package com.twitter.search.common.schema.earlybird;

import javax.annotation.Nullable;

import com.twitter.search.common.config.Config;

public enum FlushVersion {

/\* =======================================================

\* Versions

\* ======================================================= \*/

VERSION\_0("Initial version of partition flushing."),

VERSION\_1("Added timestamps and corresponding mapper to SegmentData."),

VERSION\_2("Add column stride fields."),

VERSION\_3("Change facet field configuration."),

VERSION\_4("Add per term offensive counters to parallel posting arrays."),

VERSION\_5("Add native photo facet."),

VERSION\_6("Add UserFeature column stride field"),

VERSION\_7("Index segment optimizations; new facet data structures."),

VERSION\_8("Store statuses in memory in Earlybird."),

VERSION\_9("Index from\_user\_ids into a searchable field."),

VERSION\_10("Change from\_user\_id dictionary from fst to mphf"),

VERSION\_11("Write image and video facet in separate lucene field."),

VERSION\_12("Add retweeted status ID to the sparse CSF."),

VERSION\_13("Add isOffensive field for profanity filter."),

VERSION\_14("Fix features column stride field corruption."),

VERSION\_15("Upgrade Lucene version, which has a different FST serialization format."),

VERSION\_16("Remove maxDoc in favor of lastDocID"),

VERSION\_17("Added partition and timeslice identifiers to SegmentData."),

VERSION\_18("Per-term payloads"),

VERSION\_19("Multiple per-doc payload fields"),

VERSION\_20("Unify and fix hash codes"),

VERSION\_21("Super awesome new flexible realtime posting list format."),

VERSION\_22("Added new geo implementation."),

VERSION\_23("Upgrade to Lucene 4.0.0 Final"),

VERSION\_24("Added tweet topic ids."),

VERSION\_25("Turn on skip list for mention facet."),

VERSION\_26("Added new EncodedTweetFeaturesColumnStrideField."),

VERSION\_27("Topic ids facet field."),

VERSION\_28("From-user discover stories skiplist field."),

VERSION\_29("Move tokenized screen name to the new username field"),

VERSION\_30("Enable HF term pairs index."),

VERSION\_31("Remove reverse doc ids."),

VERSION\_32("Switch shared status id CSF to non-sparse long CSF index."),

VERSION\_33("New skip lists for optimized high df posting lists."),

VERSION\_34("Store tweet signature in EarlybirdEncodedFeatures."),

VERSION\_35("Don't store shared status id csf in archive indexes."),

VERSION\_36("Don't store norms."),

VERSION\_37("64 bit user ids."),

VERSION\_38("Index links in archive."),

VERSION\_39("Fix pic.twitter.com image link handling not setting the internal field correctly."),

VERSION\_40("Fix all archive tweets being marked as replies."),

VERSION\_41("Avoid flushing event\_ids field; event clusters are applied as updates."),

VERSION\_42("No position fields refactoring; made a few fields to not use position."),

VERSION\_43("Index private geo coordinates"),

VERSION\_44("Materialize last doc id in HighDFCompressedPostinglists", true),

VERSION\_45("Removing from\_user\_id facets support", true),

VERSION\_46("Guard against badly out of order tweets in the search archive.", true),

VERSION\_47("Added card title and description fields.", true),

VERSION\_48("Added card type CSF.", true),

VERSION\_49("Lucene 4.4 upgrade", true),

VERSION\_50("Put mem-archive back on non-lucene optimized indexes", true),

VERSION\_51("Force index rebuild to fix blank text field. See SEARCH-2505.", true),

VERSION\_52("Refactoring of docValues/CSF.", true),

VERSION\_53("Remove SegmentData.Configuration", true),

VERSION\_54("Fix bad indices caused by SEARCH-2723.", true),

VERSION\_55("Fixed non-deterministic facetIds across restarts. SEARCH-2815.", true),

VERSION\_56("Flush FacetIDMap.", true),

VERSION\_57("Remove LatLonMapper and use standard DocValues instead.", true),

VERSION\_58("Longterm Attribute Optimization.", true),

VERSION\_59("Renamed archive segment names. Current segment is no longer mutable.", true),

// Flush version 60 and 59 have the same format.

// Flush version is increased to trigger a rebuild, because we noticed incomplete segments.

// More details can be found on SEARCH-3664

VERSION\_60("Flush version change to trigger segment rebuild.", true),

VERSION\_61("Adding back from\_user\_id", true),

VERSION\_62("Add retweet facet.", true),

VERSION\_63("Switch to new index API in com.twitter.search.core.earlybird.", true),

VERSION\_64("Sort merge archive day and part-\* data. SEARCH-4692.", true),

VERSION\_65("Fix ID\_FIELD and CREATED\_AT\_FIELD sort order. SEARCH-4004 SEARCH-912 ", true),

VERSION\_66("Rebuild data for 1/5/2015. Data on HDFS fixed as part of SEARCH-5347.", true),

VERSION\_67("Upgrade to Lucene 4.10.3.", true),

VERSION\_68("Switching to Penguin v4", true),

VERSION\_69("Fix 16% archive segments: SEARCH-6073", true),

VERSION\_70("Switching to Penguin v4 for full archive cluster. SEARCH-5302", true),

VERSION\_71("Switching to Penguin v4 for ssd archive cluster.", true),

VERSION\_72("Added Escherbird annotations for full archive.", true),

VERSION\_73("Lucene 5.2.1 upgrade.", true, 0),

VERSION\_74("Hanndle geo scurbbed data and archive geo index accuracy", true, 0),

VERSION\_75("Delete from\_user\_id\_stories from indices", true, 0),

VERSION\_76("Allow multiple index extensions.", true, 0),

VERSION\_77("Removed EarlybirdCodec", true, 0),

// minor version 2: added embedded tweet features

// minor version 3: change embedded tweet features to INC\_ONLY

VERSION\_78("Added 80 bytes of extended features", true, 3),

// minor version 1: SEARCH-8564 - Reference Tweet Author ID, using

// EXTENDED\_TEST\_FEATURE\_UNUSED\_BITS\_2 and EXTENDED\_TEST\_FEATURE\_UNUSED\_BITS\_3

VERSION\_79("Renamed UNUSED\_BIT to HAS\_VISIBLE\_LINK", true, 1),

// minor version 2: SEARCH-8564 / http://go/rb/770373

// Made REFERENCE\_AUTHOR\_ID\_LEAST\_SIGNIFICANT\_INT and

// REFERENCE\_AUTHOR\_ID\_MOST\_SIGNIFICANT\_INT immutable field

VERSION\_80("Facet for links: SEARCH-8331", true, 2),

// minor version 1: added video view count

VERSION\_81("Adding LowDF posting list with packed ints", true, 1),

VERSION\_82("Enabling HighDF posting list with packed ints", true, 0),

// minor version 1: SEARCH-9379 - Added bitset for nullcast tweets

// minor version 2: SEARCH-8765 - Added visible token ratio

VERSION\_83("Add bits in encoded features for media type flags. SEARCH-9131", true, 2),

VERSION\_84("Enable archive rebuild for \_\_has\_links field. SEARCH-9635", true, 0),

// minor version 1: SEARCHQUAL-8130, add engagement v2

VERSION\_85("New archive build gen for missing geo data. SEARCH-9894", true, 1),

VERSION\_86("Added new fields to the index", true, 0),

// During this rebuild both the statuses and the engagement counts were regenerated.

// minor version 1: added quote\_count

VERSION\_87("Periodic archive full rebuild. SEARCH-9423", true, 1),

// minor version 1: make new tokenized user name/handle fields textSearchable

// (see go/rb/847134/)

// minor version 2: added has\_quote

VERSION\_88("Fixing missing day in the full archive index. SEARCH-11233", true, 2),

VERSION\_89("Index and store conversation ids.", true, 0),

VERSION\_90("Fixing inconsistent days in the full archive index. SEARCH-11744", true, 0),

VERSION\_91("Making in\_reply\_to\_user\_id field use MPH. SEARCH-10836", true, 0),

VERSION\_92("Allow searches by any field. SEARCH-11251", true, 0),

// During this rebuild we regenerated engagement counts and merged the annotations in the

// aggregate job.

VERSION\_93("Periodic archive full rebuild. SEARCH-11076", true, 0),

// minor version 1: add ThriftCSFViewSettings.outputCSFType

VERSION\_94("Indexing a bunch of geo fields. SEARCH-10283", true, 1),

VERSION\_95("Removing topic ID fields. SEARCH-8616", true, 0),

// minor version 1: add ThriftCSFViewSettings.normalizationType

VERSION\_96("Enabling conversation ID for all clusters. SEARCH-11989", true, 1),

// minor version 1: set several feature configuration to be correct double type

// minor version 2: set some more feature configuration to be correct double type

// minor version 3: add safety labels SEARCHQUAL-9561

// minor version 4: add weighted engagement counts SEARCHQUAL-9574

// minor version 5: add Dopamine non personalized score SEARCHQUAL-9743

VERSION\_97("Changing CSF type to BOOLEAN for some has\_\* flags.", true, 5),

VERSION\_98("Periodic archive full rebuild. PCM-56871.", true, 1),

VERSION\_99("Removing named\_entities field. SEARCH-13708", true, 0),

// minor version 1: add periscope features (SEARCHQUAL-10008)

// minor version 2: add raw\_earlybird\_score to TweetExternalFeatures (SEARCHQUAL-10347)

VERSION\_100("Upgrade Penguin Version from V4 to V6. SEARCH-12991", true, 2),

// minor version 1: adjust for normalizer type for some engagement counters (SEARCHQUAL-9537)

// minor version 2: add decaying engagement counts and last engaged timestamps (SEARCHQUAL-10532)

VERSION\_101("Add emoji to the index. SEARCH-12991", true, 2),

VERSION\_102("Periodic full archive rebuild. PCM-67851", true, 0),

VERSION\_103("Add liked\_by\_user\_id field. SEARCH-15341", true, 0),

// minor version 1: remove last engaged timestamp with 3-hour increment (SEARCHQUAL-10903)

// minor version 2: add fake engagement counts (SEARCHQUAL-10795)

// minor version 3: add last engaged timestamp with 1-hour increment (SEARCHQUAL-10942)

VERSION\_104("Reverting to the 20170109\_pc100\_par30 build gen. SEARCH-15731", true, 3),

VERSION\_105("Add 3 new fields to archive index for engagement features. SEARCH-16102", true, 0),

// This is the last rebuild based on /tables/statuses. Starting 9/14 this build-gen is powered

// by TweetSource. During this rebuild both statuses and engagement counts were rebuilt.

VERSION\_106("Periodic archive full rebuild. PCM-74652", true, 0),

VERSION\_107("Removing card fields from full archive index.", true, 0),

VERSION\_108("Removing the tms\_id field from all schemas.", true, 0),

VERSION\_109("Removing LAT\_LON\_FIELD from all schemas.", true, 0),

VERSION\_110("Adding the card fields back to the full archive index.", true, 1),

// minor version 1: Add composer source csf field (SEARCH-22494)

VERSION\_111("Adding composer\_source to index. SEARCH-20377.", true, 1),

VERSION\_112("Partial rebuild to fix SEARCH-22529.", true, 0),

VERSION\_113("Full archive build gen 20180312\_pc100\_par30.", true, 0),

VERSION\_114("Fix for SEARCH-23761.", true, 0),

VERSION\_115("Add fields for quoted tweets. SEARCH-23919", true, 0),

// minor version 1: Add 4 bit hashtag count, mention count and stock count (SEARCH-24336)

VERSION\_116("Bump flush version for scrubbing pipeline. SEARCH-24225", true, 1),

VERSION\_117("Add retweeted\_by\_user\_id and replied\_to\_by\_user\_id fields. SEARCH-24463", true, 0),

// minor version 1: Removed dopamine\_non\_personalized\_score (SEARCHQUAL-10321)

VERSION\_118("Adding the reply and retweet source tweet IDs: SEARCH-23702, SEARCH-24502", true, 1),

// minor version 1: add blink engagement counts (SEARCHQUAL-15176)

VERSION\_119("Remove public inferred location: SEARCH-24235", true, 1),

VERSION\_120("Flush extensions before fields when flushing segments.", true, 0),

VERSION\_121("Flush the startingDocIdForSearch field. SEARCH-25464.", true, 0),

VERSION\_122("Do not flush the startingDocIdForSearch field.", true, 0),

VERSION\_123("Renaming the largestDocID flushed property to firstAddedDocID.", true, 0),

VERSION\_124("Use the skip list posting list for all fields.", true, 0),

VERSION\_125("Use hashmap for tweet ID lookup.", true, 0),

VERSION\_126("Use the skip list posting list for all fields.", true, 0),

VERSION\_127("Flushing the min and max doc IDs in each segment.", true, 0),

VERSION\_128("Add card\_lang to index. SEARCH-26539", true, 0),

VERSION\_129("Move the tweet ID mapper to the segment data.", true, 0),

VERSION\_130("Move the time mapper to the segment data.", true, 0),

VERSION\_131("Change the facets classes to work with any doc IDs.", true, 0),

VERSION\_132("Make the CSF classes work with any doc IDs.", true, 0),

VERSION\_133("Removing smallestDocID property.", true, 0),

VERSION\_134("Optimize DeletedDocs before flushing.", true, 0),

VERSION\_135("Add payloads to skiplists.", true, 0),

VERSION\_136("Add name to int pools.", true, 0),

VERSION\_137("Add unsorted stream offset.", true, 0),

VERSION\_138("Switch to the OutOfOrderRealtimeTweetIDMapper.", true, 0),

VERSION\_139("Remove realtime posting lists.", true, 0),

VERSION\_140("Add named\_entity field. SEARCH-27547", true, 0),

VERSION\_141("Flush the out of order updates count.", true, 0),

VERSION\_142("Add named\_entity facet support. SEARCH-28054", true, 0),

VERSION\_143("Index updates before optimizing segment.", true, 0),

VERSION\_144("Refactor TermsArray.", true, 0),

VERSION\_145("Remove SmallestDocID.", true, 0),

VERSION\_146("Add entity\_id facet support. SEARCH-28071", true, 0),

VERSION\_147("Enable updating facets", true, 0),

VERSION\_148("Rename the counter for feature updates to partial updates", true, 0),

VERSION\_149("Stop flushing offsets for sorted updates DL streams.", true, 0),

VERSION\_150("Update the name of the property for the updates DL stream offset.", true, 0),

VERSION\_151("Upgrade Lucene version to 5.5.5.", true, 0),

VERSION\_152("Upgrade Lucene version to 6.0.0.", true, 0),

VERSION\_153("Upgrade Lucene version to 6.6.6.", true, 0),

VERSION\_154("Store the timeslice ID on EarlybirdIndexSegmentData.", true, 0),

VERSION\_155("Do not flush index extensions.", true, 0),

VERSION\_156("Deprecate ThriftIndexedFieldSettings.defaultFieldBoost.", true, 0),

VERSION\_157("Load CREATED\_AT\_CSF\_FIELD into RAM in archive.", true, 0),

VERSION\_158("Added directed at user ID field and CSF.", true, 0),

VERSION\_159("Changing deleted docs serialization format.", true, 0),

VERSION\_160("Add fields for health model scores. SEARCH-31907, HML-2099", true, 0),

VERSION\_161("Switch to the 'search' Kafka cluster.", true, 0),

VERSION\_162("Update Lucene version to 7.0.0.", true, 0),

VERSION\_163("Update Lucene version to 7.7.2.", true, 0),

// minor version 1: add IS\_TRENDING\_NOW\_FLAG

VERSION\_164("Collect per-term stats in the realtime segments.", true, 1),

VERSION\_165("Update Lucene version to 8.5.2.", true, 0),

VERSION\_166("Serialize maxPosition field for InvertedRealtimeIndex", true, 0),

VERSION\_167("Add field for pSpammyTweetScore. HML-2557", true, 0),

VERSION\_168("Add field for pReportedTweetScore. HML-2644", true, 0),

VERSION\_169("Add field for spammyTweetContentScore. PFM-70", true, 0),

VERSION\_170("Add reference author id CSF. SEARCH-34715", true, 0),

VERSION\_171("Add space\_id field. SEARCH-36156", true, 0),

VERSION\_172("Add facet support for space\_id. SEARCH-36388", true, 0),

VERSION\_173("Add space admin and title fields. SEARCH-36986", true, 0),

VERSION\_174("Switching to Penguin v7 for realtime-exp0 cluster. SEARCH-36068", true, 0),

VERSION\_175("Adding exclusive conversation author id CSF", true, 0),

VERSION\_176("Adding card URI CSF", true, 0),

// minor version 1: add FROM\_BLUE\_VERIFIED\_ACCOUNT\_FLAG

// minor version 2: Adding new cluster REALTIME\_CG. SEARCH-45692

VERSION\_177("Adding URL Description and Title fields. SEARCH-41641", true, 2),

/\*\*

\* This semi colon is on a separate line to avoid polluting git blame history.

\* Put a comma after the new enum field you're adding.

\*/;

// The current version.

public static final FlushVersion CURRENT\_FLUSH\_VERSION =

FlushVersion.values()[FlushVersion.values().length - 1];

public static final String DELIMITER = "\_v\_";

/\* =======================================================

\* Helper methods

\* ======================================================= \*/

private final String description;

private final boolean isOfficial;

private final int minorVersion;

/\*\*

\* A flush version is not official unless explicitly stated to be official.

\* An unofficial flush version is never uploaded to HDFS.

\*/

private FlushVersion(String description) {

this(description, false, 0);

}

private FlushVersion(String description, boolean isOfficial) {

this(description, isOfficial, 0);

}

private FlushVersion(String description, boolean isOfficial, int minorVersion) {

this.description = description;

this.isOfficial = isOfficial;

this.minorVersion = minorVersion;

}

/\*\*

\* Returns file extension with version number.

\*/

public String getVersionFileExtension() {

if (this == VERSION\_0) {

return "";

} else {

return DELIMITER + ordinal();

}

}

/\*\*

\* Returns file extension given flush version number.

\* If the flush version is unknown (e.g. higher than current flush version or lower than 0), null

\* is returned.

\*/

@Nullable

public static String getVersionFileExtension(int flushVersion) {

if (flushVersion > CURRENT\_FLUSH\_VERSION.ordinal() || flushVersion < 0) {

return null;

} else {

return FlushVersion.values()[flushVersion].getVersionFileExtension();

}

}

/\*\*

\* Returns a string describing the current schema version.

\* @deprecated Please use {@link com.twitter.search.common.schema.base.Schema#getVersionDescription()}

\*/

@Deprecated

public String getDescription() {

return description;

}

/\*\*

\* Returns the schema's major version.

\* @deprecated Please use {@link com.twitter.search.common.schema.base.Schema#getMajorVersionNumber()}.

\*/

@Deprecated

public int getVersionNumber() {

return this.ordinal();

}

public boolean onOrAfter(FlushVersion other) {

return compareTo(other) >= 0;

}

/\*\*

\* Returns whether the schema version is official. Only official segments are uploaded to HDFS.

\* @deprecated Please use {@link com.twitter.search.common.schema.base.Schema#isVersionOfficial()}.

\*/

@Deprecated

public boolean isOfficial() {

// We want the loading/flushing tests to pass locally even if the version is not meant

// to be an official version.

return isOfficial || Config.environmentIsTest();

}

/\*\*

\* As of now, this is hardcoded to 0. We will start using this soon.

\* @deprecated Please consult schema for minor version. This should only be used to build schema.

\*/

@Deprecated

public int getMinorVersion() {

return minorVersion;

}

}