

PROBLEM STATEMENT

Current data integration implies consuming data from different data services and integrating the results while meeting the user quality requirements concerning the data that is being integrated such as price, availability, privacy and provenance. Data services provide data according to specific APIs that specify parameters describing the data to be retrieved and the type of results they can produce. Moreover data provision is done according to different data quality measures regarding the conditions in which they provide data, authentication, privacy subscription and the quality of their data, for example freshness, veracity, provenance and other non-functional conditions like economic price.

While consuming data, the user can use services that may need to retrieve data and to processing requests over this data. Executing these requests can require a considerable amount of storage, memory and computing capacity that can be provided by cloud architectures. Furthermore, users could need to integrate the data provisioned by services in a homogeneous and general result. The integration process can also require important computing resources that can be obtained from the cloud, for instance.

Although the cloud is guided by an economic model, the data consumption is determined by a different economic model and quality constraints. These constraints could be the storage and computing capacity of the device that consumes the data, the data transmission bandwidth and cost, whether the data consumption is critical (time constraints) and energy consumption. The economic model of the cloud determines the type of services quality a user can expect from its services. The user profile, her quality requirements and her execution context determine the conditions in which she is expecting to consume data by using cloud services.

The contract between the level of services that the cloud can provide and the user is called SLA (Service Level Agreement). SLA specifies which computing resources the user can access and the quality guarantees that she can expect. Thus, the challenge is to find a match between both expectations, and be able to integrate data. The data should be integrated taking into consideration the service level that each service can ensure, and at the same time trying to fulfill the user expectations in terms of her quality requirements. Here, the data integration process includes looking for services that can be used as data providers, and for services required in order *(i)* to retrieve the data; *(ii)* to build an integrated result; and *(iii)* to deliver it to the user considering the user quality requirements and resources consumption.