# Spark Batch Processing and Streaming

========================================================================

* With respect to Data we will do two things either we will store the data or process the data
* Processing is nothing but for example if some data which is in the file.suppose we want to read contents of the data which is their in a file that is also called as a process

Suppose if we run a query that is also a process

Generally there are two types of process Batch process and streaming process.

### **Batch processing:**

Processing data in a particular frequency of time. It means setting a limit or a batch to process data.

* It means If the data reaches that particular limit or that particular batch then the data is processed.
* For example if we set a batch in such a way that in every business day at particular time the data should be processed or if the data reaches the particular limit say for eg: 10GB then only data should be processed.
* It is kind of a Historical data analysis it means processing 2 months data, 1year data etc

**The Travel of data In Batch processing is as follows:**



* The data which comes to landing there we can do any data cleaning and editing etc and then we can process that data. Iit means we can change the originality of data.
* One more example is that when a some electronics store wants to know the people who are coming to store are interested in which things so based on some historical data they do some sentimental analysis using some algorithm and based on that they know the things which people are more interested and the store will keep the products according to that.

========================================================================

### Stream processing:

-------------------------------------

* The data which comes from live or real data is called streaming data
* It means the data which comes it immediately gets processed

for example Transaction data when we buy something in ecommerce site

The money which we paid immediately gets updated in our account this is called stream processing.

One more example is the live streaming shows is also example of stream processing

The data which comes from live streaming gets immediately processed

The data travel inn stream processing is



In stream processing the data is first processed and then go to the landing which may be file or database.

Example is Live cricket show

The streaming data process cannot be done in hadoop so we go with spark

In spark we can do both

* Spark streaming enables high\_throughout and fault tolerant stream processing of live stream data
* Streaming context is the entry point for the spark functionality.
* Streaming context Consumes a stream of data in spark and registers an inputstream to produce a receiver object.