected indefinitely. Sensitive or regulated information can also be moved into environments without the right controls. The result can be higher migration costs, higher ongoing cloud spend, greater compliance exposure, and a more complex environment to govern after the move.

Before you give AI access to enterprise data, know what's in it.

- Datasets relevant to the AI use case
- Outdated, low-value, or irrelevant content
- Sensitive and regulated information
- Data that should be excluded from AI
- Access or governance issues requiring remediation
- Repositories with the highest-value data for the initiative

Instead of asking	How do we connect our enterprise data to AI?
Ask first	Which data should AI be allowed to use – and which data will actually improve the outcome?

Potential impact

Reduce unnecessary AI processing, improve data relevance, reduce exposure, and establish a stronger foundation for AI governance.

- Focus AI pipelines and retrieval systems on data that is relevant to the intended use case instead of processing the entire enterprise estate.
- Reduce compute, indexing, vectorization, storage, and model-related costs associated with low-value or irrelevant content.
- Identify sensitive, regulated, confidential, or inappropriate data that should be restricted, quarantined, masked, or excluded from AI.
- Improve the quality of AI outputs by reducing stale, conflicting, redundant, and irrelevant source content.
- Establish an evidence-based foundation for AI governance by documenting what data exists and which data is approved for AI use.

Cost & risk of not assessing first

Connecting AI to poorly understood enterprise data can amplify existing data problems. Models and assistants may surface information users should not see, rely on outdated or irrelevant content, or process sensitive data that was never intended for AI. Organizations may also spend heavily indexing and processing data that contributes little value. The financial risk is wasted AI infrastructure and consumption; the business risk is inaccurate outputs, data leakage, governance failures, and delayed AI adoption after issues are discovered.

04 | BACKUP, DISASTER RECOVERY & CYBER RESILIENCE

Don't protect and recover everything equally.

- Business-critical data
- Inactive and low-value data
- Data that must be retained
- Data that could be defensibly deleted
- Sensitive-data concentrations
- Areas where the protected footprint could potentially be reduced

Instead of asking	How do we protect and recover all of our data?
Ask first	What data actually requires this level of protection and recovery?

Potential impact

Reduce backup and replication requirements, prioritize critical information, simplify recovery, and reduce the footprint exposed to cyber threats.

- Reduce the amount of data that must be backed up, replicated, stored, and recovered by removing unnecessary data from the protected footprint.
- Prioritize business-critical and sensitive datasets for stronger recovery objectives and protection controls.
- Potentially reduce backup storage, replication bandwidth, recovery infrastructure, and administrative overhead.
- Improve recovery planning by distinguishing data that must return quickly from data that can be archived, restored later, or eliminated.
- Reduce the volume of unnecessary data exposed during ransomware events and simplify the scope teams must validate during recovery.

Cost & risk of not assessing first

When every file is treated as equally important, backup and recovery environments grow along with the entire data estate. IT can end up paying to replicate and retain data that no longer has business value while recovery teams face a larger footprint during an incident. More data can mean longer backup windows, more storage and bandwidth, greater recovery complexity, and potentially slower restoration of the information the business actually needs.

Before you add controls, understand where the risk actually is.

- Sensitive and regulated information
- Repositories with the greatest concentrations of risk
- Access to critical data
- Data retained beyond its useful or required life
- Governance and policy gaps
- Unnecessary data increasing exposure

Instead of asking	What security or governance technology do we need?
Ask first	Where is our actual data risk – and what controls will reduce it?

Potential impact

Prioritize security investment, reduce unnecessary exposure, improve governance, and focus remediation where it can have the greatest impact.

- Prioritize remediation around repositories and datasets with the highest concentrations of sensitive or regulated information.
- Reduce the attack surface by eliminating unnecessary data and identifying information that is broadly accessible or retained without a business need.
- Improve the effectiveness of security and governance investments by applying controls based on actual data risk rather than repository assumptions.
- Support retention, disposition, privacy, and compliance programs with evidence about what data exists and where it resides.
- Help security, legal, privacy, and IT teams focus limited resources on the risks most likely to create material exposure.

Cost & risk of not assessing first

Organizations can spend heavily on security and governance technology while the underlying data remains unknown. Sensitive information may continue to sit in unexpected locations, unnecessary data can expand breach impact, and retention violations can persist unnoticed. Without data-level intelligence, teams may over-control low-risk repositories while missing higher-risk areas, increasing tool spend and remediation effort without proportionately reducing exposure.

06 | DATA CENTER & INFRASTRUCTURE MODERNIZATION

Don't build your next environment around yesterday's data footprint.

- True capacity requirements
- Data that can be eliminated
- Data that can be archived or tiered
- Workloads requiring high-performance infrastructure
- Sensitive data requiring specific controls
- Where data should live in the future-state architecture

Instead of asking	How do we upgrade our existing environment?
Ask first	What should our future environment actually be designed to support?

Potential impact

Right-size infrastructure, reduce unnecessary capital and operating expense, simplify modernization, and avoid carrying legacy data problems forward.

- Establish a more accurate future-state capacity requirement before hardware, hosting, or architecture decisions are finalized.
- Reduce the volume of data that must be moved, re-platformed, tested, protected, and supported during modernization.
- Match data to the appropriate performance and cost tier instead of reproducing legacy placement decisions.
- Lower the infrastructure, migration, backup, and operational costs associated with the future-state environment.
- Use the modernization event as an opportunity to eliminate accumulated data debt rather than transferring it into the next architecture.

Cost & risk of not assessing first

Modernization can become an expensive way to preserve yesterday's inefficiencies. If IT designs the new environment around the existing footprint, unnecessary data can drive capacity requirements, migration effort, backup architecture, and long-term operating costs. Once the data is moved, the organization may be locked into another cycle of supporting low-value content on newly purchased infrastructure, making future optimization harder and more expensive.

Five questions IT leaders should ask before the next major investment

1**Do we know how much of our current data is actually active and valuable?****2****Do we know how much could be deleted, archived, or moved to a lower-cost tier?****3****Do we know where sensitive and regulated data resides?****4****Do we know which data truly needs to be migrated, protected, processed, or made available to AI?****5****Are we sizing this investment around what we truly need – or simply around what we have today?**

If you can't confidently answer these questions, assess before you invest.

Understand your data before your next major IT investment.

Congruity360 helps organizations establish a clear, actionable understanding of their unstructured data estate before making costly infrastructure, cloud, security, governance, resilience, and AI decisions.

>> REQUEST A DATA ASSESSMENT <<