Advanced Science and Technology Letters
Vol.54 (Database 2014), pp.100-108
http://dx.doi.org/10.14257/astl.2014.54.25

# OLAPing and Mining Big Data: Large-Scale, Long-Running, Serendipitous Computations within Next-Generation Clouds

Alfredo Cuzzocrea

ICAR-CNR and University of Calabria, I-87036 Cosenza, Italy
cuzzocrea@si.deis.unical.it

**Abstract.** *OLAPing and Mining Big Data* is among one of the most attracting research contexts of recent years. Essentially, this puts emphasis on how classical *OLAPing and Mining algorithms* can be extended in order to deal with novel features of Big Data, such as *volume, variety* and *velocity. This* novel challenge opens the door to a widespread number of challenging research problems that will generate both academic and industrial spin-offs in future years. Following this main trend, in this paper we provide a brief discussion on most relevant open problems and future directions on the fundamental issue of OLAPing and Mining Big Data.

## 1 Introduction

In recent years, the problem of *Mining Big Data* (e.g., [27,40]) is gaining momentum due to both a relevant push from the academic and industrial worlds. Indeed, mining Big Data is firstly relevant because there are a number of big companies, such as Facebook, Twitter and so forth, that produce massive, Big Data and hence need advanced Data Mining approaches for mining such data, as a first-class source of knowledge that can be further exploited within the same companies to improve their own business processes (e.g., [33]).

Similarly, *OLAPing Big Data* (e.g., [17, 14,52,25]), which can be reasonably considered an area that is strongly related to the one above, is another relevant research context, which has attracted lot of attention from the research community. Due to the intrinsic nature of Big Data (e.g., [22,2, 11,23]) and their typical application scenarios (e.g., [30,9]), it is natural to adopt *Data Warehousing and OLAP methodologies* [29] with the goal of collecting, extracting, transforming, loading, warehousing and OLAPing such kinds of data sets, by adding significant add-ons supporting *analytics over Big Data* (e.g., [1,22,30, 14]), an emerging topic in Database and Data Warehousing research.

On the other hand, supporting *OLAPing and Mining Big Data* within *next-generation Clouds* ha also a *"serendipitous"* nature, meaning that, due to the same decentralized, delocalized nature of *Cloud Computing* infrastructures (e.g., [51]), applications and tools implementing these methodologies are likely to discovery surprising, interesting knowledge in remote Cloud nodes efficiently (e.g., [31,42]). This is a novel

ISSN: 2287-1233 ASTL

and critical research perspectives that will surely impact relevantly in future Big Data research efforts.

The first, critical result which inspired our research is recognizing that classical OLAPing and Mining algorithms are not suitable to cope with Big Data, due to both methodological and performance issues. As a consequence, there is an emerging need for devising innovative models, algorithms and techniques capable of mining Big Data while dealing with their inherent properties, such as *volume*, *variety* and *velocity [39]*. Inspired by these motivations, this paper provides a brief discussion on most relevant open problems and future directions on the fundamental issue of OLAPing and Mining Big Data.

## 2 Current Research Issues

### 2.1 Open Problems on OLAPing Big Data

Several research problems arise when computing OLAP data cubes over Big Data. Among these, we identify the following ones.

*Size:* fact tables can easily become huge when computed over Big Data sets this adds severe computational issues as the size can become a real bottleneck from practical applications (e.g., *[32])*.

*Complexity:* building OLAP data cubes over Big Data also implies complexity problems which do not arise in traditional OLAP settings (e.g., in relational environments) for instance, the number of dimensions can really become explosive, due to the strongly unstructured nature of Big Data sets, as well as there could be *multiple* (and *heterogeneous) measures* for such data cubes.

*Design: design methodologies* for OLAP data cubes have been of relevant interest for Database and Data Warehousing research. In the specific case of designing methodologies of OLAP over Big Data, the *performance aspect* must be taken into greater consideration, due to obvious spin-offs given by such design task in this case, designers must move the attention on the following critical questions: (*i*) what is the *overall building time* of the data cube to be designed (computing aggregations over Big Data may become prohibitive)?; (*ii*) how the data cube should be updated? which *maintenance plan* should be selected?; (*iii*) which *building strategy* should be adopted (e.g., *divide & conquer* (e.g., *[53])?*

*Computing methodologies:* due to the enormous size, computing OLAP data cubes over Big Data will turn (again!) into a challenging research problem, similarly to what happened for early OLAP data cube computing experiences (e.g., *[12])* in this case, the most promising technology to follow seems to be the emerging *Cloud Computing* paradigm, perhaps inspired by classical *parallel computing* methodologies (e.g., *[24, 4])*.

*In-memory representation:* how an OLAP data cube over Big Data should be mapped in memory? This is a critical challenge to be considered, due to the fact that the very high number of dimensions in such cubes easily convey to explosive cell car-dinalities as a consequence, solutions based on tertiary memory should be deeply investigated.

 

*Innovative hardware support:* it is natural to figure-out that innovative *hardware solutions,* such as *GPU-based data processing* (e.g., [48]), will play an important role with respect to the issue of computing OLAP data cubes over Big Data.

*Query languages and optimization:* classical *MDX approaches* do not incorporate *optimization solutions* prone to deal with Big Data needs; future investigations must focus the attention on optimization issues given by processing Big Data in a multidimensional fashion.

*End-user performance:* OLAP data cubes computed over Big Data tend to be huge, hence end-user performance easily becomes poor on such cubes, especially during the *aggregation* and *query phases* therefore, it follows that end-user performance must be included as a critical factor within the *design process* of OLAP data cubes over Big Data.

*Quality:* quality aspects will more and more become a critical factor in next generation Data Warehousing and OLAP methodologies over Big Data in fact, due to the strongly unstructured nature of Big Data sources, aggregations computed on such data sources can easily turn out to be poor; hence, it is easy to understand how much important controlling the quality of final data cubes will become.

*Usability:* OLAP data cubes over Big Data must, prominently, be processed and managed to extract and build useful analytics this aspect opens the door to a wide family of research problems, such as devising methodologies to measure how much usable an OLAP data cube built on Big Data repositories is.

*Visualization:* as Big Data expose explosive size, visualization issues of OLAP data cubes (e.g., [20,21]) over Big Data play a first-class role in this research field as a consequence, a novel class of *visualization metaphors, methodologies* and *solutions* must be devised, in order to cope with emerging challenges posed by visualizing massive OLAP data cubes over Big Data; real-time visualization of extracted core data, visualization of mashuped data, and effective visualization over mobile devices should also be considered.

*Interactive exploration:* coupled with visualization issues, *interactive exploration issues* are severe milestones to traverse in the context of OLAP data cubes over Big Data research in fact, enormous-in-size data cubes are difficult to explore (e.g., under the execution of a fixed analytical process) while extracting useful knowledge, with important implications such as *conceptual navigation, concept drift, interaction metaphors* (e.g., [47]), and so forth.

*Analytics:* analytics over Big Data (cubes [22]) represent a topic of emerging interest for the Database and Data Warehousing research community in this case, there exist several problems to be investigated, running from how to design an analytical process over OLAP data cubes computed on top of Big Data to how to optimize the execution of so-obtained analytical processes, and from the seamless integration of OLAP (Big) data cubes with other kinds of unstructured information (within the scope of analytics).

*Integration with classical data-intensive platforms:* an important issue is represented by how to *integrate* models, techniques, algorithms and computational platforms devised for OLAP over Big Data with classical data-intensive platforms, in the view of a seamless vision of comprehensive large-scale data-intensive systems.

 

*Development tools:* last but not least, *suitable tools* for supporting the design and the development of OLAP data cubes over Big Data, according to a methodology able of incorporating all the aspects discussed above, represent a non-secondary challenge to be dealt with.

### 2.2 Open Problems on Mining Big Data

Currently, a wide family of open research problems in the are of Mining Big Data exists. These open problems are inspired by both methodological and practical issues, with particular regards to algorithm design and performance aspects. Here, we discuss some of them.

The first problem to be investigated in the Mining Big Data area is just represented by the issue of *understanding Big Data* (e.g., [46]), as an initial step of any arbitrary mining technique over Big Data. Indeed, in real-life application scenarios, scientists and annalists first need to understand Big Data repertories, so that being capable of capturing and modeling their intrinsics features such as *heterogeneity* (e.g., [49]), *high-dimensionality* (e.g., [22]), *uncertainty* (e.g., [45]), *vagueness* (e.g., [13]) and so forth. This problem has been recently investigated, and several proposals appeared in literature.

Another relevant issue is represented by the problem of dealing with the *streaming nature* of Big Data. Indeed, a very high percentage of actual Big Data repositories are generated by streaming data sources. To give some examples, it suffices to think of Twitter data, or sensor network data, and so forth. In all these application scenarios, data are collected in a *streaming form,* hence it is natural to imagine novel methods for *effectively and efficiently acquiring Big Data streams* despite their well-known *3V* nature. Here, algorithms and techniques need to deal with some well-understood properties of Big Data streams (e.g., [3]), such as *massive volumes, multi-rate arrivals, hybrid behaviors,* and so forth.

*Big Graph Mining* (e.g., [35]) is another critical research challenge in the Mining Big Data research area. This essentially because graph data arise in a number of real-life applications, ranging from *social networks* (e.g., [38]) to *biomedical systems* (e.g., [43]), and so forth. In this specialized research context, one of the most relevant issue to be faced-off is represented by *devising effective and efficient algorithms that scale well on large Big Graph instances* (e.g., [41]). Similarly to the previous research aspect, *mining Big Web Data* (e.g., [34]) is playing a leading role in the community, due to the fact that Big Web Data are very relevant in the actual Web and expose a very wide family of critical applications, such as *Web advertisement* and *Web recommendation.*

*Privacy-Preserving Big Data Mining* is another important problem that deserves significant attention. Basically, this problem refers to the issue of *mining Big Data while preserving the privacy of data sources that input the target (Big) Data Mining task.* This problem has attracted a great deal of attention from the research community, with a variegate class of proposals (e.g., [54]), and, in particular, it has been addressed via both *exact and probabilistic approaches.*

 

# 3 Future Research Perspectives

## 3.1 Trends in OLAPing Big Data

From the analysis of open research problems of Data Warehousing and OLAP over Big Data, several future research directions to be considered turn out. Among these, we highlight the following ones.

*Innovative Methodologies for Designing OLAP Data Cubes over Big Data.* There is a strong need for methodologies capable of dealing with requirements posed by designing and modeling OLAP data cubes over Big Data.

*Innovative Solutions for Computing Aggregations.* Computing aggregations, which is an annoying problem in classical Data Warehousing and OLAP research, gets worse when considered in the context of OLAP data cubes over Big Data from this, it follows an emerging need for innovative solutions capable of dealing with challenging requirements of OLAP (Big) data cubes, such as *curse of dimensionality, irregular data sets* (e.g., by numerousness of dimensional members), *multi-way aggregations,* and so forth.

*Novel Computational Paradigms for Effectively and Efficiently Computing OLAP Data Cubes over Big Data.* Computing OLAP data cubes over Big Data is very resource-consuming, hence computational paradigms for innovative aspects (e.g., *context-aware resource scheduling (*e.g., [50]) are necessary to this end.

*Powerful High-Performance Architectures for Implementing OLAP Data Cubes over Big Data.* The exploitation of hardware solutions for supporting the implementation of OLAP data cubes over Big Data (e.g., GPU [48]) is a promising direction for next generation scientific computing applications.

*Complex OLAP Data Cubes over Big Data.* Due to the intrinsic complexity of Big Data sets, it follows the need for defining and exploiting *complex* OLAP data cubes over Big Data, tailored to support advanced data-intensive large-scale scientific applications (e.g., [21]).

*Customizable MDX Predicates.* OLAP data cubes over Big Data must also be*flexible,* due to the requirements posed by modern scientific applications in this respect, the classical MDX language for querying multidimensional data should be extended as to incorporate *customizable predicates* catching the necessary flexibility in OLAP (Big) data cube processing.

*Semantically-Rich OLAP (Big) Data Cubes.* Exploiting *semantics-based techniques* for modeling classical OLAP data cubes has been a successful experience in the context of next-generation complex information systems similarly, we believe that these techniques can provide significant achievements in the context of OLAP (Big) data cubes as well, due to the fact semantics-based methods (e.g., *Ontologies [37,36])* can improve the access, browsing and delivery experiences over such data cubes.

*Process-Oriented Definition Languages for Analytics.* Analytics define complex functions over very-large amounts of data, even with the exploitation of modern *NoSQL platforms* [5] this activity may turn to be very problematic when OLAP (Big) data cubes are considered so that it follows the need for devising rigorous process-oriented definition languages for analytics over OLAP (Big) data cubes with the goal of achieving *standardization and interoperability.*

 

*Security and Privacy Issues.* Security and privacy aspects, which are relevant for classical OLAP data cubes as well (e.g., [18, 19, 15]), play a very relevant role for the case of OLAP data cubes over Big Data, due to the fact these structures are open by definition as a consequence, future efforts must focus on the emerging requirement of enforcing and emphasizing the security and the privacy of OLAP (Big) data cubes in open information systems and over the Cloud.

*Applications.* Finally, due to the same nature of OLAP (Big) data cubes, applications over these structures play a first-class role mostly, future research efforts should be focused on testing the effectiveness and the reliability of OLAP (Big) data cubes in a wide range of application scenarios ranging from bio-medical applications to social networks, from Cloud-enabled systems to sensor- and stream-based frameworks (e.g., [16, 13]), and so forth.

### 3.2 Trends in Mining Big Data

Mining Big Data is an emerging research area, hence a plethora of possible future research directions arise. Among these, we recognize some important ones, and we provide a brief description in the following.

*Massive (Big) Data.* Dealing with massive Big Data repositories, and hence achieving *scalable data processing solutions,* is very relevant for Big Data Mining algorithms. Indeed, this requires to investigate innovative indexing data structures as well as innovative data replication and summarization methods.

*Heterogenous and Distributed (Big) Data.* Big Data Mining algorithms are *likely* to execute over Big Data repositories that are *strongly heterogenous in nature, and even distributed.* This calls for innovative paradigms and solutions for adapting classical Data Mining algorithms to these novel requirements.

*Analytics over Big Data.* Big Data repositories are a surprising source of knowledge to be mined (e.g., [1,22, 30, 14]). Unfortunately, classical *algorithmic-oriented solutions* turn to be poorly effective to such goal, while, by the contrary, analytics over Big Data, which argue to support the knowledge discovery process over Big Data via *functional-oriented solutions* (still incorporating algorithmic-oriented ones as basic steps) seems to be the most promising road to be followed in this context.

*Big Data Visualization and Understanding.* Not only effectively and efficiently mining Big Data is a strong requirement for next-generation research in this area, but also *visualizing and understanding mining results* over Big Data repositories (e.g., [7]) will play a critical role for future efforts. Indeed, the special nature of Big Data makes classical data visualization and exploration tools unsuitable to cope with the characteristics of such data (e.g., multi-variate nature, (very) high dimensionality, incompleteness and uncertainty, and so forth).

*Big Data Applications.* We firmly believe that, in the particular context of mining Big Data, *applications over Big Data* will play a critical role for the success of Big Data research. This because they are *likely* to suggest several research innovations dictated by the same *usage* of mining Big Data in real-life scenarios, which will surely uncover challenging research aspects still unexplored. Examples of successful Big Data applications which make use of Data Mining techniques are: *biomedical tools over Big Data* (e.g., [8,44]), *e-science and e-life Big Data applications* (e.g., [6,26]), *intelligent tools for exploring Big Data repositories* (e.g., [10,28]).

# 4 Conclusions

Inspired by recent, emerging trends in Big Data research, in this paper we have provided a brief discussion on most relevant open problems and future directions in the Mining Big Data research area. Our work has puts emphasis on some important research aspects to be considered in current and future efforts in the area, while still opening the door to meaningful extensions of classical Data Mining approaches as to make them suitable to deal with the innovative characteristics of Big Data. Also, one non-secondary aspect to be considered is represented by developing novel Big Data Mining applications in different domains (e.g., Web advertisement, social networks, biomedical tools, and so forth), which surely will inspire novel and exciting research challenges.

# References

1. A. Abouzeid, K. Bajda-Pawlikowski, D. J. Abadi, A. Rasin, and A. Silberschatz. Hadoopdb: An architectural hybrid of mapreduce and dbms technologies for analytical workloads. *PVLDB,* 2(1):922–933, 2009.
2. D. Agrawal, S. Das, and A. El Abbadi. Big data and cloud computing: current state and future opportunities. In *EDBT,* pages 530–533, 2011.
3. X. Amatriain. Mining large streams of user data for personalized recommendations. *SIGKDD Explorations,* 14(2):37–48, 2012.
4. L. Bellatreche, A. Cuzzocrea, and S. Benkrid. Effectively and efficiently designing and querying parallel relational data warehouses on heterogeneous database clusters: The f&a approach. *J. Database Manag.,* 23(4):17–51, 2012.
5. R. Cattell. Scalable sql and nosql data stores. *SIGMOD Record,* 39(4):12–27, 2010.
6. Y.-W. Cheah, S. R. Canon, B. Plale, and L. Ramakrishnan. Milieu: Lightweight and configurable big data provenance for science. In *BigData Congress,* pages 46–53, 2013.
7. K. Chen. Optimizing star-coordinate visualization models for effective interactive cluster exploration on big data. *Intell. Data Anal.,* 18(2):117–136, 2014.
8. X. Chen, H. Chen, N. Zhang, J. Chen, and Z. Wu. Owl reasoning over big biomedical data. In *BigData Conference,* pages 29–36, 2013.
9. Y. Chen, S. Alspaugh, and R. H. Katz. Interactive analytical processing in big data systems: A cross-industry study of mapreduce workloads. *PVLDB,* 5(12):1802–1813, 2012.
10. D. Cheng, P. Schretlen, N. Kronenfeld, N. Bozowsky, and W. Wright. Tile based visual analytics for twitter big data exploratory analysis. In *BigData Conference,* pages 2–4, 2013.
11. J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton. Mad skills: New analysis practices for big data. *PVLDB,* 2(2):1481–1492, 2009.
12. A. Cuzzocrea. Providing probabilistically-bounded approximate answers to non-holistic aggregate range queries in olap. In *DOLAP,* pages 97–106, 2005.
13. A. Cuzzocrea. Retrieving accurate estimates to olap queries over uncertain and imprecise multidimensional data streams. In *SSDBM,* pages 575–576, 2011.
14. A. Cuzzocrea. Analytics over big data: Exploring the convergence of data warehousing, olap and data-intensive cloud infrastructures. In *COMPSAC,* pages 481–483, 2013.

 

15. A. Cuzzocrea and E. Bertino. Privacy preserving olap over distributed xml data: A theoretically-sound secure-multiparty-computation approach. *J. Comput. Syst. Sci.,* 77(6):965–987, 2011.
16. A. Cuzzocrea and S. Chakravarthy. Event-based lossy compression for effective and efficient olap over data streams. *Data Knowl. Eng.,* 69(7):678–708, 2010.
17. A. Cuzzocrea, R. Moussa, and G. Xu. Olap*: Effectively and efficiently supporting parallel olap over big data. In *MEDI,* pages 38–49, 2013.
18. A. Cuzzocrea, V. Russo, and D. Saccà. A robust sampling-based framework for privacy preserving olap. In *DaWaK,* pages 97–114, 2008.
19. A. Cuzzocrea and D. Saccà. Balancing accuracy and privacy of olap aggregations on data cubes. In *DOLAP,* pages 93–98, 2010.
20. A. Cuzzocrea, D. Saccà, and P. Serafino. A hierarchy-driven compression technique for advanced olap visualization of multidimensional data cubes. In *DaWaK,* pages 106–119, 2006.
21. A. Cuzzocrea, D. Saccà, and P. Serafino. Semantics-aware advanced olap visualization of multidimensional data cubes. *IJDWM,* 3(4):1–30, 2007.
22. A. Cuzzocrea, I.-Y. Song, and K. C. Davis. Analytics over large-scale multidimensional data: the big data revolution! In *DOLAP,* pages 101–104, 2011.
23. J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. *Com-mun. ACM,* 51(1):107–113, 2008.
24. F. K. H. A. Dehne, T. Eavis, and A. Rau-Chaplin. The cgmcube project: Optimizing parallel data cube generation for rolap. *Distributed and Parallel Databases,* 19(1):29–62, 2006.
25. F. K. H. A. Dehne, Q. Kong, A. Rau-Chaplin, H. Zaboli, and R. Zhou. A distributed tree data structure for real-time olap on cloud architectures. In *BigData Conference,* pages 499–505, 2013.
26. A. G. Erdman, D. F. Keefe, and R. Schiestl. Grand challenge: Applying regulatory science and big data to improve medical device innovation. *IEEE Trans. Biomed. Engineering,* 60(3):700–706, 2013.
27. W. Fan and A. Bifet. Mining big data: current status, and forecast to the future. *SIGKDD Explorations,* 14(2):1–5, 2012.
28. N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Trans. Vis. Comput. Graph.,* 19(12):2149–2158, 2013.
29. J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh. Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub totals. *Data Min. Knowl. Discov.,* 1(1):29–53, 1997.
30. H. Herodotou, H. Lim, G. Luo, N. Borisov, L. Dong, F. B. Cetin, and S. Babu. Starfish: A self-tuning system for big data analytics. In *CIDR,* pages 261–272, 2011.
31. W. Hummer, B. Satzger, and S. Dustdar. Elastic stream processing in the cloud. *Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery,* 3(5):333–345, 2013.
32. D. Jiang, B. C. Ooi, L. Shi, and S. Wu. The performance of mapreduce: An in-depth study. *PVLDB,* 3(1):472–483, 2010.
33. K. Kambatla, G. Kollias, V. Kumar, and A. Grama. Trends in big data analytics. *J. Parallel Distrib. Comput.,* 74(7):2561–2573, 2014.
34. U. Kang, L. Akoglu, and D. H. Chau. Big graph mining for the web and social media: algorithms, anomaly detection, and applications. In *WSDM,* pages 677–678, 2014.
35. U. Kang and C. Faloutsos. Big graph mining: algorithms and discoveries. *SIGKDD Explorations,* 14(2):29–36, 2012.
36. S. Khouri and L. Bellatreche. Dwobs: Data warehouse design from ontology-based sources. In *DASFAA (2),* pages 438–441, 2011.

37. S. Khouri, L. Bellatreche, and N. Berkani. Modetl: A complete modeling and etl method for designing data warehouses from semantic databases. In *COMAD*, page 113, 2012.
38. H.-C. Kum, A. Krishnamurthy, A. Machanavajjhala, and S. C. Ahalt. Social genome: Putting big data to work for population informatics. *IEEE Computer*, 47(1):56–63, 2014.
39. D. Laney. 3D data management: Controlling data volume, velocity, and variety. Technical report, META Group, February 2001.
40. J. Lin and D. V. Ryaboy. Scaling big data mining infrastructure: the twitter experience. *SIGKDD Explorations*, 14(2):6–19, 2012.
41. Z. Lin, D. H. P. Chau, and U. Kang. Leveraging memory mapping for fast and scalable graph computation on a pc. In *BigData Conference*, pages 95–98, 2013.
42. F. J. Meng, X. Zhuo, B. Yang, J. M. Xu, P. Jin, A. Apte, and J. Wigglesworth. A generic framework for application configuration discovery with pluggable knowledge. In *IEEE CLOUD*, pages 236–243, 2013.
43. A. O'Driscoll, J. Daugelaite, and R. D. Sleator. 'big data', hadoop and cloud computing in genomics. *Journal of Biomedical Informatics*, 46(5):774–781, 2013.
44. M. Paoletti, G. Camiciottoli, E. Meoni, F. Bigazzi, L. Cestelli, M. Pistolesi, and C. Marchesi. Explorative data analysis techniques and unsupervised clustering methods to support clinical assessment of chronic obstructive pulmonary disease (copd) phenotypes. *Journal of Biomedical Informatics*, 42(6):1013–1021, 2009.
45. J. Pei. Some new progress in analyzing and mining uncertain and probabilistic data for big data analytics. In *RSFDGrC*, pages 38–45, 2013.
46. D. J. Power. Using 'big data' for analytics and decision support. *Journal ofDecision Systems*, 23(2):222–228, 2014.
47. S. Sarawagi, R. Agrawal, and N. Megiddo. Discovery-driven exploration of olap data cubes. *InEDBT*, pages 168–182, 1998.
48. E. A. Sitaridi and K. A. Ross. Ameliorating memory contention of olap operators on gpu processors. In *DaMoN*, pages 39–47, 2012.
49. Y. Sun and J. Han. Mining heterogeneous information networks: a structural analysis approach. *SIGKDD Explorations*, 14(2):20–28, 2012.
50. A. Thusoo, J. S. Sarma, N. Jain, Z. Shao, P. Chakka, N. Zhang, S. Anthony, H. Liu, and R. Murthy. Hive - a petabyte scale data warehouse using hadoop. In *ICDE*, pages 996–1005, 2010.
51. A. N. Toosi, R. N. Calheiros, and R. Buyya. Interconnected cloud computing environments: Challenges, taxonomy, and survey. *ACM Comput. Surv.*, 47(1):7, 2014.
52. M. Weidner, J. Dees, and P. Sanders. Fast olap query execution in main memory on large data in a cluster. In *BigData Conference*, pages 518–524, 2013.
53. Y. Yuan, X. Lin, Q. Liu, W. Wang, J. X. Yu, and Q. Zhang. Efficient computation of the skyline cube. In *VLDB*, pages 241–252, 2005.
54. X. Zhang, C. Liu, S. Nepal, C. Yang, W. Dou, and J. Chen. Sac-frapp: a scalable and cost-effective framework for privacy preservation over big data on cloud. *Concurrency and Computation: Practice and Experience*, 25(18):2561–2576, 2013.

 