**Inventory procurement strategy:**

1. Review the sales data for last 3 month, so that we get the trending product for upcoming sale.
2. Avoid backorder (out of stock) product in the sale.
3. Include overstock product in the sale, we can set a threshold if unit of product is greater than this threshold then it is overstock and it should be mandatorily available in the sale.
4. Finding out most sold category from previous 3-month sales data and include all product within this category in Big day sale except backorders.

We will be using above 4 points for deciding which product will show up in Big sale day.

**Sample Data:**

* I have created a sample data for Product, Category, Order tables as Product.csv, Category.csv, Order.csv respectively.
* Order table (Order.csv) contain past order details which we will use for procurement strategy.
* For simplicity we will reading Order.csv and Product.csv file directly from Spark code instead of creating JDBC connection or using Sqoop.

**Why PySpark?**

* We will be using pyspark to develop this module.
* As data is structured /semi structure we will be using spark Data frame API for transforming data, we are use Data Frame API so that we can be benefited by

Catalyst Optimizer.

* For simplicity we will be saving output in csv format.