

# **AN APPROACH TO EXTRACTING DISTRIBUTED DATA FROM THE INTEGRATED ENVIRONMENT OF WEB TECHNOLOGIES BASED ON SET THEORY**

Cheikh Ould El Mabrouk and Karim Konaté

Department of Mathematics and Computing, University Cheikh Anta DIOP,  
Dakar - Senegal

## **ABSTRACT**

*The composition and extraction of the information distributed by the different web technologies allow a design of a multifunction imported data environment. However, the problems of integration and communication between the heterogeneous data web always (lodges) on distributed servers.*

*In fact, Web technologies have been proposed to meet certain needs related to heterogeneous and distributed information systems for communicating and exchanging computerized data on the Web. Further more, web generations go through client-server communication (from a static, dynamic and semantic web human client) to server-server communication.*

*For this purpose, Web services are technology application integration by excellence across the Internet. They operate independently of the heterogeneities of the system components on which they are based and are weakly coupled software components interacting with each other.*

*This paper aims to achieve a management approach to extracting and modeling distributed information based on set theory and calculate the execution time of a query to this distributed data.*

## **KEYWORDS**

*Web Technologies, Data, Distribution, Integration, Modeling, Query, Optimization*

## **1. INTRODUCTION**

The distribution of company data across multiple servers requires thinking about creating some homogeneous web technology architectures to integrate data at what distance the company itself from its objects.

Web Technologies are a model for integration systems. This integration model is an important evolution compared to previous models such as Electronic Data Interchange (EDI) that relied on exchange of ownership formats and each Information System. [1] Indeed, the typologies of the Company Information System are generally based on software and heterogeneous data sources and distributed on several computers, consecutive by the use of the various information and communication technologies.

This distribution has considered the problem of integration as a major issue when companies have sought to communicate the different departments or services between them or between other external companies to optimize their business processes of their production chain.

This physical distribution is achieved through a distribution of data and / or applications allowing the implementation at the local level of certain data and / or certain treatments of the information system.

In this context, there are online information systems playing an important role in the distribution of web integrations (e.g., e-commerce, e-learning, e-government and e-health ...) [20].

Companies need to make different applications of their information systems. The techniques of cooperation between these various proprietary systems can be studied as an information system whose content is heterogeneous and distributed.

These interaction or cooperation techniques between business processes and applications are generally referred to as computer data integration techniques within an organization's Information System. For example:

- An application A which periodically and / or regularly needs access to the data of application B;
- An application A which requires a function of the application B;
- A and B applications that need to share a set of data to avoid redundancies;
- Applications A and B which must collaborate by data exchange;
- Applications A and B which are orchestrated by an ad hoc application (workflow engine) to carry out a business process;
- etc. [21]

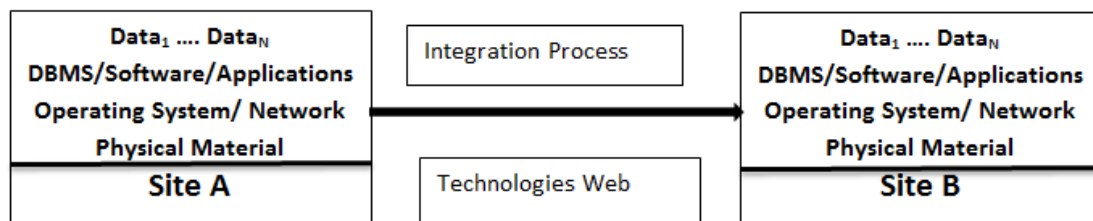


Figure 1: integration and exchange of data and applications

Integrated environment is an interdisciplinary approach which has concepts now on which it is possible to build new data and applications. It's a collaborative and interdisciplinary composition process, supporting on knowledge, methods and techniques resulting from the sciences and experiment put in to define a system which satisfies a need identifies, and is acceptable for the environment, while seeking has to balance total economy of the solution, on all the aspects of the problem in all the phases of the development and the life of the system. Integrated environment concept for the complex problems passes by their decomposition under more limiting problems to which one can bring a solution. [23]

In this paper, we make a breakdown of our work as follows: the first part is this introduction.

The second part illustrates the state of art on the evolution of web technologies. The third part of this paper focuses primarily on a conception of mathematical models of set theory. The fourth part makes the proposed procedure for managing and optimizing distributed data queries.

Finally, the fifth part presents the conclusion and the report of the work of this paper.

## 2. STATE OF THE ART ON THE EVOLUTION OF WEB TECHNOLOGIES

The Web which had been developed at CERN in Switzerland in 1989 is a phenomenal source of information. It is a hypermedia information system allowing access to the resources of the

internet network. Which supports heterogeneous and distributed multimedia documents (text, image, video,...). The web today leads a new normative production mode of information exchange in the world, more than a billion connected individuals. [2]

The web has known a remarkable evolution in recent years which have passed by the three generations below:

- The Static Web produce by HTML technologies the contents of the sites do not change and remain the same on each page when using a browser, the webmaster the only one authorized to modify the contents.
- The Dynamic Web: set of technologies (HTML + CGI: Common Gateway Interface) allowing Internet users to produce, publish, broadcast; share any type of multimedia information content.
- The Semantic Web: set of technologies (HTML + Data models + Server script language) to create a web page anywhere in the world and share with anyone. What has the merit of being practical, simple, useful, fast and effective...

The table below summarizes the main existing web technologies of each type of generation and the date of publication:

Table 1: Evolution of technology web

<b>web technologies</b>	<b>type of generation</b>	<b>the date of publication</b>
HTML	The Static Web	1989
(X)HTML	The Static Web	1992
CSS	The Static Web	1996
PERL	The Dynamic Web	1986
PHP	The Dynamic Web	1994
COLDFUSION	The Dynamic Web	1995
JAVA	The Dynamic Web	1995
ASP	The Dynamic Web	1996
JSP	The Dynamic Web	1996
RDF Model	The Semantic Web	2004
RDF SCHEMA	The Semantic Web	2004
OWL	The Semantic Web	2007
SPARQ	The Semantic Web	2008

## 2.1 Web 1.0

It's the first generation of Web applications (Web 1.0) before 2003 [4] [5]. The contents of the multimedia information (text, image, video, sound...) are produced and hosted by a company that owns the site. They are static; the content of the pages is rarely changed. The user is only the reader of the information. The user is passive. However, the appearance of new scripting languages (such as PHP or ASP) coupled with databases, some sites have become dynamic. That content of the multimedia information is managed by a Content Management System (CMS). They allow multiple individuals to work and edit information on a single document. These technologies also allow the separation of management from form and content. [6]

## 2.2 Web 2.0

The second generation of web applications (Web 2.0) after web 1.0 [from 2003 to 2008] [4] [5] is an information system visible all over the world and in which any user can be active. Internet users became actors of all types of multimedia content (text, image, video, sound...). The author [7] defines the privileges of Web 2.0 by: “Write, comment, copy and paste, mix, publish, share or exchange photos, videos, links and tags, on sites presenting self and its world relational ...” In this context, appeared social networks and the phenomenon of syndication appeared: the famous RSS feeds, AJAX, MySpace, ..

Web 2.0 is an integral part of corporate information and is redefining how organizations and individuals communicate and collaborate with each other for content management and new ways of communicating and interacting with users. [8] [9]

## 2.3 Web 3.0

The third generation of the web produced after web 2.0 from 2008 to today [4] [5]. Web 3.0 could be detached from the servers to promote exchanges between Peer to Peer Internet users, an interactivity that would happen in real time. In addition, Web 3.0 enhances Data mining of distributed and heterogeneous databases, and intelligent search options across different types of multimedia information. [10] As soon as you type the first letter of the search, the Google instantly proposes the list of the most frequently searched terms starting with this letter. In addition, one could simply click on Google Maps and get images that others have uploaded to Flickr.... [3] In this context, a series of new Web services appears (such as Second Life, Twitter, Exalted...). Web 3.0 is considered the next frontier of intelligent Web technology (Web 4.0) that sees the potential of intelligent information and semantic web. [11].

The figure below represents this evolution of the different generations of web and its forward-looking web 4.0 (smart web):

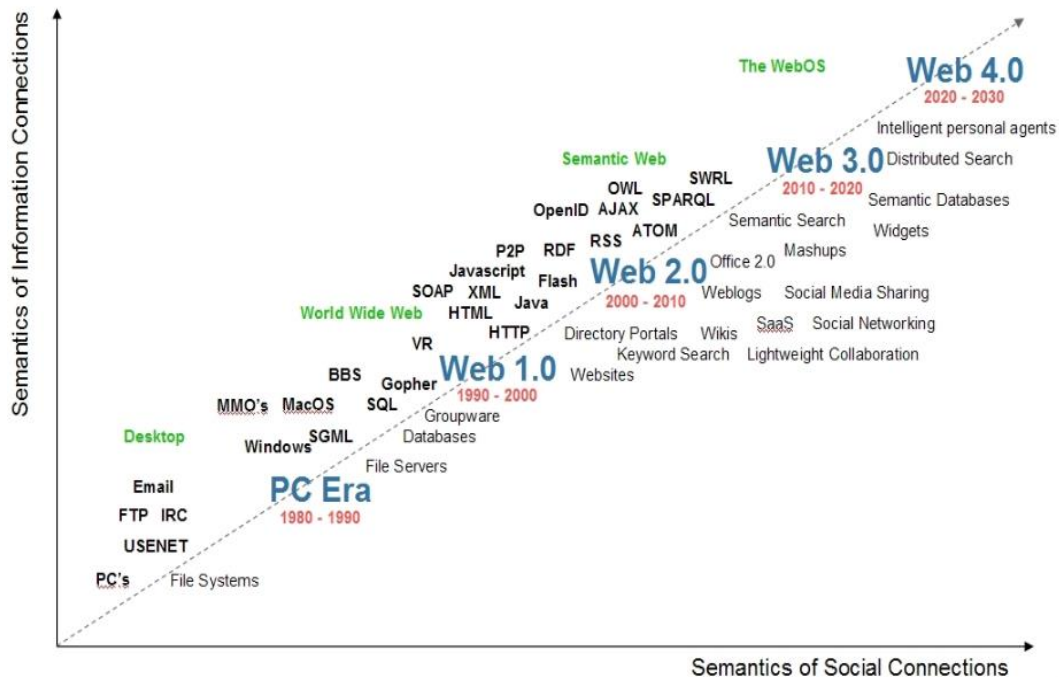


Figure 2: Evolution of Web generations in source time (Source: Nova Spivack Site)

### 3. DESIGN MODEL FOR AN INTEGRATED ENVIRONMENT

#### 3.1 Design of an Information Environment

The computerized data environment is the set of computer files grouped in a directory. It also includes the databases and its models of design and implementation of the company.

In this context, the integration today is involved in the implementation of applications to ensure the development and deployment of data. The integration of information systems has become a necessity due to the fact that large companies have developed over time, for the most part, distributed web information. [17] However, an integrated information environment becomes a development space, offering a way to enrich the value of the operational process. It must manage the movement of their data from different information systems. [18] As for example, a data table looks a lot like rows and columns to a paper chart: it consists of rows of data. Then, tables are the central objects of a design of a computerized data environment. But these are not the only objects that exist in an information system. Several other types of objects can be created in order to make the use and management of data more efficient or practical such as:

- Views, query requests
- Functions, Operators,
- Data Types,
- Domains Triggers and rewrite rules
- Management rules

As a result, the evolution of the tools of an information environment is the design of a new tool or the replacement of an existing tool, based on the import of web data. For this fact, the structure of the integrated environment brings together common object bases and the tools of data integration and reuse of existing objects.

#### 3.2 Mathematical Model

In this section we have made mathematical reminders of set theory and discussion of our model of creating a data environment based on information imported by web technologies:

##### 3.2.1 Recall of Set Theory

**Definition (1):** A mathematical model is a translation of reality in order to apply mathematical tools, techniques and theories, the translation of mathematical results obtained into predictions or operations in the real world.

**Definition (2):** A set A is a subset or part of a set B or B is superset of A, if every element of subset A is also element of superset B.

$$\forall e \in A \implies e \in B$$

**Definition (3):** We call combination of (p) elements of a finite set E  $\{a_1, \dots, a_n\}$  of (n) elements, every subset A of (p) elements of E.

The number of combinations of **p** elements of a set of **n** elements is:  $C_n^p = \frac{n!}{(n-p)!p!}$

### 3.2.2 Discussion of the Model

The model of the creation of a set of n elements expresses how is constituted a relation or a structured file: the number of attributes is a different name for each attribute, the domain in which it takes its values in our constitutive environment by recurrence.

For n = 1, E = {a1} the number of possible non-empty files is unique  $x = 1 = 2^1 - 1$

For n = 2, E = {a1, a2} the possible files are {a1}, {a2}, {a1, a2}  $x = 3 = 2^2 - 1$

For n = 3, E = {a1, a2, a3} the possible files are {a1}, {a2}, {a3}, {a1, a2}, {a1, a3}, {a2, a3}, {a1, a2, a3} the number is  $x = 7 = 2^3 - 1$ .

In general, n elements of the set  $E = \{a_1, a_2, \dots, a_p, \dots, a_n\}$  can construct the non-empty files of the numbers of different properties  $p \ E_p = \{a_i, \dots, a_p\}$

$$C_n^p = \frac{n!}{(n-p)!p!} = \frac{n(n-1)\dots(n-p+1)}{p!}$$

The sum of all possible non-empty files is the sum of all elementary compositions of the set  $E = \{a_1, a_2, \dots, a_p, \dots, a_n\}$  such that  $(i \leq p \leq n)$

$$x = \sum_{p=1}^n C_n^p = \sum_{p=0}^n \frac{n!}{p!(n-p)!} - C_n^0 = 2^n - 1 \quad \text{So, } X = 2^n - 1.$$

As a consequence of this conceptual modeling based on set theory:

n: elements, attributes, properties, columns ...

x: sets, files, tables, relationships, entities ...

The table below gathers the relation between the number of the elements (n) and the number of the possible composition sets (x) and not empty.

Table 2: Result of the composition following set theory

Elements	1	2	3	4	5	...	n
Sets	1	3	7	15	31	...	$2^n - 1$

## 4. DISTRIBUTED INFORMATION MANAGEMENT PROCEDURE

Computerized data is generally stored in web databases distributed on several sites, geographically distant, connected by a network, which are often heterogeneous in technical, syntactic and semantic mode, ... [12] [14]

The meeting and integration of these data by web technologies form a complete database. However, the problems of web data integration are characterized by the heterogeneity of different distributions of web data sources.

Web services are a tremendous technological evolution, ensuring the interoperability of heterogeneous and distributed data on the Internet.

In this section, we have managed the distribution of web information by data warehouse, data fragmentation techniques and distributed query optimization techniques, and finally an algorithm is proposed for computation of the execution time by the composition and the distribution according to the overall set of theory.

#### **4.1 Data Warehouse and Distribution**

The Data Warehouse is a database dedicated to the storage of all data used in decision-making and decision analysis. The Data Warehouse is exclusively for this purpose. It is fed with data from the production bases thanks in particular to ETL Extract Transform Load tools. The Data Warehouse is not just a copy of the production data. It refers to a kind of database used to collect and store temporary information from other databases. Each piece of information collected is assigned a date, or a version number, to avoid overwriting information already in the database and to track the evolution of this information over time. In fact, integration of Data Warehouse requires Extraction, Transformation and Loading processes (ETL). These must consider several data quality related distribution, as for instance duplicated data, possible data inconsistency, creation of new variables using transformations, and the integrated environment etc. [25]

The information collected can come from several databases. Sometimes, information from different databases of an enterprise is collected in a single data warehouse, or there are different breakdowns of data warehouses depending on the subject or business related to each information data (Data Mart) and according to the distribution multi-dimensional (Cube Data). [22]

##### **4.1.1 Subset Distribution (Data Mart)**

A Data Mart (data warehouse) is a subset of a relational database used in business intelligence. It is used in companies to provide information targeted to a specific business and constituting for the latter, a set of indicators to guide an activity and decision support. Data Marts are also called Decision Warehouses.

Data Marts are simplified extracts the detail of the company data, they are of interest only for identified and repetitive requests; it is easier for the system to query a Data Mart that contains only the necessary than to have to identify and sort the entire database. In addition, Data Marts classify and clarify information, so that each business has access to figures corresponding to its functional expectations, without being polluted by contiguous data.

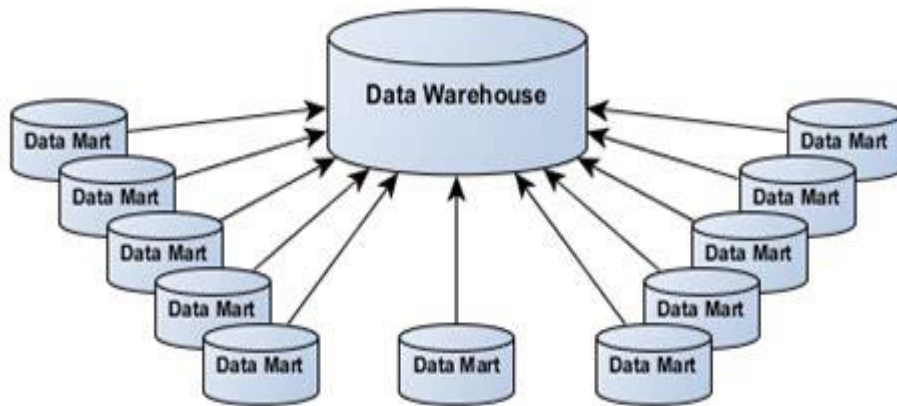


Figure 3: Data Mart

#### 4.1.2 Multidimensional Distribution (Data Cube)

The multidimensional distribution of data makes it possible to observe the data in several views. Its analysis views facilitate access to the data. In this distribution, rows and columns are replaced by dimensions, as representative concepts, and by measures that act as quantitative values. The multidimensional distribution assumes that the major objective is the multidimensional view of the data by the data cube provides a distribution very close to the way the analyst sees and interrogates the data. It organizes the data into one or more dimensions that determine a measure of interest. A dimension indicates how one looks at the data to analyze it, whereas a measurement is an object of analysis. Each dimension is formed by a set of attributes and each attribute can take different values.

Dimensions typically have associated hierarchies that organize attributes at different levels to observe data at different granularities. A dimension can have multiple associated hierarchies, each specifying different order relationships between its attributes. As for example Time {Day, Month, Year}.

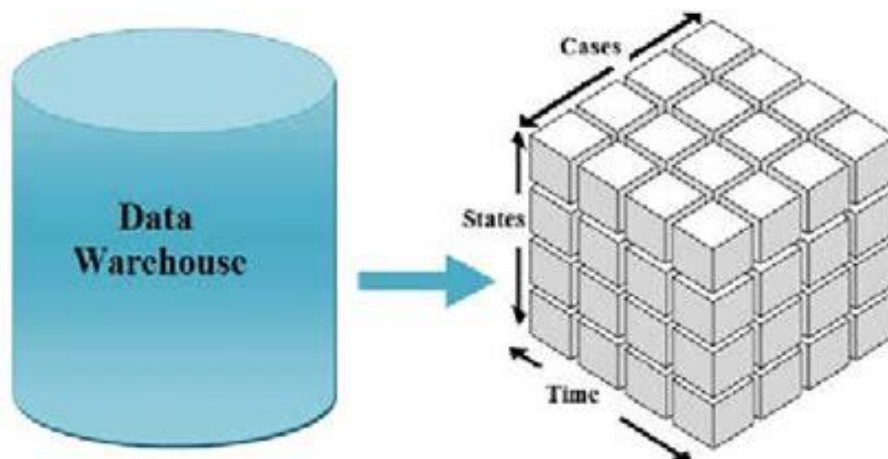


Figure 4: Data Cube



## 4.2 Data Fragmentation Techniques

Web data fragmentation is a distributed data management technique without loss of information, if the logical database can be fully reconstructed from multiple fragments. This technique is expressed using the instructions of the data manipulation language associated with the data model used (for example, algebra or SQL for a relational Database). Indeed, fragmentation consists of division into several fragments (subsets of the relationship) that can be non-disjoint. [12].

In general, a database can be fragmented from different possible segments, both at the representation schema and at the data level. The fragmentation unit determines the type of distribution element of fragmented horizontal, vertical, and hybrid (combination of horizontal and vertical) data.

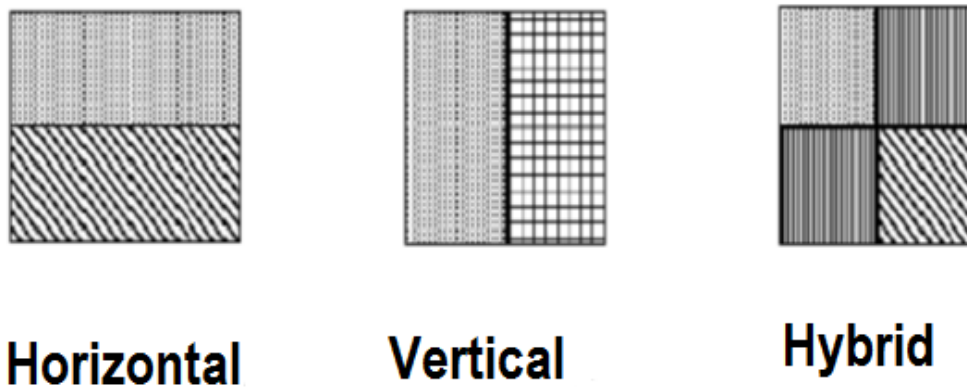


Figure 5: Fragments of Web Data

Fragmentation is the task of dividing a table into a set of smaller tables. The subsets of the table are called fragments. Fragmentation can be of three types: horizontal, vertical, and hybrid. Horizontal fragmentation can further be classified into two techniques: primary horizontal fragmentation and derived horizontal fragmentation. In fact, fragmentation should be done in a way so that the original table can be reconstructed from the fragments. This is needed so that the original table can be reconstructed from the fragments whenever required.

### Advantages of Data Fragmentation

- Since data is stored close to the site of usage, efficiency of the database system is increased.
- Local query optimization techniques are sufficient for most queries since data is locally available.
- Since irrelevant data is not available at the sites, security and privacy of the database system can be maintained.

### Disadvantages of Data Fragmentation

- When data from different fragments are required, the access speeds may be very high.
- In case of recursive fragmentations, the job of reconstruction will need expensive techniques.
- Lack of back-up copies of data in different sites may render the database ineffective in case of failure of a site. [24]

#### 4.2.1 Vertical Fragmentation

In vertical fragmentation, the fields or columns of a table are grouped into fragments. In order to maintain constructiveness, each fragment should contain the primary key field(s) of the table. Vertical fragmentation can be used to enforce privacy of data.

For example, let us consider that a University database keeps records of all registered students in a Student table having the following schema.

*STUDENT (Regd\_No, Name, Course, Address, Semester, Fees, Marks)*

Now, the fees details are maintained in the accounts section. In this case, the designer will fragment the database as follows –

```
CREATE TABLE STD_FEES AS  
SELECT Regd_No, Name, Fees  
FROM STUDENT;
```

#### 4.2.2 Horizontal Fragmentation

Horizontal fragmentation groups the tuples of a table in accordance to values of one or more fields. Horizontal fragmentation should also confirm to the rule of constructiveness. Each horizontal fragment must have all columns of the original base table.

For example, in the student schema, if the details of all students of Computer Science Course needs to be maintained at the School of Computer Science, then the designer will horizontally fragment the database as follows –

```
CREATE COMP_STD AS  
SELECT * FROM STUDENT  
WHERE COURSE = "Computer Science";
```

#### 4.2.3 Hybrid Fragmentation

In hybrid fragmentation, a combination of horizontal and vertical fragmentation techniques are used. This is the most flexible fragmentation technique since it generates fragments with minimal extraneous information. However, reconstruction of the original table is often an expensive task.

Hybrid fragmentation can be done in two alternative ways –

- At first, generate a set of horizontal fragments; then generate vertical fragments from one or more of the horizontal fragments.
- At first, generate a set of vertical fragments; then generate horizontal fragments from one or more of the vertical fragments. [24]

In this paper, we have compared the execution times of a query before and after fragmentation by the results of the author's work [13] [14] [19] [24]. The advantage is a great different on the input / output time and response time based on the minimization of information search space after each horizontal, vertical and hybrid fragmentation. Nevertheless, the disadvantage is the obligation to homogenize the structure of heterogeneous files for horizontal and hybrid fragmentation and the requirement to increase the storage space by duplicating the foreign key of all the files structured for the vertical and hybrid fragmentation.

### 4.3 Optimization of Distributed Queries

In web database query processing, we generate the algebraic tree of the SQL query that we optimize with the same criteria. Each leaf of the tree represents a relation, and each intermediate node represents a logical operation (and, or ...) to interpret the algebraic clauses. The optimization of an algebraic query imposes to minimize the complexity of the execution on the various sites of the distributed data, and then to enrich the results of the distribution of the data on the distributed sites, and on the place (site) where each operation the query must be executed.

Suppose that the optimization of a query (SELECT Attributes (A1, A2, ... An), FROM Relations (R1, R2, ... ..Rn), WHERE (C1, C2, ... .Cn) is a composition of integrated query in web services.

There are therefore no tools yet fewer models to check the integration time of the web data, and the time T remains estimated.

The total execution time of a distributed query is the sum of the CPU calculation time needed to execute the algebraic operations (joins, selections, projections ...) on different relations (R1, R2, ... ..Rn) for each site Si (Ti), with the transfer times (Query Transformation) data between two distinct sites Si and Sj (Tij). So the time T is:

$$T = \sum_{i=1}^n (T_i + \sum_{j=1}^n T_{ij} )$$

Processing consists of executing a set of system tasks simultaneously. The expression of time will be:

$$\begin{aligned} T [\text{site}] &= \max (T_1, T_2, \dots, T_n) \\ T [\text{transfer}] &= \max (T_{i1}, T_{i2}, \dots, T_{in}) \\ T &= T [\text{site}] + T [\text{transfer}]. \end{aligned}$$

The work of the authors [15], [16], preferred to give a comparison of the execution time values of a query before and after optimization in some DBMS. The advantage is that calculating the execution time of a distributed query automatically adjusts its execution plan to the most efficient solution, regardless of different writes. However, the disadvantage is that with this random estimated value, some DBMS would not be able to make some transformations by themselves; the use of the most efficient scripts will increase the transfer times.

### 4.4 The Proposed Algorithm

In this contribution, we proceeded to a new procedure proposed for the problem solving of the distribution: called optimal integration procedure.

As a result of research, we have grouped in a calculation algorithm the execution time of a query of the distributed data with the distribution calculated by the set theory.

#### 4.4.1 Distribution Management Algorithm

```

BEGIN
    ❖ Definition of the query (Relations (R1, R2, ... ..Rn), attributes (A1, A2, ... An),
    conditions (C1, C2, ... .Cn))
    ❖ The X distribution according to the set of theory:

$$x = \sum_{p=1}^n C_n^p = \sum_{p=0}^n \frac{n!}{p!(n-p)!} - C_n^0 = 2^n - 1$$

    FOR i from 1 to x do
        BEGIN
            ❖ Management of the distribution (Ri relationship of the Si site)
            ❖ Algebraic Tree Representation (Root, Nodes, Leaves)
            ❖ Optimization of the request (specific treatment of each site before the grouping)
            ❖ The time required for the aggregation of the sum of all data transfer times between
            Sites (Tij).

            IF simultaneous execution
                FOR i from 1 to n do
                    T[site] = max (T1, T2, ..., Tn)
                    T[transfer] = max (Ti1, Ti2, ..., Tin)

                    T = T[site] + T[transfer].

                    Else  $T = \sum_{i=1}^n (t_i + \sum_{j=1}^n t_{ij})$ 
                END FOR
            END IF
        END
    END FOR
END
    
```

#### 4.4.2 Example of Query

Relations R1(A,B,C,K) , R2(A,D,G), ... , Rn(B,E,F)

Select R1.A, R1.B, R2.D, ..., Rn.E

From R1, R2, ... , Rn

Where (R1.A = R2.A) and (R1.B= Rn.B) and C > c and D = d and E <> e

**Data distribution**

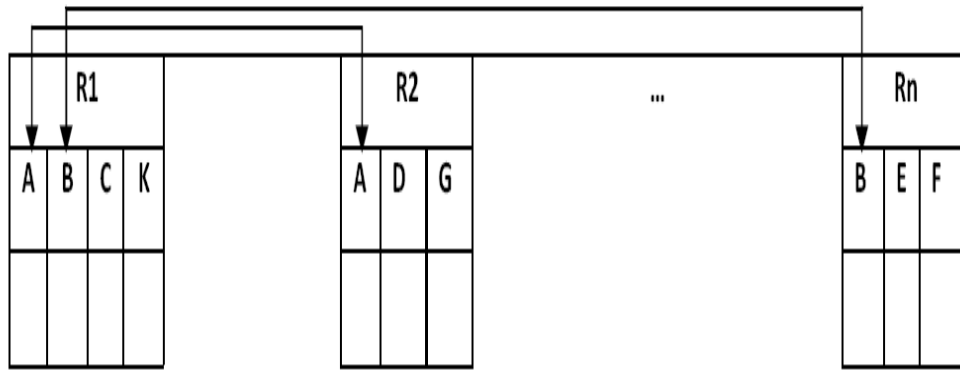


Figure 6: Data Distribution

**Algebraic tree**

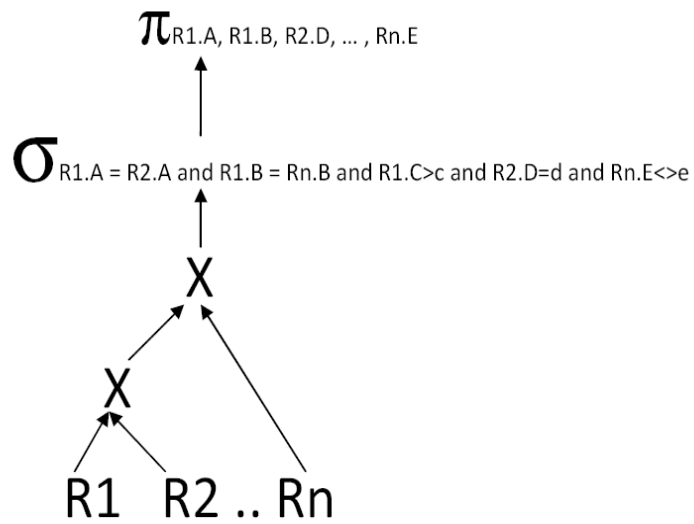


Figure 7: Algebraic Tree of Presentation

## Optimization

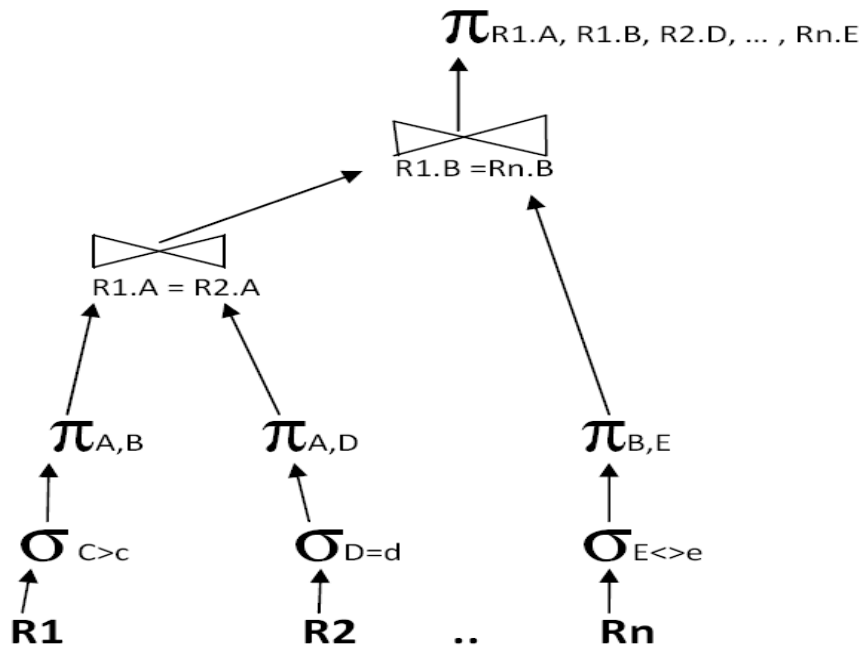


Figure 8: Algebraic Tree Optimization

### 4.4.3 Discussion

The solution we propose offers a distribution management algorithm, which looks like all the processes needed to optimize the integration of distributed data across multiple sites. In our algorithm, we proceeded to the combination of the two optimization approaches: set of theory and algebraic tree processing.

This combination will allow a design of any possible composition of partial gatherings towards the global integration of the information environment. By integrating the identification and description of databases distributed by projection, selection and join properties and assignments that groups together the data environment.

The process begins with the application of set theory, which is considered a preparatory step of the possible division phase, where we will decide whether it is possible or even recurrence to continue the given composition given to an information environment. This step is very important for composing and managing this distribution of a data environment, hence the clear and thorough understanding of the queries of the data distribution, link trees and their optimizations.

In the same run, we calculate the execution time on the time value of grouping by the transfer time between the district sites. From these we estimate the transfer time by the maximum of the simultaneous execution times. This distribution of data bases is realized by the set theory modeling. The final phase is then the integration of the data in the optimal distribution of the information system.

## 5. CONCLUSION AND FUTURE WORK

In this paper, we made a presentation of all developments in web technologies, the different generations and uses of static web, dynamic, semantic, and modes of operation, as well as the most recent tools used for communications between Internet users.

Likewise, we have approached the theory of sets for the composition of an information environment and data imported by web services then, a web data management procedure and the advantages and disadvantages of each optimization technique. This paper ends with a proposed algorithm for calculating the execution time of a query based on the composition generated by set of theory and the data transfer times between the distributed sites. This approach will allow a design of any possible composition of partial gatherings towards the global integration of the information environment.

Our future work will address the security issues of imported data, changing the properties of each system relative to another system, recovering data after a site crash, calculations and concatenations, and exploring virtual properties of distributed database.

We hope that this work will be useful, especially for distributed data, which must formulate and manage integrations between the data distribution of the company, because it is clear that the proposed approach must be supported in the evolution web technologies to manage new sources and data distributed on multiple sites.

## REFERENCES

- [1] Nabil Ahmed Sultan, The Evolving Model for Software Delivery: The Case of Web and Semantic Services International Journal of Web Services Practices, Vol. 3, No.1-2 (2008), pp. 57-65
- [2] Anderson, P. (2007). All That Glisters Is Not Gold' -- Web 2.0 And The Librarian. Journal of Librarianship and Information Science, 39 (4), 195–198.
- [3] Isaías, P., Miranda, P., & Pifano, S. (2009). Critical Success Factors for Web 2.0 – A Reference Framework. Heidelberg: Springer-Verlag Berlin.
- [4] Web 1.0 à Web 3.0 les étapes d'évolution, 2008, L'Atelier Informatique – Référencement – Création de sites <http://www.atelier-informatique.org/internet/evolution-web-10-web-20-web-30/358/>
- [5] Macnamara, Jim (2010). University of Technology Sydney, Public communication practices in the Web 2.0-3.0 mediascape: The case for PRevolution. PRism 7(3): <http://www.prismjournal.org>
- [6] Le monde du web Gérard Savoie, 2010 <http://gerardsavoie.ca/web.html>
- [7] Laurence Allard, Revue MediaMorphoses n°21, septembre 2007, Armand Colin/INA Synthèses sur le Web 2.0
- [8] Sankar, K., & Bouchard, S. A. (2009). Enterprise Web 2.0 Fundamentals. Indianapolis: Cisco Systems, Inc.,.
- [9] Boulakfouf, S., & Zampunieris, D. (2008). Integrating Web 2.0 Features into a Learning Management System. Paper presented at the 7th European Conference on e-Learning (ECEL'08), (pp. 127-133). Cyprus.
- [10] Myhill, M., Shoebriidge, M., & Snook, L. (2009). Virtual research environments – A Web 2.0 cookbook? Library Hi Tech , 27 (2), 228-238.

- [11] Dwivedi, Yogesh; Williams, Michael; Mitra, Amit; Niranjan, Suraj; and Weerakkody, Vishanth, "Understanding advances in web technologies: evolution from web 2.0 to web 3.0" (2011). ECIS 2011 Proceedings. Paper 257.
- [12] Principles of Distributed Database Systems M. T. Ozsu and P. Valduriez.: Second Edition. Prentice Hall, 1999.
- [13] Techniques d'optimisation des requêtes dans les data warehouses Ladjel Bellatreche LISI/ENSMA, 2003
- [14] Optimisation de requêtes OLAP en entrepôts de données Approche basée sur la fragmentation génétique, El houssain ZIYATI, 2010
- [15] Pearson Education France – Optimisation des bases de données – Laurent Navarro, 2010  
[http://www.pearson.fr/resources/titles/27440100612040/extras/2412\\_chap06.pdf](http://www.pearson.fr/resources/titles/27440100612040/extras/2412_chap06.pdf)
- [16] Abdul Mateen, Basit Raza, Mian Muhammad Awais, Muhammad Sher Selection of the Best DBMS: A Fuzzy based Multi-objective Decision Making Approach, Proceedings of the International MultiConference of Engineers and Computer Scientists 2011 vol I IMECS 2011, March 16 – 18, Hong Kong
- [17] D. Litan, M. Velicanu, L. Copcea (Teohari), M. Teohari, A. M. Mocanu (Virgolici), I. Surugiu, and O. Raduta, Business' New Requirement:Information Systems Integration – Methods and Technologies, INTERNATIONAL JOURNAL OF COMPUTERS AND COMMUNICATIONS Issue 3, Volume 5, 2011
- [18] Inter systèmes Ensemble: Intégration pour l'entreprise en évolution rapide d'aujourd'hui, livre blanc, Sandra Rogers, IDC, Décembre 2003.
- [19] BARR Mohamed, BELLATRECHE Ladjel, Approche dirigée par les fourmis pour la fragmentation horizontale dans les entrepôts de données relationnels, Nature & Technologie, page, 16 – 24, 2012
- [20] Web Services in Distributed Information Systems: Availability, Performance and Composition; Zhao Xia, and al.,International Journal of Distributed Systems and Technologies, one (1), 1-16, January-March 2010 1
- [21] Erl T., "Service-Oriented Architecture - A Field Guide to Integrating XMLand Web Services". Printice Hal., 2004.
- [22] A.M. LAW & W.D. KELTON, Simulation Modelling and Analysis. Edition Mac Graw Hill Book Company, 1999
- [23] FROM PLM TO ERP : A SOFTWARE SYSTEMS ENGINEERING INTEGRATION  
Nafisa Osman1 and Abd-El-Kader Sahraoui AlmedtechInc..and University SUST Khartoum, Sudan 2LAAS-CNRS, Université de Toulouse, UT2J, Toulouse, France International Journal of Software Engineering & Applications (IJSEA), Vol.9, No.1, January 2018
- [24] DISTRIBUTED DBMS - DESIGN STRATEGIES  
[HTTPS://WWW.TUTORIALSPOINT.COM/DISTRIBUTED\\_DBMS/DISTRIBUTED\\_DBMS\\_DESIGN\\_STRATEGIES.HTM](https://www.tutorialspoint.com/distributed_dbms/distributed_dbms_design_strategies.htm)
- [25] DATA WAREHOUSE AND BIG DATA INTEGRATION Sonia Ordoñez Salinas and Alba Consuelo Nieto Lemus Faculty of Engineering, Distrial F.J.C University, Bogotá, Colombia, International Journal of Computer Science & Information Technology (IJCSIT) Vol 9, No 2, April 2017