Framework for Client-Server Distributed Database System for Population Census.

Olabode Olatubosun, Ph.D.* and O.C. Akinyokun.

Computer Science Department, Federal University of Technology, Akure, Nigeria, GSM: 08033511257.

*E-mail: olabode_olatubosun@yahoo.co.uk

ABSTRACT

With the great awareness of the growing need for information on population dynamics, a comprehensive study of demography in Nigeria with particular reference to population census and projection is a desirable venture. Existing sources of demographic data in Nigeria is restricted to census, surveys, and special inquires. There are no national systems of registration of vital events at the levels of completeness sufficient to provide useful demographic information. The time taken to summarize, analyze, and publish census results are usually very long. This observable problem can be associated with the conventional approach to census collection taken in Nigeria. In this study, an attempt is made to design a framework for a distributed database system for population census collection in Nigeria. This framework includes a model for the analysis of demographic parameters of respondents in a study area. Basically, the demographic parameter or population variables are: relation to head of household, sex, age, disability, home local government area, home place, address, literacy, educational qualifications, work status, type of employment, sector of employment, marital status, and statistics of marriage. Microsoft Visual Basic® and SQL is considered in the design.

(Keywords: distributed database system, population census, fragmentation, replication, vital registration)

INTRODUCTION

The greatest asset of any nation is her citizenry. The population of a nation's citizens is the net value of the birth and death together with migration. Politically, socially, and economically the strength of a nation is measured by the size of its population.

In Heisel (1969), it is stated that the accurate and up to date information on the component and patterns of change of population is a primary requisite for intelligent decision making in any society. It is a fundamental requirement that any nation should estimate her population size, rate of growth, and the components patterns of fertility and mortality. Planning for growth obviously requires the best possible knowledge of the characteristics and prospects of the target population.

With the great awareness of the growing need for information on population dynamics, a comprehensive study of demography in Nigeria with particular reference to population census is a desirable venture.

On the effort by the Nigerian Government to estimate her population; a lot of money has been expended without any acceptable results by the Federal Government. One reason for its failure is the negative attitude and behavior of the body concerned to adapt to the current trends in information technology today. Information and telecommunication technology is a driving force in development in many facets of life. A population census should not be left behind in the advancement of information technology.

A study of the Nigerian 2006 census suggests that an automated population census is still lacking in Nigeria as of today. To this end, an attempt is made to design a framework for a distributed database system for population census in Nigeria. The framework shall include models for the analysis of demographic parameters of respondents in a study area. Basically, the demographic parameter or population variables are namely: relation to head of household, sex, age, disability, home local government area, home place, address, literacy, educational qualification, work status, type of employment, sector of employment, marital status, and statistics of marriage.

According to Navathe (2000) a distributed database (DDB) as a collection of multiple logically interrelated databases distributed over a computer network, and a distributed database management system (DDBMS) as a software system that manages a distributed database while making the distribution transparent to the user. From the definition of Navathe (2000), the elementary unit of a distributed system is a computer that is networked with other computers. The computer is autonomous in the way it carries out its actions. Computers are linked to one another over a communications network that enables an exchange of messages between computers. The client-server architecture used as a platform for database application development is an approach which has successfully been used to solve some lingering problem in the society today. The Client Server system is adopted in this research work.

PLANNING FOR POPULATION CENSUS

A population census is the total process of collecting, compiling, evaluating, publishing, and disseminating demographic, economic, and social data at a specific period. The essential features of a population census, as stated in Bernard, 1968 are individual enumeration, universality within a defined territory, simultaneity, and defined periodicity. According to Bernard, a national census should cover the country's entire territory and all people resident or present depending upon whether the basis of enumeration is *de jure* or *de facto*. A national census or major survey involves a vast amount of preparatory work, some aspects of which may be in process a number of years before the enumeration date. Preliminary activities should include the following:

a. Geographic work, such as preparing maps and lists of places of living of people

b. Determining the data needs of the national and local governments, business, labor, and the public

c. Deciding on the method of enumeration

d. Choosing the questions to be asked and the tabulations to be made

e. Designing the questionnaire

f. Testing the forms and procedures

g. Planning the data processing procedures

h. Procurement, installation of the equipment to be used for geographical survey, data collection, and storage, retrieval and statistical analysis

The list of recommended items for census questionnaire as presented in United Nations, 2001 are listed below:

1. Place where found at time of census
2. Place of usual residence
3. Place of birth
4. Duration of residence
5. Place of previous residence
6. Place of work
7. Sex
8. Age
9. Relationship to head of household
10. Marital status
11. Age at marriage
12. Duration of marriage
13. Marriage order
14. Children born alive
15. Children living
16. Citizenship
17. Literacy
18. School attendance
19. Educational attainment
20. Educational qualifications
21. Language
22. Religion
23. Type of activity
24. Occupation
25. Status (as employer, employee, etc.)

Various procedures should be put in place for quality control of the enumeration of persons during a population census. Supervisors should occasionally accompany enumerators to review, verify, and validate their enumeration. Special Forms for recording the results and review of enumerators' work and procedures for dismissing those whose work is unacceptable should be introduced (United States 2000).

In the processing of census results, a full-scale homogeneous distributed database system should be implemented. Thus, the replication of a distributed database system in geographic areas, such as Country level, State, LGA, Localities,

Survey Areas (SAs), and EAs should be considered. The information to be stored should include estimated number of households, estimated persons in institutions and total estimated population of the EAs, LGAs, State, and the country. The data interface can be designed using any suitable database management system on a powerful mainframe at work stations.

THEORY OF DISTRIBUTED DATABASE AND DDBMS

A *distributed system* is an information-processing system that contains a number of independent computers that cooperate with one another over a communications network in order to achieve a specific objective (Kay-Romer et al. 2006).

Distributed databases bring the advantages of distributed computing to the database management domain. A distributed computing system consists of a number of processing elements, not necessarily homogeneous, that are interconnected by a computer network, and that cooperate in performing certain assigned tasks.

A physical view of a distributed system includes computers as nodes of the communications network along with details about the communications network itself. In contrast, a logical view of a distributed system highlights the applications aspects. Figure 1 can therefore also be interpreted as a set of cooperating processes.

The distribution aspect refers to the distribution of state (data) and behavior (code) of an application. The process encapsulates part of the state and part of the behavior of an application, and the application's semantics are achieved through the cooperation of several processes. The logical distribution is independent of the physical one.

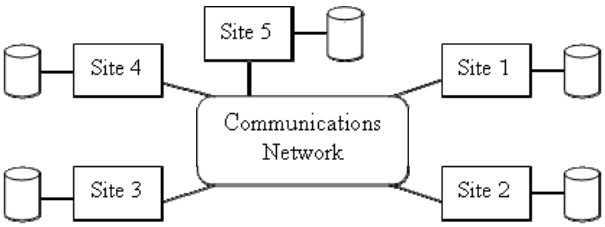

Figure 1: A Distributed Database Architecture

Distributed database management has been proposed for various reasons ranging from organizational decentralization and economical processing to greater autonomy. Navathe (2000) highlighted some of these advantages as presented below.

a. *Management of distributed data with different levels of transparency such as:*

   - *Distribution or network transparency:* This refers to freedom for the user from the operational details of the network. This parameter may be divided into location transparency and naming transparency. Location transparency refers to the fact that the command used to perform a task is independent of the location of data and the location of the system where the command was issued. Naming transparency implies that once a name is specified, the named objects can be accessed unambiguously without additional specification.

   - *Replication transparency:* In this case, copies of data may be stored at multiple sites for better availability, performance, and reliability. Replication transparency makes the user unaware of the existence of copies.

   - *Fragmentation transparency:* Two types of fragmentation are possible. Horizontal fragmentation distributes a relation into sets of tuples (rows). Vertical fragmentation distributes a relation into sub-relations where each sub-relation is defined by a subset of the columns of the original relation. A global query by the user must be transformed into several fragment queries. Fragmentation transparency makes the user unaware of the existence of fragments.

b. *Increased reliability and availability:* These are two of the most common potential advantages cited for distributed databases. Reliability is broadly defined as the probability that a system is running (not down) at a certain time point, whereas availability is the probability that the system is continuously available during a time interval. When the data and DBMS software are distributed over several sites, one site may fail while other sites continue to operate. Only the data and

software that exist at the failed site cannot be accessed. This improves both reliability and availability. Further improvement is achieved by judiciously *replicating* data and software at more than one site. In a centralized system, failure at a single site makes the whole system unavailable to all users. In a distributed database, some of the data may be unreachable, but users may still be able to access other parts of the database.

c. *Improved performance:* A distributed DBMS fragments the database by keeping the data closer to where it is needed most. Data localization reduces the contention for CPU and I/O services and simultaneously reduces access delays involved in wide area networks. When a large database is distributed over multiple sites, smaller databases exist at each site. As a result, local queries and transactions accessing data at a single site have better performance because of the smaller local databases. In addition, each site has a smaller number of transactions executing than if all transactions are submitted to a single centralized database. Moreover, inter-query and intra-query parallelism can be achieved by executing multiple queries at different sites, or by breaking up a query into a number of sub-queries that execute in parallel. This contributes to improved performance.

d. *Easier expansion:* In a distributed environment, expansion of the system in terms of adding more data, increasing database sizes, or adding more processors is much easier.

In Navathe (2000), the term distributed database management system can describe various systems that differ from one another in many respects. The main thing that all such systems have in common is the fact that data and software are distributed over multiple sites connected by some form of communication network. Different types of DDBMSs and the criteria and factors that make some of these systems different have been discussed in some literatures.

The first factor considered is the degree of homogeneity of the DDBMS software. If all servers (or individual local DBMSs) use identical software and all users (clients) use identical software, the DDBMS is called homogeneous; otherwise, it is called heterogeneous. Another factor related to the degree of homogeneity is the degree of local autonomy. If there is no provision for the local site to function as a stand-alone DBMS, then the system has no local autonomy. On the other hand, if *direct access* by local transactions to a server is permitted, the system has some degree of local autonomy.

At one extreme of the autonomy spectrum, we have a DDBMS that "looks like" a centralized DBMS to the user. A single conceptual schema exists, and all access to the system is obtained through a site that is part of the DDBMS—which means that no local autonomy exists.

At the other extreme we encounter a type of DDBMS called a *federated DDBMS* (or a *multi-database system*). In such a system, each server is an independent and autonomous centralized DBMS that has its own local users, local transactions, and DBA and hence has a very high degree of *local autonomy.* The term federated database system (FDBS) is used when there is some global view or schema of the federation of databases that is shared by the applications.

On the other hand, a multi-database system does not have a global schema and interactively constructs one as needed by the application. Both systems are hybrids between distributed and centralized systems and the distinction we made between them is not strictly followed. We will refer to them as FDBSs in a generic sense.

In a heterogeneous FDBS, one server may be a relational DBMS, another network DBMS, and a third an object or hierarchical DBMS; in such a case it is necessary to have a canonical system language and to include language translators to translate sub-queries from the canonical language to the language of each server. The type of heterogeneity present in FDBSs may arise from several sources. Viz; differences in data models, differences in constraints and differences in query languages. Semantic heterogeneity occurs when there are differences in the meaning, interpretation, and intended use of the same or related data. Semantic heterogeneity among component database systems (DBSs) creates the biggest hurdle in designing global schemas of heterogeneous databases.

There are three alternative approaches to separating functionality across different DBMS-related processes; these alternative distributed

DBMS architectures are called Client-Server, Collaborating Server, and Middleware.

A **Client-Server** system has one or more client processes and one or more server processes, and a client process can send a query to any one server process. Clients are responsible for user-interface issues, and servers manage data and execute transactions. Thus, a client process could run on a personal computer and send queries to a server running on a mainframe.

In a **Collaborating Server** system, we can have a collection of database servers, each capable of running transactions against local data, which cooperatively execute transactions spanning multiple servers. When a server receives a query that requires access to data at other servers, it generates appropriate sub-queries to be executed by other servers and puts the results together to compute answers to the original query. Ideally, the decomposition of the query should be done using cost-based optimization, taking into account the costs of network communication as well as local processing costs.

The Middleware architecture is designed to allow a single query to span multiple servers, without requiring all database servers to be capable of managing such multi-site execution strategies. It is especially attractive when trying to integrate several legacy systems, whose basic capabilities cannot be extended. The idea is that we need just one database server that is capable of managing queries and transactions spanning multiple servers; the remaining servers only need to handle local queries and transactions. We can think of this special server as a layer of software that coordinates the execution of queries and transactions across one or more independent database servers; such software is often called **middleware**. The middleware layer is capable of executing joins and other relational operations on data obtained from the other servers, but typically, does not itself maintain any data.

FRAME WORK FOR CLIENT-SERVER DISTRIBUTED DATABASES SYSTEM FOR POPULATION CENSUS

The primary types of system architectures for information processing include: Service Oriented Architecture (SOA), distributive (client–server), and centralized information systems processing more commonly associated with mainframe and midrange computers. The client–server architecture is considered in this case. This is a network architecture in which each computer or process on the network is either a client or a server. Servers are powerful computers or processes dedicated to managing disk drives (file servers), printers (print servers), or network traffic (network servers). Clients are PCs or workstations on which users run applications. Clients rely on servers for resources, such as files, devices, and even processing power.

Interaction between client and server might proceed as follows during the processing of an SQL query:

a. The client parses a user query and decomposes it into a number of independent site queries. Each site query is sent to the appropriate server site.

b. Each server processes the local query and sends the resulting relation to the client site.

c. The client site combines the results of the sub-queries to produce the result of the originally submitted query.

In this approach, the SQL server has also been called a transaction server (or a database processor (DP) or a back-end machine), whereas the client has been called an application processor (AP) (or a front-end machine). The interaction between client and server can be specified by the user at the client level or via a specialized DBMS client module that is part of the DBMS package. For example, the user may know what data is stored in each server, break down a query request into site sub-queries manually, and submit individual sub-queries to the various sites. The resulting tables may be combined explicitly by a further user query at the client level. The alternative is to have the client module undertake these actions automatically.

In a typical DDBMS, it is customary to divide the software modules into three levels:

a. The server software is responsible for local data management at a site, much like centralized DBMS software.

b. The client software is responsible for most of the distribution functions; it accesses data distribution information

from the DDBMS catalog and processes all requests that require access to more than one site. It also handles all user interfaces.

c. The communications software (sometimes in conjunction with a distributed operating system) provides the communication primitives that are used by the client to transmit commands and data among the various sites as needed. This is not strictly part of the DDBMS, but it provides essential communication primitives and services.

The client is responsible for generating a distributed execution plan for a multi-site query or transaction and for supervising distributed execution by sending commands to servers. These commands include local queries and transactions to be executed, as well as commands to transmit data to other clients or servers. Hence, client software should be included at any site where multi-site queries are submitted. Another function controlled by the client (or coordinator) is that of ensuring consistency of replicated copies of a data item by employing distributed (or global) concurrency control techniques. The client must also ensure the atomicity of global transactions by performing global recovery when certain sites fail.

The conceptual diagram of framework presented in Figure 1 is viewed as a system comprising of a distributed database, Inference engine (Analyser) and Decision Support engine

KNOWLEDGE BASE

The Database is composed of the quantitative (structured) and qualitative (unstructured) knowledge of population dynamic acquired by the human. The database is a network of semantically related static and dynamic objects, each of which is modeled, in a relational form. The structured knowledge is concerned with facts, rule and events of human population dynamics, which are commonly agreed upon by experts in demography. The unstructured knowledge is that knowledge which is acquired by demographic expert from experience and population census survey. It is the heuristic knowledge or that which is acquired by good practice, guess and judgments.

Data may be organized using a relational, hierarchical network or object orientated database model. The databases may be accessed via networks using technologies like client-server.

The prominent form of database organization described as relation allows the user to think in the form of two dimensional tables which is the way many people see data reports. It takes its name from the mathematical theory of relations. The data in the database are stored together with minimum of redundancy to serve multiple applications, so the database is independent of the computer program that uses it and the type of hardware where it is stored.

The general description of a relation is given in (Date, 1986). a **relation** (or **relation state**) r of the relation schema $R(A_1, A_2, . . ., A_n)$, also denoted by r(R), is a set of n-tuples $r = \{t_1, t_2, . . ., t_m\}$. Each **n-tuple** t is an ordered list of n values $t = <v_1, v_2, . . ., v_n>$, where each value v_i, 1 1 i 1 n, is an element of dom(A_i) or is a special **null** value. The i^{th} value in tuple t, which corresponds to the attribute A_i, is referred to as $t[A_i]$. One of the relations in this database is as presented below:

ZONE [zone_id, zone_name, zone_head_quarter]

STATE [state_zone_id, state_id, state_name, state_capital],

LOCAL GOVERNMENT [lg_state_id, lg_id, lg_name, lg_headquarter]

TOWN[town_lg_id, town_id, town_name, Head_of_community]

WARD[ward_town_id, ward_id, ward_name]

STREET[street_ward_id, street_id, street_name]

HOUSEHOLD[household_id, street_id, ward_id, town_id, lg_id, state_id, zone_id, house_description, head_of_household]

BIO[respondent_id, house_id, respondent_name, relation_to_household_head, sex, age, tribe, nationality, disability, duration of residence, previous_residential_address, present_residential_address]

EDUCATION[respondent_id, literacy, highest_educational_qualification]

ECONOMIC[respondent_id, work_status, type_of_employment, sector_of_employment]

MARITAL[respondent_id, marital_status,
age_at_marriage, no_of_children_born,
no_of_death_of_children_in_the_last_1_year]

INFERENCE ENGINE

The inference engine is concerned with the adoption of an appropriate line of reasoning, leading to the solution of a given demographic problem or the formulation of a body of consultative advice on a given demographic phenomenon.

Carrying out population census tabulation programs involves a determination of the number of different levels of geographic details. Data are presented for administrative divisions and subdivisions of the country, Federal government, state government, and local government in various categories.

For small geographic area, such as villages, the results as a rule are limited to a report of the number of inhabitants or perhaps the population by sex only. At another level, tabulation may provide only inventory statistics that is simply a count of persons in the various categories of age, marital status, economic activity and so on, with but little cross-classification with other characteristics.

In some cases, most subjects are cross-tabulated by age and sex and often there are cross-classification with social and economic characteristics, such as educational attainment by economic activity or employment status by occupation

Data from the population census questionnaire are edited to correct inconsistencies and to eliminate omissions. Responses from census questionnaire are coded. In other word, converting entries on the questionnaire into symbols, which can be used as, input data to system. Some entries required no coding. For example, sex, marital status and questions which are answered in terms of a numerical entry such as age number of children ever born and so on.

The list of demographic statistics which can be generated from the system are listed below:

a. Population distribution by age group and sex

b. Population distribution of type of employment by age group and sex

c. Population distribution of marital statistics by age group and sex

d. Population distribution of work status by age group and sex

e. Population distribution of literacy by age group and sex

f. Population distribution of relationship to household head by age group and sex

g. Population distribution of disability by age group and sex

h. Population distribution of educational attainment by age group and sex

i. Population distribution of sector of employment by age group and sex

j. Population distribution of tribe by age group and sex

k. Population distribution of death by age group and sex

l. Population distribution of marital status by age group and sex

m. Population Distribution by Date of Birth and Months of year

n. Population Distribution by Birth Rate and Months of the Year

o. Population Distribution of Marital Status by Age Group for Male and Female.

   i. Distribution on Single marital status by sex

   ii. Distribution on Married marital status by sex

   iii. Distribution on Divorced marital status by sex

   iv. Distribution on Widowed marital status by sex

p. Population Distribution of Employment Status by Age Group and Sex.

i. Distribution on self employed employment status by sex

ii. Distribution on public sector employment status by sex

iii. Distribution on private sector employment status by sex

iv. Distribution on unemployed employment status by sex

q. Population Distribution of Birth in Wards by Year Group and Months of the Year.

i. Population distribution of birth in the ward by year of birth and month

ii. Population distribution of birth rate of the ward by year of birth and month

iii. Population distribution of ward by year group, male, female and total of sexes

iv. Population distribution of ward by age group, male, female and total of sexes

v. Population distribution of the ward by relative density

r. Population Distribution by PSLC, JSS, SSS, OND, HND, NCE, B.SC, MSC and PHD for Male and Female

s. Human life table

t. Population projection

The demographic statistics stated above are largely determined by the age group of sampled population. The generation of each statistics is determined by a unique function, that is, each function differs from the other. For example, the function for generating the population distribution of marital status by age group and sex is different from that function for generating the population distribution of employment status by age group and sex.

Therefore there is the existence of a function F_k, $k = 1, 2, 3, \ldots, n$ which is a vector defined on age group of population denoted by g_i, $i = 1, 2, 3, \ldots, m$.

Where n is the number of statistical/demographic information

m is the number of age group of the sample population

INFERENCE PROCEDURE FOR GENERATING POPULATION DISTRIBUTION BY AGE GROUP AND SEX

Let F_1 represents the function defined on g_i to generate the population distribution by age group and sex. The types of sex considered in this research are:

a. Male
b. Female

Hence, $F_1(g_i)$ produces a matrix defined by:

$A_{i,j}$; $i = 1, 2, \ldots, m$; $j = 1, 2$

where m is the number of age group and $A_{i,j}$ is defined by:

$$A_{i,j} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ " & " \\ " & " \\ " & " \\ a_{m,1} & a_{m,2} \end{pmatrix}$$

$a_{i,j}$ represents the frequency of ith age group for jth type of sex and it is generated by:

$$a_{i,j} = \sum_{k=1}^{n} h_k(g_i);$$

where

$i = 1, 2, 3, 4, ..., 16$; $j = 1, 2$

n represent the population census sample size

The pseudocode for generating $a_{i,j}$ is given as follows:

```
loop k from 1 to m                    /* m
represents age group

loop j from 1 to n                    /* n
represents number of
attributes of statistics

        set a_i,j to 0
        loop i from 1 to p
          /* p represents the sample population
                if h_k(g_i) = 1 then   /* h_k(g_i) represents
                                          the statistics of
                                          i^th respondent in
                                          k^th group
                    a_i,j = a_i,j + 1
                endif
        endloop
    endloop
endloop
```

DIALOGUE MANAGEMENT

The dialogue component refers to the hardware and software that create the user interface for the framework. Two type of dialogue style can be adopted. The menu dialogues and the form-based and question-answer dialogue and graphical user dialogue styles. The hardware support for both dialogues emphasized keyboard monitor and mouse for input. The pull-down menu Dialogue and the form-based dialogue are designed in Visual Basic programming language while the Microsoft excel can be integrated to include a spreadsheet for graphical user interface. The Graphical User-interface can tabulate population census data and present results in pictorial form. With this dialogue management, data can be entered, retrieve, updated and stored. The windows based user-interface can have four levels of pull-down menus, viz:

File Maintenance (File_Create, Edit, Update, Save, Delete, Exit)

Consultation (Cognitive, Emotional, Crude Rate Analysis)

Population Distribution (Sorting, Classification, Tabulation, Presentation)

Report (Screen Display, Hard Copy, Soft Copy, Back up)

Chart (Bar, Line)

Help (What about Census-Manager)

COMPONENTS OF THE DISTRIBUTED SYSTEM

The distributed system have Internet facing Web-based applications that can be accessed remotely by the users either within the confines of the organization or remotely. The following are list of information technology (IT) infrastructure components.

Firewall: A system designed to prevent unauthorized access to or from a private network. Firewalls can be implemented in both hardware and software, or a combination of both. Firewalls are frequently used to prevent unauthorized Internet users from accessing private networks connected to the Internet, especially intranets. All messages entering or leaving the intranet pass through the firewall, which examines each message and blocks those that do not meet the specified security criteria. There are several types of firewall techniques:

- *Packet Filter*: Looks at each packet entering or leaving the network and accepts or rejects it based on user-defined rules. Packet filtering is fairly effective and transparent to users, but it is difficult to configure. In addition, it is susceptible to Internet Protocol (IP) spoofing.

- *Application Gateway*: Applies security mechanisms to specific applications, such as File Transfer Protocol (FTP) and Telnet servers. This is very effective but can impose performance degradation.

- *Circuit-Level Gateway*: Applies security mechanisms when a Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) connection is established. Once the connection has been made, packets can flow between the hosts without further checking.

- *Proxy Server*: Intercepts all messages entering and leaving the network. The proxy server effectively hides the true network addresses.

Router: A router is a special purpose computer or software device that enables two or more dissimilar networks to communicate. Routers route traffic, which consists of Transmission Control Protocol/Internet Protocol (TCP/IP) packets.

Host: A computer that is connected to a TCP/IP network, including the Internet.

Server s: A server is a dedicated computer that allows other computers to connect to it. Various types of servers exist which include the following:

- Domain Name System
- Web servers
- Internet banking servers
- E-mail servers
- Proxy servers

PC Workstations: In networking, workstation refers to any computer connected to a local area network. It could be a workstation or a personal computer.

Intrusion Detection Systems: Intrusion detection is fundamentally the process of monitoring computer networks and systems for violations of computer policy.

NETWORKS STRUCTURE

A metropolitan area network (MAN) which can cover a city is best for this kind of system. This system grew from earlier community antenna systems used in areas with poor over-the-air television reception. In these early systems, a large antenna was placed on top of a nearby hill and signal was then piped to the subscribers' houses. Figure 2 describes the structure of this framework.

CONCLUSION

The primary source of information about the population of a country is the population census. This has to deal with a process of collecting, compiling and publishing demographic, economic, and social data pertaining at a specified time or times, to all persons in a country or delimited territory.

There have been many developments in demographic analyses of census results in recent years. Most of them fall within the same broad direction of orientation, namely the provision of more information about the social and economic characteristics of populations and about the pattern of social and economic organization of communities.

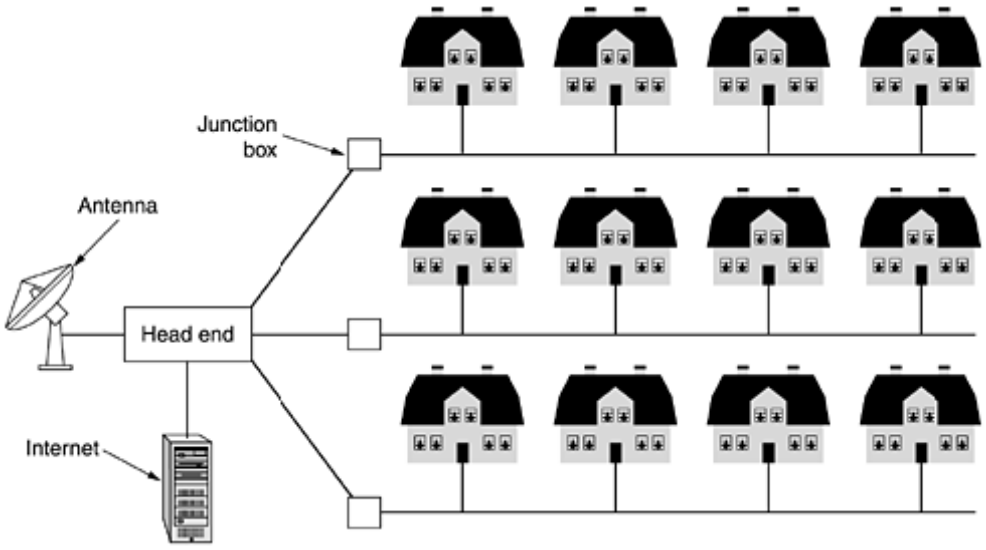

Figure 2: A Metropolitan Are Network (MAN) Based System.

With the great awareness of the growing need for information on population dynamics, a comprehensive study of demography in Nigeria with particular reference to population census In this study, an attempt is made to design a framework for distributed database system for population census and projection in Nigeria. This includes the analysis of demographic parameters of respondents in a study area.

Basically, the demographic parameters or population variables are namely: relation to head of household, sex, age, disability, home local government area, home place, address, literacy, educational qualification, work status, type of employment, sector of employment, marital status, statistics of marriage. An attempt is also be made to formulate a demographic Model for the population growth and projection in Nigeria, with a view to determining how growth impacts future planning for governmental decision making.

REFERENCES

1. Bernard, B. 1968. *Demographic Analysis*. George Allen and Unwin Ltd. pp. 21 – 30.

2. Dates, C.J. 1986. *An Introduction to Database System, Fourth Edition.* Addison-Wesley Publishing Company: San Diego, CA.

3. Heisel, D.F. 1969. *Measuring Current Population Changes Population of Tropical Africa*. Lowe & Bryone Ltd.: London, UK. pp. 34-49.

4. Kay-Romer, P., Pilhofer, F., and Arno, P. 2006. *Distributed System Architecture.* Morgan Kaufmann Publishers is an imprint of Elsevier: San Francisco, CA.

5. Nigerian National Population Commission. 1991. *NNPC Enumeration Manual*. National Population Printing Press: Lagos, Nigeria.

6. Navathe, E. 2000. *Fundamentals of Database Systems. Third Edition.* Teturo Sawada, Exclusive Publisher and Distributor.

7. Olabode, O. 1999. "Computer Aided System for Demography in Nigeria". Unpublished M.Tech Thesis. Federal University of Technology: Akure, Nigeria.

8. Owusu, D.J. 1969. *Taking a Population Census in Tropical Africa: Population of Tropical Africa*. Lowe & Bryone Ltd.: London, UK. pp. 134-141.

9. Ramakrishnan, R. and Gehrke, J. 2004. *Database Managenment Systems. Second Edition*. McGraw-Hill: New York, NY. http://www.cs.wisc.edu/~dbbook

10. Tanenbaum, A.S. 2008. *Computer Networks, Foruth Edition*. Prentice Hall: Princeton, NJ. http://www.cs.vu.nl/~ast/

11. United Nations. 2001. *The State of the Demographic Transition in Africa: Executive Summary*. Economic Commission For Africa, Joint ECA/OAU/ADB Secretariat In Collaboration With UNFPA. FSSDD/ICPD/FC.4/01/3: Yaounde, Cameroon.

ABOUT THE AUTHORS

Olabode Olatubosun, Ph.D., is professional computer scientist. He obtained his B.Tech. degree in Ind. Maths degree in 1991, M.Tech. degree in computer science in 1999, and PhD. Degree in computer science in 2005 from the Federal University of Technology, Akure, Nigeria. Presently, he is a lecture in computer science department, Federal University of Technology, Akure, Nigeria. He has since been teaching courses and researching in the areas of information system, databases, expert systems, image processing, software engineering, data mining, advance micro-computing, probability theory, and e-commerce. Dr. Olabode has great aptitude for programming and solving numerical problems, particularly those that are related to industry. He has numerous journal publications and conference proceedings to his credit. Presently, he is a consultant to Ondo State Government and Rufus Giwa Polytechnic, Owo. He has served as a technical resource expert in many workshops organized for Local Governments in Ondo State. He is also a consultant to JT Canada. He is a member of research groups in the University where he is currently working.

O.C. Akinyokun, is a researcher in the Computer Science Department, Federal University of Technology, Akure, Nigeria with interests in information technology and applications.

SUGGESTED CITATION

Olatubosun, O. and O.C. Akinyokun. 2008. "Framework for Client-Server Distributed Database System for Population Census". *Pacific Journal of Science and Technology.* 9(2):395-406.