Root Cause Analysis on Sudden Restarts of POS Devices and Forecast and Insights to Make Decision on Better POS Manufacturing

Vijayanandan Vithulan, Jegasingam Jeyanthasingam

Department of Computer Science & Engineering
University of Moratuwa

vithulan.20@cse.mrt.ac.lk
jeyayrthasingam.20@cse.mrt.ac.lk

***Abstract*— POS Systems are critical components for businesses and flaws in a system may lead to various consequences. In this report, we present insights for sudden restarts in a POS system. This project is carried out using small parts of the data to prove the hypothesis of relationship between alerts and restarts. We collected 3 different datasets and descriptively, diagnostically and predictively analysed on top of these datasets. We present insights on the resource utilization and improvements. Via diagnostic analysis, we find the alerts associated with sudden restarts . We have predicted the pattern of alerts and restarts for a device and provide insights for the business to predict the future. With the positive results from this experiment, we recommend analysis on large dataset for better results.**

***Keywords*— POS Systems, Inferring causality, Time series analysis, Pattern detection, Future Prediction**

I. Introduction

POS (Point of Sale) system is the system which handles customer payments for the goods offered by a company. This plays a crucial part in a company since all the transactions will be undergone through this POS system.

Flawless operation is expected from a POS system as it is not only associated with processing transactions but also improves the customer experience. In this paper, we have analysed 3 datasets from 10 POS devices. These datasets include restart information, alert information and resource usage information. The dataset information is detailed under the Data Collection Section.

We have used different technologies for this project. We use Kibana for analytics and visualization. This mainly used to provide insights on resource utilizations. Different data analysis techniques are used to validate the relationship between alerts and restarts. Further we went ahead with a conditional probability based approach to find the root cause for the sudden restarts. We used the prophet library provided by facebook to analyze and predict alerts and shutdowns in the future.

The objective of this research is mainly to provide insights on the root cause for sudden restarts including insights on resource usage and critical alerts generated. In this project, we have implemented the following

- Merged three different datasets to get more insights.
- Analysed the resource usage pattern
- Day wise analysis on alerts and restarts
- In-Depth mathematical analysis to infer causality
- Future prediction of restart and alert for a device.
- Insights on the potential root cause and actionable items.

II. Data collection

We have time-series data which starts from 2018 March. Each data point is populated in different frequencies (1 min, 3 min and event-based) and we have these data for approximately 9000 devices. Thus we have approximately 100800 data points per device and 907200000 data points in total up to now (It is a growing data). This data is stored in Elasticsearch Cluster. These data are stored in the Hot Warm and Cold architecture concept. That is, new data (data within the 2 past months) are being stored in Hot nodes and old datas will be moved to Warm and Cold nodes accordingly.

We have decided to use the past 3 months of data of approximately 10 devices to prove the hypothesis and then we can extend our study to further devices. In order to pull the data out from Elasticsearch, we have written a Python script which will retrieve, anonymize and store it in a CSV file for data analytics purposes.

We collected 3 types of data for this project. Restart dataset contains the information about restarts of a device. This includes the occured time, mac address, software versions, etcc… Since there is a scheduled restart everyday, this will have a record everyday.

We collected the datasets containing all the alerts from devices. There are predefined rules for the POS systems for flawless performance. An alert will be generated if the device does not satisfy any of the rules. The alert data consists of alert details and occured time.

We have collected the health dataset for the POS system. This health data contains the resource usage snapshot at a given time. These data are collected with different frequencies.

In order to proceed with these analyses, we had to merge these datasets of different frequencies to one common frequency. For Descriptive analysis we have used the data itself with the default frequency, for Diagnostic analysis, we had to merge 2 different data sets by looking at a 3-hour moving time window and then finally for predictive analysis we had to calculate the daily average and reduce the frequency of the data to per day. All these merged and derived data sets were temporarily stored in Pandas Data Frames during the analysis period.

III. Descriptive Analysis

In this section, we have explored the data and analyzed resource utilization and critical alerts. Since the volume is high and also data is already in the Elasticsearch we have connected a Kibana instance with the existing Elasticsearch cluster. Kibana is an open source analytics and visualization platform designed to work with Elasticsearch. We can use Kibana to search, view, and interact with data stored in Elasticsearch indices. You can easily perform advanced data analysis and visualize your data in a variety of charts, tables. We have used Kibana for the following descriptive analysis subsections.

A. Resource utilization

We have conducted an analysis of CPU, RAM and HD usages in the cause of finding any patterns.

1) CPU Utilization

CPU usage is varying with time but it shows a similar trend every week as shown in Figure 1. We can observe high CPU spikes in POS operation hours and a small spike after the operation hour. It was because of the Couch purge happening every day. We figured it out after discussing with the application team.

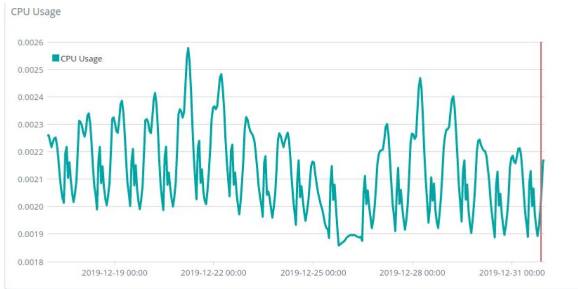

Fig. 1 CPU variation with time

Also, we can observe that there are high CPU spikes on weekends compared to weekdays, which implies higher sales in the restaurants.

2) Memory Usage

Here in Fig. 2, we can observe that some applications in POS have memory leak issues, because of that Memory is constantly increasing with the time and resets after every day scheduled restarts. We should collect more data and identify which application causes the memory leak.

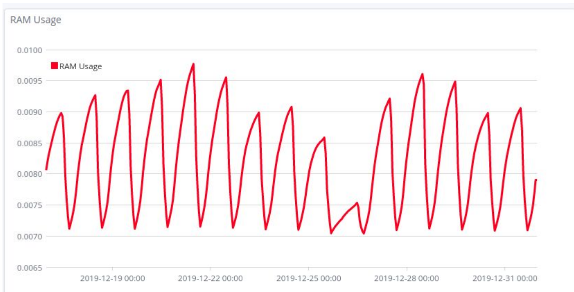

Fig. 2 Memory variation with time

3) Hard disk usage

Here in Figure 3 also we can observe a similar pattern like memory usage (Fig. 2) every day. Ideally, HD usage should increase but here we can see a reduction in HD usage every day because of the Couch purge (We discovered that with Figure 1).

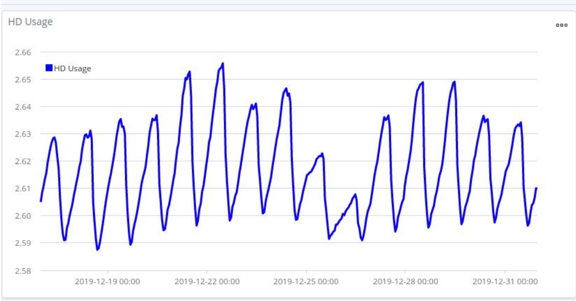

Fig. 3 Hard disk usage against time

B. Predefined rules

An alert will be generated when a predefined rule is violated. We have created a word cloud with most occurrences of alerts for all devices.

Fig. 4 Alerts wordcount by occurrences

As we can see in Fig. 4, Hard disk bad sectors, Touch driver issues and Card reader not found are the top few rule violations we encounter in these POS devices.

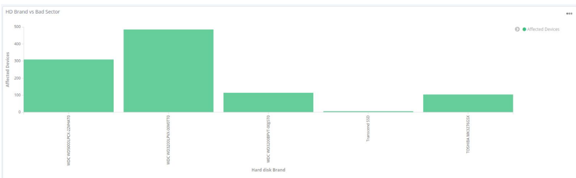

Fig. 5 - Hard disk type vs Bad sector alerts

After identifying that the bad sector is the most burning issue in these POS devices, we dug deeper to identify which brand of the hard disk is causing more bad sector problems. Turns out, as we can see in Figure 5, the WDC brand of hard disks have caused a lot of bad sector problems. Transcend and Toshiba perform better compared to WDC. Therefore we can advise the POS hardware team not to invest more in WDC hard disks and to invest in SSD hard disks.

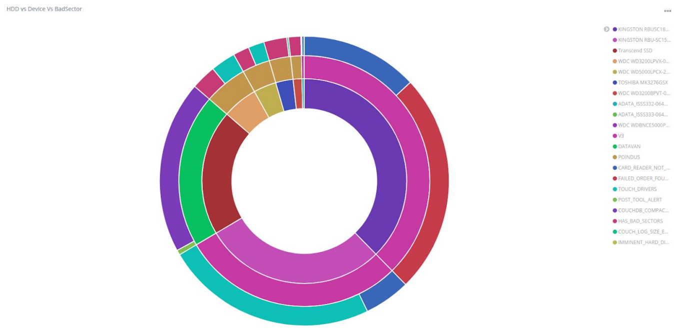

Fig. 6 - Distribution of POS device types against Hard disk model with hard disk bad sector issues

Fig. 6 shows how hard disk brands are distributed among different POS device types, also how this bad sector alert is distributed. We can see that Poindus types of POS devices are installed with WDC hard drives and these are the ones that are experiencing a lot of bad sector problems.

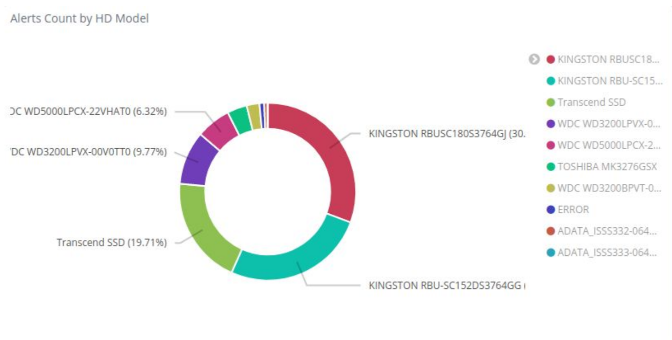

Fig. 7 Alerts distribution with Hard disk type

Fig. 7 shows how alerts are distributed with hard disk models. By looking at this we can see that Kingston type of Hard disk installed devices are facing a lot of issues compared to Transcend or Toshiba.

Therefore we can come to a decision, that it is better to go for Transcend or Toshiba types of hard disks or to Datavan manufacturers. As we have already discussed above the WDC type of Hard disk model has the highest issues in Bad sectors and Kingston has a lot of alerts and Poindus manufacturers are coming with WDC hard disks and V3 manufacturers come with Kingston hard disks.

Then we analysed which restaurants are facing a lot of issues compared to others.

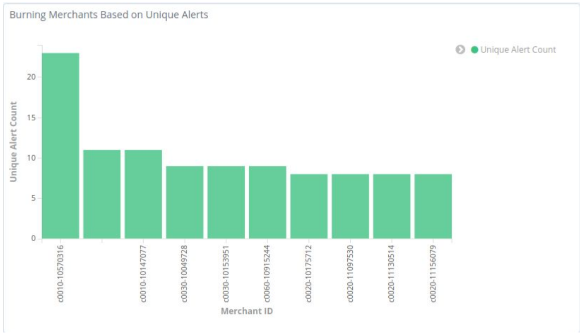

Fig. 8 - Restaurant vs Alert count

By looking at Figure 8, the Customer support team in the organization can reach out to the restaurant and fix their problems and make sure their operation is not obstructed. This may help to reduce the fallouts count. Then we found another pattern in rules violations with sudden restarts.

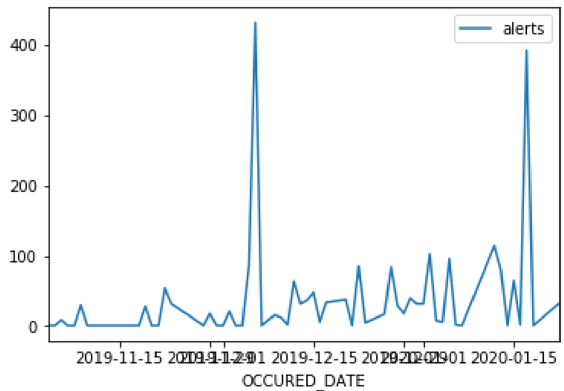

Fig. 9 - Date vs Alert count

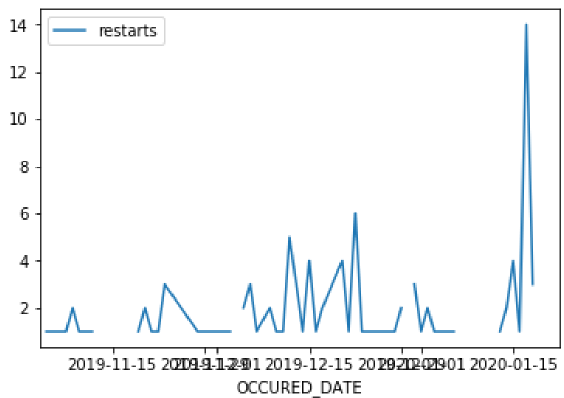

Fig. 10 - Date vs Sudden restarts count

After looking at Fig. 9 and Fig. 10 we can see that both alerts and sudden restarts have a similar pattern and correlations. In other words, sudden restart has something to do with the alerts. Therefore we have decided to find out relations with sudden restarts with respect to the alerts.

IV. Diagnostic Analysis

In this section, we have identified root causes for sudden restarts.

Restart data is created in an event-based manner, i.e whenever there is a restart, a record will be created in the Elasticsearch. Alerts data is created in 3 minutes frequency. The crucial challenge we have faced is to combine these different frequency data sets. Then the second challenge is that alerts and restarts won't occur at the same exact time. There can be 3 hours - 1 Second window that any issue has occurred and the POS might have faced a sudden restart. The third challenge is that the velocity and the volume of the data are high, so it is hard to analyze and find out root causes from all 9000 devices.

In order to overcome the above challenges, we have collected the past three months of data of 10 random devices. Then we merged the 3-minute frequency time series data with the Restarts data by having 3 hours of moving time window. After combining both data we have calculated the conditional probability of each event (alerts) to see the likelihood of both sudden restart and a given alert occurring together. Following are the filtered alerts that occurred with restarts.

TABLE I

Alerts and likelihood of occurring with sudden restarts

Alert	**Likelihood of occurring with sudden restart P(alert|sudden restart)**
NO_IP_ASSIGNED	0.5020
POWER_BUTTON_RESTARTS	0.4296
ROUTER_UNREACHABLE	0.1865
NETWORK_CABLE_UNPLUGGED	0.1663
NETWORK_UNREACHABLE	0.1379

HOST_UNREACHABLE	0.0519
COUCH_REFUSED	0.0465
MASTER_NOT_FOUND	0.0179
CARD_READER_NOT_FOUND	0.0135
RECIEPT_ID_WARNING	0.0075

Therefore, based on these conditional probabilities in Table 1, we can say 50% of the time sudden restarts may have been caused by IP not being assigned to the POS device by the router.

Note that these alerts are not mutually exclusive. There can be correlations within these alerts too.

TABLE 2
OTHER SUDDEN RESTART ALERTS CORRELATION WITH POWER BUTTON RESTART

Alert	**Correlation**
NO_IP_ASSIGNED	0.8520
ROUTER_UNREACHABLE	0.6298
HOST_UNREACHABLE	0.3523
NETWORK_CABLE_UNPLUGGED	0.7975

By looking at Table 2, we can assume that the POS operator may have done a power button restart after POS may have faced the above-mentioned issues in Table 2.

TABLE 3
OTHER SUDDEN RESTART ALERTS CORRELATION WITH IP NOT ASSIGNED ISSUE

Alert	**Correlation**
POWER_BUTTON_RESTART	0.8520
ROUTER_UNREACHABLE	0.3021

Table 3 suggests that some of the IP not assigning issues might have occurred because the router is unreachable, but the other reasons behind NO_IP_ASSIGNED alert remain hidden due to the lack of data.

However we can ask the company to replace the router with a newer version or the restaurant owners to restart the router everyday to reduce NO_IP_ASSIGNED issues. So the probability of POS encountering a sudden restart may reduce.

V. PREDICTIVE ANALYSIS

In this section, we detail the steps followed for the Predictive Analysis. The objective of this section is to predict the future alert count and restart count for a device.

One scheduled restart will be happening for every device. So there should be at least one entry for every day. Since we have data for different devices, we choose one with proper entries so that we do not have any missing values.

Once we extract the data, we process the data to generate the day based information from the alert and restart information. We used the fbprophet library to implement the predictive model.

The dataset passed to the model includes the date and no of alerts/restarts for the day. Further, the future is predicted for one month. All the diagrams from the prediction are given below.

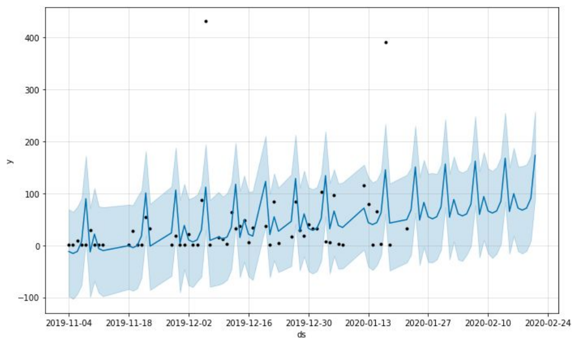

Fig. 11 This shows the existing 3 months pattern and the prediction for the next month for the number of alerts generated per day.

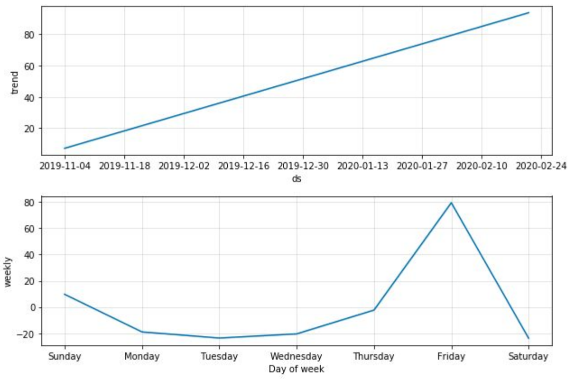

Fig. 12 This shows the trend and seasonality for the prediction of the number of alerts generated per day.

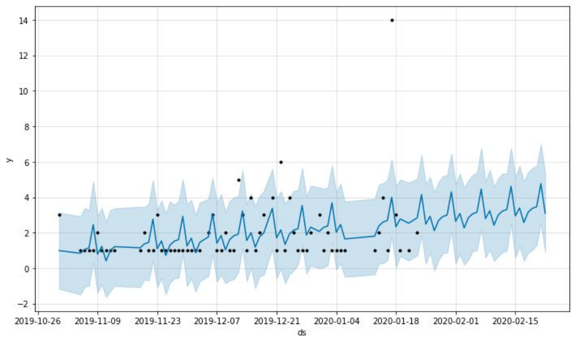

Fig. 13 This shows the existing 3 months pattern and the prediction for the next month for the number of restarts per day.

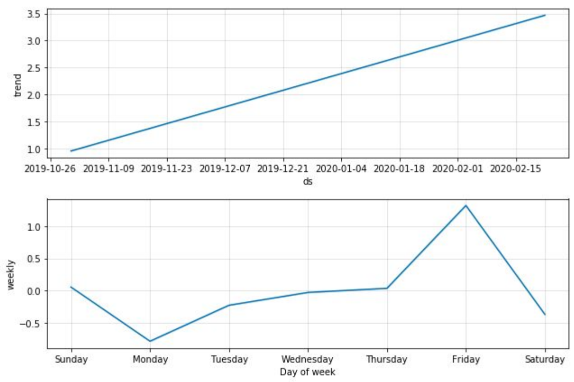

Figure 14 - This shows the trend and seasonality for the prediction of the number of restarts per day.

These predictions can be used to plan for future risks using the POS systems. We can observe an increasing pattern in the trend. This is a piece of important information to take into consideration since in the future the number of restarts and alerts are likely to increase. (trend in Fig. 10 and Fig. 12).

We can see the good seasonality throughout the dataset. This can be further analysed to find the root cause for the sudden restarts. For example, We see high restarts and alerts on Friday. This may be because of the heavy load on Friday as people tend to buy products for the weekend. Since we do not have enough data to analyse this, we could not find the root cause. But this opened avenues to analyse this problem.

VI. Limitations

One of the main limitations of this research is that all the insights are provided based on past information. So any unseen trends or sudden external factors might affect the predictions. For instance, the increasing coronavirus cases would have created a drastic reduction to the predicted values.

Another limitation of this research is fewer data. Since this is a small experiment to prove a hypothesis, this is conducted on a small section of the overall data. These relationships and patterns do not necessarily exist in the data as a whole. So to validate the outcomes we have to conduct the same steps on the large dataset.

VII. Conclusion

Through this research, we have given many insights to sudden restarts in the 10 POS machines. During the descriptive analyses, we found patterns in the resource usage trends and suggested some actionable items to overcome them. Then we have done an in-depth diagnostic analysis, through which we identified the main alert type associated with the sudden restarts and suggested techniques to minimize this. In the Predictive analyses, we have predicted the values for one month and found the trend and seasonality. We have pointed to the risks associated with the trend as well as deduction methodology for which we do not have enough data.

VIII. Future works

As pointed out in the limitations section these insights using a small part of the data does not need to be in the large dataset. So for better performance, these steps have to be duplicated over the whole data set. Prediction for a device can be improved taking many devices and generalizing.

Analyzing the seasonality in the and the trends in depth can be used to find the root cause from the business perspective.

Even the inferring causality is done in this research it is not in-depth. So an in-depth inferring causality analysis (using FBL, CTIR, CCM, etc) will provide better results.