Near Real–Time Call Detail Record ETL Flows
(Industrial Paper)

Munir Cochinwala and Euthimios Panagos
Telcordia Applied Research
One Telcordia Drive, Piscataway, NJ 08854
{munir,thimios}@research.telcordia.com

Abstract. Telecommunication companies face significant business challenges as they strive to reduce subscriber churn and increase average revenue per user (ARPU) by offering new services and incorporating new functionality into existing services. The increased number of service offerings and available functionality result in an ever growing volume of call detail records (CDRs). For many services (e.g., pre-paid), CDRs need to be processed and analyzed in near real-time for several reasons, including charging, on-line subscriber access to their accounts, and analytics for predicting subscriber usage and preventing fraudulent activity. In this paper, we describe the challenges associated with near real-time extract, transform, and load (ETL) of CDR data warehouse flows for supporting both the operational and business intelligence needs of telecommunication services, and we present our approach to addressing these challenges.

Keywords: Real-time business intelligence, real-time ETL.

1 Introduction

Today, telecommunication companies are facing significant challenges in reducing subscriber churn and increasing average revenue per user (ARPU) as they try to stay competitive and expand their service offerings. Gathering intelligence via real-time visibility into subscriber usage is a critical element in both business activity monitoring and real-time decision support solutions. Timeliness of the intelligence gained from real-time visibility into subscriber usage is important for revenue gain as well as protection from revenue loss. Consider, for example, the case where the absence of a rule in the logic associated with real-time charging of voice calls results in free international calls for subscribers who dial a specific prefix. Obviously, not being able to detect this abnormal condition in a timely fashion could result in substantial loss of revenue. On the other hand, revenue enhancement opportunities are time sensitive and, often, event based, such as access to tickets for a band or a video of a missed goal in the soccer world cup.

Subscriber usage in telecommunications is recorded in call detail records (CDRs). CDRs represent a wealth of information that can be mined in order to discover patterns related to both calling behavior and service feature usage (e.g., SMS, MMS, etc.). The majority of CDRs are generated by the telecommunication switches.
(numbering around 25,000 in the United States) and intelligent network (IN) call processing nodes that handle specific service transactions, such as toll free telephone number translation and mobile number portability. CDR files are fetched from switches and IN call processing nodes on a periodic basis (e.g., every two minutes) and stored in a staging area that offers reliable near-term storage functionality since telephone switches and other network elements have limited storage capacity. Generally, the near-term CDR storage component does not provide any advanced CDR analysis functionality because the emphasis is on billing (including real-time charging) which requires efficient and timely off-loading of the original CDR files from inline session management network components.

Enabling near real-time CDR mining and analytics requires that the infrastructure responsible for extracting CDRs from the near-term storage component and inserting them into longer term storage (for billing applications as well as data warehouses for decision support and business intelligence) is able to operate in near real time. Such extractions are typically performed by customized extract, transform and load (ETL) flows. These ETL flows must address many challenges not typically found in traditional data warehousing projects. In particular:

1. Handle introduction of new service CDRs in a timely manner and with minimum changes in processing flows;
2. Handle changes to existing CDR attributes with minimal impact in processing flows;
3. Support multiple CDR versions for the same service (this occurs when different service functionality is enabled for sub-sets of subscribers);
4. Accommodate introduction of new near-term CDR storage repositories (required when the number of subscribers exceeds specific thresholds);
5. Support multi-tenant solutions where either multiple services co-exist (e.g., CDMA and GSM versions of a pre-paid offering) or different versions of the same service are being offered to different subscriber groups.

In this paper, we present our work in the area of real-time processing of CDRs in the context of a hosted pre-paid service for mobile virtual network operators (MVNOs). In particular, we present a flexible and extensible CDR extraction, transformation, and load solution that handles dynamic changes to CDR attributes, introduction of new service CDRs, and versioning of CDR in a simple and efficient manner. The solution can be easily generalized to support services in other domains that exhibit the same characteristics with respect to the data that needs to be processed within strict timing constraints. The remainder of this paper is organized as follows. Section 2 provides background information with regard to the MVNO business model. Section 3 discusses the problem we address in this paper. Section 4 outlines our solution. Section 5 covers related work and, finally, Section 6 concludes our paper.

2 MVNO Background

The mobile telecommunications market is considered to be a very lucrative one. However, building a mobile network and purchasing spectrum licenses can be very