Design and Development of Data Warehousing for Bookstore Using Pentaho BI Tools

C. S. Anagha[1,2(✉)] and Siddhaling Urolagin[1,2]

[1] Department of Computer Science, APP Center for AI Research (APPCAIR), Birla Institute of Technology and Science, Pilani, India
[2] Dubai International Academic City, Dubai, United Arab Emirates

Abstract. The bookstore data warehouse is proposed to analyze the book selling patterns to increase the turnover by means of best sellers and other factors which affect a bookstore. This analysis will help bookstore owners to purchase books according to their customers' interests, language preferences, book types, and other factors. This paper tries to overcome the obstacles the bookstore faces in collecting data for analysis. The data will be in different formats and may contain false values or empty cells. So, analytical query response time will not be efficient. The data needs to be cleaned and integrated before analysis. The data needs to be changed to a standard format which will speed up the query response time. Data warehouse design is very dynamic. A data warehouse can be used for analysis, decision making, and for future planning of a bookstore. The schema design and storage of data is done using the MySQL relational database management system. The Pentaho Server, MySQL workbench, and Data Integration tools are used for data integration and transformation into a standard form. The multi-dimensional database modelling approach used is STAR Schema, which uses dimension tables and a fact table. The data was collected from different bookstores online. The collected information is cleaned and transformed at the bookstore data warehouse. The Pentaho Server is used for querying the cube using the MDX query for transforming the data into tabular and chart representation. Based upon the data transformed and descriptions, the decision-maker or stakeholders can decide on upcoming strategies for successfully achieving target turnover.

Keywords: Data extraction · Data analysis · Data warehouse · Pentaho

1 Introduction

If one can understand the products on a deeper level, it will be easier to order and maintain a stock of those items. It is challenging to analyze the best sellers in a bookstore. A bookstore has several branches, opened in different parts of the world. Analyzing the best sellers is not an easy task in a bookstore. The number of books is increasing day by day, as well as the number of authors. In different places, according to the language, people, etc., the best sellers may vary. Nowadays, many people prefer audio books or kindles rather than printed versions. Since there are different types of single book available on the market, people will choose from these according to their interests. Selling methods

K. R. Venugopal et al. (Eds.): ICInPro 2021, CCIS 1483, pp. 195–210, 2021.
https://doi.org/10.1007/978-3-030-91244-4_16

have also changed. Online ordering of books and homedelivery services also need to be considered along with employee performance for the better performance of a bookstore. The statistical analysis will help a bookstore owner to make better decisions to ensure better turnover and service for the bookstore. This analysis will help to predict the collection of books to be stored in the bookstore. The data from heterogeneous sources is collected and pushed using services to the data warehouse. This study focuses on more aspects which affect a bookstore. This study utilizes data integration tools and the Pentaho Schema workbench. A recent study says 39% read printed books, 7% read digital books and 29% read both printed and digital books. Most surveys come up with the finding that printed book format readers are always higher in number compared to e-books and audio books. A study shows the Holy Bible is the best-selling book in the world. Some readers like a particular author, while some readers like to read a particular type of book. Some publishers' books sell more than others. This is all information that concerns a bookstore owner for decision making.

We identified the need for a data warehouse for a bookstore. The objective of this paper mainly focuses on Datawarehouse architecture for bookstores, where it will be very useful for doing analysis to find the selling pattern. We developed a data warehouse model for the bookstore using Pentaho tools. This model is useful for bookstores to analyze their selling patterns and make decisions to improve their sales and thus make more profit. It is hard to see a Datawarehouse model developed particularly for a bookstore. Our Datawarehouse model discusses more dimensions, and it can be adapted to any bookstore, which will be helpful for any sellers. In the following section, we will describe the background of the bookstore warehouse. Data warehouse architecture is discussed in Sect. 3, schema design in Sect. 4, Extract Transformation and Loading in Sect. 5, Experimental Results in Sect. 6, followed by conclusion in Sect. 7.

2 Literature Survey

A data warehouse is the main repository of an organization's historical data. The data warehouse is optimized for reporting and analysis [1]. A growing number of large enterprises select a data warehouse to help with their decision-making analysis. With a data warehouse, enterprises can understand the information of customers, business conditions, sales channels, and make timely and effective decisions, thereby reducing operating costs, improving customer satisfaction, increasing operating profits and expanding market share [2]. OLAP (Online Analytical Processing) tools offer the possibility of archiving, management, analysis, and multidimensional modelling [3]. Before applying the Business Intelligence technique, data is processed and normalized to avoid data redundancy [4]. The Multidimensional Data Model can be used for the creation of multiple data marts and the design of an ETL process for populating the data marts from the data source [5]. The use of dynamic ETL process using metadata ETL is required when the ETL process is dealing with the operational system and to address the increasing requirements for reporting from users [6]. Business Intelligence tools (BI tools) can be used to analyze large amounts of data [7]. A top-down approach makes it possible to elicit and consolidate user requirements and expectations [8]. The star is composed of two kinds of basic tables: fact tables and dimension tables. The fact table includes

operational transactions, or the analysis values wanted, and dimension tables include the description information related to these transactions and values. The star schema exists widely in database application systems [9]. Pentaho data integration can give much better results [10].

In paper [11] we discuss the constraint of a bookstore in processing data, since data increases with increasing time. A bookstore has problems analyzing bestsellers because of the increasing number of authors and book publishers who work together. This paper found a solution for GIS Bookstore by implementing business intelligence that plays a role in the processing of raw data into well-structured data information. The results are collected and processed in a data warehouse scheme. Even though the model needs to be improved for the current situation, with online book purchase, implementation changes are needed to improve the analysis of the old model. This model analyses only 7 dimensions. There are more dimensions which affect the sale of books in bookstores. So, this model cannot be considered as a common model which is applicable to any bookstores. We need to develop a common model for bookstores. In paper [12], they developed a snowflake schema for the library system. which is used to decide on the purchase of books for the library according to students' needs. But this cannot be applicable in the case of a bookstore. This model includes dimensions for selection and management of books. Here, data dimensions can be broken down in more detail. But the snowflake schema requires normalization of tables.

3 Data Warehouse Architecture for Bookstore Data

The Data Warehouse Architecture consists of different parts, as shown in Fig. 1. Extraction Transformation Loading (ETL), Data Warehouse, Data Marts, and Business Intelligence (BI) tools. The data is collected from heterogeneous sources like the operational system, ERP, CRM, and flat files. ETL tools are used for extracting, transforming, and loading data. The data is first stored in the staging data warehouse. Once confirmed and verified, the data is pushed from staging to the production Data Warehouse. The data warehouse consists of metadata, summary data, and raw data. This data will be in the form of a DataMart. By using OLAP tools, the data is presented and visualized. The same data is used for the mining process as well as for future analysis, prediction, and automation.

The Bookstore Datawarehouse shown in Fig. 2 consists of two major parts. One is Online Transactional Processing (OLTP) and the other is Online Analytical Processing (OLAP). This paper deals extra with the OLAP part. The source data is taken from various internal and external sources. Various bookstores and publishers have their own transaction sources, which is known as Online Transactional Processing (OLTP). In this study, we are generating data from bookstore-based applications and spreadsheets. The data from various sources is integrated. This process is well known as Extract Transformation Loading (ETL). The ETL process is practical by using tools. We are using MySQL and Pentaho Tools (Pentaho Server, Schema Workbench, and Pentaho Data Integration Tools). On the Bookstore Data Warehouse design and implementation, we apply the top-down approach to designing the data warehouse.

The data includes information related to bookstores, such as bookstore information, book information, publisher information, customer information, employee information,

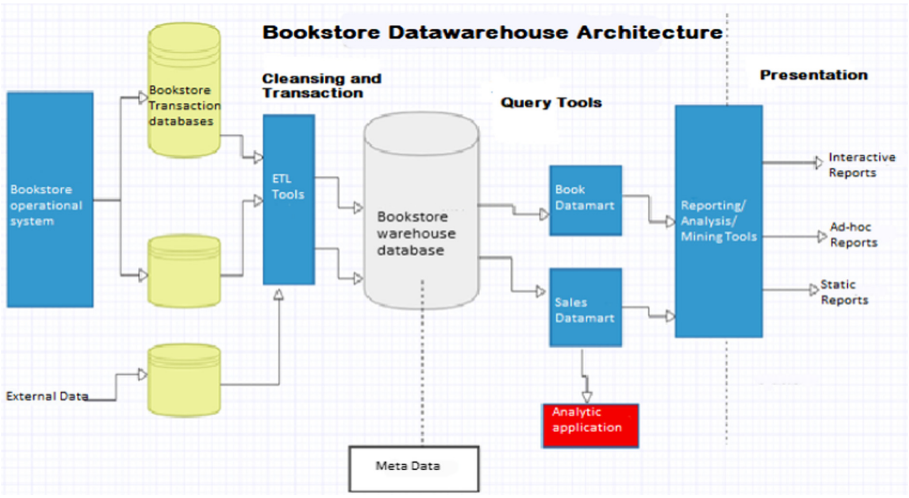

Fig. 1. Datawarehouse architecture

etc. This data is stacked and approved to ensure the extracted information is correct. This is done in the staging area. Here, ETL operations will take place which will utilize ETL devices. Then the information is pushed to the information distribution center. Various deletion, aggregation, and summarization techniques will be done on extracted data, and then loaded back into the data warehouse. After all the rundown is finished, some more changes will be made to the data to make the information structure characterized by the data marts. The source data from different sources is extracted and pushed to the bookstore data warehouse. The data warehouse is built on a relational database management system (RDBMS), and the star schema style is used to develop data warehouses and dimensional data marts. The Cube schema is generated using the Pentaho Schema workbench tool. For querying the data from the bookstore data warehouse database, the MDX (Multidimensional Expression) query is used. For transforming and showing the bookstore statistics in tabular and chart format, the Pentaho Data Integration or JPivot View on the Pentaho BI tool is used.

4 Schema Design

The Star schema has become a common term used to connote a dimensional model. Database designers have long used the term star schema to describe dimensional models because the resulting structure looks like a star and the logical diagram looks like the physical schema. The star model is the basic structure for a dimensional model. It typically has one large central table (called the fact table) and a set of smaller tables (called the dimension tables) arranged in a radial pattern around the fact table. It is more powerful for dealing with straight-forward questions. Usually, the truth tables in a star model are in the third normalized form (3NF), while dimensional tables are in the denormalized form. The star model is the most used model nowadays and is suggested by Oracle.

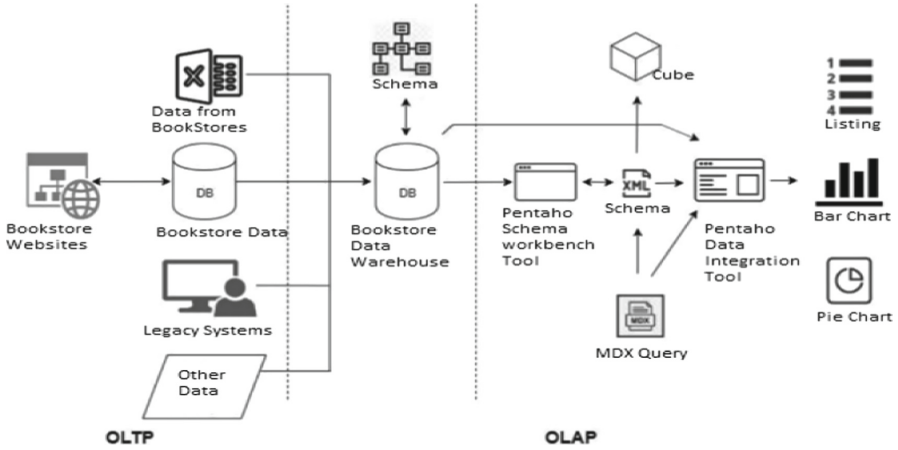

Fig. 2. Bookstore data warehouse architecture

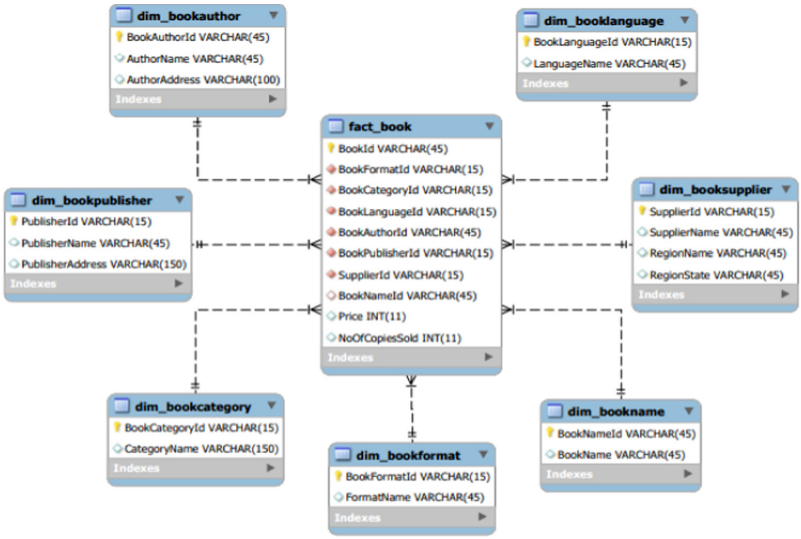

Fig. 3. Book DataMart

In this study, the multidimensional data is illustrated by the Star Schema. The multidimensional database consists of dimensions and fact tables. A fact table has two sorts of sections: foreign keys to measurement tables and measures those numeric attributes of a fact, representing the performance or behavior of the bookstore relative to the dimensions. A fact table can contain a fact's data in detail or gathered level. Dimensions are the parameters over which we want to perform Online Analytical Processing (OLAP). For example, time, location, customers, salespeople, etc… Figure 3 shows one of the data marts in the bookstore data warehouse. It consists of a fact table (fact_book) and dimension tables (dim_bookname, dim_bookcategory, dim_bookformat, dim_booklanguage,

Fig. 4. Sales DataMart

dim_bookauthor, dim_bookpublisher, dim_supplier). The book_facttable contains the measure price, number of copies sold, and the foreign keys to each dimensional table. The dim_bookname table contains a list of all books with their id. The dim_category table includes categories of books such as fiction, non-fiction, kids, etc. The dim_format includes details of the format of the book. The format can be kindle, audio, paperback etc. The dim_bookLanguage table consists of different languages in which the book is published. The dim_author table consists of the author's details, such as name and address. The dim_publisher table consists of different book publisher names and details. The dim_supplier table consists of the supplier's region and name. Figure 4 shows the sales data mart. The sale_fact table consists of keys to the dimensional tables along with the measure price and number of copies sold for each book. The dimensional tables are dim_store, dim_employee, dim_bookforstore, dim_datetime, dim_typeofsale and dim_customers. The dim_store contains details of each store, like name, region, etc. The dim_bookforstore contains the id of each book and the name of the book. The dim_employee consists of employee related data. The dim_typeofsale consists of an id for the sale type and a description. The dim_date dimension includes the date of the transaction (the year, month, week etc.). The dim_customer stores the details about the customer, such as name, region, email etc.

5 Extract Transformation and Loading (ETL)

ETL means Extract Transformation and Loading. Extraction is the process by which data is grabbed from heterogeneous sources and saved in another repository to analyse it later. The transformation of data changes the form of the data from the form in which it is stored to another form which suits the database where the data is going to be stored. It can be done using data filtration-selection of relevant data, by joining data from different

sources, by removing duplicate data, etc. For data transformation, specific rules and query tables can be utilized. Data is extracted from different sources and stored on excel sheets. The csv file for each dimensional table as well as fact tables are created. Then the csv files are imported into the Bookstore Schema. Sample data for the Book_Fact table is shown in Table 1. Sample data for Sales_Fact is shown in Table 2.

Tables 3 and 4 show sample data for dimensional tables Dim_BookCategory from Book DataMart and Dim_Date from Sales DataMart. The Mondrian Schema was created using the Pentaho workbench tool. The Mondrian Schema is a logical model that consists of cubes, different levels of hierarchies, and members.

Table 1. Sample data for book fact data

BookId	Book-FormatId	BookCat-egoryId	BookLan-guageId	BookAu-thorId	Book-PublisherId	Sup-plierId	Book-NameId	Book-PublishedDate	Price	NoOfCopies-Sold
B1	F40	BC53	L14	AU65	P8	SU2	PZNJBW	06/04/1914	1344	93470014
B10	F40	BC144	L33	AU261	P50	SU7	VPPNOR	15/05/1932	1545	74678639
B100	F15	BC52	L25	AU199	P37	SU4	SINUQV	19/04/1989	384	95219250
B1000	F27	BC3	L28	AU122	P51	SU5	MXBUCO	26/03/2017	1314	75782165
B1001	F6	BC200	L7	AU147	P53	SU1	BQGKKE	24/05/1901	1779	37631794
B1002	F13	BC173	L13	AU131	P38	SU3	KJEXGK	08/08/2017	1564	14974535
B1003	F41	BC51	L24	AU245	P58	SU2	MQRYSI	14/01/2011	157	11015760
B1004	F5	BC230	L27	AU212	P49	SU6	FBDVLV	07/01/1974	259	18519626
B1005	F1	BC148	L33	AU95	P18	SU3	SXXFED	10/02/1922	1932	19190162
B1006	F35	BC33	L23	AU190	P56	SU1	OYICFW	14/12/1912	884	15665714

Table 2. Sample data for sales fact data

TransactionId	StoreId	DateId	BookId	CustomerId	EmployeeId	Price	Quantity	TypeOfSaleId
T1	ST3	534	B1663	348	EM14	1582	12	Online
T10	ST1	249	B753	462	EM5	331	10	Online
T100	ST4	47	B2138	204	EM7	1548	12	Online
T1000	ST2	1685	B1911	347	EM19	99	19	Online
T1001	ST3	121	B1858	406	EM9	1393	18	WalkIn
T1002	ST1	1471	B326	227	EM9	1819	9	Online
T1003	ST4	1204	B1011	112	EM11	1004	14	Online
T1004	ST5	1336	B697	172	EM16	843	3	WalkIn
T1005	ST5	840	B1436	335	EM14	1883	12	Online
T1006	ST4	1743	B2106	166	EM6	1551	10	Online

Table 3. Sample data for book category dimension data

Book categoryId	Category name
BC1	Prolegomena. Fundamentals of knowledge and culture. Propaedeutics
BC10	Computer communication
BC100	Astronomy. Astrophysics. Space research. Geodesy
BC101	Physics
BC102	Mechanics
BC103	Optics
BC104	Heat. Thermodynamics. Statistical physics
BC105	Electricity. Magnetism. Electromagnetism
BC106	Condensed matter physics. Solid state physics
BC107	Physical nature of matter

Table 4. Sample data for date dimension data

DateId	Date	Day	Month	Year	Day of month	Week of year	Month of year	Quarter of year	Half of year
1	01/01/2001	Monday	January	2001	1	1	1	Q1	H1
10	10/01/2001	Wednesday	January	2001	10	2	1	Q1	H1
100	10/04/2001	Tuesday	April	2001	10	15	4	Q2	H1
1000	27/09/2003	Saturday	September	2003	27	39	9	Q3	H2
1001	28/09/2003	Sunday	September	2003	28	40	9	Q3	H2
1002	29/09/2003	Monday	September	2003	29	40	9	Q3	H2
1003	30/09/2003	Tuesday	September	2003	30	40	9	Q3	H2
1004	01/10/2003	Wednesday	October	2003	1	40	10	Q4	H2
1005	02/10/2003	Thursday	October	2003	2	40	10	Q4	H2
1006	03/10/2003	Friday	October	2003	3	40	10	Q4	H2

6 Experimental Results

We have created about 2500 records in the Book Datamart and about 3000 records in the Sales Datamart. The sample data is randomly generated in Excel documentation (CSV format) and converted to star schema as shown in Figs. 3 and 4. The Mondrian schema and cubes are created using the Pentaho workbench tool as shown in Fig. 5. The created schema is validated and then published to the Pentaho repository. The JPivot view on the Pentaho BI server is used for the bookstore data analysis. The statistics are viewed in tabular or various kinds of chart formats. The MDX Query is used for querying the

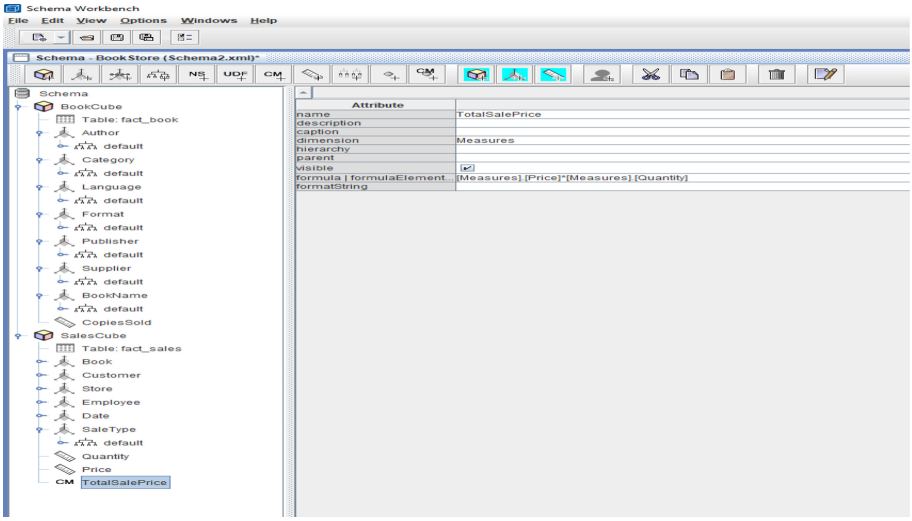

Fig. 5. The bookstore Mondrian schema

Bookstore Datawarehouse database. Some test MDX queries and display results are shown below.

Number of Copies Sold for Each Book. Figure 6 shows the number of copies sold for each book. This report helps to find the best sellers. From the figures, we can find copies of some books sold for 80000000, some are sold for less than 2000000.

MDX Query: select NON EMPTY {[Measures].[CopiesSold]} ON COLUMNS, NON EMPTY [BookName].[All BookNames].Children ON ROWS from [BookCube].

Books Sold According to Author. The report in Fig. 7 indicates which author's books are more in demand. The author's name is shown in the x axis and the copies of books sold in the y axis. This report will help the analyser to include more books by a particular author.

MDX Query: select NON EMPTY {[Measures].[CopiesSold]} ON COLUMNS, NON EMPTY [Author].[All Authors].Children ON ROWS from [BookCube].

Books Sold According To Language. Figure 8 shows a pie chart in which each slice represents a language. It can help to find out the sales of books according to their customers' preferred language. This helps to incorporate books in customers' preferred languages.

MDX Query: select NON EMPTY {[Measures].[CopiesSold]} ON COLUMNS, NON EMPTY [Language].[All Languages].Children ON ROWS from [BookCube].

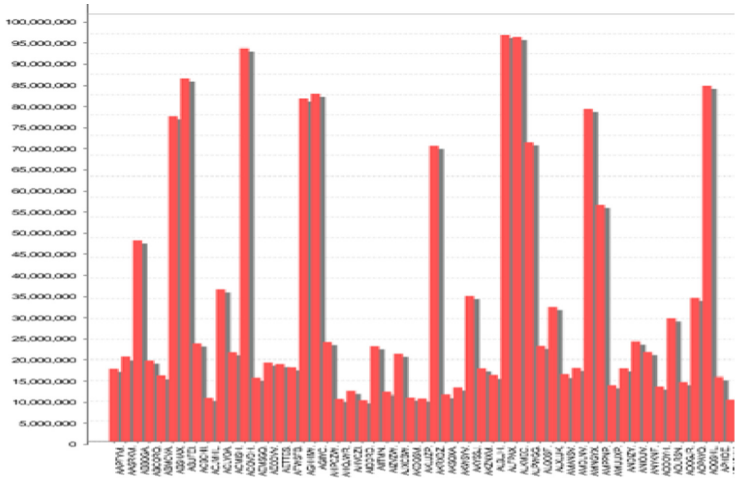

Fig. 6. Number of copies sold for each book

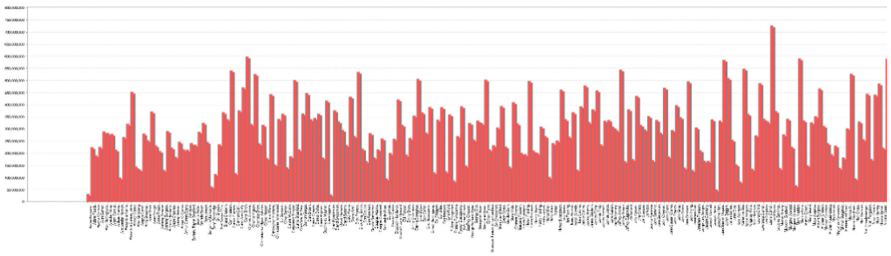

Fig. 7. Number of Books sold according to authors

Sales of a Selected Store in the Year 2001 Quarter Q1 Month of February (Slicing the Cube). Figure 9 shows the results of sales of a particular store in a particular year in a particular month. Here the sales of a single branch, AK Bookstore, are represented for the year 2001 in the quarter Q1 of the month of February. For each day of the month, the sales are represented using a bar chart.

MDX Query: select NON EMPTY Crossjoin({[Store].[AK Bookstore]}, [Date].[2001].[Q1].[February].Children) ON COLUMNS, NON EMPTY {[Measures].[Quantity]} ON ROWS from [SalesCube].

Total Sales of Each Book for the Company. Figure 10 shows the total sales of each book for the company. The result is in a tabular form. It helps to find out the total sale price of each book for the company. The total sale price is calculated by multiplying the sale quantity by the price.

MDX Query: select NON EMPTY {[Measures].[Quantity], [Measures].[Price], [Measures].[TotalSalePrice]} ON COLUMNS, NON EMPTY [Book].[All Books].Children ON ROWS from [SalesCube].

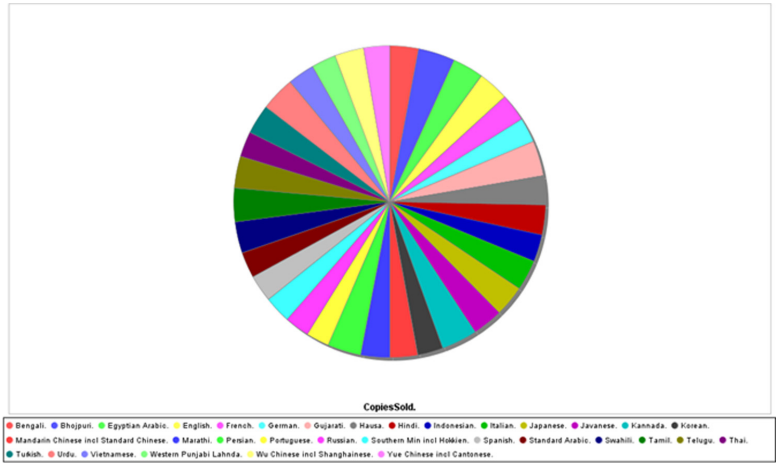

Fig. 8. Number of books sold according to the language

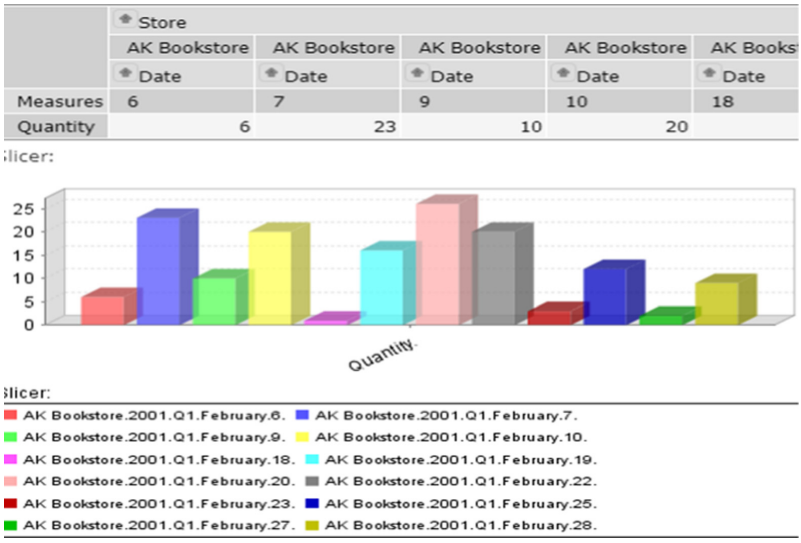

Fig. 9. Sales of a selected store

Performance of a Selected Employee For Each Year (Slicing the Cube). An employee's performance can be calculated by the total sales he made to the company. The bar chart in Fig. 11 shows the sales made by a single employee for each year. It is basically slicing the cube with the dimensions of sales time and employee name (Fig. 12).

MDX Query: select NON EMPTY {[Measures].[TotalSalePrice]} ON COLUMNS, NON EMPTY

	Measures		
Book	Quantity	Price	TotalSalePrice
AAPFYM	19	$1,481	$28,139
AASRKM	12	$1,704	$20,448
ABCQRQ	7	$1,328	$9,296
ABMCVA	29	$4,900	$142,100
ABUFEI	12	$1,548	$18,576
ACBCHI	49	$5,840	$286,160
ACJMHL	11	$3,468	$38,148
ACLYOA	11	$1,545	$16,995
ACQVDH	15	$751	$11,265
AETTES	35	$4,144	$145,040
AGHHIW	14	$1,284	$17,976
AGIIYC	1	$1,308	$1,308
AHQJWR	15	$504	$7,560
AIITMN	13	$39	$507
AIZVZW	48	$2,922	$140,256
AJXCBW	13	$1,204	$15,652
AKDQSM	5	$631	$3,155
AKLUZP	5	$2,480	$12,400
AKRXDZ	18	$1,676	$30,168
AKSQXA	3	$1,146	$3,438
AKWSYV	20	$2,232	$44,640
AKYSSJ	24	$198	$4,752
AKZNXM	6	$866	$5,196
ALFPAX	8	$844	$6,752
ALKMEC	17	$2,324	$39,508
ALQOSF	33	$5,619	$185,427
ALXLUK	22	$2,674	$58,828
AMANSV	45	$4,098	$184,410
AMNGYX	4	$1,760	$7,040
AMPPNP	17	$162	$2,754
ANDIZY	26	$234	$6,084
ANYKNF	13	$418	$5,434

Fig. 10. Total sales of each book for the company

Crossjoin({[Employee].[Abrahams]}, [Date].[All Dates].Children) ON ROWS from [SalesCube].

Performance of a Selected Employee for a Selected Year (Dicing the Cube). The chart in Fig. 13 shows the performance of a particular employee for a particular period. It is basically dicing the cube.

MDX Query: select NON EMPTY {[Measures].[TotalSalePrice]} ON COLUMNS, NON EMPTY ([Employee].[Abrahams], [Date].[2001])} ON ROWS from [SalesCube].

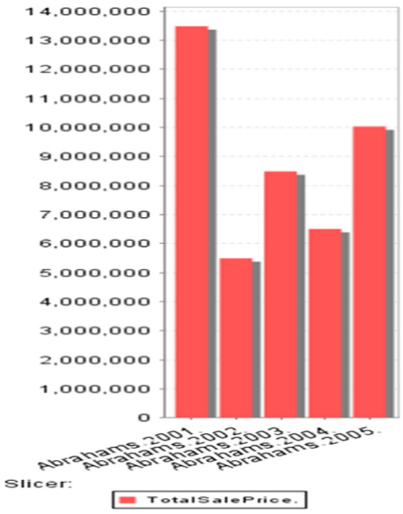

Fig. 11. Performance of a selected employee

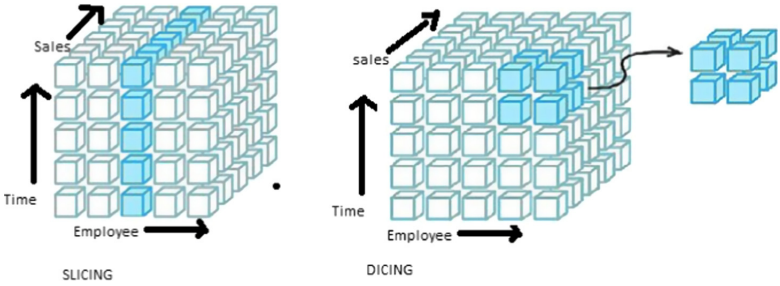

Fig. 12. Slicing and dicing of a cube

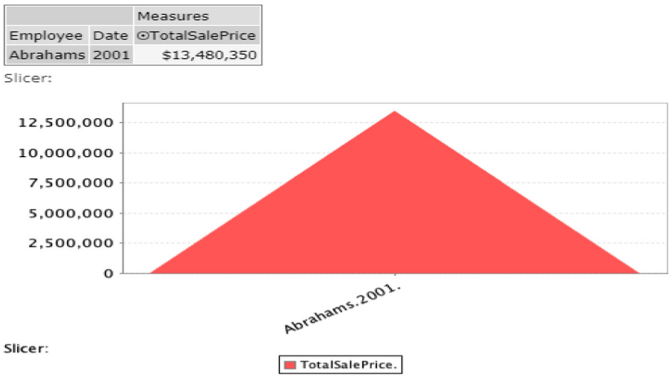

Fig. 13. Performance of a selected employee for a selected year

Sales According to the Sale Type. The bookstore sells books online as well as in person. If online book sales are more, the bookstore has to take decisions about the management of their employees to deliver the books to the customer hassle-free. Figure 14 shows the results of finding out the number of online sales vs walk-in sales.

MDX Query: select NON EMPTY [SaleType].[All SaleTypes].Children ON COLUMNS, NON EMPTY {[Measures].[Quantity]} ON ROWS from [Sales-Cube].

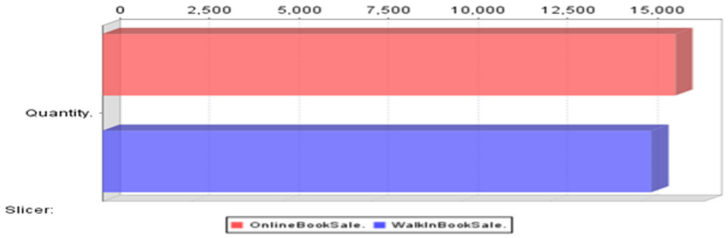

Fig. 14. Online and walk-in sale statistics

Sales Statistics According to Customers. Some customers buy more books from the store. Finding out about them is a part of business for further decision making according to their customer interests. This analysis shown in Fig. 15 helps to find out the customers who bought the most books from the store.

MDX Query: select NON EMPTY {[Measures].[Quantity]} ON COLUMNS, NON EMPTY [Customer].[All Customers].Children ON ROWS from [Sales-Cube].

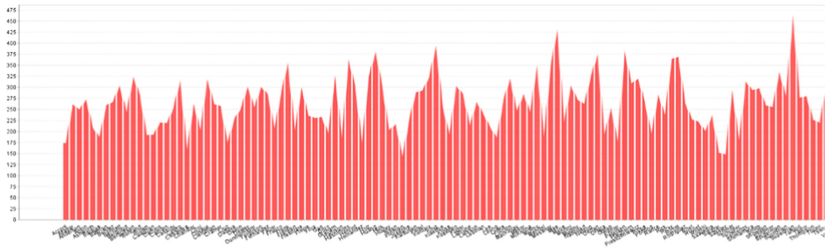

Fig. 15. Sales statistic according to customers

Sales in the Year 2001 in Quarter Q1 of the month of January (Drill down). The result in Fig. 16 shows the total sales for each day of the month of January in quarter Q1 of the year 2001 for each branch of the company. This will help to find out the monthly income for the company from all the stores.

MDX Query: select NON EMPTY {[Measures].[Quantity], [Measures].[Price], [Mea-
sures].[TotalSalePrice]} ON COLUMNS, NON EMPTY Crossjoin([Store].[All Stores].Children, Date].[2001].[Q1].[January].Children) ON ROWS from [Sales-Cube].

		Measures		
Store	Date	⊙Quantity	⊙Price	⊙TotalSalePrice
AK Bookstore	2	25	$2,559	$63,975
	3	26	$999	$25,974
	4	9	$323	$2,907
	15	12	$661	$7,932
	25	20	$1,958	$39,160
	29	13	$314	$4,082
	31	5	$1,439	$7,195
Anchor Bookstore	1	8	$248	$1,984
	7	1	$1,459	$1,459
	8	9	$718	$6,462
	11	18	$719	$12,942
	15	20	$1,121	$22,420
	23	23	$2,186	$50,278
	26	6	$641	$3,846
	27	20	$1,876	$37,520
	30	4	$1,160	$4,640
Follet Books	10	1	$1,331	$1,331
	13	3	$1,607	$4,821
	17	5	$332	$1,660
	21	6	$537	$3,222
	24	9	$366	$3,294
	28	11	$1,160	$12,760
Greenleaf Books	2	18	$456	$8,208
	8	19	$439	$8,341
	17	16	$1,120	$17,920
	25	5	$1,659	$8,295
	27	10	$1,541	$15,410
Hachette Book Group	6	23	$1,628	$37,444
	26	9	$1,598	$14,382

Fig. 16. Sales in the year 2001 in quarter Q1 in the month January

7 Conclusion and Future Work

Using Data warehouse and Pentaho tools, the data can be grouped into the newly developed data source. The Analysis of data can be done by using tabular and chart forms. This result can be generated more orderly and easily. The data can be made ready to process by Pentaho schema workbench. Using Pentaho Business Intelligence (BI) Server the dashboard can be prepared. This Panels helps to make decisions for the Analyzer. The Datawarehouse could analyze the bookstore data successfully. It could produce analysis results for number of books sold according to author, format of the book, written language, category, publisher, and supplier. The system enables to find out the best sellers

in a bookstore. The Sales data mart enables the analyzer to understand the current sale of the books as well as past sale records. The employee performance could monitor using this system. The online transactions and walk-in transactions also analyzed successfully. The Data warehouse extracted and cleaned data, which can be used for datamining related activities. This data mart can be used by business organizations to plan things accordingly for the future.

References

1. Rifaie, M., Kianmehr, K., Alhajj, R., Ridley, M.J.: Data warehouse architecture and design. In: 2008 IEEE International Conference on Information Reuse and Integration, Las Vegas, NV, USA, 2008, pp. 58–63 (2008).https://doi.org/10.1109/IRI.2008.4583005
2. Song, Q., Liu, L.: Research on the design of a new data warehouse system. In: 2009 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, 2009, pp. 462–465 (2009). https://doi.org/10.1109/ICCSIT.2009.5234503
3. Boutkhoum, O., Hanine, M., Tikniouine, A., Agouti, T.: Integration approach of multicriteria analysis to OLAP systems: Multidimensional model. In: 2013 ACS International Conference on Computer Systems and Applications (AICCSA), Ifrane, 2013, pp. 1–4 (2013). https://doi.org/10.1109/AICCSA.2013.6616464
4. Wang, J., Kourik, J.L.: Data warehouse snowflake design and performance considerations in business analytics. J. Adv. Inf. Technol. **6**(4), 212–216 (2015). https://doi.org/10.12720/jait.6.4.212-216
5. Seah, B.K., Selan, N.E.: Design and implementation of data warehouse with data model using survey-based services data. In: Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014), Luton, 2014, pp. 58–64 (2014). https://doi.org/10.1109/INTECH.2014.6927748
6. Wijaya, R., Pudjoatmodjo, B.: An overview and implementation of extraction-transformation-loading (ETL) process in data warehouse (Case study: Department of agriculture). In: 2015 3rd International Conference on Information and Communication Technology (ICoICT), Nusa Dua, 2015, pp. 70–74 (2015). https://doi.org/10.1109/ICoICT.2015.7231399
7. Moscoso-Zea, O., Andres-Sampedro, Luján-Mora, S.: Datawarehouse design for educational data mining. In: 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET), Istanbul, 2016, pp. 1–6 (2016). https://doi.org/10.1109/ITHET.2016.7760754.
8. https://dl-acm-org-bpdc.new.knimbus.com/journal/tosem
9. Zhang, C., Zhuang, L.: Association rules expand technology research based on attributes. In: 2013 IEEE 4th International Conference on Software Engineering and Service Science, Beijing, pp. 133–135 (2013). https://doi.org/10.1109/ICSESS.2013.6615272
10. Dwiandriani, F., Kusumasari, T.F., Hasibuan, M.A.: Fingerprint clustering algorithm for data profiling using pentaho data integration. In: 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, pp. 359–363 (2017). https://doi.org/10.1109/ICITISEE.2017.8285528
11. Harvy, G.A., Matitaputty, A.S., Girsang, S.M., Isa, S.M.: The use of book store GIS data warehouse in implementing the analysis of most book selling. In: 2019 7th International Conference on Cyber and IT Service Management (CITSM), Jakarta, Indonesia, 2019, pp. 1–5 (2019). https://doi.org/10.1109/CITSM47753.2019.8965404
12. Dahlan, A., Wibowo, F.W.: Design of library data warehouse using snowflake scheme method: case study: library database of campus XYZ. In: 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 318–322. IEEE (2016).