

VaultIQ

Storage-Native AI Memory for Exascale Systems

Krishna Maheshwari

Krishna@vaultiq.ai

MSST 2026

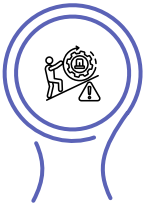
The exascale AI challenge



How do we find relevant artifacts across billions of files?

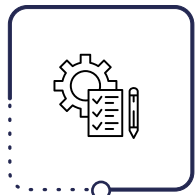


How do we reconstruct scientific workflows from distributed outputs?



How do we do both without breaking existing security boundaries?

Interrogation of scientific repositories



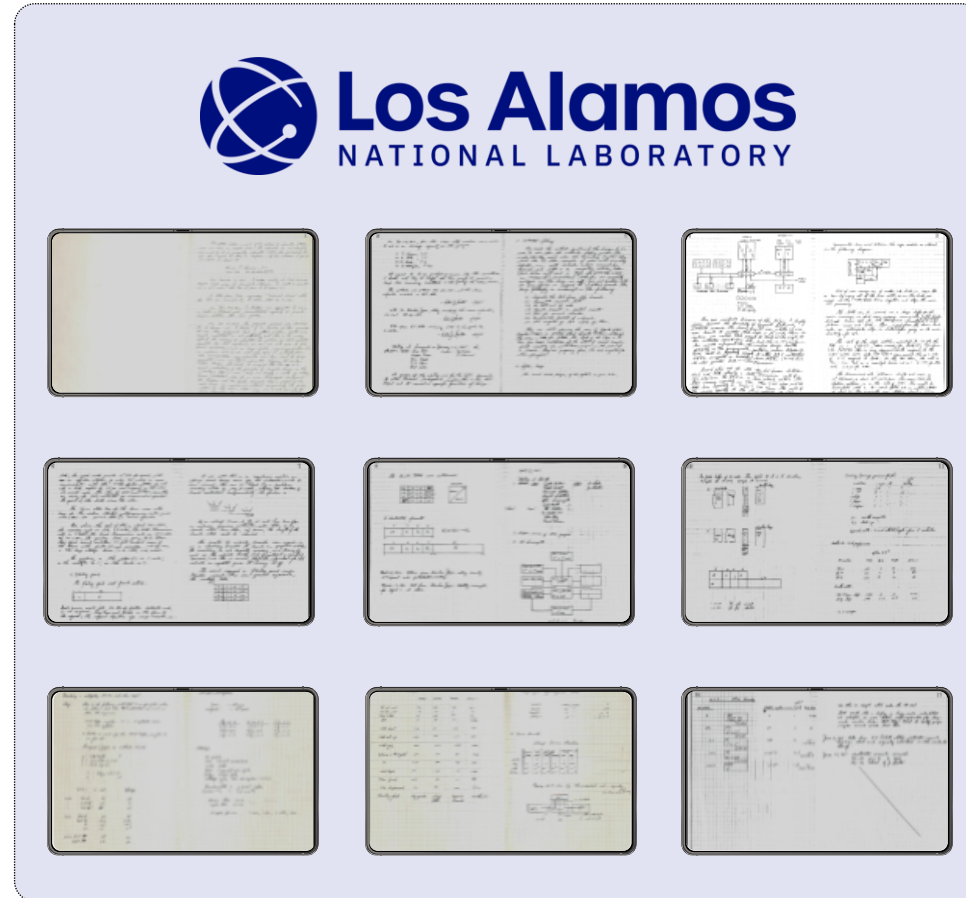
Requirement to limit data & metadata access to cleared personnel only



Exabytes of data resides on-prem



Untapped research still physically held (e.g., 100 years of lab notebooks)



“

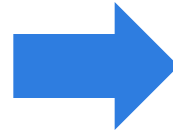
For the first time, we can ask questions of a century of notebooks—safely. VaultIQ promises to preserve our security posture and accelerates research

— Gary Grider, Sr. Director for Computing Technology,
Los Alamos National Laboratory

Enterprise scale assumptions break at exascale

Common AI architectures

- Data can be copied into an AI platform
- Retrieval operates over a centralized corpus
- Security and compliance are application concerns
- Governance relies on manual classification and tagging
- Vector similarity alone is sufficient



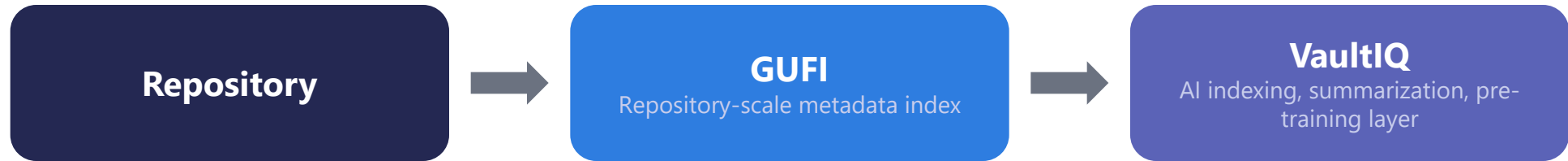
Exascale reality

- Data often cannot leave its storage boundary
- Billions of files span multiple repositories, storage media and permission domains
- Security must inherit from existing storage controls
- Manual governance and data classification do not scale
- Vector similarity alone is insufficient

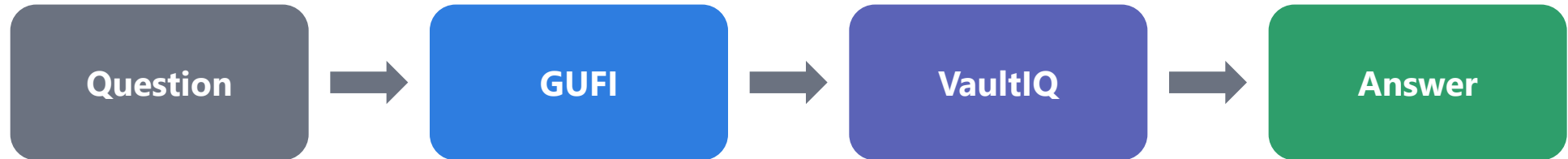
At exascale, security, compliance, and data governance must be intrinsic architectural properties

Repository intelligence meets AI

Indexing phase



Query phase



Indexing and querying are completely disjoint

VaultIQ transforms repository intelligence into permission-aware knowledge access

Why repository intelligence matters

Repository-scale indexing

Billions of files across multiple repositories without moving or copying the data

Permission-aware retrieval

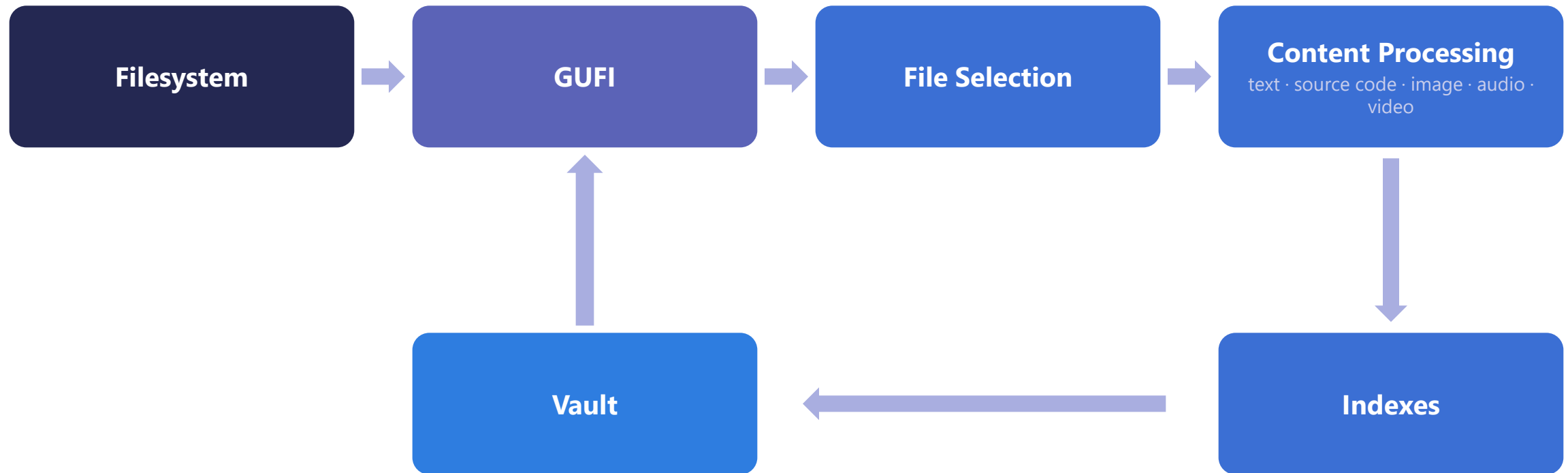
Every query runs as the asking user, so honoring existing POSIX permissions

Metadata-driven retrieval

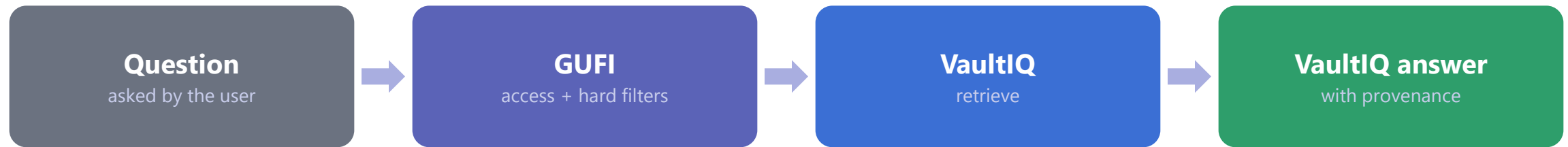
Narrows billions of artifacts to the relevant set before semantic retrieval begins

Repository-scale metadata is the foundation for exascale AI retrieval.

Building repository intelligence with GUFI



Permission-aware retrieval

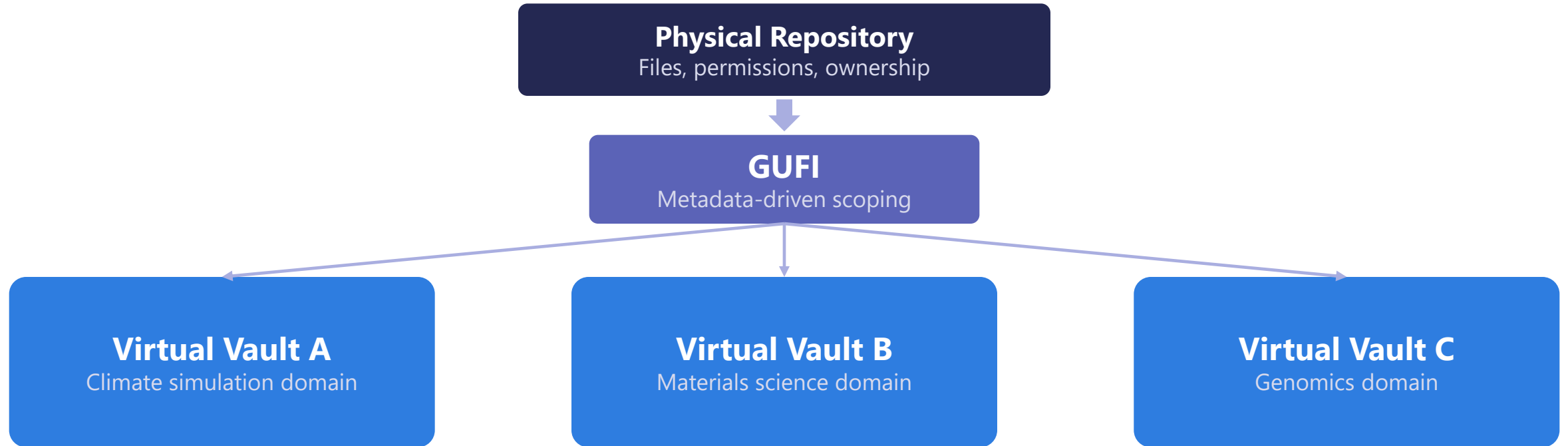


AUTHORIZATION REMAINS OUTSIDE THE AI

- GUFI executes as the requesting user.
- POSIX permissions are enforced before retrieval.
- VaultIQ never sees unauthorized content.

Virtual Vaults: one repository, many scoped views

Virtual Vaults are metadata-defined views over a shared repository.



Security boundaries

Each vault enforces its files' storage permissions — nothing leaks across.

Delegation

Owners grant scoped access without exposing the full repository.

AI isolation

Models and indexes stay separated by design, per vault.

Domain specialization

Vaults map to domains for focused, expert AI behavior.

Early experience: lessons learned

Design principles that emerged from deployment

1

Security must remain outside the AI

Authorization belongs to the storage and OS layer — never delegated to the model.

2

Metadata filtering must come before semantic retrieval

At repository scale, narrowing by metadata first is essential — semantic search alone cannot scan billions of files.

3

Permission boundaries are only the first filter

Repository structure, metadata, and provenance further define retrieval scope.

Key takeaways

-  **Repository-scale intelligence is a prerequisite for exascale AI.**
-  **Security must remain outside the AI.**
-  **Virtual Vaults enable pruning prior to retrieval.**
-  **AI can unlock scientific knowledge trapped in repositories.**

Questions?