

git_comments:

1. * Tests that we can retrieve an HA enabled cluster by only specifying the application id if no * other high-availability.cluster-id has been configured. See FLINK-20866.
2. set cluster-id to app id if not specified

git_commits:

1. **summary:** [FLINK-20866][yarn] Set high-availability.cluster-id to application id if not configured
message: [FLINK-20866][yarn] Set high-availability.cluster-id to application id if not configured In order to easily connect to an HA enabled Yarn cluster, this commit sets the high-availability.cluster-id to the application id if not configured. This is symmetric to how we deploy HA enabled clusters and makes it unnecessary to explicitly set the high-availability.cluster-id if one tries to reconnect to the cluster. This closes #14577.
label: code-design

github_issues:**github_issues_comments:****github_pulls:**

1. **title:** [FLINK-20866][yarn] Set high-availability.cluster-id to application id if not configured
body: ## What is the purpose of the change In order to easily connect to an HA enabled Yarn cluster, this commit sets the high-availability.cluster-id to the application id if not configured. This is symmetric to how we deploy HA enabled clusters and makes it unnecessary to explicitly set the high-availability.cluster-id if one tries to reconnect to the cluster. ## Verifying this change Added `YARNHighAvailabilityITCase.testClusterClientRetrieval` ## Does this pull request potentially affect one of the following parts: - Dependencies (does it add or upgrade a dependency): (yes / **no**) - The public API, i.e., is any changed class annotated with `@Public(Evolving)` : (yes / **no**) - The serializers: (yes / **no** / don't know) - The runtime per-record code paths (performance sensitive): (yes / **no** / don't know) - Anything that affects deployment or recovery: JobManager (and its components), Checkpointing, Kubernetes/Yarn/Mesos, ZooKeeper: (**yes** / no / don't know) - The S3 file system connector: (yes / **no** / don't know) ## Documentation - Does this pull request introduce a new feature? (yes / **no**) - If yes, how is the feature documented? (**not applicable** / docs / JavaDocs / not documented)
label: code-design

github_pulls_comments:

1. Thanks a lot for your contribution to the Apache Flink project. I'm the @flinkbot. I help the community to review your pull request. We will use this comment to track the progress of the review. ## Automated Checks Last check on commit bdf70e96ab2231aabf6e6af3801d314510bfc157 (Thu Jan 07 10:08:04 UTC 2021) ✓ no warnings _{Mention the bot in a comment to re-run the automated checks.} ## Review Progress * ? 1. The [description] looks good. * ? 2. There is [consensus] that the contribution should go into to Flink. * ? 3. Needs [attention] from. * ? 4. The change fits into the overall [architecture]. * ? 5. Overall code [quality] is good. Please see the [Pull Request Review Guide](https://flink.apache.org/contributing/reviewing-prs.html) for a full explanation of the review process.<details> The Bot is tracking the review progress through labels. Labels are applied according to the order of the review items. For consensus, approval by a Flink committer or PMC member is required <summary>Bot commands</summary> The @flinkbot bot supports the following commands: - `@flinkbot approve description` to approve one or more aspects (aspects: `description`, `consensus`, `architecture` and `quality`) - `@flinkbot approve all` to approve all aspects - `@flinkbot approve-until architecture` to approve everything until `architecture` - `@flinkbot attention @username1 [@username2 ..]` to require somebody's attention - `@flinkbot disapprove architecture` to remove an approval you gave earlier </details>
2. <!-- Meta data { "version" : 1, "metaDataEntries" : [{ "hash" : "bdf70e96ab2231aabf6e6af3801d314510bfc157", "status" : "DELETED", "url" : "https://dev.azure.com/apache-flink/98463496-1af2-4620-8eab-a2ecc1a2e6fe/_build/results?buildId=11736", "triggerID" : "bdf70e96ab2231aabf6e6af3801d314510bfc157", "triggerType" : "PUSH" }, { "hash" : "615f3b2a84fb56c696287dd8c1afcff014451d2e", "status" : "SUCCESS", "url" : "https://dev.azure.com/apache-flink/98463496-1af2-4620-8eab-a2ecc1a2e6fe/_build/results?buildId=11745", "triggerID" : "615f3b2a84fb56c696287dd8c1afcff014451d2e", "triggerType" : "PUSH" }] }--> ## CI report: * 615f3b2a84fb56c696287dd8c1afcff014451d2e Azure: [SUCCESS](https://dev.azure.com/apache-flink/98463496-1af2-4620-8eab-a2ecc1a2e6fe/_build/results?buildId=11745) <details> <summary>Bot commands</summary> The @flinkbot bot supports the following commands: - `@flinkbot run travis` re-run the last Travis build - `@flinkbot run azure` re-run the last Azure build </details>
3. Thanks for the review @zentol. I've addressed your comments and updated this PR.

github_pulls_reviews:

1. Should we deduplicate/sync this block with the one in `#startAppMaster`?
2. maybe rename the `clusterId` variable to `appId`
3. **body:** Ok, let me see what I can simplify here.
label: code-design
4. Jup, good idea.
5. **body:** Ok, I've removed `YarnClusterDescriptor.zookeeperNamespace` as it seems to be no longer used.
label: code-design

jira_issues:

1. **summary:** Add how to list jobs in Yarn deployment documentation when HA enabled
description: We suggest users to list the running jobs in a Flink cluster via the following commands. {code:java} List running job on the cluster ./bin/flink list -t yarn-per-job -Dyarn.application.id=application_XXXX_YY {code} However, it could not work when the HA is enabled. The root cause is that GenericCLI do not override the "high-availability.cluster-id" with specified application id. The correct command could look like following. {code:java} flink list --target yarn-per-job -Dyarn.application.id=\$application_id -Dhigh-availability.cluster-id=\$application_id {code} Find more information in the discussion ML[1].
[1]. https://lists.apache.org/thread.html/r0ed905699a182d7b9180ed52d765fcd40caa659ef9640d60aa2ad20e%40%3Cuser.flink.apache.org%3E

jira_issues_comments:

1. I think we can change the behaviour that we always set `{{high-availability.cluster-id}}` to the application id if it has not been set. Then accessing the cluster is symmetric to deploying a cluster. Only if one has chosen an explicit cluster-id, one needs to specify it then.
2. Make sense. Then users do not need to specify the `{{high-availability.cluster-id}}` which does not mean to be set explicitly.
3. Fixed via 1.13.0: 1c719c9287f852bbe2f5e2bfc68085cf9152c74e 1.12.1: 46514f38a2e364da30a96444ece45eceb1e09762 1.11.4: 1d15e3afd6a72526ebc571fc535907c77c203c0d