

vinchin

破解海量文件 迁移难、备份慢难题

主讲人：吴瑶



目 录

PART 01 非结构化数据保护背景与挑战

PART 02 解决方案及核心优势

PART 03 应用场景

PART 01

非结构化数据保护 背景与挑战



非结构化数据快速增长带来的数据保护难题



IDC 预测，到2028年全球数据量将增长至**393.8ZB**，相比于2018年增长**9.8**倍。
中国区数据量将达到接近**100ZB**。

非结构化数据分布特点

vinchin



非结构化数据存储分布

海量小文件特点

- **文件小**：单个文件大小为几KB到几MB之间
- **文件数量多**：达千万级、数亿级、数十亿级
- **数据总量大**：达数百TB、PB
- **层次深**：一般数层、数十层

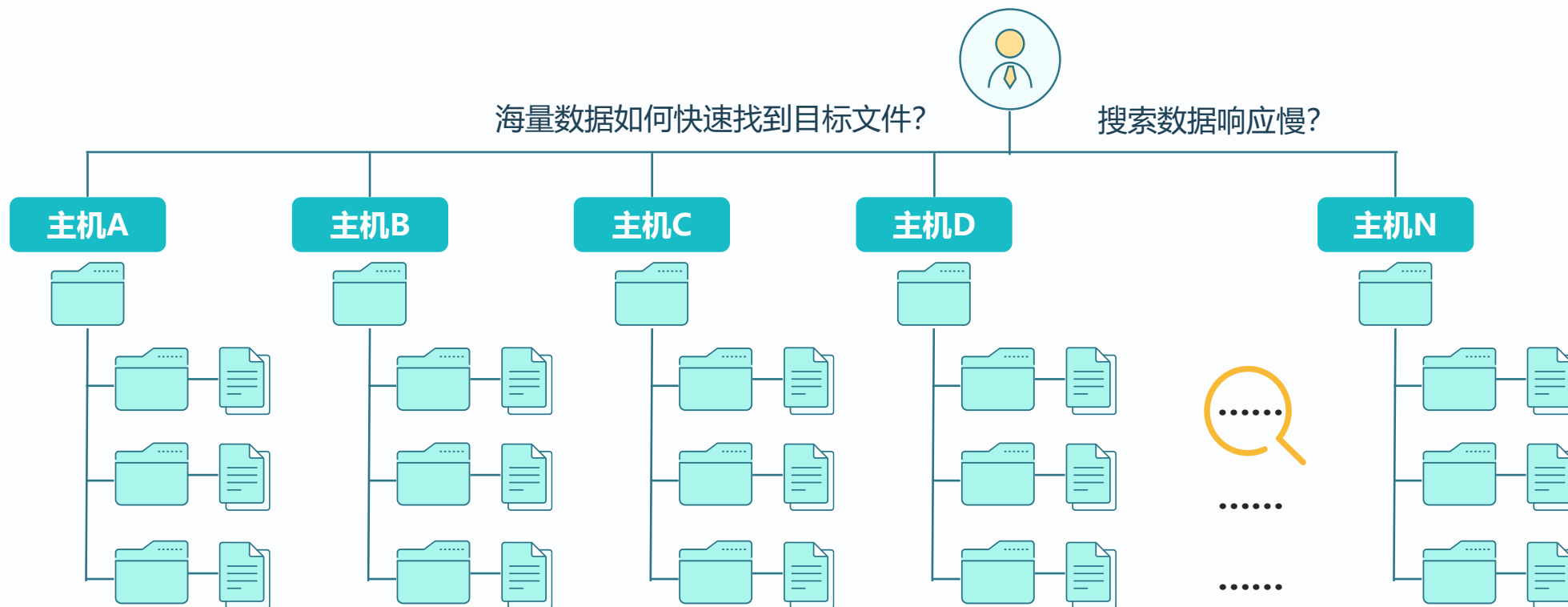
海量小文件为什么备份慢？

文件数据量多，且为小文件

文件备份是打开文件，读取文件，再关闭文件的过程，在海量小文件场景中，打开关闭的过程耗时长，频繁打开关闭导致扫描缓慢。

文件目录层级深且分布散乱

层级目录的深度及目录结构复杂性，会增加扫描的复杂性，导致扫描和传输效率低下。



目标文件匹配效率低

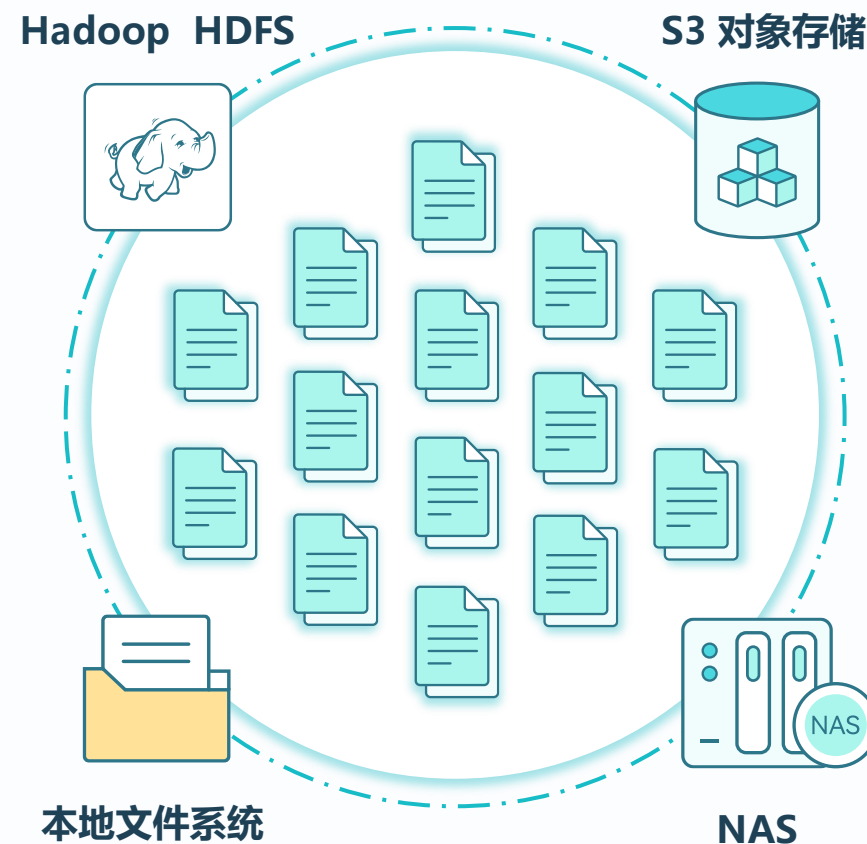
海量非结构数据中，因目录结构复杂和层级深等原因，定位目标文件时间久，整体效率低下。

恢复文件查找慢

备份海量数据之后，如需单独恢复某个小文件，对其查找等待时间久，极大降低恢复效率。

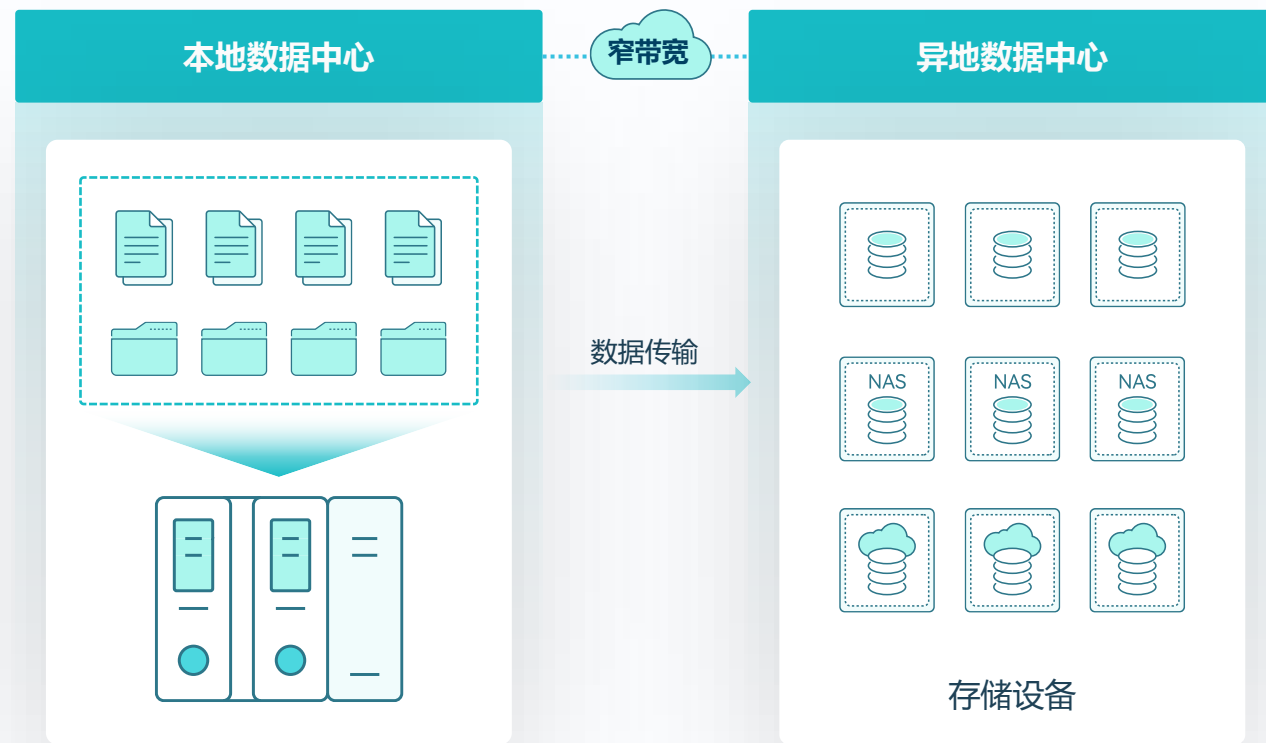
数据异构迁移流动性差

- **异构迁移困难**：非结构化数据分散存储在各类平台中，如NAS，本地文件系统，对象存储等。在后续设备变动或数据整合时，无法平滑的实现数据迁移。
- **迁移时间长**：非结构化数据迁移整体过程因数据量大小，网络波动等原因，造成的业务迁移时间不定。



窄带宽场景下无法满足备份时间窗口要求

窄带宽通常是指网络连接的带宽较低，即网络传输速率较慢。这种带宽通常不足以支持高速数据传输和高带宽需求的应用。



备份周期长

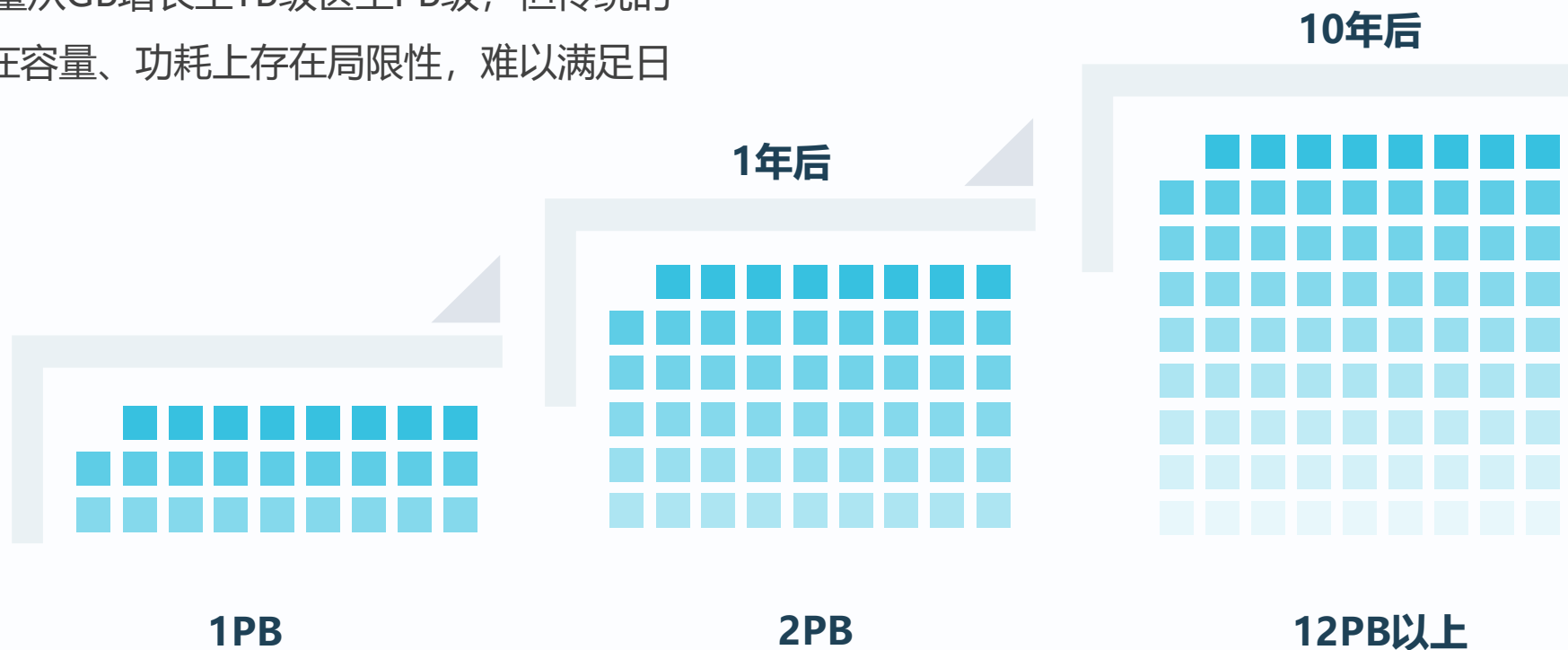
带宽有限的情况下，且数据量大时，执行备份任务，数据传输时间延长，执行完全备份时间长。

例：备份近**100TB**的数据量，带宽为**100Mbps**左右，最大速度约为12.5MB/s，整体备份时间约需**97天**。

数据增长带来的存储成本急剧上升

- **存储成本增加**：随着数据量的逐年激增，企业面临着数据规模达到数百TB甚至PB级别的挑战，这导致了对存储资源的急剧增长需求。

- **传统存储局限**：数据量从GB增长至TB级甚至PB级，但传统的存储介质如HDD和SSD在容量、功耗上存在局限性，难以满足日益增长的需求。



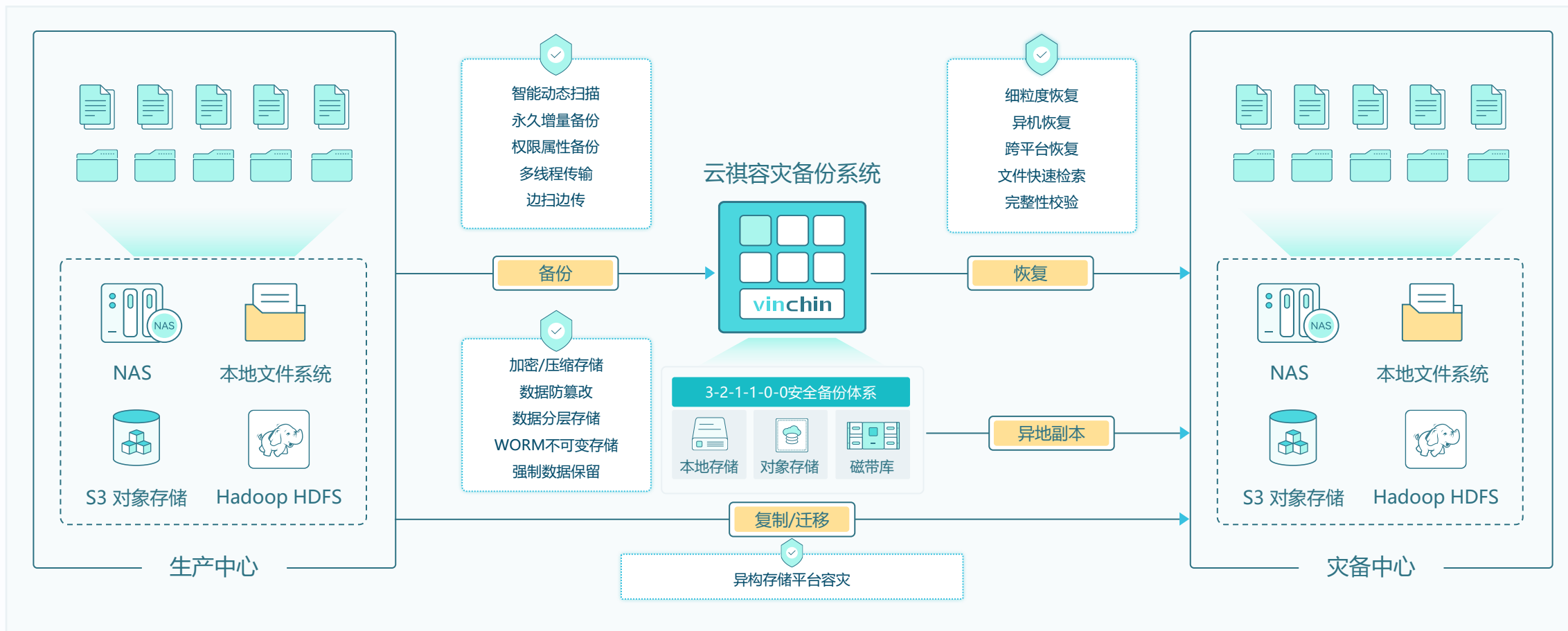
PART 02

解决方案及核心优势



非结构化数据保护整体解决方案

vinchin



高性能备份恢复

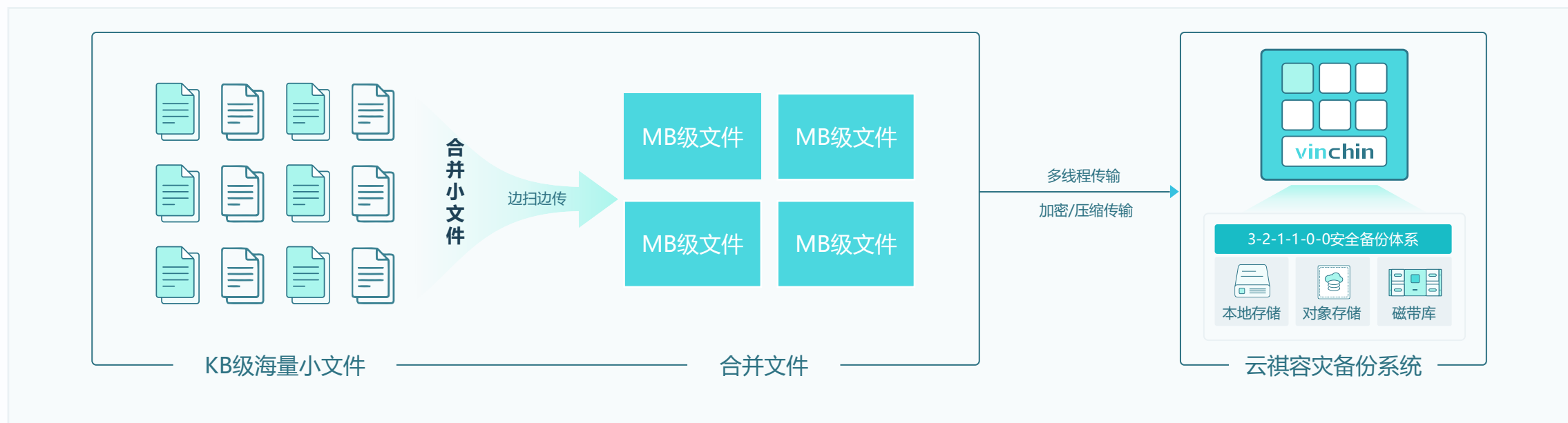
异构平台迁移

降低存储成本

保障业务连续性

高效处理小文件提升任务效率

vinchin



智能动态扫描

深广度自适应扫描引擎可根据生产环境目录结构智能调整扫描策略，结合多线程扫描和边扫边传实现海量文件的高速扫描。

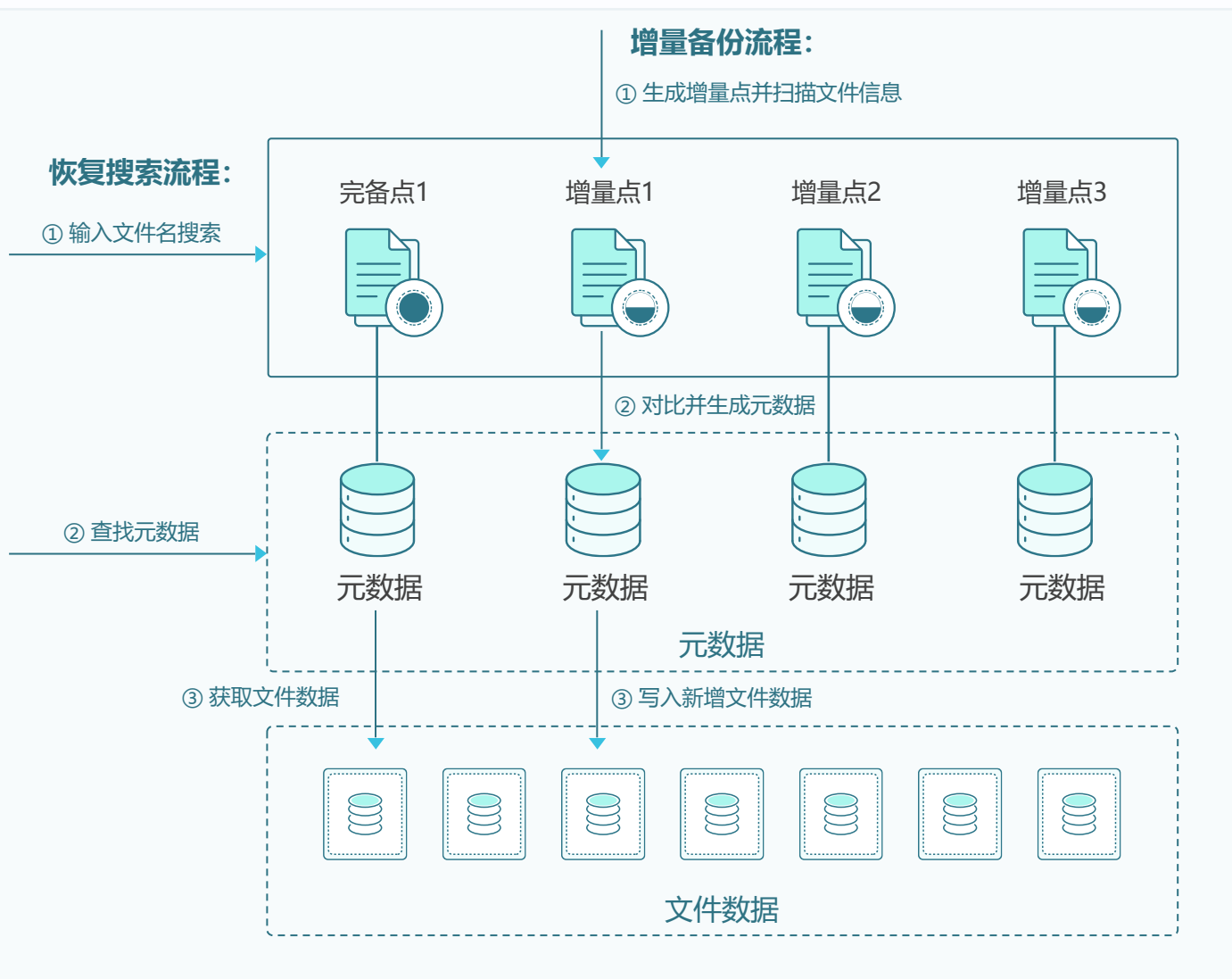
多线程并发处理

支持单任务配置多线程，可同时结合多任务并发、多节点并发、边扫边传等大幅提升备份效率。

快速增量检测

通过云祺独创专利技术，构建元数据索引系统，通过将文件指纹与备份集对比实现增量数据快速识别，提升增量备份效率。

元数据索引系统提升响应速度



快速增量检测

通过云祺专利技术，构建独创的元数据索引系统，扫描对比上一时间点的元数据来快速识别增量数据。

恢复快速检索

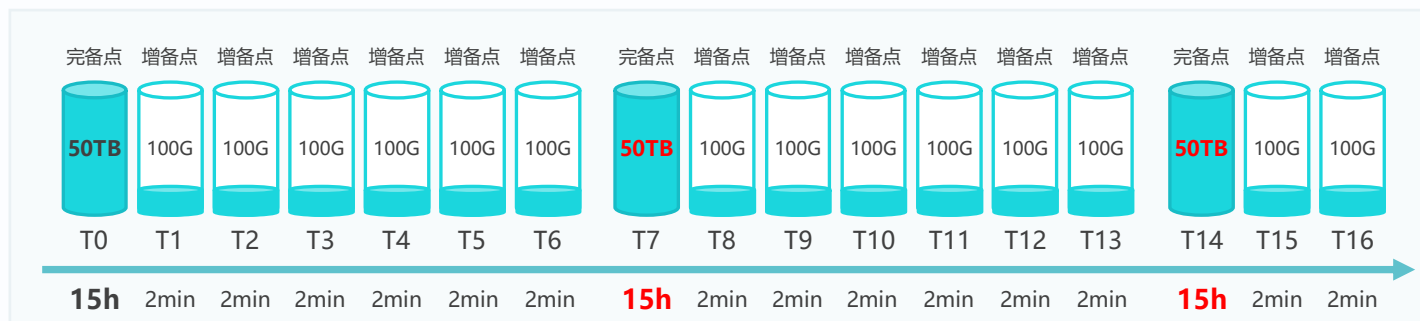
支持在恢复时，结合元数据索引系统，在已勾选恢复目录下进行关键词搜索，亿级别文件最快可在秒级内检索出目标文件。

文件永久增量减少时间窗口

总数据量：151.37TB

总时间：45.47H

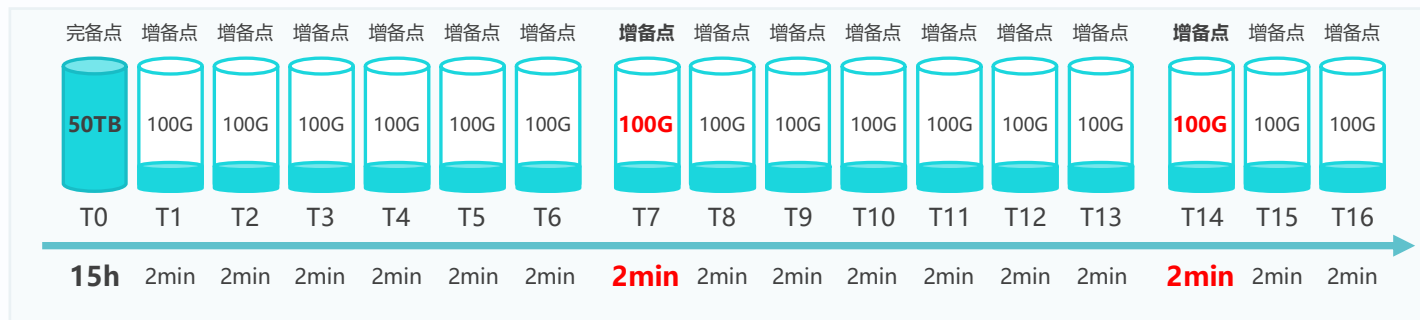
传统备份



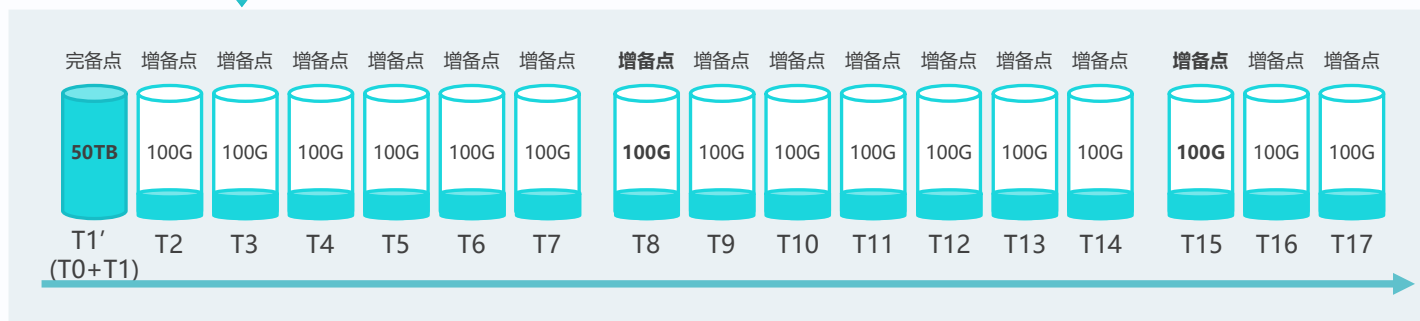
总数据量：51.56TB

总时间：15.53H

永久增量



保留17个备份点，第18次备份时触发保留策略，T0和T1数据合并



总数据大小：传统备份≈2.94永久增备

总时间花费：传统备份≈2.93永久增备

永久增量备份：只执行一次完备，后续每次备份过程中只对新增的数据进行备份，减轻数据传输负荷，减少备份时间和备份存储介质的空间占用。

例：若采用万兆网完备数据量大小50TB，备份时间窗口为15H，采用永久增量备份方式除第一次备份后，增量备份仅需要2min，T7时间点备份总窗口较传统方式快450倍；17个备份点的备份总窗口较传统方式快3倍，数据传输总量减少66%，备份更高效、成本更节省



高效数据传输

结合压缩传输提升传输效率，智能增量对比获取变化数据并同步，减少数据传输量。

异构平台复制

支持不同设备之间进行文件数据同步，并迁移基本权限。

数据一致性保障

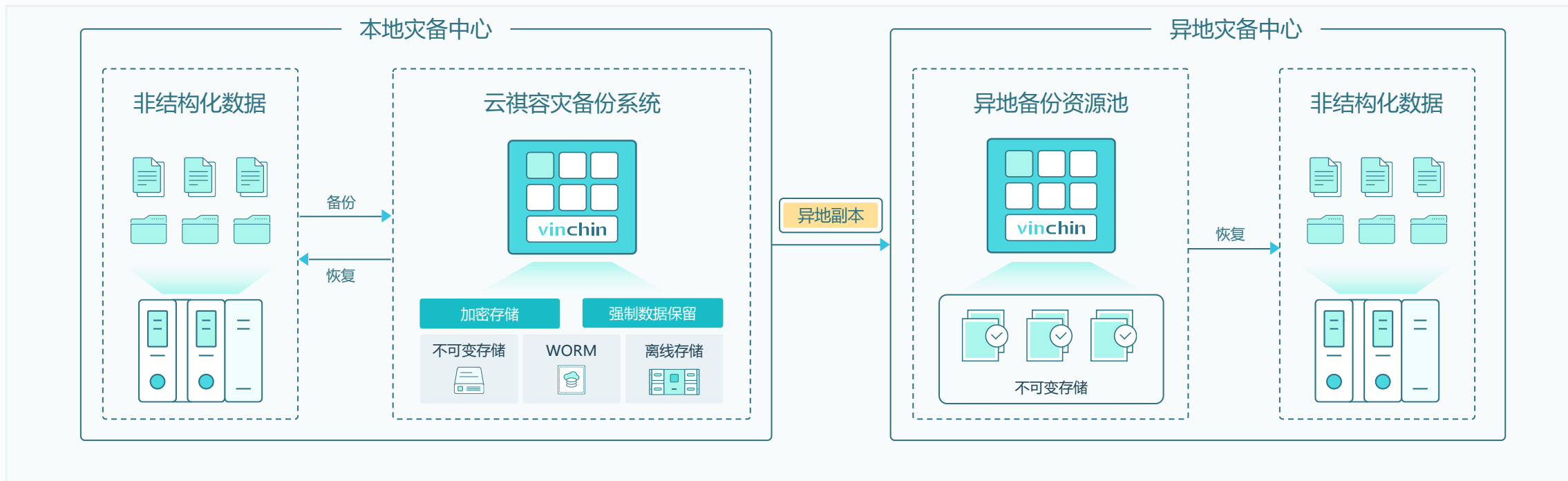
对比文件校验值来检测源端和目标端的差异性，并提供对比差异结果。

PART 03

应用场景



场景一：数据本地备份及异地容灾



场景描述

业务系统运行在本地并对数据进行本地保护，同时建立异地灾备中心已近乎为基础要求。以避免自然灾害导致的数据丢失，保证在生产数据和本地备份数据同时丢失的情况下，仍然可以从异地备份中快速恢复，启动业务运行。

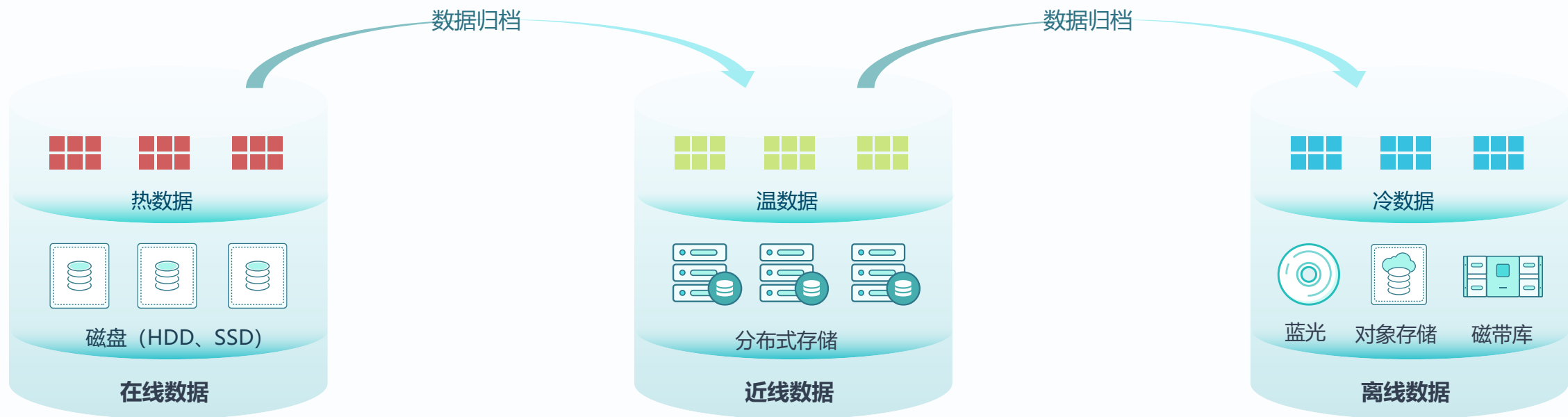
场景优势

海量数据高性能：通过小文件合并打包传输、边扫边传、多线程并行处理等多种技术处理海量小文件，极大的提升了备份效率。

传输安全保障：支持SM2和RSA的链路对副本数据加密，确保公网数据传输安全。

数据保护自动化：通过任务编排方式，确保备份完成后自动触发副本任务，实现备份-副本一体流程化操作，极大提高任务效率。

场景二：数据分层存储



场景描述

根据数据访问频次，重要性等方面采取不同的存储方式。如将不常用且数据量大的数据归档至磁带等更为经济的存储中，达到释放生产存储空间和降低存储成本目的。

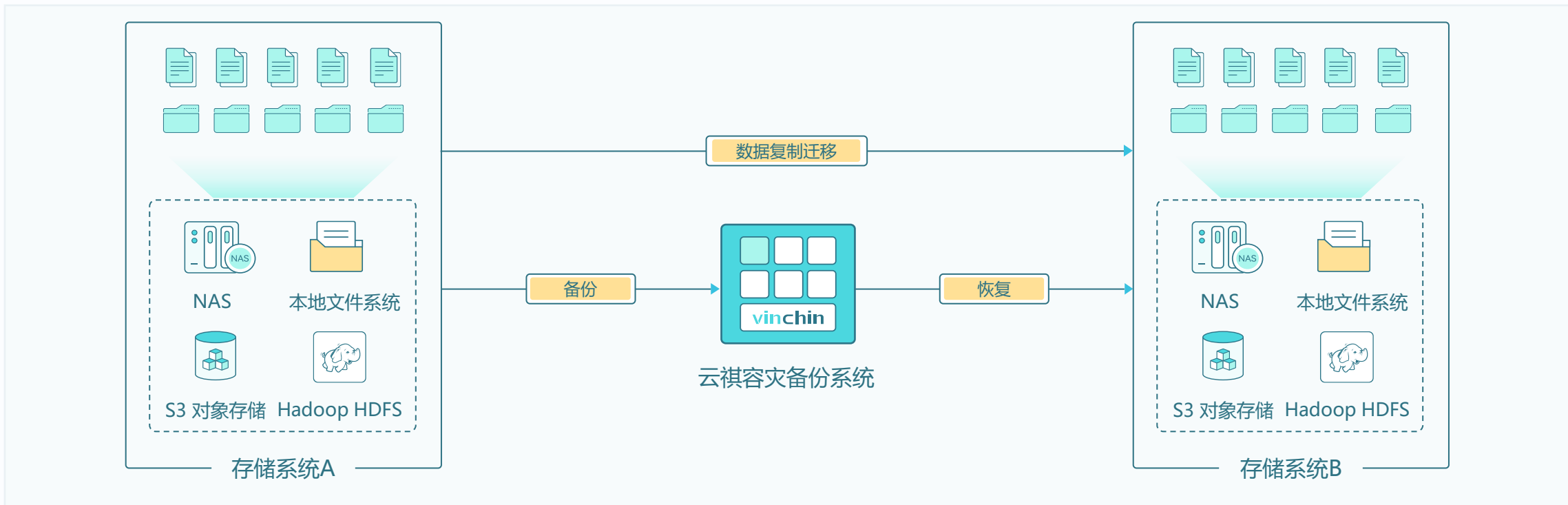
场景优势

节约成本：归档数据将减少对生产存储的占用，支持将数据放置在磁带、对象存储等介质。

数据长期保留：节约本地存储资源的同时满足行业法规等要求，实现数据的长期保留。

存储介质兼容性广：兼容主流公有云平台存储、对象存储及磁带，实现对业务数据归档上云或归档到磁带的保护，满足业务数据保护需求。

场景三：异构平台数据迁移



场景描述

业务系统运行在各种存储平台之中，若有业务调整或数据整合要求，需要将业务数据从现有存储平台迁移至新的数据存储平台之中。

场景优势

异构平台广泛兼容：支持本地文件系统、NAS设备、对象存储、Hadoop HDFS、分布式存储互相跨平台恢复，满足不同存储平台之间的文件数据迁移需求。

文件验证对比：对比源端和目标端的文件校验值，并提供对比差异结果，有效防止传输过程中可能出现的数据丢失或损坏，确保迁移的高可靠性。

细粒度迁移：可根据需求自主对文件进行排除或选定并提供基于时间点的筛选，同时可迁移基本读写权限至异构平台。

场景四：业务数据容灾



场景描述

存储平台承载着各种关键业务，对企业正常运行至关重要，需要定期复制业务数据到灾备中心，以避免业务系统发生故障时，实现快速容灾。

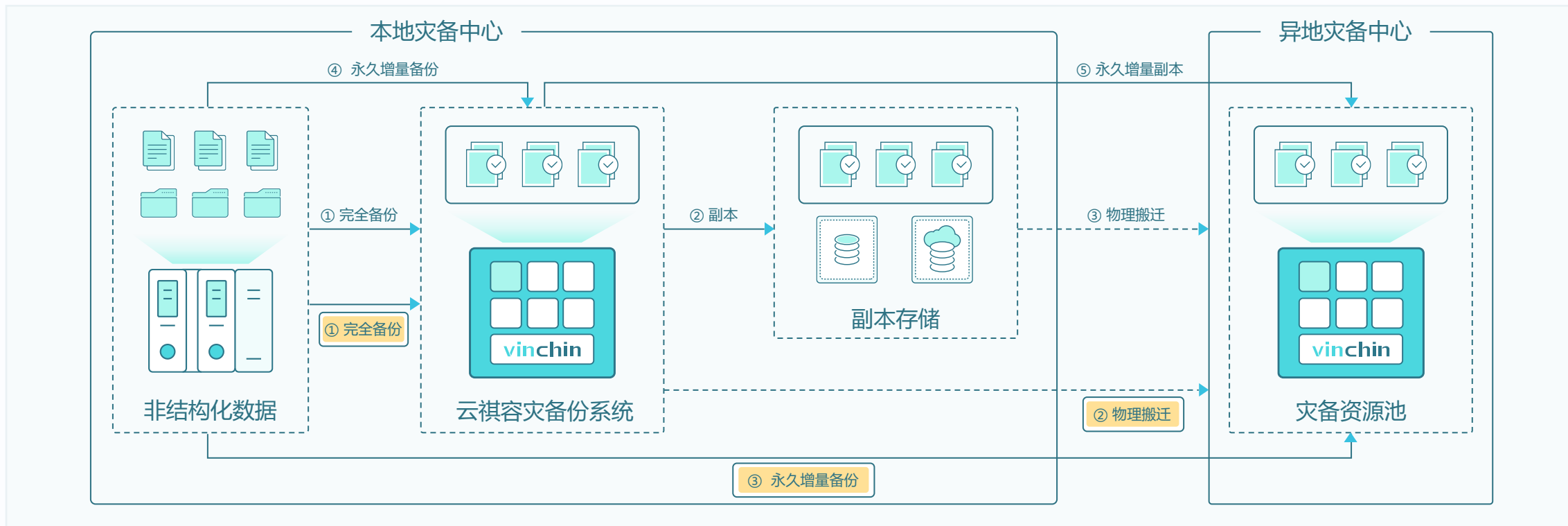
场景优势

数据持续保护：通过智能增量对比获取变化数据并同步至备机，确保灾备端同步更新，最大限度地减少数据丢失。

一致性保障：对比源端和目标端的文件校验值，并提供对比差异结果，有效防止了传输过程中的数据丢失或损坏，提升一致性保障。

数据安全保障：业务数据传输过程采用加密算法，全面保障数据安全，避免数据泄露。

场景五：窄带宽场景下的异地容灾



场景描述

随着非结构化数据量的激增，多地之间的带宽往往有限，难以支撑大规模数据的灾备需求。为了确保业务连续性，异地灾备需要在有限的带宽条件下，实现高效且安全的数据传输至远程灾备站点。

场景优势

异地传输效率高：针对异地带宽有限环境，通过源端压缩和4个级别的压缩算法，并结合永久增量技术，减少数据传输，提高传输效率。

传输安全保障：支持国密SM2和RSA的链路对副本数据加密，确保公网数据传输安全。

提升传输可靠性：支持在备份数据传输过程中，网络中断后从已处理的备份点继续进行传输。

vinchin

THANKS



云祺公众号



云祺视频号

