

git_comments:

1. * The default buffer size in bytes ({@value #DEFAULT_BUFFER_SIZE}).

git_commits:

1. **summary:** LUCENE-2453: fix javadocs
message: LUCENE-2453: fix javadocs git-svn-id:
https://svn.apache.org/repos/asf/lucene/dev/branches/branch_4x@1437312 13f79535-47bb-0310-9956-ffa450edef68

github_issues:

github_issues_comments:

github_pulls:

github_pulls_comments:

github_pulls_reviews:

jira_issues:

1. **summary:** Make Index Output Buffer Size Configurable
description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up it's buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of it's maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align it's output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.
label: code-design
2. **summary:** Make Index Output Buffer Size Configurable
description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up it's buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of it's maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align it's output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.
label: code-design
3. **summary:** Make Index Output Buffer Size Configurable
description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up it's buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of it's maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align it's output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

4. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

5. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

6. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

7. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

8. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

9. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered

index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

10. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

11. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

12. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

13. **summary:** Make Index Output Buffer Size Configurable

description: Currently, the buffered index input class allows sub-classes and users thereof to specify a size for the input buffer, which by default is 1024 bytes. In practice, this option is leveraged by the simple file and compound segment index input sub-classes. By the same token, it would be nice if the buffered index output class could open up its buffer size for users to configure. In particular, this would allow sub-classes thereof to align the output buffer size, which by default is 16348 bytes, to that of the underlying directory's data unit. For example, a network-based directory might want to buffer data in multiples of its maximum transmission unit. To use an existing use-case, the file system-based directory could potentially choose to align its output buffer size to the operating system's file block size. The proposed change to the buffered index output class involves defining a one-arg constructor that takes a user-defined buffer size, and a default constructor that uses the currently defined buffer size.

jira_issues_comments:

1. **body:** Patch looks good ! Few comments: * buffer can still be final (and should) since it's only initialized in the ctor * I'd inline checkBufferSize in the ctor * I think that adding the same level of control to BufferedIndexInput would be useful too? In general, I think the size of the buffer (1024) is set like that because larger buffer sizes did not improve the performance. Can you perhaps run on the benchmark

indexing algorithms, w/ the buffer size set to larger values and report the results? It'd be interesting to note if there are any improvements before we open up the API like that.

label: code-design

2. **body:** Hi Shai, To answer your comments: * buffer can still be final (and should) since it's only initialized in the ctor [K] Agreed. It's not like we want to allow the size of the buffer to be changed once it has been instantiated. * I'd inline checkBufferSize in the ctor [K] Done. Again, we only need to check the buffer size one time in the ctor. * I think that adding the same level of control to BufferedIndexInput would be useful too? [K] Actually, the BufferedIndexInput already allows this level of control, and then some. In fact, I plagiarized the #checkBufferSize method from that class, where it is used twice, once in the ctor, and then again in the #setBufferSize method. In theory, we could allow the size of the BufferedIndexOutput's buffer to be reset as well, but in case the buffer is made smaller, we'll have to take care to flush some of the "older" bytes that no longer fit in the buffer. IMO, that was not worth the risk and hassle. I will update the patch momentarily based on the comments above, and keep you posted on the benchmark results. Regards, Karthick

label: code-design

3. Karthick, I'm interested in moving forward with this and LUCENE-2456. Could you perhaps prepare an updated patch that incorporates the comments above?
4. Hi Andrzej, The patch has already been updated to incorporate the comments above. Please let me know if you need anything else.
5. bq. Hi Andrzej, The patch has already been updated to incorporate the comments above. Please let me know if you need anything else. Karthick , you should not delete old patches from the issue, as it makes it hard to follow the issue. Just upload the new patch with same filename and JIRA will automatically gray the old one out, but its still visible.
6. bringing this up-to-date.... I think this is pretty useful for downstream apps though. I plan to commit this soon too.
7. [trunk commit] Simon Willnauer <http://svn.apache.org/viewvc?view=revision&revision=1437295>
LUCENE-2453: Make Index Output Buffer Size Configurable
8. [branch_4x commit] Simon Willnauer <http://svn.apache.org/viewvc?view=revision&revision=1437304>
LUCENE-2453: Make Index Output Buffer Size Configurable
9. [branch_4x commit] Simon Willnauer <http://svn.apache.org/viewvc?view=revision&revision=1437312>
LUCENE-2453: fix javadocs
10. [trunk commit] Simon Willnauer <http://svn.apache.org/viewvc?view=revision&revision=1437311>
LUCENE-2453: fix javadocs
11. Closed after release.