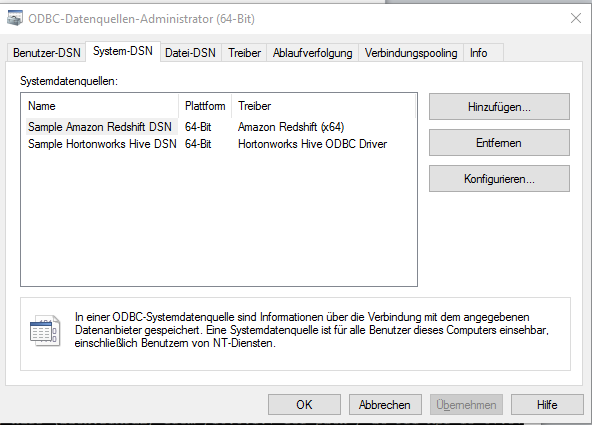
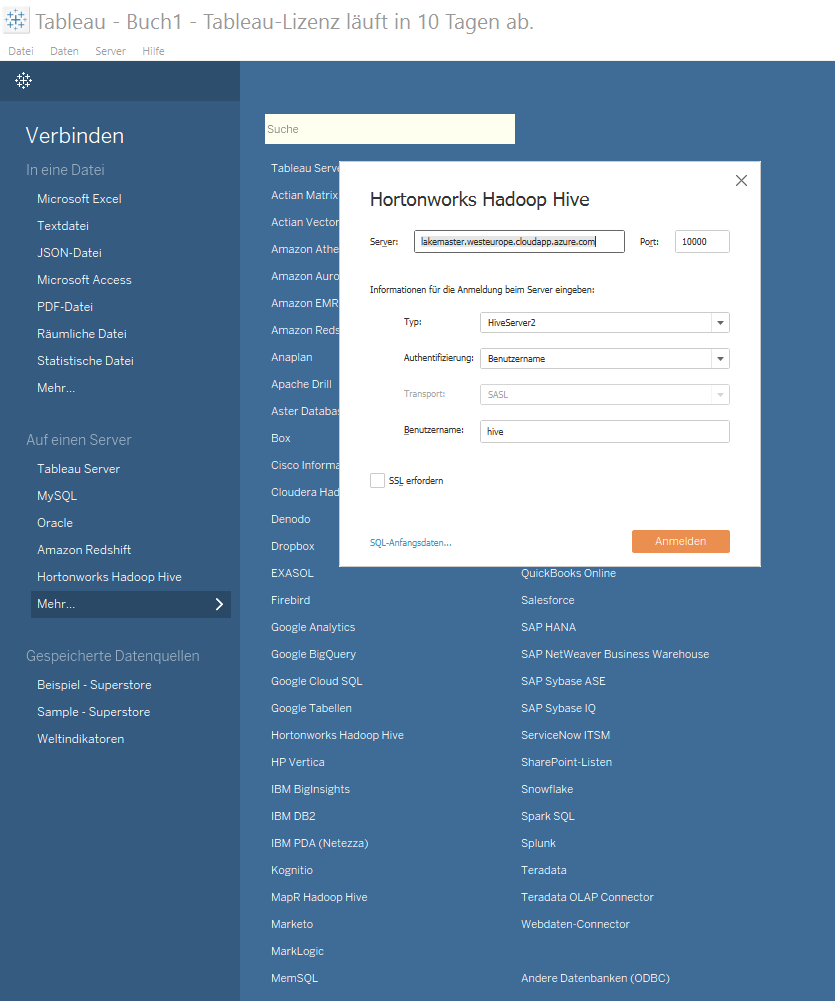
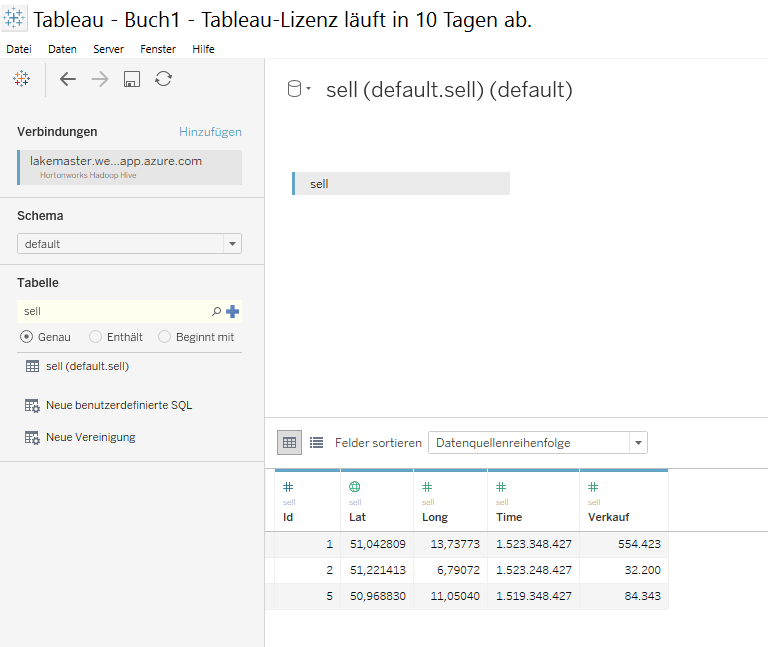
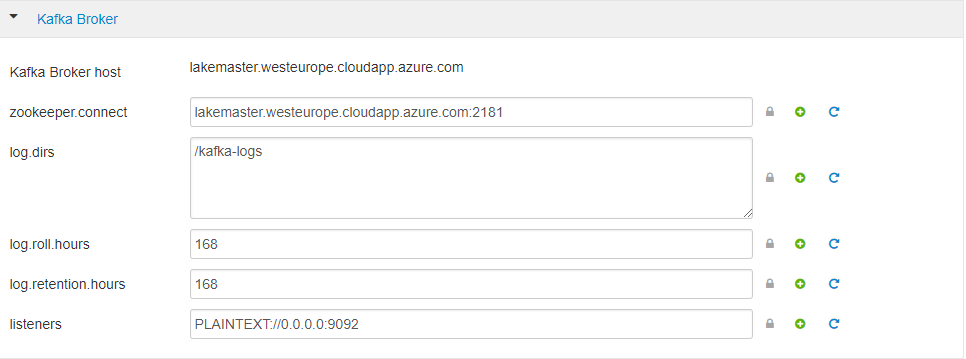
**Dokumentation:**

**Hadoop(Hive) – Tableau (am Beispiel von Tableau Desktop unter Windows):**   
(Anleitung: <https://onlinehelp.tableau.com/current/pro/desktop/en-us/examples_hortonworkshadoop.html>)

* Port von Hive (Standard-Port: 10000) freigeben
* Tableau Desktop installieren
* OBDC-Treiber installieren
  + Download: <https://de.hortonworks.com/downloads/>   
    Unter: Add-ons für Hortonworks Data Platform - Hortonworks ODBC Driver for Apache Hive (v2.1.12)
  + Treiber installieren
  + Treiber Konfigurieren - unter ODC-Datenquellen-Administrator(64-Bit) – Unter System-DSN  
    
* Tableau mit Hive verbinden
* 
* Schema auswählen
* Tabelle auswählen  
  

**Kafka – Flume – Hive:**

* Kafka Konfigurieren
  + Port: 9092
  + URL: 0.0.0.0 (damit von localhost und von außerhalb Kafka angesprochen werden kann
  + 
* Kafka Topic erstellen
  + bin/kafka-topics.sh --create --zookeeper lakemaster.westeurope.cloudapp.azure.com:2181 --replication-factor 1 --partitions 1 --topic data
* Hive Table erstellen
  + [evtl. müssen noch folgende Attribute gesetzt werden]
    - SET hive.support.concurrency=true;
    - SET hive.enforce.bucketing=true;
    - SET hive.exec.dynamic.partition.mode=nonstrict;
    - SET hive.txn.manager=org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;
  + create table flume\_test(id string, message string) clustered by (message) into 1 buckets STORED AS ORC tblproperties ("orc.compress"="NONE", "transactional"="true");
* Flume Agent erstellen
  + Variablen vorher setzen und Java 1.8 muss installiert sein
    - export HIVE\_HOME=/usr/hdp/current/hive-server2
    - export HCAT\_HOME=/usr/hdp/current/hive-webhcat
    - JAVA\_HOME=/usr/jdk64/jdk1.8.0\_112
    - export JAVA\_HOME
  + Flume konfigurieren ( z.b. unter /home/data/flume/test.conf)
    - flumeagent1.sources = source\_from\_kafka
    - flumeagent1.channels = mem\_channel
    - flumeagent1.sinks = hive\_sink
    - flumeagent1.sources.source\_from\_kafka.type = org.apache.flume.source.kafka.Kafka Source
    - flumeagent1.sources.source\_from\_kafka.zookeeperConnect = lakemaster.westeurope.c loudapp.azure.com:2181
    - flumeagent1.sources.source\_from\_kafka.topic = flume\_test
    - flumeagent1.sources.source\_from\_kafka.groupID = flume
    - flumeagent1.sources.source\_from\_kafka.channels = mem\_channel
    - flumeagent1.sources.source\_from\_kafka.interceptors = i1
    - flumeagent1.sources.source\_from\_kafka.interceptors.i1.type = timestamp
    - flumeagent1.sources.source\_from\_kafka.consumer.timeout.ms = 1000
    - flumeagent1.sinks.hive\_sink.type = hive
    - flumeagent1.sinks.hive\_sink.channel = mem\_channel
    - flumeagent1.sinks.hive\_sink.hive.metastore = thrift://lakemaster.westeurope.clou dapp.azure.com:9083
    - flumeagent1.sinks.hive\_sink.hive.database = default
    - flumeagent1.sinks.hive\_sink.hive.table = flume\_test
    - flumeagent1.sinks.hive\_sink.useLocalTimeStamp = false
    - flumeagent1.sinks.hive\_sink.round = true
    - flumeagent1.sinks.hive\_sink.roundValue = 10
    - flumeagent1.sinks.hive\_sink.roundUnit = minute
    - flumeagent1.sinks.hive\_sink.serializer = DELIMITED
    - flumeagent1.sinks.hive\_sink.serializer.delimiter = ","
    - flumeagent1.sinks.hive\_sink.serializer.serdeSeparator = ','
    - flumeagent1.sinks.hive\_sink.serializer.fieldnames = id,message
    - flumeagent1.sinks.hive\_sink.hive.txnsPerBatchAsk = 2
    - flumeagent1.sinks.hive\_sink.batchSize = 10
    - flumeagent1.channels.mem\_channel.type = FILE
    - flumeagent1.channels.mem\_channel.transactionCapacity = 1000000
    - flumeagent1.channels.mem\_channel.checkpointInterval 30000
    - flumeagent1.channels.mem\_channel.maxFileSize = 2146435071
    - flumeagent1.channels.mem\_channel.capacity 10000000
    - flumeagent1.sources.source\_from\_kafka.channels = mem\_channel
  + Flume starten:
    - bin/flume-ng agent -c conf -f /home/data/flume/test.conf --name flumeagent1 -Dflume.root.logger=INFO,console

Troubleshooting:

Falls Fehlermeldungen kommen, dass JAR-Dateien fehlen, diese in Ambari bei Hive mit einfügen:

