Hadoop Common / HADOOP-16829 Über-jira: S3A Hadoop 3.3.1 features / HADOOP-13873

log DNS addresses on s3a init

Details

Type:	Sub-task	Status:	**RESOLVED**
Priority:	Minor	Resolution:	Fixed
Affects Version/s:	2.8.0	Fix Version/s:	3.3.0
Component/s:	fs/s3		
Labels:	None		
Target Version/s:	3.4.0		

Description

HADOOP-13871 has shown that network problems can kill perf, and that it's v. hard to track down, even if you turn up the logging in hadoop.fs.s3a and com.amazon layers to debug.

we could maybe improve things by printing out the IPAddress of the s3 endpoint, as that could help with the network tracing. Printing from within hadoop shows the one given to S3a, not a different one returned by any load balancer.

Issue Links

relates to

HADOOP-13871 ITestS3AInputStreamPerformance.testTimeToOpenAndReadWholeFileBlocks performance...   **RESOLVED**

HADOOP-14551 S3A init hangs if you try to connect while the system is offline   **RESOLVED**

links to

GitHub Pull Request #1955

Activity

Andres Perez added a comment - 08/Dec/16 19:00 - edited

Would adding this to `S3AFileSystem#initialize` be enough?

```
...
    bucket = name.getHost();

    if(LOG.isDebugEnabled()) {
      LOG.debug("Bucket endpoint: "
          + InetAddress.getByName(bucket).toString());
    }
...
```

Steve Loughran added a comment - 05/Sep/17 08:58

yes, that'd work...we wouldn't be logging it except on debug, so eliminating rdns on the main path

but, we should use `NetUtils.createSocketAddrForHost()` for its handling of problems, even if that doesn't quite match what's going on in URL handling.

Steve Loughran added a comment - 05/Sep/17 09:04

Actually, HADOOP-14551 highlights that the AWS SDK doesn't seem to handle DNS lookup failure (and this isn't an rDNS, is it?), so an early check for hostname lookup and fail fast if wrong could actually be better.

Steve Loughran added a comment - 12/Feb/18 12:56

+list https proxy settings being picked up in httpclient

Steve Loughran added a comment - 07/Apr/20 13:06

do in bucket existence check or async -DNS delays shouldn't hold up init

---

Hudson added a comment - 17/Apr/20 13:44

SUCCESS: Integrated in Jenkins build Hadoop-trunk-Commit #18154 (See https://builds.apache.org/job/Hadoop-trunk-Commit/18154/)
HADOOP-13873. log DNS addresses on s3a initialization. (stevel: rev 56350664a76b1ea8e1a942a251880ae3fab12f0c)

- (edit) hadoop-tools/hadoop-aws/src/main/java/org/apache/hadoop/fs/s3a/S3AFileSystem.java
- (edit) hadoop-tools/hadoop-aws/src/main/java/org/apache/hadoop/fs/s3a/Constants.java
- (edit) hadoop-tools/hadoop-aws/src/main/java/org/apache/hadoop/fs/s3a/impl/NetworkBinding.java

People

Assignee:

Mukund Thakur

Reporter:

Steve Loughran

Votes:

0   Vote for this issue

Watchers:

4   Start watching this issue

Dates

Created:

08/Dec/16 11:40

Updated:

17/Apr/20 17:06

Resolved:

17/Apr/20 13:21