Scale CES S3 with watsonx.ai & Db2 DWH & watsonx.discovery

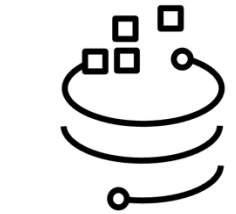
IBM Storage Scale Days 2025 DE

March 19th – 20th, 2025 | Heidelberg, Germany

Harald Seipp, IBM Client Engineering

seipp@de.ibm.com

Disclaimer

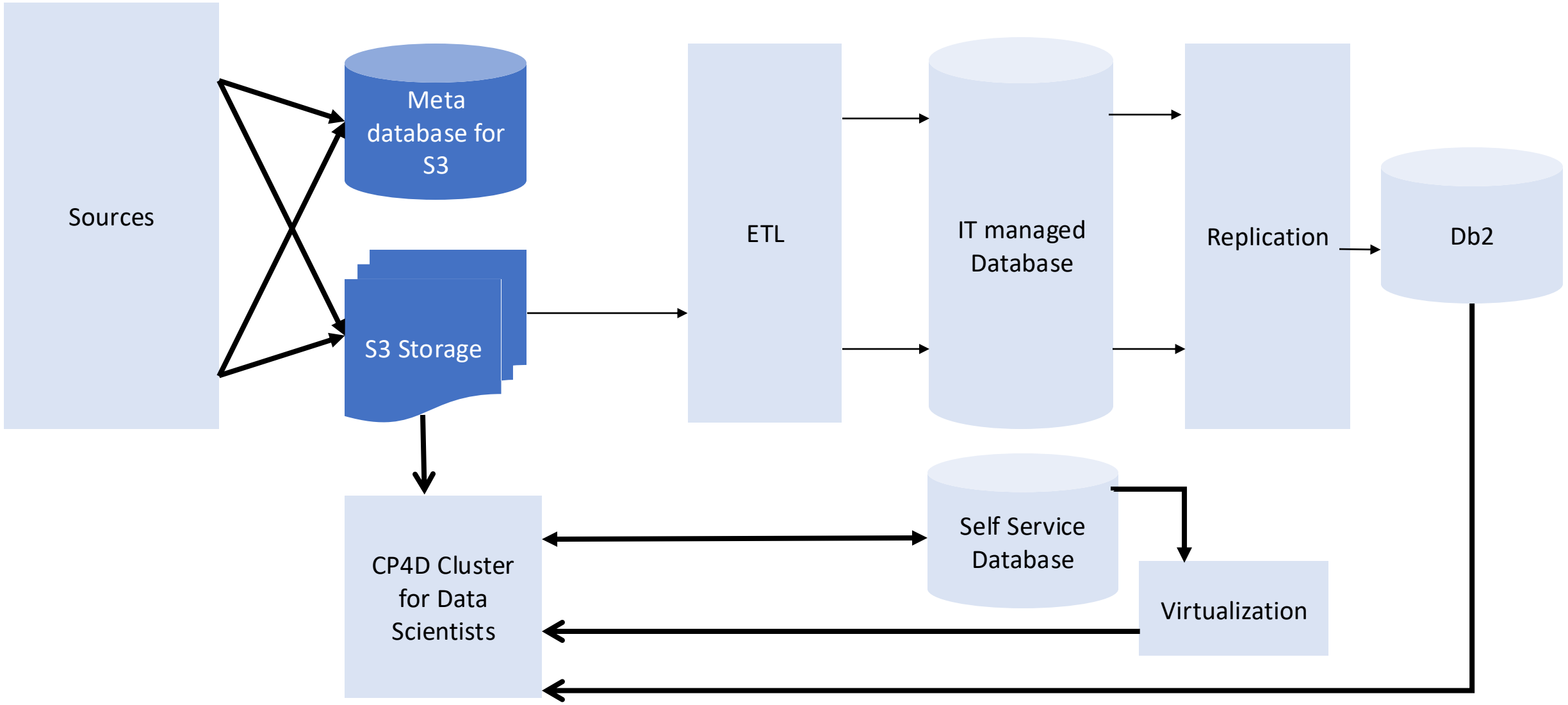


- IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.
- IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

Use Case 1: S3 and Generative AI

- Runtime environment for RAG application (frontend and backend): OpenShift
- LLM provider: watsonx.ai, IBM’s AI & data platform
- Document database : Elasticsearch Enterprise, a search solution enhanced with vector embedding and hybrid search capabilities (provided through IBM watsonx discovery)
- IBM CES S3 as central object store for documents, Elasticsearch snapshots, …

Use Case 2: S3 as central „Landing Zone“



Main Requirements for Solution Architecture

Data must not leave the data center

On-Premise

Some documents that will be analyzed contain sensitive information. Even sending parts of documents as context to LLM hosted externally is currently not possible.

Existing toolset

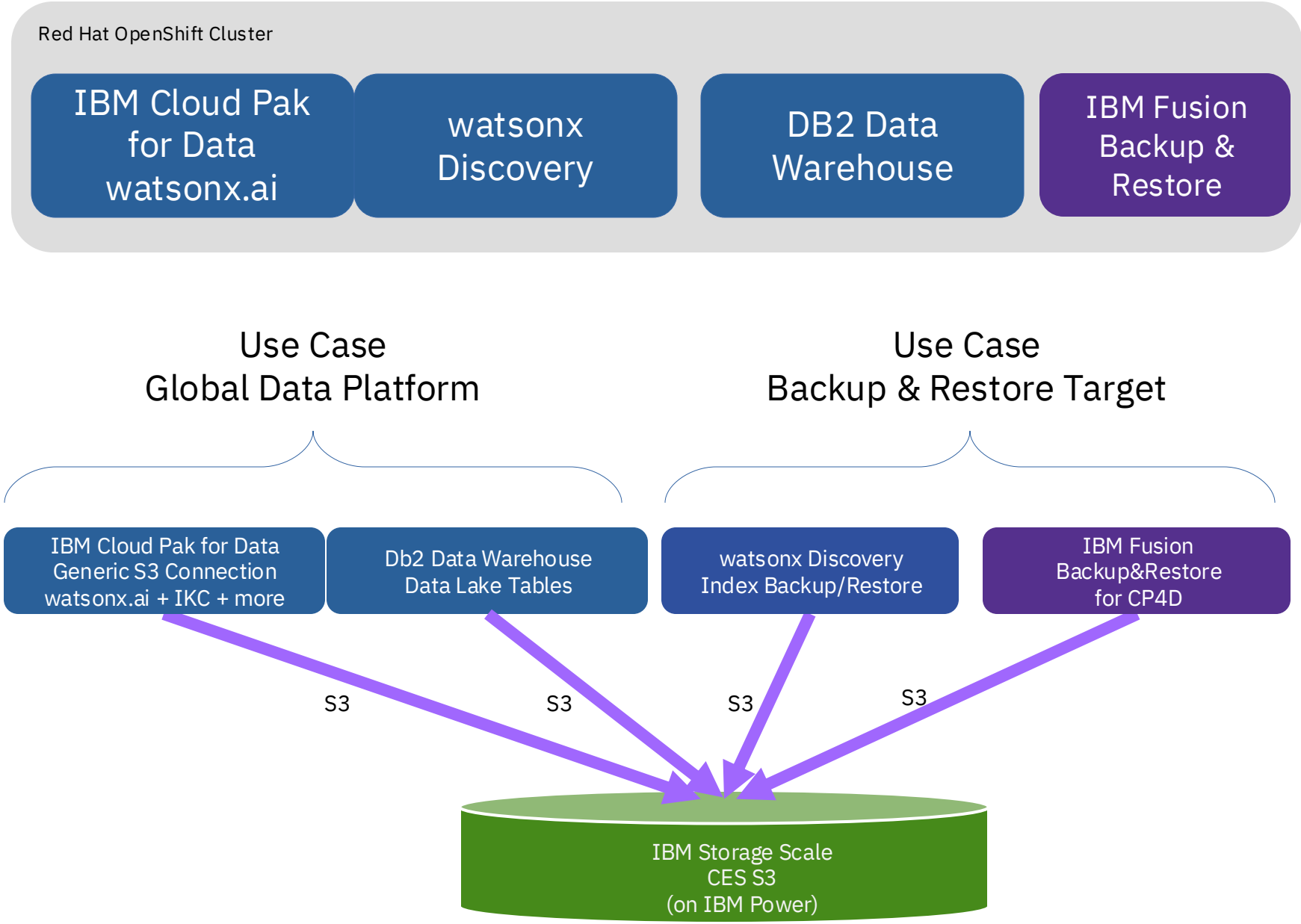
Existing CP4D cluster as endpoint for all “classical” ML deployments.

Private cloud

Consolidate workload around a private cloud setup on top of Cloud Pak for Applications (OpenShift).

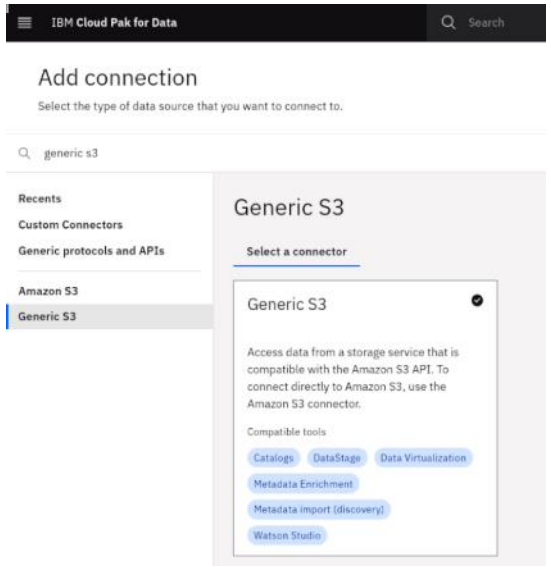
Pilot Overview

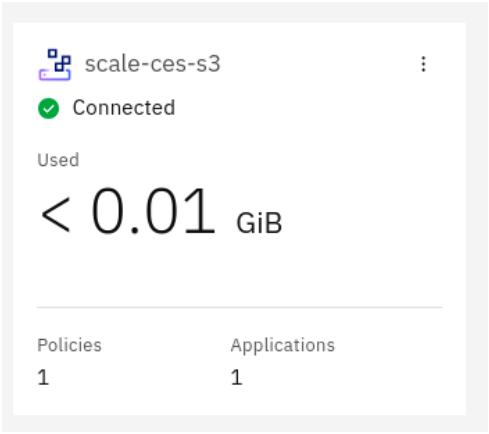
- Pilot motivation: Although IBM Software products state S3 protocol support, interoperability validation with Storage Scale CES S3 was required
- Pilot target: Prove that product integration works to get IBM Product Management support statements
- Pilot methodology: test the four products in separate sprints with weekly playbacks
  - 3 products tested at IBM
  - wx.discovery at client
- Overall duration: 4 wks



Pilot Summary

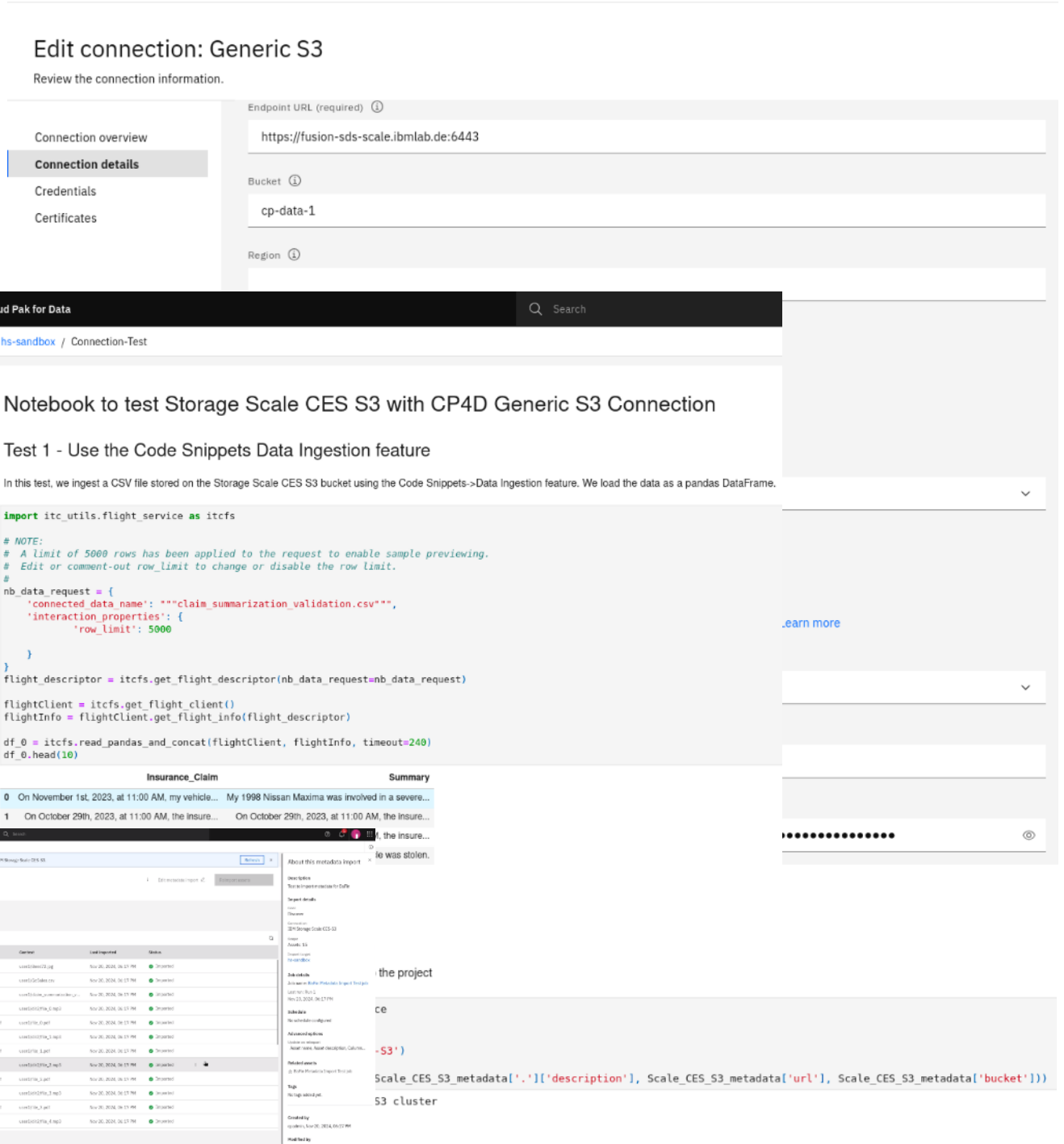
- Some challenges setting up TLS certificates for S3 clients
  - Lesson learned: Client uses an own CA, so the client software needs to be configured with either the CA certificate or the full certificate chain
- All test were successful
- The involved product teams provided a support statement for the client





Cloud Pak for Data testing

- Use CP4D Generic S3 Connection
- Jupyter Notebook deployed within CP4D to test
  - Connectivity
  - Read/Write data
  - Multipart upload
  - Bucket operations
  - Read/Write Object metadata
- Used IBM Knowledge Catalog to import Metadata from a CES S3 connection



Db2 Data Warehouse testing

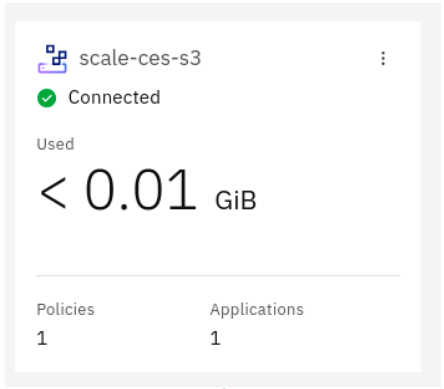
- Db2 11.5.9 with Db2 BigSQL 11.3.0, db2uinstance CR, 3 nodes
- archivelogs persistent volume must use ReadWriteMany access mode
- CES S3 TLS certificate must be imported to Db2 Java Keystore
  (default password: changeit)
- Undocumented command to set Path-style S3 access
  db2 "CALL SYSHADOOP.SET_DATALAKE_CONFIG('CORE', 'fs.s3a.bucket.user1.path.style.access', 'true')"
- Data import is supported for Parquet data but not for Iceberg tables written outside Db2

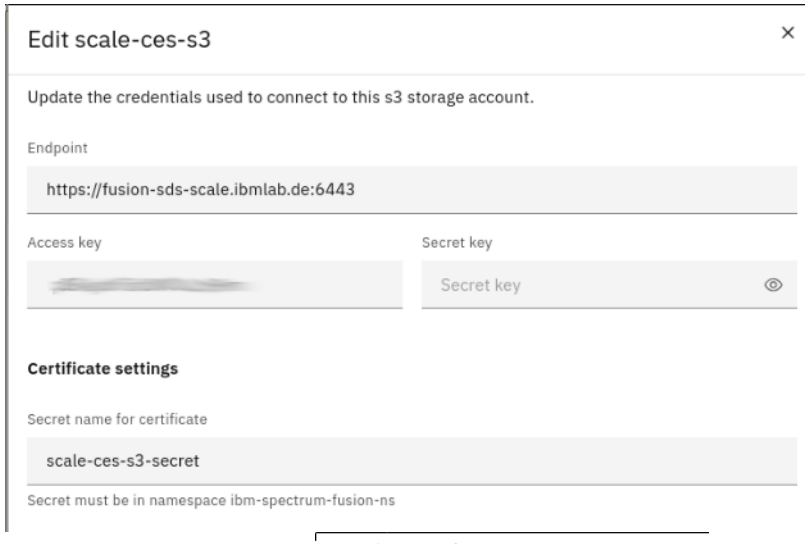
```
$ db2 "CREATE DATALAKE TABLE t2 (i INT, s VARCHAR(10)) STORED BY \
ICEBERG LOCATION 'DB2REMOTE://cesnode1-user1//t2-iceberg' \
TBLPROPERTIES ('external.table.purge'='true');"
DB20000I  The SQL command completed successfully.

$ db2 "INSERT INTO t1 VALUES (55131, 'Mainz');"
DB20000I  The SQL command completed successfully.
$ db2 "INSERT INTO t1 VALUES (55278, 'Undenheim');"
DB20000I  The SQL command completed successfully.
$ db2 "INSERT INTO t1 VALUES (55270, 'Zornheim');"
DB20000I  The SQL command completed successfully.

$ db2 "SELECT * FROM t1"
I           S
----------- ----------
      55131 Mainz
      55270 Zornheim
      55278 Undenheim
  3 record(s) selected.
```

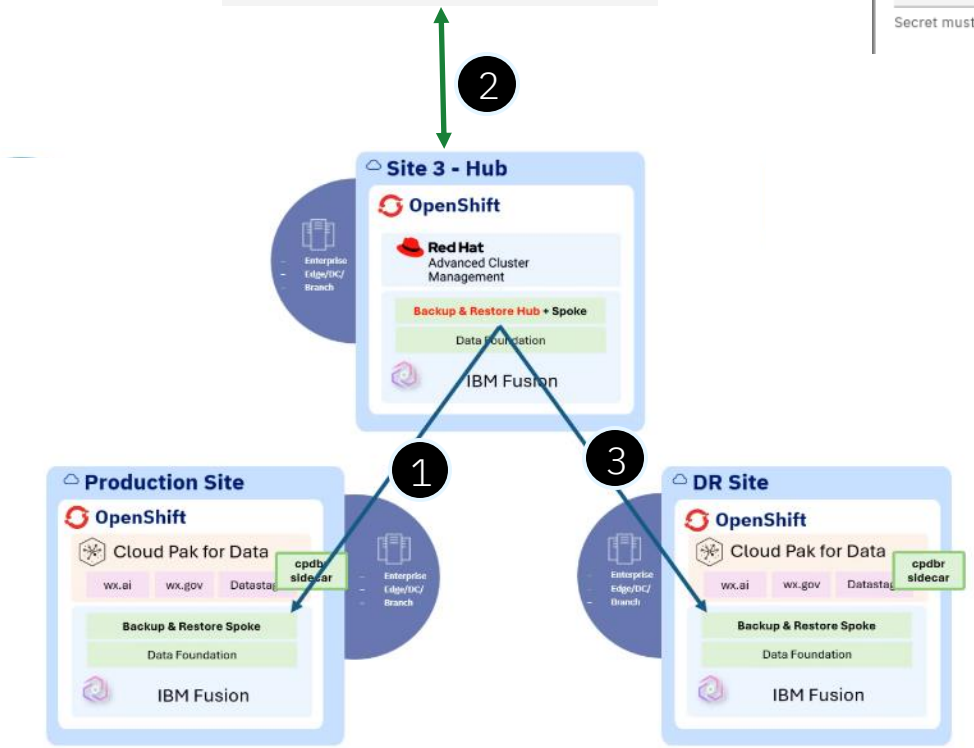
```
# ls -ltR t2-iceberg
t2-iceberg:
total 1
drwxrwx---. 2 root root 4096 Nov 21 11:19 metadata
drwxrwx---. 2 root root 4096 Nov 21 11:19 data
t2-iceberg/metadata:
total 80
-rw-rw----. 1 root root 4416 Nov 21 11:19 00003-bd1ab615-1ad4-4c12-9f3c-747311f536dc.metadata.json
-rw-rw----. 1 root root 4368 Nov 21 11:19 snap-8275370836592526925-1-0ef1dca6-5a24-4a64-8fbc-
cfae7ac2948e.avro
-rw-rw----. 1 root root 6677 Nov 21 11:19 0ef1dca6-5a24-4a64-8fbc-cfae7ac2948e-m0.avro
-rw-rw----. 1 root root 3478 Nov 21 11:19 00002-c97e146c-907f-41e6-a97d-4c802f296799.metadata.json
-rw-rw----. 1 root root 4316 Nov 21 11:19 snap-1379196754593492153-1-ba4dc5c7-d7b5-40b6-bab1-
a23cd8401545.avro
-rw-rw----. 1 root root 6679 Nov 21 11:19 ba4dc5c7-d7b5-40b6-bab1-a23cd8401545-m0.avro
-rw-rw----. 1 root root 2540 Nov 21 11:19 00001-7c66f0c6-8b6b-47ab-8be5-3f7da06e6e00.metadata.json
-rw-rw----. 1 root root 4244 Nov 21 11:19 snap-4061617877835189135-1-1444175c-403e-4c19-894d-
3f891932ba41.avro
-rw-rw----. 1 root root 6675 Nov 21 11:19 1444175c-403e-4c19-894d-3f891932ba41-m0.avro
-rw-rw----. 1 root root 1547 Nov 21 11:14 00000-86c3a4f8-4185-4fde-8b4e-88a69aacdef8.metadata.json
t2-iceberg/data:
total 24
-rw-rw----. 1 root root 685 Nov 21 11:19 239620564-0-20241121101943961-00001.parquet
-rw-rw----. 1 root root 692 Nov 21 11:19 1483219878-0-20241121101934442-00001.parquet
-rw-rw----. 1 root root 664 Nov 21 11:19 1093471490-0-20241121101918648-00001.parquet
```

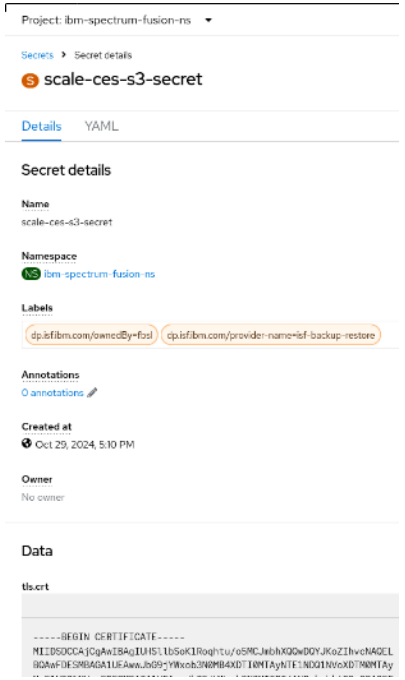
Fusion Backup & Restore testing

- Backed up CP4D with the applications the client uses on one OpenShift cluster, restored on another OpenShift cluster
- Leveraging IBM Fusion recipes for CP4D application consistency
- Ensured that all components are functional after restore









IBM Client Engineering

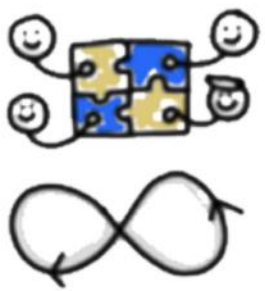
Do you have a
GREAT IDEA
to solve and improve a business challenge?

Start-up speed with Enterprise Scale

INNOVATE
in 4-10 hours

We **align** on a solution and define a **scope to prove value**.

PREPARE
in 1-5 days



We **prepare, plan, and commit** to be ready to rapidly build.

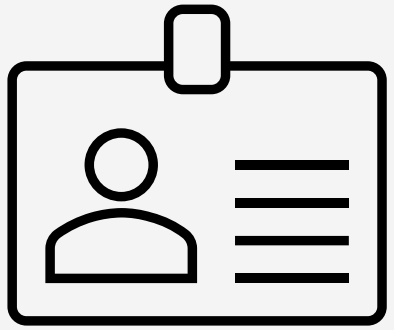
CO-CREATE
in 2-4 weeks

We **build, validate, and improve** over a series of Agile Sprints and Playbacks with you.

TRANSITION
adapted for you

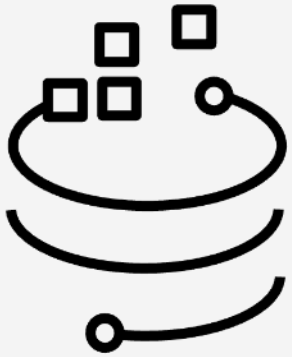
We get on path to bring our solution into **scaled production!**

Thank you

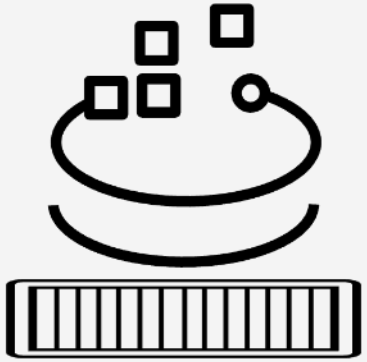


Harald Seipp, IBM Client Engineering
seipp@de.ibm.com

Thank you for using



Storage Scale



Storage Scale System