CHAPTER – 1

CLOUD COMPUTING APPROACH

1.1 Admittance to Cloud Computing Paradigm

The involvements of HTC and HPC metrics shows cloud computing stands on the base line of utility computing. The ensemble techniques and approaches of distributed computing can also be feasibly adopted in cloud computing platform. Cloud can be categorized as three types where the private cloud is purposefully oriented for an organization in which all the administrative tasks are within the administrative boundaries. Public cloud can be termed as utility model for providing of services with on demand basis. Hybrid cloud aggregates with the composite features of private and public cloud where multi tenant service utilizations are initiated with domain based rules. The adoption of cloud exhibits the advantages from two perspectives, from provider's perspective utilization of partially used resources and minute initial capital investment and with respect to consumer perspective low maintenance for hardware and software, better illusion of virtualization and scalability of resources.

The cloud paradigm admits the features of responding to the real time resource utilizations with a high availability rate on demand. Cloud computing obeys batch processing task efficiently with parallel operational analyses. The cloud paradigm emphasised with an internet medium where the applications and data can be used with the salient features of enabled virtualization. The distinguished features of cloud computing with related to grid computing are classified in managing of resources, need of virtualization, security issues, application model and business model. In the context of comparing cloud computing to the grid cloud enables flexible pay and use model, simultaneous resource sharing, providing of encapsulation with abstraction towards virtualization. The cloud service tag is mainly represented with three service models one is SaaS, examples are Google apps and salesforece.com second is PaaS, examples are Google app engine and windows azure, finally third is IaaS, examples are virtualization, Amazon EC2and S3. In notifying the advantages of cloud applications Amazon EC2 for IaaS service plays a pivot role in deploying the ZEN architecture. This Amazon EC2 uses S3 service for storage tasks. Google app engine

is mainly oriented on web application that deploys Google architecture facilities supported by java and python languages. The application from these engines follows requesting and replying approach. Windows azure is a feasible .net language library which is flexible for PaaS. The applications from windows azure can be easily deployed in Microsoft data centres. This windows azure exhibits computing service, storage facilities and fabric controlled service. The computing service supports multiple users with multiple applications and their roles are categorized as web role instance, working role instance. The storage service in windows azure is distinguished with storing of binary data, storing of unstructured data as an organized storage tables and promoting the message in queue structure. The outages in cloud computing exposes scalability, availability, reliability, power consumptions, data logs, data recovery, policy monitoring, and security concerns.[1]

The abnormal changes in computational scenario make the IT enterprise step into the outsourcing management which transforms into cloud services. The raise in technology platforms in the context of HTC and HPC produces the adoptability features of cloud in a scalable manner. The outsourcing orientation is an initiation task for globalization and win in the competitive computational markets. The achievement of outsourcing outcomes with enabling of lower cost spends for the infrastructure, instant development approach, quality assurance, performance metrics and service deploying and acquiring from different geographical locations. The cloud computing enables hardware and software services managed in the network based on pay as you use. Generally cloud computing show case different service models like SaaS, PaaS, and IaaS where SaaS is the highest abstraction level tagged with WWW service with heterogeneous applications that can be used at ERP and supply chain approaches.

PaaS is a middle level abstraction used for computational connectivity, accessing and controlling process. Operating systems play as a ideal role to justify for this service. IAAS is a hardware layer service which is enabled with heterogeneity of resources like storage, memory etc., which can be utilized according to the preferences with the concept of virtualization. The discussion of cost benefit approaches, performance related tasks can be accomplished with economic strategies which deal with utilization of resources in many to many and one to many fashion.

As the cloud computing enable virtualization a balanced infrastructure support may be established feasibly by the demand of the consumers which exhibits the salient feature of elasticity. An ideal prominent task enabled in cloud computing is total cost minimization and acquiring of ownership privileges in a shared infrastructure. Cloud computing with its sophisticated approaches mingle into different business scenarios where the business with limited budget, maximizing customer satisfaction levels with innovative strategies, complex IT operations with high capital expenditure, risk in handling of heavy operations and monitoring tasks in automatic failure systems. The adoption of cloud computing was globally accepted practice in most of the IT related computational sectors as a value added service. The provocation issues in cloud computing address privacy, security, availability, performance, data integrity, monitoring compliance and standards maintenance. [2]

1.2 Cloud Computing Beneficences

Adopting cloud computing addresses many advantages ie., cost optimization of resource utilization, security, privacy, mobility flexibilities, improved quality, preventing of data loss that occurred through risks, updating software automatically etc.

1.2.1 Software as Service: The information related to customers and companies is now stored into cloud and there where the concern towards the security of data stored in cloud raised. Every organization understood the reality that by just entering into the cloud they can increase their infrastructure resources. Cloud data security became the top challenge which reduces cloud adoption. Monitoring and managing data in cloud are not trustable which raised to many problems like accessibility issues, virtualization issues, privacy issues, caused by third party, integrity issues, data leak etc,. the following security challenges are identified to be very specific and they are third party resources security, data security, security on data transmission and application security. SaaS, PaaS and IaaS are the service delivery models which are used by the cloud to provide services to the end users. These service delivery models are arranged one above the other starting from bottom infrastructure as a service (IaaS) which is built on with platform as a service (PaaS) PaaS is built on with software as a service (SaaS).

All the three services have their individual security issues. SaaS enables the customer's level of accessing applications which are posted remotely by the service provider over the Internet. In the debate of SaaS security features the origin of authorization and authentication of services should be carefully embossed on the data sharing issues. The considerable parameters for data security in SaaS consider the access of data with segregation, integration and data locality. Most of the SaaS services enables web applications which provocative data confidentiality. The adoption of SaaS services by the consumer judges the availability rate and backup features. The concept of securing data in SaaS for small enterprises highlight the issues like access control weak points, weak storage, improper configurations and cookie manipulations. The SaaS services takes the internet as a carrier medium for service delivery where this service level should have validating check points about session management, network packet analysis and SSL configurations.

The SaaS model should maintain the data privacy issues to its check bounds and provides reliability to the customer satisfaction level. The task of data integrity in SaaS services represents database acid properties to deploy data integrity. The data segregation issue in SaaS approach is ensured with different segregations to avoid malicious modification at applicative level. The SaaS data segregation issues should be enriched with finding of faults in an SQL injection and validation of data. The data confidentiality issues in SaaS elevate privacy and personal information, location information, and legal uncertainties. As the SaaS services enables web as a major carrier medium that triggers security issues are application service at forty percent rate, known vulnerabilities as eighteen percent, operating service issue as twenty four percent etc. The administrative task in SaaS approach is identity management which can be differentiated as pure identity approach, user access approach, and service approach. Broadly identity management was segmented into independent, credential and federated. Each of these segmented tasks has their own pros and cons related to security. [3]

1.2.2 About Provenance: Provenance can be used for heterogeneous applications which show the importance of focusing on manipulated data. Provenance stores crucial information related to scientific complex data. The interest on desired data and its changes with an event of actions are captured with granularity approach as

coarse grained and fine grained. The content of the provenance and its execution behaviour can be considered by the workflow figures. The information of provenance can also be tracked by automated and non automated process and identify the information in between different layers. In the maintenance of provenance tracking of input data with a maximum authentication level and interoperability levels are validated. The system issue of provenance considers storage level. The usage of provenance termed with interactions involved by the user retrieving levels of data with queries and fault finding. [4]

In finding the provenance data links web based approaches plays a pivot role where the interoperable issues are show case frequently. The recognition of linking provenance data impacts on reproducibility of results that targets the interoperability issues. The interoperability issue challenges deals with artifacts, processes and agents. As these artifacts are generated by process, the agenda of constituting open provenance model deals relational database approaches shows their importance. Many examples on provenance data infrastructure, OPM extensions, provenance data querying are discussed to extract solution. The issue of interoperability in the content of data linking intensifies the identification challenges and infrastructure challenges. The best optimal solution strategies can be practised to solve web based provenance linking that triggers provenance data furnishing, linking data in the provenance for furnishing, enabling graph models and data centric computations. [5]

1.2.3 Trust in Provenance: Cloud elevates with the tag line of pay for the utilization of resources which relied on two factors trust and reputation. The cloud provider, cloud consumer, cloud broker and cloud auditor has to show case trust in delivering of service, utilization of service and in monitoring of quality of service. The cloud computing is segmented into three parts like services, deployment models and cloud stake holders (cloud provider, cloud consumer, cloud broker and cloud auditor and cloud carrier as per NIST). In the process of adopting cloud computing many uncertainties and many misconceptions are questioned. The optimal solutions for most of these uncertainties is establishing of trust which solves the hindrances between cloud stake holders. The SLA plays a pivot role between cloud stake holders for triggering truthful and trust worthy information.

The cloud computing enables trust facilitator by pinpointing many communication entities between stake holders which yield an economic growth. The scenario of reputation depends on empirical experiences, events and actions involved between stake holders as per their initial agreement. The trust can be established by four operational tasks using SLAs, the commitments and initiation standards can be specified at each level of services. Audits ensure lower security concerns which are correlated in between the stake holders. Metric assessment accomplished with customer behaviours, feedback and comparisons. Incorporating self assessment delivers the idea of competencies between different stake holders and discusses the capabilities and standards where the stake holders generate. In fulfilling of trust in cloud management services SLA's provide a bench mark information. Compliance is an auditor based standard, portability and interoperability depended on security and performance approaches. The need of trust may accompany in multiple contents and multiple criteria objectives. The trust can be evaluated with interior approach, exterior approach and user feedback. [6]

The mechanisms involved for the data control in cloud environment is a hurdle task to retrieve sensitive data by the consumer. The part of trust relationships play a major role in cloud environment to specify data control by empowering audit ability and encryptions which benefits to the data ownership. The inability of data owners raises the evidential questions to stream line cyber security attacks. Many realistic incidents which are prone to cyber securities breach situations like migrating mobile camera to public cloud, un authorized undertaking of G-Talk services. The model STRATUS resembles security technology returns accountability, trust and user centric service which can be comprised with four segments auditing of data with transparency , processing and storing of data with privacy concerns, rapid detection of malicious actions, rapid recovery from un known events with resilience. The proposed STRATUS approach examines the above discussed four segments with an appropriate projections which intensifies the protection levels for the above discussed data control problems in cloud [7]

1.2.4 Cloud monitoring: Effective and efficient monitoring states to be mandatory for managing huge and complex infrastructures and all the current mechanisms are in lag regarding detailed analysis on cloud monitoring. NIST and cloud community

stated the following a) essential characteristics: on demand self service, broad network access, resource pooling, rapid elasticity, measured service. b) Service models: based on the service type there are three models infrastructure as service (IAAS), platform as a service (PAAS), and software as a service (SaaS). c) There are two types of hosting internal and external d) deployment models: based on the cloud location they are categorized as private cloud, public cloud, community cloud and hybrid cloud. e) Roles: cloud developers support various roles like cloud auditor, cloud service provider, cloud service carrier, cloud service broker, cloud service consumer. Challenges in cloud computing can be a) provision replace of scalability, load balancing and quality of service b) provision and guarantee of SLA's c) managing complex infrastructures d) performance analysis. To overcome these challenges fine grain monitoring and performance measuring techniques are required.

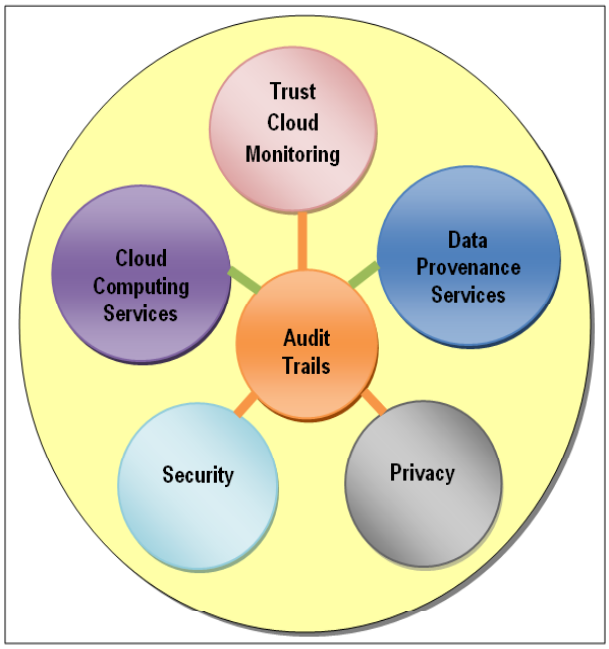

Fig 1.2.4.1 Conceptual view of Audit trails Outcomes

The above diagram gives various perspectives that address the key issues of audit trails. The audit trail is encompassed with provocation parameters like trust, security and privacy where it utilizes the cloud computing services and data provenance services as inputs. The enrichment of Audit trails for performance audit reveals monitoring of economy, effectively and efficiency of services. An audit trail is that which outcomes an independent transparent assessment of services in cloud

environment. Most of the audit trails are constructive in cloud environment which showcase trust, privacy, security and quality attributes assessment.

Need of monitoring enables with continuous monitoring of cloud and SLA which enhance performance monitoring and quality of service offerings between consumer and provider. Cloud auditor is the one who takes care of monitoring through continuous auditing. Based on the SLA specifications, cloud service provider guarantees quality of service offerings. It is very essential to monitor in fine grain manner as the virtual resources migrate between the virtual machines and also due to dynamic network conditions. To certify SLA compliance while auditing cloud monitoring is the most essential task. Monitoring plays a key role for both consumer and provider with respect to provider, monitoring helps to do billing and with respect to consumer monitoring helps to compare different service providers and also to track their own usage. The monitoring scenario of cloud computing is distinguished with the service levels and their significant utilization of resources.

Many monitoring plug-ins in cloud computing elevates the fault issues and performance issues. The utilization of resources, availability of resources considers consumers heterogeneous preferences related to the improving of performance management. The hosting services of the service provider enriched with security patterns spread in service utilizations with strict conditional monitoring approaches that elevates the security topology involvement. The identified active pieces with applicative functionalities of cloud concepts are SLA's, availability, reliability etc. are discussed in many conceptual levels of cloud layers. The major stake holders in cloud platform addresses data centres and their physical network topologies, linking patterns in between clouds, operational strategy of hardware, physical and virtual machine in operating system mode, exhibiting the prominent importance about the middle ware in terms of PASS and SaaS service model and finally applicative user correlated tasks are examined. The earlier cloud monitoring strategies encompassed with system based and guest based metrics which evaluates internal and external complexities of virtual machines for communication.

The monitoring levels are further notices as high and low depending on the utilization of resources. Monitoring at high level deals with the context of

virtualization where as monitoring at low level deals with hardware, power consumption etc. In the part of testing metric in cloud platform is categorized into computations related through throughput and network. Computation related test metric involves with concurrent use of resources, input and output, memory. This type of metrics can also be carried with the association of third parties (auditors). Network based metrics deals with bandwidth, congestion in the network, jitter functions etc. Cloud computing differs with grid and cluster paradigms where the complexities in the cloud platform exhibits more compared to grid and cluster. The transparency in different services of cloud platform maintains restricted relationship for monitoring dynamic utilization of resources at heterogeneous levels. The design architectures which are scalable in grid and cluster platforms are adopted in cloud platform to enhance the critical characteristics of elasticity and adoptability.

The properties of scalability and elasticity are resembled for cloud monitoring demonstrates the salient features of scalability with metric aggregation and filtering. The elasticity nature in cloud monitoring perspective deals with prominent issues of availability, reliability, resilience factors to track the resources and their migration levels. The property of adoptability and timelines are the paramount issues in cloud monitoring discussions deals with bulk amount of data processing, inclusion of sampling intervals and number of resources monitored. The peculiar property automaticity tagged with cloud monitoring sequence deploys triggering situation alertness and controlling the conditional looping procedures. The advance properties involved in cloud monitoring process are range of extensibility, intrusiveness and comprehensiveness. These properties undertake monitoring of different resources, trouble shooting natures occurred in resources and isolation of computations. The diplomatic property embedded with cloud monitoring is accuracy. This property deals with workload activities and data centres, metric procedures involved in virtualized environment.

The cloud monitoring platforms are diversified as commercial and open source. In the line of commercial based cloud monitoring platforms are like ANEKA, which was constituted with scalable middleware framework for application execution monitoring. It supports heterogeneity of clouds which outcomes the elasticity and scalability features. AZURE WATCH which is a user customized metric basing on

available data in the website and readily outcomes the property of adoptability and extensibility. CLOUD WATCH tool relies on virtual platform monitoring and compose a statistical indications about the behaviour of the services and its event behaviours. This tool outcome qualified metric timelines in different cloud layers. CLOUD KICK – rack space authorization involves in utilization rate of CPU in multiple cloud anatomies which can be adoptable for real-time strategies. CLOUD STATUS integrated with different web and app engines which are helpful for consumer utilization of applications to study the method of performance and its incurred faults. GROUND WORK this monitoring plat form is exposed with different heterogeneous plug-ins in a virtualization pattern of data centre. The intensity of virtualization can be monitored in multi user's strategy for minimizing of cloud risks. LOGIC MONITORING this monitoring is more adoptable for IAAS service which show case the accountable information acquired resources, provision resources and deleted resources. A dash board environment for more number of virtualized platforms can be coordinated for highlighting comprehensiveness property. MONITIS this is more flexible for cloud consumer about the resource utilization and performance analysis. NIMSOFT this provides the dash board for different cloud monitoring with the aware of SLA in different data centres which exhibits the property of scalability. In the line of open source cloud monitoring platforms are like CLOUD STACK ZEN PACK is a program scripted in java which can be feasibly adopted for huge virtual networks concluded with zenoss extensions called ZENPACK. This tool can be easily enabled for different physical devises which undergoes virtual platforms.

DARGOS this tool show the prominent resource monitoring approach in physical and virtual strategies of cloud. The two main components are monitoring node and supervising node, the monitoring node acquires the information of different resources and their utilization levels. Supervisor node takes the input of monitoring node data and sampling with time which in terms elevates the feature of intrusiveness. HYPEREC-HQ this tools was developed with the help of java script which supports for different operating system platforms for analyzing resource utilizations, service levels, and operational strategies. NIMBUS is a pack of different tools for configurations and monitoring which is very compatible for IAAS. Pull methods and

Push methods are the salient features available which outcomes the autonomic feature. NAGIOS tools supports for Amazon services and demonstrates the standardization for Ubuntu in IAAS. This tool elevates the feature of extensibility to its best.

PCMONS this targets the private cloud where the tool comprised with information gathering about a particular node involved in virtualization. This tool contains integrator module per data cluster and integrator module for data monitoring. This tool consists of BM monitor and configuration initiations which rely on NAGIOS interface. A server is dedicated for monitoring strategy of the above discussed events. SENSU tool is an integrated pack oriented for message communication with middle wares and monitoring services. A web based dashboard was enabled with querying protocol features. The addressing factors of cloud performance and their services evaluated with dependability nature can be validated with different tools. CLOUD SLEUTH tools integrates the salient features of reliability and timeliness in a web based scenario as the services are distributed to different geographical locations, this is a parallel feasible platform for IAAS and SaaS providers in evaluation of different response times. CLOUD HARMONY is mainly public cloud oriented which exhibits with different bench mark information related to operating system, memory, input output, accessing of file systems and system overhead management etc.

CLOUD STONE facilitates for web 2.0 applications and feasible for IAAS strategies. The workload strategies of different resource utilizations bounded with markov chains which mainly focus on quality of service. CLOUD CMP introduces by Duke University in coordination with Microsoft in considering cost benefit approach towards provider. This performance tool endeavours with intra cloud facilities based on cloud instances and client executions. CLOUD CLIMATE the depiction of different graphs on test runs with a web based approach for IAAS services are admitted generally. The major feature provided by this tool is resilience. Cloud monitoring scenario clearly discuss in terms of testing metrics, salient features of cloud properties, cloud monitoring in commercial platforms, cloud monitoring in open source platforms and cloud performance in service accessing. The cloud monitoring approach admits the importance of effectiveness and efficiency by cross

examining monitoring tools, architectures which outcomes the cost economic cloud model. [8]

1.3 Cloud Computing Collaborators Delineation Outline

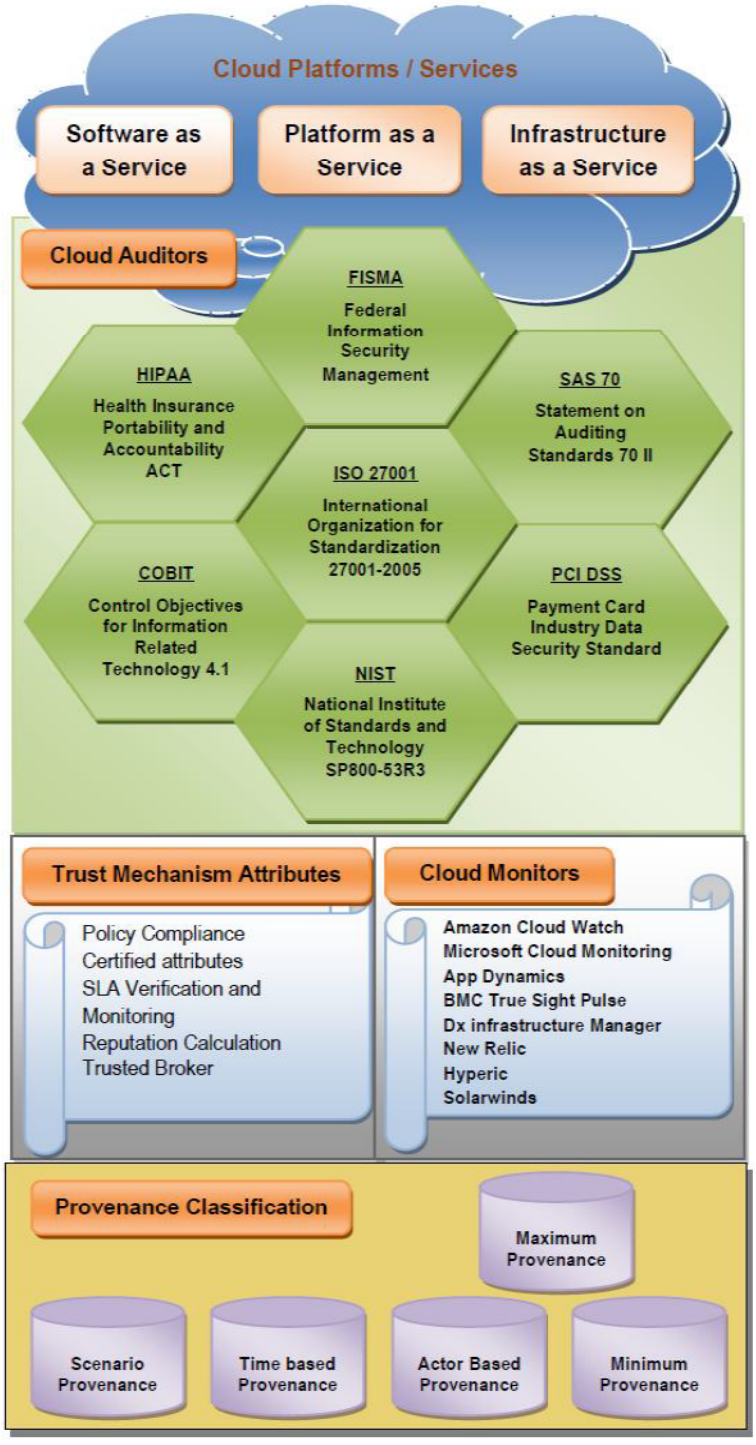

Fig 1.3.1 Cloud Computing Colloborators Outline

SaaS Provider	
Billing	Aria systems, evapt, OpSource, Redi2,Zuora
Financials	Concur, Xero, Workday, Beam4d
Legal	DirectLaw, Advologix, filos, Sertifi
Sales	Xactly, LucidEra, StreetSmarts, Success Metrics
Desktop Productivity	Zoho, IBM Lotus Live, Google Apps, Hyper Office, Microsoft Live, Cluster Seven
Human Resources	Taleo, Workday, iCIMS
Content Management	Clickability, SpringCM, Crown Point
Collaboration	Box.net, DropBox, Zoho
CRM	Zoho, NetSuite, Parature, Responsys, RightNow, Salesforce.com, LiveOps, MSDynamics, Oracle on Demand
Document Management	NetDocuments, questys, DocLanding, Aconex, Xythos, Knowledge, TreeLive, SpringCM
Backup & Recovery	JungleDisk, Mozy, Zmanda Cloud Backup, Open RSM, Syncplicity
Social Networks	Ning, Zembly Amitive
PaaS Provider	
General Purpose	Force.com, Etelos, Long Jump AppJet, Rollbase, Bungee Labs Connect, Google App engine, Engine Yard, Caspio Qrimp, MS Azure Services Platform, The rackspace Cloud sites
Intelligence Business	Aster DB, Quantivo, Cloud Analytics, Blink Logic, K2 Analytics, Log XML, Oco, Panorama, Pivot Link, Clario Analytics, Cold Light Neuron, Infobright, Vertica
Databases	Google BigTable, amazon Simple DB, fathom DB, Microsoft SDS
Integration	Amazon SQS, MuleSource Mule, On demand, Boomi, snapLogic, OpSource Connect, CastIron, Microsoft Biztalk Services, gnip, SnapLogic SaaS Solution Packs, Applan anywhere, HubSpan, Informatica On-Demand
Development and Testing	Keynote Systems, Mercury, SOASTA, SkyTap, Aptana, LoadStorm, Collabnet, Dynamsoft
IaaS Provider	
Storage	AmazonS3, Zetta, CTERA Portal, The rackspace cloud files, Nirvanix
Compute	Amazon EC2, Serve Path GoGrid, Elastra, The rackspace, Cloud Servers, Joyent Accelerators, AppNexus,Flexiscale, Elastichosts, hosting.com, Cloud Nine, Terremark, Grid Layer, iTriCiTY, Layered Tech, enomaly ECP
Services Management	Scalr, CohesiveFT, Ylastic, dynect, CloudFoundry, NewRelic, Cloud42

Table 1.3(a) Chart of SaaS,PaaS,IaaS Provider with different domains

Cloud Service Brokers	
IaaS	Right Scale, enStratus, Kaavo, Elastra, CloudKick, CloudSwitch
NaaS (Network as a Service)	Limelight Networks, Akamai, at&t, BT, Telstra, Verizon, Vodafone
IPMaaS (Identity and Policy Management as a Service)	PingIdentity, VMware Horizon, Symplified, Vordel
DaaS(Desktop as a Service)	StrikeIron, Datadirect
Consultancy Brokers	Blueworlf, Miles consulting, Cloud Avenue, 8KMiles, Booz, Allen, Hamilton, THINK strategies, Burstorm
SaaS	AppDirect, Jamcracker, Mercado Enterprise,Parallels Automation,Snow Commander,IBM Cloud Broker,Compute Next,Gravitant cloudmatrix,RackNap,DXC Agility Platform,AWS Service Catalog,StratoZone,Pax8

Table 1.3(b) Chart of Cloud Service Brokers with different services

In describing Fig 1.3.1 cloud computing colloborators are delineated to our problem boundaries and show case the outline information of different cloud platform services. we know that SaaS is meant for application service provisioning, PaaS is generally used for middle ware architectures and virtual platforms, IaaS is meant for physical service management architectures which can be used by different domains of PaaS and SaaS architectures. A bunch of cloud auditors who do assessments by following standards and guidelines for the specific domain enterprises. There are HIPAA, COBIT,SAS70, PCIDSS, NIST, FISMA and so on.

These cloud auditors are tagged with trust mechanism attributes for the evaluation of trust in different architectures which are sapplication service provisioning, PaaS is generally used for middle ware architectures and virtual platforms, IaaS is meant for physical service management architectures which can be used by different domains of PaaS and SaaS architectures. A bunch of cloud auditors who do assessments by following standards and guidelines for the specific domain enterprises. There are HIPAA, COBIT,SAS70, PCIDSS, NIST, FISMA and so on. These cloud auditors are tagged with trust mechanism attributes for the evaluation of trust in different architectures which are synchronized to cloud monitors for different compliance analysis. Trust relies on auditor assessment tasks in acqiring of data, querying of data and storing of data. The elevation of cloud auditor assessments undertake the provenance data which contains meta data and log records.

In the elevation of assessment cloud auditor can take a partnership with provenance media. Provenance can be catageroized into many forms which can be modulated for that active specific applications. In Table 1.3(a) SaaS provider, PaaS provider, and IaaS provider are exampled with some specific domains reference to that specific applications, tools and softwares. In Table 1.3(b) Different cloud services canbe arbitrage by Cloud service brokers with some available tools, services, protocols and softwares. From Table Char 1 and 2 lucid advantages are outlines to that particular providers and brokers.

1.4 Provenance Challenging issues in Cloud Computing:

The adoption of Provenance to the current technologies of Cloud environment exposes some challenges in the concept of data workflows, privacy issues, and security issues. We know provenance hold the metadata information which cannot be modified and the querying, analyzing and monitoring of data in provenance play a vital role in triggering of provocation issues of provenance.

1.4.1 Workflow issues in Provenance: In the scenario of open provenance model trident approach shares the major role in managing of workflows in a provenance. As trident workflow is a benchmarking platform as it derives the history of data products with verification strategy. The workflows of the provenance consider by the trident composer with the help of dot net activity enabling. The flow of data with dependencies control flow can be compared to the previous control flow activities. The trident bench mark hosted at HPC clusters, remote machines and local machines basing on elements generated by work flow engine activities. The registry of this trident bench mark can be shared which consists of workflow logs, libraries, external data resources etc.

The collection of provenance was carried in this trident benchmark which can be used for monitoring infrastructure services. The native provenance collection approach records the memory events, data flows, work flows and notifications parallels stored in the registry. The Meta data of native provenance collected by trident is used for comparing the OPM open provenance model associated with different approaches like integrating with tightly coupled provenance data, integrated storage and querying which is loosely coupled. This trident benchmark expresses that

attribute of maintaining data consistency, storage overhead performance and query performance with an interoperability features. [9]

Provenance maintains Meta data where we can derive the history of workflows which can be interpreted for diagnosing changes in relational databases. In the part of showcasing the provenance work flows at storage levels an entity relationship diagram was proposed on data base schema. The identification different identifiers source, destination, key attributes, role of attributes, relationship type can be obtained from this ER based scenario. The evaluation of above discussed ER points makes us burden less in dealing with complex situations. The agenda of open provenance model deals with the above discussed techniques. The insertion of data into the database in a relational spectrum defines the time complexity, data mapping etc. Most of the relational database management approaches deals with the data retrieval based on SQL. Reasoning and Querying based on rules and set of logic methods helps in retrieving of provenance data in an optimal manner. [10]

A framework with an integrated provenance approach to solve the work flows in the distributed environment evaluates with many complexities. The outline of data quality and maintaining the reliability are the key points for the integration of provenance. The workflow in the distributed environment in the concern of provenance is a primary task in identification of user roles. The approach of workflow monitoring in an integrated environment exhibits the highest level of interactions. The distinguishing of internal and external provenance says that internal provenance conforms the workflow instances and relationships where as external provenance shows the input data and its workflows. A provenance indexed service was accomplished on internal and external provenance to abolish the complexities involved in redundancy of data objects. The architecture of integrated framework in service oriented distributed architectures tracks the provenance with scalable level. The provenance indexed service solves the challenges by making independent provenance services, recording provenance services. Provenance index services deployed with data model, domain model and provenance service model. The data model demonstrates the workflows of the input data objects which assures data quality. The domain model demonstrates the external provenance data objects and

their mapping workflows. The provenance service model encompasses with different useful interfaces in showcasing of data object parameters. [11]

The current day computing environment exhibits the characteristics of HTC and HPC where atomicity plays a prominent role in provenance which consists of workflows in a pipelined fashion. Generally atomicity refers either the event or transaction execution full or null. The discussions of this atomicity in traditional transaction processing have an optimal solution where as in provenance the solutions are specific. As data involved in provenance determined in a form of tuples with the help of query the data can be retrieved. The management of atomicity in hierarchal workflows determines the actor, composite actor, data tokens and data channel represented in the context of use cases. This atomicity management system deals with data dependencies and rounds where they tell us about the whole event at particular point of time. There are some specific operations like reset for terminating the current round which can also be defined explicitly. The token dependency graph was show cased about the rounds and its work flows. The atomicity term with commit and abort operations on rounds for an explicit defined transactions. The commit and abort operations on rounds having their own respective function in concluding of data basing of log information. The data channel in this atomicity management represents reliability with recoverable queue and fault tolerance queue. The evaluation of this atomicity maintenance evaluated by stating the boundaries of round, conflicts between transactions in the rounds, protocols assured for isolation and roll back of operations to achieve performance. The provenance system with an atomicity feature records the event logs of different channel, events, and rounds in a form of provenance ontology using serialize ability. [12]

1.4.2 Scientific Workflow issues in Provenance: The context of scientific workflows in a provenance delivers different perceptive on researchers side and practitioner side. Scientific workflows aim scalability, provenance support, automation and adoption (SPAA). The scientific workflows in the provenance raises a question on workflow model computation and the research perception states about the requirements of the work flow model. The work flow executions in a provenance undertake the research issue about the scalable workflow execution, parallelism, workflow optimization. The workflow designed in a provenance enables the research

issue, simplification of design, less programming, linear model etc. The provenance work flows for reuse targets the research issue in adoption of business process querying language (BPQL), content management etc. The provenance models raise the research issues about the important specification of open provenance model, PROV and W3C models. Tracking of provenance addresses some research issues like provenance over head maintenance, key values from different interfaces while capturing provenance. The issue of querying data from storage of provenance addresses query reachable and regular path queries (RPQs). The research perceptions included for querying provenance storage represents mining and discovering of workflows to the expected outcomes. Most of the current day provenance applications deal the provenance data with privacy aware approaches. The involvement of security mechanisms and control mechanisms balance the security tags involved in provenance applications notifying of overheads on provenance data, debugging and fault tolerances enables to promote the efficiently of workflows with timestamps.[13]

The provenance contains different scientific workflows which are queried for later uses as provenance is tightly coupled with an integration of heterogeneous procedures. In the discussion of scientific workflow management the knowledge of distributed tasks, domain information and relationships between data objects helps for the derivation of data linage. Scientific workflow life cycle incorporates with composition, execution and analysis phases. In the scenario of composition decisions are taken in hypothesis way. This phase admits conception and reusable activities. The execution pattern deploys on monitoring and distribution activities. The analyses phase demonstrates query and virtualization procedures. The data involves in the provenance can be captured with different mechanisms based on internal structures and external services in heterogeneous environment. Provenance data can be traced by lazy and eager approaches propose be captured with different mechanisms based on internal structures and external services in heterogeneous environment. Provenance data can be traced by lazy and eager approaches proposed buneman's work. In capturing of provenance different levels are stated like operating system level, activity level etc.

The techniques of capturing the provenance facilitates with annotation and inversion strategies. A prospective provenance can be collected at composition phase and a retrospective provenance is collected at execution phase. In the debate of the granularity of a provenance coarse grain provenance generates the whole picture of workflows that involves with external environment. The provenance can be accessed with querying of data based on query by example methods usually. Querying of provenance depends on logical programming and semi structured approaches eg., COMAD and XML files. The concept of storage in provenance state the scalability at centralized and distributed scenarios with a classification of homogeneous and heterogeneous systems. The strategy of coupling delivers the provenance data associations. In archiving provenance time stamping process delta sequence methods are used on workflows to showcase the efficiently of provenance used.[14]

1.4.3 Privacy issues in Provenance: Analyzing of Scientific workflows, querying of provenance data encompasses in revealing of private information that exists in the provenance. Many workflow systems deploy their own flow execution models which are evidenced in monitoring the provenance. Open provenance availability leads privacy concerns. The privacy issues related to the scientific workflows of provenance are notified in a Black Box module, where the information of the provenance is not leaked. Data mining, privacy preserving approaches makes the perfect data recordings by identifying the attributes.

In the content of privacy preserving, auditing queries and perturbation of techniques are taken into account to exhibit the uniqueness of privacy. The work flow model of these privacy issues comprises with modules, Input and output data, input ports/output ports and parameters for the modules. The privacy managing workflow models raises three issues about the privacy of module, privacy of data and privacy of the provenance. In preserving the privacy of module all the modules can be termed as private and hide the subset data. These workflow modules which raise the privacy issues are depicted in the form of graphs where edges and nodes are incurred with values. [15]

The concept of privacy in a provenance raises many questions about analyzing of data and provenance queries. The work flow specifications in a provenance are

picturized by using directed acyclic graph which comprised of edges, modules, and data flows. The output of this DAG graph resembles workflow specifications that are carried on provenance data items which help to derive privacy levels. In the part of considering data privacy which is identified workflow modules of provenance is evaluated with verifiability parameters. The module privacy adopts the intermediate data and exhibits it to the user at certain levels. The concept of sub graph exhibit vertices and edges is notified as structural privacy and preserving the privacy levels to an extent. The combination approach of searching with privacy encapsulates the unwanted workflows to the query. [16]

1.4.4 Security issues for Provenance: The involvement of service oriented architecture at web and grid environment evaluates the utilization of provenance. Providing security concerns to the provenance is dependent on environmental architecture to that applications pertaining to SOA services. GRIPHYN consists of set of tools which shows the workflows in grid environment. The architectural models in grid and web environments attached with components, inputs, outputs, workflows, messages, and services interact with each other to execute the workflow as a process. The logical components in SOA architectures denoted as provenance store where the process concludes with queries, p-assertions and actors. Security issues can be distinguished based on access control over process documentation, trust ability frameworks, accountability of p-assertions, sensitivity of p-assertions, and storage procedures of p-assertions. [17]

Querying of data linage from the data provenance can be retrieved under the concern of security view. In the scenario of maintaining confidentiality the scientific workflows are analyzed as intra genomic combinations pertaining workflow comparisons, unique identifications, size of the workflow and value of the workflow was demonstrated. As the workflow in the provenance is hierarchal security view is the dependent post doc approach which results on provenance information. This security new scientific workflow provenance model encompassed with atomicity task, composite task, workflow execution provenance task. These tasks show case the granularity and dependency levels for evaluating security model for the work flow specifications. SECPROV is a framework which deals the above discussed security views in provenance workflows [18]

1.5 Chapter Summary

In this chapter we covered admittance of cloud paradigm from its root level changes to the demanding computing situations. Cloud computing facilities and advantages are mentioned through SaaS and importance of provenance with trust was clearly structured. The pivot role of monitoring which helps for cloud auditor assessment supported with the picture of cloud computing collaborators, table of different service providers and brokers are justified to the maximum extent. Provenance challenges are distinguished with security privacy and workflows were addressed clearly.