**Problem Statement**

A computational system records transaction These transactions are stored in a time series log. Now consider following scenarios,

 **Scenario 1:** In a grocery store, a sales purchase is made for some goods. This transaction with relevant information is recorded in a log by a computer system. This log then contains historical information of all the goods purchased in that store.

 **Scenario 2:** A Cloud platform continuously records the resource utilization of instances per minute and stores this information in the resource utilization log. For example, Instance A, has its CPU, Memory, Disk and Network Bandwidth utilization with a timestamp stored in a log file. This log file then contains the historical data of resource utilization from the time that Instance A was created and till the time it was destroyed.

Scenario 1 and Scenario 2 are completely different and are only related by the fact that forecasting is a useful tool to conduct strategic and optimization planning. Assume relevant variables to understand the problem statement, for example, in scenario 1, overall sales forecasting would require transactional logs of all the sales, and cosmetic sales forecasting would require logs related to cosmetics. Where as in case of scenario 2, predicting future resource utilization for a single instance would require the resource utilization of that single (here) instance A, where as to predict resource utilization of a project/group, resource utilization logs of all the instances in that group (say) instance A, B, C, D, E, F will be required.

Given scenario 1 and 2, How can we use Machine Learning constructs, to implement an intelligent system, that allow the user to see a predicted forecast of transaction. Input to the system is transactional log files. The system is expected to perform required data cleaning and transformation. Then the system is expected to perform required analysis on the processed data and generate results in terms of future predictions. The output of the system should be a predicted transactional log.

**Description:**

As given in problem statement, we are given the transaction logs containing the historical data of the resource utilization per minute of a particular instance from the time it was created and destroyed, various features like Memory allocated, CPU Allocated,

CPU Used, etc. are given in logs. There are groups also consisting of various instances.

A group data will consist many instances and their log files**, I am assuming here that each instance of a group is independent of each other** , My understanding of the problem is that we have to forecast the future log file means we have to predict the future resource utilization for a given instance or given group , like how much memory it is going to take how much CPU will it use , We will predict a transactions log for the future .As we have multiple instances per group we will predict the transaction for each instance of that group.

**Proposed solution:**

As we have a transaction log per instance provided in the dataset , we will use each instance to predict future transaction per instance , As we have data provided per minute utilization of resources we will take this problem as a time series problem , First we will read the data from an instance and load into the data frames for group we have to read the data for each instance and keep it ready for making a prediction for each instance but first we have to train our instance data and keep ready for prediction for each instance by making the pickle file for each instance model and keeping it in a folder per group .

Here I have created two Instance Prediction using VAR model code is in pynb file

**Instance1**

**Instance2**

and a group wise prediction in the file

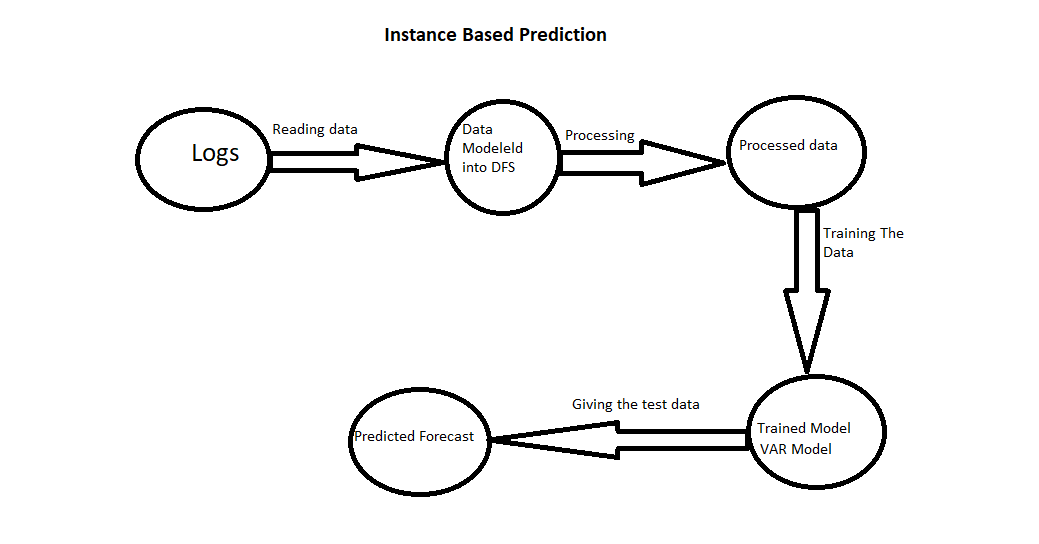
**Group\_wise\_prediction**

Now we come how we will predict the transaction for each instance , as we have loaded the data into data frames ,we will modify our data like adding the columns that match the logs , changing the timestamp to the data time and making it index column , changing the data types etc ,after our data is modeled according to our requirement , we will pre processed our data we will do that by plotting the graphs seeing the information of our data like in our data we saw there are some values which will remain constant all along and are not changing with time , we will drop them and will add them back , Now after our data is processed we will search for the Algorithms on which we have to train our data , as we have multiple features to predict we cant use simple ARIMA or ARMA or SARIMA model so here we are using Vector Auto Regression(VAR) model for making the predictions ,

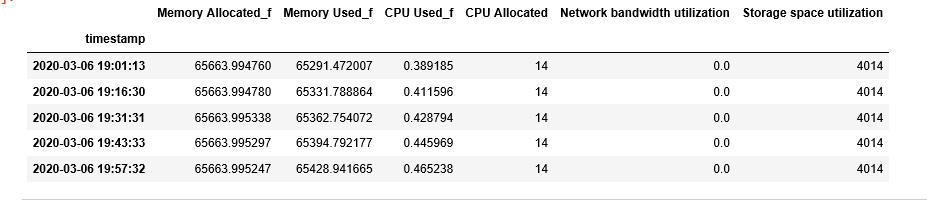
After training the algorithm on our training data we will make predictions on our test data and decide whether it is predicting right or not .As the data is huge I have predicted for two instances of a group and then used it to predict the transactions for that group and each instance ,

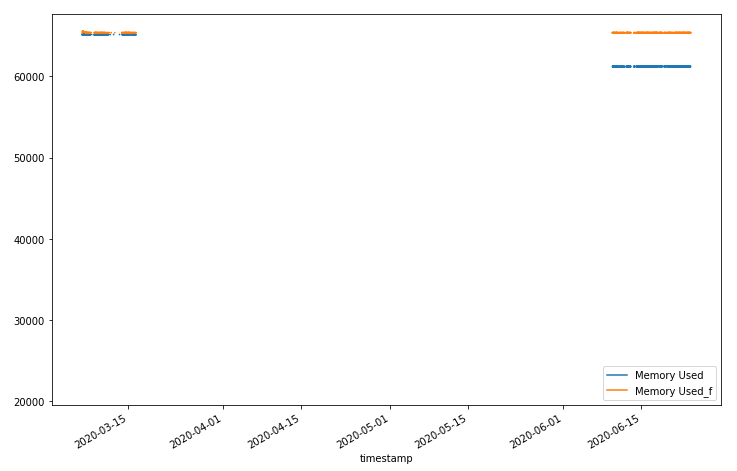
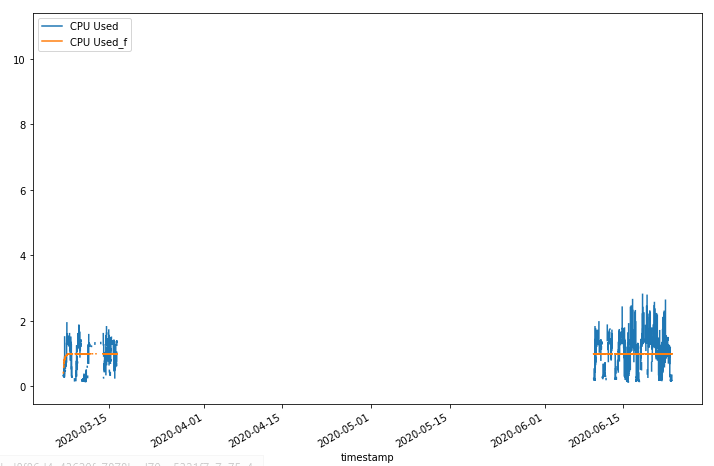
Also we will store the pickle file our models to use them back again for making the predictions.

**Block Diagram:**



**Sample Test Results:**



**Future scope:**

For Better improvement of this model we have very bigger scope, Here we have used simple Vector Auto Regression (VAR) model as our data has more than one feature and if we use ARIMA or ARMA model it will not work efficiently , We can improve VAR model by tuning our lag operator it will reduce our error and make our data more fitter ,

As our data consist of multiple groups and thousand instance and we have work only on 2 instance due to memory restrictions , if we want to make our prediction in one go we have to use Deep learning and libraries like Keras and powerful GPUs to train our model , we can use clustering to train our data parallelly , There are several time series models available in DL like LSTM model for time series model , but for now I have used simpler models and codes to make predictions.

Thanks