NOVASTOR

PetaByte Challenge

Backing up a Billion files to the cloud –
and restoring the data

Where did the admin go?

He ran som ware.

NOVASTOR

Reasons why | We can’t access our data!

01

False security

We don’t need backup, we have high-availability/redundancy

02

False promises

Our data is in the cloud so we cannot loose it

03

Having faith

We have a backup …

but no one ever did a restore test

04

Missing knowledge

I didn’t know this data is (also) important

05

Over confidence

Our employees especially IT are well trained, everything is under control

NOVASTOR

Petabyte - Billion files | Data sources

DMU visualization data

ERP databases **AI vector DBs**

IoT data **CAD/CAM construction data**

User/ Tracking data

Test readings i.e. from automotive driving from thousands of sensors and thousands of miles of test driving

Quotes, Orders, Receipts

Gigabye – Petabyte | What's the difference?

Gigabyte

- Volume
- Identifying relevant data
- Classify importance of the data / impact on loss
- Assigning resources for transport and storage
- Granularity of restorability
- Defining RPO (How much can you loose?) and RTO (How fast you need to be back online?)
- Bandwidth limitations (i.e. IONOS PublicCloud 700Mbit/s max.)
- Cost of storage and data retrieval (1ct/GB - 2.5ct/GB)

Petabyte

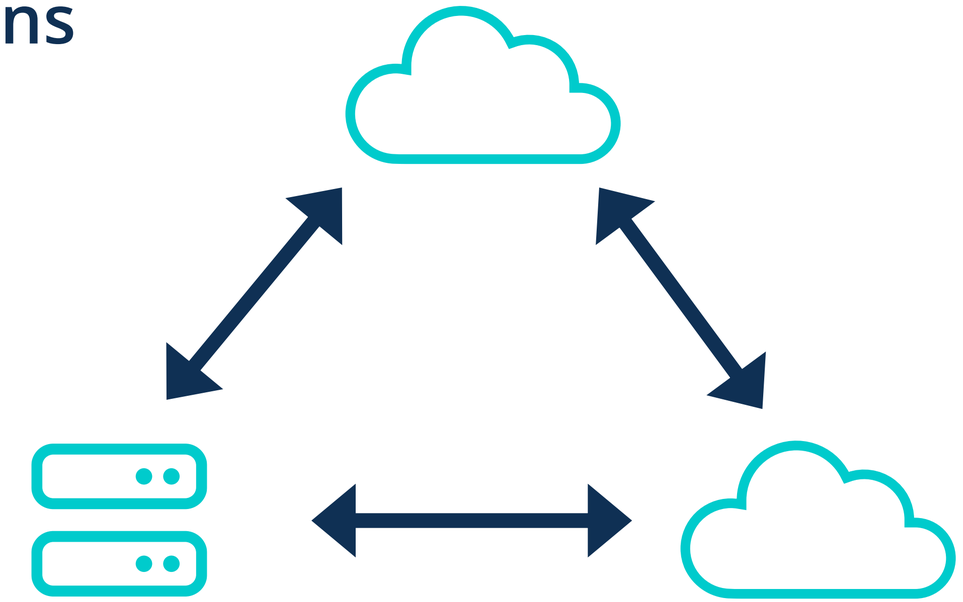
NOVASTOR

Petabyte - Billion files

Considerations

- 1 Million files
  - A regular path length of 64 characters, i.e.: *C:\ProgramData\YourApp\8239ed25-0198-4858-85b1-f9a8634ef7f7.abc*
  - Information about the owner, ACLs, creation/modification times etc.
  - Result: At least 128MB Metadata
- 100 Million files: 12.8 GB just for metadata
- 1 Billion files: 128 GB just for metadata
- Standard tasks
  - Find all files modified before/after a certain date
  - Retrieve all data from one customer
  - Remove backups from a former customer

Petabyte - Billion files | Technical Solutions



- D2D2C, D2C2C, C2C

- Reducing data by
  - Compression, source level
  - Deduplication, source level

- Accessing data by
  - Smart indexing, no matter if native files, databases or VMs

- Securing data by
  - Encryption on transport (TLS >=1.2) and storage (AES)
  - Immutable storage (Object locking, immutable file system)

- Making backup resilient by
  - Integrating your backup app with SIEM (Security Information and Event Management)

Petabyte - Billion files | Structural Solutions

Backup server locally + cloud storage

- No cloud PCUs needed
- All administration/ configuration locally
- C2C not possible without transferring the data through a local proxy

Backup server in the cloud + cloud storage

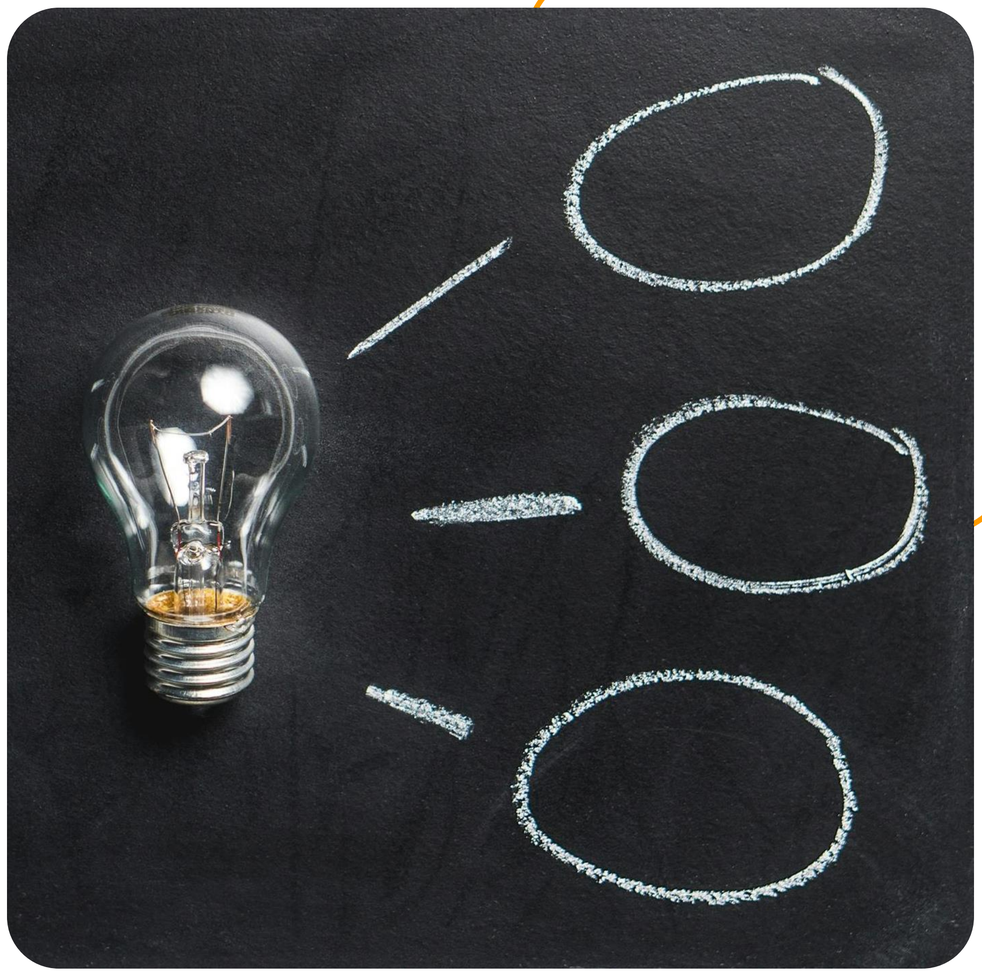
- Additional cost through usage of cloud resources
- C2C capabilities
- Split brain

NOVASTOR

Petabyte - Billion files

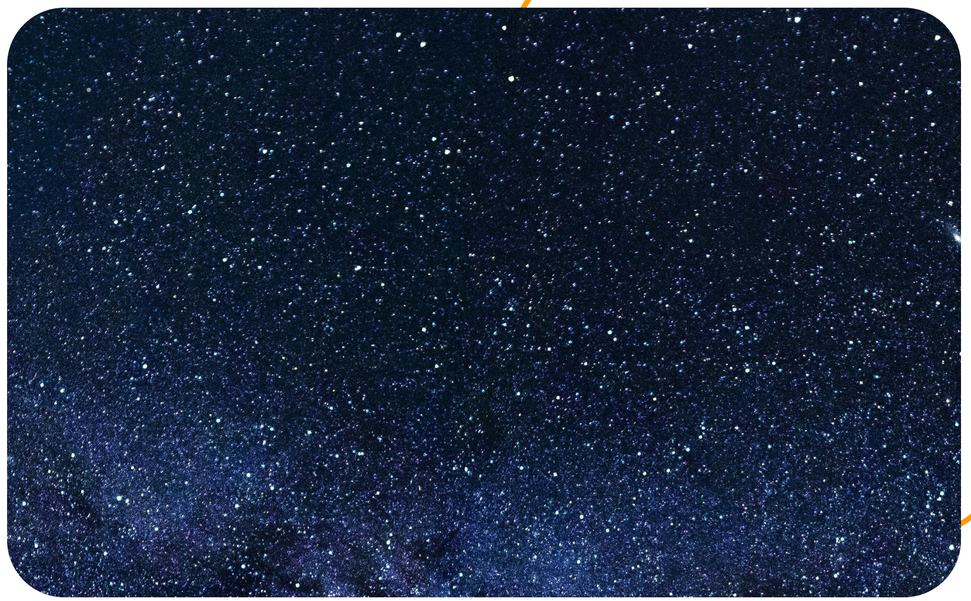
Techniques

- Deduplication
  - Block level instead of file level
  - Fully configurable storage, optimized for your file sets (100GB video files or receipt scans)
  - Deduplication between backups
- Indexing
  - Index location
  - Split index



Your PByte - Billion file project

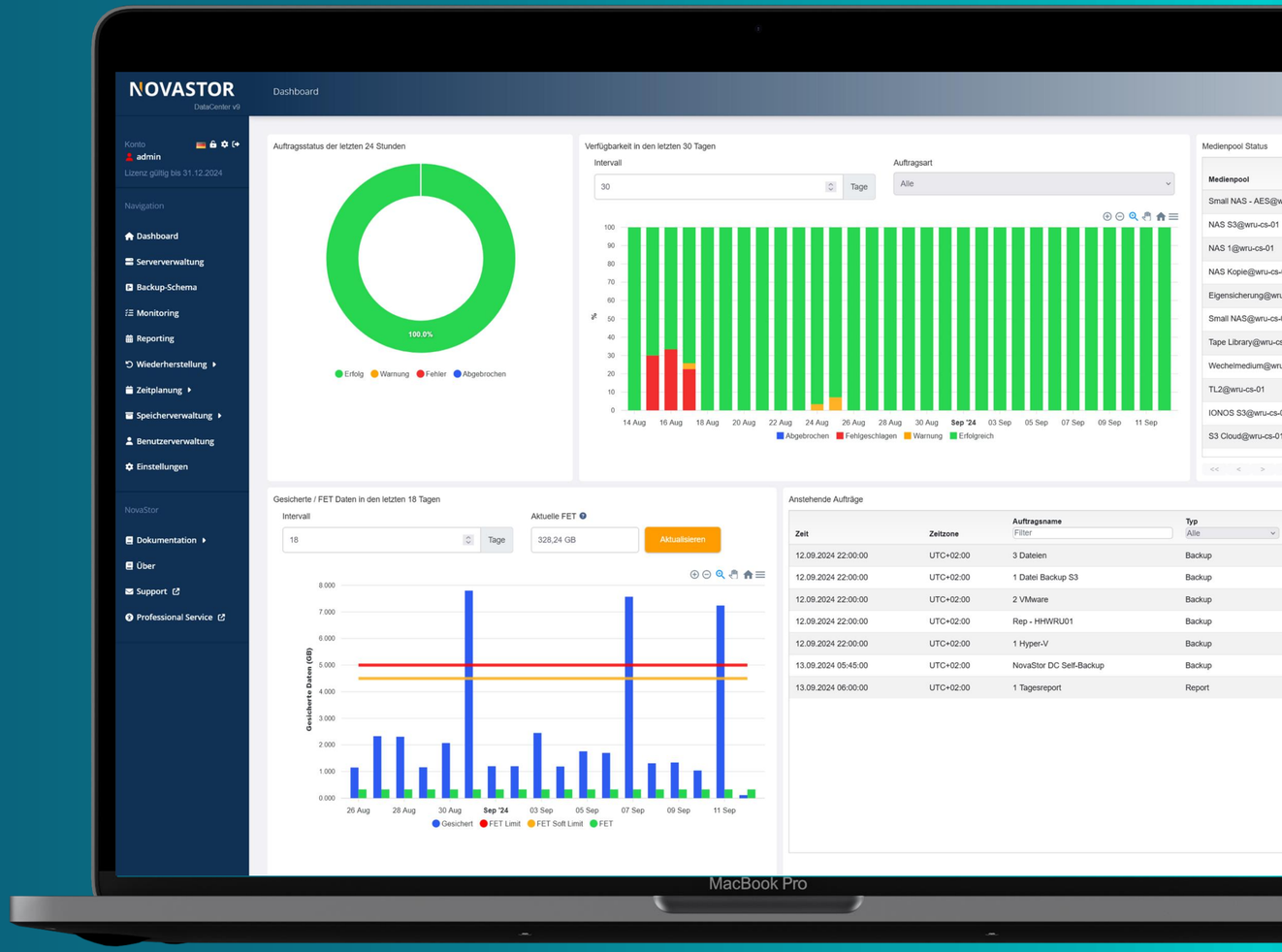
A real world example



- Customer with SAP interface for document management
- 50M documents (each around 256k) and growing
- Data stored on NAS system in a public cloud environment
  - Replication to another location
- Backup needed for disaster recovery, S3 Cloud storage to be used
  - Initial full backup in 24-36 hours
  - Incremental backups once a day within 2-4 hours
  - No relevant impact on production
  - Full restore with prioritization

Dear admin – your weekend is yours!

- Simple data backup & recovery
- 100% Made in Germany
- Comprehensive advice and support
- Technical support directly from the manufacturer



NOVASTOR

The path to the mountain top starts with your question.

bernd.pohlmann@novastor.com

+49 (40) 63809 980

www.novastor.de

Neumann-Reichardt-Straße 27-33
22041 Hamburg | Germany

Bernd Pohlmann

CTO @ NovaStor

NOVASTOR