Trusted SLA guided data integration on multi-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): polyglot approaches. Yet, having such heterogeneous schema-less 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.

Furthermore, security aspects are crucial in integrating big data. Indeed, the integration task must guarantee the integrity and privacy of data, the willingness of the clouds to well contribute to the security process through acceptable runtime environment conditions, and the adequacy of the proposed security levels to the services behind the integration operation.

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 that will be exported as a DaaS by a cloud provider. Therefore we propose strategies for computing integrated SLAs according to agreed SLAs

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

proposed by services and optimized and adaptable query rewriting and bi data sets integration according to user preferences. Given the computational cost of a query evaluation, our approach uses automatic learning techniques for generating knowledge out of every task and reducing its economic cost.

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.

2 RELATED WORK

- security concerns - SLA
- 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

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. Computation of a global SLA given the existing possible data providers that can be called for answering the query.
2. 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
   (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 that can compute each subquery and can integrate the global result. Each subquery is optimized with respect to user profile

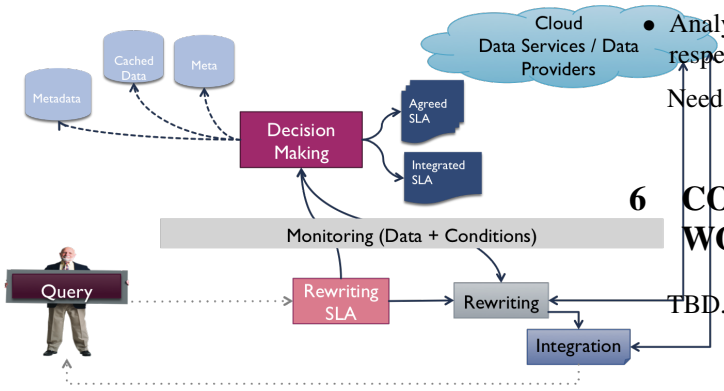

Figure 2: Energy exchange.

(b) Dispatch the execution of subqueries, integrate them into a result

⟶ 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.