BigData Analytics

Installation

sohailimran2@yahoo.com

Ref: Web

APACHE DRILL

Apache Drill is an innovative **schema-free SQL query engine** for Hadoop, NoSQL, and cloud storage. This open-source software enables users to analyze large-scale datasets from numerous sources directly, without needing to shift data across systems. Apache Drill provides the **flexibility** of on-the-fly data discovery, making it significantly adaptable to changing data formats and structures.

Drill supports a variety of NoSQL databases and file systems, including HBase, MongoDB, MapR-DB, HDFS, MapR-FS, Amazon S3, Azure Blob Storage, Google Cloud Storage, Swift, NAS, and local files. A single query can join data from multiple datastores. For example, you can join a user profile collection in MongoDB with a directory of event logs in Hadoop.

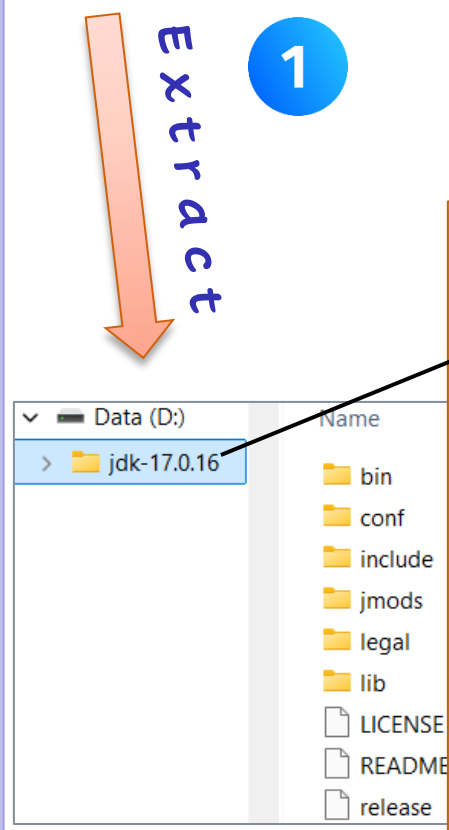
Drill's datastore-aware optimizer automatically restructures a query plan to leverage the datastore's internal processing capabilities. In addition, Drill supports data locality, so it's a good idea to co-locate Drill and the datastore on the same nodes.

Setup JAVA_HOME

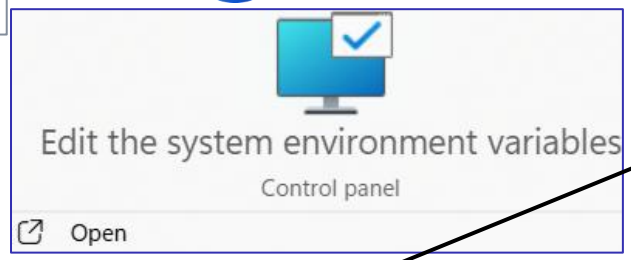
Java

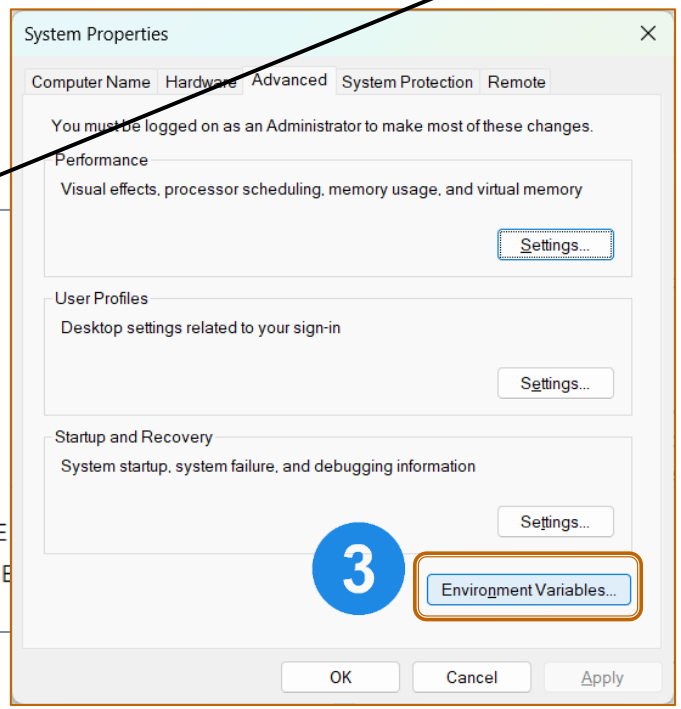
Extract

1



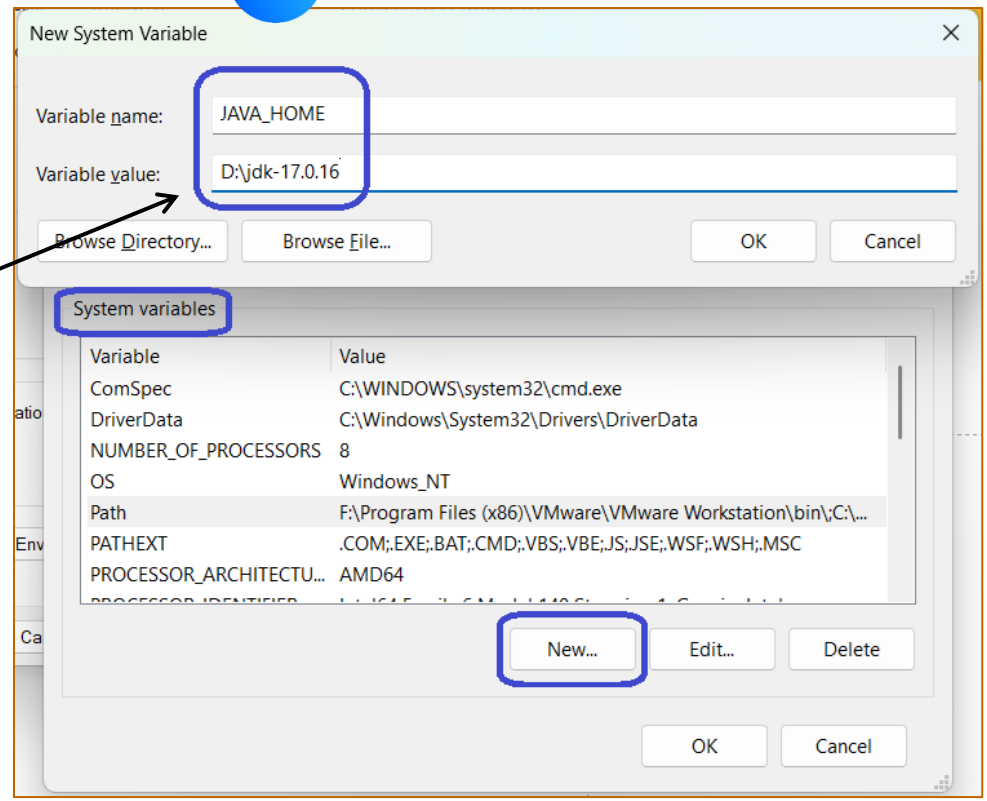
2



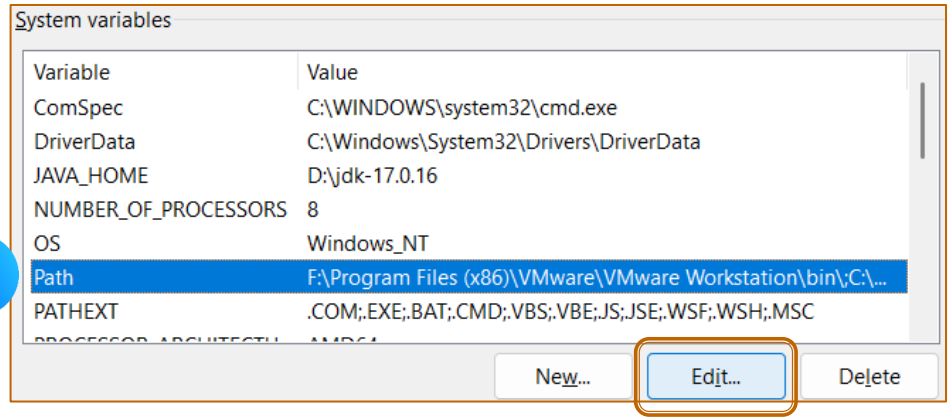


3

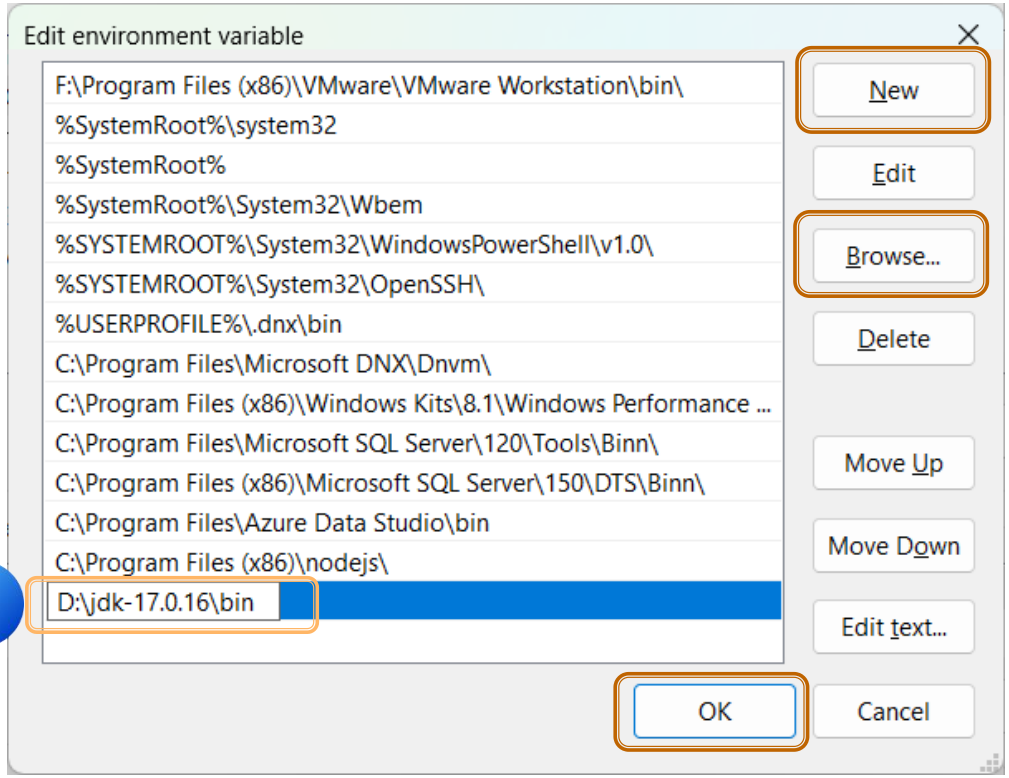
4



Setup - PATH



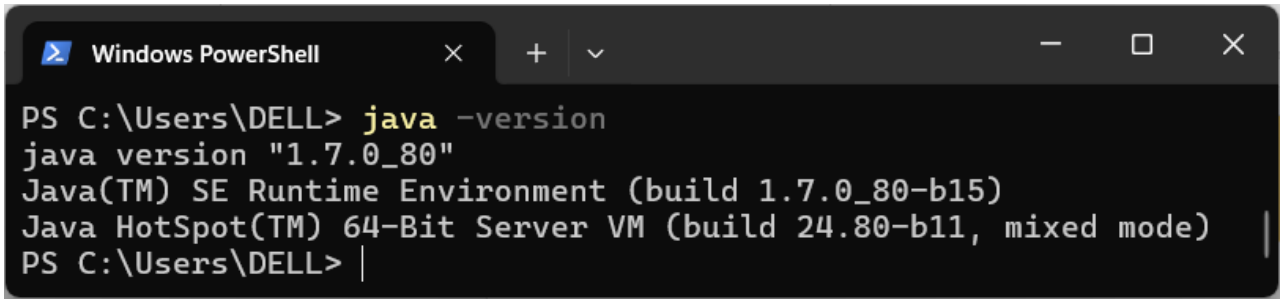
1



2

3

java -version



download and extract

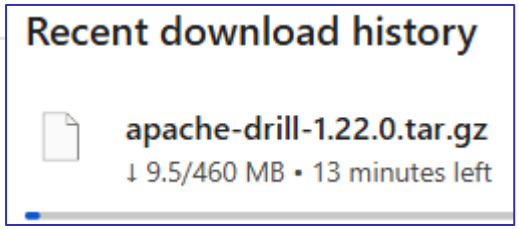
https://drill.apache.org/download/

Download

Drill 1.22.0 was released on June 28, 2025.

Drill for Hadoop 3 and non-Hadoop environments

Find an Apache Mirror | Direct File Download | Client Drivers (ODBC/JDBC)



Drill for Hadoop 2 environments

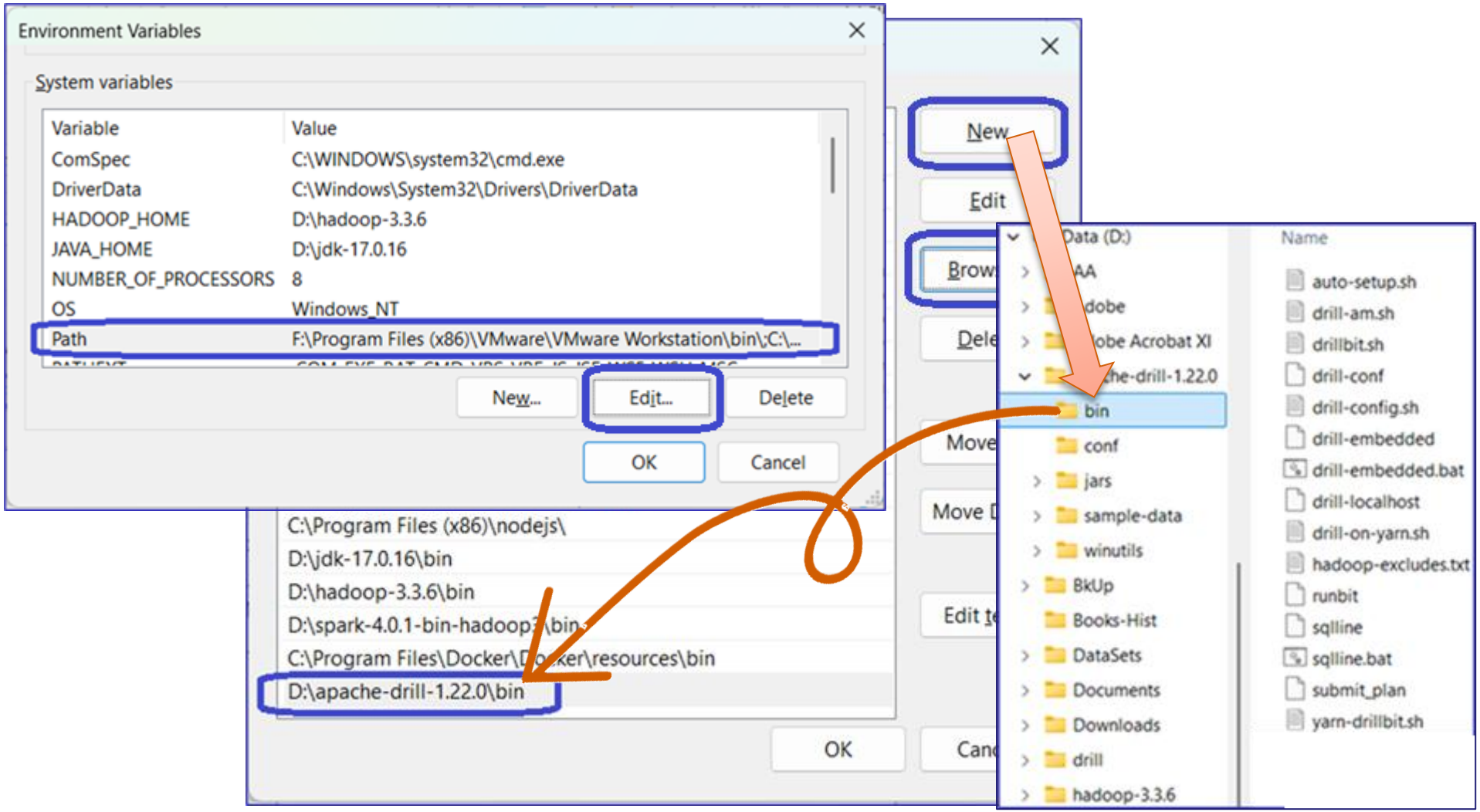
Offical support for Hadoop 2 was dropped along with support for Java 8 with the release of Drill 1.22.0. Users who would like to create their own builds of Drill 1.22.0 or newer for Java 8 and Hadoop 2 are directed to the developer documentation.

Additional resources for Drill 1.22.0:

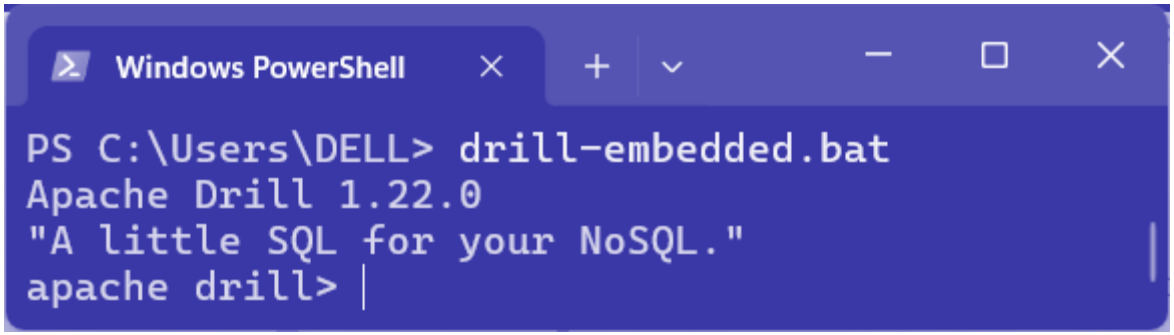
- Blog post
- Release notes (JIRA)
- Source code: tar.gz ZIP.
- Verification Links:
  - SHA
  - ASC
  - KEYS
  - How to Verify Downloaded Files

If you're looking for an older release, see the release notes.

Environment Variables

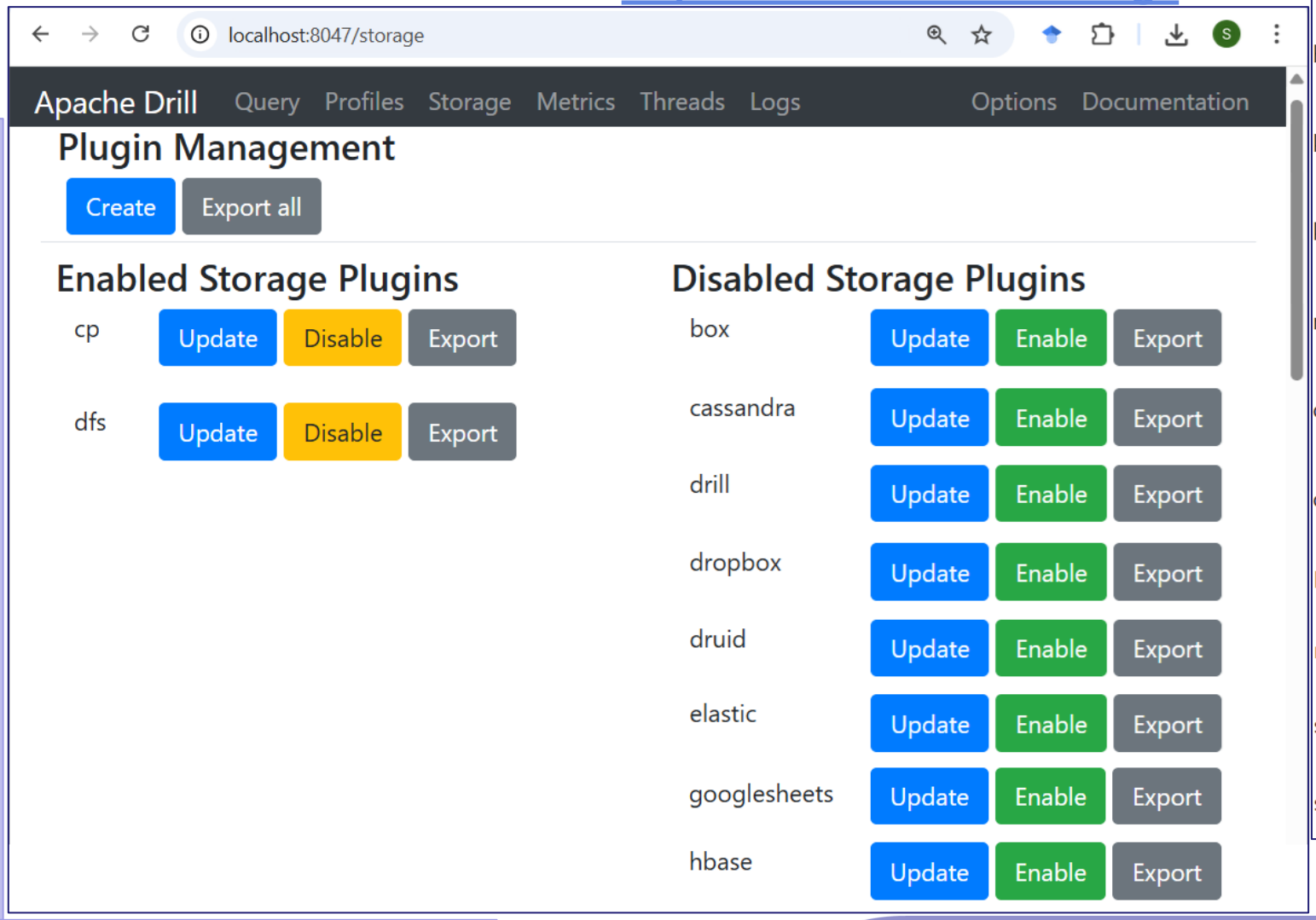


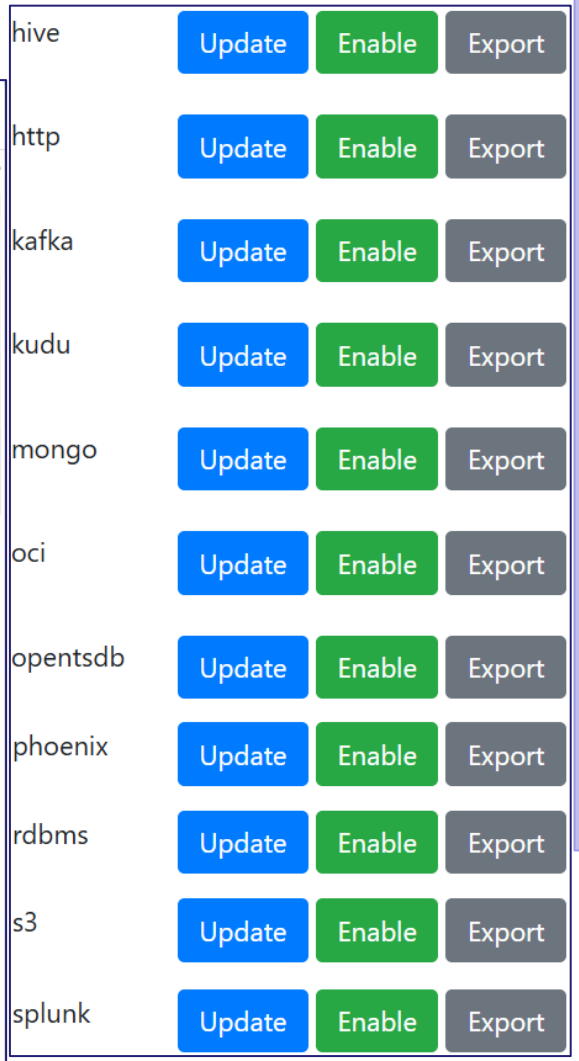
Start


Windows PowerShell
PS C:\Users\DELL> drill-embedded.bat
Apache Drill 1.22.0
"A little SQL for your NoSQL."
apache drill>

Browser

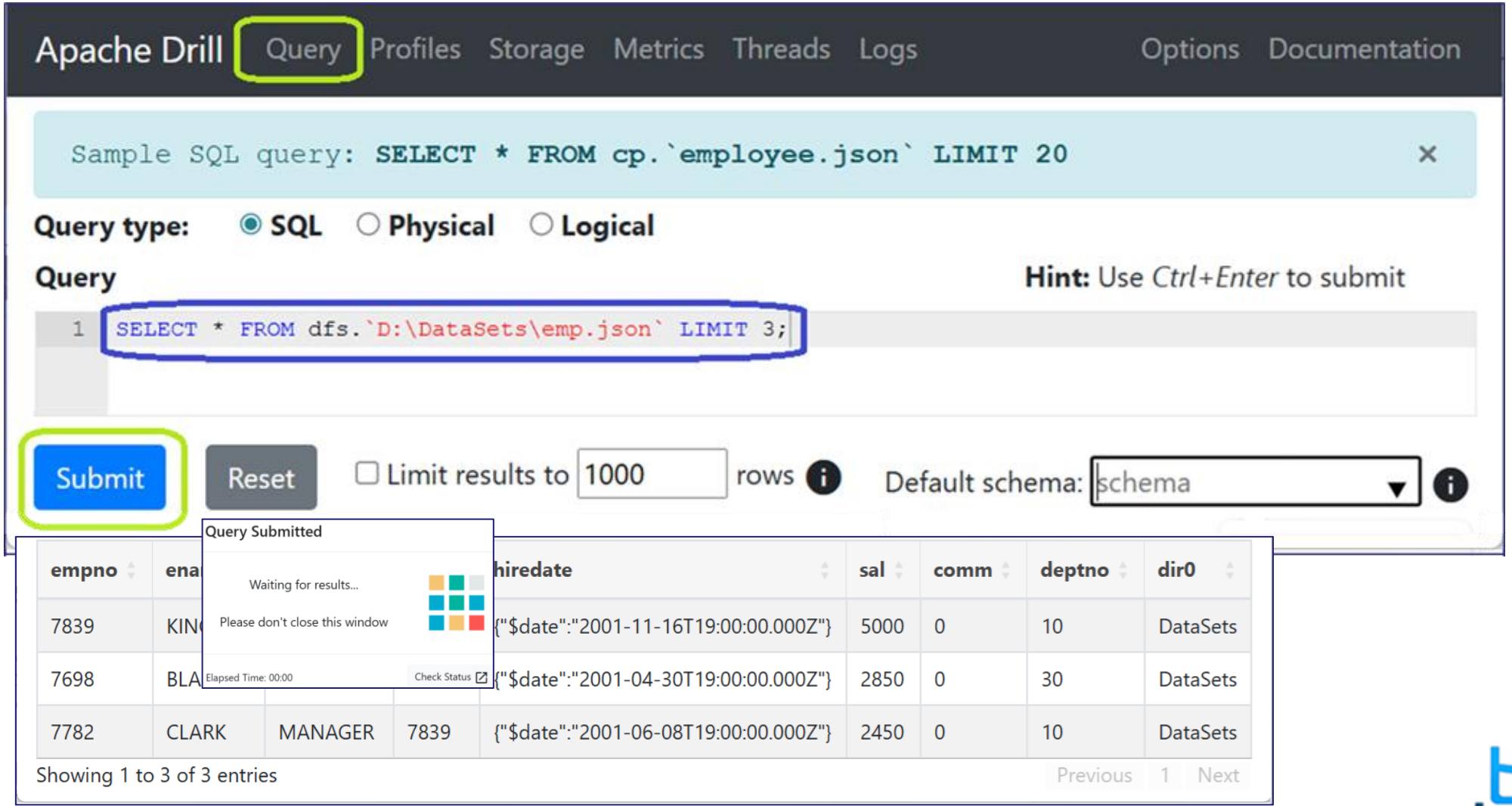
http://localhost:8047/storage



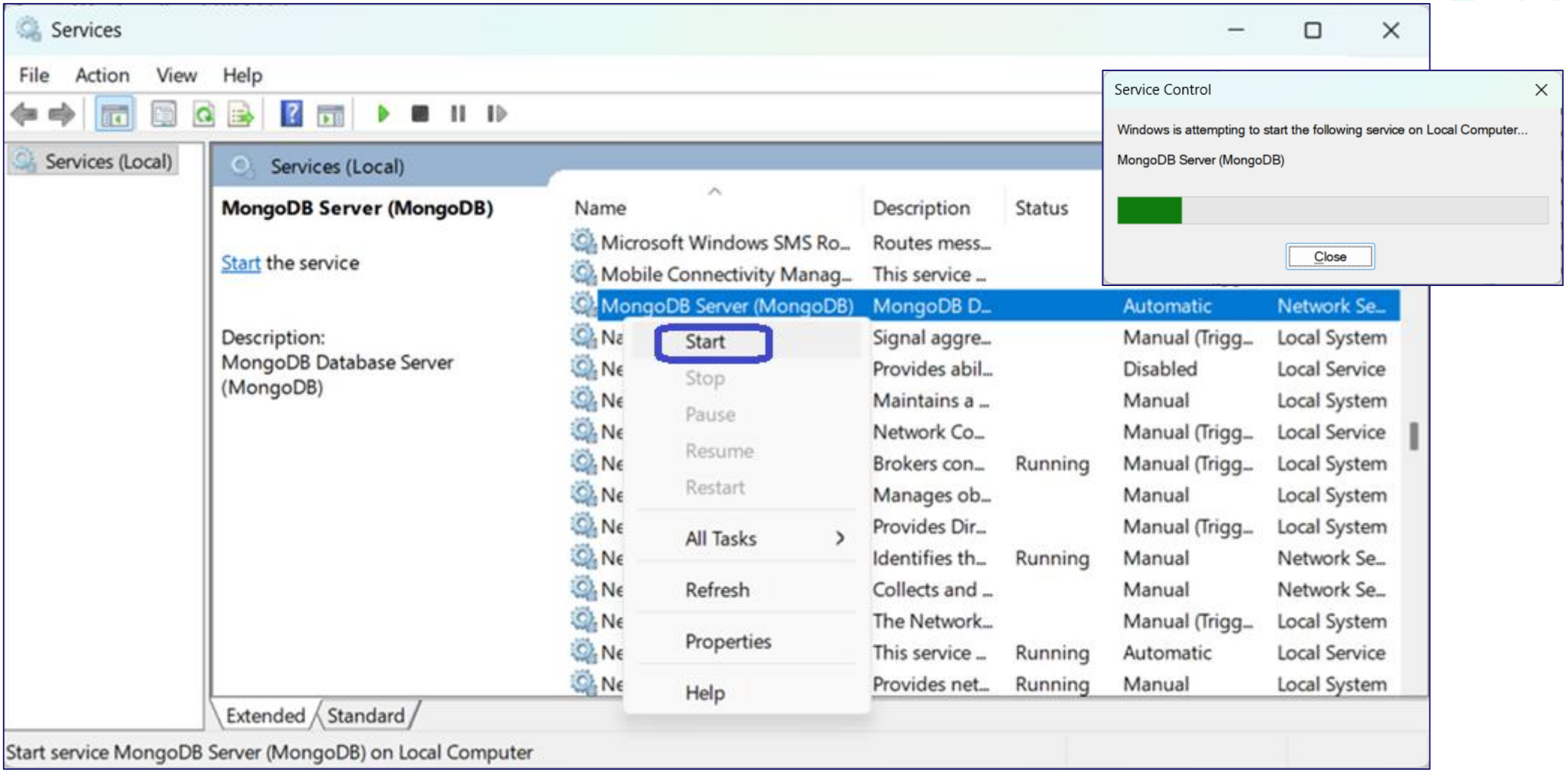


Query - file

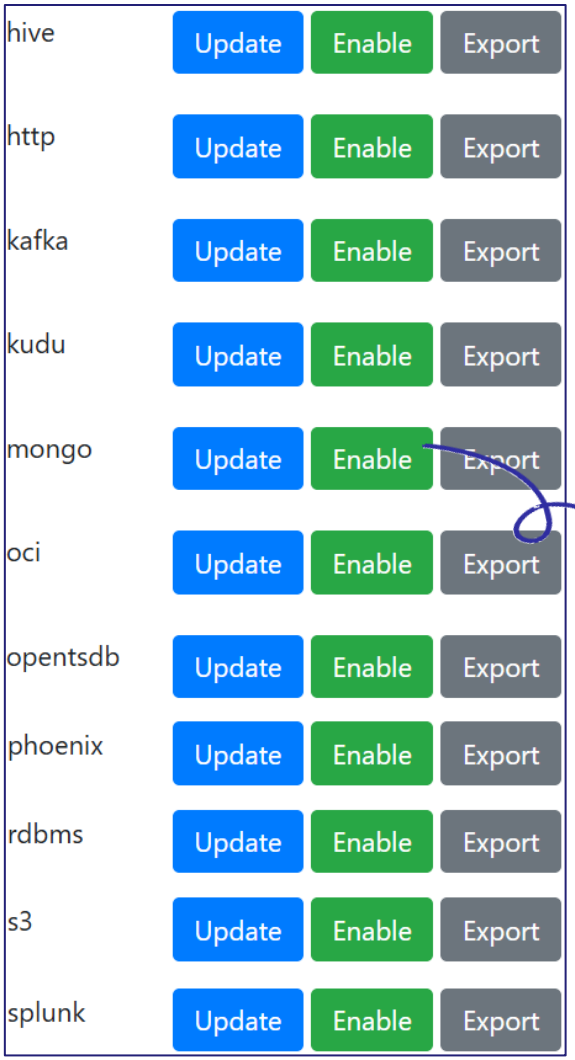
```
SELECT * FROM dfs.`D:\DataSets\emp.json` LIMIT 3;
```

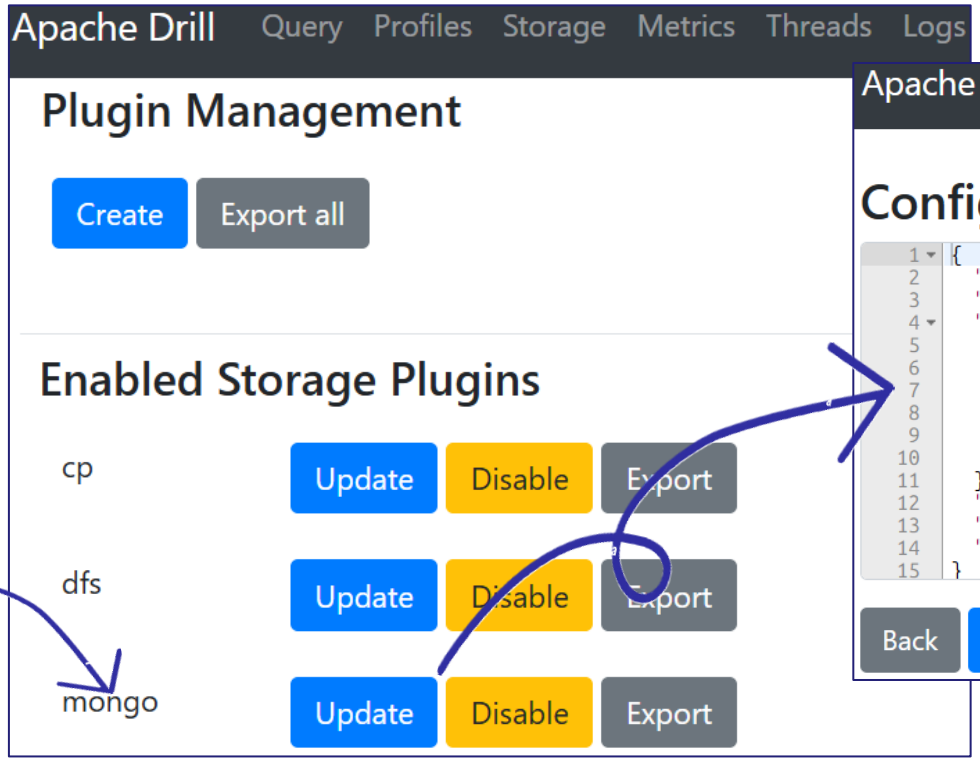


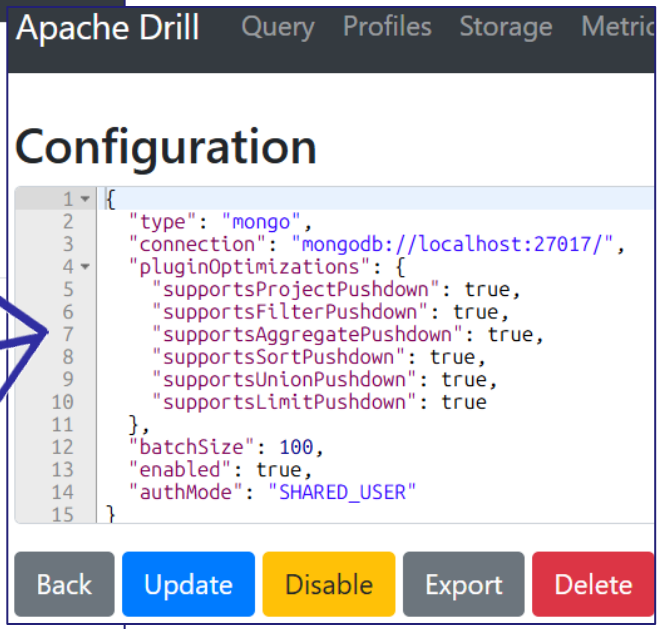
MongoDB service



Enable Mongo







Query - MongoDB

```
SELECT * FROM mongo.mydb.emp LIMIT 3;
```

Query type: ◉ SQL ○ Physical ○ Logical

Query

```
1  SELECT * FROM mongo.mydb.emp LIMIT 3;
```

_id	empno	ename	job	hiredate	sal	deptno	mgr
h\xE1\x1BQh\x97\xF3\xAD\xAE\xDC\xDBz	7839	KING	PRESIDENT	2001-11-17T00:00	5000	10	null
h\xE1\x1BQh\x97\xF3\xAD\xAE\xDC\xDB{	7698	BLAKE	MANAGER	2001-05-01T00:00	2850	30	7839
h\xE1\x1BQh\x97\xF3\xAD\xAE\xDC\xDB|	7782	CLARK	MANAGER	2001-06-09T00:00	2450	10	7839

Showing 1 to 3 of 3 entries Previous 1 Next