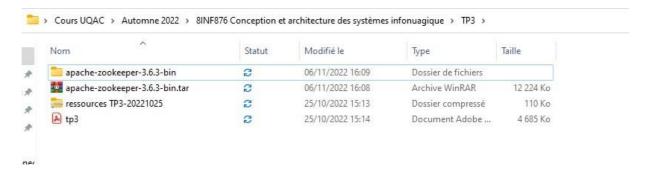
Zookeeper, distributed systems & cloud

I) Introduction

ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services. All these kinds of services are used in some form or another by distributed applications. Each time they are implemented there is a lot of work that goes into fixing the bugs and race conditions that are inevitable. Because of the difficulty of implementing these kinds of services, applications initially usually skimp on them, which make them brittle in the presence of change and difficult to manage.

ZooKeeper architecture supports high availability through redundant services. The clients can thus ask another ZooKeeper leader if the first fails to answer. ZooKeeper nodes store their data in a hierarchical name space, much like a file system or a tree data structure. Clients can read from and write to the nodes and in this way have a shared configuration service. ZooKeeper can be viewed as an atomic broadcast system, through which updates are totally ordered. The **ZooKeeper Atomic Broadcast** ¹(ZAB) protocol is the core of the system.

Zookeeper installation and configuration



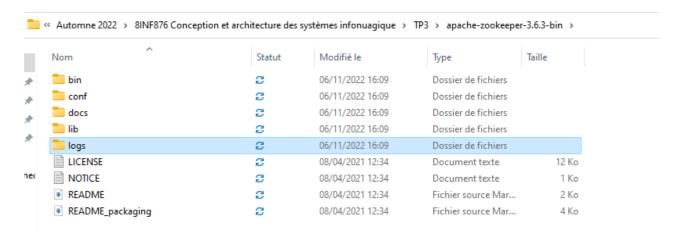
We first installed **Apache Zookeeper** bin.

¹ ZooKeeper Atomic Broadcast: messaging system that keeps all the servers in sync

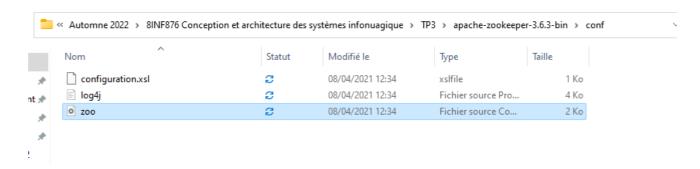
C:\Users\moadb>Javac -version javac 17.0.1

C:\Users\moadb>echo %JAVA_HOME%. C:\Program Files\Java\jdk-17.0.1.

There are Java Development Kit ² and Java Home³.



Then, we create logs folder.



We rename the file zoo_sample.cfg by zoo.cfg.

²Java Development Kit: is a distribution of *Java* Technology by Oracle Corporation. It implements the *Java* Language Specification

³ **Java Home**: is an operating system (OS) environment variable which can optionally be set after either the Java Development Kit (JDK)

```
C: > Users > moadb > OneDrive > Bureau > Cours UQAC > Automne 2022 > 8INF876 Conception et architecture des systèmes infonuagique > TP3 > apache-zookeeper-3.6.3-bin > conf > © zoo.cfg

1  # The number of milliseconds of each tick
2  tickTime=2000
3  # The number of ticks that the initial
4  # synchronization phase can take
5  initimit=10
6  # The number of ticks that can pass between
7  # sending a request and getting an acknowledgement
8  synclimit=5
9  # the directory where the snapshot is stored.
10  # do not use /tmp for storage, /tmp here is just
1  # example sakes.
12  dataOir=/Users/moadb/OneOrive/Bureau/Cours UQAC/Automne 2022/8INF876 Conception et architecture des systèmes infonuagique/TP3/apache-zookeeper-3.6.3-bin/logs
13  # the port at which the clients will connect
14  clientPort=2181
15  # the maximum number of client connections.
16  # increase this if you need to handle more clients
17  #maxClientCnors=60
```

We modify of the **zoo.cfg** file: (modification of the logs directory).

```
C:\Users\moadb\OneDrive\Bureau\Cours UQAC\Automne 2022\BINF876 Conception et architecture des systèmes infonuagique\TP3\apache-zookeeper-3.6.3-bin\bin>zkServer.cmd

C:\Users\moadb\OneDrive\Bureau\Cours UQAC\Automne 2022\BINF876 Conception et architecture des systèmes infonuagique\TP3\apache-zookeeper-3.6.3-bin\bin>zkServer.cmd

C:\Users\moadb\OneDrive\Bureau\Cours UQAC\Automne 2022\BINF876 Conception et architecture des systèmes infonuagique\TP3\apache-zookeeper-3.6.3-bin\bin\bin>zkServer.cmd

C:\Users\moadb\OneDrive\Bureau\Cours UQAC\Automne 2022\BINF876 Conception et architecture des systèmes infonuagique\TP3\apache-zookeeper-1.6.3-bin\bin\bin\...\logs\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\models\mode
```



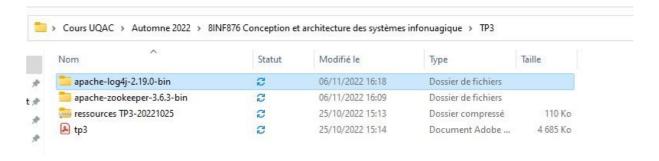
We launch zkServer.cmd.

Then, we **launchzkCli.cmd** in another terminal.

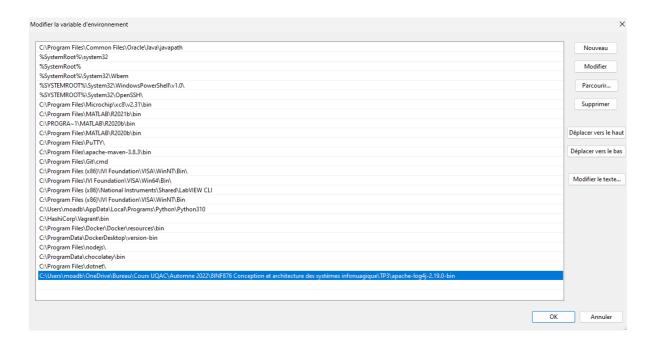
II) Part 1

Section: Zookeeper Client Threading Model

Step 1: Installation log4i

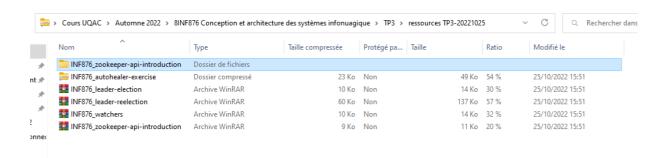


We downloaded Apache-log4j-2.19.0.

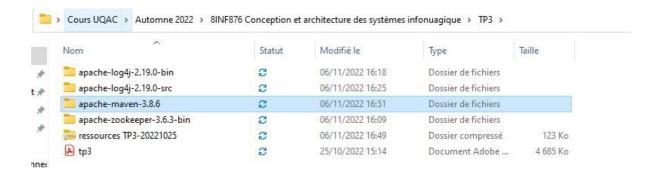


There is the environment variable.

Step 2: Installation Zookeeper API and Unzip



Step 3: Installation Maven



There is the project on Intellij 4: ZooKeeper Java API

⁴Intellij: IDE for Java

```
| CleaderElection.java × | M pom.xml (leader.election) × | CartifactId>maven-compiler-plugin</artifactId> | CartifactId>maven-compiler-plugin</artifactId> | CartifactId>maven-compiler-plugin</artifactId> | CartifactId>maven-compiler-plugin</artifactId> | CartifactId> | Cartif
```

There is the **pom.xml**.

There are the log4j.properties folder: PAGE 17

III) PART 2

<u>Section: Leader election implementation</u>

Package and test our first distributed application: page 26

1.1. Add the volunteer for leadership method inside that method, and define the **znode**⁵ prefix which is going to be the election namespace /c underscore C stands for candidate

⁵**Znode**: Every node in a *ZooKeeper* tree is referred to as a *znode*. It maintains a stat structure that includes version numbers for data changes, acl changes

```
🎯 LeaderElection.java 🚿
                   抗 log4j.properties × 🎹 pom.xml (leader.election)
                   <plugin>
                       <groupId>org.apache.maven.plugins
                       <artifactId>maven-assembly-plugin</artifactId>
                       <version>2.2-beta-5
                       <executions>
                           <execution>
                               <phase>package</phase>
                               <qoals>
                                   <goal>single</goal>
                               </goals>
                           </execution>
                       </executions>
                       <configuration>
                           <archive>
                               <manifest>
                                   <mainClass>LeaderElection</mainClass>
                           </archive>
                           <descriptorRefs>
                               <descriptorRef>jar-with-dependencies</descriptorRef>
                           </descriptorRefs>
                       </configuration>
                   </plugin>
```

1.2. We call Zookeeper create method which takes znode prefix then the data we want to put inside the znode which is an empty byte array in this case

1.3. Next parameter is the access control list



IV) PART 3

Section: Watchers, triggers, and introduction to failure detection

We can see our application successfully connected to Zookeeper, and we have nothing yet since we haven't created the target znode.

```
C:\WINDOWS\system32\cmd.exe — X

ead@1183] - SASL config status: Will not attempt to authenticate using SASL (unknown error)
2022-11-12 17:52:09,824 [myid:localhost:2181] - INFO [main-SendThread(localhost:2181):ClientCnxn$SendThread@1013] - Socket connection established, initiating session, client: /[0:0:0:0:0:0:0:0:1]:61762, server:
localhost/[0:0:0:0:0:0:0:0:1]:2181

JLine support is enabled
2022-11-12 17:52:09,843 [myid:localhost:2181] - INFO [main-SendThread(localhost:2181):ClientCnxn$SendThread@1448] - Session establishment complete on server localhost/[0:0:0:0:0:0:0:1]:2181, session id = 0x100
09a31a740001, negotiated timeout = 30000

WATCHER::

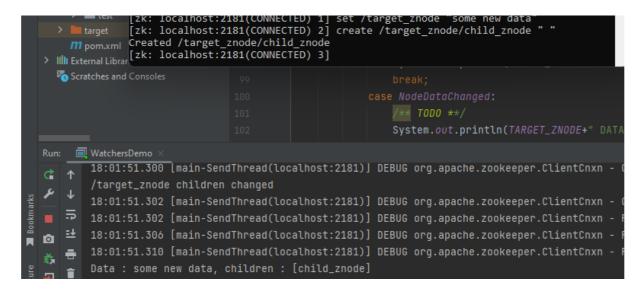
WatchedEvent state:SyncConnected type:None path:null
[zk: localhost:2181(CONNECTED) 0] ls /
[election, workers_service_registry, zookeeper]
[zk: localhost:2181(CONNECTED) 1] create /target_znode "some test data"

Created /target_znode
[zk: localhost:2181(CONNECTED) 2] ____
```

Create a target znode with "some test data" in it.

```
[zk: localhost:2181(CONNECTED) 2] create /target_znode "some data test"
[zk: localhost:2181(CONNECTED) 3] set /target znode "some new data"
cZxid = 0x6e
ctime = Fri Nov 11 23:35:48 EST 2022
mZxid = 0x6f
mtime = Fri Nov 11 23:36:51 EST 2022
pZxid = 0x6e
cversion = 0
dataVersion = 1
aclVersion = 0
ephemeralOwner = 0x0
dataLength = 13
[zk: localhost:2181(CONNECTED) 4]
   23:36:51.279 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Got notification sessions.
   23:36:51.279 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Got WatchedEvent stat
   /target_znode data changed
   23:36:51.282 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Reading reply session
   Data : some new data children : []
```

We get notified about a **new data**.



We then create a children znode.

```
[zk: localhost:2181(CONNECTED) 7] rmr /target_znode
[zk: localhost:2181(CONNECTED) 8]
23:38:53.687 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn -
23:38:53.688 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn -
23:38:53.688 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn -
/target_znode was deleted
```

We delete the target znode.

V) Part 4

Section: Leader Reelection Implementation

we update the "electLeader()" method.

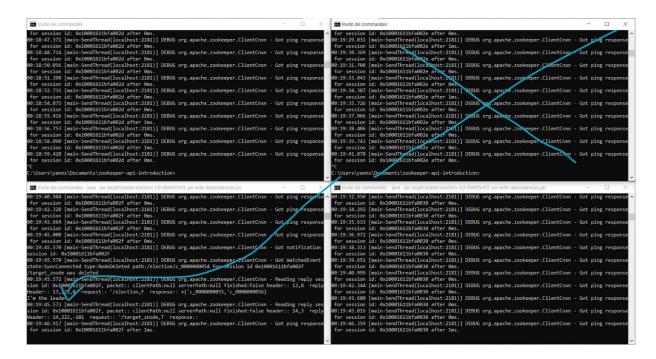


WATCHERs⁶ 1 leader and child watching previous.



WATCHERs kill leader - reelection first child.

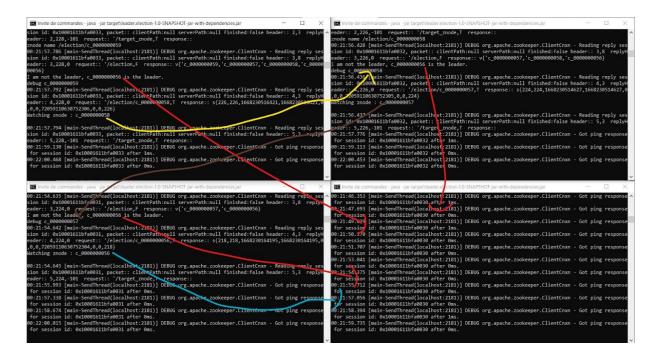
⁶ Watchers: The Zookeeper Watcher utility class is a circuit breaker design pattern around Apache Zookeeper that allows you to watch a one or two files in order to be notified when they are updated. The circuit breaker allows your application the ability to function normally while your connection is interrupted to Zookeeper, and reestablish the connection and watches automatically



WATCHERs kill next leader - reelect first child.



WATCHERs kill next leader - reelect last child.



WATCHERs - launch new children - Same leader, and watching list.



WATCHERS - Kill leader - Reelect again first child in list.



Section: practice- cluster auto-healer using zookeeper

```
private void launchWorkersIfNecessary() throws InterruptedException, KeeperException, IOException {
    /**
    * Implement this method to watch and launch new workers if necessary
    */
    List<String> children;
    do {
        startNewWorker();
        Thread.sleep( millis: 1000);
        children = this.zooKeeper.getChildren(AUTOHEALER_ZNODES_PATH, watcher: this);
        System.out.println(children);
        System.out.println(children.size()<=this.numberOfWorkers);
    } while (children.size()<=this.numberOfWorkers);
}</pre>
```

```
C:\Windows\System32\cmd.e X
21:23:59.381 [main] INFO org.apache.zookeeper.ClientCnxn - zookeeper.request.timeout value is θ. feature enabled=false 21:23:59.389 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.SaslServerPrincipal - Canonicalized address to
21:23:59.393 [main-SendThread(localhost:2181)] INFO org.apache.zookeeper.ClientCnxn - Opening socket connection to serve
r localhost/[0:0:0:0:0:0:0:1]:2181.
21:23:59.393 [main-SendThread(localhost:2181)] INFO org.apache.zookeeper.ClientCnxn - SASL config status: Will not attem
pt to authenticate using SASL (unknown error)

21:23:59.399 [main-SendThread(localhost:2181)] INFO org.apache.zookeeper.ClientCnxn - Socket connection established, ini
tiating session, client: /[0:0:0:0:0:0:0:0:1]:60105, server: localhost/[0:0:0:0:0:0:0:1]:2181

21:23:59.401 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Session establishment request sen
t on localhost/[0:0:0:0:0:0:0:0:1]:2181
21:23:59.427 [main-SendThread(localhost:2181)] INFO org.apache.zookeeper.ClientCnxn - Session establishment complete on server localhost/[0:0:0:0:0:0:0:0:1]:2181, session id = 0x100002e22470000, negotiated timeout = 4000
 Successfully connected to Zookeeper
21:23:59.438 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Reading reply session id: 0x10000
2e22470000, packet:: clientPath:null serverPath:null finished:false header:: 1,3 replyHeader:: 1,132,-101 request:: '/
 workers.F response::
 21:23:59.464 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Reading reply session id: 0x10000
 2e22470000, packet:: clientPath:null serverPath:null finished:false header:: 2,1 replyHeader:: 2,133,0 request:: '/workers,.v{s{31,s{'world,'anyone}}},0 response:: '/workers
 21:24:00.795 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Got ping response for session id: 0x100002e22470000 after 2ms.
 21:24:02.132 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Got ping response for session id:
 0x100002e22470000 after 2ms
21:24:03.470 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Got ping response for session id:
  0x100002e22470000 after 2ms.
21:24:04.806 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Got ping response for session id:
 0x100002e22470000 after 2ms.
 21:24:06.148 [main-SendThread(localhost:2181)] DEBUG org.apache.zookeeper.ClientCnxn - Got ping response for session id:
  0x100002e22470000 after 2ms.
```

As you can see, there is the result of the project completed.

CONCLUSION:

Apache ZooKeeper is a coordination service for distributed application. It has become the solution for high availability for many other projects. Some of Apache's well known open-source distributed services include:

- Apache Hadoop (an umbrella of projects including many components for Big Data processing such as Hadoop Common, Hadoop Distributed File System (HDFS), Hadoop YARN (yet another resource negotiator) and Hadoop MapReduce)
- Apache HBase: non-relational database on top of HDFS
- Apache Hive: data warehouse with SQL-like interface
- Apache Kafka: stream processing
- Apache Nifi: automated data flow processing.

Some of them, such as Nifi, has an embedded implementation of ZooKeeper ensemble if there isn't a separate ensemble. There is some limitation with embedded Zookeeper ensemble. First, we cannot start ZooKeeper without starting Nifi service on the same server. Second, we need to orchestrate the configuration so that the ZooKeeper ensemble does not grow too large. We need to keep in mind that the ZooKeeper ensemble is a separate cluster of its own, and it is not recommended to have more than 7 nodes on ZooKeeper.