Open issues in Big Data Warehouse design

Sandro Bimonte*

*TSCF, Irstea
9 Avenue Blaise Pascal CS20085, 63178 Aubière, France
sandro.bimonte@irstea.fr

Abstract. Data Warehouse and OLAP systems allow analyzing huge volumes of data represented according to the multidimensional model. In the era of Big Data, NoSQL systems have been proved to be an effective Business Intelligence solution. Some works recently study warehousing and OLAPing Big Data. (Un)Lucky these works exclusively investigate time performance related to the Volume and Velocity features of Big Data. Therefore, in this paper we investigate the impact of other Big Data features: Variety, Veracity and Value on warehousing and OLAP analysis. Then, we go beyond computation performance and we highlight new Big Data Warehouses design issues.

1 Introduction

Data Warehouses (DWs) and OLAP systems allow analyzing huge volumes of data represented according to the multidimensional model, which defines the concept of dimension (the analysis axes) and fact (the analysis subject) (Kimball, 1996). OLAP relational and multidimensional architectures have been widely studied in the last 30 years (Kimball, 1996).

Conceptual, logical and physical design issues have been extensively investigated by academic and industrial communities (Malinowski and Zimányi, 2006), (Kimball, 1996). Nowadays, DWs and OLAP systems have reached a great maturity for the analysis of Small Data (Miller, 2010). They have been successfully applied in several domains such as marketing, health, agriculture, etc.

However, with the advent of Big Data (Davis, 2012)(Media, 2014) (new) analytical possibilities are offered to decision makers for (new) application domains. In the era of Big Data NoSQL systems have been proved to be an effective Business Intelligence solution (Chen et al., 2012). Different types of NoSQL systems exist: Key value, Extensible record, and Document, Graph (Bugiotti et al., 2014) (Stonebraker et al., 2007), (Floratou et al., 2012). A key value database is a collection of data without a schema and organized as a collection of key value pairs. Data is accessed using the key and its value represents data. Extensible record databases represent data with tables where each row can present different attributes (different columns). Document databases store information as documents having a complex structure. In particular, some works recently study warehousing and OLAPing data using NoSQL systems, since they allow scaling in time and space (e.g. (Dehdouh et al., 2014a) (Dehdouh et al., 2014b) (Chevalier et al., 2015a) (Chevalier et al., 2015b)). Although these works show the feasibility of triggering OLAP operators on the top of Big Data Warehouses, several issues remain unexplored (Cuzzocrea et al., 2013) (Cuzzocrea et al., 2011). In this paper, we try to deeply study the main concepts of Data Warehouse and Big Data to provide an adequate new definition of Big Data Warehouse that effectively integrates all main features (the five 'V') of Big Data (Davis, 2012) into Data Warehouses. Indeed, to the best of our knowledge existing works on Big Data Warehouses are limited to Volume and Velocity.

Moreover, we point out open issues related to the design of Big Data Warehouses.

2 Main concepts

In this section we present main concepts of DWs and OLAP (Sec 2.1), and Big Data (Sec 2.2).

2.1 Data warehouse and OLAP

A Data Warehouse has been defined as "*A subject oriented, integrated, time variant and non volatile collection of data in support of management's decision making process*" (Kimball, 1996). In details:

— Subject Oriented: A DW is used to analyze a particular subject area, for example in the retail domain sales are the analysis subject.

— Integrated: A DW integrates data from multiple data sources. For example, two different stores may have different ways of identifying a product, but in the DW there will be only one way of identifying a particular product.

— Time Variant: Historical data are kept in a DW. For example, one can retrieve data from 6 months, 12 months, or even older data.

— Non volatile: Once data is loaded into the DW, it will not be removed or updated. So, warehoused data are historical data.

OLAP systems allow answering multidimensional analytical queries using warehoused data (Kimball, 1996) . Main features of OLAP analysis are:

— Online queries: Queries results should be provided to decision makers under 10 seconds (Minsky, 1993).

— Multidimensional queries: Queries are defined using dimensions and aggregate measures. These aggregated values are considered as believable and high quality data.

— Simple representation: Queries results should be represented using usable simple pivot tables and/or graphical displays. Decision makers are first end users of OLAP queries (Stolte et al., 2008).

— Explorative: Queries are used in a data exploration process. Sometimes decision makers do not know in advance relevant warehoused data (Stolte et al., 2008).

2.2 Big Data

Several definitions of Big Data have been proposed in literature (Emani et al., 2015), such as (Davis, 2012) (Media, 2014), combining from big size to big dimensionality. Nowadays, academic and industrial communities agree to define Big Data using the 5 V: Volume, Variety, Velocity, Value and Veracity. In details:

— Volume: Using huge volumes of data improve analysis since it allows having better models. Therefore, companies collect vast amount of data to improve their decision making process.
— Variety: Data cannot have a predefined schema. By consequence, structured, semi structured and unstructured data could be transparently used for analysis.
— Velocity: Data should be available as soon as possible. In other terms, when new data arrive they should be stored and then analyzed in real time.
— Value: Data should be used to generate economic value.
— Veracity: Data can present quality problems (inconsistency, accuracy, etc.), but their analysis should provide high quality results.

3 What does Big Data Warehouse mean?

In this section we discuss the meaning of the integration of Big Data features in DW and OLAP systems.

In details, we study how the 5 'V' of Big Data are present or not in the actual definition of DW and OLAP.

Value: By definition a DW is a decision support system aiming to take benefit from data. This benefit can be economic, social, organizational, etc. Therefore, DWs and OLAP systems provide additional Value from data sources.

Volume: DW usually refers to huge volume of data. Therefore, particular storage and querying methods have been defined (Stonebraker et al., 2007) (Cuzzocrea et al., 2011). However, the Volume of Big Data refers to a size that makes ineffective existing DBMS's technologies. Therefore, Big DW refers to volume of stored data that makes ineffective existing Relational, Multidimensional and Hybrid architectures. This new feature for Big DWs raises new challenges: scalability and time performance, which concern almost all existing work, and "usability". In this paper we refer to "usability" as the capacity for decision makers to explore for understanding and analyzing voluminous data. Indeed, it has been widely recognized that when data is too much huge (Cuzzocrea and Mansmann, 2009), it is difficult to look for the right (useful) data using classical OLAP operators. This implies that only necessary data should be warehoused.

Velocity: This issue is not explicitly present in the definition of DW. However, usually DWs are composed of historical data, without focusing on real time data. Indeed, warehousing data as soon as possible has been defined as Real time DW. An important literature exists (Fischer et al., 2012). Therefore, Big Data Warehouse should include real time DWs features. The other meaning of Velocity concerns efficient time performance. However, this issue is already present in the definition of DW since they are decision making support systems to support online (i.e. OLAP) analysis.

Variety: An important difference between warehoused data and Big data is the Variety. A DW is a set of integrated uniformed data. In the classical architecture the Extraction Transformation Load (ETL) process integrates and homogenizes different data sources (Kimball, 1996). Data marts are then extracted from this data. Sometimes this process of transforming data can lead to loss of quality. For example, transforming a textual measure (e.g. a pdf bill) into a numerical one can imply some translation errors. Usually, this kind of transformations can generate quality problems, and therefore they are not integrated in the existing ETL processes. Indeed, handling quality into the warehousing process is an open issue (Berrahou et al., 2015). This approach is antonymic with Big Data analytic where the variety of data is considered as a base for analysis methods, which try to take benefit from this variety. Moreover, the Variety is also associated to the flexibility of NoSQL systems, which allow easy integrating new different data sources since they are schemaless systems. OLAP queries on the top of classical DWs are limited to dimensions and measures defined in the DWs. However, to provide effective Big Data analytics, decision makers should be able to integrate any useful information, even if it is not compliant with the predefined measures and members types (we refer to this issue as "multirepresentation data"), and with the multidimensional model (we refer to this issue as "contextual queries").

Veracity: Contrary to Big Data storage policies, warehoused data are high quality data since they are processed using ETL tools, where quality problems have to be solved (Kimball, 1996). By consequence, a Big Data Warehouse should integrate only high quality data issued. Therefore, we can conclude that Veracity is already included in the definition of classical DW.

Therefore, we can define a Big Data Warehouse as a: "Subject oriented, *multirepresentation*, integrated, time variant, non volatile, *necessary*, *voluminous*, *real time* collection of data in support of management's decision making process". In this definition the new terms refer to the different 'V' of Big Data as previously discussed:

— *Necessary, Voluminous for Volume*
— *Multirepresentation for Variety*
— *Real time for Velocity*

Once the data storage redefined, also the definition of OLAP analysis should be adapted. In this way, we define: "*OLAP over Big DW as an approach to answering multidimensional, online, analytical, usable and contextual queries*". In this definition the new terms refers to the different 'V' of Big Data as previously discussed:

— *Usable for Volume*
— *Contextual for Variety*

4 Challenges and opportunities in NoSQL Big Data Warehouses design

In this section, we present open issues related to the design of NoSQL Big DWs according to our definition.

4.1 Logical modeling

Several works investigate the implementation of DWs into Relational OLAP (ROLAP) systems. Dimensions and facts are translated into particular well known logical models: star and snowflake schemas. However, ROLAP implementations suffer from scaling up to very large data volumes (i.e. "Big Data") (Stonebraker et al., 2007), and handling massive real time data. Therefore, nowadays some authors study the logical modeling of transactional systems using NoSQL systems (Bugiotti et al., 2014), and also OLAP systems. In (Zhao and Ye, 2013), the authors implement a DW in the HBase column oriented store system. Moreover, they propose an implementation of OLAP queries using MapReduce like functions. (Chevalier et al., 2015a) formalizes this implementation. Document based NoSQL systems for DWs implementation are investigated in (Chevalier et al., 2015a). However, contrary to the ROLAP context where several works address complex multidimensional models (Malinowski and Zimányi, 2006), these logical NoSQL models address only simple multidimensional design issues. Indeed, real DW projects are characterized by complex dimensions and facts (Malinowski and Zimányi, 2006) (Iftikhar and Pedersen, 2014), such as non strict or non covering hierarchies, non onto hierarchies, and multigranular facts. Therefore, *in our opinion the definition of logical models for NoSQL DWs supporting complex multidimensional models is an open promising issue*.

Moreover all those works do not support the variety of members and measures as defined at Section 3. Therefore, *some approaches that investigate complex DWs (spatial, multimedia, textual) and multirepresentation of data* (Bédard et al., 2002) should represent a possible base for handling Big Data's Variety in Data Warehouse.

As DWs integrate several data sources, it appears evident that different NoSQL databases can be useful at the same time, for example for handling a slice operator using a graph predicate on a graph database, and a rollup operator of a text measure using a document database. A similar approach has been previously adopted for spatial relational DWs for handling GIS and OLAP queries at the same time (Gómez et al., 2008). Solving this issue is not trivial. Two scenarios should be studied. The first one consists in the transformation of a multidimensional model of a NoSQL database family into another one (Chevalier et al., 2015a), and the second one is to provide a mediation mechanism among multidimensional models of different NoSQL database families inside the same DW. For the first scenario, the transformation can imply losses of analysis capabilities (for example: is (and how) it possible to provide OLAP graph queries on the top of document database?). For the second scenario, a mechanism for translating the same OLAP queries in different query languages should be provided. This mediation process will also have consequences of performances. These questions should be necessary solved.

ROLAP logical models are conceived to support dimensions and facts representation, and at the same time to improve performance of OLAP queries, for example by denormalizing dimensions in the star schema (Kimball, 1996). Therefore, since a de facto standard for NoSQL logical model does not exist, then a dedicated benchmark for comparing the different proposal should be proposed. In (Dehdouh et al., 2014b), the authors propose a dedicated benchmark for NoSQL systems. However, this benchmark does not fit with our vision of Big Data Warehouse. Therefore, the *definition of OLAP workloads and logical models (i.e. benchmarks) (Cuzzocrea and Moussa, 2013) for Big Data Warehouse represent a mandatory future work*.

Finally, NoSQL systems are characterized by the absence of Integrity Constraints (ICs). ICs are defined as rules allowing the correctness of data. RDBMSs natively implement a set of ICs on tables attributes (not null, primary key, etc.) and relations references (e.g. foreign key). Others ICs can be easily implemented using triggers. In the context of relational DWs, ICs are defined on data using RDBMSs ICs and triggers, on aggregated values using triggers, and on visualization using OLAP client visualization policies (Boulil et al., 2014). Nowadays, some recent works investigate the implementation of ICs on the top of NoSQL DBMS. (Georgiev, 2013) implements using Map Reduce documents references ICs. (Curé et al., 2011) define a mapping framework, with an associated query language, for relational and NoSQL DBMS. Some other works study the translation from conceptual/relation schema to documents (Chevalier et al., 2015a). However, to the best of our knowledge, no work provides a mechanism to grant that the warehoused data is conform to the multidimensional model in schemaless DBMSs. In other terms, the *definition of ICs granting well formed warehoused data is an open issue* (how and where (in the NoSQL system or in the ETL system) these ICs should be implemented?).

4.2 Implementation

On one hand several efforts have been provided to address performance issues, leading to mature storage and computing technologies. On the other hand, the design of Big Data Warehouse has been few investigated. Therefore, in the same way as relational DWs (Mazón and Trujillo, 2009), we believe that Big Data Warehouses can also benefit from software engineering technologies, such as Model Driven Architecture (MDA). MDA is an approach for software development by using models. This framework separates the specification of system functionality in a Platform Independent Model (PIM) from the specification of the implementation of that functionality on a specific technology in a Platform Specific Model (PSM). Furthermore, the system requirements are specified in a Computation Independent Model (CIM). MDA not only allows the development of these models in a formal and integrated way by using a standard notation, but also the specification of model transformations in order to obtain the final software product.

As defined in (Mazón and Trujillo, 2009) in the context of relational DW, the PIM is represented by the conceptual multidimensional model (Mazón and Trujillo, 2009). The PSM usually represents the implementation of the relational star and snowflake schemas. Therefore, the PSM implementation is achieved by SQL scripts in a particular DBMS. Several conceptual (PIM) models for DWs have been proposed in the last years (Boulil et al., 2015). Several works propose an automatic transformation of those conceptual models to relational logical and physical models (PSMs) (Boulil et al., 2015), using sometime MDA (Mazón and Trujillo, 2009). Indeed, automatic implementation allows an error free implementation, which is translated into economic gains. In the context of NoSQL Big data warehouse no work studies the use of MDA for two main reasons: i) NoSQL technologies reached maturity on the last years, and ii) each NoSQL DBMS presents a particular PSM model with an associated ad hoc query language, which implies that there is not a direct mapping between the logical model and the PSM. By consequence, as defined in (Chevalier et al., 2015b), a new MDA layer should be investigated: the Logical PIM (LPIM), which represents a formal data structure for the multidimensional model which is related to a category of NoSQL database, but independent from the NoSQL DBMSs. Figure 1 shows an example of the layers of an hypothetical MDA implementation for a document DBMS.

4.3 Opportunities

Solving the previous described challenges offers different opportunities.

For first, with the MDA approach, it should be possible design and implement *new AGILE prototyping OLAP tools*. Indeed, it has been widely recognized that DWs implementation can effectively benefit from prototyping methodologies and tools (Bimonte et al., 2013) to reduce engineering efforts, which correspond to important time and economic gains.

Secondly, well established multidimensional NoSQL logical models will make possible to define *complete OLAP architectures*. Indeed, classical Relational OLAP architectures are com-

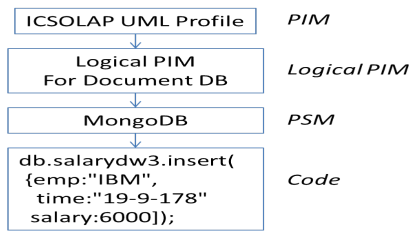

FIG. 1 – *MDA for NoSQL data warehouses*

posed of three tiers: the data warehouse tier where data is stored using a relational DBMS according to the star or snowflake schema, the OLAP server that translates MDX queries in SQL queries executed in the data warehouse tier; and the OLAP client that allows decision makers to visualize and trigger MDX queries by simply interacting with pivot tables and graphic displays. In this sense, well established multidimensional logical models will permit to industrial OLAP server providers to define transformation rules of MDX queries into NoSQL queries. Nowadays, some commercial OLAP suites, such as Pentaho, support this feature but the NoSQL DW has to be modeled as a relational one.

Thirdly, reference multidimensional NoSQL logical models could be used as base for the definition of *physical structures* such indexes, materialized views to speed up OLAP queries. Indeed, in the context of classical DWs several works propose indexes (such as bitmap, etc.) to improve response time of group by SQL queries on the top of the star and snowflake schemas.

5 Conclusion

Data Warehouse and OLAP systems allow analyze huge volume of data represented according to the multidimensional model. In the era of Big Data, NoSQL systems have been proved to be an effective solution for Business Intelligence. Some works recently study warehousing and OLAPing data stored using NoSQL systems, since they allow to scaling in time and space. These works do not take into account all features of Big Data (the 5 'V'). By consequence in this vision paper, we present our new definition of Big Data Warehouse. Moreover, we present the open issues related to the design of Big Data Warehouses. These issues are our current and future work.

6 Acknowledgment

This work has been partially supported by the Irstea's Captiven project of the ANR 11 CNRT 0003 programme "investissements d'avenir" valorisation PME.

References

Berrahou, L., N. Lalande, E. Serrano, G. Molla, L. Berti-Equille, S. Bimonte, S. Bringay, F. Cernesson, C. Grac, D. Ienco, F. L. Ber, and M. Teisseire (2015). A quality-aware spatial data warehouse for querying hydroecological data. *Computers & Geosciences 85*, 126–135.

Bimonte, S., É. Edoh-Alove, H. Nazih, M. Kang, and S. Rizzi (2013). Protolap: rapid OLAP prototyping with on-demand data supply. In *Proceedings of the sixteenth international workshop on Data warehousing and OLAP, DOLAP 2013, San Francisco, CA, USA, October 28, 2013*, pp. 61–66.

Boulil, K., S. Bimonte, and F. Pinet (2014). Spatial OLAP integrity constraints: From uml-based specification to automatic implementation: Application to energetic data in agriculture. *Journal of Decision Systems 23*(4), 460–480.

Boulil, K., S. Bimonte, and F. Pinet (2015). Conceptual model for spatial data cubes: A UML profile and its automatic implementation. *Computer Standards & Interfaces 38*, 113–132.

Bugiotti, F., L. Cabibbo, P. Atzeni, and R. Torlone (2014). Database design for nosql systems. In *Conceptual Modeling - 33rd International Conference, ER 2014, Atlanta, GA, USA, October 27-29, 2014. Proceedings*, pp. 223–231.

Bédard, Y., M.-J. Proulx, S. Larrivee, and E. Bernier (2002). Modeling multirepresentations into spatial data warehouses: A uml-based approach. In *Symposium on Geospatial Theory, Processing and Applications, Ottawa*.

Chen, H., R. H. L. Chiang, and V. C. Storey (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly 36*(4), 1165–1188.

Chevalier, M., M. E. Malki, A. Kopliku, O. Teste, and R. Tournier (2015a). Implementation of multidimensional databases in column-oriented nosql systems. In *Advances in Databases and Information Systems - 19th East European Conference, ADBIS 2015, Poitiers, France, September 8-11, 2015, Proceedings*, pp. 79–91.

Chevalier, M., M. E. Malki, A. Kopliku, O. Teste, and R. Tournier (2015b). Implementation of multidimensional databases with document-oriented nosql. In *Big Data Analytics and Knowledge Discovery - 17th International Conference, DaWaK 2015, Valencia, Spain, September 1-4, 2015, Proceedings*, pp. 379–390.

Curé, O., R. Hecht, C. L. Duc, and M. Lamolle (2011). Data integration over nosql stores using access path based mappings. In *Database and Expert Systems Applications - 22nd International Conference, DEXA 2011, Toulouse, France, August 29 - September 2, 2011. Proceedings, Part I*, pp. 481–495.

Cuzzocrea, A., L. Bellatreche, and I. Song (2013). Data warehousing and OLAP over big data: current challenges and future research directions. In *Proceedings of the sixteenth international workshop on Data warehousing and OLAP, DOLAP 2013, San Francisco, CA, USA, October 28, 2013*, pp. 67–70.

Cuzzocrea, A. and S. Mansmann (2009). OLAP visualization. In *Encyclopedia of Data Warehousing and Mining, Second Edition (4 Volumes)*, pp. 1439–1446.

Cuzzocrea, A. and R. Moussa (2013). Multidimensional database design via schema transformation: Turning TPC-H into the tpc-h*d multidimensional benchmark. In *19th International Conference on Management of Data, COMAD 2013, Ahmedabad, India, December 19-21, 2013*, pp. 56–67.

Cuzzocrea, A., I. Song, and K. C. Davis (2011). Analytics over large-scale multidimensional data: the big data revolution! In *DOLAP 2011, ACM 14th International Workshop on Data*

Warehousing and OLAP, Glasgow, United Kingdom, October 28, 2011, Proceedings, pp. 101–104.

Davis (2012). *Ethics of Big Data: Balancing Risk and Innovation*. O Reilly Media.

Dehdouh, K., F. Bentayeb, O. Boussaid, and N. Kabachi (2014a). Towards an OLAP environment for column-oriented data warehouses. In *Data Warehousing and Knowledge Discovery - 16th International Conference, DaWaK 2014, Munich, Germany, September 2-4, 2014. Proceedings*, pp. 221–232.

Dehdouh, K., O. Boussaid, and F. Bentayeb (2014b). Columnar nosql star schema benchmark. In *Model and Data Engineering - 4th International Conference, MEDI 2014, Larnaca, Cyprus, September 24-26, 2014. Proceedings*, pp. 281–288.

Emani, C. K., N. Cullot, and C. Nicolle (2015). Understandable big data: A survey. *Computer Science Review 17*, 70–81.

Fischer, U., D. Kaulakiene, M. E. Khalefa, W. Lehner, T. B. Pedersen, L. Siksnys, and C. Thomsen (2012). Real-time business intelligence in the MIRABEL smart grid system. In *Enabling Real-Time Business Intelligence - 6th International Workshop, BIRTE 2012, Held at the 38th International Conference on Very Large Databases, VLDB 2012, Istanbul, Turkey, August 27, 2012, Revised Selected Papers*, pp. 1–22.

Floratou, A., N. Teletia, D. J. DeWitt, J. M. Patel, and D. Zhang (2012). Can the elephants handle the nosql onslaught? *PVLDB 5*(12), 1712–1723.

Georgiev, K. (2013). Referential integrity and dependencies between documents in a document oriented database. *GSTF Journal of Computing 2*.

Gómez, L. I., A. A. Vaisman, and S. Zich (2008). Piet-ql: a query language for GIS-OLAP integration. In *16th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2008, November 5-7, 2008, Irvine, California, USA, Proceedings*, pp. 27.

Iftikhar, N. and T. B. Pedersen (2014). Using a time granularity table for gradual granular data aggregation. *Fundam. Inform. 132*(2), 153–176.

Kimball, R. (1996). *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. John Wiley.

Malinowski, E. and E. Zimányi (2006). Hierarchies in a multidimensional model: From conceptual modeling to logical representation. *Data Knowl. Eng. 59*(2), 348–377.

Mazón, J. and J. Trujillo (2009). An MDA approach for the development of data warehouses. In *XIV Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2009), San Sebastián, Spain, September 8-11, 2009*, pp. 208–208.

Media, O. R. (2014). *Big Data Now: 2014 Edition*. O Reilly Media.

Miller, H. (2010). The data avalanche is here. shouldnŠt we be digging? *Journal of Regional Science 50*(1), 181–201.

Minsky, M. (1993). Allen newell, unified theories of cognition. *Artif. Intell. 59*(1-2), 343–354.

Stolte, C., D. Tang, and P. Hanrahan (2008). Polaris: a system for query, analysis, and visualization of multidimensional databases. *Commun. ACM 51*(11), 75–84.

Stonebraker, M., S. Madden, D. J. Abadi, S. Harizopoulos, N. Hachem, and P. Helland (2007).

The end of an architectural era (it's time for a complete rewrite). In *Proceedings of the 33rd International Conference on Very Large Data Bases, University of Vienna, Austria, September 23-27, 2007*, pp. 1150–1160.

Zhao, H. and X. Ye (2013). A practice of TPC-DS multidimensional implementation on nosql database systems. In *Performance Characterization and Benchmarking - 5th TPC Technology Conference, TPCTC 2013, Trento, Italy, August 26, 2013, Revised Selected Papers*, pp. 93–108.

Résumé

Les Entrepôts de données et les systèmes OLAP permettent d'analyser d'énormes volumes de données représentées selon le modèle multidimensionnel. À l'ère du Big Data, les systèmes NoSQL se sont montrés être une solution de Business Intelligence efficace. Certains travaux étudient l'entreposage et lŠanalyse en ligne du Big Data. (Mal)Heureusement ces travaux étudient exclusivement les performances du temps liées au volume et la vélocité du Big Data. Par conséquent, dans cet article, nous étudions l'impact des autres caractéristiques du Big Data : variété, véracité et valeur sur l'entreposage et l'analyse en ligne. Ensuite, nous allons au-delà des problématiques des performances de calcul, et nous mettons en évidence les problématiques ouvertes liées à la modélisation des entrepôts de données de Big Data.