Multi Variate Linear Regression Analysis For Sonatype Nexus Repository Space Management

Kishore Kumar Jinka*[1], Dr. B. Purnachandra Rao[2]

Abstract: Sonatype Nexus is a powerful repository manager widely used in DevOps and continuous integration/continuous delivery (CI/CD) pipelines. It is designed to manage, store, and retrieve binary artifacts efficiently. Nexus plays a crucial role in software development by providing centralized storage and management for build artifacts, dependencies, and containers, enabling teams to collaborate more effectively and maintain control over their software supply chain. As part of the business activity on daily basis , lots of artifacts will be uploaded to number of repositories to each nexus instance. The space will be consumed proportional to the volume of artifacts. It is administrator's responsibility to clear the space as and when it reaches to beyond the thresh hold limit. If the admin misses to clear the space it leads to the situation to shuts down the server, which will effect the entire business. Admin can use the available automated tasks from the nexus admin tasks, but they are having some limitations on the deletion of artifacts, we need to opt only unwanted artifacts. But in the case of high usage of wanted artifacts the space will get decreased and we can't know when the server will go down. This paper resolves this issue by providing the solution using machine learning algorithm (multivariate Linear Regression Analysis) on the usage of space by users. It will findout the regression equation for the given data , so that we can find out prediction value for each actual value. Using the regression equation we can predict the value.

Keywords: *Nexus Repository Manager, NXRM, Release Repository, Snapshot Repository, Docker registry, npm repository, Maven, Nuget, LDAP, Univariate Linear Regression, Feature Engineering, Linear Regression Analysis.*

1. Introduction

Nexus Repository is a centralized platform used to store and manage software artifacts [1], such as libraries, packages, and dependencies, in various formats. It acts as a universal repository manager that supports formats like Maven, npm, NuGet, Docker [2], and more, allowing teams to store, retrieve, and share these artifacts during the software development lifecycle. Nexus [3] stores build artifacts like JAR files, Docker images, etc., making them accessible to development and deployment pipelines. It can proxy external repositories (e.g., Maven Central) and cache dependencies locally, which reduces build times and network usage. Allows teams to store multiple versions of artifacts and manage them through consistent versioning policies. Nexus can act as a private Docker registry [4] for storing and managing container images, ensuring secure access to containerized applications.

Nexus integrates seamlessly with continuous integration and continuous deployment tools like Jenkins [5] and Bamboo, automating the retrieval and publication of build artifacts. It offers fine-grained role-based access control (RBAC) and can scan repositories for vulnerabilities, ensuring compliance with security policies. Nexus is widely used in DevOps pipelines to manage dependencies [6], store build outputs, and ensure a secure software supply chain. Each business unit will have one nexus instance and the users will start uploading artifacts to

[1]*GlobalLogic Inc, VA,USA*
[2] *HCL Technologies , Bangalore, Karnataka, India*
* Corresponding Author Email: kjinkaiitb@gmail.com

corresponding repository. There are snapshot and release repositories. All non prod environment artifacts will be created as snapshot artifacts and the prod artifacts will be created as non snapshot artifacts called release artifacts. ReactJS and Node JS artifacts will be stored in npm hosted repositories.

2. Literature Review

2.1 Sonatype Nexus Repository

Nexus instances are connected to load balancer as shown in the Fig 1. There are different types of clients available to work with nexus repository manager (NXRM). Based on the dev team request NXRM admin [7] will create the hosted repositories for the respective type of deliveries. For Java development hosted repositories SNAPSHOT , RELEASE will be created, for nodes js npm repositories will be created. To upload the docker images docker registries will be created. The Fig: 1 shows the two instances backed to one load balancer and marinating the consistency of the data using blob store and database postgres SQL [8]. NXRM admin will provide access to respective users. If there are number of users for the same repository he will create Active Directory (AD) Groups and add the number of users to the same AD Groups so that he no need to add each user to the repository for access.

The AD group [9] will be given with the corresponding permissions. Java development team will create artifact

type as jar file. This will be uploaded to snapshot or release repository based on the version of the artifact. If it is fixed version it will be uploaded to release repository , if it is snapshot version it will be uploaded to snapshot repository. For web development npm packages will be created and these will be uploaded to npm hosted repository. Dot Net development packages will be uploaded to Nuget repository. Docker images will be uploaded to Docker registry.
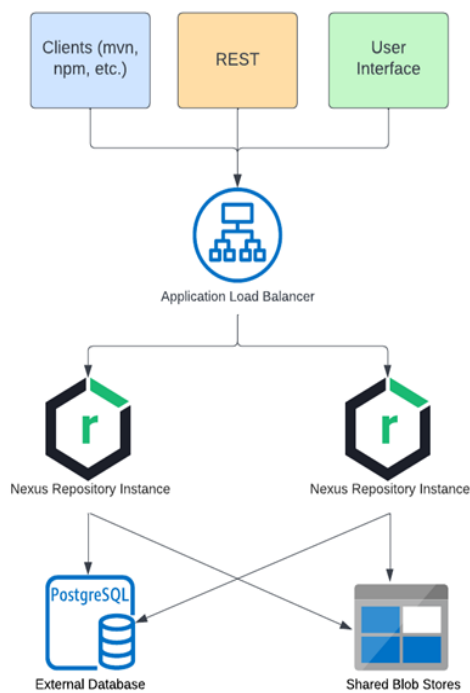

Fig. 1. Sonatype Nexus Repository Management Architecture diagram

2.2   Artifact Storage

On daily basis artifacts from the different deveopers will be uploaded to corresponding repositories. The access has been given by the NXRM admin team to work with the uploads. We can not re upload if it is NON SNAPSHOT artifact. If it is SNAPSHOT version it will be re uploaded.

2.3   Snapshot Repository

In the development environment the code needs to be corrected frequently , so we can not finalize the artifact in the initial cycles. For each delivery or the completion of cycle the artifact will be uploaded. So in this case we need to use SNAPSHOT (Ex: 1.0.0-SNAPSHOT) version so that for each upload of the same artifact the time stamp will be appended to the artifact. For example if we are creating the artifact on the date 10/17/2024 21:03 then the artifact name is abc-1.0.0-171020242103. In this case abc is the name of the artifact. If we do the same build again after 5 minutes then the artifact name is 1.0.0-171020242603. I have assumed the same minutes for easy

explanation. In this case for each build it will create new artifact name. So the rebuild is happening without any issue. Snapshot repository allows rebuild operation. But this is technically wrong statement, why because we are not uploading same artifact each time. Every time we are uploading different artifact.

2.4   Release Repository

Once all cycles are completed in dev environment i.e If we feel that the build is working fine and we are ready to create the artifact , we can change the version of the artifact as fixed version. In this case it doesn't matter about the date of working. The name of the artifact will be abc-1.0.0. No time stamp will be added to artifact name. In this case re upload is not allowed unless and until we need to enable the option to re upload the same artifact. Next release the artifact name will be abc-1.0.1. This will come from the snapshot artifact abc-1.0.1-SNAPSHOT [10].

2.5   Repository Proxying

Different teams are having different repositories in the Nexus instance. Suppose team A wants to use the artifacts of team B then they can use the repositories inside Maven settings.xml [11] file so that these repositories will be scanned based on the dependencies mentioned at pom.xml file. We need to mention groupId , artifactId and version at pom.xml if we want to use any artifact. If the artifact is internal to organization then it will be downloaded from the repos mentioned at corresponding tags in settings.xml file. If the artifact is available at outside of the organization then it will throw the error if the corresponding location is not mentioned at settings.xml file. We need to mention the external urls as proxies inside settings.xml file. Maven will look into local cache for artifact. If it is not available at local cache It will connect with urls internal to organization. If it is not available at internal , then it will connect to external proxies to get the artifacts to local. As soon as it finds the artifacts either at local organization or external to organization it will be downloaded to local cache for future reference and it will be used in the subsequent builds.

2.6   Version Control

Suppose the current artifact name is nexus while creating the development branch in the github the version is nexus-1.0.0-SNAPSHOT. It will be uploaded to nexus SNAPSHOT repository as nexus-1.0.0-171020241136. If we do next build then the artifact name is nexus-1.0.0-171020241138 (this buld is after 2 mins). Once we are done with all these cycles if we fix the version by removing the SNAPSHOT then the build version is nexus-1.0.0 and this will be uploaded to releases repository. Once we are done with this artifact if any one in the organization refers the GAV [12]of this artifact , this will be downloaded to their local from this nexus release repository. For the next cycle the new git hub branch will created with the version at development as nexus-1.0.1-SNAPSHOT. For eachj development build time stamped artifact will be created like nexus-1.0.1-171020241142 , nexus-1.0.0-171020241145 etc. The release artifact nexus-1.0.1 will be created at release build and it will be uploaded to nexus release repository. We can refer any of these versions using the GAV (GroupId, artifactId and version) parameters at pom.xml file.

2.7 Private Docker Registry

Nexus can act as a private Docker registry for storing and managing container images, ensuring secure access to containerized applications. We need to create docker registry [13] at nexus instance as we have created like snapshot and release repositories, and login to this docker registry from the terminal and start pushing the docker images to same registry. We can pull the images back to our location based on our deployments.

2.8 Integration with CI/CD Pipelines

Nexus integrates seamlessly with continuous integration [14] and continuous deployment tools like Jenkins and Bamboo, automating the retrieval and publication of build artifacts. Once the build starts by getting the github code into local , the artifact will get created. This needs to be uploaded to Nexus repository using the Nexus plugin which is already integrated to Jenkins. This should be authenticated using nexus credentials. Once the development team commit the code it will automatically trigger the job to run for build so that the code will be checked out to the base of the Jenkins (CI/CD tool). Once the build is successful , this process will create the artifact based on the type of the pom.xml file i.e it might be jar , war or zip .

This will be saved at the local cache of the machine where the build is happening. As part of the next phase of the build the artifact will be uploaded to Nexus repository. If the build created the snapshot artifact , then it will be uploaded to snapshot repository. If it is release artifact , it will be created to release repository. Even we can do the same process for npm artifacts or docker images. As soon as the docker image got created it will get pushed to docker registry available at the nexus repository. Please find the architecture of the flow what we have discussed so far on the code commit till artifact uploaded to NXRM repository at Fig 2.
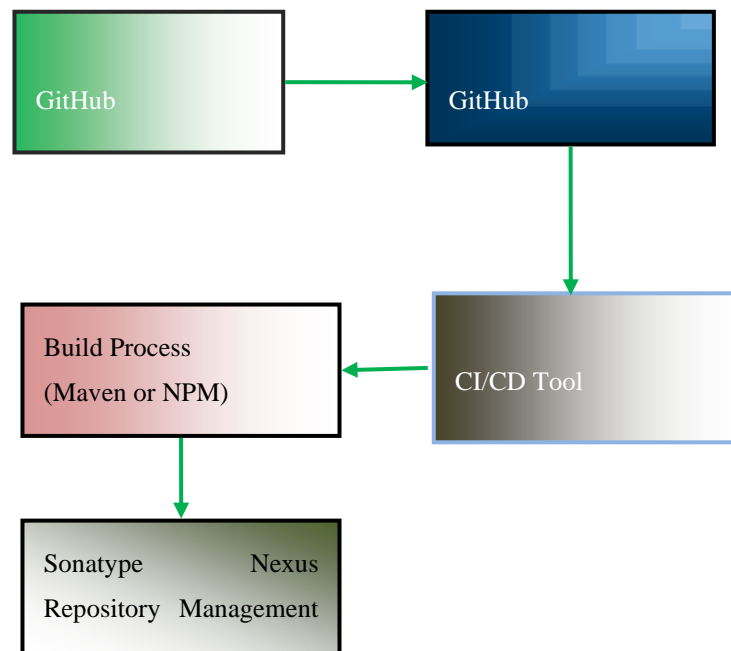

Fig. 2. CI/CD Pipeline Integration

2.9 Security and Access Control:

We can have admin creds for configuring the NXRM instance. Once we are done with config operation we will be asked to update the creds by providing the password in encrypted form. We can have our own creds. We can integrate Linear Data Access Protocol (LDAP) [15] to the NXRM instance so that we can add AD Groups as well to the instance. For each project team one AD Group will be created and added to instance. User management becomes very easy by adding and removing the members to the ADGroup.

SNo	System IP	Time Stamp	Operation	URL	Size	Status
1	10.0.*	100924	PUT	url	150	success
2	10.1.*	110924	DEL	url	200	success
3	10.2.*	120924	GET	url	250	success
4	10.4.*	130924	PUT	url	300	success
5	10.3.*	140924	GET	url	350	success
6	10.6.*	150924	PUT	url	400	success
7	10.5.*	160924	DEL	url	450	success
8	10.4.*	160924	GET	url	500	success

Table 1. Nexus Repository log file entries

For number of repositories Table 1 shows the different type of metrics like name of the repository , time stamp , sub group of the repository, type of the operation GET, PUT or DELETE [16] , volume of the upload, success or failure of the upload. In this we need to segregate the metrics using the type of the operation. Among these operations only PUT operation will directly influence the space of the disk. We need to figure out the total volume of the uploads happening on the daily , weekly and monthly analysis [17]. For example we have taken sample values only having PUT operation for each sub group from the repository. Please find the same in Table 2. The values which we have mentioned in the Table 1 might not be related to values from Table 2.

subgroup	URL	Time Stamp	Size (Gb)	Operation	Status
Sb1	url1	100924	5	PUT	success
Sb2	url2	120924	6	PUT	success
Sb3	url3	130924	10	PUT	success
Sb4	url4	140924	3	PUT	success
Sb5	url5	150924	4	PUT	success
Sb6	url6	160924	8	PUT	success
Sb7	url7	170924	10	PUT	success
Sb8	url8	180924	4	PUT	success

Table 2: Business unit vs Upload operations
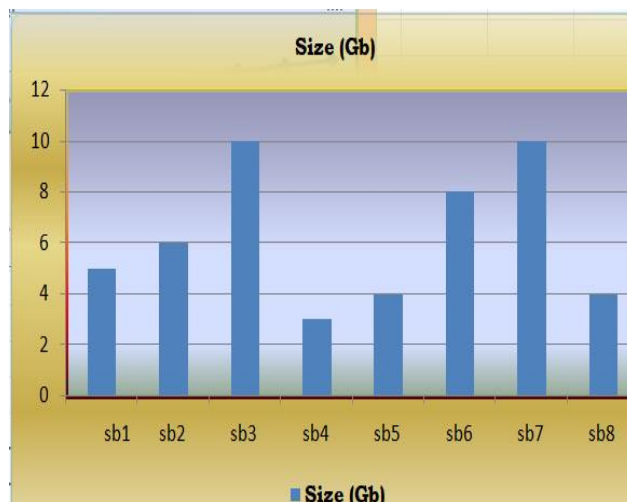

Graph 1: Business unit vs Upload operations

Please observe the usage on daily, weekly , monthly basis. We need to manually delete the unwanted artifacts like snapshots and very old fixed version volumes. If we didn't delete the same , then the upcoming uploads will not be successful which will effect the huge number of developers. Either admin has to delete them manually or by using admin task available at NXRM or he needs to connect with development team to pick the unwanted artifacts and remove the same. The amount of space gaining in the deletion of unwanted artifacts should be in sync with the amount of volume uploaded to NXRM repositories. Since the entire NXRM instance is shared by number of business units and there will be very high volume consumers or low volume consumers. Either we need to provide separate instance for high volume consumers or have them to provide unwanted artifacts list in high frequency so that admin can delete the same to gain the space for further activities. Failing of any of these actions will lead to abnormal failure of uploads by other business units though they are consuming very less amount of space.

Even though we have huge volume available in the beginning of the NXRM activities , one day we need to provide the solution for any of these issues. Instead of working on this process where the manual intervention is very much demanding, if we can predict the usage based on the daily , weekly and monthly usage analysis , it would be easy forecast the expected space required for further activities. If we know well in advance that the available space sufficient only for certain period of time , then it is easy for us to take the appropriate action like alerting the developers to provide unwanted artifacts or increasing the disk space.

If we follow this procedure this will avoid the abnormal incidents like NXRM is not available for operations. This will save the huge number of man hours from development team. If there are 20 members in one team and 10 to 15 teams are depending on the same NXRM then 2400 man hours we will save on daily basis. If there are any abnormal activities because of not having this prediction analysis then we need to face the wastage of man hours. Please find the table shows the man hours analysis.

SNo	Num of developers	Hours	Hours Saved
1	10	80	80
2	20	160	160
3	30	240	240
4	40	320	320

Table 3: Number of developers vs Man hours
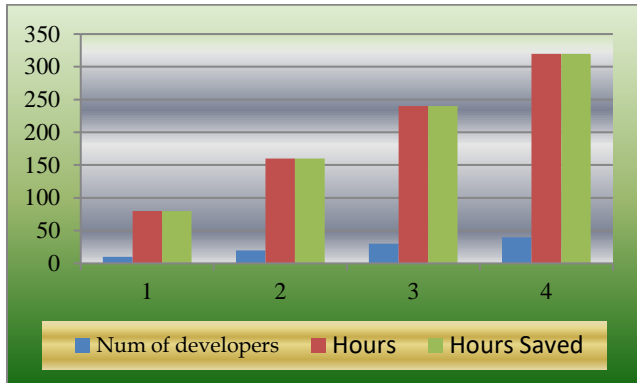
Graph 2: Number of developers vs Man hours

Graph 2 shows that number of hours saved by the developers

once we get any solution to avoid abnormal incidents. What ever the man hours available in each team , 100% savings we can observe.

3. Proposal Method

3.1 Problem Statement

Usage of space using number of uploads will lead to abnormal incidents of the NXRM instance. Instead of waiting till the end of space and facing the abnormal incident we need to predict the space usage on daily , weekly or monthly basis will help us to patch the disk space so that we can save the number of man hours.

3.2 Proposal

Machine Learning algorithm will be used to predict the usage based on the historical analysis of the space usage. We need to consider past one year log files to analyze and figure out the dependent attributes and independent attributes (Feature Engineering). NXRM log files carries repository info in the form of url , time stamp , type of operation , status of the operation, size of the upload, time required to complete the upload operation. We need to analyze the data python data analytics tools like Pandas and numpy to get the exact repository name from the url which the dev team has used to upload the artifacts to nexus repository manager. Finally we will get the repo name , time stamp , artifact size , type of operation and time required to perform the corresponding activity.

In the second time analysis we need to filter the data using type of operation. With this we will get data for PUT

operation (upload operation). Like this we need to get the data for the entire month for each repository, Ex: If there are repositories like TCT , RFR , CFC , PRD , GST, TGB, and SCY. Number of users are available in each business unit (repository name for each business unit), and each user is uploading artifacts to NXRM repository. In the analysis we need to filter the data on repo wise , followed by analysis on monthly basis for each business unit. Please find the data at Table 3 having detailed analysis on what we have described so far.

Ui , middle , back like this they have different type of repo names created inside TCT business unit. The main problem is TCT business unit is facing the space issue. We have started the analysis on getting the repo name and type of operation along with time stamp , size of the artifact for uploading to repository. Monthly wise we will get total volume has been uploaded to each repository. This is how we will get the total space consumption by one business unit (including all repos ui, middle , back end).

SNo	URL	Name	Activity	Size(Gb)	Status
1	http://ui	Ui	PUT	4	Success
2	http://back	Back	GET	---	Success
3	http://middle	Middle	DEL	---	Success
4	http://ui	Ui	GET	---	Success
5	http://back	Back	GET	---	Success
6	http://ui	Ui	GET	---	Success
7	http://middle	middle	PUT	6	Success

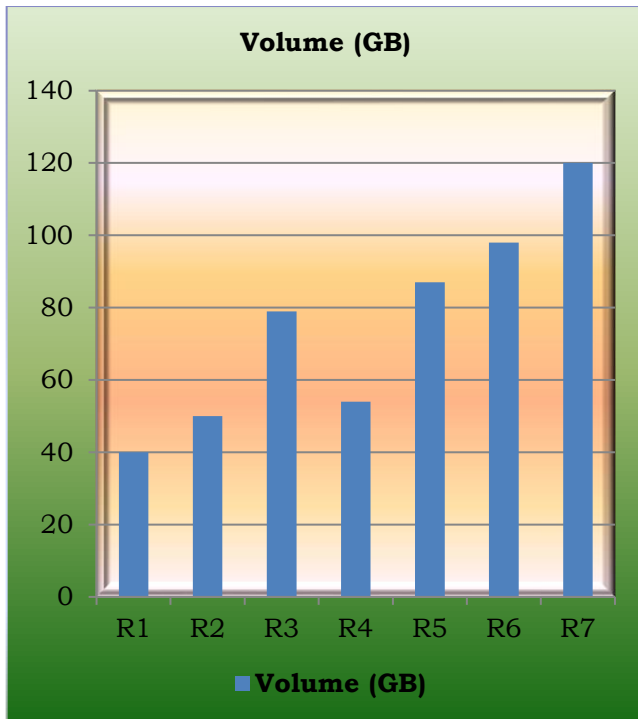**Table 4**: Nexus log file data attributes and values

3.3 IMPLEMENTATION

We have used python numpy , pandas , Scikit learn to analyze the data . The data has been taken based on daily basis , weekly basis and monthly basis usage analysis. Using this we can get the approx space required to run the NXRM instance without facing the abnormal incidents.

Once we have historical usage for the single business unit on the monthly basis , we can use the same parameters in the machine learning algorithm to predict the dependent variable for future activities. Like this we need to do for all business unit and apply the same feature engineering process and figure out the parameters to apply inside machine learning algorithm (Linear Regression Analysis). This will provide prediction on dependent variable . In our case time, CPU usage and number of users are the independent variables and the size of the artifact (total size consumed by all artifacts business unit wise in a month) is the dependent variables.

SNo	Repo Name	Volume (GB)
1	Repo1	40
2	Repo2	50
3	Repo3	79
4	Repo4	54
5	Repo5	87
6	Repo6	98
7	Repo7	120

Table 5: Repo vs Volume consumption

Day	CPU Usage	Number of users	Total space consumption
1	35	1000	12
2	45	1050	14
3	50	1100	16
4	55	1200	18
5	60	1300	21
6	70	1350	27
7	75	1400	27
8	80	1450	29
9	85	1500	36
10	90	1550	39
11	92	1600	42
12	95	1650	55
13	97	1700	60
14	98	1750	65
15	99	1800	70
16	100	1850	75
17	100	1900	80
18	100	1950	85
19	100	2000	90
20	100	2050	95

Table 6: Space consumption on daily basis
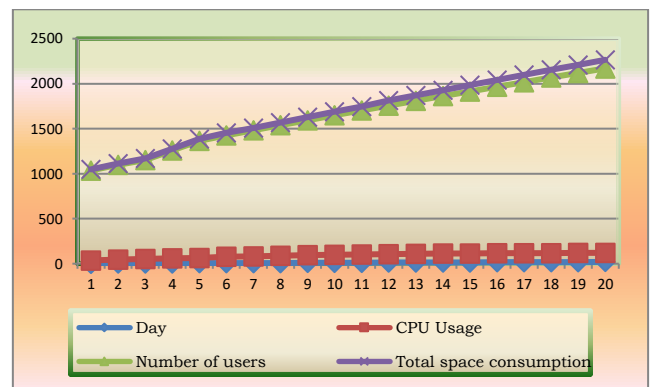
Graph 4: Day, CPU usage, number of users vs Volume consumption

Graph 3: **Repo vs Volume consumption**

Graph3 shows the volume consumption for each repo (Repo1→ R1, Repo2→R2, Repo→ R3, Repo4→ R4, Repo5→ R5, Repo6→ R6 and Repo7→R7). Please find the usage data for all repos in one business unit per one month.

We need to take the consolidate amount space consumed by one business unit in one month. As part of the analysis we have taken daily basis usage for 12 days, so that we will predict the volume required for next x number of days. Please find the table as follows for 12 days in October month. Here we would like to take two more independent variables CPU usage and number of users.

In multivariate linear regression, the formula for estimating the intercept ($\beta 0$) and slopes ($\beta 1, \beta 2, \ldots, \beta n$) involves using the least squares method. This method aims to minimize the sum of the squared differences between the observed and predicted values of the dependent variable. The equation for multivariate regression is

$Y = \beta 0 + \beta 1 X 1 + \beta 2 X 2 + \ldots + \beta n X n + \epsilon$

Y is the dependent variable (e.g., Volume Consumption in our case).

$X1, X2, \ldots, Xn$ are the independent variables (e.g., Number of Users and CPU Usage).$\beta 0$ $0 \beta 0$ is the intercept (the value of Y when all X variables are zero). $\beta 1, \beta 2, \ldots, \beta n, \beta 1, \beta 2, \ldots, \beta n$ are the slopes (coefficients) of the independent variables. ϵ epsilon is the error term.

Y= Total Space Consumption (Gb), X1= Day, X2= CPU Usage, X3= Number of users

This formula allows the calculation of predicted values based on the independent variables (Day, CPU Usage, and Number of Users).

To compute the values of m and b, we use the least squares method [18], which minimizes the sum of the squared differences between the actual and predicted values. The formulas for the slope (m) and intercept (b) are

intercept (β0): 1.6414, Slope for X1 (β1): 1.0881, Slope for X2 (β2): 0.6888

X1 = 1, X2 = 12 The regression equation is

Y=1.6414+1.0881*X1+0.6888*X2 ,

Y=1.6414+1.0881*1+0.6888*12,Y=1.6414+1.0881+8.2656=11.27

X1 = 2, X2 = 14 The regression equation is

Y=1.6414+1.0881*X1+0.6888*X2 ,

Y=1.6414+1.0881*2+0.6888*14,Y=1.6414+2.1762+9.6432=14.20

X1 = 5, X2 = 21 The regression equation is

Y=1.6414+1.0881*X1+0.6888*X2 ,

Y=1.6414+1.0881*5+0.6888*21,Y=1.6414+5.4405+14.4648=21.90

Please find the calculation for couple of actual values 1, 2, 5.

Table 7: shows the actual and predicted values for the daily usage of volume based on the regression equation derived in the previous step.

| Day	Actual values (Gb)	Predicted Values (Gb)
1	12	11.27
2	14	14.2
3	16	16.04
4	18	18.97
5	21	21.9
6	27	25.92
7	27	26.53
8	29	29.47
9	36	33.63
10	39	39.83
11	42	46.02
12	55	52.22

Table 7: Actual vs Predicted values
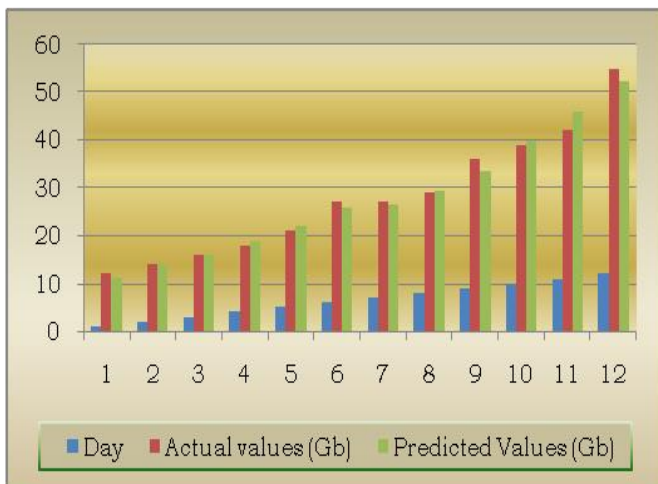
Graph 5: **Actual vs Predicted values**

Table 8 shows the space consumption monthly wise along with cpu usage and number of users.

This shows the data for one year. If required we can go for more samples as well. In this data time , cpu usage and number of users are the independent variables and space is the]dependent variable. We need to find out how much consumption will be there after 4 months or 5 months so that it is easy for us to predict the expected space.

Month	CPU Usage (%)	Users	Space Consumption (GB)
1	30	1000	500
2	32	1050	576
3	34	1100	556
4	36	1150	634
5	38	1200	659
6	40	1250	798
7	42	1300	892
8	45	1350	908
9	47	1400	1300
10	50	1450	1600
11	53	1500	2000
12	55	1550	2400

Table 8: Monthly, CPU usage, number of users vs Volume consumption (First Sample)
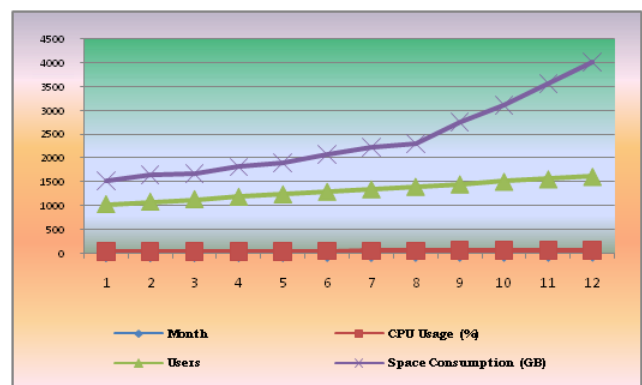
Graph 6: Monthly, CPU usage, number of users vs Space consumption (First Sample)

Graph 6 shows the graph representation of monthly usage analysis. Usage is getting increased on monthly basis even though we have removed number of unwanted artifacts. This is because dev team is not ready to give back old artifacts and they are expecting to preserve them for couple of years.

multivariate Linear Regression:

Y is dependent variable [20] and CPU Usage and Users are independent variables. y=b0+b1·CPU Usage+b2·Users regression

equation where we need to find out the slope b1, b2 and interceptor b0. y is the dependent variable (space consumption), b0 is the intercept, b1 is the coefficient for , CPU usage, b2 is the coefficient for number of users. In our example if we can find out the interceptor and slope , we can predict the volume required for 4 months or 5 months.
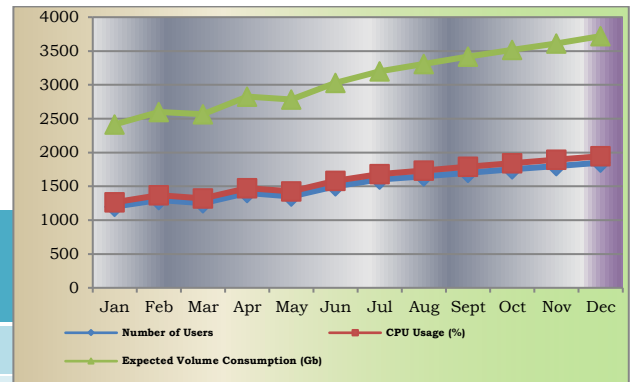
$y=\beta_0+\beta_1\cdot Day+\beta_2\cdot CPU\ Usage+\beta_3\cdot Number\ of\ Users$ Volume$=500+0.5*CPU+1.2*Number\ of\ Users$. Please find the calculation for couple of values:

$500+0.5*500+1.2\cdot473.8=1318.56$,

$500+0.5*892+1.2\cdot787.23=1890.676$,

$500+0.5*1300+1.2\cdot1260.67$, here we are calculating for month one , month 7 and month 9.

Month	Actual Space Consumption (Gb)	Predicted Space Consumption (Gb)
1	500	473.8
2	576	526.04
3	556	578.28
4	634	630.52
5	659	682.76
6	798	735
7	892	787.23
8	908	1208.43
9	1300	1260.67
10	1600	1681.87
11	2000	2103.07
12	2400	2155.31

Table 9: Actual vs Predicted values for monthly analysis (First Sample)

Graph 7: Actual vs Prediction values (First Sample)

Table 11: Actual vs Predicted values for monthly analysis (Second Sample)

Graph 9: Actual vs Prediction values (Second Sample)

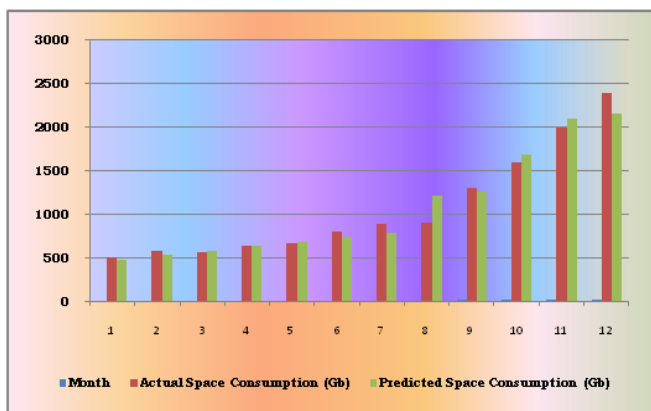Volume Consumption$=-100+0.5\times$(Number of Users)$+10\times$(CPU Usage).

Prediction: Using this equation, you can predict the space consumption for any month within the provided data range.

The positive slope indicates that space consumption tends to increase over time, suggesting a growing usage trend.

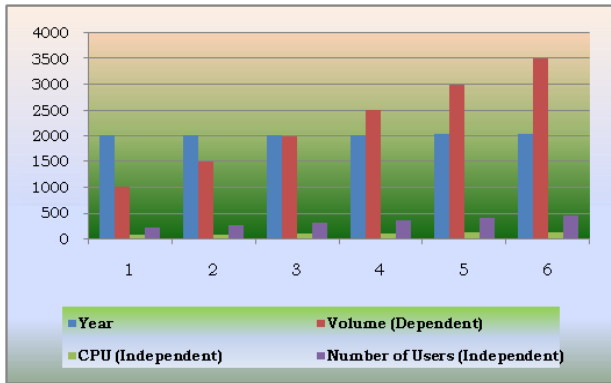We have worked on CFC sample having the following dataset.

Year	Volume	CPU	Number of Users
2020	1000	70	200
2021	1500	80	250
2022	2000	90	300
2023	2500	100	350
2024	3000	110	400
2025	3500	120	450

Table 12: Yearly volume consumption

Graph 10: Yearly volume consumption

Regression Analysis:

Regression equation obtained from the data is

volume=β0+β1·CPU+β2·Numberofsers=200.58 volume is the predicted total space consumption (in Gb).

Year	Actual value	Expected value
2020	1000	1090
2021	1500	1580
2022	2000	2070
2023	2500	2560
2024	3000	3050
2025	3500	3540

Table 13: Actual vs Prediction values
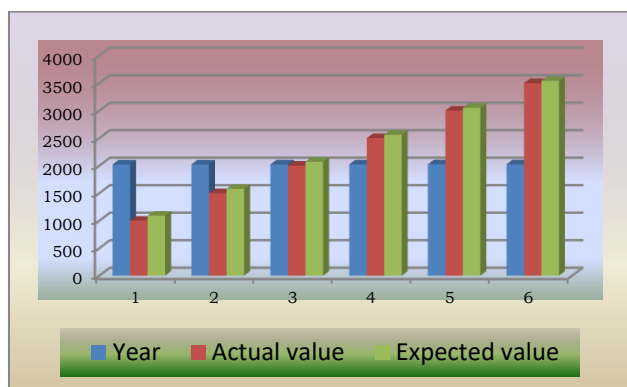
Graph 11: Actual vs Prediction values

Table 13 and Graph 10 shows that yearly volume consumption and Table 12 and Graph 11 shows the actual and prediction value after working on the regression equation.

4. Conclusion

NXRM instance has been facing issues with usage of the space . To avoid the abnormal incidents we have prepared the linear regression analysis so that we can predict the space required for the next 3 months 4 months or whatever the duration you need it. Once we have the prediction analysis we can take the necessary precautions like connecting with dev team to get the list of unwanted artifacts so that we can remove the list to gain the space. If it is not satisfies with the predicted values , then extra space request need to be created so that there will not be any abnormal incidents on NXRM. If we doesn't have this type of analysis , we will not be known until the system goes down. Having this type of analysis saves lot of man hours.

We have number of options like connecting with dev team to get the unwanted artifacts to remove and gaining the space or running the cron tasks from the admin tasks. But all these things workout until there is no addition of business. On day to day, business will get increased or the developers are not ready to give the artifacts proportional to uploading the artifacts to NXRM. As long as there is proportionality between upload artifacts and delete artifacts definitely we need to add extra space to disk , else one day we need to face the system hang issue , where all the developers start connecting admin for the help or quick resolution which will not be possible. To avoid all these scenarios the solution which we have provided will help us a lot. Future work includes working on feature engineering to get any other attributes influencing the final volume and try to implement the logic to get the expected volume to avoid any system issues so that we can have our dev team to work on with out any interruption.

References

[1] Apache Maven cookbook by RaghuRam Barathan.

[2] Docker in Action by Jeff Nickoloff, Stephen Kuenzli, Second Edition , Manning Publications.

[3] Sonatype Nexus Repository Management documentation.

[4] Improving Docker Registry Design based on Production Workload Analysis, Ali Anwar, Mohamed Mohamed, Vasily Tarasov, Michael Littely, Feb 2018.

[5]Jenkins 2: Up and Running- Evolve Your Deployment Pipeline for Next Generation Automation, Brent Laster.

[6] The Maven repository dataset of metrics, changes, and dependencies, Steven Raemaekers, Arie van Deursen, Joost Visser.

[7] Linux Administration a Beginners Guide, Wale Soyinka, 1 July 2017..

[8] THE IMPLEMENTATION OF POSTGRES, Michael Stonebraker, Lawrence A. Rowe and Michael Hirohama EECS Department, University of California, Berkeley

[9] International Research Journal of Engineering and Technology (IRJET) , Volume: 06 Issue: 04 | Apr 2019

[10] An Empirical Study of Maven Archetype, Xinlei Ma Yan Liu School of Software Engineering, Tongji University, Shanghai, China.

[11] MAVEN mission perspectives and approaches to inclusion, S.M. Curry.

[12] A comprehensive study of bloated dependencies in the Maven ecosystem, Nicolas Harrand, Martin Monperrus & Benoit Baudry , 25 March 2021, Volume 26, article number 45, (2021).

[13] An Introduction to Docker and Analysis of its Performance, Babak Bashari Rad, Harrison John Bhatti, Mohammad Ahmadi.

[14] CI/CD Pipeline Using Jenkins Unleashed: Solutions While Setting Up CI/CD Processes, by Pranoday Pramod Dingare.

Understanding and Deploying LDAP Directory Services, by Timothy A. Howes Ph.D.

[15] B. Purnachandra Rao, Dr. N. Nagamalleswara Rao, HDFS Write Operation Using Fully Connected Digraph DataNode Network Topology, International Journal of Applied Engineering Research ISSN 0973- 4562 Volume 12, Number 16 (2017) pp. 6076-6090, © Research India Publications., http://www.ripublication.com

[16] Hadoop Distributed File System Write Operations, Renukadevi Chuppala, Dr. B. Purnachandra Rao, International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING, ISSN:2147-6799.

[17] Method of Least Square, Alexandria Engineering Journal, 2011, Science Direct.

[18] Linear regression analysis study, January 2018Journal of the Practice of Cardiovascular Sciences , kushbu kumari , Suniti Yadav.

[19] Analytical study of two feature extraction methods in comparison with deep learning methods for classification of small metal objects, Somaieh Amraee, Maryam Chinipardaz & Mohammadali Charoosaei , Article number: 13 (2022).