CHAPTER 4: RESEARCH METHODOLOGY

To examine the proposed hypotheses and the research model, this study needs to collect and analyze relevant data. A questionnaire will act as a survey instrument to collect the relevant data. Such survey instrument may be developed with the help of validated questionnaires that were already used in previous studies. To collect data based on the questionnaire, suitable sample is to be selected. The collected data needs to be analyzed with the suitable analytical techniques. This chapter describes all these stages under the name of research methodology.

4.1 Survey Instrument

The proposed research model consists of 9 constructs and a total of 39 items or indicators. The measurement scales for all the constructs adapted from previous studies are widely recognized and used. The summary of constructs and their measures is given in Table 1 of chapter 3. As per those details a detailed questionnaire has been developed to collect data relevant to the study. To measure the construct of perceived usefulness, the five items of usefulness, efficiency, effectiveness, performance and productivity, as reported by Davis et al. (1989) and Venkatesh & Davis (2000), have been transformed into five meaningful questions. Similarly the six items of perceived ease of use and three items of behavioral intention to use as reported by the above authors have been transformed into the respective questions under the two constructs specified. The three measures of the construct, Job Relevance, as adopted from Venkatesh & Davis (2000) have been converted into three questions. All the specified literature sources are linked to technologies other than cloud computing and hence the measures adopted from such studies have been modified, so as to reflect a cloud computing context.
In the present work, to measure the dependent variable, actual usage, three questions were developed based on the three items, namely, longevity, intensity, and frequency of use, as reported by Davis et al (1989). Similarly, three questions have been developed to measure perceived ubiquity of cloud computing by using the three items of providing communication and network accessibility, anytime-and-anywhere communication and connectivity, and using technology for personal and business purposes, adopted from Kim & Garrison (2009). In the same way, following the research report of Hsu et al. (2014), six items of perceived benefits and seven items of perceived risks have been transformed into same number of meaningful questions. The above scales of measurement were found applicable to cloud computing and they have been adopted without any changes in this study. Following the reports of Premkumar & Roberts (1999) and Kuan & Chau (2001), the three measures of the construct of Perceived Costs (set-up, training and running/maintenance of software and hardware) have been transformed into three relevant questions in this study.

All these 39 developed questions in the form of a survey instrument, used to collect data on cloud computing adoption, are shown in the questionnaire given in the Appendix 1.

4.2 Sample selection

In the present study, there are 39 items included in the measurement instrument and hence, at least 390 responses should be collected to test the hypotheses and maintain 1:10 ratio between an item and respondents as recommended by Hair et al. (2010). The study targeted senior managers (CIO, IT manager and other senior staff) of firms as sample population, who are responsible for
making IT decisions in the organizations and at least two years of experience in using cloud computing. The firms may be adopters or non-adopters of cloud computing. The respondents are selected from the database of a project consultancy company, NIIR (National Institute of Industrial Research), which includes 7448 of SMEs and large firms of India. These firms belong to the sectors of IT, service, manufacturing, finance and telecommunication. The selection of these industries is in accordance with the CIO report (2010) that these industries have high cloud computing adoption rate. The locations of the companies are Hyderabad, Bangalore, Mumbai, Chennai and Delhi. This study used the survey instrument to gather data from these organizations.

4.3 Data analysis techniques

The data collected as per the survey instrument developed and from the sample selected should be processed and analyzed with the help of suitable statistical techniques. This study used two important statistical techniques - Exploratory Factor Analysis (EFA) and Structural Equation Modeling (SEM) for data analysis. SEM comprises two important stages – Confirmatory Factor Analysis (CFA) and Path Analysis. When there is very less a priori knowledge of structural model developed, it is best to conduct EFA before proceeding to CFA (Ruscio & Roche, 2012). So an initial EFA was performed to test the fundamental structure of the factor. Then CFA was performed to examine construct reliability and validity and also to evaluate the goodness of model fit. Finally, path analysis has been applied to examine the proposed hypotheses and also to assess the structural model fit.
4.3.1 Exploratory Factor Analysis (EFA)

Exploratory Factor Analysis (EFA) brings together inter-correlated variables and groups them under some common factor. According to Van Hout & Rietveld (1993), the objective of the factor analysis is dimension reduction, in which the dimensionality of the original space is reduced and new space with fewer dimensions is created. Factor analysis offers not only the option of achieving an understandable view of the data, but also the likelihood of using the output in consequent analyses (Van Hout & Rietveld, 1993). In this study, EFA is used to establish unidimensionality of all the factors by determining the number of factors that are required to explain correlations among the items.

There are different factors extraction methods available and they include Principal Component Analysis (PCA), Maximum Likelihood (ML) and Principal Axis Factor method. PCA is a method used for dimension reduction, usually used when some variables are inter-correlated. Each measured variable is defined as a linear function of principal components containing both common and unique variance; and linear combinations among such variables are determined to retain as much information as possible (Fabrigar et al., 1999). ML method analyzes the maximum likelihood of sampling the observed correlation matrix (Tabachnick & Fidell, 2007) and is more helpful to CFA (Yong & Pearce, 2013). In Principal Axis factor method, all variables will be treated as belonging to the starting group and after its extraction as a factor, a residual matrix is calculated. The successful extraction of factors continues till correlation matrix acquires a large amount of variance (Tucker & MacCallum, 1997). This method is useful when the assumption of multivariate normality is violated by the data (Costello & Osborne, 2005).
PCA can be used as the initial step for data reduction (Yong & Pearce, 2013). This study used PCA as data reduction technique.

In the case of EFA models having factors more than one, a single solution will not be possible and hence, one solution should be selected out of infinite equally fitting solutions (Fabrigar et al., 1999). The factors should be rotated in multi-dimensional space to get a solution with the best simple structure (Thurstone, 1947). Factor rotation helps in removing ambiguity among factors in terms of cross and negative loadings and generating a simple optimal structure of factors, where each variable loads on as few factors as possible (Rummel, 1970). Several analytic rotation methods were reported by several authors like Crawford & Ferguson (1970), Dielman et al. (1972) and Hofmann (1978). Orthogonal rotations lead the factors uncorrelated, whereas oblique rotations allow correlations among factors. Authors like Nunnally (1978) preferred orthogonal rotation due to its simplicity and clarity in concept. Out of several orthogonal rotational methods developed, the Varimax method developed by Kaiser (1958) is recognized as the best and widely used method. This study used orthogonal Varimax rotation method to reduce the number of variables loading highly on each factor.

To find the number of factors to be extracted or retained, there are two criteria which are commonly used - Eigen values and Scree plot. The amount of variance explained by each of the principal component is indicated by Eigen values. Eigen vectors are the weights used to calculate components’ score. Kaiser (1960) suggested the criterion of retaining the factors whose Eigen values are above one. Statistical software generally consists of a default cut off value of 1 for Eigen value. In this study, components with Eigen values over 1 were only retained. The cut off
point of 1 is kept for Eigen value of factors in the model. Sometimes, going by this rule may not result in the best solution and in such case, help may be taken from a scree plot, which can be optionally generated in software like SPSS. A Scree test consists of Eigenvalues and factors (Cattell, 1978) and is reliable when there is a minimum sample size of 200 (Yong & Pearce, 2013). A scree plot always shows a downward curve. In a scree plot, the Eigen values are plotted on y-axis and number of scores on the x-axis. At some point in the plot, the slope of the curve can be seen to level off; this part is known as the elbow. This elbow indicates the number of factors to extract.

In addition, there is a need to maintain cut-off for a factor loading. According to Stevens (2002), when the sample size is larger, smaller loadings are acceptable for a factor to be considered significant. Tabachnick & Fidell (2007) suggested a rule of thumb for a factor loading to be at least 0.32 for a sample size of 300. In this study, it was assumed that all items were uncorrelated with one another and items with loadings less than 0.5 were not retained. Therefore, the factors with loadings equal to or above 0.5 are only considered. In this study SPSS version 20 is used to conduct EFA.

4.3.2 Structural Equation Modeling (SEM)

SEM is a multivariate statistical method used to test direct and indirect associations between a predictor variable with one or more outcome variables. It examines the structure of interrelationships among multiple variables in terms of a series of multiple regression equations. Perhaps the most important strength of SEM is that such relationships can be examined in a way that reduces the error in the model (Hair et al., 2010). A structural equation model is composed
of two types of models, namely, measurement model and structural model. A measurement model represents the relationships among observed and unobserved variables. A structural model represents the relationships among unobserved variables (constructs) only. In SEM, latent variables are free of random error, with common variance.

CFA addresses the measurement model, whereas path analysis addresses the structural model. The regression or path coefficients represent the relationships between the theoretical constructs. Structural equation models are shown graphically by a path diagram. Path analysis tests the relationships among latent variables. It shows the relationships that exist between constructs according to theory and each of these relationships or paths represent a hypothesis included in the study.

The Research model is depicted as Structural Equation Model in Analysis of Moment Structure (AMOS) software version 20 and the hypotheses are tested with the help of derived path coefficients.

### 4.3.2.1 Confirmatory Factor Analysis (CFA)

CFA evaluates and establishes evidence of sufficient construct validity. It also tests whether measures of research constructs are consistent or fit the validity of the research model (Creswell, 2003). As a special type of factor analysis, CFA is an initial step for a complete test of a structural model (Hair et al., 2010). It consists of a sequence of various stages of analysis. Those stages include establishment of construct validity, convergent and discriminant validity and goodness of fit.
Construct validity ensures the suitability of inferences made on the basis of observations to check whether a test measures the intended construct. Campbell & Fiske (1959) proposed two aspects to assess the construct validity - Convergent validity and Discriminant validity. Under Convergent validity, the items that are indicators of a particular construct should converge or share high proportion of variance in common. It will be established by construct reliability and average variance extracted (AVE). Discriminant validity is the extent of true distinction of one construct from others. It can be assessed by checking whether the square root of AVE is greater than the correlation coefficient between the discussed construct and other constructs (Fornell and Larcker, 1981). The higher the discriminant validity, the higher will be the evidence that a construct is unique. Chin (1998) reported that the constructs having AVE loading greater than 0.5 establish discriminant validity.

There are mainly three important categories of goodness of fit indices – absolute, incremental and parsimonious. According to McDonald & Ho (2002), absolute fit indices test how good the apriori model fits, or reproduces data. They include chi-square ($\chi^2$), GFI, AGFI, RMR, and SRMR. $\chi^2$ value is a measure to evaluate overall model fit and is a function of the sample size and the difference between observed covariance and model covariance matrices. It is called CMIN in AMOS. The degrees of freedom (df) is the quantity of mathematical information existing for estimating model parameters. Hair et al. (2010) reported the values of relative chi-square ($\chi^2$/df) between 1 and 3 as representing good model fit. The test of $\chi^2$/df is used to examine whether distribution of categorical variables differ from one another.
According to Byrne (2010), the incremental fit indices measures the balanced enhancement in fit by matching the default model with a more restricted baseline or saturated model. The examples of incremental fit indices are normed fit index (NFI), comparative fit index (CFI), tucker lewis index (TLI), incremental fit index (IFI) and non-normed fit index (NNFI). Hu & Bentler (1999) reported a value of 0.90 or above for NFI, CFI, TLI, IFI and NNFI as indicating acceptable model fit. As relative fit indices, parsimony indices are alteration to most of the fit indices and serve as a criterion for choosing between alternative models. Their examples include PGFI, PNFI PNFI2 and PCFI with acceptable values above 0.90.

CFI denotes the ratio of inconsistency of target model to the inconsistency of independence model. Generally, CFI indicates the degree to which the concern model is better than independence model. Values close to 1 indicate acceptable fit. IFI is the ratio of difference between the $\chi^2$ of independent and target models and the difference between $\chi^2$ of target model and df for target model. The ratio of these values indicates the IFI. Values above 0.90 are regarded as acceptable. NFI is equal to the difference between $\chi^2$ of target and null model and divided by $\chi^2$ of null model. Values above 0.90 are considered as acceptable. TLI is the ratio of difference between $\chi^2$/df for the target and null models and $\chi^2$/df for null model minus one. Values above 0.90 or 0.95 are acceptable (Hu & Bentler, 1999).

Goodness of fit index (GFI) is a measure of fit between hypothesized model and observed covariance matrix. The adjusted GFI (AGFI) is the result of adjusting GFI by a ratio of degrees of freedom used in a model to the total degrees of freedom available. It is affected by the number of indicators of each latent variable by correcting GFI. According to Baumgartner, H., &
Hombur, C. (1996) and Hair et al. (2010), values between 0 and 1 are acceptable for GFI and AGFI, with a value of over 0.9 generally indicating a good model fit.

Root Mean Square Error of Approximation (RMSEA) denotes how well a model fits a population. A Lower RMSEA value indicates good fit and values ranging from 0.05 to 0.08 are acceptable (Hair et al., 2010). RMR is equal to the square root of average of covariance residuals. Zero denotes a perfect fit, value below 0.08 is acceptable (Browne & Cudeck, 1993; Hu & Bentler, 1999) and a value below 0.05 is ideal and acceptable (Stieger, 1990). For reporting model fit indices, Hu & Bentler (1999) recommended CFI as the index of choice, whereas MacCallum & Austin (2000) have strongly recommended RMSEA to be the index of choice.

The confirmatory factor models are presented as path diagrams, where the observed variables are represented by squares, latent variables by ellipses and errors in correlating variables by circles to the respective constructs. The causal relationship among the variables is shown by single-headed arrows whereas covariance between two latent variables is shown by double-headed.

In the present study, CFA has been used to test the consistency of measures of the constructs. The objective of conducting a CFA is to test whether the data fits a hypothesized measurement model. CFA is done using AMOS 20.0. The estimation procedure used for CFA is maximum likelihood estimation (MLE). The MLE is one of the widely used procedures for CFA in which the value of model parameters that maximize the likelihood function is selected. This allows for maximum agreement of model fit with measured data. Additionally, in CFA, significance of
standard regression weights or estimates shows that indicated variables are representative of their corresponding latent constructs.

### 4.3.2.2 Path analysis

Path diagrams enable researchers depict explicitly the hypothesized set of relationships, represented by the model, needed for the analysis and hence they are fundamental to SEM (Ullman, 2001). Path analysis is used to analyze causal relationships among two or more variables and estimate a set of simultaneous regression equations. According to Ullman (2001), path analysis allows researchers to examine relationships among independent or dependent variables, either continuous or discrete in both cases. It deals with only measured or observed variables without error.

In the present study, the software AMOS version 20 was used for both CFA and path analysis under SEM.

### 4.4 Data collection

Emails were sent to the senior managers to know whether they have experience in using cloud computing and how long their firm has been using cloud computing. They were also asked to express their interest to fill the questionnaire on cloud computing. The questionnaire for the survey is represented in Appendix 2. Initially, the questionnaire was administered through the web links provided by Google forms and www.surveymonkey.com. Around 71 responses were obtained from these web links. 25 responses were collected from personal visits to the
companies, namely TCS, CA technologies, Franklin Templeton, NIIT, TSBCL (Telangana State Beverages Corporation Limited), Wipro and Delloite etc. Total 96 responses received from web links Google forms (36) and Surveymonky.com (35) and personal visit (25).

Additional data were collected through the Hyderabad based research agency, namely, MarQ Research Solutions. The research agency was instructed to collect the data from the senior managers of the firm having at least two years of experience in using cloud computing. The agency was also instructed to note either email or mobile number of the respondents as a mandatory requirement. The agency hired 5 people to collect data with the questionnaire. As reported by the Agency, the average time taken by a respondent for filling the questionnaire was 15 minutes. To cross check whether the data collected by the Agency were genuine and to verify that the questionnaire was indeed filled by the proper person, 50 responses were randomly selected and contacted over mobile phone, as well as through email. The total duration of data collection was 5 months and 24 days. With the responses collected by the Agency and those collected from web links and personal visits made the total responses as 550. Out of 550 filled questionnaires, 12 cases were incomplete and hence they were removed. So the final sample size came down to 538.

Out of these 538 responses, 69 respondents provided only their email-id without contact numbers and 6 respondents provided only their contact numbers without email-id. Whereas, 463 respondents provided both email ids and contact numbers. All these responses were collected from metropolitan cities of India, exclusively from Hyderabad (486). The responses from non-adopter firms belong to Hyderabad city (120) only, whereas the responses from adopter firms
cover not only Hyderabad (366), but also the cities of Bangalore (42), Delhi (6), Chennai (2) and Mumbai (2).

The data analysis techniques as explained in section 4.3 have been applied to the data collected for detailed analysis. Chapter 5 presents and discusses the results of data analysis in detail.