

Paper 5

CASE STUDY ON IMPLEMENTATION OF BIG DATA AT ADOBE

K. M. Kiran Raj^{1, 2} & Krishna Prasad K³¹Research Scholar College of Computer Science and Information Science, Srinivas University, Mangalore, India²Lecturer, Department of Computer Science, Sri Venkataramana Swamy College, Bantwal, India³College of Computer Science and Information Science, Srinivas University, Mangalore, IndiaE-mail: kiranraj224@gmail.com**Abstract**

This paper speaks about implementation of the Big Data in Adobe Inc. Adobe is software multinational. The company was established in San Jose, California, USA in December 1982, by Charles Geschke and John Warnock. Shantanu Narayen is current CEO of the company. Adobe is placed at the Fortune 500 list from 3 years with 339 as its current position. It has the revenue of \$ 9,030 Million, profit of \$ 2,590.8 Million with 21,357 employees. Earlier Adobe had focused on creation of multimedia and creativity software product whereas now it is focusing on digital marketing software. Adobe is implementing new technologies which are able to improve efficiency and customer experience like Cloud Technology is implemented by using Adobe Creative Cloud, AI and Machine learning is implemented by Adobe Sensei. In this paper First part focuses on Implementation of Big Data in Adobe where it also focuses on deployment of Hadoop as a service on VMware and vSphere. Using virtual components like VMware, vSphere and VMware and cloudera can be used to virtualize hadoop which also helps in optimization of both hardware and software tools. Paper will also discuss how Adobe is using Big Data Technique. Adobe plans to offer more tools to help its customers improve their customer commitments in real time. Adobe collects the data and stores them at different places which in turn makes harder to analyze it in short period of time. Capacity, type and importance of Data are growing at rapid rate as time progresses. To store and process data using traditional techniques is creating problems because of it Adobe is implementing Big Data techniques to increase its efficiency in its field. We will also discuss some of the tools which are used / provided for analysis of the Big Data for customer. Next part focuses on financial analysis and PEST analysis of the company. Financial analysis is done by using Annual report of 2015 – 2018 of the company and we will notice even though there are several management issues both revenue and profit of company is growing at rapid rate. This will help in analyzing the company status and progress of it. PEST analysis will help in discovering; understanding several macro factors that will affect company and necessary measures can be taken for it.

Keywords: Big Data, Adobe Inc., Hadoop, VMware, vSphere, PEST analysis,**1. INTRODUCTION**

In recent years, due to improvement in computer sector, social media data is generated at exponential pace. Since the generated data complex and can differ in types it is making difficult to store, analyze and use in an effective manner. Decades ago data used to be very simple and were generated at rather slow rate which was very easy to store, processed and analyzed using the traditional way. Big Data cannot be stored and analyzed using the traditional methods and they are not effective. Since Big Data is Complex in nature and comprises of the characteristics like volume, variety, velocity, veracity and value. Companies want to analyze the data at fast and effective way so that they can attract and provide services in the effective manner which is why most of the companies are shifting towards Big Data environment tools Adobe Inc is one such company. Adobe Systems Incorporated was later changed to Adobe Inc. Adobe is international software company headquartered at San Jose California founded by John Warnok and Charles Geschke in 1982 [1]. Shantanu Narayen is the current CEO [2]. Because of innovative thinking and power the grasp and evaluate market trends it is able to grow at rapid rate with annual revenue of \$ 9,030 million and profit of \$ 2,590 million. Adobe is employing 21,317 people across globe. It is placed in fortune 500 list for the consecutive 3 years and current position is 339 [3]. Adobe focuses on the categories of products like Graphic Design Software, Web Design Programs, Video Editing, Animation, and Visual Effects, Audio Editing Software, eLearning Software, Digital Marketing Management Software, Server software, Formats and Web-hosted Services. Along with Big Data it is also implementing new technologies like Cloud and Artificial Intelligence so that it can give better performance [4].

In this paper we will be mainly focusing on how adobe is implementing Big Data in their company along with virtualization of Hadoop. Section1. INTRODUCTION gives the basic information regarding big data and Adobe. Section2. FINANCIAL ANALYSIS gives the information about financial stability of the company. In this section we will discuss how it was able to maintain healthy growth. Section3. VIRTUALIZATION OF HADOOP tells the need of the virtualization, type of virtualization, Application virtualization and storage mechanism used by adobe. Section4. PLATFORM FOR HADOOP gives information about how vmware vsphere is implemented, its uses and adobe's virtual private cloud. Section5. ARCHITECTURE technical architecture of the how hadoop is implemented on the vmware and cloud implementation is taken into consideration. Architecture will also give explanation about its working and components used. Section6. BENEFITS- different benefits are listed because of implementation of concepts like big data, virtualization and cloud technology. Section7. PEST ANALYSIS does the analysis on Adobe Company by considering political, economical, social and technological aspects.

2. FINANCIAL ANALYSIS

By analyzing the financial status of the company we can determine company's progress. Solutions can be developed by analyzing Financial Status of the company. Here we will be analyzing the financial status of the Adobe from year 2014 – 2018 using its financial report. A financial detail of Adobe Inc. is shown in Table 1 which contains the details like revenue, cost, profit and employees with respect to year. Even though Adobe has competition from IBM, Autodesk, Sparkroom and Unified Adobe is continuously growing both in terms of revenue and profit. In the year 2018 company had annual revenue of U.S. \$ 9.03 Billion in spending cost of U.S. \$ 1.19 Billion where it had profit of U.S. \$ 7.83 Billion. The number of employees working

for Adobe is also growing in a healthy manner. From the below table we can also notice that its expenditure is very less compared with the profit. There are mainly 2 reasons for its exponential growth; first reason is it has very less competitors. Before developing any product Adobe surveys the market and checks which type of product is in demand in the market and also checks for the role players in that field. If it finds there are less competitors then it will invest in that field. Second reason is it adapts to the new technologies like Big Data, Cloud Technology and Artificial Intelligence. Big Data is mainly used for analysis of information so that it provides services effectively. Cloud Technology is implemented by using Adobe Creative Cloud where every product of Adobe is included and conveniently used by the user, Artificial Intelligence and Machine learning is implemented by Adobe Sensei and powers up features of the Adobe products [5]. Adobe has a total addressable market of U.S. 108 B, it is estimated in 2021 the total addressable market in Adobe Document Cloud will reach U.S. \$ 7.5 B, Creative Cloud with a total addressable market of U.S. \$ 29.2 B, Digital Experience with a total addressable market of U.S. \$ 71.2 B with the help of artificial intelligence it is expected to grow more [6].

Table 1: Financial Analysis of Adobe Inc. between 2014 -2018 [7][8][9][10]

| Year | Revenue (in U.S. \$ Billions) | Cost (in U.S. \$ Billions) | Profit (in U.S. \$ Billions) | Employees |
|------|-------------------------------|----------------------------|------------------------------|-----------|
| 2018 | 9.03 | 1.19 | 7.83 | 21,357 |
| 2017 | 7.30 | 1.01 | 6.29 | 17,973 |
| 2016 | 5.85 | 0.81 | 5.03 | 15,706 |
| 2015 | 4.79 | 0.74 | 4.05 | 13,893 |
| 2014 | 4.14 | 0.62 | 3.52 | 12,499 |

3. VIRTUALIZATION OF HADOOP

Data is generated at very fast speed with more complexity. Traditional method of storing and analyzing the data is mostly useful when the data is simple, textual or numeric format where SQL Query can be used. When data is in format of image or graph analyzing becomes complicated. Big Data analysis should be used here. For implementing new technologies like Big Data companies cannot afford to start-over and build all new application every time. Virtualization can be used under these circumstances, where one physical system can be used for many virtual systems. Resources can be utilized between virtual systems as per requirement. Virtualization optimizes computing environment by standardization and automation [11]. Many types of virtualization are present where Adobe is making use of Application virtualization where deployment of individual application is needed for the analysis of Big Data. In Fig. 1: diagram of virtualization of application is shown. Even though an extra layer is added application virtualization removes dependency of individual application from physical computer system by encapsulating applications. This will help in improving individual overall manageability and portability of the applications [12].

Earlier stakeholders did not like the idea of virtualization, because of SAN based virtualization. Where bottleneck is easily reached since computer is on one side and

storage on another side. Adobe uses VMware and cloud to virtualize Hadoop. VMware supports direct attached storage which is very useful since Hadoop is deployed on

private cloud. VMware can also support direct attached, SAN based and shared storage. VMware has certified vsphere big data extension on cloud platform [13].

| | | | |
|------|------|------|------|
| App1 | App2 | App3 | App4 |
|------|------|------|------|

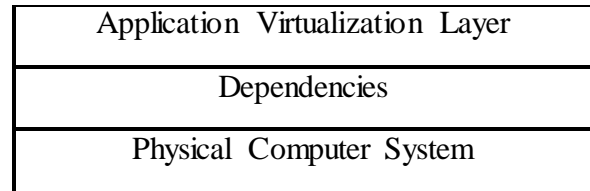


Fig. 1: Application Virtualization

4. PLATFORM FOR HADOOP

Implementation of Big Data at Adobe is mainly in Adobe digital marketing business unit. Other units of Adobe are also implementing Big Data at Adobe. Adobe digital marketing business unit manages Digital assets of its customers email, social media and web presence generation of report. As it mentioned in the earlier section vsphere is used as a platform for hosting adobe's hadoop based application. Vsphere is utilized in the Infrastructure as a Service of cloud, where it can be used for server virtualization. It will help in managing the application in an operating environment with security [14]. Reason for using vsphere is

- In vsphere hadoop cluster is open for end users and community development which reduces time to insight into data
- Vsphere is applied at the IAAS of cloud, which is used for server virtualization. It helps in utilization of existing hardware resources.
- Provides more security, cost effective by developing in-house platform service.

Adobe has created private cloud in infrastructure as a service named as Adobe virtual cloud is also supported by vsphere technologies. Implementing cloud also helped adobe in reducing expenditure on data centre. Adobe cloud applications are helping company to grow at rapid rate since data can be accessed from any remote locations. Big data extensions technique is combined with hadoop platform which is present in the platform as a service across data center. Here hadoop is provided as a service. Since it is provided on cloud and virtualized users can easily get access to service and user does not have to bother about background. Users can customize the services they are going to use; along with basic application users can add application according to their requirement. User is also permitted to use different topologies and distribution method [15].

5. ARCHITECTURE

Technical architecture of implementation of hadoop at Adobe is shown below in fig. 2 which shows outlined architecture. Deployment of hadoop in adobe is done by vmware vsphere with the help of cloud technology. In the figure we can observe 3 layers of cloud are used they are infrastructure as a service (IAAS), platform as a service (PAAS) and software as a service (SAAS). IAAS is a bottom layer in cloud which provides hardware, storage and network as a service. vSphere is implemented in this layer and works as basic platform for complete system and helps in virtualizing infrastructures. Infrastructure provided in IAAS layer is scalable in nature. vSphere provides infrastructure for applications. vSphere is ideal choice for cloud environment when virtualization is required. When hadoop is started by the user daemons of hadoop are also started. There are five daemons Name Node, Secondary Name Node, Data Node, Job Tracker and Task Tracker. Daemons start to run in different virtual machines that are

managed by vSphere. Resource Manager, Node Manager and Application Manager are mapped in different virtual machines with the help of Big Data Extensions in PAAS layer. Using vSphere increases performance, availability and efficiency.

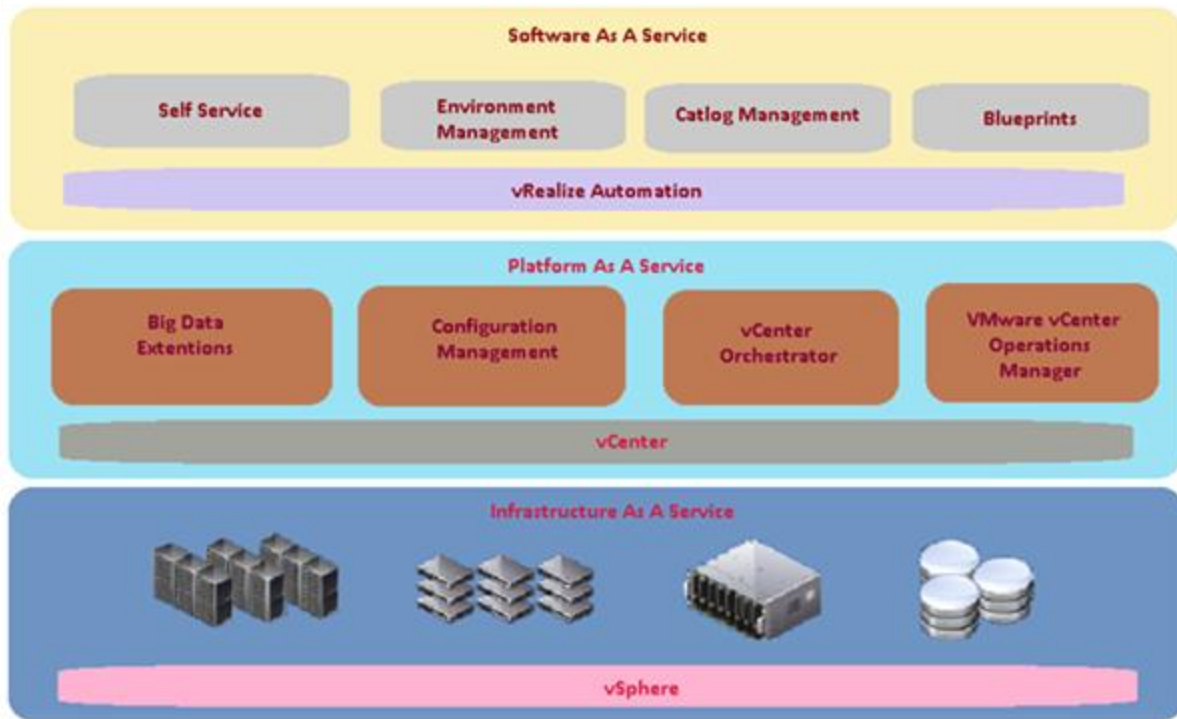


Fig. 2: Outlined Technical Architecture for Adobe Hadoop Provisioning Portal [16]

Second layer is Platform As A Service (PAAS) which is supported by vCenter. PAAS consists of components like Big Data Extensions, Configuration Management, vCenter Orchestrator, VMware vCenter Operations Manager. vCenter manages the ESXi hypervisor which is used to manage virtual machines. Big Data Extensions give different analysis method and automated elasticity to architecture. vCenter Orchestrator is IT processor automation tool which is graphical, powerful and extensible. Workflows are created and customized to perform certain tasks. It has numerous plugin like vCenter, vCloud, vCenter chargeback manager, vCenter Configuration, Manager, Active Directory and 3rd party plugins. vCenter Operations helps in trouble shooting problems related to performance where workload, health and capacity is considered. vCenter Operations is used to determine working of Data Center.

Top layer in cloud is Software As A Service (SAAS) using it software can be used from anywhere. For the existing of SAAS it needs support of both PAAS and IAAS. SAAS can be used for the interaction between system and users whereas PAAS and IAAS cannot be used for communication. User interface and control flow is built using vRealize automation software. vRealize automation supports SAAS. vRealize automation runtime engine is connected with workflows that are customized by vCenter Orchestrator. vRealize invoke workflows which in turn invokes Big Data Extensions API. Big data extensions will implement hadoop cluster. Big data extension with the help of vCenter enables vsphere administrator to supply hadoop clusters

to virtual machines. Workflows, blueprints, frameworks and plugins are customized in vRealize automation, vCenter orchestrator, vCenter operation manager and Big data extensions so that it can handle all requirements of Adobe [16].

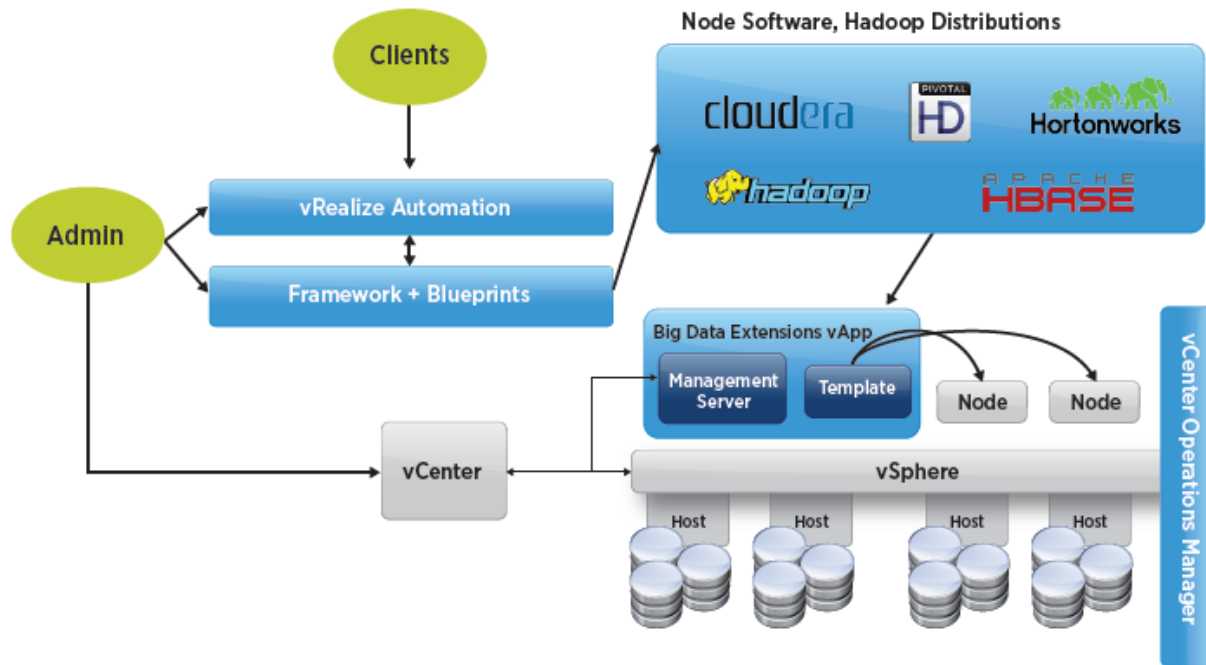


Fig. 3: Detailed Technical Architecture for Adobe Hadoop Provisioning Portal [16]

Fig. 3 gives detailed architecture of how adobe has deployed hadoop in vmware vSphere. As it is mentioned earlier client cannot directly connect with PAAS Layer or IAAS Layer. Clients can only communicate directly with vRealize Automation which is present in SAAS Layer. vRealize Automation and Frameworks, Blueprints is handled by administrator. Clients use framework, blueprints which are customized in vRealize to interact with Big data extensions through vCenter. When blueprint is invoked control is given to hadoop distribution. When big data extension is invoked using management server clones of template is created which acts as virtual machine. Required infrastructure is provided by using vSphere.

6. BENEFITS

Hadoop is implemented in Adobe Inc. using several technologies like Private cloud and virtualization. Virtualization is provided by vmware vsphere and private cloud is created by the use of cloud technologies. Using of virtualization and private cloud is cost effective and improves performance, minimal downtime, fault tolerance, flexibility and scalability. Some of the advantages are listed below

Competitive Performance

Performance of virtual hadoop cluster is comparable or better than the physical cluster. Performance of cluster also depends on the architecture of the system. Performance can be easily increased by increasing number of virtual machines.

- **Rapid Deployment**

Deployments of physical hadoop clusters are complicated because of setup and configuration. In virtual machine hadoop clusters can be easily deployed by cloning other virtual machine nodes. Virtual machines needs less time for configuration. Clusters can be added and removed very easily.

- **Improved Management and Monitoring**

Hypervisor present in virtual machine (vCenter) can be used to manage the resources according to the needs of user. Hadoop is provided as a service to the user, in the platform we can customize the workflows, monitoring system etc.

- **Scalability**

Nodes can be added or removed as per requirement. The concept of load balancing becomes very important here. vSphere has built in load balancer which will be used when any nodes are added or removed.

- **Resource Utilization**

When hadoop is virtualized the resources allocated is shared between different tasks and not dedicated to one particular hadoop cluster as in non-virtualized system. The physical resources are completely utilized by all the virtual machines for their tasks. In non-virtualized system resources were underutilized since resources completely dedicated to the single hadoop cluster.

- **High Availability and Fault Tolerance**

Hadoop clusters have high availability and virtualization provides very good fault tolerance. In virtualized systems high availability solutions monitor hosts and virtual machines to detect failures in resources like hardware and operating system used by virtual machines and hosts. If server outage or network fails the virtual machines from failed host is restarted in new host. When operating system fails virtual machines are automatically restarted. Hadoop cluster is uniformly managed and configured for high availability. Fault tolerance maintains availability of virtual machine without any restrictions. If virtual machine fails it starts secondary virtual machine and if hardware fails it starts failover procedure.

- **Hardware Flexibility**

Hadoop provides flexibility and built in protection by using commodity hardware. By virtualization consumer need not bother about the hardware allocation. Since we are using cloud it can use direct attached storage, storage attached network or network attached storage.

- **Configuration Flexibility**

The one hypervisor can have either one or more virtual machines. The virtual machines can run the computational tasks freely without thinking about allocation of resources. Configuration flexibility creates elastic scalability and user can add or remove VM's as per requirement [17].

7. PEST ANALYSIS

PEST analysis helps to understand the factors that affect any organization. PEST analysis gives the information about political, economical, social and technological factor that affect the company. By using PEST analysis necessary steps can be taken to improve the situation of the company and understand strong and weak points of company.

In Political Analysis it can be noticed that stability and policies implemented by governments directly affect companies. Governments promoting digitization has helped the company to grow steadily. Since it is a multinational company polices, trade or taxation may contradict each other.

Economical analysis of adobe done here and gives positive impact. Annual revenue of Adobe is growing at very good rate. Expenditure cost is very low compared to the profit it is generating. Because of growing social network digital media data is also growing at rapid rate. Adobe is mainly working on digital data and taking complete advantage of growing social network [18] [19].

Social Analysis we can notice the demography of the population varies from place to place like educational level, gender proportionality, cost of living etc, company has to create or popularize different products based on demographic analysis. Main social concern is data security since company has huge amount of data. To provide security company is adopting new technologies [20].

In Technological Analysis we can see that adobe is implementing new trending technologies to keep up with the market. New technologies like cloud, Artificial intelligence, machine learning and big data analysis etc are implemented because of company is growing at rapid rate. New technologies help in processing the data at good speed because of which it attracts huge amount of customers [21].

8. CONCLUSION

Adobe is a rapidly growing company with huge amount of data with U.S. \$ 9.03 Billion annual revenue. Adobe is growing at a faster rate by implementing new technologies like Big Data, Artificial Intelligence and Cloud Technology. To process and analyze big data efficiently hadoop is implemented by virtualizing it on the cloud platform. Hadoop is provided as a service using which big data analysis can be easily done. Virtualization of hadoop provides flexibility and efficiency to the system. Implementation of cloud technology helps in accessing the services from any remote place and helps in implementation of virtual private cloud. The factors affecting the company like political, economical, social and technological aspects are also working in favor of company because of its policy and adaptability to new technologies.

9. REFERENCES

- [1].Norr, H. R. (2019). *Adobe Inc.* 1–8. Retrieved on October 2, 2019. From <https://www.britannica.com/topic/Adobe-Systems-Incorporated>
- [2].SuccessStory. (2019). *Adobe Systems Incorporated SuccessStory.* 1–3. Retrieved on October 2, 2019. From <https://successstory.com/companies/adobe-systems-incorporated>
- [3].Fortune Media IP Limited. (2019). *Adobe.* 1–5. Retrieved on October 3, 2019. From <https://fortune.com/fortune500/2019/adobe/>
- [4].Wikipedia contributors. (2019, November 20). Adobe Inc.. In *Wikipedia, The Free Encyclopedia.* Retrieved 03:14, October 2, 2019. From https://en.wikipedia.org/w/index.php?title=Adobe_Inc.&oldid=927095624
- [5].Nair, S. (2019). *The Big Data Infrastructure Underpinning Powering Autocomplete for Adobe Lightroom.* 1–10. Retrieved on October 5, 2019. From <https://medium.com/adobetech/the-big-data-infrastructure-underpinning-powering-autocomplete-for-adobe-lightroom-41c752bd51c6>

-
- [6]. Adobe Inc. (2019). *Adobe Investors Presentation*. Retrieved on October 7, 2019. From <https://www.adobe.com/content/dam/acom/en/investor-relations/pdfs/ADBE-Investor-Handout-January2019.pdf>
- [7]. Adobe Systems Incorporated. (2015). *ADOBE SYSTEMS INCORPORATED*. 1–115. Retrieved on October 5, 2019. From <https://www.adobe.com/content/dam/acom/en/investor-relations/pdfs/ADBE-10K-FY15-FINAL-CERTIFIED.pdf>
- [8]. Adobe Systems Incorporated. (2016). *ADOBE SYSTEMS INCORPORATED*. 1–112. Retrieved on October 5, 2019. from <https://www.adobe.com/content/dam/acom/en/investor-relations/pdfs/ADBE-10K-FY16-FINAL-CERTIFIED.pdf>
- [9]. Adobe Systems Incorporated. (2017). *ADOBE SYSTEMS INCORPORATED*. 1–107. Retrieved on October 5, 2019. From <https://www.adobe.com/content/dam/acom/en/investor-relations/pdfs/ADBE-10K-FY17-FINAL-CERTIFIED.pdf>
- [10]. Adobe Inc. (2018). *ADOBE INC*. 1–112. Retrieved on October 5, 2019. From <https://www.adobe.com/content/dam/acom/en/investor-relations/pdfs/ADBE-10K-FY18-FINAL-CERTIFIED.pdf>
- [11]. VMWare. (n.d.). *Virtualizing Hadoop ® on VMwre vSphere ®*. Retrieved on October 3, 2019. From <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/vsphere/vmware-hadoop-deployment-guide.pdf>
- [12]. Razbonyali, C., & Güvenoğlu, E. (2016). Traditional Data Storage Methods and the Big Data Concepts. *International Research Journal of Engineering and Technology (IRJET)*, 03(06), 2556–2561.
- [13]. VMWare. (2015). *Virtualized Hadoop Performance with VMware*. 1–20.
- [14]. VMware. (2019). *Secure Applications and Infrastructure with vSphere*. 1–16. Retrieved on October 10, 2019. From <https://www.vmware.com/in/products/vsphere.html>
- [15]. Jlassi, A., Martineau, P. (2016). *Virtualization Technologies for Hadoop-based applications*. The 31st ACM Symposium on Applied Computing (SAC 2016), Apr 2016, Pise, Italy. HAL Id : hal-01325039
- [16]. VMWare. (April, 2015). *Adobe Deploys Hadoop as a Service on VMware vSphere ®*. Retrieved on October 6, 2019. From <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/files/whitepaper/vsphere/vmware-vsphere-adobe-deploys-haas-cs-white-paper.pdf>
- [17]. Webster, C. (February, 2015). *Hadoop Virtualization*. O'Reilly Media Inc. Retrieved on October 6, 2019. From <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmware-hadoop-virtualization.pdf>
- [18]. Cao, D. (2018). *Adobe Systems Inc. Strategic Analysis and Recommendation*. Retrieved on October 7, 2019. From <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1068&context=honorstheses>
- [19]. *SWOT & PESTLE.com* (2019). *Adobe Systems SWOT & PESTLE Analysis - SWOT & PESTLE.com*. Retrieved on October 13, 2019. From <https://www.swotandpestle.com/adobe-systems/>
-

- [20]. Murphy, F. (2019). *Adobe Systems PESTEL Analysis & Environment Analysis*. 1–7. Retrieved on October 13, 2019. from <https://www.essay48.com/term-paper/13131-Adobe-Systems-Pestel-Analysis>
- [21]. Fern Fort University. (2019). *Adobe Systems Incorporated PESTEL & Environment Analysis*. 1–12. Retrieved on October 12, 2019. From <http://fernfortuniversity.com/term-papers/pestel/nyse4/3311-adobe-systems-incorporated.php>