Resource- and Human-constrained Data Processing

Panos K. Chrysanthis & Alexandros Labrinidis
Advanced Data Management Technologies Laboratory
Department of Computer Science, University of Pittsburgh
panos@cs.pitt.edu & labrinid@cs.pitt.edu

Motivation

In the last 20 years, the mobile data management research community has produced techniques and solutions to effectively support data processing under the resource constraints inherent to mobile devices, e.g., communication bandwidth, battery, and size of the display. These solutions assumed small (user-device) interactions and the existence of a support infrastructure for sharing data processing between the mobile device and stationary computing devices. The emergence of the Computing Cloud, which has the potential to facilitate the services that comprise the assumed mobile support infrastructure, is making these solutions more feasible.

The hypothesis of this position paper is that the proliferation of mobile devices, and in particular smart phones, and the emergence of the social networks that generate content which is location-specific and potentially time-sensitive, require new solutions that go beyond just handling of the resource constraints to effectively handle large dynamic interactions, while preventing information overload for users.

The New Scalability Challenge

The "traditional" challenge characterizing mobile data management is that data processing in this environment has been (severely) resource-constrained. Although human cognitive capacity (referred to as *human processing* from now on) has also been a constraint in the past, social networks, combined with advances in mobile device technology, have made human processing an equally important challenge in mobile data management. So, the quest is not only in content efficiently to the user (given resource constraints), but doing so in a way to allow the user to process it (in a meaningful way). Context-awareness is playing a big role in identifying relevant content, but is not enough. The new challenge is how to leverage existing techniques and develop new ones, in order to perform (collaborative) data processing in this environment, **given both the inherent resource constraints and also the human processing limitations,** and to do so by taking advantage of the Computing Cloud.

Towards a Solution

We believe that the driving principle in addressing the new scalability challenge is the examination of the trade-offs imposed by the resource- and human- constrained processing. For instance, when choosing what content to push to the mobile devices, we need to consider both the bandwidth/energy requirements and the relevance of the content to the user (i.e., to prevent both energy depletion and mental overload). We believe the solution to this challenge to be at the intersection/integration of three research thrusts.

Materialized views as a mechanism for personalization and controlling information overloading
The idea is to build on the powerful notion of views which provide a means to present different users with different portions of the database, based on the users' perspective (i.e., context, and preferences). In order to support flexible query processing in meeting the needs of the users and also support asynchronous and disconnected communication to minimize energy consumption, one can explore customization and localization properties of the materialized views. Further, we propose to use the view maintenance options to personalize information sharing among users in a social network. Such personalization should consider both the types of content that are downloaded (i.e., be the most relevant for the specific user) and also the data availability/freshness preferences of the users (i.e., by allowing finer grain of control in maintaining consistency, controlling inconsistencies, and considering the ensuing trade-offs).

Such personalized view maintenance [1] can be supported by a cloud computing infrastructure which offers differentiated levels of view maintenance service (view holders) through multi-tenancy in a scalable way.

Computing Cloud as a Mobile Support Infrastructure

View maintenance is typically an expensive proposition [2] and it is becoming increasingly more expensive given the high incoming data rates, for example, due to real-time data feeds from sensor networks, from the Web, or from user-submitted data (on social networks). This necessitates sharing of (personalized view maintenance) computation between mobile devices and the cloud. This type of sharing can increase efficiency (esp. in terms of resource consumption), enable higher levels of sophistication at the cloud (given the computing power), and allow for larger volumes of data to be considered in tandem with user preferences (and possibly aggregated/filtered out at the cloud and not pushed to the mobile clients). The ability to only get notifications of events and data of interest is generalized through the notion of continuous queries, which are implemented through a Data Stream Management System (DSMS) [3, 4].

All these benefits are made available in an elastic way and in a distributed fashion, whereas the push-based data processing paradigm is implemented via a DSMS as a service (DSMSaaS) deployment and reliability is provided by means of relaxed transactional semantics (both at the cloud and the mobile device [5]). Although DSMS or DSMSaaS are often sufficient to handle simple interactions and single-user data processing requests, complex interactions and processing requirements involving multiple users (i.e., the result of collaboration) need a new paradigm. This is the paradigm supported by continuous workflows, i.e, the product of the “marriage” of traditional workflow systems and DSMS [6, 7].

Continuous Workflows as a facilitator of social network (open) collaborations

Being able to work collaboratively increases efficiency and generates better results (decisions) and innovative ideas. Social networks enable such collaboration (for fun and social gain), beyond the confinements of institutional or enterprise barriers. On the other hand, social-network-based collaborations share many similarities with virtual enterprises, effectively supported by workflows which define roles and tasks for each participant in the collaboration, as well as defining the steps needed for content to be developed, deployed, and shared. Collaborations in this context typically have real-time characteristics (in terms of both data and tasks) and involve continuous (i.e., never-ending) data exchanges. For this reason, we propose to implement continuous workflows as a cloud-hosted service (CWaaS), similarly to DSMSaaS. Successful deployment of CWaaS will require consideration of timeliness (by addressing scheduling) and quality of service and results (by addressing uncertainty, ranking of the answers and annotations).

Conclusions

In this position paper, we identified human- and resource-constrained data processing as the main challenge behind the promise of combining social networking, mobility, and the cloud. We proposed to decompose this challenge by utilizing Materialized views as a mechanism for personalization and controlling information overloading, using Computing Cloud as a Mobile Support Infrastructure, and by employing Continuous Workflows as a facilitator of social network (open) collaborations.

References

[1] Susan Weissman Lauzac, Panos K. Chrysanthis, Personalizing Information Gathering for Mobile Database Clients, Proc. of the 17th ACM Annual Symposium on Applied Computing (SAC'02), pp. 49-56, Mar 2002

[2] Alexandros Labrinidis, Nick Roussopoulos, Exploring the tradeoff between performance and data freshness in database-driven Web servers, The VLDB Journal, 13(3):240-255, Sep 2004.

[3] Mohamed A. Sharaf, Panos K. Chrysanthis, Alexandros Labrinidis, Kirk Pruhs, Algorithms and Metrics for Processing Multiple Heterogeneous Continuous Queries, ACM Transactions in Database Systems (TODS), 33(2):5.1-5.44, Mar 2008.

[4] Thao N. Pham, Lory Al Moakar, Panos K. Chrysanthis, Alexandros Labrinidis, DILoS: A Dynamic Integrated Load Manager and Scheduler for Continuous Queries, The Sixth International Workshop on Self-Managing Database Systems (SMDB'11), pp. 10-15, Apr 2011.

[5] Panos K. Chrysanthis, Krithi Ramamritham, Synthesis of Extended Transaction Models Using ACTA, ACM Transactions on Database Systems (TODS), 19(3):450-491, Sep 1994.

[6] Panayiotis Neophytou, Panos K. Chrysanthis, Alexandros Labrinidis, Towards Continuous Workflow Enactment Systems, Proc. of the 4th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom'08), Nov 2008.

[7] Panayiotis Neophytou, Panos K. Chrysanthis, Alexandros Labrinidis, CONFLuEnCE: CONtinuous workFLow ExeCution Engine, Proc. of the 30th ACM International Conference on Management of Data (SIGMOD'11), pp. 1311-1314, Jun 2011