Graph-based Campaign Amplification in Telecom Cloud

M. Saravanan[1], Sandeep Akhouri[1] and Loganath Thamizharasu[2]
[1]Ericsson R&D, Ericsson India Global Services, Chennai, India
[2]College of Engineering Guindy, Anna University, Chennai, India

Keywords: Telecom Cloud, Campaign Management, NoSQL Graph Database, Map Reduce Framework.

Abstract: Majority of telecom operators are making a transition from a monolithic, stove-pipe approach of creating services to a more flexible architecture that provides them agility to rapidly develop and deploy services. New revenue streams require an ability to rapidly identify and target dynamic shifts in traffic patterns and subscriber behaviour. As subscriber behaviour morphs with plans, promotions, devices, location and time, this presents challenges and opportunities for an operator to create and launch targeted campaigns. The enormous volume of data being generated requires a scalable platform for processing massive xDR (eg. Call Detail Records). This paper proposes graph databases in a telecom cloud environment for quickly identifying trends, isolating a targeted subscriber base and rapidly launching campaigns. We also highlight the limitations of a conventional relational database in terms of capturing complex relationships as compared to a NoSQL graph database and the benefits of automatic provisioning and deployment in the cloud environment.

1 INTRODUCTION

As mobile penetration worldwide reaches near saturation level, telecom operators are looking beyond net additions of subscribers for revenue growth. Improvements in operational efficiency leading to increased consolidation and managed services will help reduce costs. However, the growth in revenues will be driven by innovative service offerings and service differentiation. Tiered data plans and Quality of Service (QoS) offerings are steps in this direction. Monetizing new revenue streams and leveraging partnerships will require operators to significantly improve subscriber intelligence.

We explore generating dynamic campaigns that can be launched in parallel. A BigData platform in a hybrid cloud environment provides ability to process large volumes of data while leveraging the advantages inherent with cloud platform (Zhang et al., 2010). To enhance the process, we have used another upcoming technology - graph databases in a highly scalable Hadoop (Tom, 2009) framework. Sophisticated queries were efficiently modelled using a NoSQL graph database for generating various types of campaigns in a telecom cloud environment (Leavitt, 2010). We also explored the graph indices available with the graph database for generating the node-level summary.

The main contributions of this paper are: (i) a newly developed software framework for modeling cloud computing environments (ii) An end-to-end cloud network architecture that supports the huge telecom volume and various services and (iii) It exposes powerful graph-based features that could easily be extended for modeling custom cloud computing environments (like automatic campaigns) and other applications.

2 BACKGROUND

Cloud computing is a promising paradigm for the provisioning of IT services in any new applications (Paul and Dan, 2010). Recent advances in cloud computing, BigData platforms such as Hadoop and NoSQL technologies present a unique opportunity for telecom operators to improve upon their subscriber intelligence. Call Detail Records (CDRs/xDRs) contains the details of the calls made by the customers. The data related to the customers interest added to the existing dataset. This synthetic data is added for launching targeted campaigns In this paper, we use HDFS and MapReduce

Saravanan M., Akhouri S. and Thamizharasu L..
Graph-based Campaign Amplification in Telecom Cloud.
DOI: 10.5220/0004050801950198
In *Proceedings of the International Conference on Data Technologies and Applications* (DATA-2012), pages 195-198
ISBN: 978-989-8565-18-1

framework to develop a reliable and high performance platform for processing, aggregation and storage of CDRs. We leverage the automatic provisioing and utility pay-as-you-use model provided by a cloud computing (Armbrust et al., 2010). We experimented with several graph databases and chose InfiniteGraph (Wood, 2011) for modeling the inherent call graph structure.

2.1 Telecom Cloud

Majority of telecom operators manage a large computing infrastructure comprising of a diverse set of hardware, software and applications. They operate in a high availability, performance and secure environment. In this paper, we propose using a private cloud to offload the batch processing and storage of xDR. This provides an ability to handle peak workloads and release excess computing capacity for other applications. As discussed, the private cloud environment can be setup internally. In order to cost efficiently scale the graph databases, we are using Amazon's Infrastructure as a Service (IaaS) (Garfinkei, 2007) for public cloud setup. The graph database can be exposed as Software as a Service (SaaS) for a variety of applications, customers and partners. The proposed model tackles the pre-processing and batch operations in an on-premise private cloud. It hosts the graph database in a public cloud for generating campaigns.

A campaign lifecycle comprises of 3 major phases – identify the subscribers, launch the campaign and follow-up. Potential subscribers for a campaign can be identified based on segmentation performed on behavioural characteristics e.g. high spend, low usage etc. Triggers associated with business rules can also be used to identify a target set e.g. send international rate plan promotion to subscribers who make an international call.

With the available information, campaign computing based on graphs allow us to quickly identify subscribers with a certain interest e.g. subscribers who are interested in a major sporting event can be identified by traversing the graph with a search criteria. In addition to simply identifying the subscribers, we can also get detailed information on the 'neighbourhood' of subscribers who fit these criteria. We can explore if these people are part of a smaller network etc.

Campaign Computing allows operators to perform a deeper level of analysis. It is tightly integrated with social networking and forms the basis of more dynamic marketing models. Current campaign management systems can design rules to identify subscribers who are travelling to a different city and send a promotion. However, graph based systems can allow the operator to send a reference of another subscribers from the subscribers circle of friends who is already using this promotion.

2.2 Graph based Analytics

Graph analytics is the study and analysis of data that is formed as a network (graph) of objects and relationships (Modani et al., 2010). Unlike traditional analytics that focuses on summarizing, grouping and filtering data, Graph analytics addresses traversing and navigating the network. The benefit of using graph database is to quickly find relationships, connections and patterns between objects (Saravanan et al., 2011).

InfiniteGraph (Wood, 2011) is a java API used for creating and managing graph database which provides fast and efficient query by traversal operation across massive and distributed data stores. It supports breaking down navigational queries into subtasks and executing them across clusters. The federated InfiniteGraph architecture (collection of distributed databases) allows elastic scalability of both processing and data in the cloud environment. Other major graph database alternatives are AllegroGraph, graphDB, FlockDB and Pregel.

Telecom call graphs can be modeled with edges to denote voice calls, connections to websites, relationship between users etc. Edge weights can be used to denote call duration, call cost etc. Nodes designate subscribers. Subscriber information such as demographics, home location etc. can be stored at the node level. Often, campaigns are launched using aggregated data attributes e.g. usage for past 3 months > 100 minutes. In order to accommodate these cases, aggregated and derived data can also be stored at the node level.

3 TELECOM CLOUD ENGINEERING

In this section, we discuss building a telecom cloud. We use open source Eucalyptus (Nurmi et al., 2009) to build the private cloud. Amazon Web services are used for public cloud setup. The internal private cloud provides an operator complete control over the environment and enforces stringent security requirements. The public cloud provides scalability and access to enterprise users, subscribers and partners. The confidential data and its pre-processing are maintained in the private cloud and not exposed in the public cloud. Fig. 1 describes the proposed modeling of triggering campaigns.

3.1 Infrastructure

The telecom cloud infrastructure was designed as a hybrid cloud with commodity servers and Amazon EC2 instances. A variety of cloud computing platforms are available including Eucalyptus, OpenStack, Opennebula and Nimbus (Peng et al., 2009). Eucalyptus was chosen to build an on-premise private cloud environment based on the cloud types, interfaces, compatibility, and implementation & deployment requirements. Eucalyptus helps provide a highly available, scalable, environment compliant with Amazon EC2 APIs. Since the architecture of eucalyptus is simple, flexible and modular it allows administrators to deploy and manage machine images and other users can launch and access virtual instances of the machine images. Major components in Eucalyptus architecture are Cloud Controller, Walrus Controller, Cluster Controller, Storage Controller and Node Controller.

The Cloud Controller provides access point into the entire cloud infrastructure for users and administrators. It queries the node managers for information about resources, making high level scheduling decisions and implements them by making requests to cluster controllers. Walrus Controller (WS3) is a persistent put/get storage service. It provides a mechanism for storing and accessing machine images. Cluster Controller gathers the information about and schedules VM execution on specific node controllers, as well as manages the virtual instances networking. Storage Controller (SC) provides persistent block. It allows creating snapshot volumes. Node Controller (NC) controls the execution, inspection and termination of VM instances at the host on top of the hypervisor.

3.2 VM Allocation Modeling

Virtualization is the backbone technology for cloud computing. Hypervisor provides a layer of abstraction for the Virtual Machines (VM). A VM encapsulates the resources like Storage, Network, Hardware, Operating System and pre-Installed applications. In private cloud environment users are provided the functionalities to run and control VM instances and cloud users can access VMs across various compute nodes. An individual VM core image was built using KVM hypervisor bundled with Java, Graph database, Tomcat web server, Apache Hadoop. In Euca2ools ram-disk, kernel of a respective platform (OS) is bundled-uploaded-registered to the private cloud so that the images with its respective kernel, ram-disk with specified storage can be spawned as a VM.

3.3 Modeling Dynamics

The large volumes of CDR's (eg. 200 GB capacity) were pre-processed using Map Reduce framework (Jeffrey and Sanjey, 2004) in the telecom cloud. The subscriber call graphs were partitioned by location for manageability and performance evaluation. Each node is named with unique random label. An MSMapper program written to do simple socket communication between cloud and operator server that is maintained in the operator side to map mobile numbers with unique random labels. Upon returning the results of the query, the results can be mapped back to the particular mobile number to trigger campaigns.

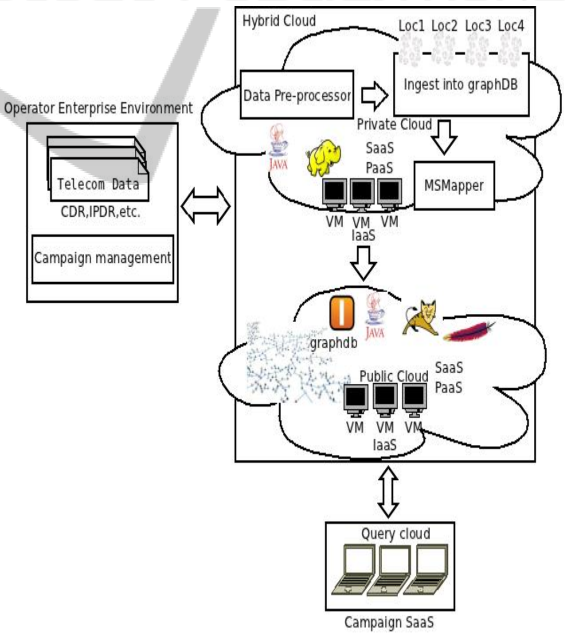

Figure 1: Telecom cloud modelling.

3.4 Security

In order to ensure data confidentiality, critical data is maintained in the on-premise private cloud. Eucalyptus provides ingress filtering through the concept called security groups which are associated with individual rules with respect to IP/Network, Protocol type, destination ports etc.

Additional users who need to access the cloud are placed into respective security groups. Communication with eucalyptus instances occurs via command line through secured socket shell connection using public key cryptography.

In addition to the Eucalyptus security model, MSMapper engine ensure that the confidential subscriber details are never shared outside the private internal cloud. Access to the graph database in the public cloud is governed through standard user authentication mechanism. The extracted graph data from private cloud is securely migrated to the public cloud for faster processing (scalable environment) of customer details.

4 GRAPH CAMPAIGN EXPERIMENTS

Initially, the telecom graph data's are preloaded in the cloud instance with server cluster with high memory for processing graph campaign experiments. The telecom data that are visualized in graph can be trivially represented in relational database. Querying relational database involved in larger joins incorporating higher costs for greater degree of separation. For example, a simple query is executed to find immediate friends and FOAF with specific conditions are represented in SQL as well as in Graph Database and evaluated. The graph based experiments for immediate friends compared with respect to RDBMS.

On average to traverse 4673 FOAF paths it took 0.0184s in graph database comparing with RDBMS took a time of about 0.468s. When the degree of separation increased to 3, time taken to traverse 160000 dynamic paths of total 667723 edges in a graph database around 0.957s while in RDBMS it took 50.59s. On increasing the degree of separation and total traversal amount RDBMS started to perform poorly compared to the graph database. The graph query performance was better when compared to SQL queries on the provided cloud infrastructure.

5 CONCLUSIONS

This paper explains the exploitation of NoSQL graph databases in a cloud context for the provision of campaigns by telecom operators towards their mobile clients. It also discusses the software framework for modeling Cloud Computing environments and an end-to-end Cloud network architecture for telecom oriented services. Campaign Computing in Cloud environment using graph database provides operators a cost-effective, scalable, secured, advanced analytics platform to target the telecom customers. This allows operators to create dynamic, real-time marketing models. The proposed framework quickly identifies trends, isolating a targeted subscriber base and rapidly launching campaigns.

REFERENCES

Zhang, Cheng, Lu, and Boutaba, R., 2010. Cloud computing: State-of-the-art and research challenges, *Journal of Internet Serv.Appl.* 1: 7-18.

Tom, W., 2009. *Hadoop: The Definitive Guide,* O'Reilly Media / Yahoo Press, California, USA, 2nd edition.

Leavitt, N., 2010. Will NoSQL Databases Live Up to Their Promise? *Computer*, 43:12–14.

Paul Hofmann and Dan Woods, 2010 Cloud Computing: The Limits of public clouds for Business Applications, *IEEE Internet Computing*, 14:90-93.

Armbrust, M., Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M., 2010. A view of cloud computing. *Communications of the ACM*; 53(4):50–58

Darren Wood, Introduction to InfiniteGraph, The distributed and scalable graph database, 2011. *NoSQL Now!,* San Jose, USA

Garfinkei, S, 2007. *An Evaluation of Amazon's Grid Computing Services: EC2, S3 and SQS.* Tech. Rep. TR-08-07, Harvard University.

Modani, N, Dey, K, Mukherje, S, and Nanavati, A, 2010. Discovery and analysis of tightly knit communities in telecom social networks, *PIBM Journal of Research and Development*, 7:1-7.

Saravanan M., Prasad G., Karishma S., and Suganthi D, 2011. Analyzing and Labeling of Telecom Communities using Structural Properties, *International Journal of Social Network Analysis and Mining*, Springer Netherlands, 1-16.

Nurmi, D., Wolski, R., Grzegorczyk, C,. Graziano, O,. Soman, S,. Youseff, Lamia,. Zagrodnov, Dmitri., 2009. The Eucalyptus Open-Source Cloud – Commputing System, *9th International Symposium on Cluster Computing and Grid*.

Peng, J., Zhang, X., Lei, Zhou., Zhang, Wu., Li, Q., 2009. Comparison of several cloud computing platforms. *2nd International Symposium on Information Science and Engineering*.

Jeffrey Dean, Sanjay Ghemawat, 2004. MapReduce: simplified data processing on large clusters, *Opearting Systems Design & Implementation*, San Francisco, CA, p.10-10.