SLA guided data integration on cloud environments
Sharing energy consumption in social networks

Keywords: The paper must have at least one keyword. The text must be set to 9-point font size and without the use of bold or italic font style. For more than one keyword, please use a comma as a separator. Keywords must be titlecased.

Abstract: This paper proposes data integration (lookup, aggregation, correlation) strategies adapted to the vision of the economic model of the cloud such as accepting partial results delivered on demand or under predefined subscription models that can affect the quality of the results; accepting specific data duplication that can respect privacy but ensure data availability; accepting to launch a task that contributes to an integration on a first cloud whose SLA verifies security requirement rather that a more powerful cloud but with less security guarantees in the SLA.

Deadline 15/11-

Conférences cibles:
CLOUD 2014: http://www.thecloudcomputing.org/2014/cfp.html Work-in-progress abstract: 17/01, full paper: 22/01

CLOSER14: http://closer.scitevents.org/: deadline : 12 Decembre pour les poisitions paper

Workshops:
https://sites.google.com/site/clouddb2014/ deadline 13 Novembre 2013 (To discard)
http://endm2014.endm.org/: December 7, 2013

A completer...

1 INTRODUCTION

The emergence of new architectures like the cloud opens new challenges to data processing. The possibility of having unlimited Access to cloud resources and the "pay as U go" model make it possible to change the hypothesis under which current technology and solutions address the processing of huge volumes of data. Instead of designing processes and algorithms taking into consideration the limits on resources availability, the cloud sets the focus on the economic cost implied of using resources and producing results by parallelizing the use of computing resources while delivering data under subscription oriented cost models.

Current data management approaches on the cloud tend to use NoSQL stores for managing huge heterogeneous data collections (graph, key-value, tables, relational). Yet, having such heterogeneous schemaless data stores, calls for efficient methods for integrating (correlating, associating, filtering) heterogeneous data collections taking into consideration their "structural" characteristics (due to the different data models) but also their quality of service, e.g., trust, freshness, provenance, partial or total consistency. Existing data integration techniques based on mappings, views management, ontologies have to be revisited in order to integrate data that are weakly curated and described through metadata or schemas.

Our work intends to address data integration on a hybrid cloud guided by the SLA exported by different cloud providers and by several QoS measures associated to data collections properties: trust, privacy, economic cost. We aim to address big data integration d in a multi-cloud hybrid context. This implies several granularities of SLA: first, at the cloud level; the SLA ensured by providers regarding data; then at the service level, as unit for accessing and processing data, to be sure to fit particular service needs; and finally at the integration level i.e the possibility to process, correlate and integrate big data collections distributed along different cloud storage supports, providing different quality properties to data (trust, privacy, reliability, etc).

The objectives of our work are to propose an SLA guided continuous data provision and integration system exported as a DaaS by a cloud provider. Therefore we propose strategies for computing integrated SLAs according to agreed SLAs proposed by services and optimized and adaptable query rewriting for integrating data sets according to user profiles.

This paper proposes data integration (lookup, aggregation, correlation) strategies adapted to the vision of the economic model of the cloud such as accepting partial results delivered on demand or under predefined subscription models that can affect the quality of the results; accepting specific data duplication that can respect privacy but ensure data availability; ac-

cepting to launch a task that contributes to an integration on a first cloud whose SLA verifies security requirement rather that a more powerful cloud but with less security guarantees in the SLA.

Accordingly, the remainder of this paper is organized as follows. Section 2 presents related works that address SLA modelling, integration and SLA guided data management processes. Section 3 gives an overview of our approach for integrating data sets provided by services (i.e., DaaS) by concilating SLA's provided by services and user's profiles expressing QoS preferences about the data they want to consume and the conditions in which they must be processed and delivered. Section 4 introduces on demand and incremental data integration strategies. Section 5presents a use case for illustrating the interest and use of our approach. Finally 6 concludes the papers and discusses future work.

2 RELATED WORK

- Cloud computing represents a novel on-demand computing approach where resources are provided in compliance to a set of predefined non-functional properties specified and negotiated by means of Service Level Agreements (SLAs). SLA currently exploited in the cloud allows service providers and cloud client to fix the resource level that should be used either by a service for the service provider or by the client in the case of service invokation. SLA could concern all types of the services on the cloud (IAAS, PAAS or SAAS). One of the weakness of SLA for the client is that they express low level clauses like storage amount use or virtualization level. In fact when a client launches a task on the cloud, she has unlikely Q & S requirement rather than low level features. Recently we notice the existing of a lot of works with the objective to fill the gap. Emeakaroha and al propose in (Emeakaroha et al., 2010) a cloud component that acts autonomously based on low level monitored features after analysing them towards high level clauses expressed by the user. Moreover, another difficulty to enforce an SLA is to measure and identify, starting from a high level SLA clause, how could it be declined at different layers in the cloud. Therefore in (Dastjerdi et al., 2012) the authors describe a semanticSLA which can be understood by all parties including providers, requestors, and monitoring services. One of the major objectives in the cloud is to anticipate SLA violation and to assess SLA failure cascading on violation detection. (Dastjerdi et al., 2012) proposes an SLA dependency modeling using Web Service Modeling Ontology (WSMO) to build a knowledge database. (Emeakaroha et al., 2010) proposes to anticipate failure by analysing the monitored feature and to act by anticipation. in (Brandic et al., 2010), the authors propose LAYSI, a layered solution that minimizes user interactions with the system and prevents violations of agreed SLAs. On the other hand SLA contracts do not cover all client requirement. There is still a lack in some domain.....(to be completed)
- data integration (cf. travaux Gonzalez)

3 SLA BASED DATA INTEGRATION

Overview of our approach that will include: - An SLA Model: including security issues
- A multi cloud environment representation
- On demand incremental data integration strategies

Figure 1 shows the general architecture of an SLA guided data integration system that is supported by data services which are data providers deployed in a cloud and that provide agreed SLAs. These descriptions are stored in a directory together with meta-data about the way queries are evaluated for producing results. The system uses this information by query processing and monitoring modules for rewriting queries according to given quality of service (QoS) preferences expressed by a data consumer, for example a user.

3.1 SLA model

- Expression haut niveau du SLA en termes de prfrences qui doit converger avec le SLA technique des services.
  - (Souhait de temps de rponse, cot des services, espace de stockage,
  - Templates pour exprimer le SLA
  - Intgration: modle pivot de SLA
- SLA violation contrle avec des mechanisms de monitoring.
- Que ce que devient lintgration de donnes par rapport au SLA
- Cration dynamique de SLA niche de march: tant donne deux SLA fabriquer un SLA dintgration

In order to propose an SLA guided continuous data provision and integration, we need to think about

Figure 1: General architecture of an SLA guided data integration system.

possible steps from the request to the delivery of the result sets. Indeed, let consider a request R launched by a user who specifies a number of constraints on the environment execution. Executing this query requires first a semantic analysis which will subdivide R into a set of sub-queries, in such a way that each sub-query can be processed by a DataService deployed on the cloud. One may think to a first filter to remove the individual services which do not meet some or all of the constraints expressed by the user. This first stage can be defined as a vertical mapping SLAs given the high level SLA described by the user (i.e. macroscopic constraints: execution time, pay / no pay, data reliability, data source). The system should be able to find relevant service compositions that respond to the query and, when combined, meet the constraints imposed by the user (High level SLA). To meet this objective, it is necessary to compare the SLA services to combine, in order to check if their joint use is compatible with the individual SLA. This step may lead either to the rejection of integration in case of total incompatibility, or to a negotiation between SLA which will lead us to the proposal for a negotiated SLA integration and thus the need for an adaptive Template. The negotiation of this type of SLA depends strongly on the request sent and the services deployed at the arrival time of the application on the cloud. This negotiation can be expensive and may not scale well. It is therefore crucial to provide proactive mechanisms for optimizing the production of such SLA. We believe that the optimization of this process can occur at two levels, firstly at the level of SLA previously traded, and secondly at the level of partial or total results. Indeed, queries requesting the same services compositions will have clauses in their SLAs that are more conditions of use of the infrastructure (ie not touching the data). For two different queries, they will be negotiated in the same way. These previously negotiated SLAs are reusable. In a second time, we think optimizing this process on the data storage mechanisms to cache intermediate results, individually or in partial or complete combination depending on the terms of SLA services (data access, intermediate storage capacity , cost of storage , etc ...). Given this proposal, we identify several issues: - Level modeling would require a model that allows the representation of SLA integration. - There should also be a template for representing the requirements expressed by the user - A mining component to identify, from the requirements expressed in the template by the user, and before the analysis of the application, the candidate integration SLA to use or adapt according to the request. This implies mapping between property and expressed clauses being.

3.2 Query Rewriting

Here Martin will explain the rewriting problem.

3.3 Data and query models

TBD.

4 ON DEMAND AND INCREMENTAL DATA INTEGRATION STRATEGIES

Given a requirement expressing a query and quality of service preferences : cost, provenance, reputation, time the system processes it according to the following steps:

1. See whether a similar SLA has been computed before
   $\longrightarrow$ **yes:** use it
   $\longrightarrow$ **else:** compute the total or partial SLA and store it in the history
2. Computation of a global SLA given the existing possible SLA agreed by data providers that can be called for answering the query. Data providers are filtered in this way, since only those agreed SLA that can be combined into a global SLA that can fulfill the user preferences are considered for retrieving data.
   (a) Filter the data providers that can potentially participate in the evaluation of the query taking into consideration the preferences associated to the query
   (b) Rewrite the query into n subqueries that can compute a partial answer
3. see wether a similar Q has been already rewritten
   $\longrightarrow$ **yes:** use the rewriten queries
   $\longrightarrow$ **else:** compute it and store the result
   (a) Generate a service coordination search space that can compute each subquery and can integrate the global result. Each subquery is optimized with respect to user profile
   (b) Dispatch the execution of subqueries, integrate them into a result

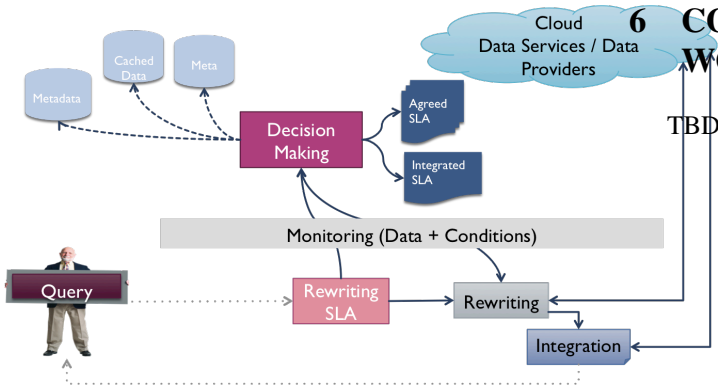

Figure 2: Energy exchange.

⟶ at execution: monitor the consumption of resources, the execution, time, behaviour of the services and make decisions

5 USE CASE

Consider a smart city that aims at being energy self-sustainable and produce and consume as much as possible, energy within its geographic area. Producers are characterized according to their location, the amount of energy in kWatts-hour that they can sell, the cost, and the time window in which they can produce it, with a given service level agreement concerning their availability and fault tolerance. Consumers, give also their location, their energy requirements during a certain interval of time, the maximum total cost they are ready to pay, and quality of service requirements such as availability and how critical it is to consume this amount of energy. A energy exchange market is established in order to continuously monitor energy provision/consumption ensuring that all consumers will have the energy they require at every moment.

In our approach energy producers are modelled as services with associated "agreed" SLAs for a given time window. In general, we assume that several producers will be able to supply energy for a given period of time given specific QoS preferences expressed by a consumer. An energy request is expressed as a query that specifies an energy requirement with QoS preferences independently of the possible producers.

- Processing big data implied in the energy consumption observation
- Computing energy consumption behavior models
- Analyse and optimize energy consumption versus respecting the comfort requirement of inhabitants

Need of efficient data processing solutions

6 CONCLUSIONS AND FUTURE WORK

TBD.

REFERENCES

Brandic, I., Emeakaroha, V., Maurer, M., Dustdar, S., Acs, S., Kertesz, A., and Kecskemeti, G. (2010). Laysi: A layered approach for sla-violation propagation in self-manageable cloud infrastructures. In *Computer Software and Applications Conference Workshops (COMPSACW), 2010 IEEE 34th Annual*, pages 365–370.

Dastjerdi, A. V., Tabatabaei, S. G. H., and Buyya, R. (2012). A dependency-aware ontology-based approach for deploying service level agreement monitoring services in cloud. *Softw. Pract. Exper.*, 42(4):501–518.

Emeakaroha, V., Brandic, I., Maurer, M., and Dustdar, S. (2010). Low level metrics to high level slas - lom2his framework: Bridging the gap between monitored metrics and sla parameters in cloud environments. In *High Performance Computing and Simulation (HPCS), 2010 International Conference on*, pages 48–54.