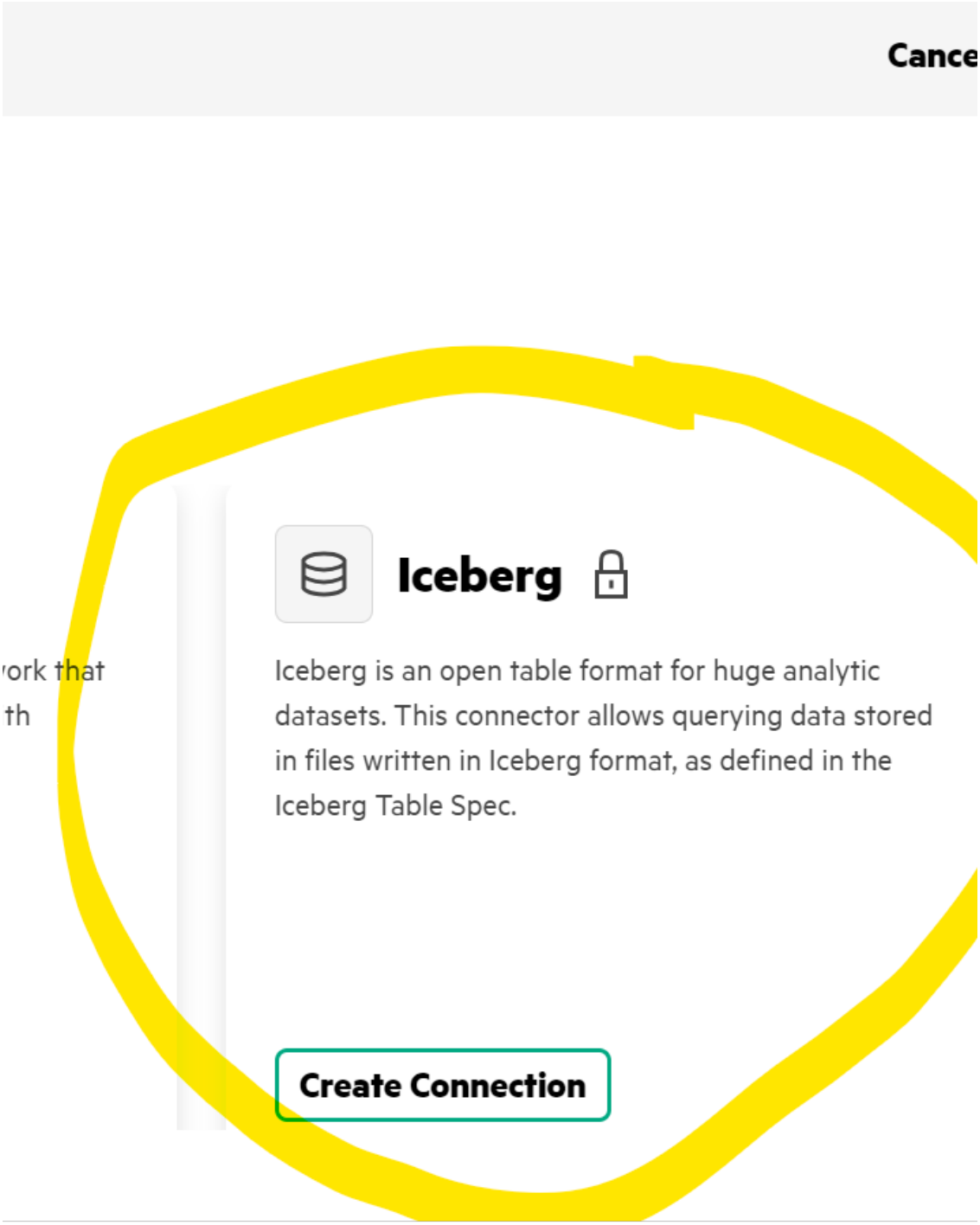
How to add an Iceberg data connection using Hadoop as the catalog type in EZUA 1.5

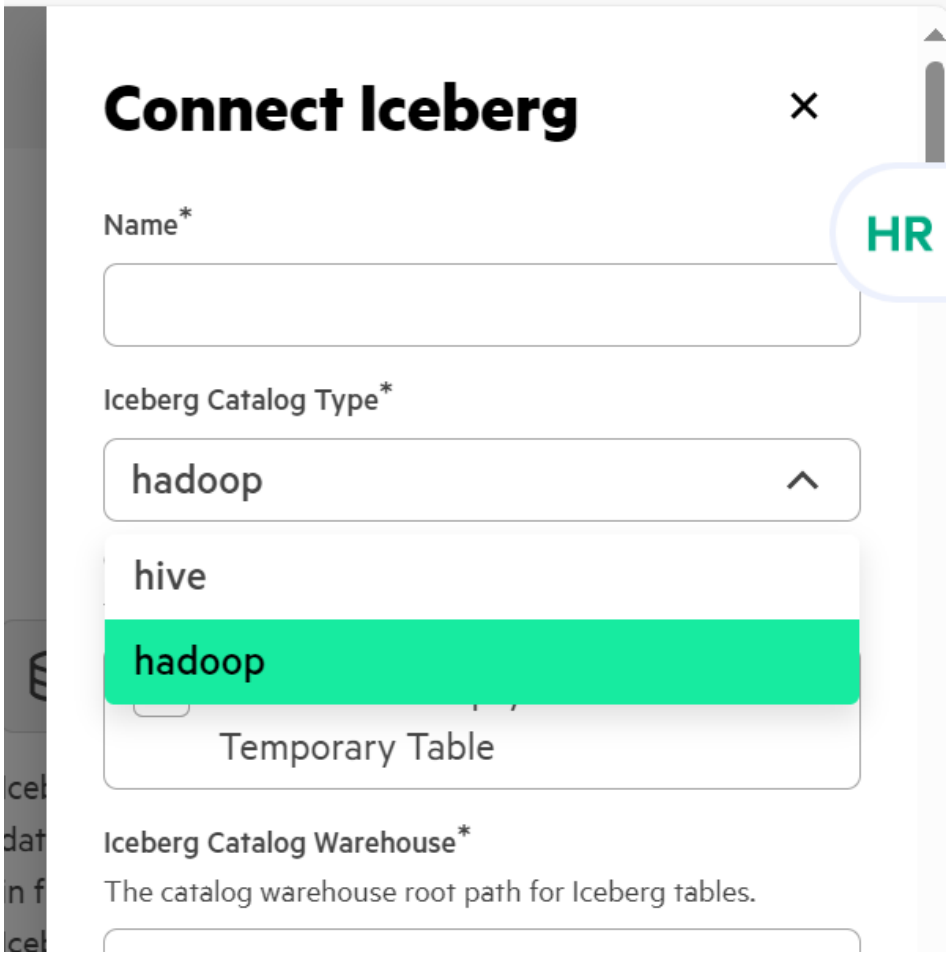
Table of contents

Issue

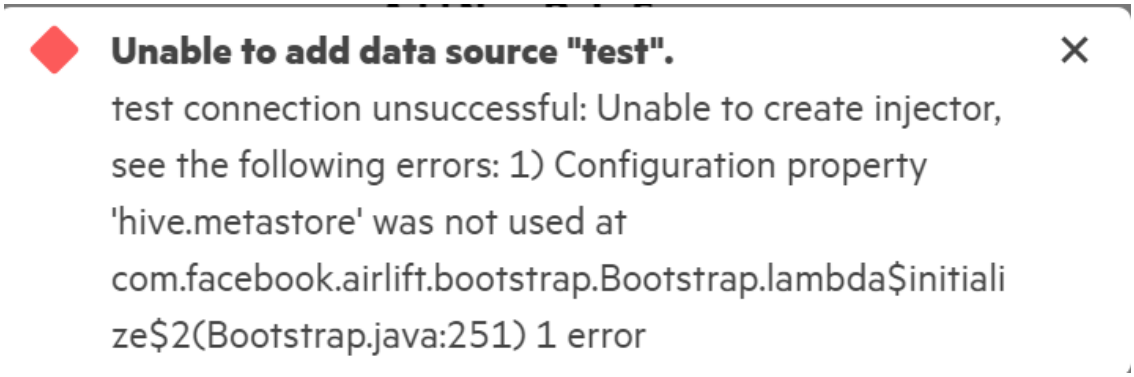
A user wants to add a DataSource for an Iceberg connection using "hadoop" as the Iceberg Catalog Type in EZUA 1.5 by selecting "Create Connection" as shown below:



They further choose "hadoop" as the Iceberg Catalog Type as showed below:



Then the user adds other details and selects "Create Connection". The following error is returned:



Environment

EZUA 1.5

Cause

Recent changes introduced by Open Source PrestoDB cause Iceberg data connections to fail in Unified Analytics when the Catalog Type is Hadoop. This appears to be a bug as it expects Hive MetaStore values whether Hive was selected or not.

Resolution

You can add the connection from the CLI using API calls and curl. You need to run the below curl command from your K8s master host (i.e. the second control plane) and NOT from the coordinator host.

```
curl -u user:password --location '
https://ezpresto.<domain_name>/v1/catalog'
--header 'Content-Type: application/json' --insecure --data '{

    "catalogName": "test1",
    "connectorName": "iceberg",
```

```
    "properties": {
        "iceberg.catalog.type": "hadoop",
        "iceberg.file-format": "PARQUET",
        "iceberg.compression-codec": "GZIP",
        "iceberg.max-partitions-per-writer": "100",
        "iceberg.minimum-assigned-split-weight": "0.05",
        "iceberg.catalog.warehouse": "s3://devbucket/iceberg_data",
        "iceberg.catalog.cached-catalog-num": "10",
        "hive.s3.aws-access-key": "<>",
        "hive.s3.aws-secret-key": "<>",
        "hive.s3.endpoint": "<>",
        "hive.s3.path-style-access": true,
        "hive.s3.ssl.enabled": false
    },
    "fileProperties": {
        "iceberg.hadoop.config.resources": [
            "<>"
        ]
    }
}'
```

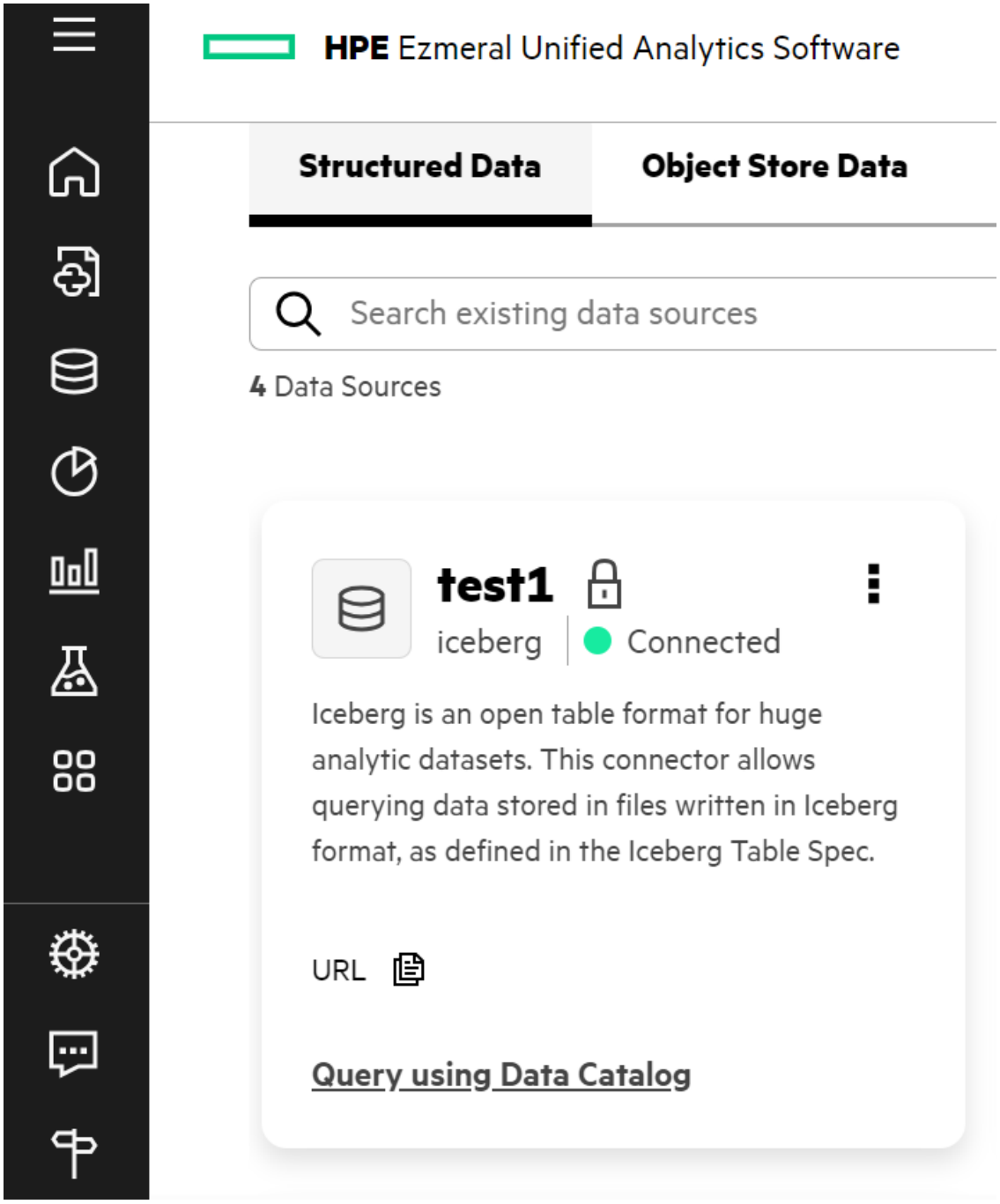
Important Notes:

1. In the above curl command, it is asking for the user/password which you use for the EZUA platform or portal login.
2. You can change the catalog name, as shown in the example above (the example uses the name "test1")
3. Do not change the connector name
4. Change the iceberg.catalog.warehouse, hive.s3.aws-access-key, hive.s3.endpoint, and hive.s3.aws-secret-key to match your environment.
5. For the iceberg.hadoop.config.resources property, you need to create the below file and supply the values accordingly. Then change the whole content of the file to base64 format and provide it inside the "<>"

The file contents which need to be changed to base64:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
    <property>
        <name>presto.s3.ssl.enabled</name>
        <value>false</value>
    </property>
    <property>
        <name>presto.s3.access-key</name>
        <value>**************************</value>
    </property>
    <property>
        <name>presto.s3.secret-key</name>
        <value>**********************************</value>
    </property>
    <property>
        <name>presto.s3.endpoint</name>
        <value>****************************************</value>
    </property>
    <property>
        <name>presto.s3.path-style-access</name>
        <value>true</value>
    </property>
    <property>
        <name>fs.s3.impl</name>
        <value>com.facebook.presto.hive.s3.PrestoS3FileSystem</value>
    </property>
    <property>
        <name>fs.s3a.impl</name>
        <value>com.facebook.presto.hive.s3.PrestoS3FileSystem</value>
    </property>
</configuration>
```

After running the above command, you will see a new tile in the Datasource window for Iceberg connection where the catalog type is Hadoop as shown below:


HPE Ezmeral Unified Analytics Software
Structured Data
Object Store Data
Search existing data sources
4 Data Sources
test1
iceberg
Connected
Iceberg is an open table format for huge analytic datasets. This connector allows querying data stored in files written in Iceberg format, as defined in the Iceberg Table Spec.
URL
Query using Data Catalog