Set Up Kafka with given dataset Producer:

##Prequisites- Set up KAFKA ###Step 1: Download the code Download the 0.8.2.0 release and un-tar it.

```
> tar -xzf kafka_2.10-0.8.2.0.tgz
> cd kafka_2.10-0.8.2.0
```

###Step 2: Start the server Kafka uses ZooKeeper so you need to first start a ZooKeeper server if you don't already have one. You can use the convenience script packaged with kafka to get a quick-and-dirty single-node ZooKeeper instance.

```
> bin/zookeeper-server-start.sh config/zookeeper.properties

[acadgild@localhost ~]$ cd kafka
[acadgild@localhost kafka]$ ls -l
total 55640
drwxr-xr-x. 7 acadgild acadgild     4096 Jun  7 13:48 kafka_2.11-1.1.0
-rw-rw-r--. 1 acadgild acadgild 56969154 Jun  7 13:40 kafka_2.11-1.1.0.tgz
[acadgild@localhost kafka]$ tar -xzf kafka_2.11-1.1.0.tgz
[acadgild@localhost kafka]$ cd kafka_2.11-1.1.0
[acadgild@localhost kafka_2.11-1.1.0]$ bin/zookeeper-server-start.sh
config/zookeeper.properties
```


Now start the Kafka server: In other tab

```
> bin/kafka-server-start.sh config/server.properties

[acadgild@localhost ~]$ cd kafka
[acadgild@localhost kafka]$ ls -l
total 55640
drwxr-xr-x. 7 acadgild acadgild     4096 Mar 24 04:24 kafka_2.11-1.1.0
-rw-rw-r--. 1 acadgild acadgild 56969154 Jun  7 13:40 kafka_2.11-1.1.0.tgz
[acadgild@localhost kafka]$ bin/kafka-server-start.sh config/server.properties
-bash: bin/kafka-server-start.sh: No such file or directory
[acadgild@localhost kafka]$ tar -xzf kafka_2.11-1.1.0.tgz
[acadgild@localhost kafka]$ cd kafka_2.11-1.1.0
```

```
[acadgild@localhost kafka_2.11-1.1.0]$ bin/kafka-server-start.sh
config/server.properties
```



```
Last login: Mon Jun 18 17:12:52 2018 from 10.0.2.2
[acadgild@localhost ~]$ cd kafka
[acadgild@localhost kafka]$ ls -l
total 55640
drwxr-xr-x. 7 acadgild acadgild     4096 Mar 24 04:24 kafka_2.11-1.1.0
-rw-rw-r--. 1 acadgild acadgild 56969154 Jun  7 13:40 kafka_2.11-1.1.0.tgz
[acadgild@localhost kafka]$ bin/kafka-server-start.sh config/server.properties
```



```
[acadgild@localhost kafka]$ tar -xzf kafka_2.11-1.1.0.tgz
[acadgild@localhost kafka]$ cd kafka_2.11-1.1.0
[acadgild@localhost kafka_2.11-1.1.0]$ bin/kafka-server-start.sh config/server.properties
[2018-06-18 18:43:34,293] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2018-06-18 18:43:34,972] INFO starting (kafka.server.KafkaServer)
[2018-06-18 18:43:34,974] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2018-06-18 18:43:35,002] INFO [ZooKeeperClient] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2018-06-18 18:43:35,012] INFO Client environment:zookeeper.version=3.4.10-39d3a4f269333c922ed3db283be479f9deacaa0f, built on 03/23/2017 10:13 GMT (org
e.zookeeper.ZooKeeper)
[2018-06-18 18:43:35,012] INFO Client environment:host.name=localhost (org.apache.zookeeper.ZooKeeper)
```

Now create Topic:

[acadgild@localhost ~]$ cd kafka

[acadgild@localhost kafka]$ ls -l

total 55640

drwxr-xr-x. 7 acadgild acadgild   4096 Mar 24 04:24 kafka_2.11-1.1.0

-rw-rw-r--. 1 acadgild acadgild 56969154 Jun 7 13:40 kafka_2.11-1.1.0.tgz

[acadgild@localhost kafka]$ tar -xzf kafka_2.11-1.1.0.tgz

[acadgild@localhost kafka]$ cd kafka_2.11-1.1.0

[acadgild@localhost kafka_2.11-1.1.0]$ bin/kafka-topics.sh --create --zookeeper localhost:2181 --
replication-factor 1 --partitions 1 --topic dataset_producer

WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide.
To avoid issues it is best to use either, but not both.

Created topic "dataset_producer".


```
Last login: Mon Jun 18 18:42:30 2018 from 10.0.2.2
[acadgild@localhost ~]$ cd  kafka
[acadgild@localhost kafka]$ ls -l
total 55640
drwxr-xr-x. 7 acadgild acadgild     4096 Mar 24 04:24 kafka_2.11-1.1.0
-rw-rw-r--. 1 acadgild acadgild 56969154 Jun  7 13:40 kafka_2.11-1.1.0.tgz
[acadgild@localhost kafka]$ tar -xzf kafka_2.11-1.1.0.tgz
[acadgild@localhost kafka]$ cd kafka_2.11-1.1.0
[acadgild@localhost kafka_2.11-1.1.0]$ bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic dataset_producer

WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either, but not
both.
Created topic "dataset_producer".
```

We can now see that topic if we run the list topic command:

```
> bin/kafka-topics.sh --list --zookeeper localhost:2181
```

##Details: KafkaFileProducer performs following tasks:

1. Read data from dataset/dataset_producer.txt file

2. For each line in the file it creates a message of type (Key,Value)

##Validate: Run KafkaFileProducer and verify that data is loaded into Kafka using following command

```
> bin/kafka-console-consumer.sh --zookeeper localhost:2181 --topic dataset_producer --from-beginning
```