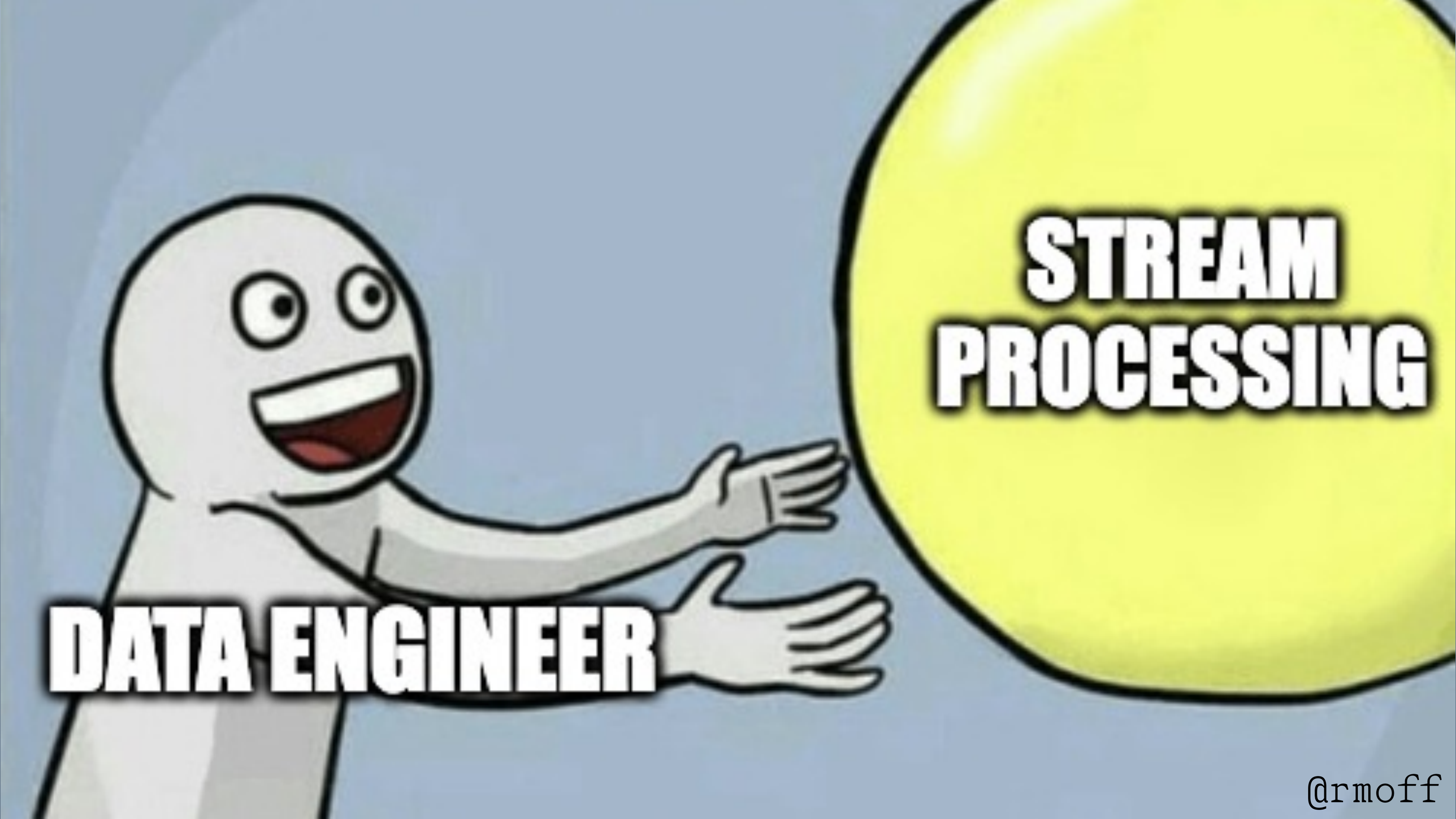
@rmoff / 18 Sep 2024 / #current24

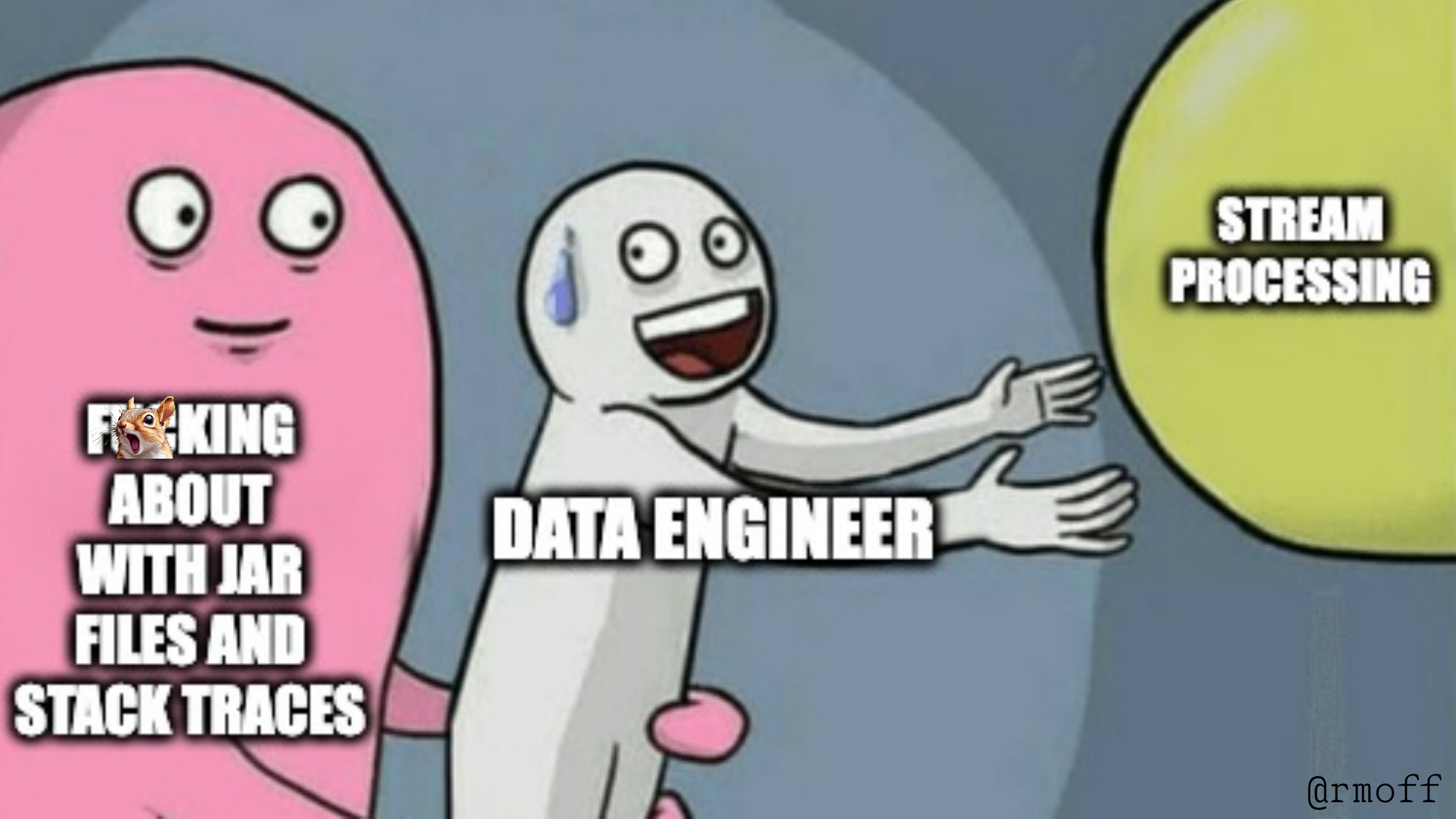
The Joy of JARs

and Other Flink SQL Troubleshooting Tales

Robin Moffatt, Principal DevEx Engineer @ Decodable



@rmoff


FKING
ABOUT
WITH JAR
FILES AND
STACK TRACES
DATA ENGINEER
STREAM
PROCESSING
@rmoff

Could not execute SQL statement.
Reason:

java.lang.**ClassNotFoundException**

```
org.apache.flink.core.fs.
UnsupportedFileSystemSchemeException:

Could not find a file system
implementation for scheme 's3'
```

Could not find any factory for identifier 'hive' that implements 'org.apache.flink.table.factories.CatalogFactory' in the classpath.


MY GOD
IT'S FULL OF JAVA

Troubleshooting

"…now I will jiggle things randomly until they unbreak" is not acceptable

—Linus Torvalds

https://libquotes.com/linus-torvalds/quote/lbr1k4j

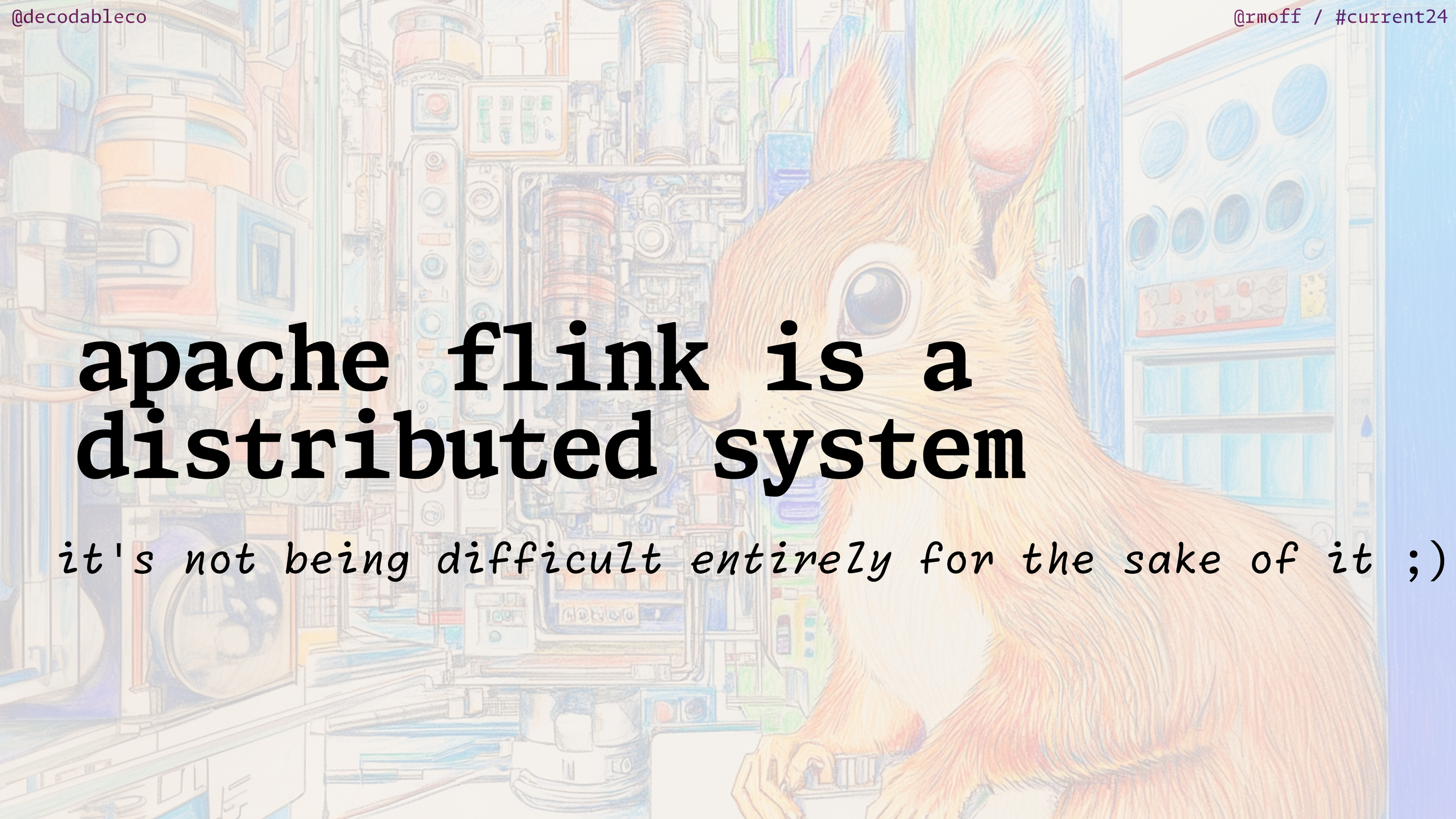
Troubleshooting Flink SQL

- What is the *root* error?
  - Java loves stack traces!
  - SQL Client often only gives you the top level view
- *Where* is the error coming from?
- Is what you're doing *supported*?
  - e.g. JDBC Catalog is read-only, some table formats don't support UPDATE, etc.

Things to always check

- Versions
  - Flink vs libraries
- Dependencies, e.g.
  - Required JARs
  - Java version

What Runs Where?

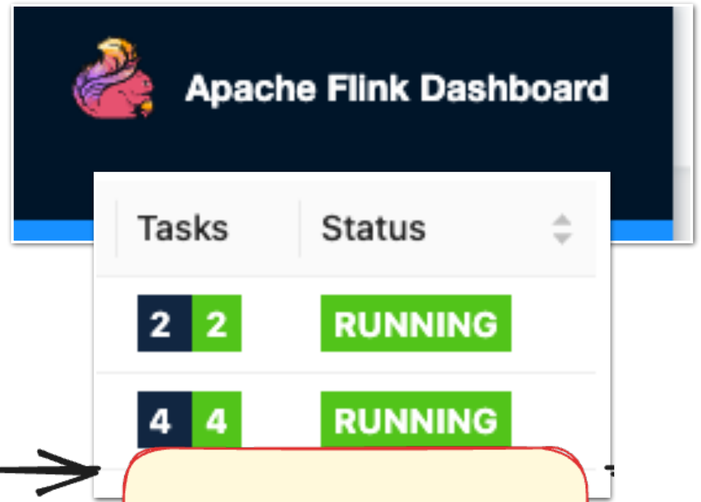


apache flink is a distributed system

it's not being difficult entirely for the sake of it ;)

Ref: https://nightlies.apache.org/flink/flink-docs-master/docs/deployment/overview/

Flink SQL>

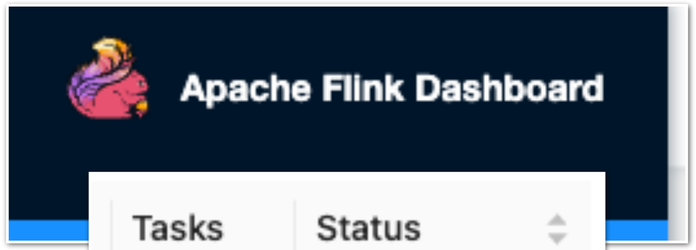


SQL Client

Job Manager

Ref: https://nightlies.apache.org/flink/flink-docs-master/docs/deployment/overview/

Flink SQL>



SQL Client

Job Manager

Task Manager

Ref: https://nightlies.apache.org/flink/flink-docs-master/docs/deployment/overview/

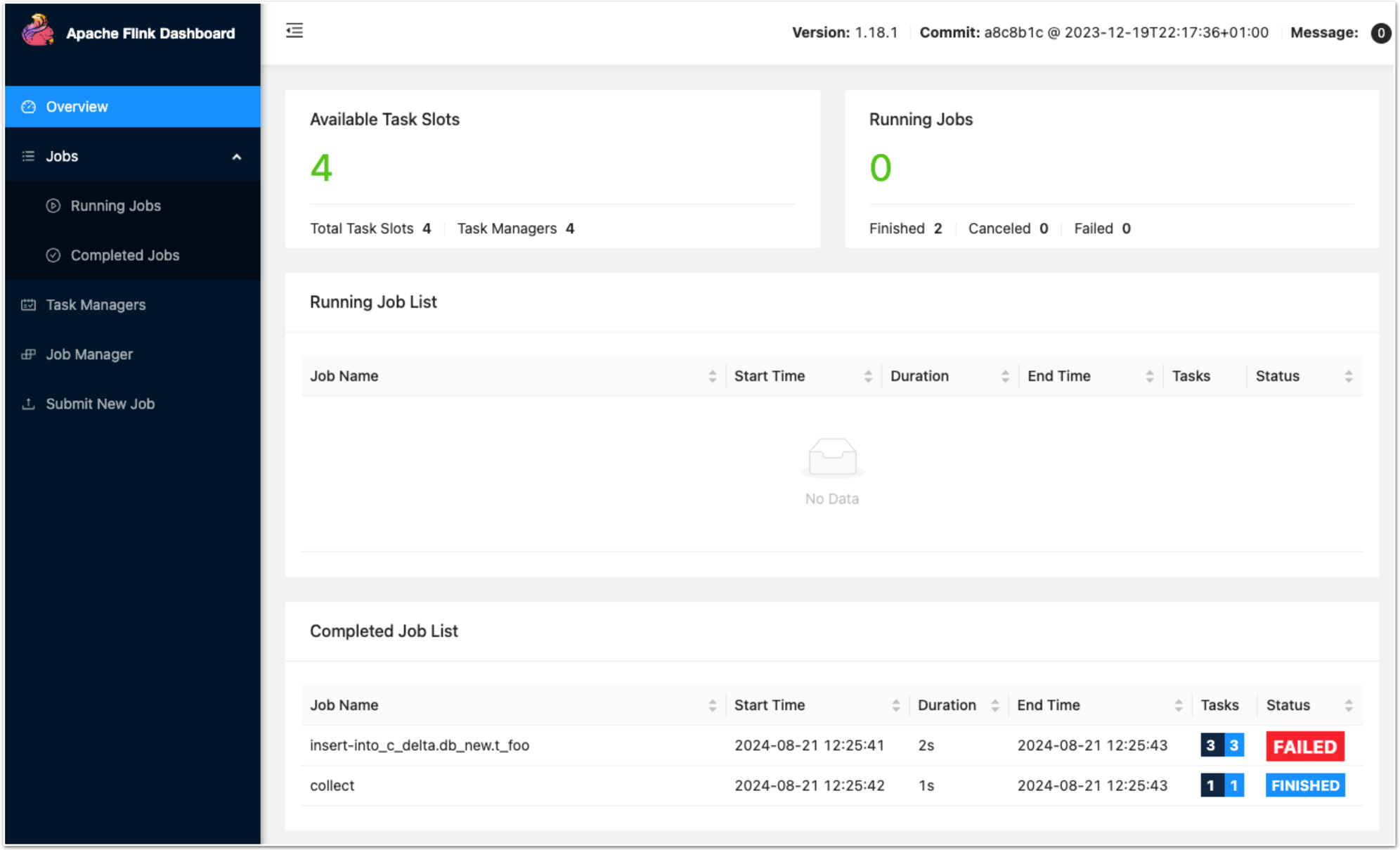
What went wrong?

```
Flink SQL> CREATE DATABASE 'db01';

[ERROR] Could not execute SQL statement.
Reason: MetaException(message:java.lang.RuntimeException:
java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found)
```

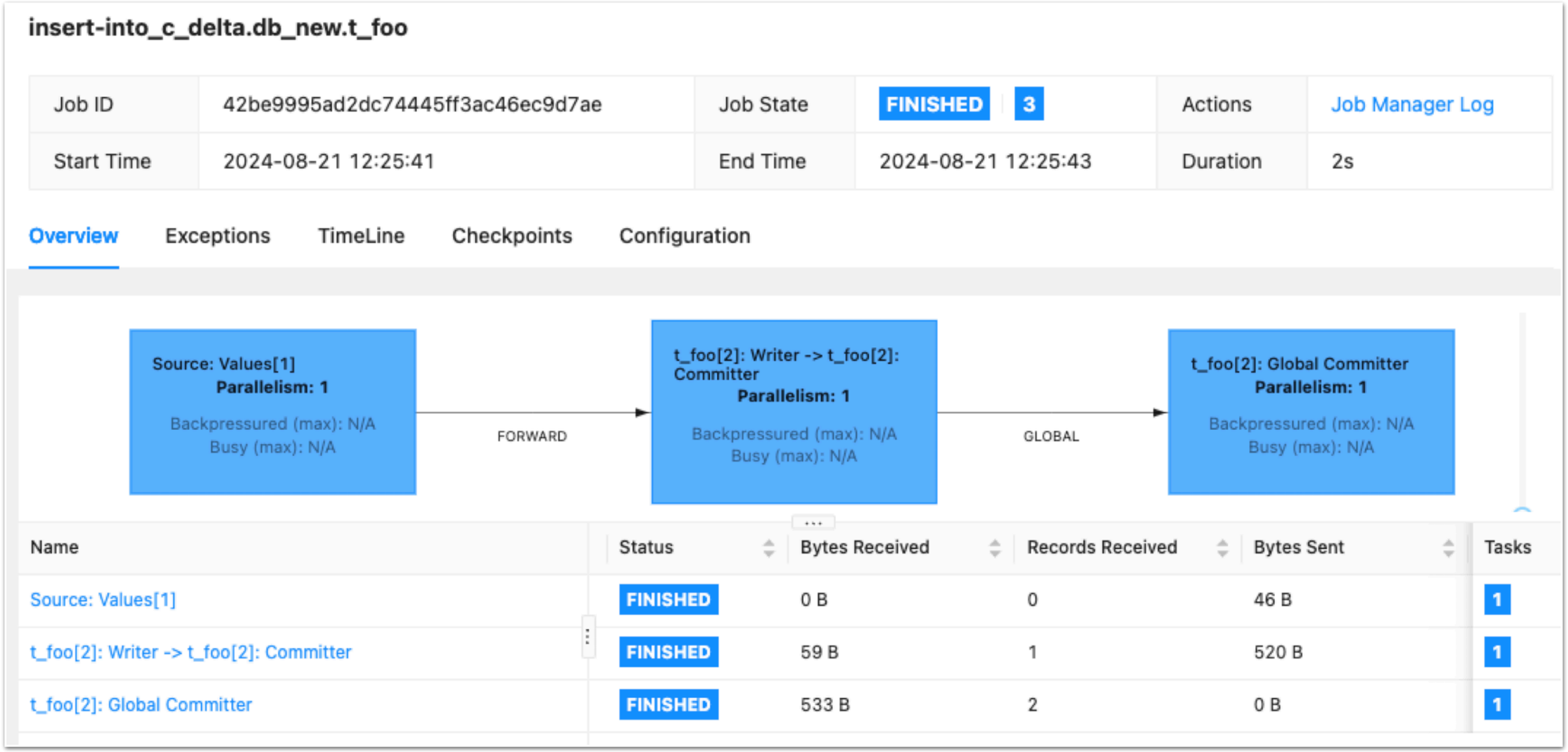
```
Flink SQL> INSERT INTO foo VALUES ('a',42);

[INFO] Submitting SQL update statement to the cluster...
[INFO] SQL update statement has been successfully submitted
       to the cluster:
Job ID: cc43d32a6bb0e2faab5270e542c70499
```

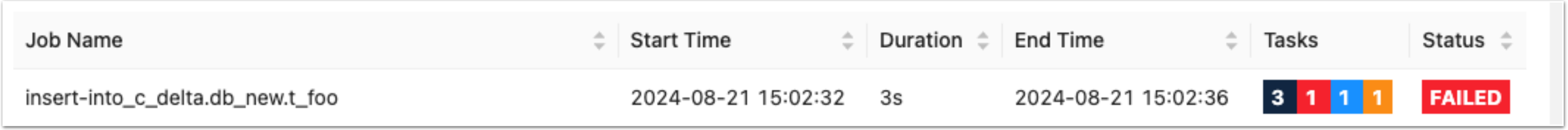
```
Flink SQL> INSERT INTO foo VALUES ('a',42);

[INFO] Submitting SQL update statement to the cluster...
[INFO] SQL update statement has been successfully submitted
       to the cluster:
Job ID: cc43d32a6bb0e2faab5270e542c70499

Flink SQL> SHOW JOBS;
+----------------------------------+----------------------+----------+-----
|                           job id |             job name |   status |
+----------------------------------+----------------------+----------+-----
| cc43d32a6bb0e2faab5270e542c70499 | insert-into_table | FAILED   | 2024
| 0b7f3f8cca2322de2366a1fd059fe5d9 |              collect | FINISHED | 2024
```

Flink Dashboard



Flink Dashboard





The truth is the log

h/t Pat Helland https://queue.acm.org/detail.cfm?id=2884038

The truth is the log

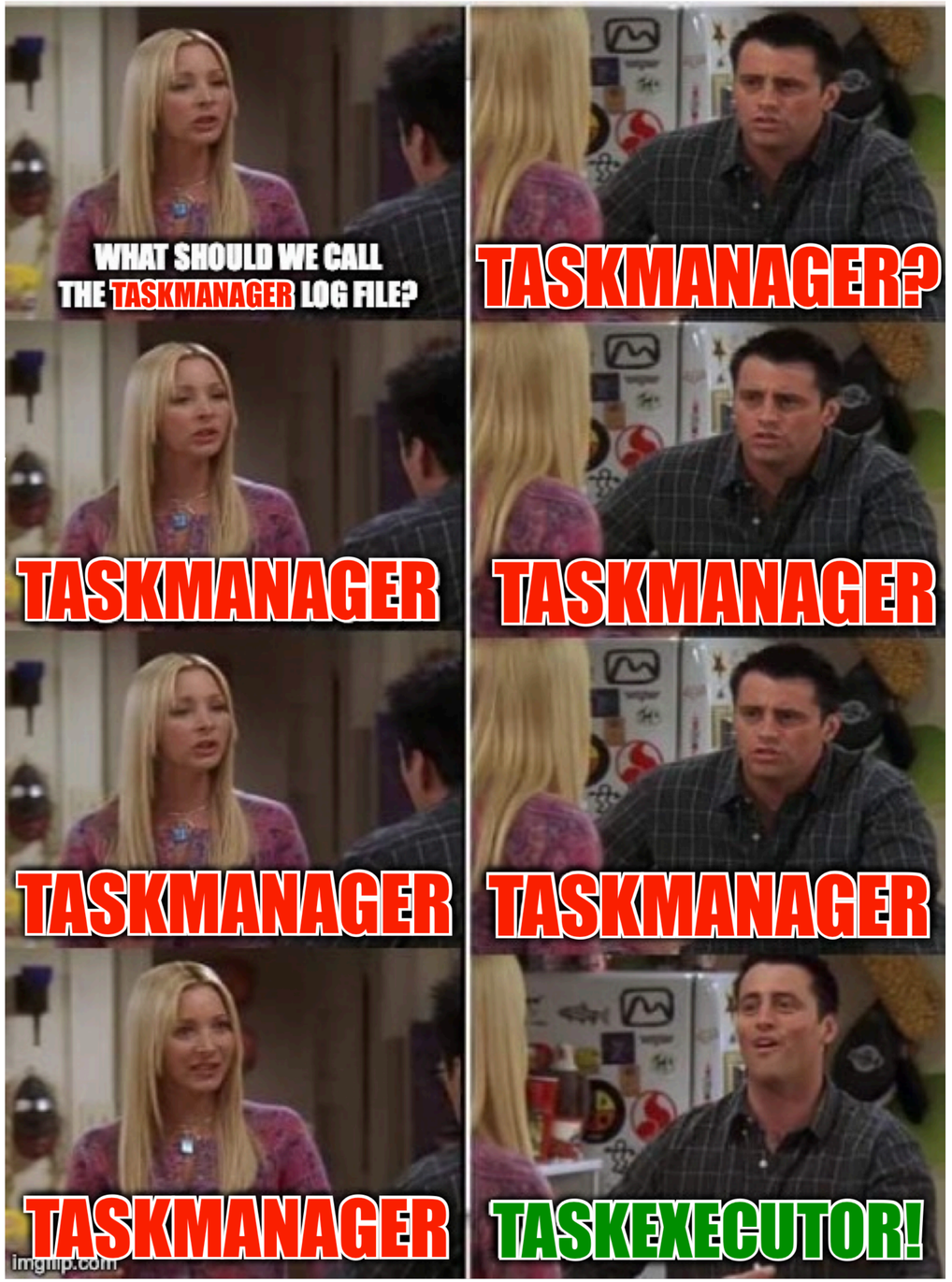
```
flink-1.20.0
└── log
    ├── flink-rmoff-sql-client-asgard08.log
    ├── flink-rmoff-standalonesession-1-asgard08.log
    ├── flink-rmoff-taskexecutor-1-asgard08.log
    └── flink-rmoff-taskexecutor-2-asgard08.log
```

The truth is the log

```
flink-1.20.0
└── log
    ├── flink-rmoff-sql-client-asgard08.log
    ├── flink-rmoff-standalonesession-1-asgard08.log
    ├── flink-rmoff-taskexecutor-1-asgard08.log
    └── flink-rmoff-taskexecutor-2-asgard08.log
```

Naming things is hard (apparently)

Naming things is hard (apparently)



Log file names

Configured through the log variable in the launch scripts

- SQL Client *(sql-client.sh)*:
  flink-$FLINK_IDENT_STRING-sql-client-$HOSTNAME.log
- Job Manager / Task Manager *(flink-daemon.sh)*
  flink-$FLINK_IDENT_STRING-$DAEMON-$id-$HOSTNAME.log

$USER *if not already set*

the command used to launch the Flink Component: ***standalonesession*** *or* ***taskexecutor***

the instance of the component (e.g. for two task managers this will be ***0*** *and* ***1****)*

Putting Flink in the ./bin

```
❯ ls bin
bash-java-utils.jar             migrate-config-file.sh
bash-java-utils.sh              pyflink-shell.sh
config-parser-utils.sh          sql-client.sh
config.sh                       sql-gateway.sh
find-flink-home.sh              standalone-job.sh
flink                           start-cluster.sh
flink-console.sh                start-zookeeper-quorum.sh
flink-daemon.sh                 stop-cluster.sh
historyserver.sh                stop-zookeeper-quorum.sh
jobmanager.sh                   taskmanager.sh
```

start-cluster.sh

jobmanager.sh start

TMWorkers start

config.sh

TMWorkers()

taskmanager.sh start

taskmanager.sh

jobmanager.sh

flink-daemon.sh

start-cluster.sh

jobmanager.sh start

TMWorkers start

config.sh

TMWorkers()

taskmanager.sh start

taskmanager.sh

flink-daemon.sh start taskexecutor

jobmanager.sh

flink-daemon.sh start standalonesession

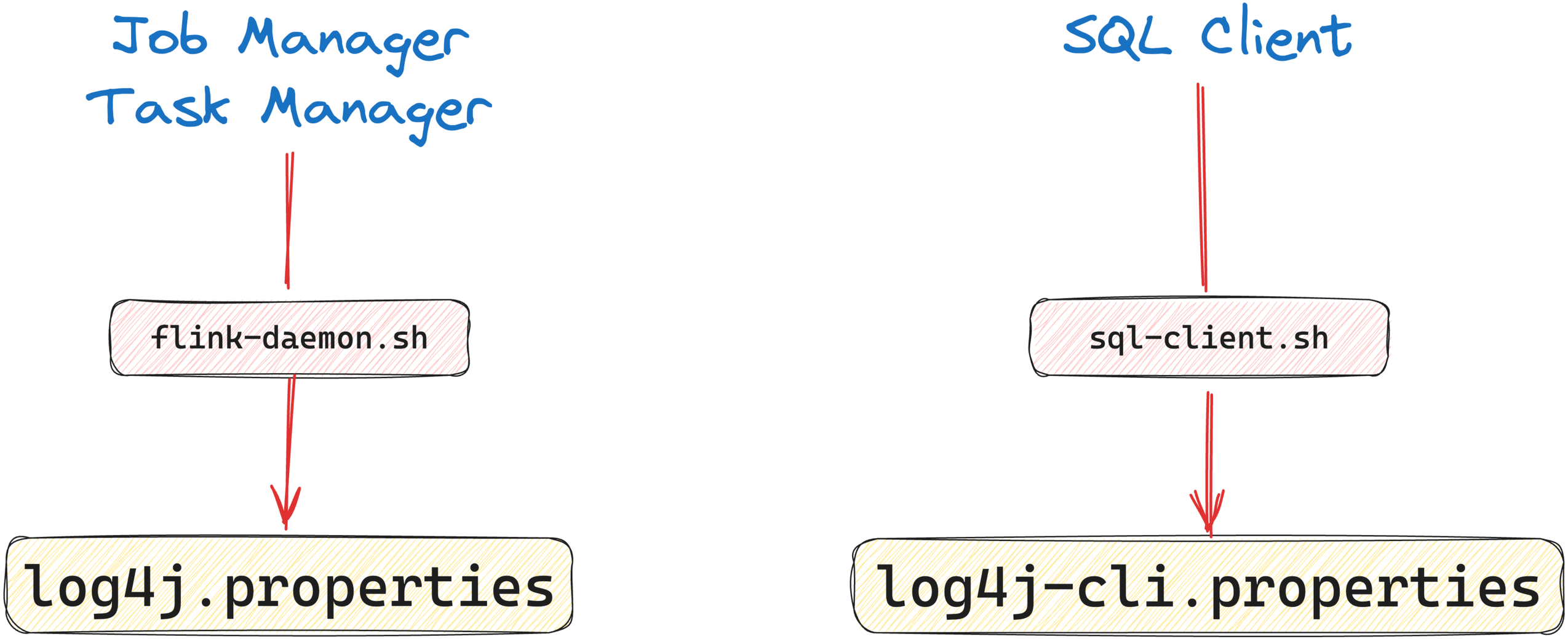
flink-daemon.sh

java $JVM_ARGS [...] org.apache.flink.runtime.taskexecutor.TaskManagerRunner

java $JVM_ARGS [...] org.apache.flink.runtime.entrypoint.StandaloneSessionClusterEntrypoint

Logging configuration (Log4j 2)

Which configuration file?



Log level

- ***Pre-1.19*** Update each **log4j configuration** directly:

  rootLogger.level = TRACE

- ***>= 1.19***: Set as a central configuration option in **Flink's config.yaml**:

  env.log.level: TRACE

Customising log level for specific components

```
logger.<id>.name = <logger name>
logger.<id>.level = <ERROR|WARN|INFO|TRACE|DEBUG>
```


```
logger.catalog.name = org.apache.flink.table.catalog
logger.catalog.level = TRACE
```

Log format

- Default is **PatternLayout**
- Others available including JSON, CSV, YAML


```
{
  "instant" : {
    "epochSecond" : 1725463441,
    "nanoOfSecond" : 471575000
  },
  "thread" : "flink-rest-client-netty-thread-1",
  "level" : "DEBUG",
  "loggerName" : "org.apache.flink.runtime.rest.RestClient",
  "message" : "Received response {\"properties\":{\"flink.hadoop.fs.s3a.secret.key\":\"hunter2\"
  "endOfBatch" : false,
  "loggerFqcn" : "org.apache.logging.slf4j.Log4jLogger",
  "threadId" : 19,
  "threadPriority" : 5
}
```

MOAR data plz

```
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Creating endpoint configuration for ""
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Using default endpoint -no need to generate a c
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] fs.s3a.endpoint.region="us-east-1"
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Using default endpoint; setting region to us-ea
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Creating endpoint configuration for "http://loc
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Endpoint http://localhost:9000 is not the defau
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Endpoint URI = http://localhost:9000
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Region for endpoint http://localhost:9000, URI
```

I can haz more data?

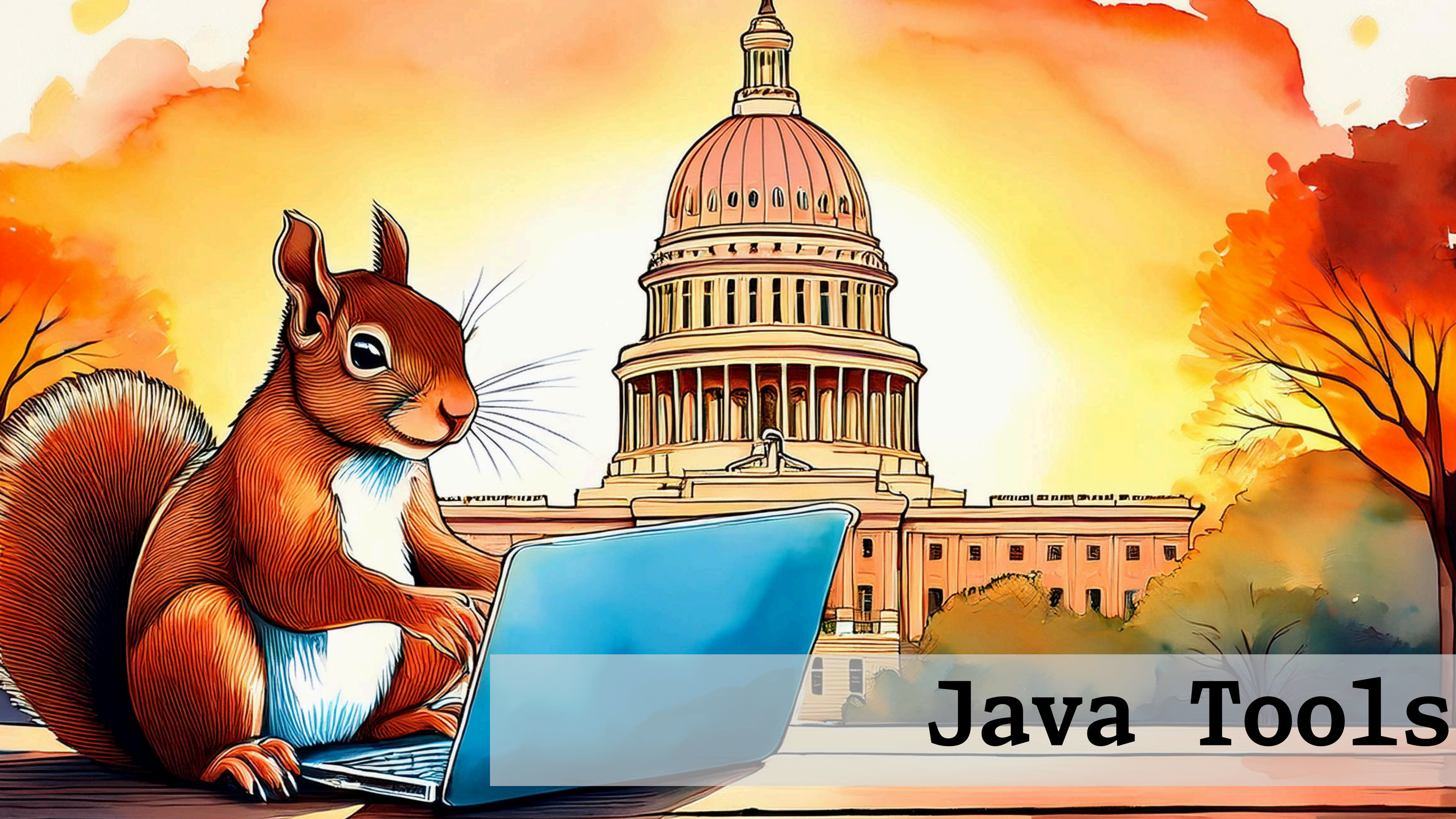
```
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Creating endpoint configuration for ""
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Using default endpoint -no need to generate a c
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] fs.s3a.endpoint.region="us-east-1"
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Using default endpoint; setting region to us-ea
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Creating endpoint configuration for "http://loc
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Endpoint http://localhost:9000 is not the defau
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Endpoint URI = http://localhost:9000
BUG org.apache.hadoop.fs.s3a.DefaultS3ClientFactory   [] Region for endpoint http://localhost:9000, URI
```

```
appender.main.layout.pattern =
  %d{yyyy-MM-dd HH:mm:ss,SSS} %-5p %-60c %x tid:%tid [%-60threadName] - %m %n
```

MOAR data!

```
ory    [] tid:79 [t_foo[2]: Writer -> t_foo[2]: Committer (1/1)#0 ] - Creating endpoint configuration fo
ory    [] tid:79 [t_foo[2]: Writer -> t_foo[2]: Committer (1/1)#0 ] - Using default endpoint -no need to
ory    [] tid:79 [t_foo[2]: Writer -> t_foo[2]: Committer (1/1)#0 ] - fs.s3a.endpoint.region="us-east-1"
ory    [] tid:79 [t_foo[2]: Writer -> t_foo[2]: Committer (1/1)#0 ] - Using default endpoint; setting re
ory    [] tid:80 [t_foo[2]: Global Committer (1/1)#0              ] - Creating endpoint configuration fo
ory    [] tid:80 [t_foo[2]: Global Committer (1/1)#0              ] - Endpoint http://localhost:9000 is
ory    [] tid:80 [t_foo[2]: Global Committer (1/1)#0              ] - Endpoint URI = http://localhost:900
ory    [] tid:80 [t_foo[2]: Global Committer (1/1)#0              ] - Region for endpoint http://localho
```

```
ayout.pattern =
-dd HH:mm:ss,SSS} %-5p %-60c %x tid:%tid [%-60threadName] - %m %n
```



Java Tools


ps -ef |
grep java

ps -ef |
grep java

jps

```
$ jps
84163 SqlClient
83671 StandaloneSessionClusterEntrypoint
83944 TaskManagerRunner
```

```
$ jinfo 84163
```

```
$ jinfo $(pgrep -f SqlClient)
```

$ jinfo $(pgrep -f SqlClient)

```
Java System Properties:
#Thu Sep 05 11:18:14 BST 2024
gopherProxySet=false
log4j.configuration=file\:/Users/rmoff/flink/flink-1.18.1/conf/log4j-cli.properties
awt.toolkit=sun.lwawt.macosx.LWCToolkit
java.specification.version=11
sun.cpu.isalist=
sun.jnu.encoding=UTF-8
java.vm.vendor=Eclipse Adoptium
sun.arch.data.model=64
java.vendor.url=https\://adoptium.net/
```

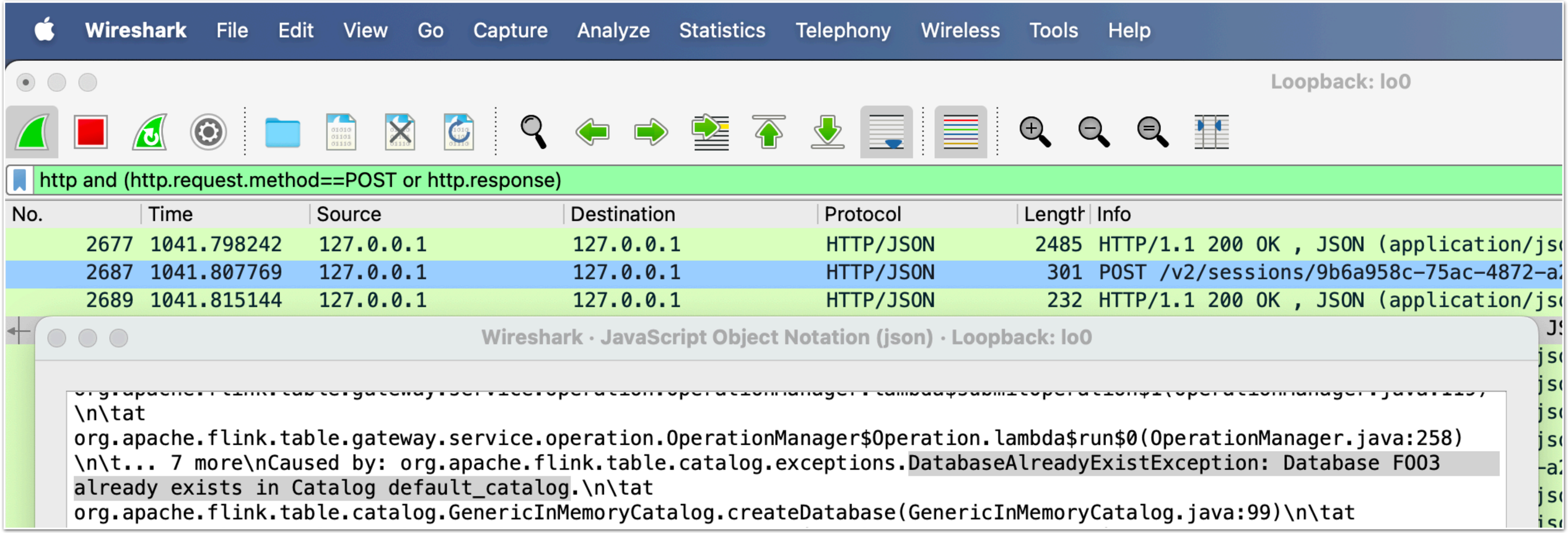
```
$ jinfo $(pgrep -f SqlClient)
[…]
VM Flags:
-XX:CICompilerCount=4 -XX:ConcGCThreads=2
-XX:G1ConcRefinementThreads=9 -XX:G1HeapRegionSize=4194304
-XX:GCDrainStackTargetSize=64 -XX:+IgnoreUnrecognizedVMOptions
-XX:InitialHeapSize=1073741824 -XX:MarkStackSize=4194304
-XX:MaxHeapSize=17179869184 -XX:MaxNewSize=10305404928
-XX:MinHeapDeltaBytes=4194304 -XX:NonNMethodCodeHeapSize=5836492
-XX:NonProfiledCodeHeapSize=122910874 -XX
:ProfiledCodeHeapSize=122910874 -XX:ReservedCodeCacheSize=251658240
-XX:+SegmentedCodeCache -XX:-UseAOT -XX:+UseCompressedClassPointers
-XX:+UseCompressedOops -XX:+UseG1GC
```

$ jinfo $(pgrep -f SqlClient)

```
[…]
VM Arguments:
java_command: org.apache.flink.table.client.SqlClient --jar /Users/
rmoff/flink/flink-1.18.1/opt/flink-sql-client-1.18.1.jar
java_class_path (initial): /Users/rmoff/flink/flink-1.18.1/lib/delta/
aws-java-sdk-bundle-1.12.648.jar:/Users/rmoff/flink/flink-1.18.1/lib/
delta/delta-flink-3.2.0.jar:/Users/rmoff/flink/flink-1.18.1/lib/
delta/delta-standalone_2.12-3.2.0.jar:/Users/rmoff/flink/
flink-1.18.1/lib/delta/delta-storage-3.2.0.jar:/Users/rmoff/flink/
flink-1.18.1/lib/delta/flink-sql-parquet-1.18.1.jar:/Users/rmoff/
flink/flink-1.18.1/lib/delta/hadoop-aws-3.3.4.jar:/Users/rmoff/flink/
flink-1.18.1/lib/delta/shapeless_2.12-2.3.4.jar:/Users/rmoff/flink/
```

```
$ jinfo $(pgrep -f SqlClient) \
    | grep java_class_path
java_class_path (initial): /Users/rmoff/flink/flink-1.18.1/lib/delta/
aws-java-sdk-bundle-1.12.648.jar:/Users/rmoff/flink/flink-1.18.1/lib/
delta/delta-flink-3.2.0.jar:/Users/rmoff/flink/flink-1.18.1/lib/
delta/delta-standalone_2.12-3.2.0.jar:/Users/rmoff/flink/
flink-1.18.1/lib/delta/delta-storage-3.2.0.jar:/Users/rmoff/flink/
flink-1.18.1/lib/delta/flink-sql-parquet-1.18.1.jar:/Users/rmoff/
flink/flink-1.18.1/lib/delta/hadoop-aws-3.3.4.jar:/Users/rmoff/flink/
flink-1.18.1/lib/delta/shapeless_2.12-2.3.4.jar:/Users/rmoff/flink/
flink-1.18.1/lib/flink-cep-1.18.1.jar:/Users/rmoff/flink/
flink-1.18.1/lib/flink-connector-files-1.18.1.jar:/Users/rmoff/flink/
flink-1.18.1/lib/flink-csv-1.18.1.jar:/Users/rmoff/flink/
```

🕵️ Wireshark 🦈


Wireshark File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Loopback: lo0
http and (http.request.method==POST or http.response)
No. Time Source Destination Protocol Length Info
2677 1041.798242 127.0.0.1 127.0.0.1 HTTP/JSON 2485 HTTP/1.1 200 OK , JSON (application/js
2687 1041.807769 127.0.0.1 127.0.0.1 HTTP/JSON 301 POST /v2/sessions/9b6a958c-75ac-4872-a
2689 1041.815144 127.0.0.1 127.0.0.1 HTTP/JSON 232 HTTP/1.1 200 OK , JSON (application/js
Wireshark · JavaScript Object Notation (json) · Loopback: lo0
\n\tat
org.apache.flink.table.gateway.service.operation.OperationManager$Operation.lambda$run$0(OperationManager.java:258)
\n\t... 7 more\nCaused by: org.apache.flink.table.catalog.exceptions.DatabaseAlreadyExistException: Database FOO3
already exists in Catalog default_catalog.\n\tat
org.apache.flink.table.catalog.GenericInMemoryCatalog.createDatabase(GenericInMemoryCatalog.java:99)\n\tat

tshark

```
Flink SQL> SHOW TABLES;

Empty set
```

```
$ tshark -i lo -Y "http and (http.request.method==POST or http.response)"

Capturing on 'Loopback: lo0'

127.0.0.1 → 127.0.0.1    HTTP/JSON 2263 HTTP/1.1 200 OK , JSON
(application/json)

127.0.0.1 → 127.0.0.1    HTTP/JSON 292 POST /v2/sessions/
8827d93e-0deb-43b5-90ab-8d1c5bce8bd1/statements HTTP/1.1 , JSON
(application/json)

127.0.0.1 → 127.0.0.1    HTTP/JSON 232 HTTP/1.1 200 OK , JSON
(application/json)
```

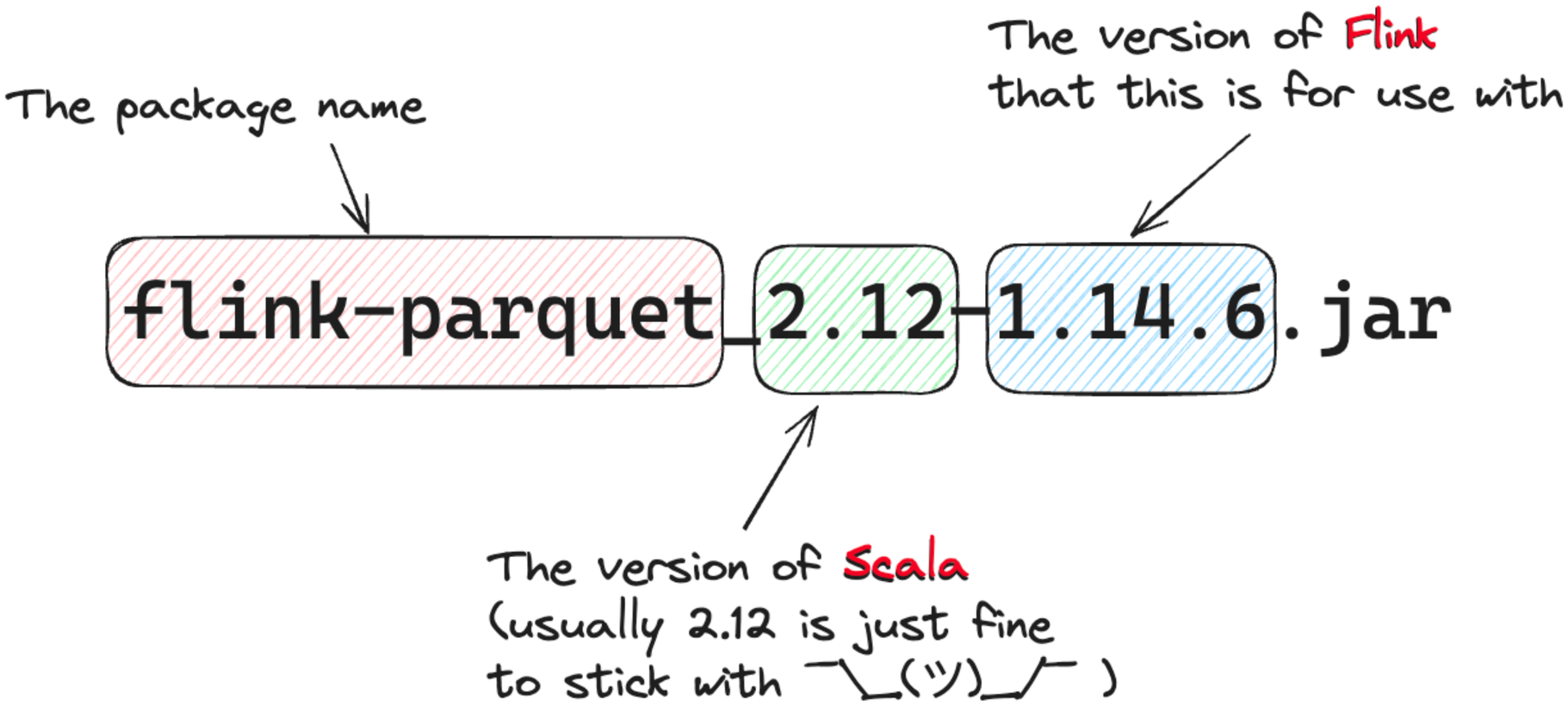
Could not execute SQL statement.
Reason:

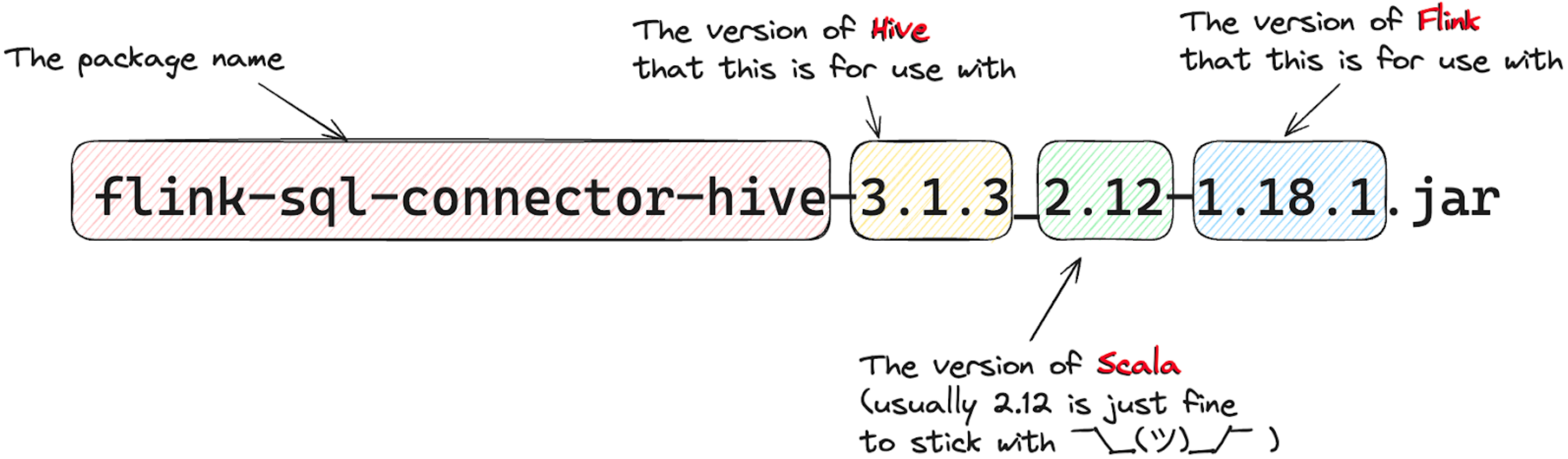
java.lang.**ClassNotFoundException**

Finding JARs

Usually the docs will tell you which JAR you need.

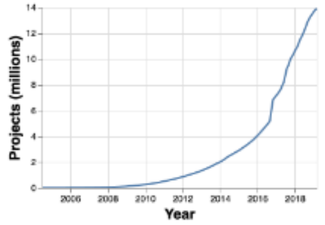
JARs are very specific to the versions of the tools that you're using.




The package name
The version of Hive that this is for use with
The version of Flink that this is for use with
flink-sql-connector-hive-3.1.3_2.12-1.18.1.jar
The version of Scala (usually 2.12 is just fine to stick with ¯_(ツ)_/¯)

MVN REPOSITORY | Search for groups, artifacts, categ | Search | Categories | Popular | Contact Us

Indexed Artifacts (38.1M)



Popular Categories

- Testing Frameworks & Tools
- Android Packages
- Logging Frameworks
- Java Specifications
- JSON Libraries
- JVM Languages
- Language Runtime
- Core Utilities
- Mocking
- Web Assets
- Annotation Libraries
- HTTP Clients
- Logging Bridges
- Dependency Injection
- XML Processing
- Web Frameworks

Home » org.apache.flink » flink-parquet » 1.18.1

Flink : Formats : Parquet » 1.18.1

Flink : Formats : Parquet

License	Apache 2.0
Tags	parquet flink serialization apache column
Date	Jan 16, 2024
Files	pom (17 KB) jar (174 KB) View All
Repositories	Central
Ranking	**#9617 in MvnRepository (See Top Artifacts)**
Used By	**42 artifacts**
Vulnerabilities	**Vulnerabilities from dependencies:** CVE-2023-2976 CVE-2022-26612 CVE-2020-8908

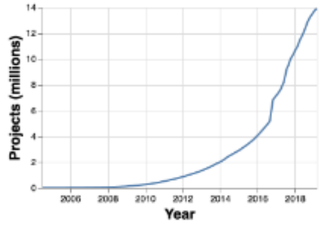
Maven | Gradle | Gradle (Short) | Gradle (Kotlin) | SBT | Ivy | Grape | Leiningen | Buildr

```
<!-- https://mvnrepository.com/artifact/org.apache.flink/flink-parquet
-->
<dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-parquet</artifactId>
    <version>1.18.1</version>
    <scope>provided</scope>
```

☑ Include comment with link to declaration

MVN REPOSITORY | Search for groups, artifacts, categ | Search | Categories | Popular | Contact Us

Indexed Artifacts (38.1M)



Popular Categories

- Testing Frameworks & Tools
- Android Packages
- Logging Frameworks
- Java Specifications
- JSON Libraries
- JVM Languages
- Language Runtime
- Core Utilities
- Mocking
- Web Assets
- Annotation Libraries
- HTTP Clients
- Logging Bridges
- Dependency Injection
- XML Processing
- Web Frameworks

Home » org.apache.flink » flink-parquet » 1.18.1

Flink : Formats : Parquet » 1.18.1

Flink : Formats : Parquet

License	Apache 2.0
Tags	parquet flink serializatio apache column
Date	Jan 16, 2024
Files	pom (17 KB) jar 194 KB) View All
Repositories	Central
Ranking	#9617 in MvnRepository (See Top Artifacts)
Used By	42 artifacts
Vulnerabilities	Vulnerabilities from dependencies: CVE-2023-2976 CVE-2022-26612 CVE-2020-8908

Maven | Gradle | Gradle (Short) | Gradle (Kotlin) | SBT | Ivy | Grape | Leiningen | Buildr

```
<!-- https://mvnrepository.com/artifact/org.apache.flink/flink-parquet
-->
<dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-parquet</artifactId>
    <version>1.18.1</version>
    <scope>provided</scope>
```

☑ Include comment with link to declaration

Put JARs in the ./lib folder

```
flink-1.20.0
 ├──── lib
 |     ├──── ext
 |     |     └──── flink-sql-parquet-1.20.0.jar
 |     ├──── flink-cep-1.20.0.jar
 |     ├──── flink-connector-files-1.20.0.jar
 |     ├──── flink-csv-1.20.0.jar
 |     ├──── flink-dist-1.20.0.jar
 |     ├──── flink-json-1.20.0.jar
 |     ├──── flink-scala_2.12-1.20.0.jar
```

Where does Flink look for JARs?

- Flink is invoked with a **classpath** argument that's built from **constructFlinkClassPath** in config.sh, which scans the **./lib** folder

```
java_class_path (initial): /Users/rmoff/flink/flink-1.20.0/lib/ext/flink-sql-parquet-1.20.0.jar:/
Users/rmoff/flink/flink-1.20.0/lib/flink-cep-1.20.0.jar:/Users/rm
off/flink/flink-1.20.0/lib/flink-connector-files-1.20.0.jar:/Users/rmoff/flink/flink-1.20.0/lib/
flink-csv-1.20.0.jar:/Users/rmoff/flink/flink-1.20.0/lib/flink-jso
n-1.20.0.jar:/Users/rmoff/flink/flink-1.20.0/lib/flink-scala_2.12-1.20.0.jar:/Users/rmoff/flink/
flink-1.20.0/lib/flink-table-api-java-uber-1.20.0.jar:/Users/rmoff
/flink/flink-1.20.0/lib/flink-table-planner-loader-1.20.0.jar:/Users/rmoff/flink/flink-1.20.0/
lib/flink-table-runtime-1.20.0.jar:/Users/rmoff/flink/flink-1.20.0/l
ib/log4j-1.2-api-2.17.1.jar:/Users/rmoff/flink/flink-1.20.0/lib/log4j-api-2.17.1.jar:/Users/
rmoff/flink/flink-1.20.0/lib/log4j-core-2.17.1.jar:/Users/rmoff/flink/
flink-1.20.0/lib/log4j-slf4j-impl-2.17.1.jar:/Users/rmoff/flink/flink-1.20.0/lib/flink-
dist-1.20.0.jar::::
```

Hadoop classpath

```
$ export HADOOP_CLASSPATH=$(/opt/hadoop/bin/hadoop classpath)
```

(or just cherry-pick the necessary JARs directly into Flink's ./lib folder)

Don't forget to restart!

What's inside a JAR?

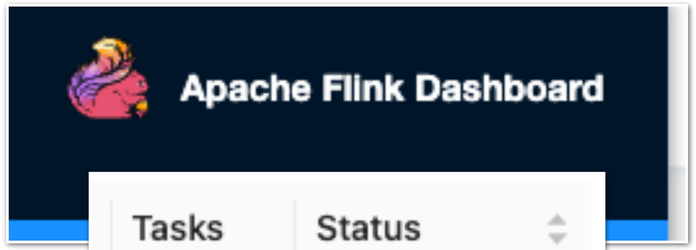
```
$ jar tf flink-sql-parquet-1.20.0.jar
org/apache/flink/formats/parquet/ParquetColumnarRowInputFormat$ColumnarRowReaderBatch.class
org/apache/flink/formats/parquet/ParquetColumnarRowInputFormat$1.class
org/apache/flink/formats/parquet/ParquetWriterFactory.class
org/apache/flink/formats/parquet/ParquetVectorizedInputFormat$1.class
org/apache/flink/formats/parquet/ParquetFileFormatFactory$ParquetBulkDecodingFormat.class
org/apache/flink/formats/parquet/ParquetVectorizedInputFormat$ParquetReaderBatch.class
org/apache/flink/formats/parquet/ParquetBuilder.class
org/apache/flink/formats/parquet/ParquetFileFormatFactory.class
org/apache/flink/formats/parquet/ParquetVectorizedInputFormat$ParquetReader.class
org/apache/flink/formats/parquet/ParquetFileFormatFactory$1.class
org/apache/flink/formats/parquet/utils/SerializableConfiguration.class
org/apache/flink/formats/parquet/utils/
ParquetFormatStatisticsReportUtil$ParquetFileRowCountCalculator.class
org/apache/flink/formats/parquet/utils/ParquetFormatStatisticsReportUtil.class
org/apache/flink/formats/parquet/utils/ParquetSchemaConverter.class
org/apache/flink/formats/parquet/utils/ParquetFormatStatisticsReportUtil$1.class
```

What got loaded from where?

```
$ FLINK_ENV_JAVA_OPTS="-verbose:class" ./bin/sql-client.sh
[20.264s][info][class,load]
org.apache.flink.table.connector.format.ProjectableDecodingFormat source: file:/
Users/rmoff/flink/flink-1.20.0/lib/flink-table-api-java-uber-1.20.0.jar
[20.264s][info][class,load]
org.apache.flink.table.connector.format.FileBasedStatisticsReportableInputFormat
source: file:/Users/rmoff/flink/flink-1.20.0/lib/flink-table-api-java-
uber-1.20.0.jar
[20.264s][info][class,load]
org.apache.flink.formats.parquet.ParquetFileFormatFactory$ParquetBulkDecodingForm
at source: file:/Users/rmoff/flink/flink-1.20.0/lib/ext/flink-sql-
parquet-1.20.0.jar
[20.264s][info][class,load]
org.apache.flink.table.factories.FactoryUtil$TableFactoryHelper$
$Lambda$971/0x0000000800779c40 source:
```


really
What Runs Where?
JOB MANAGER
SQL Cloud
SQL
Task Manager

Flink SQL>

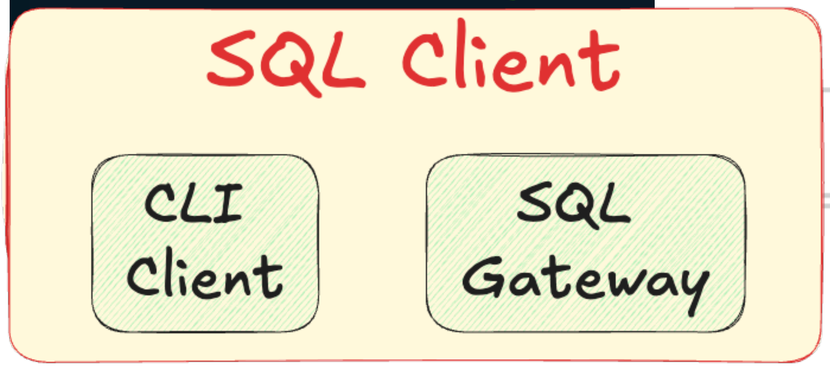


SQL Client

Job Manager

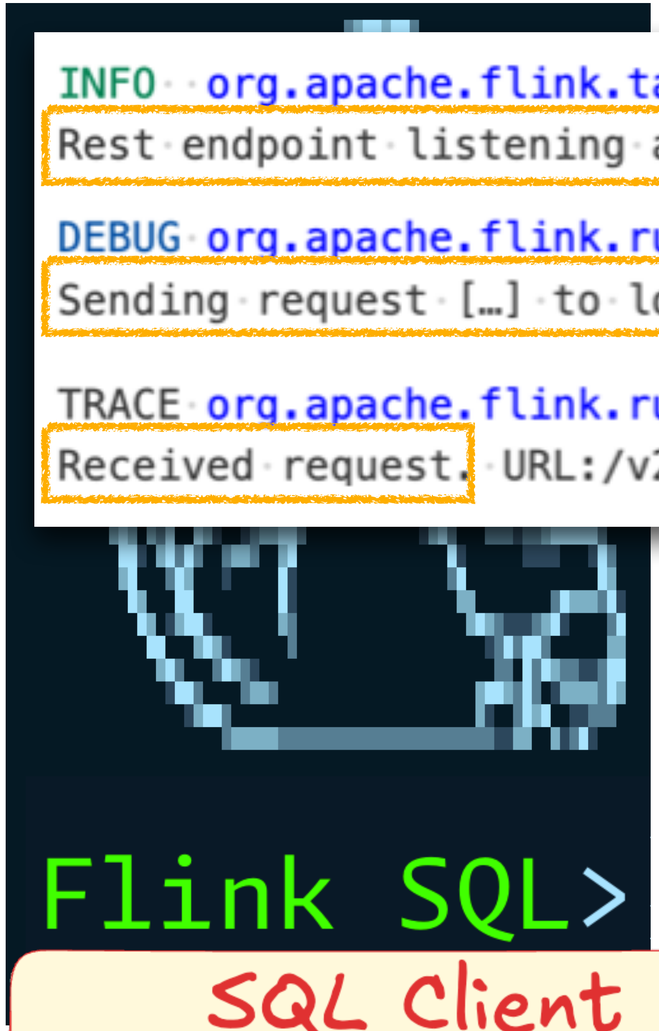
Task Manager

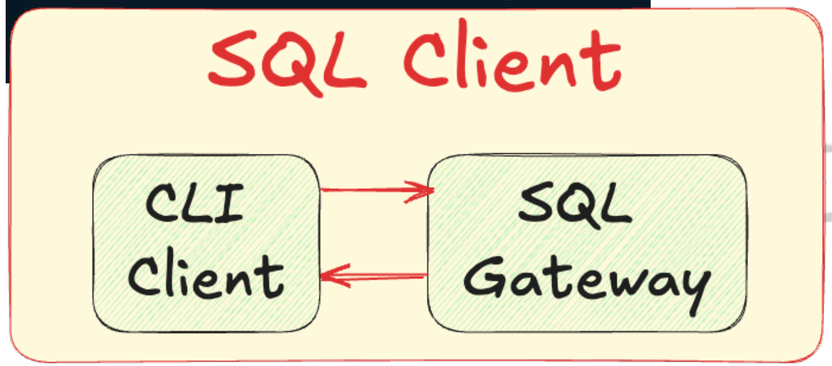
Ref: https://nightlies.apache.org/flink/flink-docs-master/docs/deployment/overview/



Ref: https://cwiki.apache.org/confluence/display/FLINK/FLIP-24:+SQL+Client

```
INFO  org.apache.flink.table.gateway.rest.SqlGatewayRestEndpoint       [] -
Rest endpoint listening at localhost:64934

DEBUG org.apache.flink.runtime.rest.RestClient                         [] -
Sending request […] to localhost:64934/v2/sessions/ba3b2ff9-e35f-4e13-8184-f94f57271e3e/statements

TRACE org.apache.flink.runtime.rest.FileUploadHandler                  [] -
Received request. URL:/v2/sessions/ba3b2ff9-e35f-4e13-8184-f94f57271e3e/statements Method:POST
```





Ref: https://cwiki.apache.org/confluence/display/FLINK/FLIP-24:+SQL+Client

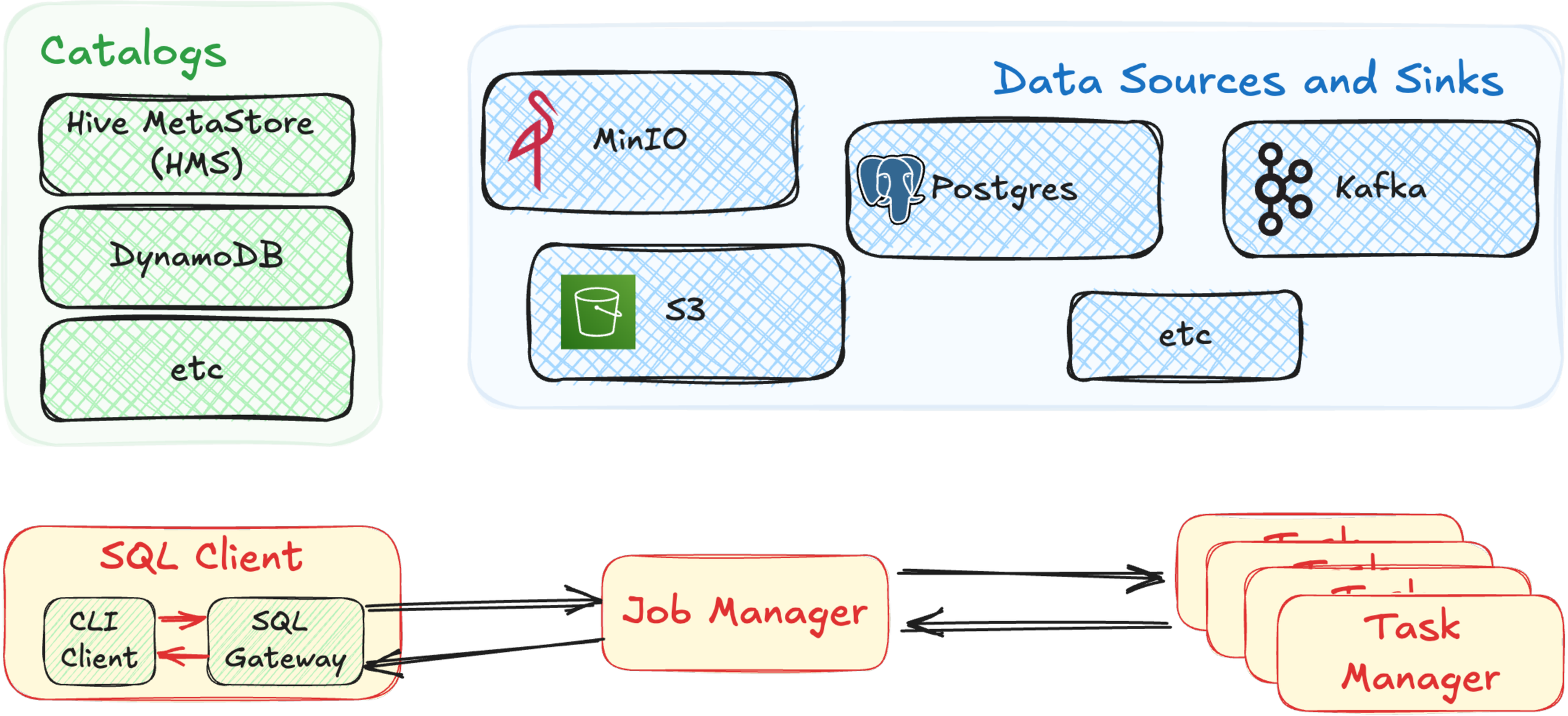
tshark

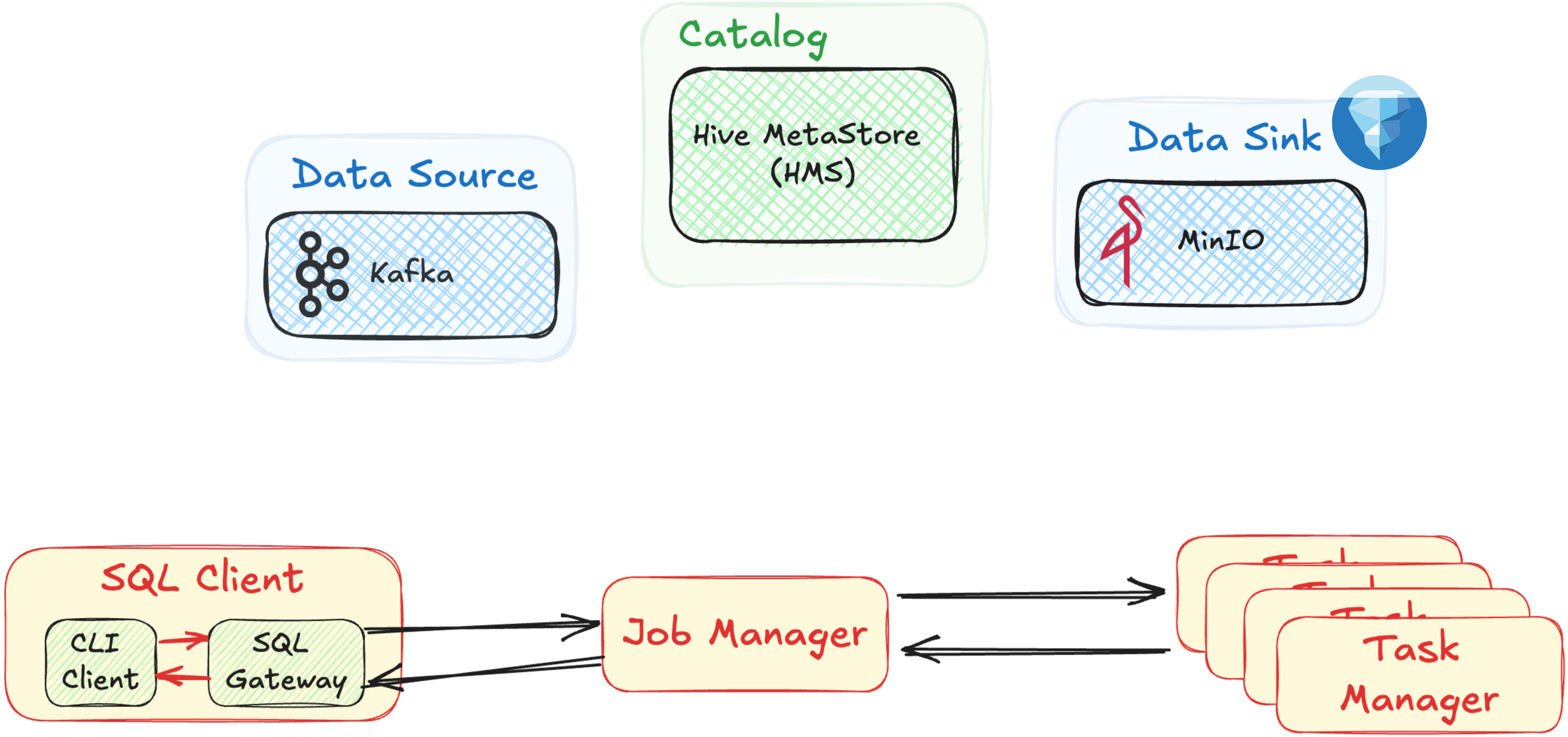
```
$ tshark -i lo \
         -Y "http and (http.request.method==POST
                       or http.response)" \
         -Tjson
```

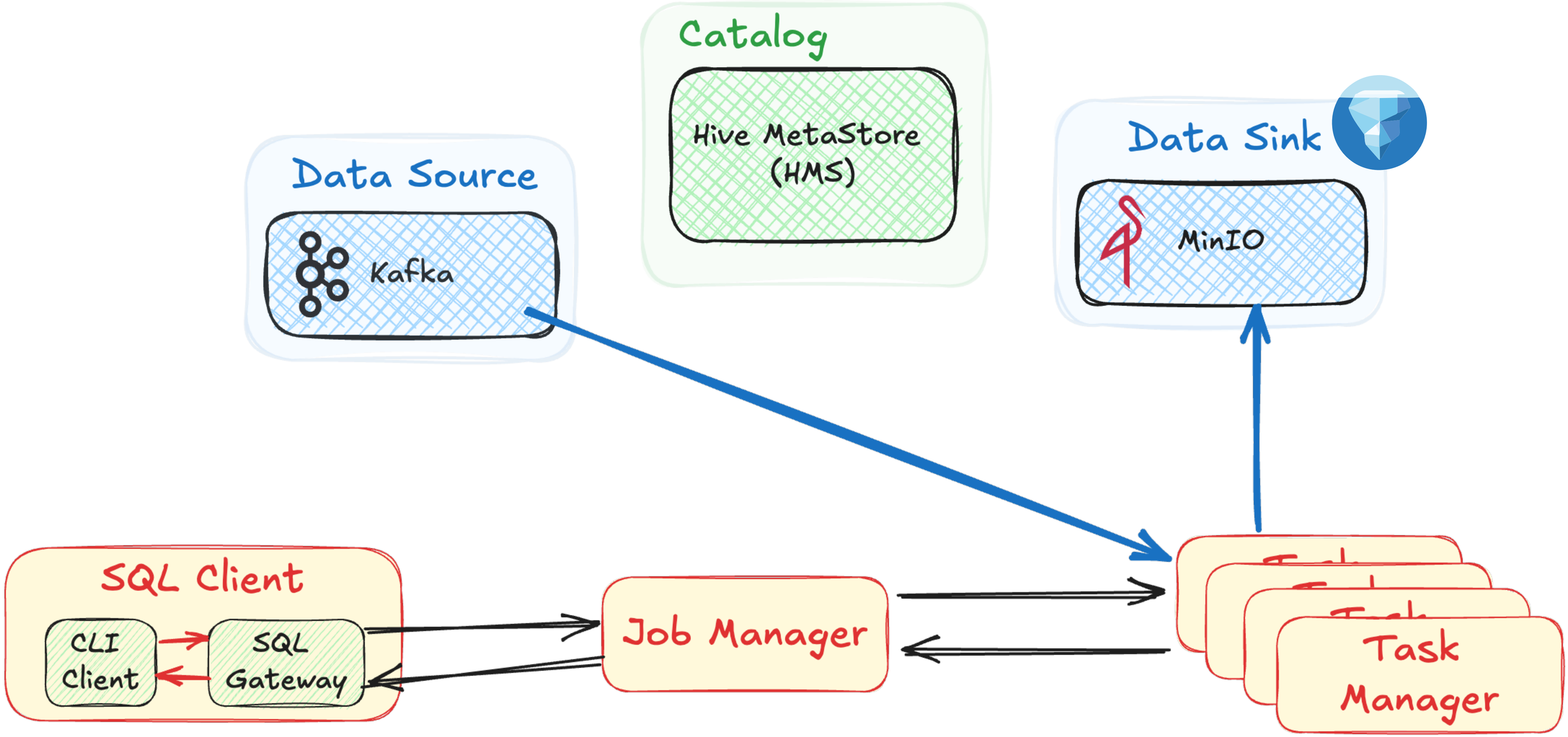

```
      "json": {
        "json.object": {
          "json.member": "statement",
          "json.member_tree": {
            "json.path_with_value": "/statement:SHOW TABLES;",
            "json.member_with_value": "statement:SHOW TABLES;",
            "json.value.string": "SHOW TABLES;",
            "json.key": "statement",
            "json.path": "/statement"
          },
```

what about external stuff?


SQL Client
CLI Client
SQL Gateway
Job Manager
Task Manager


Catalogs
Hive MetaStore (HMS)
DynamoDB
etc
Data Sources and Sinks
MinIO
Postgres
Kafka
S3
etc
SQL Client
CLI Client
SQL Gateway
Job Manager
Task Manager


Catalog
Hive MetaStore (HMS)
Data Source
Kafka
Data Sink
MinIO
SQL Client
CLI Client
SQL Gateway
Job Manager
Task Manager


Catalog
Hive MetaStore (HMS)
Data Source
Kafka
Data Sink
MinIO
SQL Client
CLI Client
SQL Gateway
Job Manager
Task Manager

Uh oh...

```
Flink SQL> CREATE CATALOG c_iceberg_hive WITH (
>             'type' = 'iceberg',
>             'warehouse' = 's3a://warehouse',
>             'catalog-type'='hive',
>             'uri'='thrift://hms:9083');

[INFO] Execute statement succeed.

Flink SQL> USE CATALOG c_iceberg_hive;
[INFO] Execute statement succeed.

Flink SQL> CREATE DATABASE `c_iceberg_hive`.`db01`;
[ERROR] Could not execute SQL statement. Reason:
MetaException(message:java.lang.RuntimeException: java.lang.ClassNotFoundException:
 Class org.apache.hadoop.fs.s3a.S3AFileSystem not found)
```

💩 rises to the top…

```
Flink SQL> CREATE CATALOG c_iceberg_hive WITH (
>             'type' = 'iceberg',
>             'warehouse' = 's3a://warehouse',
>             'catalog-type'='hive',
>             'uri'='thrift://hms:9083');

[INFO] Execute statement succeed.

Flink SQL> USE CATALOG c_iceberg_hive;
[INFO] Execute statement succeed.

Flink SQL> CREATE DATABASE `c_iceberg_hive`.`db01`;
[ERROR] Could not execute SQL statement. Reason:
MetaException(message:java.lang.RuntimeException: java.lang.ClassNotFoundException:
 Class org.apache.hadoop.fs.s3a.S3AFileSystem not found)
```

💩 rises to the top…

```
Flink SQL> CREATE CATALOG c_iceberg_hive WITH (
>             'type' = 'iceberg',
>             'warehouse' = 's3a://warehouse',
>             'catalog-type'='hive',
>             'uri'='thrift://hms:9083');

[INFO] Execute statement succeed.

Flink SQL> USE CATALOG c_iceberg_hive;
[INFO] Execute statement succeed.

Flink SQL> CREATE DATABASE `c_iceberg_hive`.`db01`;
[ERROR] Could not execute SQL statement. Reason:
MetaException(message:java.lang.RuntimeException: java.lang.ClassNotFoundException:
 Class org.apache.hadoop.fs.s3a.S3AFileSystem not found)
```


THAT'S FINE
CLASSNOTFOUND
WE CAN FIX THAT
imgflip.com

$

```
$ ls -l lib/aws
total 362080
-rw-r--r-- 1 flink flink 369799698 Sep  9 14:35 aws-java-sdk-bundle-1.12.648.jar
-rw-r--r-- 1 flink flink    962685 Sep  9 14:35 hadoop-aws-3.3.4.jar
```

```
$ ls -l lib/aws
total 362080
-rw-r--r-- 1 flink flink 369799698 Sep  9 14:35 aws-java-sdk-bundle-1.12.648.jar
-rw-r--r-- 1 flink flink    962685 Sep  9 14:35 hadoop-aws-3.3.4.jar
$ jar tf lib/aws/hadoop-aws-3.3.4.jar | grep S3AFileSystem.class
org/apache/hadoop/fs/s3a/S3AFileSystem.class
$
```

```
$ ls -l lib/aws
total 362080
-rw-r--r-- 1 flink flink 369799698 Sep  9 14:35 aws-java-sdk-bundle-1.12.648.jar
-rw-r--r-- 1 flink flink    962685 Sep  9 14:35 hadoop-aws-3.3.4.jar
$ jar tf lib/aws/hadoop-aws-3.3.4.jar | grep S3AFileSystem.class
org/apache/hadoop/fs/s3a/S3AFileSystem.class
$
```

$ more log/flink--sql-client-asgard08.log

```
2024-09-12 11:20:06,494 ERROR org.apache.flink.table.gateway.service.operation.Operati
org.apache.flink.table.api.TableException: Could not execute CREATE DATABASE: (catalog
        at org.apache.flink.table.operations.ddl.CreateDatabaseOperation.execute(Creat
        at org.apache.flink.table.api.internal.TableEnvironmentImpl.executeInternal(Ta
[…]
Caused by: java.lang.RuntimeException: Failed to create namespace db01 in Hive Metasto
        at org.apache.iceberg.hive.HiveCatalog.createNamespace(HiveCatalog.java:299) ~
        at org.apache.iceberg.flink.FlinkCatalog.createDatabase(FlinkCatalog.java:223)
        at org.apache.iceberg.flink.FlinkCatalog.createDatabase(FlinkCatalog.java:214)
        at org.apache.flink.table.catalog.CatalogManager.createDatabase(CatalogManager
        at org.apache.flink.table.operations.ddl.CreateDatabaseOperation.execute(Creat
        ... 14 more
Caused by: org.apache.hadoop.hive.metastore.api.MetaException:
java.lang.RuntimeException: java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found
```

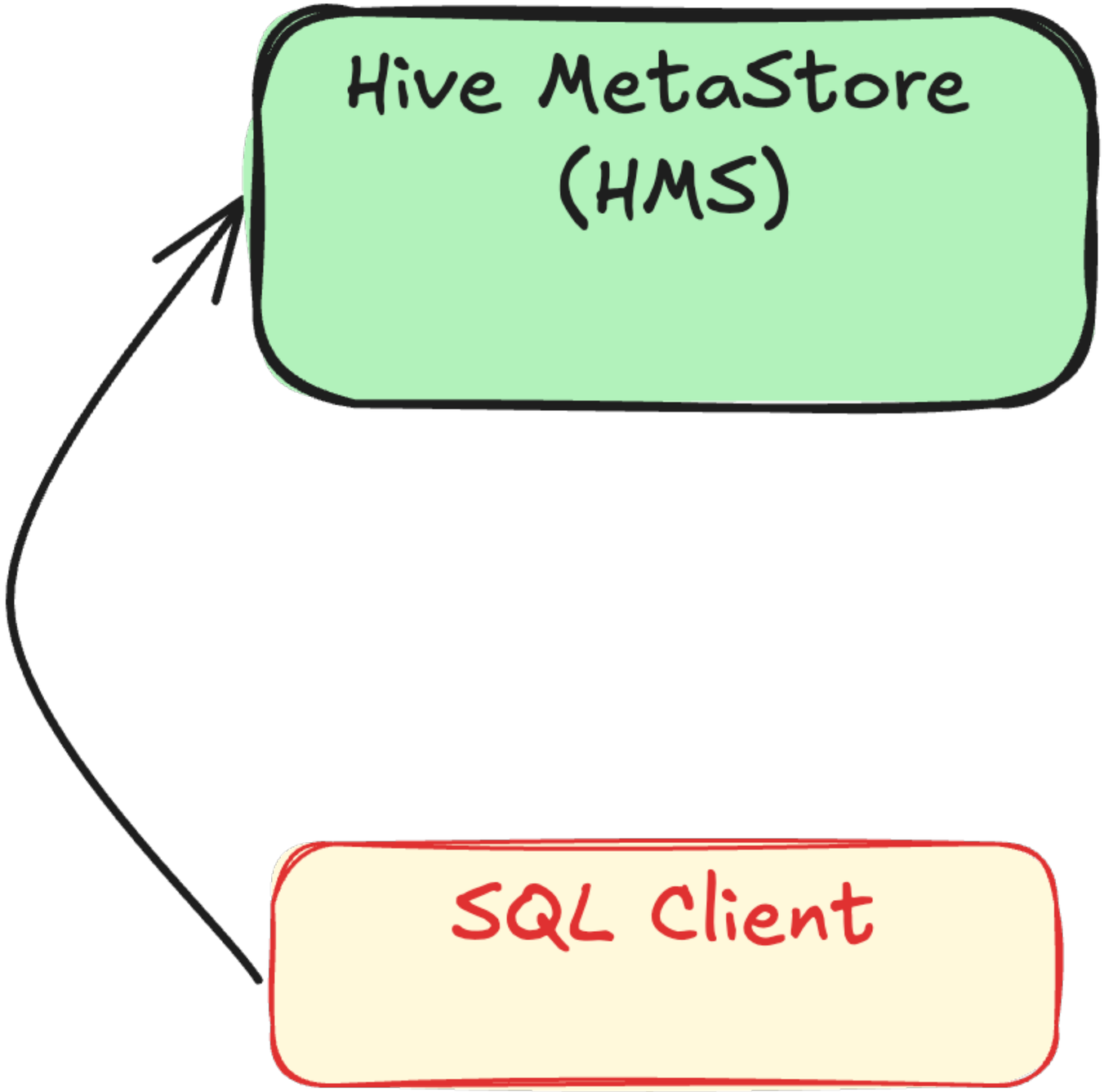
$ more log/flink--sql-client-asgard08.log

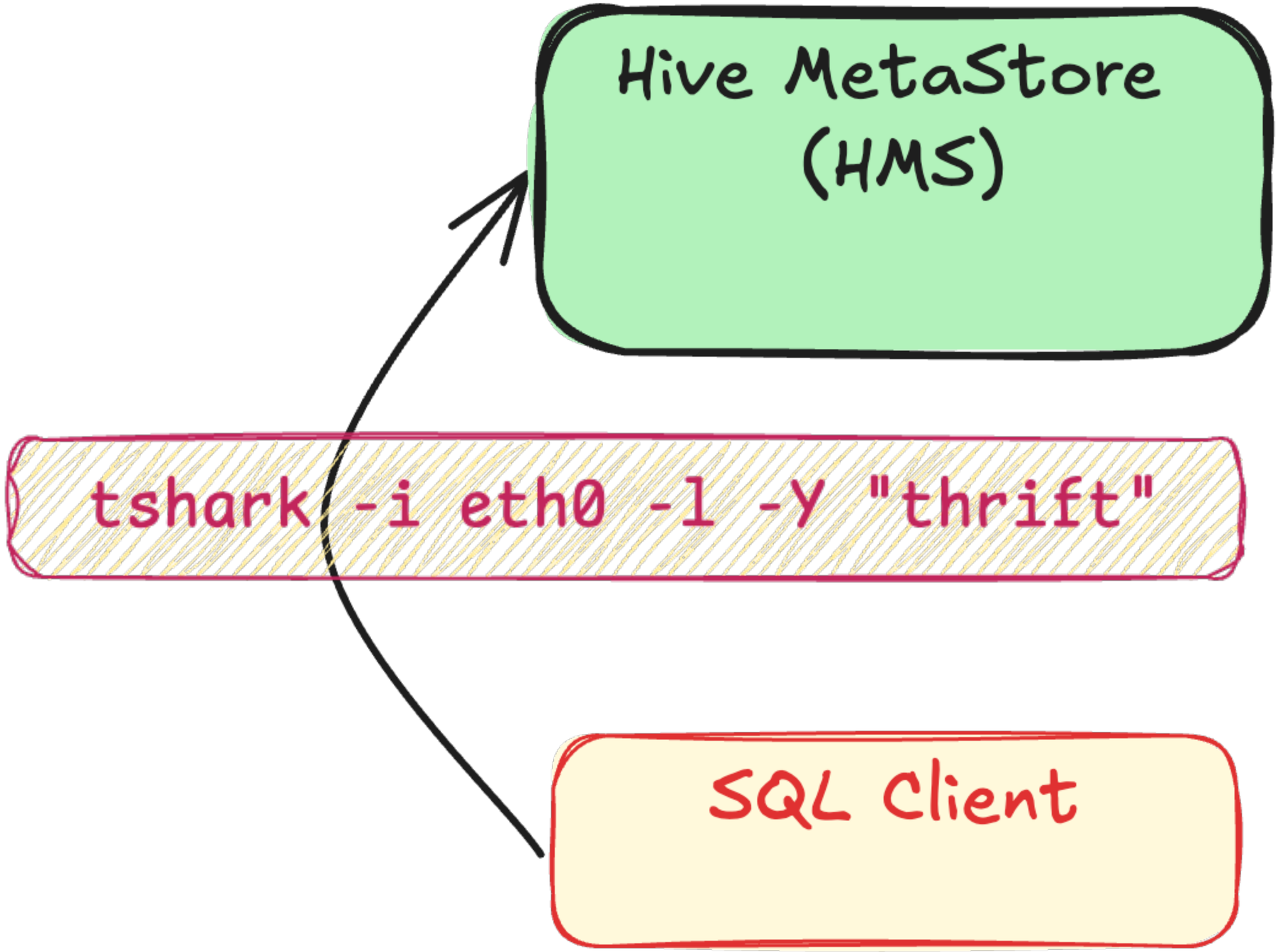
```
2024-09-12 11:20:06,494 ERROR org.apache.flink.table.gateway.service.operation.Operati
org.apache.flink.table.api.TableException: Could not execute CREATE DATABASE: (catalog
        at org.apache.flink.table.operations.ddl.CreateDatabaseOperation.execute(Creat
        at org.apache.flink.table.api.internal.TableEnvironmentImpl.executeInternal(Ta
[…]
Caused by: java.lang.RuntimeException: Failed to create namespace db01 in Hive Metasto
        at org.apache.iceberg.hive.HiveCatalog.createNamespace(HiveCatalog.java:299) ~
        at org.apache.iceberg.flink.FlinkCatalog.createDatabase(FlinkCatalog.java:223)
        at org.apache.iceberg.flink.FlinkCatalog.createDatabase(FlinkCatalog.java:214)
        at org.apache.flink.table.catalog.CatalogManager.createDatabase(CatalogManager
        at org.apache.flink.table.operations.ddl.CreateDatabaseOperation.execute(Creat
        ... 14 more
Caused by: org.apache.hadoop.hive.metastore.api.MetaException:
java.lang.RuntimeException: java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found
```

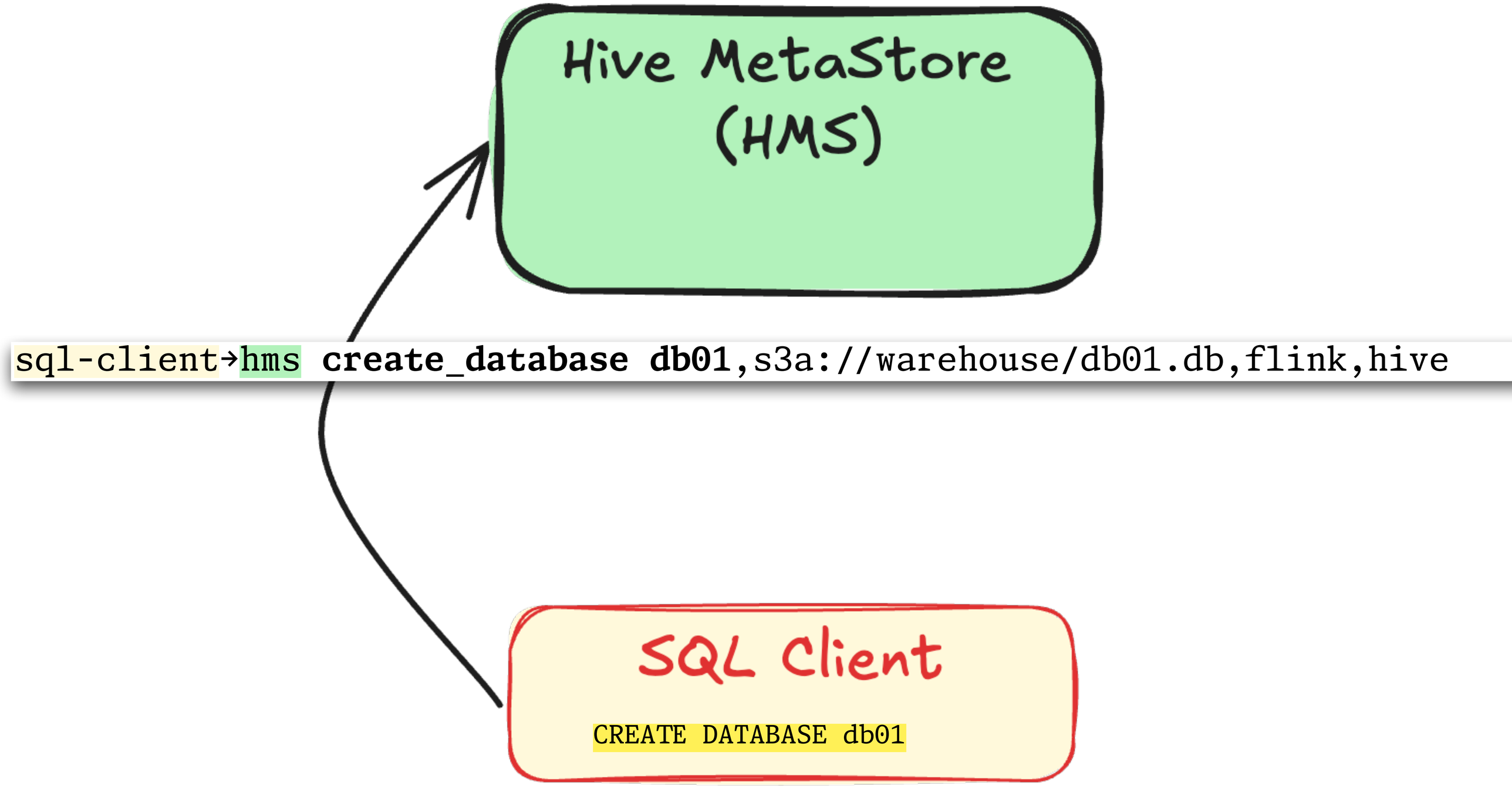
$ more log/flink--sql-client-asgard08.log

```
2024-09-12 11:20:06,494 ERROR org.apache.flink.table.gateway.service.operation.Operati
org.apache.flink.table.api.TableException: Could not execute CREATE DATABASE: (catalog
        at org.apache.flink.table.operations.ddl.CreateDatabaseOperation.execute(Creat
        at org.apache.flink.table.api.internal.TableEnvironmentImpl.executeInternal(Ta
[…]
Caused by: java.lang.RuntimeException: Failed to create database db01 in Hive Metasto
        at org.apache.iceberg.hive.HiveCatalog.createNamespace(HiveCatalog.java:209)
        at org.apache.iceberg.flink.FlinkCatalog.createDatabase(FlinkCatalog.java:23
        at org.apache.iceberg.flink.FlinkCatalog.createDatabase(FlinkCatalog.java:20
        at org.apache.flink.table.catalog.CatalogManager.createDatabase(CatalogManag
        at org.apache.flink.table.operations.ddl.CreateDatabaseOperation.execute(Operati
        ... 14 more
Caused by: org.apache.hadoop.hive.metastore.api.MetaException: java.lang.RuntimeExcep
java.lang.RuntimeException: java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found
```

```
$ more log/flink--sql-client-asgard08.log

2024-09-12 11:20:06,494 ERROR org.apache.flink.table.gateway.service.operation.Operati
org.apache.flink.table.api.TableException: Could not execute CREATE DATABASE: (catalog
        at org.apache.flink.table.operations.ddl.CreateDatabaseOperation.execute(Creat
        at org.apache.flink.table.api.internal.TableEnvironmentImpl.executeInternal(Ta
[…]
Caused by: java.lang.RuntimeException: Failed to create namespace db01 in Hive Metasto
        at org.apache.iceberg.hive.HiveCatalog.createNamespace(HiveCatalog.java:299) ~
        at org.apache.iceberg.flink.FlinkCatalog.createDatabase(FlinkCatalog.java:223)
        at org.apache.iceberg.flink.FlinkCatalog.createDatabase(FlinkCatalog.java:214)
        at org.apache.flink.table.catalog.CatalogManager.createDatabase(CatalogManager
        at org.apache.flink.table.operations.ddl.CreateDatabaseOperation.execute(Creat
        ... 14 more
Caused by: org.apache.hadoop.hive.metastore.api.MetaException:
java.lang.RuntimeException: java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found
        at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$create_database_re
        at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Client.recv_create
        at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Client.create_data
        at org.apache.thrift.TServiceClient.receiveBase(TServiceClient.java:86)
        at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Client.recv_create
        at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Client.create_data
        at org.apache.hadoop.hive.metastore.HiveMetaStoreClient.createDatabase(HiveMet
```


Hive MetaStore
(HMS)
SQL Client


Hive MetaStore
(HMS)
tshark -i eth0 -l -Y "thrift"
SQL Client

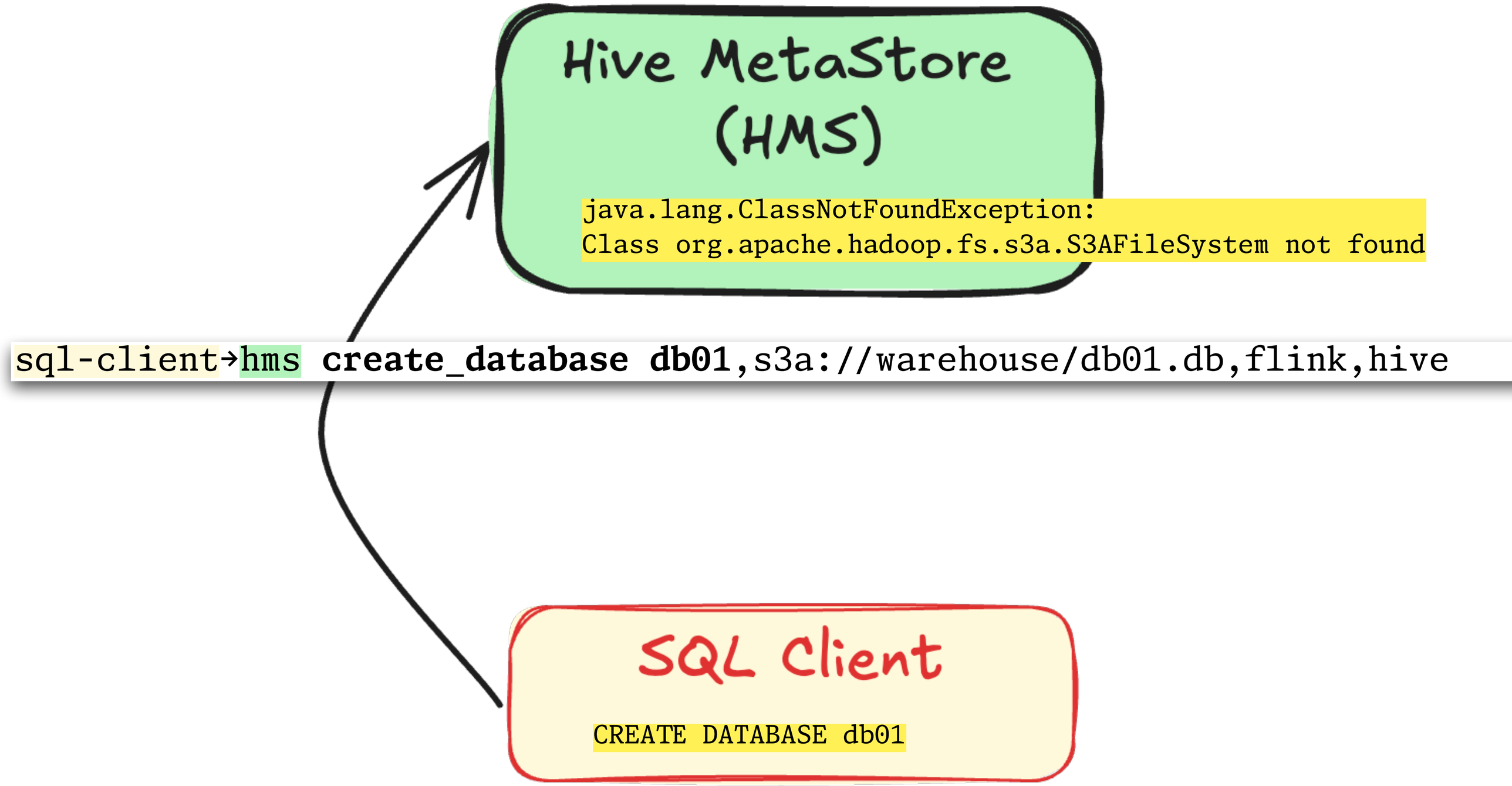

Hive MetaStore
(HMS)
sql-client→hms create_database db01,s3a://warehouse/db01.db,flink,hive
SQL Client
CREATE DATABASE db01

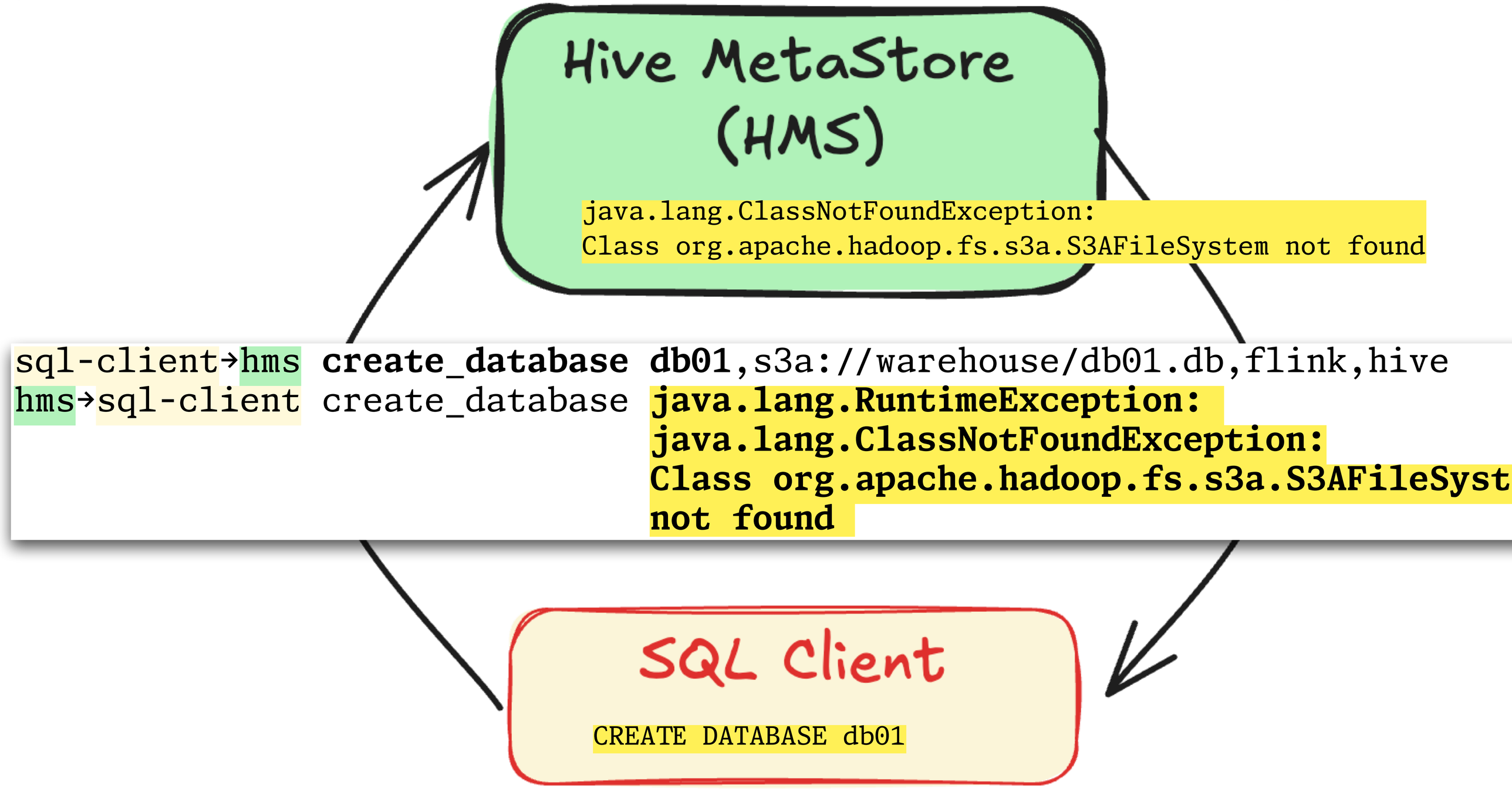
$ docker compose log hive-metastore

```
metastore.HiveMetaStore: 1: source:172.17.0.3 create_database: Database(name:db01, des
HiveMetaStore.audit: ugi=flink   ip=172.17.0.3   cmd=source:172.17.0.3 create_database
metastore.ObjectStore: Failed to get database hive.db01, returning NoSuchObjectExcepti
metastore.RetryingHMSHandler: MetaException(message:java.lang.RuntimeException:
java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found)
at org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.newMetaException(HiveMeta
at org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.create_database(HiveMetaS
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:4
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.hive.metastore.RetryingHMSHandler.invokeInternal(RetryingHMSHandl
at org.apache.hadoop.hive.metastore.RetryingHMSHandler.invoke(RetryingHMSHandler.java
at com.sun.proxy.$Proxy26.create_database(Unknown Source)
at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Processor$create_database
at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Processor$create_database
at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39)
at org.apache.hadoop.hive.metastore.TUGIBasedProcessor$1.run(TUGIBasedProcessor.java:1
at org.apache.hadoop.hive.metastore.TUGIBasedProcessor$1.run(TUGIBasedProcessor.java:1
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
```

$ docker compose log hive-metastore

```
metastore.HiveMetaStore: 1: source:172.17.0.3 create_database: Database(name:db01, des
HiveMetaStore.audit: ugi=flink   ip=172.17.0.3   cmd=source:172.17.0.3 create_database
metastore.ObjectStore: Failed to get database hive.db01, returning NoSuchObjectExcepti
metastore.RetryingHMSHandler: MetaException(message:java.lang.RuntimeException:
java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found)
at org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.                tion(HiveMeta
at org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.c                bas          taS
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAc
at sun.reflect.DelegatingMethodAccessorImpl.invoke(Deleg ing                              a:4
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.hive.metastore.RetryingHMSHandler.i                                  Handl
at org.apache.hadoop.hive.metastore.RetryingHMSHandler.inv                                java
at com.sun.proxy.$Proxy26.create_database(Unknown Source)
at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$P                             base
at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$P                             base
at org.apache.thrift.ProcessFunction.process(ProcessFunction.
at org.apache.hadoop.hive.metastore.TUGIBasedProcessor$1.run(                    r.java:1
at org.apache.hadoop.hive.metastore.TUGIBasedProcessor$1.run(T                   sor.java:1
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
```


Hive MetaStore
(HMS)
java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found
sql-client→hms create_database db01,s3a://warehouse/db01.db,flink,hive
SQL Client
CREATE DATABASE db01


Hive MetaStore (HMS)
java.lang.ClassNotFoundException: Class org.apache.hadoop.fs.s3a.S3AFileSystem not found
sql-client→hms create_database db01,s3a://warehouse/db01.db,flink,hive
hms→sql-client create_database java.lang.RuntimeException: java.lang.ClassNotFoundException: Class org.apache.hadoop.fs.s3a.S3AFileSyst not found
SQL Client
CREATE DATABASE db01

Hive MetaStore (HMS)

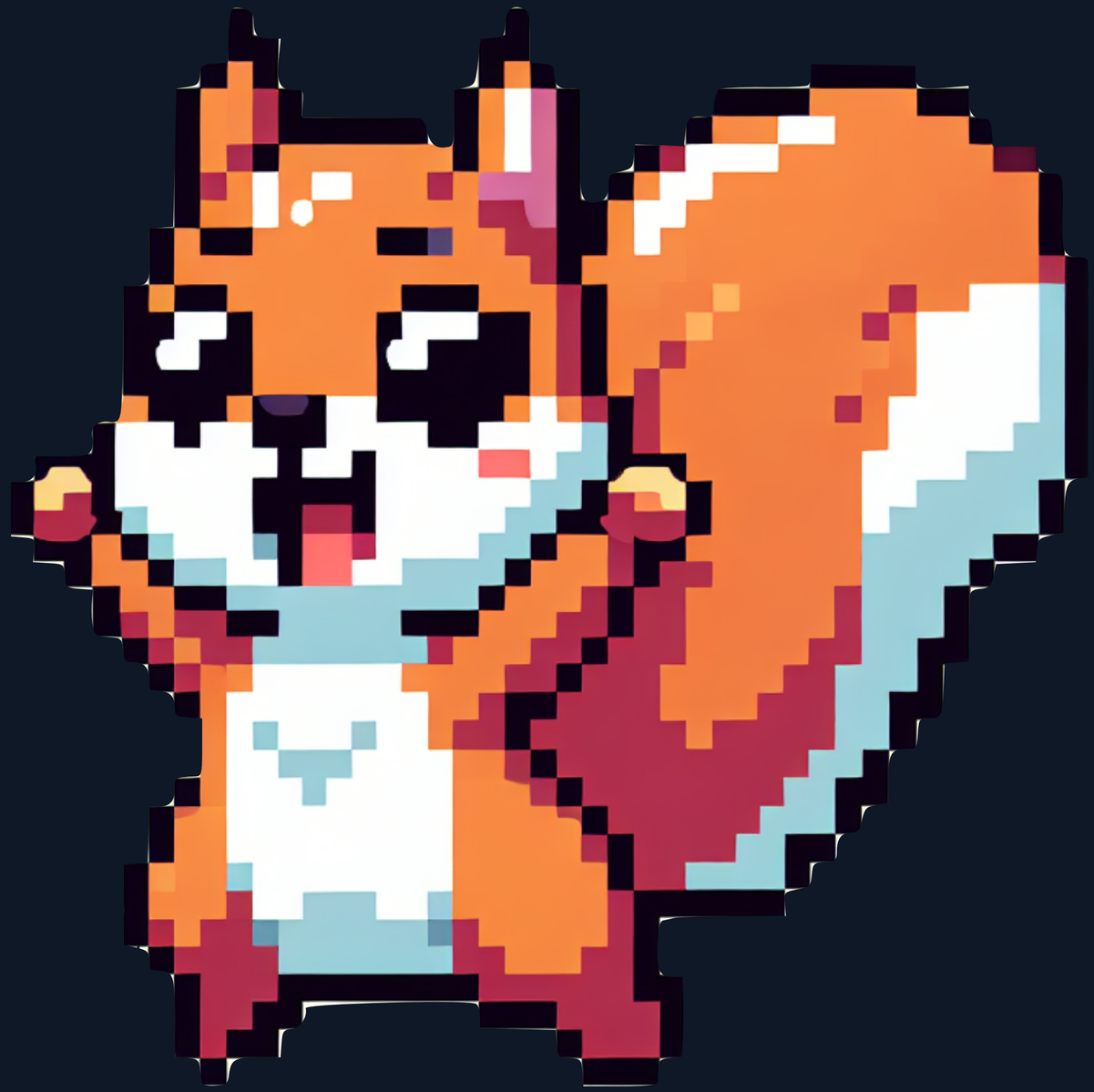
```
java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found
```

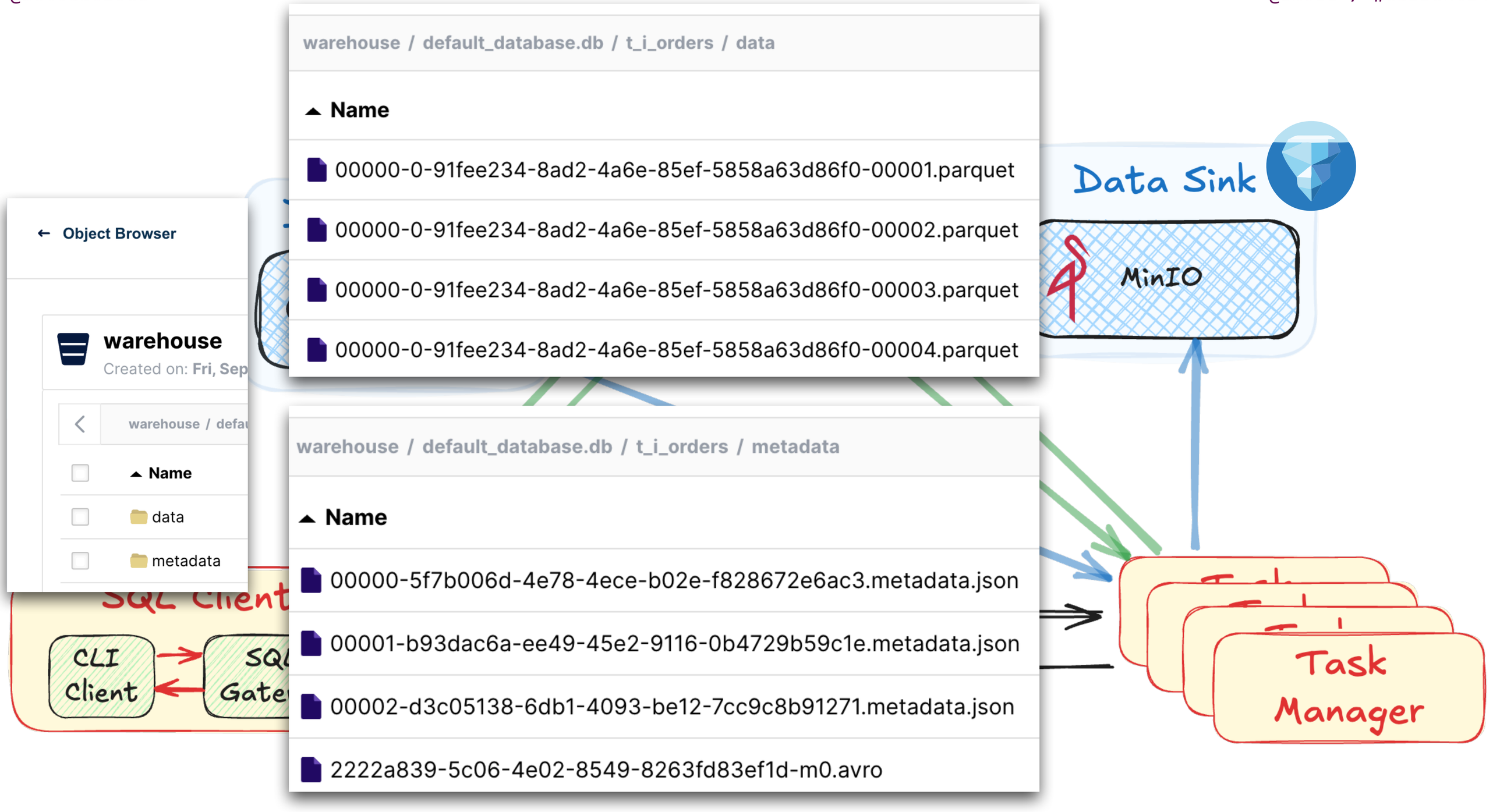
```
sql-client→hms create_database db01,s3a://warehouse/db01.db,flink,hive
hms→sql-client create_database java.lang.RuntimeException:
                                java.lang.ClassNotFoundException:
                                Class org.apache.hadoop.fs.s3a.S3AFileSyst
                                not found
```

SQL Client

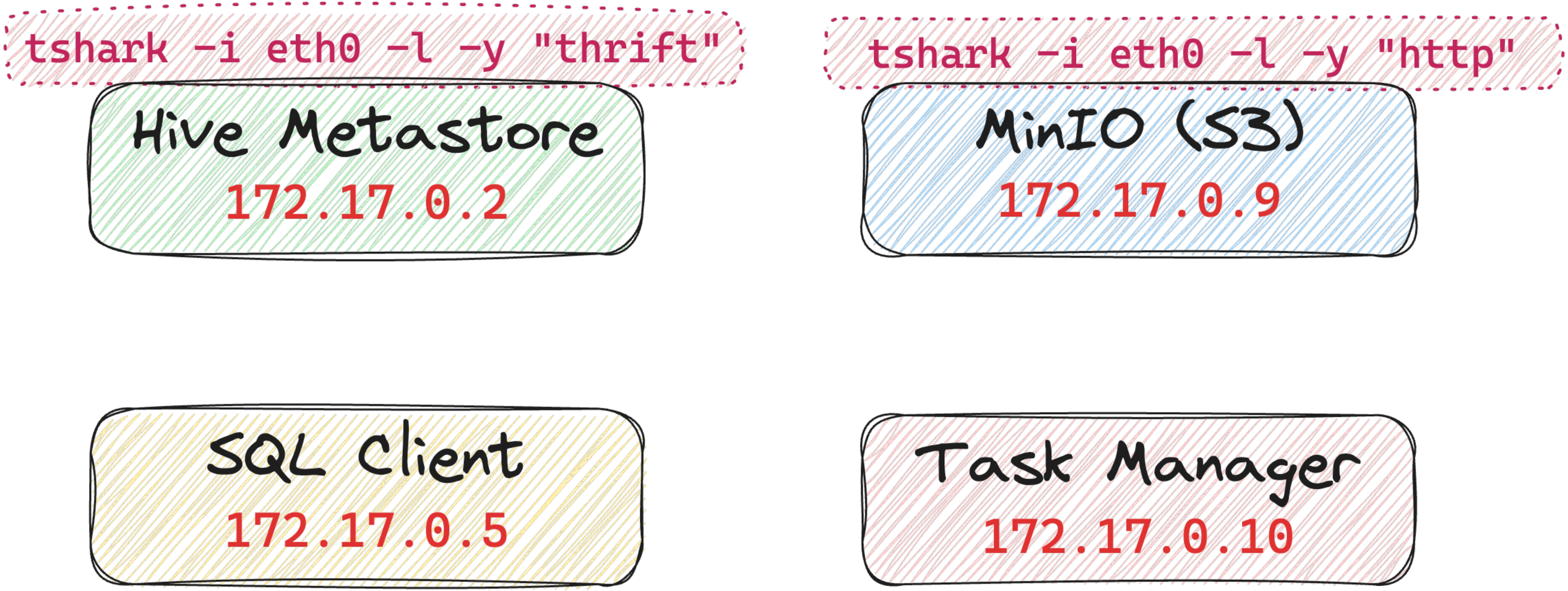
```
java.lang.ClassNotFoundException:
Class org.apache.hadoop.fs.s3a.S3AFileSystem not found
```

```
Flink SQL> CREATE DATABASE db01;
[INFO] Execute statement succeed.
```



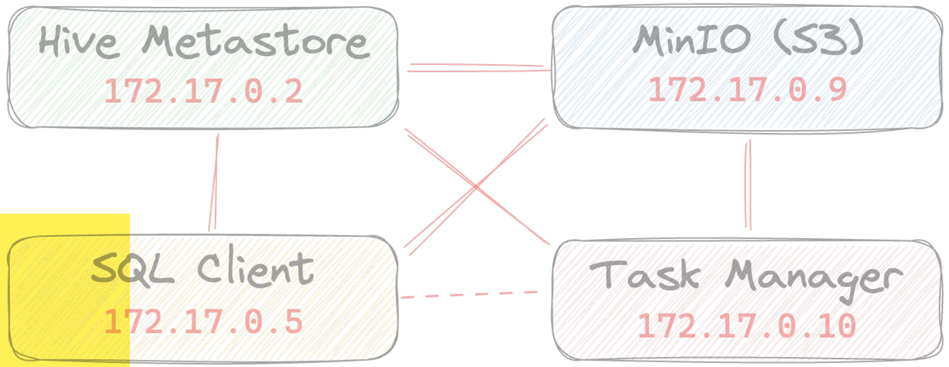

warehouse / default_database.db / t_i_orders / data
Name
00000-0-91fee234-8ad2-4a6e-85ef-5858a63d86f0-00001.parquet
00000-0-91fee234-8ad2-4a6e-85ef-5858a63d86f0-00002.parquet
00000-0-91fee234-8ad2-4a6e-85ef-5858a63d86f0-00003.parquet
00000-0-91fee234-8ad2-4a6e-85ef-5858a63d86f0-00004.parquet
Data Sink
MinIO
Object Browser
warehouse
Created on: Fri, Sep
warehouse / defau
Name
data
metadata
warehouse / default_database.db / t_i_orders / metadata
Name
00000-5f7b006d-4e78-4ece-b02e-f828672e6ac3.metadata.json
00001-b93dac6a-ee49-45e2-9116-0b4729b59c1e.metadata.json
00002-d3c05138-6db1-4093-be12-7cc9c8b91271.metadata.json
2222a839-5c06-4e02-8549-8263fd83ef1d-m0.avro
SQL Client
CLI Client
Task Manager

TCP packet capture



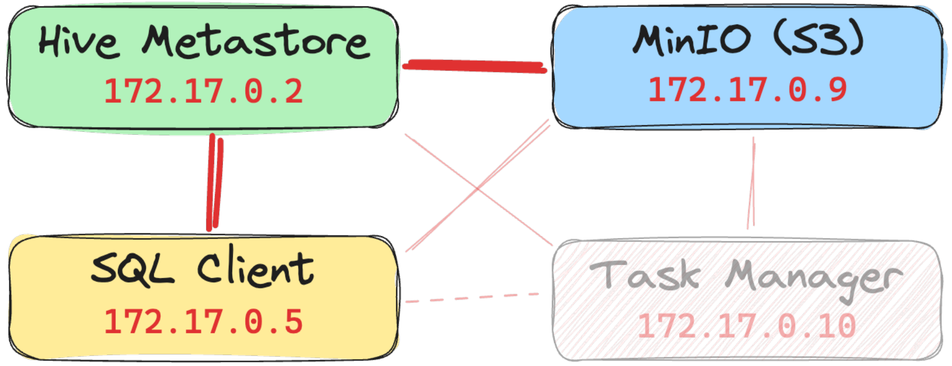
CREATE CATALOG

```
CREATE CATALOG c_iceberg WITH (
        'type' = 'iceberg',
        'catalog-type'='hive',
        'warehouse' = 's3a://warehouse',
        'hive-conf-dir' = './conf');
```



CREATE DATABASE

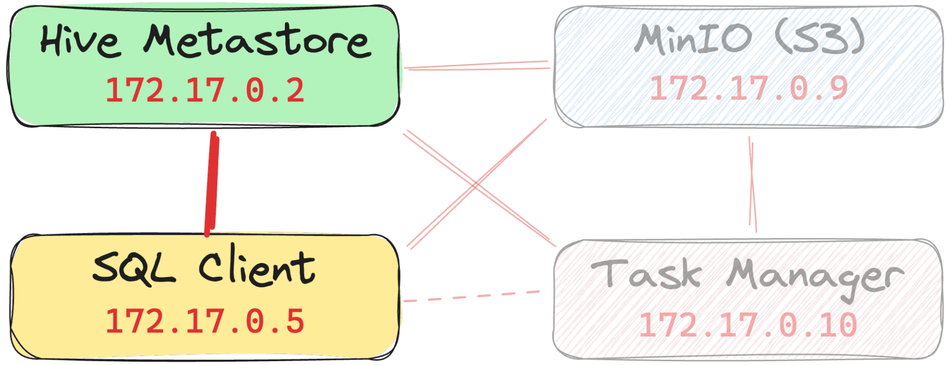
```
CREATE DATABASE c_iceberg.rmoff;
```




```
172.17.0.5→172.17.0.2 create_database rmoff,s3a://warehouse/rmoff.db,fli

172.17.0.2→172.17.0.9 HTTP 152 PUT /warehouse/rmoff.db/ HTTP/1.1  (appli
```

USE DATABASE

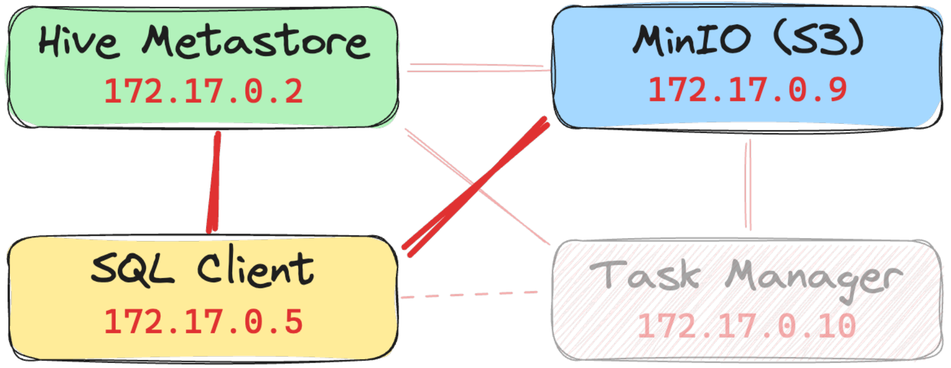
```
USE c_iceberg.rmoff;
```




```
172.17.0.5→172.17.0.2 get_database     @hive#rmoff
172.17.0.2→172.17.0.5 get_database     rmoff,s3a://warehouse/rmoff.db,fli
```

CREATE TABLE

```
CREATE TABLE foo
(c1 INT, c2 STRING);
```

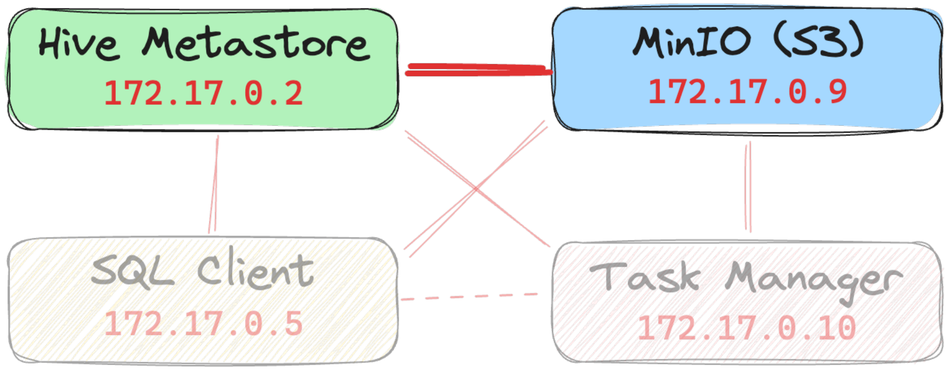



```
172.17.0.5 172.17.0.2 create_table_with_environment_context   foo,rmoff,
                      c1,int,c2,string,s3a://warehouse/rmoff.db/foo,
                      metadata_location,s3a://warehouse/rmoff.db/foo/met
                      current-schema,{"type":"struct","schema-id":0,"fie
```

```
172.17.0.5 172.17.0.9 HTTP 1229 PUT /warehouse/rmoff.db/foo/metadata/000
```

CREATE TABLE

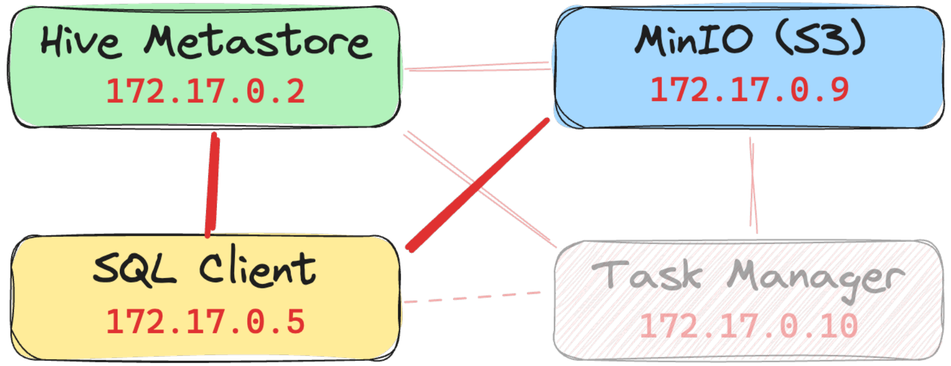
```
CREATE TABLE foo
(c1 INT, c2 STRING);
```




```
172.17.0.2 172.17.0.9 HTTP 1273 HEAD /warehouse/rmoff.db/foo HTTP/1.1

172.17.0.2 172.17.0.9 HTTP 1380 GET /warehouse/?list-type=2&delimiter=%2
```

INSERT INTO

```
INSERT INTO foo VALUES
(42, 'never gonna give you up');
```

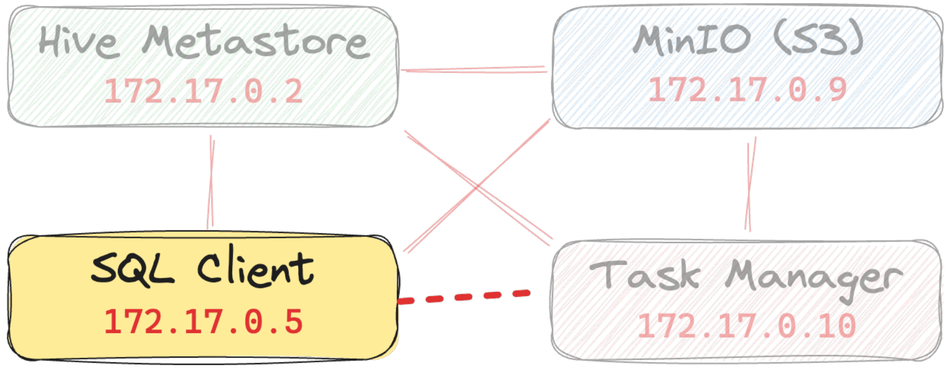



```
172.17.0.5 172.17.0.2 set_ugi flink,flink
172.17.0.2 172.17.0.5 set_ugi flink

172.17.0.5 172.17.0.2 get_table_req rmoff,foo,hive
172.17.0.2 172.17.0.5 get_table_req foo,rmoff,flink,c1,int,c2,string,s3a
                      previous_metadata_location,s3a://warehouse/rmoff.d
                      current-schema,{"type":"struct","schema-id":0,"fie
```

```
172.17.0.5 172.17.0.9 HTTP 1001 HEAD /warehouse/rmoff.db/foo/metadata/00

172.17.0.5 172.17.0.9 HTTP 1078 GET /warehouse/rmoff.db/foo/metadata/000
```

INSERT INTO

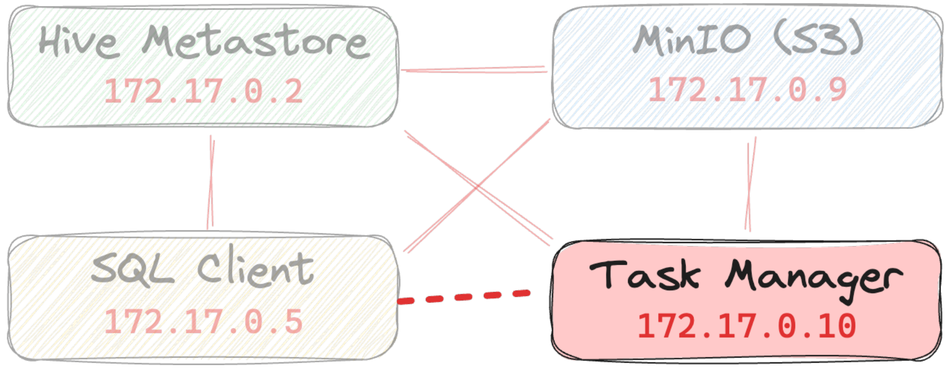
```
INSERT INTO foo VALUES
(42, 'never gonna give you up');
```




```
[INFO] Submitting SQL update statement to the cluster...
[INFO] SQL update statement has been successfully submitted to the clust
Job ID: cc43d32a6bb0e2faab5270e542c70499
```

INSERT INTO

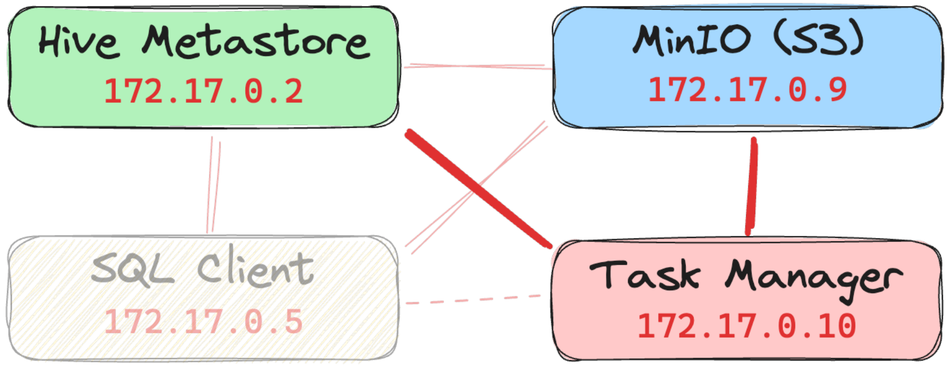
```
INSERT INTO foo VALUES
(42, 'never gonna give you up');
```




```
Establish JobManager connection for job cc43d32a6bb0e2faab5270e542c70499
Offer reserved slots to the leader of job cc43d32a6bb0e2faab5270e542c704

Received task Source: Values[3] -> IcebergStreamWriter (1/1)#0 (0b643b78

Downloading cc43d32a6bb0e2faab5270e542c70499/p-c817b883da6db5a2bfd0e5b35
from jobmanager/172.17.0.3:6124

Received task IcebergFilesCommitter -> Sink: IcebergSink c_iceberg.rmoff
```

INSERT INTO

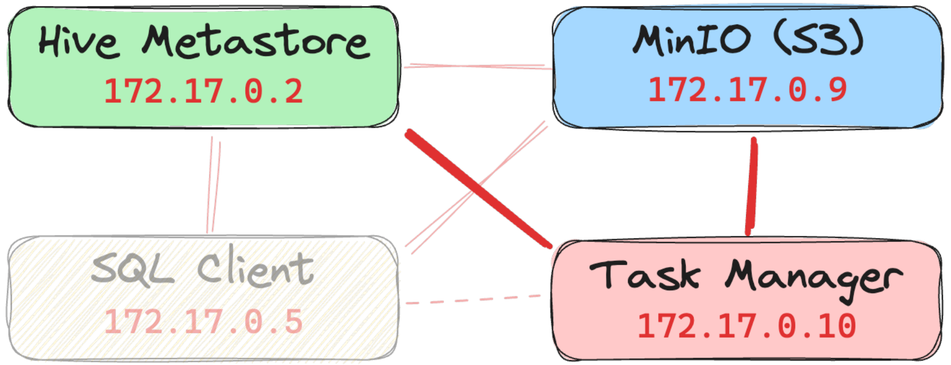
```
INSERT INTO foo VALUES
(42, 'never gonna give you up');
```




```
172.17.0.10 172.17.0.2 get_table_req rmoff,foo,hive
172.17.0.2 172.17.0.10 get_table_req foo,rmoff,flink,c1,int,c2,string,s3
                       previous_metadata_location,s3a://warehouse/rmoff.
                       current-schema,{"type":"struct","schema-id":0,"fi

172.17.0.10 172.17.0.2 alter_table_with_environment_context @hive#rmoff,
                       current-snapshot-summary,{"flink.operator-id":"90
                       "flink.max-committed-checkpoint-id":"922337203685
                       "added-files-size":"764","changed-partition-count
                       "total-records":"3","total-files-size":"2292","to
```

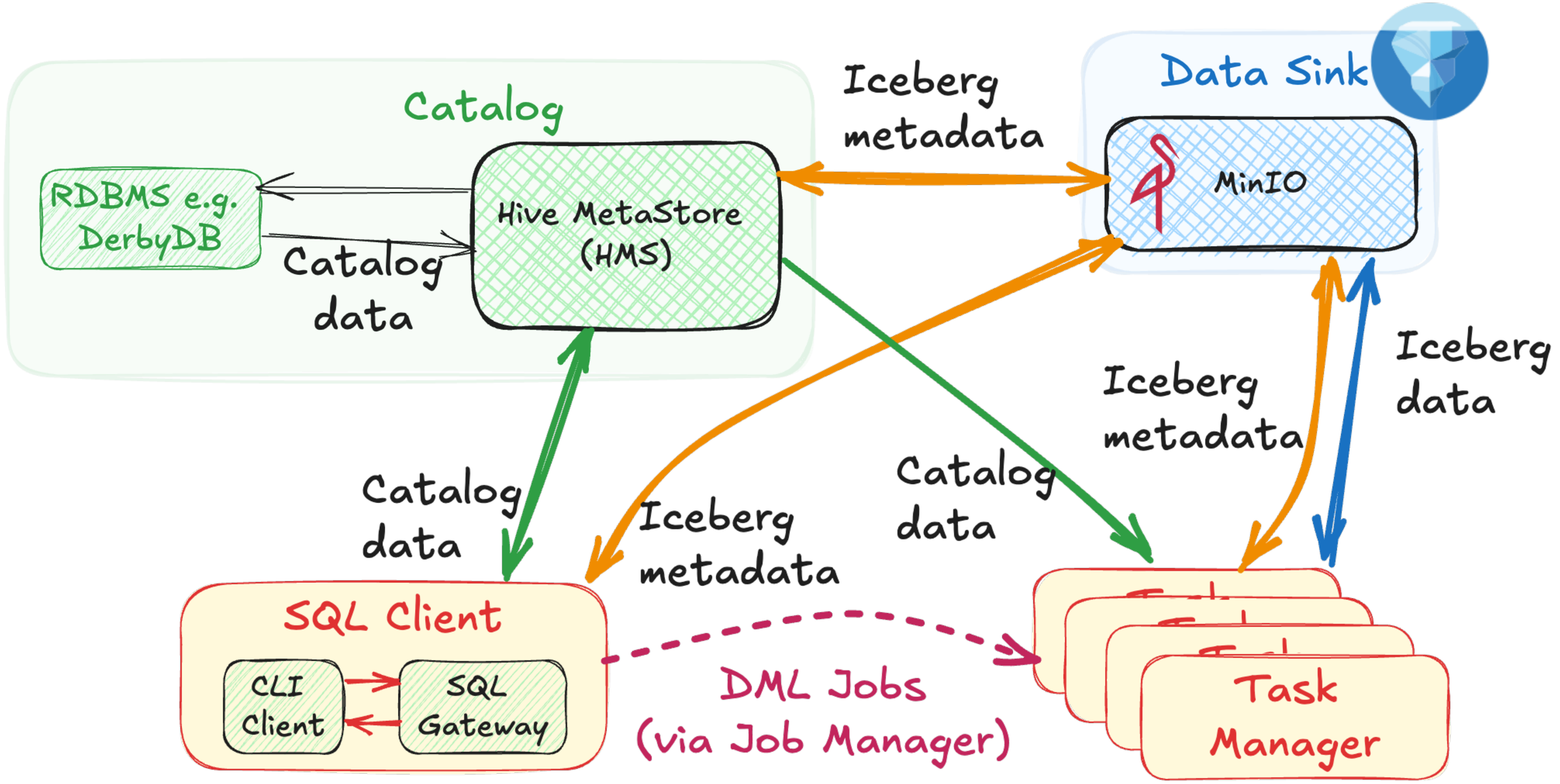
INSERT INTO

```
INSERT INTO foo VALUES
(42, 'never gonna give you up');
```




```
172.17.0.10 172.17.0.9 HTTP 1089 GET /warehouse/?list-type=2&delimiter=%
172.17.0.10 172.17.0.9 HTTP 1004 PUT /warehouse/rmoff.db/foo/data/00000-
172.17.0.10 172.17.0.9 HTTP 6902 PUT /warehouse/rmoff.db/foo/metadata/
172.17.0.10 172.17.0.9 HTTP 6914 PUT /warehouse/rmoff.db/foo/metadata/b2
```


BUT SRSLY...
Y THO?
imgflip.com


Catalog
RDBMS e.g. DerbyDB
Hive MetaStore (HMS)
Catalog data
Iceberg metadata
Data Sink
MinIO
Iceberg metadata
Iceberg data
Catalog data
Iceberg metadata
Catalog data
SQL Client
CLI Client
SQL Gateway
DML Jobs (via Job Manager)
Task Manager

What SQL runs where?

DDL

CREATE TABLE AS SELECT
CREATE
DROP
ALTER

SQL Client

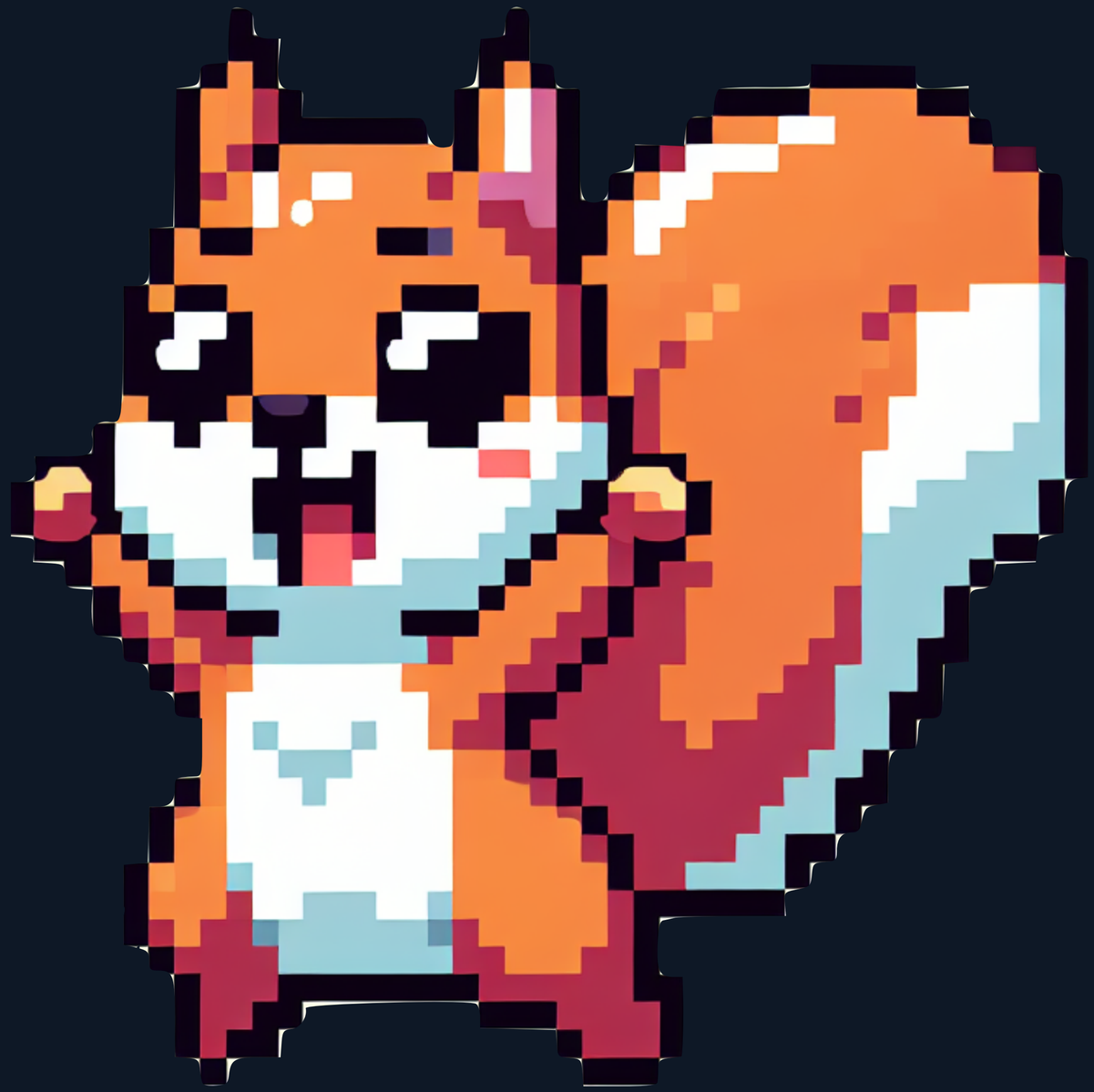
DML

SELECT
CREATE TABLE AS SELECT
INSERT INTO
UPDATE
DELETE

Task Manager

```
Flink SQL> CREATE DATABASE `c_iceberg_hive`.`db01`;
[ERROR] Could not execute SQL statement. Reason:
MetaException(message:java.lang.RuntimeException: java.lang.ClassNotFoundException
 Class org.apache.hadoop.fs.s3a.S3AFileSystem not found)
```

```
Flink SQL> CREATE DATABASE db01;
[INFO] Execute statement succeed.
```



In conclusion…

Troubleshoot methodically

Understand the architecture

Get your JARs right

Get your toolbox ready :)

decodable.co/blog

February 19, 2024 10 min read

Catalogs in Flink SQL—Hands On

A hands-on guide to using catalogs with Flink SQL, including Apache Hive, JDBC, and Apache Iceberg with different metastores. Covers installation, setup, and usage.

February 27, 2024 4 min read

Flink SQL and the Joy of JARs

A handy guide to some of the most common issues with Flink SQL and JAR dependencies, how to troubleshoot them—and how to fix them! 🎉

March 12, 2024 5 min read

Exploring the Flink SQL Gateway REST API

Learn how to use the Flink SQL gateway's REST API to send SQL statements to your Flink cluster programmatically.

February 16, 2024 8 min read

Catalogs in Flink SQL—A Primer

Explore the essentials of catalogs in Flink SQL. Catalogs store object definitions like tables and views for the Flink query engine. This primer covers the role of catalogs in managing metadata in Flink, the different catalogs available in Flink, and how to use the CatalogStore.

@rmoff / 18 Sep 2024 / #current24