Value-at-Risk Computation as-a-service on Cloud

by

Ratna Choudhury

A thesis submitted to
The Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements
of the degree of

Master of Science

Department of Computer Science
The University of Manitoba
Winnipeg, Manitoba, Canada
April 2022

Thesis advisor

Ruppa K. Thulasiram

Author

Ratna Choudhury

Value-at-Risk Computation as-a-service on Cloud

Abstract

Computing on a Cloud platform has become popular in the recent past, which people use in large scale due to of on-demand availability of compute resources. It enables users to get their job done without direct active involvement of system management. The service is available to users over the internet. Cloud's "pay-as-you-go" model makes it easier for the users to pay for the service they acquire and get the work done. Value-at-Risk (VaR) is a measure that estimates the risk on investments, which investors need to be aware of about their investments. VaR measure is widely used in many fronts from common investors to risk managers, from small business to major corporate. Cloud platforms have been providing various services and computing VaR has not been one of them, which is crucial for common investors. For my research work, I propose to provide VaR computing as a service on cloud. To compute the VaR, investors have to upload sensitive financial data to the cloud, which leads to a concern of data security for this highly valuable data. I use an efficient encryption algorithm to provide data security. Actual VaR computation is done using Monte Carlo simulation technique. By establishing link with MS Azure, my thesis work enables a means of providing VaR computation as a service, anytime, anywhere through cloud and help the investors to take their decision promptly.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to begin by thanking my advisor, my committee members, my family and all the people who have supported me along the way.

This thesis is dedicated to my daughter, Kaira.

Chapter 1

Introduction

Today's real world applications are more complex, dynamic and require quick response time for real time solutions. In this research, I have focused on one such application from the world of finance. Investors desire quick information and correct solutions to make profits. At the same time, risks involved in their investment is also of great importance and need to be analysed. However, problems in finance are too complex to be solved by a deterministic algorithm(s) in reasonable computing time. Therefore, researchers in the field of computational finance are constantly working towards developing efficient models and algorithms in order to help investors in their decision-making process and provide them with as accurate solution as possible.

In this research, my focus is on one such complex and fundamental problem in finance known as the value-at-risk (VaR) problem and use of cloud computing resources to evaluate VaR. In the rest of this chapter, I describe the VaR as well as Cloud computing.

1.1 Value-at-Risk (VaR)

In financial markets, VaR is a risk measure that finds out the possible loss on a specific investment portfolio. It is an approximate of the topmost loss of a portfolio for a specific period of time at a precise confidence level. One of the prime use of VaR analysis is in managing risk.

Risk management plays a major and careful part in nurturing financial health of an individual investor or an institution. Income or earnings from investments tend to be very volatile. Financial planners intrigue to level the revenue curves in the form of capital gains, which assists decrease taxes [1]. This is component of risk mitigation process, performed utilizing VaR analysis.

Risk management enables optimal investment too. Share holders and bond holders of an institution are continuously in disagreement with each other about the investment choices of an organization. Share holders are more inspired to grab excessive risks for good payouts while bond holders are against to take any risk. Efficient risk management enables to mitigate this disagreement for preferable management decisions [1].

Managing risk, therefore, is supremely significant for thorough performance of an institution. Although, the VaR parameters used to calculate risk are difficult to compute. If VaR is estimated inaccurately, financial investments have terrible consequences.

Mainly, VaR measures the possible loss on a portfolio for a specific period of time for a precise confidence interval [2]. As stated in [3], it permits investors to say that they are X% certain that they will not lose more than V dollars in the next N days.

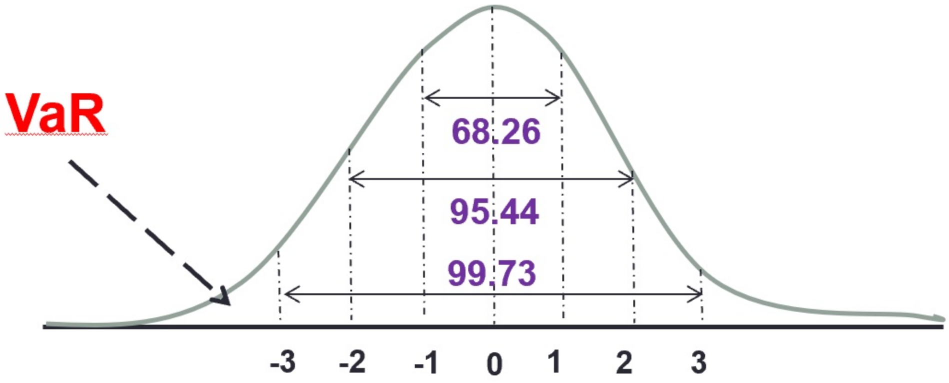


Figure 1.1: VaR for Normal distribution

Straightforwardly, VaR is an evaluation of market disclosure on a distinct investment. Think about a normally distributed pay-offs investment as appeared in Figure 1.1. The VaR with 68.26% confidence level for N business days would be the area under the curve from -1 to 1. Similarly the VaR with 99.73% confidence level for N business days would be area under the curve from -3 to 3. 95% and 99% are the industry standard for confidence/accuracy level of VaR.

1.2 Cloud Computing

With the exponential growth of computational needs of real-world problems, infrastructure and escalating scalability price of IT from the minor scale to sophisticated enterprise sectors, a practical scheme to productively lower the price related with IT infrastructure has become crucial. As IT extended from mainframe computing to Grid and now Cloud computing, this emerged in convenience such as easy operability, less migration cost and downtime, utilization of resources functionally and supervision tools. With the implementation of server based data-centers, countless significant concerns such as data management, data consistency, storage medium, security and accessibility have ultimately become central positions of concern.

There are many platforms available these days for general purpose computing, Cloud platform being one of the most popular one. A cloud platform originally included hardware and operating system (OS) from a server on an internet-based data centre. Currently many software applications are also provided as service along with hardware and OS.

1.3 VaR computing on Cloud

Cloud has become the most popular platform for the users because of its simplicity and accessibility of shared resources. Cloud computing provides services on demand for data storage and computing power. Businesses realize maintenance and investment costs as primary advantage of using Cloud services and incurring costs only for their use of such services. There are many applications of cloud such as file storage, big data analytics, backup, disaster recovery, test and development, etc [4].

People deal with financial securities such as stocks, bonds and derivatives in a financial market at certain transaction cost. The stock exchanges and commodity exchanges are the organizations which ease the deal of financial securities. It can be NYSE, which is a physical location or NASDAQ, which is an electronic system [5].

Companies face different types of financial risk while doing financial transactions [6]. To avoid these financial risks, they use different risk measures to compute the risk associated with a portfolio. In finance, various risk measures are used such as expected shortfall, value-at-risk, probability of default [7].

VaR is a statistical technique utilized to compute the risk associated with investments. With given normal market condition, for a specific period of time, it estimates how much an investment might lose with a given probability. It is usually studied and used by financial industry, investors, risk managers to determine the assets need to be protected from possible losses [8]. It helps investors to be aware of future losses. I tried to search in literature for research works related to VaR study in cloud platform. Surprisingly, this useful technique has not yet been studied utilizing Cloud resources. One reason could be the use of highly sensitive financial data. The studies which is being conducted now using various methodologies are not possible to do from any where or any time without computational resources. Cloud computing makes it easier for investors to use VaR computing service from anytime, anywhere in the world without managing computational resources and get the benefit, which is the motivation behind my study.

To compute VaR in Cloud, investors have to upload highly sensitive financial data, which creates a concern regarding data security. Cloud users face different types of data security issues such as DDoS attacks, shared services for cloud computing, negligence occurred by employee, loss of data and insufficient backup of data, attacks related to phishing, social engineering based attacks, exposure of system weaknesses are the most common [9].

Data Encryption, Two-Factor Authentication (2FA), eliminate shared accounts, well-defined shared responsibility model, standardized cloud assessment questions are some strategies to mitigate such data security issues [10]. Encryption is a possible solution to resolve the data security issue. There are different types of encryption algorithms available such as Triple DES, RSA, Blowfish, Twofish, AES, etc. [11]

Rest of my thesis is organized as follows: I have partitioned the thesis into several chapters which starts with discussions on background and related work in Chapter 2. In Chapter 3, I pointed out the motivation and specific problem statement of my proposed research. I describe the solution strategy in Chapter 4. In Chapter 5, I give the details of the experimental setup, implementation and evaluation. In Chapter 6, I give a summary on the whole work.

Chapter 2

Background and Literature Review

In this chapter, I present some approaches used and reported in the literature for measuring value-at-risk (VaR). Though I could not find any work about implementing VaR on Cloud reported in literature, I present major issues that is faced in Cloud computing in the form of data security.

2.1 Finance

Finance means to manage, create and study of money and investments. It specially handles the money related questions. How a company or individual can obtain money which is called capital in business atmosphere and how to spend or invest the obtained money. It can be classified into few groups such as personal finance, corporate finance and public finance [12].

Finance is all about the overall system of money flow. Financial markets permits the flow of money through investments and other financial instruments across and inside these areas. Thus finance also brings up the study of financial securities market which includes stocks, bonds and derivatives [12].

People deal with financial securities in a financial market at certain transaction cost. The stock exchanges and commodity exchanges are the organizations, which ease the trade of financial securities. It can be NYSE, which is a physical location or NASDAQ, which is an electronic system [5].

Investment management is a vital focal point in finance, either as money management or as asset management [12].

Risk management is the research of how to manage risks and level the profits. The procedure includes computing risk, establishing and applying approaches to control the risk. Financial risk management is the application of defending corporate value to control disclosure to risk by using financial instruments which is called ”hedging” [3]. The credit and market risk are the center of attention and operational risk is involved in banks. Credit risk happens when a debtor is unable to pay on a debt. Market risk arises when market variables like prices and exchange rates changes dramatically and creates loss to investments. Operational risk is logistical if/when there are any collapse in internal processes, people, systems or with external events. Investment managers use different risk management procedures to their portfolios to predict the risk associated with it [12].

Financial risk refers to any kind of risk related to investments and financial transactions. Sometimes only the uncertainty and possible financial loss are considered as financial risk. Financial risk includes different risks such as market risk, model risk, credit risk, liquidity risk, operational risk, etc.[6] The equity risk, interest rate risk, currency risk and commodity risk are market risk factors. Model risk happens when incorrect mathematical/statistical model is used for risk measurement, pricing or portfolio selection. If an asset cannot be exchanged in the market rapidly to prohibit a loss or make a essential profit, then it is called liquidity risk. Asset liquidity and funding liquidity are two types of liquidity risk [6].

Risk management is a pivotal process for making investment choices. The procedure includes to identify and analyze the risk associated with an investment and making decisions to take, reduce, or avoid the risks. Standard risk measures are standard deviation, alpha, beta, Value-at-Risk (VaR), conditional Value-at-Risk (CVaR) and expected shortfall etc. [13; 14]. I am focusing on computing VaR in my thesis.

2.2 Value-at-Risk (VaR)

Value-at-Risk (VaR) is commonly utilized in financial and investment sectors, and it refers to the level or degree of financial risks associated with an organization. VaR helps to speculate a financial loss for a specific period of time at a certain level of confidence, and therefore, it is useful in financial reporting and risk management [2].

VaR is a risk measure to determine the potential loss on a specific portfolio over a given period of time at a certain level of confidence. It helps individuals and institutions alike to claim that they are X(%) confident not to lose more than V dollars in next N days [3]. Here I present an example to demonstrate VaR. Assume a stock portfolio of US $1 million with VaR 1%, which implies that there is a probability of 0.01 of losing US $1 million for that portfolio on a trading day. That means, the portfolio will not lose more than US $1 million with 99% probability.

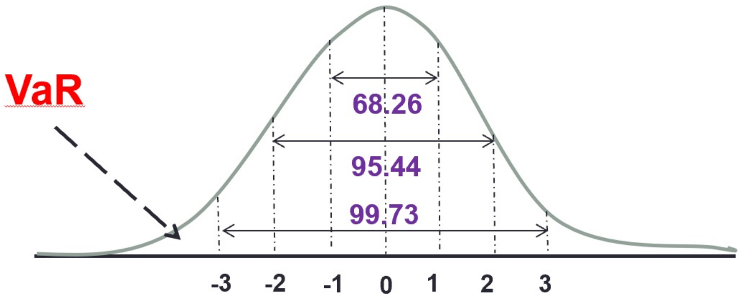


Figure 2.1: Normally distributed VaR.

Prob{△V ≤ $\$1M$} = 0.01, where the portfolio's change in value is △V or Prob{△V ≤ VaR} = 1 – c, where the level of confidence is c. Here, the confidence level is 99% [3].

In simple words, when a particular investment is made, VaR measures its exposure. Figure 2.1 shows an investment with normally distributed payoff. The VaR with 68.26% confidence level is the area under the curve from -1 to 1 for N business days. The VaR with 95.44% confidence level is the area under the curve from -2 to 2 for N business days. Similarly, the VaR with 99.73% confidence level is the area under the curve from -3 to 3 for N business days. The industry standard confidence level for VaR is usually 95 and 99 percent.

The case can be described as equity call option investment. In equity call option, payoff relies on stock (S). If there is any change in the stock after a specific threshold called strike price (K) in the positive direction, it is called the payoff from the equity call. An increase depicts profit and a decrease depicts loss in investment [3]. The pay-off from such option is proportional to the stock price rise beyond the strike price K. That is,

K	S	Payoff
10	5	-5
10	10	0
10	15	5

Table 2.1: Payoffs for different situations

$$Pay - off \propto (S - K) \tag{2.1}$$

For this investment in equity call option, the VaR is how far the price falls compared to the strike price. In other words, VaR seems proportional to the pay-off for this simple investment analysis. In equation 2.1, we suppose the constant of proportion is C.

We show an example in Table 2.1 . In this example, K is fixed to 10 for various values of S. In the case of call option, the payoff decreases if the stock price decreases. A positive payoff indicates profit and a negative payoff indicates loss. Here, we have considered payoffs for simple cases.

There are a number of methodologies used in practice to estimate VaR, including i) Variance-Covariance method, ii) Histogram simulation, iii) Monte Carlo simulation, and iv) Genetic algorithm (GA).

2.2.1 Variance-Covariance method

For VaR evaluation, variance-covariance method was proposed by [15]. It is also known as delta-normal or analytic method. For a large portfolio and normally distributed input data, this method could be fast and simple to use. However, the evaluation of probability distribution of the asset is a major problem of this method. The method works only if the input data is normally distributed, which is not realistic [15].

Let us assume a normally distributed portfolio where mean is CA $100 and standard deviation is CA $10. Considering 95% confidence level, we can say that the portfolio value will not go above CA $120, which is two times standard deviations above the mean and will not go below CA $80, which is two times standard deviations under the mean for the next year. Here, the standard deviation of 1.96 goes on both sides of the mean for 95% confidence level. For 99% confidence level, the standard deviation of 2.33 goes on both sides of the mean [2]. In simple words, we can say that within 1.96 standard deviations of the mean is the 95% area of the normal distribution.

If we have a large portfolio whose volatility estimates are good, then we can use this method. We need normally distributed input data and there cannot be any fat-tailed in data distribution. This method is quick and uncomplicated to use. Though it is challenging to implement stress test with this method.

2.2.2 Histogram Simulation

Implementing the simulation approach based on prior data to measure VaR was proposed by [16]. They create a time series using the historical data of the return on a portfolio. The data distribution is unlimited. Generally, this technique is used when the data distribution is unspecified. This is a conventional technique because it considers all the exceptions for VaR evaluation. Since there is no limitation on the data distribution, we can also use the fat-tailed distribution in this method. The method was applied to trade financial institution portfolios. The real historical data is required for VaR evaluation that may not be accessible every time and this is a major restriction of this technique [17].

2.2.3 Monte Carlo Simulation

Monte Carlo simulation is based on the concept of Brownian motion that describes the random motion of molecules in fluid. MC simulation can estimate VaR by computing the change in stock price by utilizing a numerous number of independent simulations, each of which consists of cycles. The accuracy of MC analysis can be upgraded by incrementing the cycle number [18].

MC simulation is a computational algorithm that is based on the principle of utilizing randomness to solve a problem. This technique explains the influence of risk and uncertainty in prediction and forecasting models. It works by appointing numerous values to an unknown variable, get different outcomes and average the outcomes for an computation [19]. It can be used to solve problems of various fields such as finance, engineering, supply chain and science. That is why, it is also called multiple probability simulation [19]. It uses the concept of Brownian motion that describes the random motion of molecules in a fluid, such as liquid or gas [20; 21]. Suppose $\triangle$S depicts the change in stock price. After evaluating by MC simulation, we have different values for $\triangle$S and these values are arranged in ascending order. Depending on the confidence level, a specific value of $\triangle$S is chosen. MC simulation can estimate VaR by calculating the change in stock price utilizing a numerous number of independent simulations, each of which consists of many cycles. The accuracy of MC analysis can be upgraded by incrementing the cycle numbers [18]. MC simulation is able to approximately estimate the possible loss on a financial portfolio.

When dealing with the process of making an estimation from uncertain numbers, MC simulation can be a better solution, which uses multiple values without replacing uncertain variable by a single average number. As business and finance field are tremendously troubled by random variables, MC simulation could be a proper solution to deal with the problems of uncertainty [19].

MC simulation works on stock prices for a specific time period and the simulation results are used to estimate VaR for that specific time duration at a certain confidence level [16]. For stock price simulation at t+△t time, the equation will be:

$$S(t + \triangle\ t) = S(t) + \triangle\ S \tag{2.2}$$

In this equation, the current stock price is S(t) and the change in stock price is △S. Utilizing geometric Brownian motion, we can use the following equation to calculate △S:

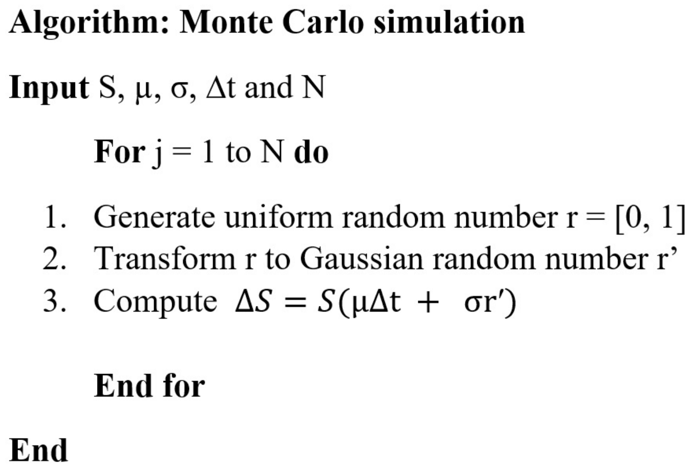
$$\triangle S = \mu S \triangle t + \sigma S r' \sqrt{\triangle t} \tag{2.3}$$

The above equation entails that △S can be assessed for time t for dollar S stock price where μ is mean return, σ is standard deviation and r' is a random number in the middle of [0, 1] [16].

To run these simulations, we need some inputs which are current stock price (S), expected return (μ), standard deviation of returns (σ), time (△t) and cycle numbers

(N) [16].

The following is the algorithm for MC simulation:



```
Algorithm: Monte Carlo simulation
Input S, μ, σ, Δt and N
    For j = 1 to N do
  1. Generate uniform random number r = [0, 1]
  2. Transform r to Gaussian random number r'
  3. Compute $\Delta S = S(\mu\Delta t + \sigma r')$
    End for
End
```

2.2.4 Genetic Algorithm (GA)

GA is a biologically inspired algorithm and founded on the idea of preferences made by nature [22]. It was developed utilizing the theory of evolution formulated by Charles Darwin. It identifies the fittest offspring that survives from a large number of possible options through the process of natural selection. To produce next generation offspring, the fittest chromosomes are selected to reproduce in the process of natural selection. The process of natural selection is controlled by two crucial functions crossover and mutation [22; 23]. The sensitivity and adaptability of the next generation, in the current environmental state, depending on the degree and extent of crossover and mutation functions [24].

There are few phases in a genetic algorithm. The process begins with the initialization phase where random population is generated. The fitness of each candidate is

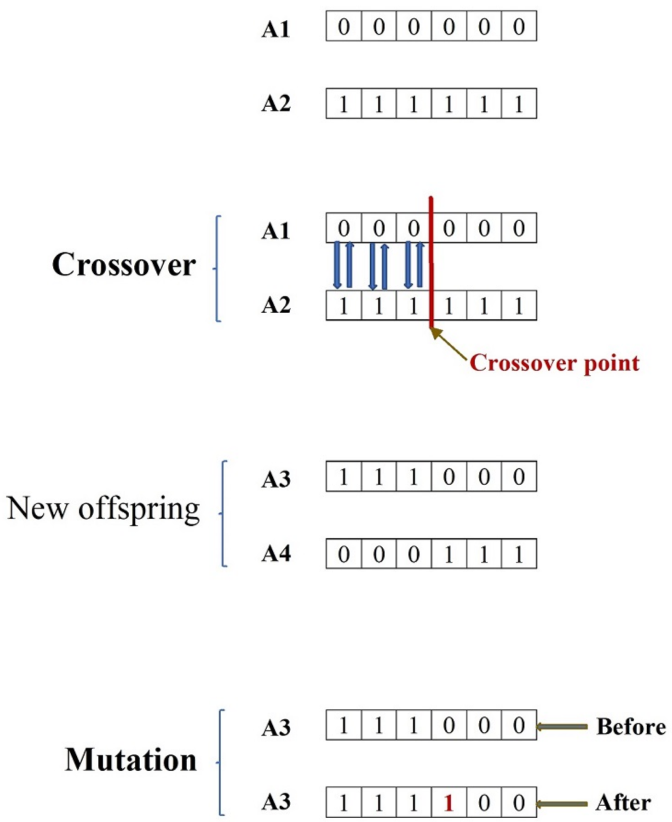
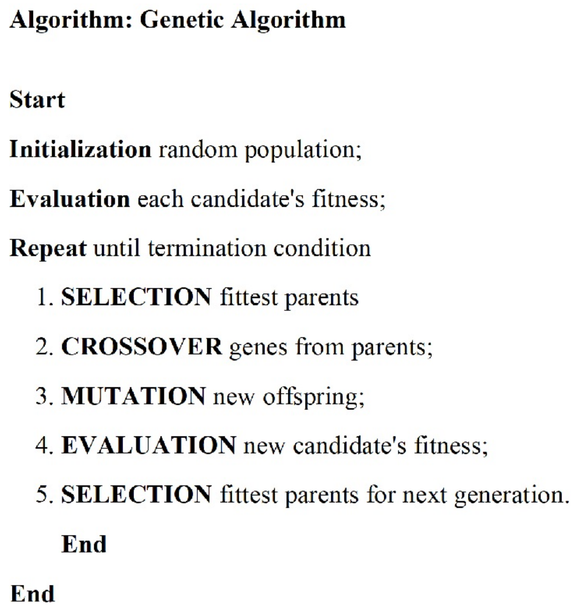


Figure 2.2: Crossover and Mutation in GA

evaluated using fitness function. It determines the fitness of the candidates and gives a fitness score to each one. Until the termination condition, the process of producing future generations is repeated. In this process, there are some more phases. First, the selection phase where fittest parents are selected to create the next generation from their genes [25].

Here is an example to illustrate crossover and mutation in Figure 2.2. Suppose a pair of parents are A1 and A2 and the crossover point is 3. Figure 2.2 shows the crossover, new offspring after crossover and mutation on a newly formed offspring.

The following is the algorithm for GA:



```
Algorithm: Genetic Algorithm

Start
Initialization random population;
Evaluation each candidate's fitness;
Repeat until termination condition
   1. SELECTION fittest parents
   2. CROSSOVER genes from parents;
   3. MUTATION new offspring;
   4. EVALUATION new candidate's fitness;
   5. SELECTION fittest parents for next generation.
      End
End
```

After the selection phase, the parents are recombined in the crossover phase. In crossover, offspring is generated by interchanging the genes from the parent chromosomes. The crossover varies the chromosomes from one to the next generation. There are different ways of using crossover function. For crossover function, we will consider commonly used single-point. A point to crossover is chosen arbitrarily for the selected parent to exchange genes between the parents [25].

Mutation is applied to some of the newly formed offspring from the crossover phase with a low random probability. It could change the chromosome at one or multiple points to improve the adaptability of offspring at a certain environmental state. It reflects the changes in the current environmental state [25]. Through mutation, the diversity is preserved within the population and also restrains early convergence [25]. The termination condition of this algorithm is when the population does not produce significantly different offspring from previous generation [25].

2.3 Cloud Computing

Cloud computing delivers computing services over the internet with a pay-as-you-go pricing model. We pay only for the cloud services we use. It helps us to minimize our operating costs, operate our infrastructure accurately and scale as per business requirement. In other words, instead of maintaining a whole datacenter, we could rent compute power and storage from cloud provider's datacenter through internet connectivity when needed. The cloud provider maintains the underlying infrastructure. We can use the resources like our own datacenter. After finishing the work, we are charged only for what we use.

Cloud environment provides a list of advantages over physical environment such as cost-effectivity, reliability, scalability, elasticity, agility, geo-distribution, security, disaster recovery, etc. There are three different cloud service models which are Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). These models help to speculate the responsibilities shared between a cloud provider and cloud tenant. Cloud has three deployment models which are public cloud, private cloud and hybrid cloud and each of them has different aspect to deploy [26].

2.4 Data Security

Preserving digital data from being stolen or damaged is data security. Different technologies are used for data security such as disk encryption, backups, data masking, data erasure, etc. [27].

Disk encryption is an encryption technique which encrypts data on a hard disk drive. It has two forms: one is software and another is hardware. It is also called On-The-Fly Encryption (OTFE) or transparent encryption [27].

If we lost data for some reason, backups is used to secure that the lost data can be retrieved from other sources. In industries, data backup is crucial and the procedure is suggested for all important files [27].

The procedure of concealing (masking) particular data in a structured database table to secure data and important details is preserved from unwanted people is data masking. Data masking may be required for users, developers, outsourcing vendors etc [27].

Data erasure is a software based overwriting technique. It entirely erases whole electronic data saved on a hard drive to secure the security of important data from an former or regenerated asset [27].

Cryptography word comes from ancient Greek language. It is the implementation and research of procedures for safe transmission with the existence of opposed actions. Basically, it establishes and investigates protocols, which stops general people to read particular communication. Current cryptography is all about data confidentiality, data integrity, authentication and non-repudiation, which are different features of information security. Current cryptography lies in the convergence of different disciplines such as mathematics, computer science, electrical engineering, communication science and physics. E-commerce, chip-based payment card, digital currency, computer password and military communication are some implementations of cryptography [28].

Cryptography is nearly identical with encryption where data is transformed and becomes clear message to meaningless rubbish. The sender hands out the the decryption process of the encrypted note with prospective receiver to prevent entrance of unwanted people [28].

In cryptography, the procedure of ciphering or encrypting data is called encryption. Plaintext, the actual portrayal of the data is transformed into ciphertext, a different form by the procedure. Theoretically, people with permission can decode the ciphertext and retrieve the actual data. Encryption does not block any trespassing but refuses the accessible content for prospective receivers [29].

An encryption algorithm produces a pseudo-random key to be utilized by an encryption technique. An encrypted data can be decrypted without using the key though significant computational materials and expertise are needed for a well-planned encryption technique. An approved receiver can decode the encrypted data by simply using the key supplied by the initiator to the receiver. Different concepts of encryption have been utilized to support cryptography. The idea of public key and symmetric key is exercised by present day encryption techniques. Present day encryption procedures guarantee safety [29].

Identical keys are used to encrypt and decrypt in symmetric key (private key) procedure. To accomplish stable transmission, transmitting groups should have the identical key. In public key procedure, the encryption key is issued publicly so that it can be used by anyone to encrypt data. Though only the authorized recipient has the decryption key to decode the data. A number of authors have proposed different schemes for public key encryption [29].

Historically, encryption has been utilized by armies and governments to ease classified transmission. Now a days, it is widely used by individuals and companies to secure their data from hacking. It can be utilized to secure ”data at rest” or ”data in transit”. Data at rest is when data is kept on computers and removable storage devices. For last few years, many incidents happen where sensitive data has been stolen from computers and backup storage devices. We can use encryption data at rest to secure this kind of stored data. If data is moved through internet, wireless microphones, mobile phones, Bluetooth devices, then encryption data in transit can secure the transmitted data. Data could be stolen during the transmission period. If the data is encrypted while moving networks, then we can secure the data from being stolen by hackers [29].

Nowadays, data encryption has become a necessary step to do to protect our data from being stolen or illegal use. It ensures privacy of the data that only authorized person can read data in transit or data at rest. It ensures security by restraining data breach. If the data at rest of any removable storage device is encrypted, then the data is still secured from being stolen or lost. Likewise, if the data in transit is encrypted, then there is no fear of losing sensitive data other while transmitting between groups. It also secures data integrity by controlling harmful actions like on path attacks. Encryption secures that communicated data is not damaged while traveling through the internet. Authentication is preserved through encryption. Through public key encryption, the users get to know if a website is real or fake. Encryption is used to preserve regulations. There are regulatory and compliance standards where encryption is essential to secure user data such as HIPAA, PCI-DSS and the GDPR

[30].

An encryption algorithm is the technique to change the plaintext into an ciphertext by using the encryption key. There are some frequently used encryption algorithms such as AES, 3-DES, SNOW for symmetric encryption algorithm and RSA, Elliptic curve cryptography for asymmetric algorithm [30].

Encryption is one of the most used techniques to solve the data security issue. In literature, we observe two types data analysis with the encryption technique. Some authors work on the encrypted data such as homomorphic encryption allows computations on encrypted data [31]. Some authors use the decrypted data for further analysis [32]. Recently, Python libraries for cryptography is being used a lot for encrpyting and decrypting data [33]. Modern day cloud service providers are also offering different security services to ensure the safety of data [34].

2.5 Related Work

There are lots of research works had been done on Value-at-Risk though VaR computation has not been studied using Cloud yet. Here, we discuss some research works which has been done to compute Value-at-Risk. Some authors worked to improve a methodology to compute VaR, some authors compared between different methodologies used to compute VaR while some authors used the VaR as a tool to do some important investigation. We also discuss some research works done for Software as a Service on cloud platform.

Hendricks evaluated various Value-at-Risk models using historical data to find out their performance in reality [35]. He used 1000 random portfolios between 1983-94 and applied VaR models to find out how close was the result with the actual numbers. He used three categories of VaR models which are equally weighted moving average, exponentially weighted moving average and historical simulation. These three categories included twelve different approaches and he also measured both 95% and 99% confidence levels for the portfolios. The experimental results showed that all the approaches provides risk measures which are not very different in size. Though historical simulation approaches performed better than variance-covariance approaches. He could not recommend one approach which is best but suggested to combine approaches to build a better VaR model [35].

Butler and Schachter proposed a different way to use historical simulation to measure Value-at-Risk by utilizing kernel quantile estimator of the probability density function of the returns for the portfolios [16]. They also utilized numerical integration on the probability density function to get statistic orders where the mean is the VaR value and the standard deviation is the confidence level. They experimented this technique by implementing it to trading portfolios. The experimental results showed that Epanechnikov kernel produced the most conservative results with moderate standard errors. The Gaussian kernel gave less conservative results but the best confidence levels. They also measured standard error by using Monte Carlo simulation for the empirical quantile estimate [16].

Glasserman et al. tried to overcome the problem of the Monte Carlo simulation to be too slow [36]. They integrated the finest attributes of two methods which were the delta-gamma approximation's speed and Monte Carlo simulation's accuracy to compute VaR. They integrated importance sampling and stratified sampling of delta-gamma approximation to decrease the scenario number to get specific precision for a simulation. They extended the method by using the term ”vega” to seize the volatility change [36].

Manganelli and Engle did a survey to evaluate the implementation of most wanted univariate VaR techniques specifically their assumptions and logical flaws. They contributed two theories [37]. First, for the Conditional Autoregressive VaR or CAViaR model, they initiated the extreme value theory. Second, they used a easy regression method to compute the expected shortfall. The implementations of the techniques were evaluated by Monte Carlo simulation. The experimental results proved that CAViaR models with heavy tailed DGP performed best [37].

Antonelli and Iovino proposed a technique to improve the performance of Monte Carlo simulation to compute VaR [38]. They used component analysis to decrease the source number of risk of portfolios. To obtain a specific precision, less price scenarios were approximated by using large deviations methods. The results proved that the presented method performed finer than the conventional Monte Carlo method [38].

Liu proposed VaR forecast using Artificial Neural Networks (ANNs) instead of controversial model selection [39]. The results of the simulation and real data proved that the performance of ANNs combinations was far more better than the individual VaR models [39].

Aniunas et al. proposed complex variance - covariance model to analyze possible risk of currency market through value-at-risk (VaR) [40]. This model helps to take investment decisions on the basis of profitability and confidence level. Considering acceptable risk level, it generates revenue, which could be found from the investments in currency market. They tested the model with real currency market data and found the model as reliable for practical use [40].

Weerd tested if the variance-covariance method produced correct VaR for assessing bank's high-risk assets [41]. To evaluate the volatility, he used the Moving Average (MA), the Exponentially Weighted Moving Average (EWMA), the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and the Threshold-GARCH (TGARCH). The author utilized the data of the AEX, CAC40, DAX and FTSE100 index from January 1985 to February 2011 and each model processed VaR prediction for one day ahead. He used the Kupiec test, the Christoffersen test and the Lopez test for back-testing the VaR results. The experimental results showed that only EWMA model provided correct VaR values for all indices. He concluded that the variance-covariance method produced correct VaR values if the EWMA method is utlized to predict the volatility [41].

Hong et al. reviewed the then improvement of MC simulation to analyze VaR and CVaR, gave the structure to recognize them and talked about the implementation of them in managing risk [42]. They mentioned that more study were required for the input uncertainty of managing risk [42].

Sharma et al. worked to find out an efficient method to compute VaR. They proposed and used Genetic Algorithm (GA) to compute VaR [24]. GA is a biologically inspired algorithm which indicates the fittest offspring which survives through the selection done by nature such as crossover and mutation. The characteristics of this algorithm is appropriate to compute VaR. They compared the findings from GA with MC simulation and the results showed that GA provides better results of VaR than

MC simulation [24].

Another work on VaR was done by Suwarno et al., where they tried to find out risk involved in Mean-Variance model [43]. It is a portfolio optimization model developed by Harry Markowitz in 1952 [44]. Analyzing various possible portfolios of specific securities, it helps to find out the most efficient portfolio. It reveals how investors can reduce their risk. Expected returns and the standard deviation of different investments are the basis of the model. In other words, the factors are mean and variance respectively, that's why it is called mean-variance model [44]. The research work considers investor risk preference with VaR model. They also compared between Single Index Model and Mean-Variance Model. They concluded that variance and data dispersion increase VaR value [43].

To determine VaR, Zhang et al. proposed GELM model which is a stochastic mapping model [45]. It incorporates Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model with the Extreme Learning Machine (ELM). According to the experimental results, the model forecasts better volatility and computes VaR with expertise and perfection compared to traditional models like GARCH, ELM and Support Vector Machine (SVM). Thus, GELM model could be used as an instrument to control threat and analyzing pressure [45].

Sumaji researched to get a resolution for Indonesia's economic issues [46]. Computing risk efficiently for stock market could decrease the problems of Indonesia's economic improvement. She inspected 9 companies using variance-covariance method of value at risk. She also used back testing technique to validate the variance-covariance technique. The tests showed that the variance-covariance technique was the proper technique to compute value at risk [46].

Goorbergh and Vlaar applied different Value-at-Risk procedures including historical simulation, Variance-covariance and tail index estimation to AEX, Dutch stock market index and Dow Jones Industrial Average [47]. They got some important findings from their research. First, volatility change of stock returns is a major feature for any VaR technique. Second, when dealing with higher confidence levels, the fat tails distribution can follow the t-distribution. Third, the capital requirement was providing appropriate incentives to the banks which was found out by the penalty scheme [47].

Linsmeier and Pearson analysed three methods which are historical simulation, the delta-normal method and MC simulation for computing Value-at-Risk [48]. They scrutinize these techniques over their pros and cons when computing VaR. They considered different situations to find out the performance of these techniques. They described about stress testing shortly. They also talked about two substitute risk measures to VaR which are sensitivity analysis and cash flow at risk (CFAR) [48].

Sarma et al. worked to select a proper VaR model by utilizing two case studies of the S&P 500 index and India's NSE-50 index [49]. The process included two step to select the model. The statistical precision of the models were analysed in the first step. The second step used loss functions to strain the remaining models from the first step. They also found out the power and delicacy of computing and analyzing VaR [49].

Steelyana did a comparative analysis to find out a better methodologies between historical simulation snd variance-covariance method for VaR computing using portfolio simulation [50]. VaR result is smaller by Variance-covariance method than VaR result by historical simulation [50].

Oppong et al. analyzed VaR computing techniques, historical simulation and Monte Carlo simulation on the Ghana Stock Exchange stocks to find out which method gave preferable VaR value [51]. They used data of ten stocks to do the comparison. Their experimental results showed that Monte Carlo simulation performed well compared to the historical simulation [51].

Berkowitz and O'brien investigated the revenues of the trading accounts of big commercial banks and VaR prediction done by the bank [52]. They assessed the execution of risk models by inspecting the precision of the VaR predictions. They claimed to be the first to give thorough analysis of the execution of VaR techniques used by big trading agencies. Their experimental results showed that the VaR prediction for some big banks performed really well while the VaR prediction of some banks are so conventional that they are not practical. The VaR prediction using the bank methods could not exceed the predictions done using APMA + GARCH technique with P&L of banks [52].

Sharma et al. used both value-at-risk and genetic algorithm to propose cloud resources pricing model called Clabacus to serve both clients and providers [53]. To develop Clabacus, the authors utilized financial option theory. The proposed approaches were based on fuzzy logic and GA to compute VaR to balance the valuation of the resource with the inherent risk for the cloud provider [53].

Gai et al. mentioned heterogeneous clouds were considered to be the solution of multimedia big data [54]. Though they were facing issues because of limited performance and price, which introduced the data allocation problem. The authors proposed Cost-Aware Heterogeneous Cloud Memory (CAHCM) model, which provided high performance cloud-based heterogeneous memory service and aimed to attain less data processing time to solve the issue. Both Greedy and Genetic algorithms were used in the CAHCM model and Dynamic Data Allocation Advance (2DA) algorithm and produced an efficient solution within a short period of time [54].

Philip et al. did research work to make Signature Recognition System (SRS) as Software as a Service (SaaS) to use for systems where authentication is required to the clients [55]. They proposed biometric-based approach for authentication purpose and used Microsoft Azure as cloud platform for flexibility, scalability and reduced overhead cost of the biometric system requirements. The Webber Local Description (WLD) process and it's algorithm was the focal point of this research, which uniquely identified each user and encapsulated security into different applications. The proposed methodology advanced the field by providing a low-cost, portable, scalable and flexible biometric-based SRS [55].

Kotas et al. evaluated and compared the performances of Amazon AWS and Microsoft Azure on two high performance computing (HPC) benchmarks, the HPC Challenge (HPCC) and the High-Performance Conjugate Gradient (HPCG) [56]. They tested their computational speed, memory bandwidth and network bandwidth. They focused on the cluster level application performance for different estimation and communication patterns. The authors purpose was to find out which cloud platform provided better performance in terms of speed, faster network, larger RAM and cheaper in cost. The experimental results showed that Azure H16r's faster network and larger

RAM was a cheaper solution for communication intensive applications [56].

In literature, no significant research work has been done on VaR computation using other bio or nature-inspired computing techniques such as Genetic Bee Colony Algorithm, Fish Swarm Algorithm (FSA), or the Artificial Algae Algorithm (AAA). However, a few research works use Particle Swarm Optimization (PSO) for VaR computation. As any research work has not been done on VaR computation utilizing Cloud resources, I could not discuss relevant literature.

Chapter 3

Problem Statement

From the discussions of Background and Related work chapter, we have seen many research works have been done on VaR. Also, Cloud platforms that started from a simple storage service has been providing many different hardware (IaaS/PaaS), software (SaaS) and application services as well as anything as a service (XaaS). From accounting to scientific, there are many applications provided by cloud these days besides military, health, and financial problems due to strong requirement for data security in these areas. One such financial computing problem as described earlier is value-at-risk (VaR) computation, the focus of my research study. VaR is an effective risk metric, which helps investors to be aware of future losses. The studies in practice using different methodologies to compute VaR need access to physical computational resources. Cloud can help investors by providing VaR computing service anytime and from anywhere in the world without managing computational resources.

For my research work, I investigate how to provide VaR computation as a service on Cloud. To compute VaR, investors have to upload highly valuable and at times proprietary financial data, which rises the concern of data security. Hence, I consider some of the existing cryptographic methods that could add security while keeping usefulness of the financial data.

Chapter 4

Solution Methodology

I present in this chapter one of the many approaches that could be used for computing VaR as well as a platform where from service could be provided. Moreover, computing VaR on Cloud requires securing data. Since the focus of my study is not in designing/developing algorithms for security of the data, I present an important approach for data security.

4.1 VaR Computing as a Service

For providing VaR as a service, the computational approach, a sample platform of study are presented in this chapter are identified first.

4.1.1 VaR Computing Methodology

There are different methodologies as described in Chapter 2 that are used to estimate VaR, such as Variance-covariance method [15], Histogram simulation [16], Monte Carlo (MC) simulation [16]. Each method has their strengths and weaknesses. My goal in this research work is to provide VaR as a service in Cloud for which I present a methodology that is being widely used in practice, Monte Carlo simulation. On real cloud implementation for VaR computation other techniques could be deployed very well.

MC Simuation

The algorithm of MC simulation produces random numbers, which are utilized to estimate a non-closed form equation. We pursue to some test and fault in selecting the random numbers what the equation allows to estimate the result. Selecting random numbers from huge numbers (a few hundred to a few million) will provide a good sign about the output formula.

Here we discuss all the aspects of using MC simulation for computing VaR. Recalling the MC algorithm for VaR where Step 3 has two special expressions. The first expression refers to the movement in the stock price and the second expression refers the shock factor. To assess the transformation in the current stock market, both of these factors are needed. Huge number of simulations are necessary for computing VaR using MC simulation [2]. 10,000 simulations are implemented to get correct outcome. The outcomes are organized in descending order. The most widely used matrices are 95 percent and 99 percent accuracy.

Utilizing normally distributed random numbers is another vital element of MC simulation. Uniform random numbers are generated from a lot of pseudo-random generators. Uniform random number are converted to normally distributed random numbers.

In the computation of VaR using MC simulation, we generate random numbers to evaluate the return (or price) of the asset without utilizing the historical data for the returns (or price) of the asset and supposing that this return can reiterate in the following interval.

In this section, we describe how the algorithm will be applied to compute VaR for stock.

Step 1: Determining the time t and split it into little time periods equally ($dt = t/n$).

As an example, we will compute VaR for one month which includes twenty two trading days. Here, we get $n = 22$ days and δt = 1 day. If we want to compute VaR for 1 day, we can split each day by the total number of minutes or seconds involved in a day.

We have to secure the value of δt is sufficient to estimate the pricing available in the financial markets. We estimate a continuous occurrence by a enormous number of discrete intervals and this procedure is called discretization.

Step 2: Select a random number from the generator and upgrade the asset price at the end of first step up.

The generator generating random returns go along with a particular conceptual distribution. This can be considered as a weak point for MC simulation. Normally we utilize the normal distribution to simulate random numbers.

To simulate the path of a return of a standard stock price model from the ith day can be written as:

$$\mathrm{R}_i = (\mathrm{S}_{i+1} - \mathrm{S}_i)/\mathrm{S}_i = \mu\triangle t + \sigma\epsilon\sqrt{\triangle t} \tag{4.1}$$

Here R_i is the ith day return of the stock, S_i is ith day stock price, S_{i+1} is the (i+1)th day stock price, μ is sample mean of the stock price, σ is standard deviation of the stock price, ϵ is a random number generated from normal distribution.

At the end of this step, we could select a random number and get the value of S_{i+1}, since other parameters can be estimated as well.

Step 3: Repeat Step 2 until the termination of analysis T with N time intervals.

In the next step (δt = 2), we select another random number and find out S_{i+2} from S_{i+1} utilizing the above equation. We will continue this approach till we hit T and find out S_{i+T}. S_{i+22} illustrates the approximate stock price for one month of the sample stock.

Step 4: Repeating Steps 2 and 3 M times to create M different paths for the stock over T.

For this stock (from i to i+22), we have created one path. By using MC simulation, we have created M different paths considering other feasible paths for the stock price for a duration of one month from its current value (S_i) to an approximate terminal price $i + T/N$. In fact, there is no distinctive approach for the stock price to go from S_i to $\mathrm{S}_{i+T/N}$. Besides, $\mathrm{S}_{i+T/N}$ is only one feasible terminal price for the stock out of infinite number of prices. In fact, there is an infinite paths for a stock price from S_i to $\mathrm{S}_{i+T/N}$, which is described on a set of positive numbers.

10,000 simulations deliver logical estimator of the terminal price for assets. In this research work, we ran 10,000 simulations.

Step 5: Sorting the M terminal stock prices from the smallest to the largest, studying the simulated value that communicates with the appropriate (1-α)% confidence level (commonly 95% or 99%) and conclude the proper VaR, which is the distinction between S_i and the lowest terminal stock price.

Suppose we would like to have the VaR with 99% confidence level. To achieve it, first we will require to sort the M terminal stock prices from the lowest to the highest.

Then we study the 1% lowest percentile in the sequence. This approximated terminal price, S_i+T1% represents that there is 1% chance the current stock price S_i could drop to S_i+T1% or less for the inspected interval with normal market conditions.

If S_i+T1% is smaller than S_i, then S_i - S_i+T1% will equivalent to a loss. This loss is the VaR with 99% confidence level.

4.1.2 Cloud Computing Platform

To implement VaR as a service on Cloud, we have to use a cloud platform. There are a number of Cloud service providers such as Amazon Web Services (AWS), Microsoft Azure, Google Cloud, Alibaba Cloud, IBM Cloud, Oracle, Salesforce, SAP, Rackspace Cloud, VMWare etc. Most popular two service providers are AWS and MS Azure. By comparing both of them according to their services, compliances, global availabilty, we have found that Microsoft Azure is the fastest-growing Cloud with the most profitable Cloud services. It has the most advanced and maximum intelligent products and services. AWS is way more expensive than Azure for Windows server and SQL server. Hence, we have decided to use Microsoft Azure cloud resource, which gives free access to popular cloud services and developer tools to students for

12 months.

Microsoft Azure

The cloud platform has more than 200 products and cloud services to deal the problems and provide solutions. One can choose the tools and framework to build, run and manage applications. It has 90+ compliance offerings which is the largest portfolio in the cloud industry. More than 95 percent of the Fortune 500 companies are doing their business on Azure. They have 60+ global regions in 140 countries. They provide better experiences for their 15.4 million customers worldwide [57].

Security and privacy are foundation for Azure. They invest 1 billion (US dollars) every year for security purpose to protect customer's data from cyberthreats. They are committed for providing highest level of trust and transparency [57].

Azure is not only for windows apps and services, it also supports open source technologies. We can use our preferred tools and technologies. By using our own device with our operating system, we can run virtually any application using our data source. We have multiple options with Azure [57].

Moreover, Azure provides flexible purchasing and pricing options for all cloud scenarios. You only pay for what you use without any upfront cost [57].

Azure is the only consistent hybrid cloud, that delivers unparalleled developer productivity. It provides comprehensive, multilayered security, which includes the largest compliance coverage of any cloud service provider. We also pay much less in Azure than other popular Cloud service providers for Windows Server and SQL Server [58].

Azure has special offer for student developers to create a free account and use their resources and learn cloud services with free 100 dollar credit for 12 months. I took the opportunity and utilized the privilege of being a graduate student from University of Manitoba and created a student account to perform my experiments using Azure's Virtual Machine (VM).

After creating my student account, I created a VM to perform my experiments. Using the student account, I have done most of the experiments for my research work.

I also opened a pay-as-you-go account as the student account doesn't provide encryption services. To test the encryption services to provide data security is another part of my research work.

To test the accuracy and efficiency of the whole process, I'm planning to use portfolio data freely available from Yahoo! Finance.

4.2 Step 2: Data Security

For this research work, my primary concern is to protect the sensitive financial data from being stolen or illegal use by attackers. From the literature review, it is clear that data analysis has been done with decryption and without decryption. I am interested to do the data analysis with decryption.

There are two different ways to secure the sensitive data. First, utilizing cloud providers encryption of data in transit service to ensure secure transmission of the data from the user to the cloud and from the cloud to the user. Second, utilizing an encryption algorithm to encrypt the data before sending it to the cloud, after the encrypted data go to the cloud decrypt the data, computing on the decrypted data, get the output, encrypt the output, send the encrypted output to the user and finally the user will decrypt it to know the actual output.

4.2.1 Encryption of Data in Transit

I am using Microsoft Azure as my cloud provider for the data computation. I have explored their security services. Azure is offering different types of security services for different security requirements such as encryption of data at rest, encryption of data in transit, in-transit encryption in VMs, Azure VPN encryption, key management with key vault etc. Azure is also offering different mechanisms to secure the data when moving from one place to another. Encryption of data in transit includes different types of services for different concerns [59].

Data-link layer encryption in Azure is used when customer traffic shifts between data centers. Microsoft does not control outside of physical boundary and IEEE 802.1AE MAC Security Standards or MACsec is used for point-to-point covering the primary network hardware. Before sending, the packets are encrypted and decrypted on the device to block physical ”man-in-the-middle” or snooping/wiretapping attacks. The technology delivers line rate encryption on the network hardware with no measurable link latency increase because of the integration on the network hardware. This MACsec is activated by default for all Azure traffic moving within a region or between regions. Customers do not need to take any action to enable the service [60].

TLS encryption in Azure is the facility where clients can utilize Transport Layer Security (TLS) protocol given by Microsoft. TLS protocol is used when data is moving between clients and the cloud services. Microsoft datacenters settled a TLS connection with client systems, which can connect to services provided by Azure. TLS gives powerful authentication, message privacy, integrity, interoperability, algorithm flexibility, ease of deployment and use. Message tampering detection, interception and forgery is included in integrity. Perfect Forward Secrecy (PFS) secures connection between client and cloud services using unique keys. RSA-based 2,048-bit encryption key length is used in connections. It is almost impossible to access the data in transit because of this fusion [60].

Azure storage transactions is another service to interchange with Azure Storage via Azure portal where all of the contracts happens above HTTPS. Storage REST API can be used above HTTPS to interchange with Azure Storage. Shared Access Signatures (SAS) is utilized to assign approach to Azure Storage objects. When using Shared Access Signatures, there is a choice to define only HTTPS protocols can be utilized. This process guarantees that the people are using appropriate protocol when transmitting links for SAS tokens. SMB 3.0 is utilized for accessing Azure Files shares and assists encryption. It can be accessible in Windows Server 2012 R2, Windows 8, Windows 8.1 and Windows 10. It authorizes desktop access and even cross-region access. Data is encrypted by client-side encryption before sending it to Azure Storage instance. Therefore, the data is encrypted when it is moving over the network [60].

SMB encryption over Azure virtual networks is utilized when SMB 3.0 in VMs is used and controlling by Windows Server 2012 or later. By using encryption data in transit above Azure Virtual Networks, data transmissions can be safe. Data encryption secures data in case of tampering and eavesdropping attacks. SMB encryption can be authorized for the whole server or for particular shares by the administrator.

When SMB encryption is activated for a share or server, after that only SMB 3.0 customers are authorized to retrieve the encrypted shares [60].

4.2.2 Encryption Algorithm

Nowadays, Python libraries for cryptography is used widely for encrpyting and decrypting data. Cryptography is a package with cryptographic recipes and primitives for Python developers. It incorporates high level recipes and low level interfaces to usual cryptographic algorithms including symmetric ciphers, message digests and key derivation functions.

Fernet implementation is the most convenient high-level secure primitive in cryptography. It is a standard for encrypting buffers because it sticks to best ways of cryptography. Large files such as gigabyte range and above are not appropriate for it. Because the entire buffer for encrypting or decrypting is to be loaded into memory at once. It assists symmetric or secret key cryptography. The key should be remain safe since the identical key is used for both encryption and decryption.

To generate a key:

$>>> k = fernet.Fernet.generate_k ey()$

$>>> type(k)$

$< class'bytes' >$

After getting the key, encryption can be done like this:

$>>> frn = fernet.Fernet(k)$

$>>> encrypted = frn.encrypt(b"xmarksthespot")$

$>>> encrypted[: 10]$

$b'gAAAAABb1'$

Decryption can be done like this:

$>>> frn = fernet.Fernet(k)$

$>>> frn.decrypt(encrypted)$

$b'xmarksthespot'$

A symmetric cipher uses the identical key for encryption and decryption. A secret key or private key is used by symmetric ciphers to convert the plaintext into ciphertext and vice versa. Some symmetric ciphers are Advanced Encryption Standard (AES), Data Encryption Standard (DES), Blowfish and Interntional Data Encryption Algorithm (IDEA). Here we discuss the Advanced Encryption Standard (AES) in detail.

The Advanced Encryption Standard (AES)

The Advanced Encryption Standard (AES) is a symmetric block cipher to encrypt electronic data. It is also known by its original name Rijndael. It is initiated by the U.S. National Institute of Standards and Technology (NIST) in 2001 to protect classified information. It is widely used throughout the world for encrypting sensitive data. It is crucial for government computer security, cyber security and electronic data protection [61].

In 1997, the NIST announced the necessity for a substitute to the Data Encryption Standard (DES) which became exposed to brute-force attacks. Hence, the NIST started forming AES.

AES consists of three block ciphers:

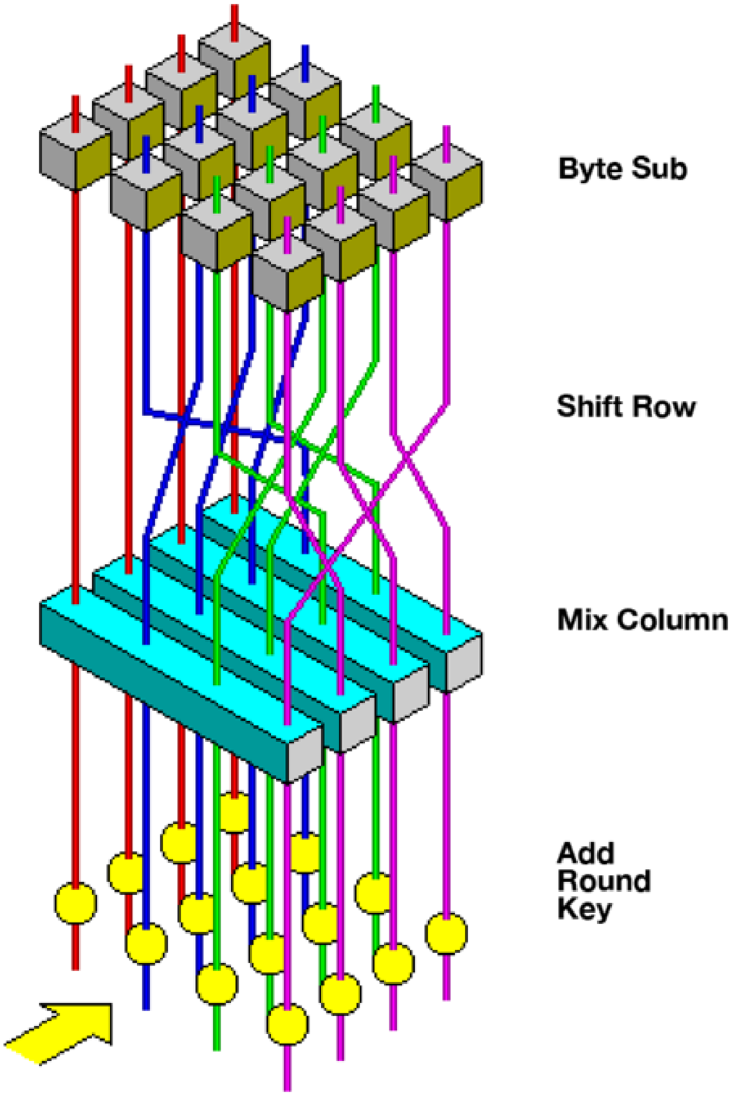


Figure 4.1: Visualization of AES round function. Source: Wikipedia.

1. AES-128: Key size 128-bit is used to encrypt and decrypt a block of messages.

2. AES-192: Key size 192-bit is used to encrypt and decrypt a block of messages.

3. AES-256: Key size 256-bit is used to encrypt and decrypt a block of messages.

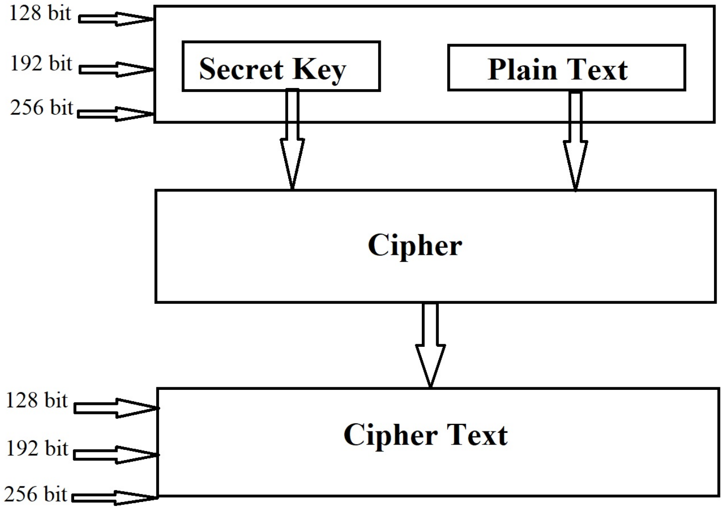
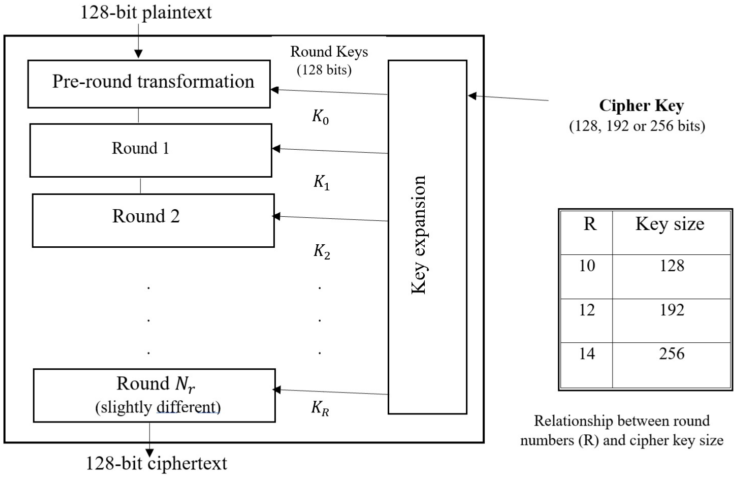


Figure 4.2: The design of AES.

For 128 bits of data blocks are encrypted and decrypted by each cipher using 128, 192 and 256 bits of cryptographic keys. There are 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. A round includes several processing steps consisting of substitution, transposition and mixing to convert the plaintext into ciphertext [61].

Some features of AES includes symmetric key with symmetric block cipher, 128-bit data with 128 or 192 or 256-bit keys, stronger and faster than Triple-DES, provides full specification and design details, software implementable etc.

The AES is an iterative cipher which is based on ”substitution-permutation network”. It involves a sequence of connected operations where some of them replace inputs by particular outputs called substitutions and some of them shuffle bits around called permutations. It executes all computations on bytes rather than bits. Thus, it



R	Key size
10	128
12	192
14	256

Figure 4.3: The structure of AES.

handles 128 bits plaintext block as 16 bytes. The 16 bytes are organized in four rows and four columns to process as a matrix. The AES encryption algorithm specifies various transformations which are implemented on the data stored in an array. The first step of the cipher is to set the data into an array. Then the cipher transformations are repeated over multiple encryption rounds. The first transformation is exchange of data using a substitution table. The second transformation moves data rows. The third mixed columns. The last transformation is implemented on each column utilizing a discrete part of the encryption key. Prolonged keys need more rounds to finish [61].

Encryption Process:

The encryption process involves several rounds. Here, we discuss a representative round of AES encryption. Each round incorporates with four sub processes. The first round can be illustrated like this [61].

Byte Substitution (SubBytes): The input 16 bytes are substituted with a fixed

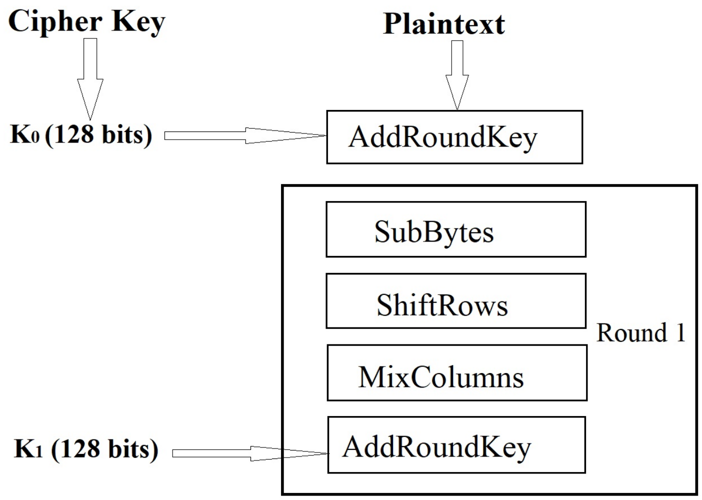


Figure 4.4: The first round of AES.

table which makes a matrix of four rows and four columns.

Shiftrows: Each row of the matrix is shifted to the left. The fallen entries are re-inserted on the right side of the row. Shift is executed by following rules. First row is not shifted. Second row is shifted by one byte position to the left. Third row is shifted by two positions to the left. Fourth row is shifted by three positions to the left. As a result, we get a new matrix with previous 16 bytes but shifted with respect to each other.

MixColumns: Each column is changed utilizing a special mathematical function. The function takes four bytes of one column as input and provides four completely new bytes as output. As a result, we have another new matrix with 16 new bytes. This step is not executed in the last round.

Addroundkey: The 16 bytes of the matrix are reviewed as 128 bits and are XOR to the 128 bits of the round key. The output of the last round is the ciphertext. For other rounds, the derived 128 bits are converted as 16 bytes and continue the process again.

Decryption Process:

The decryption process is similar to the encryption process but in an opposing direction. Each round involves four processes executed in opposing order such as Add round key, Mix columns, Shift rows and Byte substitution. Since sub-processes are in reverse manner, the encryption and decryption algorithms are implemented separately, though they are closely related [61].

In modern cryptography, AES is widely accepted and funded in both software and hardware. AES has integrated flexibility for key length which makes key searching impossible for attackers and it becomes secure against any kind of attack. The security of AES is guaranteed only if is properly implemented and key is managed professionally.

Chapter 5

Experimental Setup, Implementation and Evaluation

In this chapter, I will discuss my experimental setup, results and evaluation in detail. Figure 5.1 shows the overall high level implementation procedure.

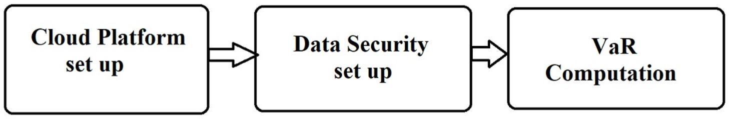


Figure 5.1: Overall implementation procedure.

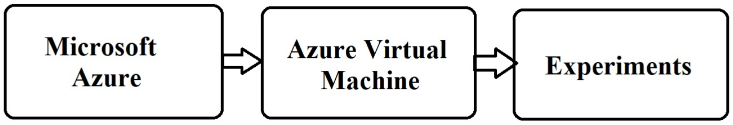


Figure 5.2: Microsoft Azure Cloud Platform set up.

5.1 Cloud Platform set up

To implement VaR as a service on Cloud, I choose to use Microsoft Azure cloud resource. For student developers, Azure provides special offer to create a free account and free access to popular cloud services and developer tools for 12 months with $100 credit. We can select the tools and framework to build, run and manage applications. It supports Windows apps and services as well as open source technologies. We can utilize our own device, our operating system to run any application virtually by using our data source. Figure 5.2 shows the overall Microsoft Azure cloud platform setup.

I utilized the opportunity of being a graduate student to open a student account. After creating my account, I created a Virtual Machine (VM) with necessary resource group, location of the VM to conduct my experiments. The specifications are as follows: resource group is computation, location is north central US, size is Standard B1s (1vcpu, 1GiB memory), operating system is Windows, plan is 2022-datacenter-azure-edition, VM generation is V2, public IP address is 65.52.18.128.

I also opened a pay-as-you-go account to perform my experiments with added Azure's data security service, encryption data in transit as the student account doesn't provide the security services. The specifications are as follows: resource group is com-

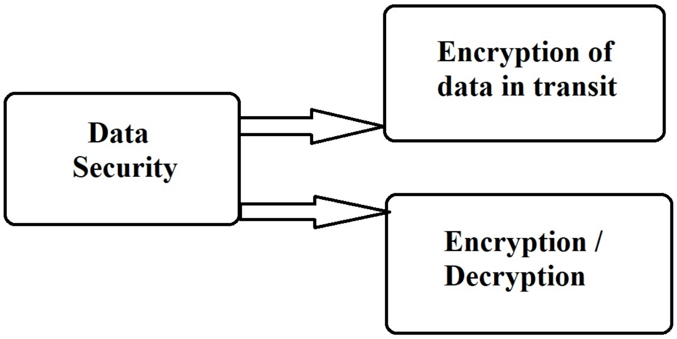


Figure 5.3: Approaches to set up for Data Security.

putation, location is north central US, size is Standard B1s (1vcpu, 1GiB memory), operating system is Windows, plan is 2022-datacenter-azure-edition, VM generation is V2, public IP address is 52.159.98.82.

I also installed Anaconda 3 with Spyder (Python 3.8) in my VM to perform the computations.

5.2 Data Security set up

To protect the sensitive financial data, I have used two different approaches. First approach is utilizing cloud providers encryption of data in transit service. Second approach is utilizing an encryption algorithm to encrypt the data. Figure 5.3 shows the approaches for data security setup.

First approach set up:

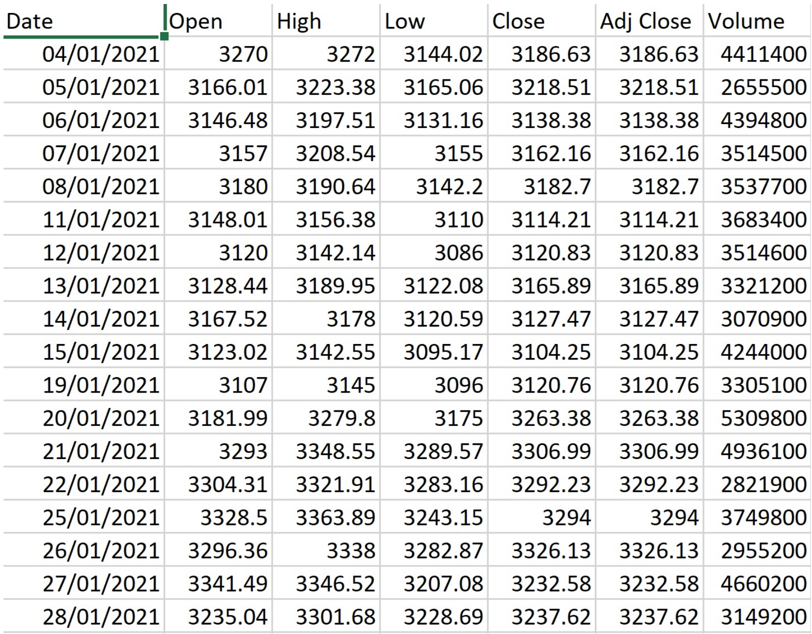
For my experiments, I am using Microsoft Azure as my cloud provider. Encryption of data in transit is one of their security services for the data which moves from one place to another. Specifically, TLS encryption in Azure is the perfect service for the research requirement where TLS protocol is used for the data moving between clients and the cloud services. PFS provides RSA-based unique keys for connections between client and cloud services. The encryption service is so powerful that it is almost impossible to get access of the data in transit.

From the Azure portal, we have to go to Security services then we have to go Microsoft Defender for Cloud, which is formerly known as Security Center. It gives a overview of the security conditions for the subscription such as secure score, recommendations etc. In the recommendations tab, it will show the secure score based on the healthy and unhealthy resources. They will show some recommendations to follow to achieve full secure score. Each recommendation provides a short description about it. You can enable the recommendations from there. Here, I have enabled Encryption of Data in Transit which will protect the data in transit from the personal computer to cloud VM and vice versa.

Second approach set up:

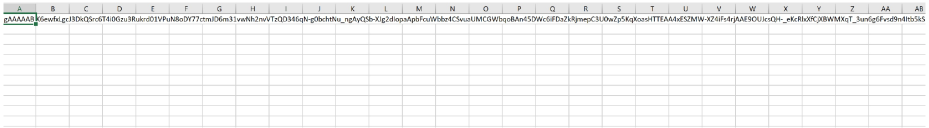
To setup the second approach, I have used Python libraries for cryptography. I build up the code for encrypting and decrypting the data in my personal computer using Python programming language.

I downloaded the required data from Yahoo!Finance as a csv file. I encrypted the data before sending it to cloud. When the encrypted data reached to the cloud, the data is decrypted before the computation since computations cannot be done on the encrypted data, which becomes strings of numbers and alphabets. After completing the computation, the result is encrypted using the encryption code. The encrypted



Date	Open	High	Low	Close	Adj Close	Volume
04/01/2021	3270	3272	3144.02	3186.63	3186.63	4411400
05/01/2021	3166.01	3223.38	3165.06	3218.51	3218.51	2655500
06/01/2021	3146.48	3197.51	3131.16	3138.38	3138.38	4394800
07/01/2021	3157	3208.54	3155	3162.16	3162.16	3514500
08/01/2021	3180	3190.64	3142.2	3182.7	3182.7	3537700
11/01/2021	3148.01	3156.38	3110	3114.21	3114.21	3683400
12/01/2021	3120	3142.14	3086	3120.83	3120.83	3514600
13/01/2021	3128.44	3189.95	3122.08	3165.89	3165.89	3321200
14/01/2021	3167.52	3178	3120.59	3127.47	3127.47	3070900
15/01/2021	3123.02	3142.55	3095.17	3104.25	3104.25	4244000
19/01/2021	3107	3145	3096	3120.76	3120.76	3305100
20/01/2021	3181.99	3279.8	3175	3263.38	3263.38	5309800
21/01/2021	3293	3348.55	3289.57	3306.99	3306.99	4936100
22/01/2021	3304.31	3321.91	3283.16	3292.23	3292.23	2821900
25/01/2021	3328.5	3363.89	3243.15	3294	3294	3749800
26/01/2021	3296.36	3338	3282.87	3326.13	3326.13	2955200
27/01/2021	3341.49	3346.52	3207.08	3232.58	3232.58	4660200
28/01/2021	3235.04	3301.68	3228.69	3237.62	3237.62	3149200

Figure 5.4: Amazon financial data for the year 2021.

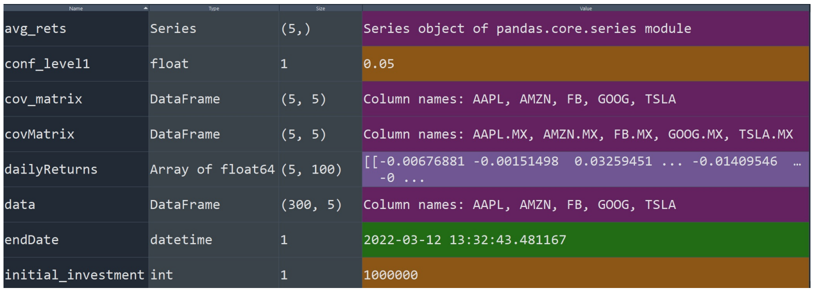


```
gAAAAABK6ewfxLgcJ3DkQSrc6T4i0Gzu3Rukrd01VPuN8oDY77ctmJD6m31vwNh2nvVTzQD346qN-g0bchtNu_ngAyQSb-XJg2dlopaApbFcuWbbz4CSvuaUMCGWbqoBAn45DWc6iFDaZkRjmepC3U0wZp5KqXoasHTTEAA4xESZMW-XZ4iFs4rjAAE9OUJcsQH-_eKcRlxXfCjXBWMXqT_3un6g6Fvsd9n4Itb5kS
```

Figure 5.5: Encrypted Amazon data for the year 2021.

result comes to my personal computer. After receiving the encrypted result, I used the decryption code to decrypt and get the actual result.

Here is an illustration to show the encryption procedure. I have downloaded 1 year financial data of Amazon for the year 2021 from Yahoo!Finance. I have applied my encryption code in Python on the data to encrypt it. In Figure 5.4, the financial data of Amazon for the year 2021 can be seen. In Figure 5.5, the encrypted data is shown after applying encryption on the data of Figure 5.4.



Name	Type	Size	Value
avg_rets	Series	(5,)	Series object of pandas.core.series module
conf_level1	float	1	0.05
cov_matrix	DataFrame	(5, 5)	Column names: AAPL, AMZN, FB, GOOG, TSLA
covMatrix	DataFrame	(5, 5)	Column names: AAPL.MX, AMZN.MX, FB.MX, GOOG.MX, TSLA.MX
dailyReturns	Array of float64	(5, 100)	[[-0.00676881 -0.00151498 0.03259451 ... -0.01409546 … -0 ...
data	DataFrame	(300, 5)	Column names: AAPL, AMZN, FB, GOOG, TSLA
endDate	datetime	1	2022-03-12 13:32:43.481167
initial_investment	int	1	1000000

Figure 5.6: Variables for VaR Computation.

5.3 VaR Computation

In previous chapter, I have described the MC simulation which is my computing methodology for VaR Computation. To perform the experiments, I have used Python programming language to build up the code in my personal laptop Dell Inspiron 7391 16GB RAM with Intel Core i7 10th generation at 1.80GHz processor. I have tested with stock data for 100 days with 10,000 simulations. To test the code and get the value of VaR, I have used stock data of major companies like Apple (AAPL), Amazon (AMZN), Facebook (FB), Google (GOOG) and Tesla (TSLA) which is freely available in Yahoo! Finance. The variables for the VaR computation can seen in Figure 5.6.

The first experiment is done in my personal laptop. Figure 5.7 represents VaR computation using MC simulation of stock portfolio performed in my personal laptop. The VaR result of this computation is 1615.61.

After performing tests in my personal laptop, I have performed same experiments in my cloud Virtual Machine (VM) in Microsoft Azure with activated encryption of data in transit. Figure 5.8 shows VaR computation using MC simulation of stock portfolio performed in cloud VM. The VaR result of this computation is 2234.24.

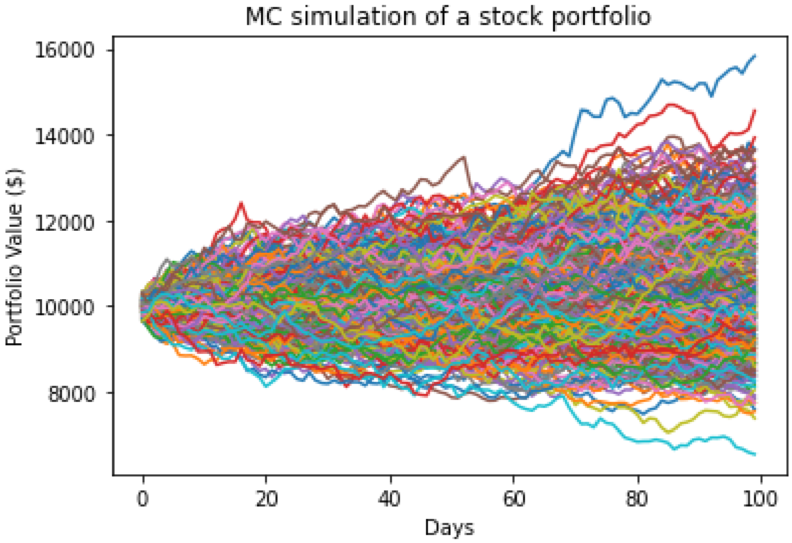


Figure 5.7: VaR Computation from personal Laptop.

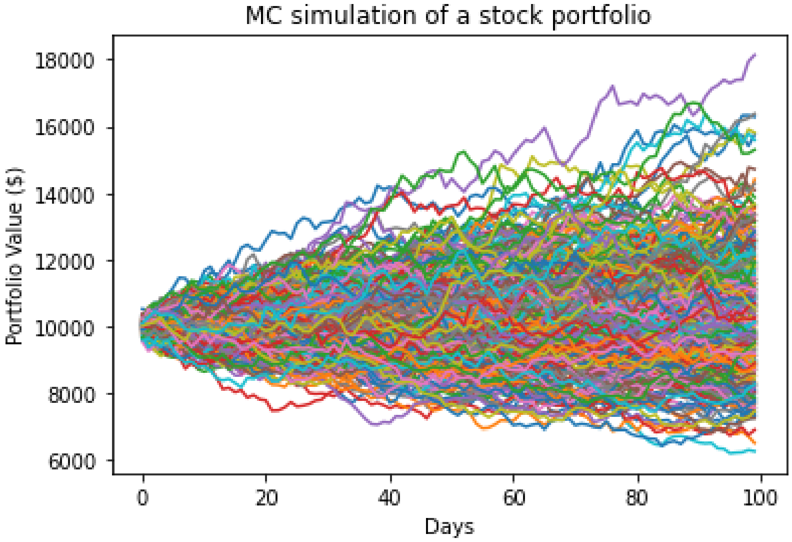


Figure 5.8: VaR Computation from Cloud VM using encryption of data in transit.

This final test is performed on the data being encrypted and decrypted before the computation. Figure 5.9 shows VaR computation using MC simulation of stock portfolio performed in cloud VM. The VaR result of this computation is 1470.94.

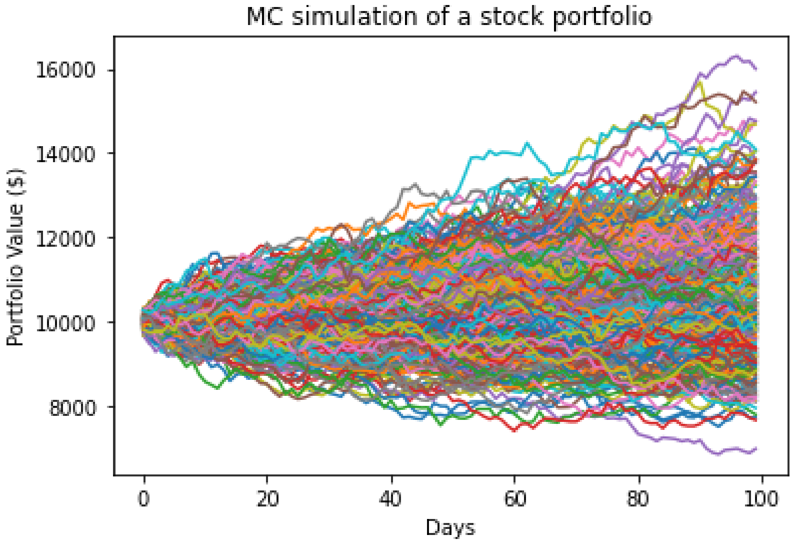


Figure 5.9: VaR Computation from Cloud VM using Python libraries for cryptography.

The experimental results prove that there is no significant difference in performance while doing VaR computation. The MC simulation provides the acceptable VaR values for stock portfolio in all three cases. Hence, VaR computation on Cloud using Encryption data in transit and encryption algorithm both are acceptable.

Chapter 6

Conclusion

This research work combines and utilizes three challenging computer science fields together which are Computational Finance, Cloud Computing and Privacy and Security. The goal of this research work is providing VaR computing on cloud platform with data security. The procedure includes MC Simulation which provides approximate VaR values while encryption procedure ensures the data security. The evaluation process checks the accuracy and usefulness of VaR value and encrypted data. My primary aim is to provide the service of VaR computation, which investors can utilize from any corner of the world through cloud whenever needed and could take their decision of investment on the spot.

I achieved this goal by setting up a VM environment with MS Azure cloud platform and implementing a MC simulation study for VaR computation.The results computed on Cloud is comparable to the stand-alone computation of the same MC algorithm.

Bibliography

[1] C. H. Cooper, *Global Association of Risk Professionals, Financial Risk Manager Exam Study Guide.* New York, USA: GARP, 2016.

[2] A. Damodaran, "http://people.stern.nyu.edu/adamodar/pdfiles/papers/VAR.pdf (undated and unpublished personal website on-line article - last accessed on March 22, 2016)," NewYork, NY, USA, 2016.

[3] J. Hull, *Options, Futures and Other Derivates.* Princeton, USA: Prentice Hall, 2011.

[4] P. P. Ray, "An introduction to dew computing: Definition, concept and implications," *IEEE Access*, vol. 6, pp. 723–737, 2018.

[5] K. Pilbeam, *Finance financial markets / Keith Pilbeam.*, 3rd ed. Basingstoke England ; New York: Palgrave Macmillan, 2010.

[6] J. Oberoi, "Interest rate risk management and the mix of fixed and floating rate debt," *Journal of Banking Finance*, vol. 86, pp. 70–86, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378426617302157

[7] G. A. Holton, "Defining risk," *Financial analysts journal*, vol. 60, no. 6, pp. 19–25, 2004.

[8] T. J. Linsmeier and N. D. Pearson, "Value at risk," *Financial analysts journal*, vol. 56, no. 2, pp. 47–67, 2000.

[9] N. Mehra, S. Aggarwal, A. Shokeen, and D. Bura, "Analyzing cloud computing security issues and challenges," in *Progress in Computing, Analytics and Networking.* Springer, 2018, pp. 193–202.

[10] B. Alouffi, M. Hasnain, A. Alharbi, W. Alosaimi, H. Alyami, and M. Ayaz, "A systematic literature review on cloud computing security: threats and mitigation strategies," *IEEE Access*, vol. 9, pp. 57 792–57 807, 2021.

[11] E. Thambiraja, G. Ramesh, and D. R. Umarani, "A survey on various most common encryption techniques," *International journal of advanced research in computer science and software engineering*, vol. 2, no. 7, 2012.

[12] R. Irons, *The Fundamental Principles of Finance.* Routledge, 2019.

[13] A. Thavaneswaran, R. K. Thulasiram, Z. Zhu, M. E. Hoque, and N. Ravishanker, "Fuzzy value-at-risk forecasts using a novel data-driven neuro volatility predictive model," in *COMPSAC (2).* IEEE, 2019, pp. 221–226.

[14] H. K. Baker and G. Filbeck, *Investment risk management.* Oxford University Press, 2014.

[15] J. Danielsson and C. G. De Vries, "Value-at-risk and extreme returns," *Annales d'Economie et de Statistique*, pp. 239–270, 2000.

[16] J. Butler and B. Schachter, "Estimating value-at-risk with a precision measure by combining kernel estimation with historical simulation," *Review of Derivatives Research*, vol. 1, pp. 371–390, 1997.

[17] X. Cui, X. Sun, S. Zhu, R. Jiang, and D. Li, "Portfolio optimization with nonparametric value at risk: A block coordinate descent method," *INFORMS Journal on Computing*, vol. 30, no. 3, pp. 454–471, 2018.

[18] C.-S. Wang and Z. Zhao, "Conditional value-at-risk: Semiparametric estimation and inference," *Journal of Econometrics*, vol. 195, no. 1, pp. 86–103, 2016.

[19] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2016.

[20] P. A. Lewis and E. J. Orav, *Simulation methodology for statisticians, operations analysts, and engineers*. Chapman and Hall/CRC, 2017.

[21] E. Gobet, *Monte-Carlo methods and stochastic processes: from linear to nonlinear*. Chapman and Hall/CRC, 2016.

[22] A. Brabazon, M. O'Neill, and D. Maringer, *Studies in Computational intelligence*. Springer, 2012, vol. 4.

[23] A. Brabazon and M. O'Neill, *Biologically inspired algorithms for financial modelling*. Springer Science & Business Media, 2006.

[24] B. Sharma, R. K. Thulasiram, and P. Thulasiraman, "Computing value-at-risk using genetic algorithm," *The Journal of Risk Finance*, vol. 16, no. 2, pp. 170–189, 2015.

[25] S. Mirjalili, "Genetic algorithm," in *Evolutionary algorithms and neural networks.* Springer, 2019, pp. 43–55.

[26] D. C. Marinescu, *Cloud computing: theory and practice.* Morgan Kaufmann, 2022.

[27] A. Waksman and S. Sethumadhavan, "Silencing hardware backdoors," in *2011 IEEE Symposium on Security and Privacy.* IEEE, 2011, pp. 49–63.

[28] M. Bellare and P. Rogaway, "Introduction to modern cryptography," *Ucsd Cse*, vol. 207, p. 207, 2005.

[29] G. C. Kessler, "An overview of cryptography (updated version 24 january 2019)," 2019.

[30] R. Bhanot and R. Hans, "A review and comparative analysis of various encryption algorithms," *International Journal of Security and Its Applications*, vol. 9, no. 4, pp. 289–306, 2015.

[31] D. Fiore, R. Gennaro, and V. Pastro, "Efficiently verifiable computation on encrypted data," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 844–855.

[32] A. A. Yazdeen, S. R. Zeebaree, M. M. Sadeeq, S. F. Kak, O. M. Ahmed, and R. R. Zebari, "Fpga implementations for data encryption and decryption via concurrent and parallel computation: A review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 8–16, 2021.

[33] Y. Acar, M. Backes, S. Fahl, S. Garfinkel, D. Kim, M. L. Mazurek, and C. Stransky, “Comparing the usability of cryptographic apis,” in *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2017, pp. 154–171.

[34] M. Almorsy, J. Grundy, and I. Müller, “An analysis of the cloud computing security problem,” *arXiv preprint arXiv:1609.01107*, 2016.

[35] D. Hendricks, “Evaluation of value-at-risk models using historical data,” *Economic policy review*, vol. 2, no. 1, 1996.

[36] P. Glasserman, P. Heidelberger, and P. Shahabuddin, “Efficient monte carlo methods for value-at-risk,” 2000.

[37] S. Manganelli and R. F. Engle, “Value at risk models in finance,” 2001.

[38] S. Antonelli and M. G. Iovino, “Optimization of monte carlo procedures for value at risk estimates,” *Economic Notes*, vol. 31, no. 1, pp. 59–78, 2002.

[39] Y. Liu, “Value-at-risk model combination using artificial neural networks,” *Emory University Working Paper Series*, 2005.

[40] P. Aniūnas, J. Nedzveckas, and R. Krušinskas, “Variance–covariance risk value model for currency market,” *Engineering economics*, vol. 61, no. 1, 2009.

[41] D. d. Weerd, “Value-at-risk using the variance-covariance approach,” Ph.D. dissertation, University of Groningen. Faculty of Economics and Business, 2011.

[42] L. J. Hong, Z. Hu, and G. Liu, “Monte carlo methods for value-at-risk and conditional value-at-risk: a review,” *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 24, no. 4, pp. 1–37, 2014.

[43] A. Suwarno and P. A. Mahadwartha, "The analysis of portfolio risk management using var approach based on investor risk preference," *KINERJA*, vol. 21, p. 129, 09 2017.

[44] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[45] H.-G. Zhang, C.-W. Su, Y. Song, S. Qiu, R. Xiao, and F. Su, "Calculating value-at-risk for high-dimensional time series using a nonlinear random mapping model," *Economic Modelling*, vol. 67, pp. 355–367, 2017.

[46] Y. M. P. Sumaji, "The calculation of value at risk using variance covariance in lq-45 companies," *Business and Finance Journal*, vol. 6, no. 2, 2021.

[47] R. Van den Goorbergh, "Value-at-risk analysis and least squares tail index estimation," *Research Memorandum WO&E*, vol. 578, 1999.

[48] T. J. Linsmeier and N. D. Pearson, "Value at risk," *Financial Analysts Journal*, vol. 56, no. 2, pp. 47–67, 2000.

[49] M. Sarma, S. Thomas, and A. Shah, "Selection of value-at-risk models," *Journal of Forecasting*, vol. 22, no. 4, pp. 337–358, 2003.

[50] W. Steelyana *et al.*, "Value at risk-which one is better: Historical simulation or variance covariance approach?(comparative analysis between historical simulation and variance covariance method with portfolio simulation at idx)," in *1st ICFERMA, International Conference on Financial Engineering and Risk Management*, 2011.

[51] S. O. Oppong, D. Asamoah, and E. O. Oppong, “Value at risk: historical simulation or monte carlo simulation,” in *International Conference On Management*, 2016.

[52] J. Berkowitz and J. O’Brien, “How accurate are value-at-risk models at commercial banks?” *The journal of finance*, vol. 57, no. 3, pp. 1093–1111, 2002.

[53] B. Sharma, R. K. Thulasiram, P. Thulasiraman, and R. Buyya, “Clabacus: a risk-adjusted cloud resources pricing model using financial option theory,” *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 332–344, 2014.

[54] K. Gai, L. Qiu, H. Zhao, and M. Qiu, “Cost-aware multimedia data allocation for heterogeneous memory using genetic algorithm in cloud computing,” *IEEE Transactions on Cloud Computing*, vol. 8, no. 4, pp. 1212–1222, 2020.

[55] J. Philip and D. Shah, “Implementing signature recognition system as saas on microsoft azure cloud,” *Data Management, Analytics and Innovation, Advances in Intelligent Systems and Computing*, pp. 479–488, 2019.

[56] C. Kotas, T. Naughton, and N. Imam, “A comparison of amazon web services and microsoft azure cloud platforms for high performance computing,” in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, 2018, pp. 1–4.

[57] M. Copeland, J. Soh, A. Puca, M. Manning, and D. Gollob, “Microsoft azure,” *New York, NY, USA:: Apress*, pp. 3–26, 2015.

[58] T. Madhuri and P. Sowjanya, “Microsoft azure v/s amazon aws cloud services:

A comparative study," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 5, no. 3, pp. 3904–3907, 2016.

[59] A. Rath, B. Spasic, N. Boucart, and P. Thiran, "Security pattern for cloud saas: From system and data security to privacy case study in aws and azure," *Computers*, vol. 8, no. 2, p. 34, 2019.

[60] P. De Tender, D. Rendon, and S. Erskine, "Azure security center," in *Pro Azure Governance and Security.* Springer, 2019, pp. 101–179.

[61] A. Bogdanov, D. Khovratovich, and C. Rechberger, "Biclique cryptanalysis of the full aes," in *International conference on the theory and application of cryptology and information security.* Springer, 2011, pp. 344–371.