

git_comments:

1. * * An abstract {@link ModelEnforcement} that provides default empty implementations for each method.
2. * Copyright (C) 2016 Google Inc. * * Licensed under the Apache License, Version 2.0 (the "License"); you may not * use this file except in compliance with the License. You may obtain a copy of * the License at * * <http://www.apache.org/licenses/LICENSE-2.0> * * Unless required by applicable law or agreed to in writing, software * distributed under the License is distributed on an "AS IS" BASIS, WITHOUT * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the * License for the specific language governing permissions and limitations under * the License.
3. * * Enforces that all elements in a {@link PCollection} can be encoded using that * {@link PCollection PCollection's} {@link Coder}.
4. * Copyright (C) 2016 Google Inc. * * Licensed under the Apache License, Version 2.0 (the "License"); you may not * use this file except in compliance with the License. You may obtain a copy of * the License at * * <http://www.apache.org/licenses/LICENSE-2.0> * * Unless required by applicable law or agreed to in writing, software * distributed under the License is distributed on an "AS IS" BASIS, WITHOUT * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the * License for the specific language governing permissions and limitations under * the License.
5. * * {@link ModelEnforcement} that enforces elements are not modified over the course of processing * an element. * * <p>Implies {@link EncodabilityEnforcment}.
6. * Copyright (C) 2016 Google Inc. * * Licensed under the Apache License, Version 2.0 (the "License"); you may not * use this file except in compliance with the License. You may obtain a copy of * the License at * * <http://www.apache.org/licenses/LICENSE-2.0> * * Unless required by applicable law or agreed to in writing, software * distributed under the License is distributed on an "AS IS" BASIS, WITHOUT * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the * License for the specific language governing permissions and limitations under * the License.
7. * * Called after a bundle has been completed and {@link TransformEvaluator#finishBundle()} has been * called, producing the provided {@link InProcessTransformResult} and * {@link CommittedBundle output bundles}.
8. * Copyright (C) 2016 Google Inc. * * Licensed under the Apache License, Version 2.0 (the "License"); you may not * use this file except in compliance with the License. You may obtain a copy of * the License at * * <http://www.apache.org/licenses/LICENSE-2.0> * * Unless required by applicable law or agreed to in writing, software * distributed under the License is distributed on an "AS IS" BASIS, WITHOUT * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the * License for the specific language governing permissions and limitations under * the License.
9. * * Called after a call to {@link TransformEvaluator#processElement(WindowedValue)} on the * provided {@link WindowedValue}.
10. * * Enforcement tools that verify that executing code conforms to the model. * * <p>ModelEnforcement is performed on a per-element and per-bundle basis. The * {@link ModelEnforcement} is provided with the input bundle as part of * {@link ModelEnforcementFactory#forBundle(CommittedBundle, AppliedPTransform)}, each element * before and after that element is provided to an underlying {@link TransformEvaluator}, and the * output {@link InProcessTransformResult} and committed output bundles after the * {@link TransformEvaluator} has completed. * * <p>Typically, {@link ModelEnforcement} will obtain required metadata (such as the {@link Coder} * of the input {@link PCollection} on construction, and then enforce per-element behavior * (such as the immutability of input elements). When the element is output or the bundle is * completed, the required conditions can be enforced across all elements.
11. * * Called before a call to {@link TransformEvaluator#processElement(WindowedValue)} on the * provided {@link WindowedValue}.
12. * * Creates {@link ModelEnforcement} instances for an {@link AppliedPTransform} on an input * {@link CommittedBundle bundle}. {@link ModelEnforcement} instances are created before the * {@link TransformEvaluator} is created.
13. * Copyright (C) 2016 Google Inc. * * Licensed under the Apache License, Version 2.0 (the "License"); you may not * use this file except in compliance with the License. You may obtain a copy of * the License at * * <http://www.apache.org/licenses/LICENSE-2.0> * * Unless required by applicable law or agreed to in writing, software * distributed under the License is distributed on an "AS IS" BASIS,

- WITHOUT * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the * License for the specific language governing permissions and limitations under * the License.
14. ** Tests for {@link EncodabilityEnforcementFactory}.
 15. * Copyright (C) 2016 Google Inc. ** Licensed under the Apache License, Version 2.0 (the "License"); you may not * use this file except in compliance with the License. You may obtain a copy of * the License at ** <http://www.apache.org/licenses/LICENSE-2.0> ** Unless required by applicable law or agreed to in writing, software * distributed under the License is distributed on an "AS IS" BASIS, WITHOUT * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the * License for the specific language governing permissions and limitations under * the License.
 16. ** Tests for {@link ImmutabilityEnforcementFactory}.
 17. * Copyright (C) 2016 Google Inc. ** Licensed under the Apache License, Version 2.0 (the "License"); you may not * use this file except in compliance with the License. You may obtain a copy of * the License at ** <http://www.apache.org/licenses/LICENSE-2.0> ** Unless required by applicable law or agreed to in writing, software * distributed under the License is distributed on an "AS IS" BASIS, WITHOUT * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the * License for the specific language governing permissions and limitations under * the License.

git_commits:

1. **summary:** Merge pull request #173 from tgroh/ippr_model_enforcement
message: Merge pull request #173 from tgroh/ippr_model_enforcement Add ModelEnforcements

github_issues:

github_issues_comments:

github_pulls:

1. **title:** [BEAM-151] Add support for RunnableOnService validation tests to runners module
body: Be sure to do all of the following to help us incorporate your contribution quickly and easily: - [x] Make sure the PR title is formatted like: `[BEAM-<Jira issue #>] Description of pull request` - [x] Make sure tests pass via `mvn clean verify`. (Even better, enable Travis-CI on your fork and ensure the whole test matrix passes). - [x] Replace "<Jira issue #>" in the title with the actual Jira issue number, if there is one. - [x] If this contribution is large, please file an Apache [Individual Contributor License Agreement] (<https://www.apache.org/licenses/icla.txt>). --- Default implementation for Dataflow requires you to specify the runnableOnServicePipelineOptions system property with a valid project/runner/... to be able to execute on Google Cloud Dataflow. Default implementation for Flink and Spark are to have it disabled. Note the whitespace formatting fix on spark pom.xml

github_pulls_comments:

1. R: @davorbonaci CC: @amitsela CC: @mxm Davor, after merge, please update Jenkins build: mvn -pl runners/google-cloud-dataflow-java -am clean verify -DrunnableOnServicePipelineOptions=... There - DrunnableOnServicePipelineOptions replaces -DdataflowOptions
2. LGTM
3. **body:** @davorbonaci Did you perform the cleanups on merge?
label: code-design
4. Yup.

github_pulls_reviews:

1. Version shouldn't be needed.
2. This profile disables automatic execution of RunnableOnService integration tests on Google Cloud Dataflow service. However, the tests will be executed if `runnableOnServicePipelineOptions` is specified.
3. used to be `both`, not `all`. Just confirming it is intentional.
4. Yes, intentional. both has been deprecated

jira_issues:

1. **summary:** Create Dataflow Runner Package

description: Move Dataflow runner out of SDK core and into new Dataflow runner maven module.

jira_issues_comments:

1. The plan is to create a java-dataflow runner maven module and move over the DataflowPipelineRunner and relevant support code over to this new module. So the directory structure would be updated as follows. runners/ flink/ java-dataflow/ pom.xml spark/
2. **body:** To be able to start moving this stuff out, there are several places in the SDK core where things assume they have access to a reference to the DataflowPipelineRunner and or one of the relevant libraries. Will need to do a case by case clean-up.
label: code-design
3. GitHub user lukecwik opened a pull request: <https://github.com/apache/incubator-beam/pull/78> [BEAM-151] Clean up reference to DataflowWorkerHarnessOptions This prevents moving DataflowWorkerHarnessOptions to Dataflow Runner maven module. You can merge this pull request into a Git repository by running: \$ git pull <https://github.com/lukecwik/incubator-beam> dataflow-runner Alternatively you can review and apply these changes as the patch at: <https://github.com/apache/incubator-beam/pull/78.patch> To close this pull request, make a commit to your master/trunk branch with (at least) the following in the commit message: This closes #78 ---- commit e8abd3f93aa0c37b7a96712276310c6c8d7b5381 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-25T22:28:43Z [BEAM-151] Clean up reference to DataflowWorkerHarnessOptions This prevents moving DataflowWorkerHarnessOptions to Dataflow Runner maven module. ----
4. GitHub user lukecwik opened a pull request: <https://github.com/apache/incubator-beam/pull/87> [BEAM-151] Work towards moving Dataflow pipeline runner to separate maven module Break out tests to Dataflow specific variants to make move of Dataflow pipeline runner to new maven module. Also, swap usage of DataflowPipelineOptions for other options interfaces which were better suited. For example, creating PubsubOptions. You can merge this pull request into a Git repository by running: \$ git pull <https://github.com/lukecwik/incubator-beam> dataflow-runner-2 Alternatively you can review and apply these changes as the patch at: <https://github.com/apache/incubator-beam/pull/87.patch> To close this pull request, make a commit to your master/trunk branch with (at least) the following in the commit message: This closes #87 ---- commit 558c5ad33e0384c0f5e9bc3ac36f6c1fde5e6a00 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-25T23:09:18Z [BEAM-151] Break out Dataflow runner dependency to separate test file This allows for moving the Dataflow specific portion of the test to the Dataflow runner maven module. commit e60e42049445b54f23e33364bb9d0828c5660b8e Author: Luke Cwik <lcwik@google.com> Date: 2016-03-25T23:19:54Z [BEAM-151] Break out Dataflow runner dependency to separate test file This allows for moving the Dataflow specific portion of the testing to the Dataflow runner maven module. commit a8a341946e17466b8faf72981c0da70fda5702b6 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-25T23:28:36Z [BEAM-151] Break out Dataflow runner specific tests to separate file This prevents moving the Dataflow runner code to its own separate maven module. commit 7b66b193ca36e6e77124eef3fd20d3e78785bfc6 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-25T23:29:48Z [BEAM-151] Fixup of comment in test file commit 7a583cb21cf830e47ced00b19e9efc9d17f133bb Author: Luke Cwik <lcwik@google.com> Date: 2016-03-28T17:02:01Z [BEAM-151] Migrate options relevant to pubsub to its own options interface This prevents moving DataflowPipelineOptions to the Dataflow runner maven module. commit aa405d2591105172c1ba9b84060d17c262ba6179 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-28T20:09:33Z [BEAM-151] Break out Dataflow transport creation to another file This prevents moving DataflowPipelineOptions into a Dataflow runner maven module. commit e88059f46183a3cd67a9da7fd1e1d493b86724f8 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-28T20:11:28Z [BEAM-151] Remove dependence on num workers in DatastoreIO This prevents moving DataflowPipelineOptions into a separate maven module. commit d4fa5f7abce953272c31da355b0c00c0be2bdd09 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-28T20:19:53Z [BEAM-151] Remove references to DataflowPipelineOptions Removed DatastoreIO dependence on numWorkers from DataflowPipelineWorkerHarnessOptions. This prevented moving DataflowPipelineOptions to a separate maven module. commit 52a4ecbeb4f2aab3db4023e10ddbced2ac70b690 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-28T20:26:19Z [BEAM-151] Remove references to Dataflow runners This prevents moving DataflowPipelineRunner and subclasses to runner specific maven module. commit ab8e54d383716cc82d49b318f0c1012157a68607 Author: Luke Cwik <lcwik@google.com> Date: 2016-03-28T21:16:34Z [BEAM-151] Remove dependence on DataflowPipelineOptions This prevents moving DataflowPipelineOptions to its own runner specific maven module. ----

5. Github user asfgit closed the pull request at: <https://github.com/apache/incubator-beam/pull/87>
6. Github user asfgit closed the pull request at: <https://github.com/apache/incubator-beam/pull/78>
7. GitHub user lukecwik opened a pull request: <https://github.com/apache/incubator-beam/pull/154> [BEAM-151] Remove DataflowPipelineRunner refs in TestPipeline/PAssert Be sure to do all of the following to help us incorporate your contribution quickly and easily: - [x] Make sure the PR title is formatted like: `[BEAM-<Jira issue #>] Description of pull request` - [x] Make sure tests pass via `mvn clean verify`. (Even better, enable Travis-CI on your fork and ensure the whole test matrix passes). - [x] Replace "<Jira issue #>" in the title with the actual Jira issue number, if there is one. - [x] If this contribution is large, please file an Apache [Individual Contributor License Agreement] (<https://www.apache.org/licenses/icla.txt>). --- You can merge this pull request into a Git repository by running: `$ git pull https://github.com/lukecwik/incubator-beam dataflow-runner-4` Alternatively you can review and apply these changes as the patch at: <https://github.com/apache/incubator-beam/pull/154.patch> To close this pull request, make a commit to your master/trunk branch with (at least) the following in the commit message: This closes #154 ---- commit e403691b01e8deabd424bade0127e873004627f9 Author: Luke Cwik <lcwik@google.com> Date: 2016-04-08T22:07:52Z [BEAM-151] Remove dependence on DataflowPipelineRunner in PAssert/TestPipeline Added the ability for TestDataflowPipelineRunner to throw an AssertionError when the pipeline fails using the first job error message. This allows for moving Dataflow to a new package. commit d84b921176e926cd71c540285f07fd37f921b0ec Author: Luke Cwik <lcwik@google.com> Date: 2016-04-08T22:49:53Z [BEAM-151] Expand javadoc in TestPipeline explaining usage Update TestDataflowPipelineRunner to capture all the failure messages. ----
8. Github user asfgit closed the pull request at: <https://github.com/apache/incubator-beam/pull/154>
9. GitHub user lukecwik opened a pull request: <https://github.com/apache/incubator-beam/pull/158> [BEAM-151] Move maven archetypes build order to be after runners Be sure to do all of the following to help us incorporate your contribution quickly and easily: - [x] Make sure the PR title is formatted like: `[BEAM-<Jira issue #>] Description of pull request` - [x] Make sure tests pass via `mvn clean verify`. (Even better, enable Travis-CI on your fork and ensure the whole test matrix passes). - [x] Replace "<Jira issue #>" in the title with the actual Jira issue number, if there is one. - [x] If this contribution is large, please file an Apache [Individual Contributor License Agreement] (<https://www.apache.org/licenses/icla.txt>). --- `sdk/java/maven-archetypes` has several dependencies on the `DataflowPipelineRunner`. Until these are refactored out or a released artifact exists, we need to modify the build order. You can merge this pull request into a Git repository by running: `$ git pull https://github.com/lukecwik/incubator-beam dataflow-runner-5` Alternatively you can review and apply these changes as the patch at: <https://github.com/apache/incubator-beam/pull/158.patch> To close this pull request, make a commit to your master/trunk branch with (at least) the following in the commit message: This closes #158 ---- commit ba4cdc5f3f753d6a3fd26f6b6d4e2f9c3d8f7e70 Author: Luke Cwik <lcwik@google.com> Date: 2016-04-09T01:37:46Z [BEAM-151] Move maven archetypes build order to be after runners Currently `sdk/java/maven-archetypes` has several dependencies on the `DataflowPipelineRunner`. This is mainly throw the utility class called `DataflowExampleUtils`. This is a temporary workaround till examples get updated to use only beam SDK. ----
10. Github user asfgit closed the pull request at: <https://github.com/apache/incubator-beam/pull/158>
11. GitHub user lukecwik opened a pull request: <https://github.com/apache/incubator-beam/pull/165> [BEAM-151] Move Dataflow runner to new Dataflow runner maven module Be sure to do all of the following to help us incorporate your contribution quickly and easily: - [X] Make sure the PR title is formatted like: `[BEAM-<Jira issue #>] Description of pull request` - [X] Make sure tests pass via `mvn clean verify`. (Even better, enable Travis-CI on your fork and ensure the whole test matrix passes). - [X] Replace "<Jira issue #>" in the title with the actual Jira issue number, if there is one. - [X] If this contribution is large, please file an Apache [Individual Contributor License Agreement] (<https://www.apache.org/licenses/icla.txt>). --- Note that Flink runner and Java examples both depend on the Dataflow runner still. So I needed to make the Dataflow runner a top level component to satisfy these dependencies. Also, we are currently shading within the Dataflow runner and using the same namespace to reduce the jar size. You can merge this pull request into a Git repository by running: `$ git pull https://github.com/lukecwik/incubator-beam dataflow-runner-5` Alternatively you can review and apply these changes as the patch at: <https://github.com/apache/incubator-beam/pull/165.patch> To close this pull request, make a commit to your master/trunk branch with (at least) the following in the commit message: This closes #165 ---- commit 0c45173b206b97825b6a9c24d095efba110eb405 Author: Luke Cwik <lcwik@google.com> Date: 2016-04-11T19:49:07Z [BEAM-151] Move a large portion of the Dataflow runner to separate maven module Note that Flink runner and Java examples both depend on the Dataflow runner still. So I needed to make the Dataflow runner a top level component to satisfy these dependencies. commit 31ca6c214e0e90a5acd0cfc895b01a97b20a165c Author: Luke Cwik

- <lcwik@google.com> Date: 2016-04-11T19:55:53Z [BEAM-151] !fixup Add comment explaining build order change ----
12. Github user asfgit closed the pull request at: <https://github.com/apache/incubator-beam/pull/165>
13. GitHub user lukecwik opened a pull request: <https://github.com/apache/incubator-beam/pull/170> [BEAM-151] Enable integration tests underneath google-cloud-dataflow-java Be sure to do all of the following to help us incorporate your contribution quickly and easily: - [X] Make sure the PR title is formatted like: `[BEAM-<Jira issue #>] Description of pull request` - [X] Make sure tests pass via `mvn clean verify`. (Even better, enable Travis-CI on your fork and ensure the whole test matrix passes). - [X] Replace "<Jira issue #>" in the title with the actual Jira issue number, if there is one. - [X] If this contribution is large, please file an Apache [Individual Contributor License Agreement] (<https://www.apache.org/licenses/icla.txt>). --- Added surefire.dependencyToScan and surefire.testPipelineOptions configuration options surefire.dependencyToScan adds an additional dependency used to find tests. This allows for java-sdk-all to be added to all runners for running the @Category(RunnableOnService.class) tests. surefire.testPipelineOptions configures the pipeline options used while constructing instances of a TestPipeline. A json list of arguments is expected, for example: ["--runner=MyRunner", "--myRunnerOption=myRunnerValue"] You can merge this pull request into a Git repository by running: \$ git pull <https://github.com/lukecwik/incubator-beam> dataflow-runner-6 Alternatively you can review and apply these changes as the patch at: <https://github.com/apache/incubator-beam/pull/170.patch> To close this pull request, make a commit to your master/trunk branch with (at least) the following in the commit message: This closes #170 ---- commit ec90f511c3e3308ceaa06a34666c2b654b7ef59d Author: Luke Cwik <lcwik@google.com> Date: 2016-04-12T01:58:00Z [BEAM-151] Enable integration tests underneath google-cloud-dataflow-java Added surefire.dependencyToScan and surefire.testPipelineOptions configuration options surefire.dependencyToScan adds an additional dependency used to find tests. This allows for java-sdk-all to be added to all runners for running the @Category(RunnableOnService.class) tests. surefire.testPipelineOptions configures the pipeline options used while constructing instances of a TestPipeline. A json list of arguments is expected, for example: ["--runner=MyRunner", "--myRunnerOption=myRunnerValue"] ----
14. Github user lukecwik closed the pull request at: <https://github.com/apache/incubator-beam/pull/170>
15. GitHub user lukecwik opened a pull request: <https://github.com/apache/incubator-beam/pull/173> [BEAM-151] Add support for RunnableOnService validation tests to runners module Be sure to do all of the following to help us incorporate your contribution quickly and easily: - [x] Make sure the PR title is formatted like: `[BEAM-<Jira issue #>] Description of pull request` - [x] Make sure tests pass via `mvn clean verify`. (Even better, enable Travis-CI on your fork and ensure the whole test matrix passes). - [x] Replace "<Jira issue #>" in the title with the actual Jira issue number, if there is one. - [x] If this contribution is large, please file an Apache [Individual Contributor License Agreement] (<https://www.apache.org/licenses/icla.txt>). --- Default implementation for Dataflow requires you to specify the runnableOnServicePipelineOptions system property with a valid project/runner/... to be able to execute on Google Cloud Dataflow. Default implementation for Flink and Spark are to have it disabled. Note the whitespace formatting fix on spark pom.xml You can merge this pull request into a Git repository by running: \$ git pull <https://github.com/lukecwik/incubator-beam> dataflow-runner-7 Alternatively you can review and apply these changes as the patch at: <https://github.com/apache/incubator-beam/pull/173.patch> To close this pull request, make a commit to your master/trunk branch with (at least) the following in the commit message: This closes #173 ---- commit 60d5471f74df1da5979dc31a6a7ee11c6fe1a850 Author: Luke Cwik <lcwik@google.com> Date: 2016-04-13T00:49:07Z [BEAM-151] Add support for RunnableOnService validation tests to runners package Default implementation for Flink and Spark are to have it disabled. Default implementation for Dataflow requires you to specify the runnableOnServicePipelineOptions system property with a valid project/runner/... to be able to execute on Google Cloud Dataflow. commit 86e61c5f91f02340b42e3608b0b949f6384ae3d9 Author: Luke Cwik <lcwik@google.com> Date: 2016-04-13T02:50:49Z [BEAM-151] !fixup Add the runnable-on-service-tests execution commit c1b4f3350b830f726e169999cfe6db0d6b337022 Author: Luke Cwik <lcwik@google.com> Date: 2016-04-13T02:54:45Z [BEAM-151] !fixup Fix comment in TestPipeline ----
16. Github user asfgit closed the pull request at: <https://github.com/apache/incubator-beam/pull/173>
17. Done?
18. No, I still need to update package-info.java and make another pass over the remaining classes to see if there are any stragglers.
19. GitHub user lukecwik opened a pull request: <https://github.com/apache/incubator-beam/pull/243> [BEAM-151] Move over some more Dataflow specific classes. Be sure to do all of the following to help us

incorporate your contribution quickly and easily: - [x] Make sure the PR title is formatted like: `[BEAM-
<Jira issue #>] Description of pull request` - [x] Make sure tests pass via `mvn clean verify`. (Even better, enable Travis-CI on your fork and ensure the whole test matrix passes). - [x] Replace `

<https://github.com/apache/incubator-beam/pull/243.patch> To close this pull request, make a commit to your master/trunk branch with (at least) the following in the commit message: This closes #243 ----
commit 305c002ecbbe031458dcc2da66b4a2a0b5f41174 Author: Luke Cwik <lcwik@google.com> Date: 2016-04-26T18:43:26Z [BEAM-151] Move over some more Dataflow specific classes. Note that users should use proto ByteString instead of RandomAccessData since it provides a safer version of the same functionality. I hoped that I would be able to move over more of the *Cloud* classes and their helpers but they are embedded part of coders. Nothing more can be done here until there is an official Beam representation of a coder decoupled from Dataflow CloudKnownTypes. ----

20. Github user asfgit closed the pull request at: <https://github.com/apache/incubator-beam/pull/243>

21. **body:** Note there are a few more objects that could move over such as CloudObject but they are so tightly integrated with Coders and other Beam primitives that at this point there is not much that is feasible to move. Closing this as done, a future refactor/clean-up issue can be opened as the goals of this issue have been met.

label: code-design