Chapter 4
Research Methodology
CHAPTER 4
RESEARCH METHODOLOGY

Research is a process of scientific investigation with a specific aim and agenda, focused at gaining further information and transforming that into knowledge through testing new ideas and techniques. This process involves:

- Statement of research objectives
- Collection of facts and figures related to the research objectives
- Systematic arrangement of gathered/collected data through proper tables, charts, figures & diagrams
- Data Analysis
- Deduction and Interpretation of results
- Driving conclusions and making recommendations on the basis of the work carried out

Methodology dwells on the research problem and explains the significance of the problem. It lays the foundation of the research work, sets the environment & directions and discusses the implications of the work once it culminates. In a nutshell, it deals with the finest details from conception to completion of the research work.

4.1 Research Objectives

The major objectives of this research work can be enumerated as:

- To study the diverse aspects related to cloud computing, focusing on the different processes, prospects and environment affiliated to it.
- To examine the economic aspect of cloud computing comprising of the diverse nature of benefits, costs, business opportunities, employment opportunities, etc. associated with the migration of traditional IT services towards that of cloud computing.
To explore the causes, effects and interrelationship of cloud computing and different macroeconomic variables like investment, consumption expenditure, government expenditure, growth rate, etc.

The current research work is a result of an exhaustive study of numerous research papers, articles, reports and white papers of different institutions, government bodies, research organizations and scholars. In order to seek elaborate knowledge about the economics of cloud computing and affiliated issues, detailed study was conducted.

The study has probed both the theoretical and practical facets.

First objective

- It includes the study of the concept as such with getting deeper knowledge about its definitions, characteristics, prospects and models
- The advancement of cloud computing as the most promising technology in the ICT field

Second objective

- To explore the concept of cloud economics, the economic aspect of cloud computing, in detail
- To analyze the economic benefits of cloud computing which are compelling enough to get the technology adopted in different sectors of any economy
- To classify the costs affiliated with the migration process of client-server technology to cloud computing
- To examine the nature of the related costs in order to comprehend the true economic nature of cloud computing
- To gain insight of the diverse economic aspects such as demand/supply side behaviour or economies attached with the cloud computing
To study the impact of changing technology on the nature of employment and job scenario on economies in general
To investigate the consequences of the gradual zealous adoption of cloud computing on different markets/industry/sectors

Third objective

To explore the effects of cloud computing on diverse macroeconomic variables viz. national income, employment, government expenditure, investment expenditure, consumption expenditure, growth rate of gross domestic product, etc.
By making use of different statistical tools and techniques, to analyse the relationship between diverse macroeconomic variables and cloud computing aspects
To represent the corresponding analysis in diagrammatical and tabular form in order to comprehend the underlying economics in a better way
To examine the effects of cloud computing on growth and reallocation in different sectors and economies
To unravel the already discussed issues related to cloud computing with special reference to Indian Economy

4.2 Research Design & Data Sources

4.2.1 Research Design

In order to probe into the economic perspective of cloud computing, first of all appropriate cloud computing aspects (relevant for the study) are decided.
After deciding the cloud computing aspects, such macroeconomic variables are selected, which are instrumental in explaining the impact of cloud computing on the economies in an effectual manner.
As cloud computing is the base for the internet economy, therefore, further with the help of the selected cloud computing aspects and macroeconomic variables, interrelationships are set up. The interrelationships between macroeconomic variables and cloud computing aspects are then methodically worked upon and tested. Appropriate and relevant statistical tools and techniques are employed to bring about the true relationship (whether cause and effect or otherwise) to the surface. The analysis work also involves categorical usage of tables, charts, figures and diagrams.

The analytical work is done in various stages involving separate statistical techniques for each stage.

### 4.2.2 Data Sources

This research work is based on secondary data. To carry out the research work in a meaningful way, to fulfill the research objectives and to make the results more credible in nature, secondary data is used. Such efforts are made in the direction of collection of data so that it serves the purpose of maintaining reliability, accuracy and appropriateness.

The nature of the data collected is spacio-temporal. The collected data is concerned with several macroeconomic variables and facts, figures and trends related to cloud computing of different countries for a relevant time period as per the requirements of the analysis. In the current study, it was unfeasible to collect first-hand data on the concerned issues and thus secondary data is used.

Sources of data:

- Reports/official publications of international institutions and bodies such as World Bank, International Telecommunication Union, United Nations, etc.
- Reports/publications of Central and State Governments
- Private publications of trade and professional bodies
- Financial & Economic Journals
- Private publications of renowned consultancy firms
- Publications brought out by research scholars, research agencies (both national & international)

The raw data is further processed, organized and presented in a readily comprehensible manner in order to facilitate the process of statistical analysis.

4.3 Research Methodology

The first stage of exploration comprises of an analysis of the interrelationship between various macroeconomic variables and cloud/internet based aspects.

This is carried out by way of examining effect of change in macroeconomic variables on internet economy and the growth of the overall economy.

In the second stage, for the sake of detailed analysis, the macroeconomic variables are segregated, degree of relationship is examined and different aspects of relationship sets are prepared.

The different aspects of relationship sets represent the correlation between:

- Growth Rate of Real GDP (2014) and Cloud Computing Jobs (2014)
- Percentage Increase in Government Expenditure on Internet Economy and CAGR of Internet Economy (2010-2016)
- Percentage Increase in Investment Expenditure on Internet Economy and CAGR of Internet Economy (2010-2016)
- Percentage Increase in Consumption Expenditure on Internet Economy and CAGR of Internet Economy (2010-2016)
- CAGR of Internet Economy (2010-2016) and Average Growth Rate of Real GDP (2011-2016)
- Increase in Percentage of Online Retail (as Percentage of Total Retail) (2010-15) and Increase in Internet users (as Percentage of total population) (2010-15)

The analysis is made not only on the basis of the degree of relationship but also on the significance of that relationship. In the third stage, the significance of correlation is tested by applying t-test.

The fourth stage comprises of the further analysis of the nature of relationship between the different variables by making use of the technique of regression analysis. The fifth stage is solely devoted to analysis focused on Indian scenarios.

4.4 Statistical Techniques

4.4.1 Correlation

Correlation analysis literally means the relationship between two or more variables which vary in sympathy so that the movements in one variable tend to be accompanied by movement in the other variable. Here the primary objective is to measure the strength or degree of linear association between two variables. Sometimes the correlation is literally obvious, but the strength of the relationship cannot be known for sure until and unless analyzed methodically. It is important to bear in mind that correlation does not necessarily mean presence of a cause and effect relationship.

Thus correlation is that statistical technique which provides information about:

- Nature of relationship – positive or negative
- Strength of relationship
4.4.1.1 Correlation Coefficient

It is a mathematical method to measure the intensity of linear relationship between two variables. As developed by Karl Pearson, it is known as Pearsonian coefficient of correlation between two variables ‘X’ and ‘Y’. It is denoted by “r”.

4.4.1.2 Karl Pearson’s Coefficient of Correlation

\[ r = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y} \]

Where
- \( r \): Karl Pearson’s Coefficient of Correlation
- \( \text{Cov}(X,Y) \): Covariance between variable X and Y
- \( \sigma_x \): Standard Deviation of variable X
- \( \sigma_y \): Standard Deviation of variable Y

4.4.1.3 Properties of Correlation Coefficient

- The value of “r” lies between – 1 and +1 i.e., \(-1 \leq r \leq +1\).
- The correlation coefficient is a number independent of any unit of measurement.
- The correlation coefficient is independent of change of scale and origin.

<table>
<thead>
<tr>
<th>Characterizations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>– 1 or +1</td>
<td>Perfect correlation</td>
</tr>
<tr>
<td>-0.75 to -1 or 0.75 to 1</td>
<td>Strong correlation</td>
</tr>
<tr>
<td>-0.5 to -0.75 or 0.5 to 0.75</td>
<td>Moderate correlation</td>
</tr>
<tr>
<td>0 to - 0.5 or 0 to 0.5</td>
<td>Weak correlation</td>
</tr>
<tr>
<td>0</td>
<td>No correlation</td>
</tr>
</tbody>
</table>

Table 4.4.1.3: Strength of Relationship of ‘r’
4.4.2 Regression Analysis

The statistical technique used for investigating relationship between different variables is known as regression analysis. Regression analysis clearly indicates the cause and effect relationship. The variable(s) constituting cause is taken as the explanatory or independent variable(s) and the variable constituting the effect is taken as the explained or dependent variable.

Regression analysis simply expresses average relationship between two or more than two variables. Such relationship is generally expressed by a line of regression. There are always two lines of regression, in case of two variables (variables being ‘X’ and ‘Y’), one with ‘X’ as dependent variable and the other with ‘Y’ as dependent variable. In both the cases, the other variable acts as the independent variable.

The two regression equations are not reversible. As the assumptions and bases behind deriving these equations are different in nature, hence they become irreversible.

4.4.2.1 Equations of the Lines of Regression

\[ Y = a + bX \]  
\[ X = c + dY \]

In the above equations, equation (1) represents ‘Y’ on ‘X’ and equation (2) represents ‘X’ on ‘Y’.

In the first equation,

\[ Y : \text{Dependent Variable} \]
\[ X : \text{Independent Variable} \]
\[ a : \text{Intercept term} \]
b : Regression coefficient of variable ‘Y’ on variable ‘X’

In the second equation,

X : Dependent Variable

Y : Independent Variable

c : Intercept term

d : Regression coefficient of variable ‘X’ on variable ‘Y’

4.5 Hypothesis Testing

4.5.1 Test of Hypothesis

Hypothesis testing is a process meant for testing the significance of population parameter on the basis of a sample (drawn from that population). In the test of hypothesis, first a statistic is computed from the sample and on this basis it is seen whether the sample so drawn belongs to that population with certain specific characteristics. The process begins with making an assumption defined as ‘hypothesis’ that is made about a population parameter. Conventionally, there is not one but two hypothesis.

4.5.2 Setting of Hypothesis

Null Hypothesis: This is the hypothesis which is actually tested for possible rejection under the assumption that it is correct. It states that there is no significant difference between the population parameter and the sample statistic. ‘No significant difference’ means that if any difference is there, then it is merely because of sampling fluctuations. The null hypothesis is denoted by ‘H₀’.

Alternative Hypothesis: The hypothesis that is complementary in nature to the null hypothesis is known as the alternative hypothesis. The common statement
thereby – there is significant difference between population parameter and the sample statistic. It is denoted by ‘H<sub>A</sub>’ or ‘H<sub>1</sub>’.

It is significant to note that these hypotheses are made in advance before actually testing the significance of the parameter/statistic. The next step concerns with the computation of an appropriate test statistic. It is required to test whether the null hypothesis earlier made is going to be accepted or rejected.

As per the research objectives, the hypothesis to be tested is made about the significance of the correlation between the following sets:

**Relationship Set I**
- Growth Rate of Real GDP (2014)
- Cloud Computing Jobs (2014)

**Relationship Set II**
- GDP per capita (2014)
- Cloud Computing Jobs (2014)

**Relationship Set III**
- Percentage Increase in Government Expenditure on Internet Economy (2010-2016)
- CAGR of Internet Economy (2010-2016)

**Relationship Set IV**
- Percentage Increase in Investment Expenditure on Internet Economy (2010-2016)
- CAGR of Internet Economy (2010-2016)
Relationship Set V

- Percentage Increase in Consumption Expenditure on Internet Economy (2010-2016)
- CAGR of Internet Economy (2010-2016)

Relationship Set VI

- CAGR of Internet Economy (2010-2016)
- Average Growth Rate of Real GDP (2011-2016)

Relationship Set VII

- Increase in Percentage of Online Retail (as Percentage of Total Retail) (2010-15)
- Increase in Internet users (as Percentage of total population) (2010-15)

4.6 Test of Significance

When correlation coefficient is derived, it should be checked for its strength and significance. The significance of correlation coefficient is thus tested by making use of a statistical test. It decides whether the relationship shown in the sample data is strong enough to derive conclusions about the parent population. The test makes use of the correlation coefficient and the sample size.

4.6.1 ‘t-test’

In many research problems, taking a small sample becomes a necessity. In such a situation, in order to test the significance of the sample statistic, a special test for small samples is used. This test is known as ‘t-test’ or ‘student’s t-test’. The concept of ‘t-test’ was given by Sir William Gosset (known by the pen name ‘student’) and it was further developed by Prof. R A Fisher.
4.6.1.1 Conditions necessary for applying t-test

- The sample size should be less than 30 i.e. n < 30.
- The population from which the sample is drawn is ‘normal’.
- The population standard deviation is unknown.

4.6.1.2 Characteristics of t-distribution

- It ranges from \(-\infty\) to \(+\infty\).
- It is symmetrical.
- The mean of the distribution is zero.
- It has a greater dispersion than the standard normal distribution.
- As the sample size approaches 30, the t-distribution, approaches the ‘normal’ distribution.

4.6.1.3 Value of t-statistic

For testing significance of correlation coefficient, the value of t-statistic is calculated as follows:

\[ t = |r| \sqrt{\frac{n-2}{1-r^2}} \]

where

- \( t \) : t statistic
- \( r \) : correlation coefficient
- \( n \) : sample size
4.7 Sampling

In accordance with the research objectives, the sample used for the research work comprises of the G 20 nations. The group includes the following nations:

- Argentina
- Australia
- Brazil
- Canada
- China
- France
- Germany
- India
- Indonesia
- Italy
- Japan
- Mexico
- Russia
- Saudi Arabia
- South Africa
- South Korea
- Turkey
- United Kingdom
- United States
- European Union

In order to fulfill the research objectives and to justify the research work, the sample should be representative of the population. Hence to study the interplay of cloud computing techniques and economic phenomenon, the sample comprises of the G 20 nations.

4.7.1 Sample Characteristics

- G 20 countries represent 90% of the world economy.
- G 20 countries represent 70% of the world population.
- G 20 countries correspond to 80 % of international trade.
- The countries represent different regions/groups of the world – APEC, OECD, BRICS, MENA, Latin America, South Asia, etc.
- The countries include major developed economies of the world.
- The countries include major emerging economies of the world.
- These countries also represent different political systems/regimes.
### 4.7.2 Core Sample Data

The data for reference and analysis has been collected from the following sources:

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Data Category</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>GDP Growth Rate</td>
<td>(&quot;Domestic Product&quot;, 2015), (&quot;Global Economic&quot;, 2015)</td>
</tr>
<tr>
<td>3.</td>
<td>Cloud Computing Jobs</td>
<td>(Columbus, 2014)</td>
</tr>
<tr>
<td>5.</td>
<td>Mobile Phone/Internet Users</td>
<td>(McCarthy, 2013), (&quot;The Data&quot;, 2015)</td>
</tr>
<tr>
<td>S.No.</td>
<td>Data Category</td>
<td>Source</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>SME Sector</td>
<td>Lundmark, O'Day, Pineda &amp; Zwillenberg, 2012)</td>
</tr>
</tbody>
</table>

Table 4.7.2: Core Sample Data Sources
4.7.2.1 Economy Size (GDP Ranking of G 20 Nations)

As per GDP ranking by the World Bank, the G 20 nations are ranked as follows:

<table>
<thead>
<tr>
<th>World Ranking</th>
<th>Country</th>
<th>GDP (millions of US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>United States</td>
<td>1,74,19,000</td>
</tr>
<tr>
<td>2</td>
<td>China</td>
<td>1,03,54,832</td>
</tr>
<tr>
<td>3</td>
<td>Japan</td>
<td>46,01,461</td>
</tr>
<tr>
<td>4</td>
<td>Germany</td>
<td>38,68,291</td>
</tr>
<tr>
<td>5</td>
<td>United Kingdom</td>
<td>29,88,893</td>
</tr>
<tr>
<td>6</td>
<td>France</td>
<td>28,29,192</td>
</tr>
<tr>
<td>7</td>
<td>Brazil</td>
<td>23,46,076</td>
</tr>
<tr>
<td>8</td>
<td>Italy</td>
<td>21,41,161</td>
</tr>
<tr>
<td>9</td>
<td>India</td>
<td>20,48,517</td>
</tr>
<tr>
<td>10</td>
<td>Russia</td>
<td>18,60,598</td>
</tr>
<tr>
<td>11</td>
<td>Canada</td>
<td>17,85,387</td>
</tr>
<tr>
<td>12</td>
<td>Australia</td>
<td>14,54,675</td>
</tr>
<tr>
<td>13</td>
<td>South Korea</td>
<td>14,10,383</td>
</tr>
<tr>
<td>15</td>
<td>Mexico</td>
<td>12,94,690</td>
</tr>
<tr>
<td>16</td>
<td>Indonesia</td>
<td>8,88,538</td>
</tr>
<tr>
<td>18</td>
<td>Turkey</td>
<td>7,98,429</td>
</tr>
<tr>
<td>19</td>
<td>Saudi Arabia</td>
<td>7,46,249</td>
</tr>
<tr>
<td>24</td>
<td>Argentina</td>
<td>5,37,660</td>
</tr>
<tr>
<td>33</td>
<td>South Africa</td>
<td>3,50,085</td>
</tr>
</tbody>
</table>

Table 4.7.2.1: World Bank’s ranking of G 20 Nations based on GDP (2014)
4.7.2.2 GDP per capita (2014) of G 20 Nations

Figure 4.7.2.2: GDP per capita of G 20 nations in 2014

As per GDP per capita, the classification of G 20 nations can be done as:

**High Income Countries:** Australia, United States, Canada, Germany, United Kingdom, France, Japan, Italy, South Korea, Saudi Arabia, Russia

**Middle Income Countries:** Argentina, Brazil, Turkey, Mexico, China, Indonesia, South Africa, India
4.7.2.3 GDP Growth Rate (2014 & 2015) of G 20 Nations

Figure 4.7.2.3: Growth Rate of GDP in G 20 Nations in 2014 & 2015

4.8 Organization of the Thesis

The structure of the present research work can be summarized as follows:

Chapter 1: The thesis starts with introduction of the concept of Cloud Computing and the importance of the economic aspect of cloud computing
technology. It is followed by overview of cloud, and definitions and characteristics of cloud computing. Thereafter, SWOT analysis of cloud computing, relevance of Jevons’ Paradox for the expanding scope and usage of cloud technology, meaning and discussion about the cloud ecosystem, and certain growth indicators of cloud computing have been discussed.

Chapter 2: This chapter contains two sections, devoted to the literature on the technical and economic aspect of the cloud technology respectively. The first section covers the deployment models and service delivery models of the cloud. The second section deals with the economic benefits of cloud computing in the form of economies of scale and scope, cost reduction, business creation effects, effects on business innovation, business development effects and multiplier effects; different types of costs affiliated with the technology, elasticity, employment aspect of cloud computing and associated externalities.

Chapter 3: The third chapter encapsulates the significant cloud computing aspects concerning the economy of India discussing about the significance of cloud computing for Indian economy, role of small and medium enterprises in growth and adoption of cloud computing, Government of India initiatives in this direction and the supply side of cloud computing in reference to India.

Chapter 4: This chapter comprises of the research methodology including the objectives, research design and data sources, methodology, statistical techniques, hypothesis testing, test of significance, sampling including sample characteristics and data, and organization of the thesis.

Chapter 5: This chapter is devoted to data processing and analysis under the heads of diagrammatic presentation and analysis of the macroeconomic variables / cloud based aspects, correlation analysis, hypothesis testing, regression analysis and analysis regarding cloud and economic parameters in reference to India.
Chapter 6: This chapter deals with the deduction and interpretation of results concerning availability of cloud computing jobs, status of internet economy, cloud computing & retail and SME sector.

Chapter 7: The research work culminates with the suggestions based on the research carried out on the cloud computing aspects and the economic aspects included in the study.

Chapter 8: In this chapter, the relevance and limitations of the research work have been stated.

References: This section comprises of the list of various research papers, articles, reports, books, magazines, journals etc. of various organizations, groups and governments referred to in this research work.

Appendices: There are two sections in the appendices. The first section includes different tables related to the data used in the thesis and the second section contains the research papers worked upon by the author of the current thesis.