KIP-749: Add --files and --file-separator options to the ConsoleProducer

This page is meant as a template for writing a KIP. To create a KIP choose Tools->Copy on this page and modify with your content and replace the heading with the next KIP number and a description of your issue. Replace anything in italics with your own description.

Status

Current state: *Under Discussion*

Discussion thread: here

JIRA: here

Please keep the discussion on the mailing list rather than commenting on the wiki (wiki discussions get unwieldy fast).

Motivation

If we have some offline log files, we need to enter data into kafka for analysis and use in subsequent processes.

This situation may not be frequent, but it does exist.

Currently, the producer command line tool can read the content of the file to produce data in this way:

e.g kafka-console-producer.sh --bootstrap-server `hostname`:9092 --topic test-01 < 1.txt (Only supports reading a single file)

Or use a more complex shell to enter the contents of multiple files into kafka.

As far as user habits are concerned, when encountering such problems, they always ask Kafka whether it supports such use at the first time, and then look for other alternative solutions.

If the application layer of the kafka producer command line tool supports reading multi-file data, it will be more convenient for users.

Even if we still only need to read the data of a single file, at least from the application layer to support this function, there will be more choices for users, which does not do any harm.

On the other hand, I want to separate the multi-file data reading from the shell, so that the kafka tool itself has this function.

On different operating system platforms, using different interpreters, reading multi-file scripts needs to consider compatibility, such as the difference between bash and bat.

If the producer command line tool naturally supports reading multiple files, it will be very convenient for users to use this function on the windows platform without learning bat.

So We want to introduce **--files** to the producer command line tool to support reading data from a given **multi-file**,

Multiple files are separated by **comma.** Comma is the only delimiter supported. We will add this prompt information to the option comment.

Public Interfaces

kafka-console-producer.sh:
The producer command will accept **--files** option.

E.g
1.The content in the 1.txt file is:
a
b
c

2.The content in the 2.txt file is:
1
2
3

3.Execute producer command line tool:
./bin/kafka-console-producer.sh --bootstrap-server `hostname`:9092 --topic test-01 --files 1.txt,2.txt

4.Execute consumer command line tool:
./bin/kafka-console-consumer.sh --bootstrap-server `hostname`:9092 --topic test-01 --from-beginning
console output:
a
b
c
1
2
3

Proposed Changes

Implement **FileMessageReader** for the **MessageReader** interfaceThe original **LineMessageReader** will still be the default reader.

When the **--files** option is used, it will switch to **FileMessageReader**.

The current implementation scheme is not complicated, it does not have much difference with Line MessageReader, but in order to distinguish it from Line, I extracted FileMessageReader.

FileMessageReader

FileMessageReader will be used to detect the availability of the input file list and read multiple files.

Compatibility, Deprecation, and Migration Plan

The new options have no effect on existing usage.

Rejected Alternatives

1.Discard the **--files-separator** option: The **comma** separator conforms to the habit of most users, and this habit does not change frequently. Similarly, we will use the comma as the default separator and just follow this rule.