

1. Paper submission (05-09-2017)
2. Paper Notification – Minor Revision (16-10-2017)
3. Paper resubmission – Final version (24-10-2017)

1. Paper submission (05-09-2017)



Leo Willyanto <leo.willyanto@gmail.com>

ISICO 2017 submission

EasyChair <noreply@easychair.org>
To: Leo Santoso <leo.willyanto@gmail.com>

Tue, Sep 5, 2017 at 1:11 PM

Dear Leo Santoso,

Leo Santoso <leo.willyanto@gmail.com> submitted the following paper to ISICO 2017:

Academic Data Warehouse with Big Data Technology for Higher Education

You are listed as one of the authors of this paper. To enter the ISICO 2017 Web pages you should visit

<https://easychair.org/conferences/?conf=isico2017>

and enter your EasyChair user name and password.

If you forgot your user name or password, please visit

<https://easychair.org/account/forgot.cgi>

and specify leo.willyanto@gmail.com as your email address.

Best regards,
EasyChair Messenger.

Please do not reply to this email. This email address is used only for sending email so you will not receive a response.



Leo Willyanto <leo.willyanto@gmail.com>

ISICO 2017 submission 143

ISICO 2017 <isico2017@easychair.org>
To: Leo Santoso <leo.willyanto@gmail.com>

Tue, Sep 5, 2017 at 1:11 PM

Dear authors,

We received your paper:

Authors : Leo Santoso
Title : Academic Data Warehouse with Big Data Technology for Higher Education
Number : 143

The paper was submitted by Leo Santoso <leo.willyanto@gmail.com>.

Thank you for submitting to ISICO 2017.

Best regards,
EasyChair for ISICO 2017.

2. Paper Notification – Minor Revision (16-10-2017)



Leo Willyanto <leo.willyanto@gmail.com>

ISICO 2017 notification for paper 143

ISICO 2017 <isico2017@easychair.org>
To: Leo Santoso <leo.willyanto@gmail.com>

Mon, Oct 16, 2017 at 12:30 PM

Dear Mr/Ms. Leo Santoso,

On behalf of ISICO Scientific Committee, we are pleased to inform that your paper with the following details:

Paper ID: 143

Title: Academic Data Warehouse with Big Data Technology for Higher Education

has been ACCEPTED as FULL PAPER for publication in Information Systems International Conference (ISICO) on November 6 – 8, 2017 in Sanur, Bali, Indonesia. Congratulations!

Please refer to our website for detail information on the submission type

[\[http://www.isico.info/?page_id=152\]](http://www.isico.info/?page_id=152).

Please find the reviews of your paper attached at the bottom of this e-mail.

The reviews are also accessible in the Easychair system for ISICO 2017

[\[https://easychair.org/conferences/?conf=isico2017\]](https://easychair.org/conferences/?conf=isico2017).

Please note that this status is conditional unless all the feedbacks from reviewers are incorporated in the final version of your full paper (camera-ready version). All authors should take the review comments carefully into account when preparing for the camera-ready version of their paper. You will receive a detailed description of camera-ready copy preparation in a separate email message, and the details on registration are available at our website [\[http://www.isico.info/?page_id=161\]](http://www.isico.info/?page_id=161).

If you have more questions, please do not hesitate to contact us at ask@isico.info. Please notify your co-authors about the reviews because only corresponding authors receive this e-mail.

Best regards,
ISICO 2017 Organizing Committees

----- REVIEW 1 -----

PAPER: 143

TITLE: Academic Data Warehouse with Big Data Technology for Higher Education

AUTHORS: Leo Santoso

Overall evaluation: 0 (borderline paper)

----- Overall evaluation -----

Abstract:

- too general, we could not find any specific statement that outline the Background Research, Research Questions, the Research Objective, Methodologies used, Problems found, Findings & Results.

- Each can be written as one-two sentences, that will wrapped up a good abstract

INTRODUCTION

- Main Text/ or should be written as Introduction

- Introduction should express and write the background research and references, especially on Big

Data

The title is: Academic Data Warehouse with Big Data Technology for Higher Education

- no Big Data references on all section, the paper focuses on the operability of DW and DSS systems in XYZ university. It only express about the Academic Systems process in a university. What makes it Big Data, is there any evidence based on the process? What makes it special? Did it record millions of transaction from that, or just thousands of hundred thousands?

- If you say it so, you should provide a table that listed all the tables, with approximate record and the big data sides from that. Not just only said this is big data but there is no further information about the transactions or number of transactions occurred.

- From what business process the millions of transactions is generated? What Department? What Business Functions? From Academic process? From How many stakeholders/ users? How many transactions?

In what period? Daily, Weekly, Hourly? No further details.

- All Figures are too small.

- There are Figures also written (half) in Indonesian.

- This paper only implement a normal database and data warehouses for an academic information systems for higher edu. All universities now mostly have it. What makes it special, you need to address this. What part of big data from this project can be written? No further explanation about it. References mostly on Data Warehouses. What Big Data technology you used in this project? Which reference say it it is a big data problem?

- Paper needs major revision.

- The novelty from this paper is not written clearly.

----- REVIEW 2 -----

PAPER: 143

TITLE: Academic Data Warehouse with Big Data Technology for Higher Education

AUTHORS: Leo Santoso

Overall evaluation: 1 (weak accept)

----- Overall evaluation -----

1. "Before starting the application development, analysis and design of the student academic records and the needs of Top-level management are done"

The paper haven't discuss about top-level management requirements / needs.

2. "1. Main text"

Typo

3. The paper doesn't explained whether the screen capture is part of the application that being developed or third party tools (Fig 4,5,6,7,8,10). Especially when there's inconsistencies of look and feel within the captured UI. If it's third party tools, it should be mentioned accordingly.

4. Implementation phase doesn't explain the technology that being used except for PostgreSQL database.

5. Require citations for unexplained terms (e.g. MDX, OLAP tools, etc)

----- REVIEW 3 -----

PAPER: 143

TITLE: Academic Data Warehouse with Big Data Technology for Higher Education

AUTHORS: Leo Santoso

Overall evaluation: 1 (weak accept)

----- Overall evaluation -----

Editor comment

The author should discuss the result and elaborate it in detail. Fail to put an additional discussion materials to the paper resulting to the withdrawal from Procedia Computer Science Submission.

Figure 2 - 11 is too small and have low quality, the author should increase the quality of the figure and increase the size.



Leo Willyanto <leo.willyanto@gmail.com>

ISICO 2017 CAMERA READY PREPARATION GUIDELINE

ISICO 2017 <isico2017@easychair.org>
To: Leo Santoso <leo.willyanto@gmail.com>

Mon, Oct 16, 2017 at 12:35 PM

Dear Leo Santoso,

Congratulation! Your paper has been accepted for publication in ISICO 2017, and now you are in the camera-ready preparation stage. Scientific Committees of ISICO 2017 will conduct second review process of your paper with the following details:

ID : 143

TITLE : Academic Data Warehouse with Big Data Technology for Higher Education

AUTHOR(S) : Leo Santoso

Please pay attention to the following details when preparing your camera ready version:

- 1) Reviewers' comments. Please incorporate all reviewers' comments and feedbacks in your camera ready paper (only corresponding author received the review result). Please provide a HIGHLIGHT to indicate which parts of your paper that have been modified. You can simply use Ms. Word feature to add the highlight.
- 2) Title. Make sure that the title of your paper is written correctly. If reviewers require you to change the title, please make sure to change the title both on your camera ready paper and in the EasyChair system.
- 3) Corresponding Author. The corresponding author is the one to whom all the correspondences (queries, revisions, etc) are sent. Corresponding Author should be one of the authors. Please make sure that email and contact number of the corresponding author are correct. Besides, don't forget to double check the authors' name written on the paper.
- 4) Maximum page length. Please ensure that you adhere to the maximum page length limits already provided in the paper guideline (http://www.isico.info/?page_id=152), since there is an extra fee for each additional page. Please ensure that your paper format has followed the conference guideline (http://www.isico.info/wp-content/uploads/2017/08/PROCS-ISICO_2017.docm or http://www.isico.info/wp-content/uploads/2017/08/PROCS-ISICO_2017_TeX_Template.zip).
- 5) Please ensure that the revised version of your paper has been properly formatted into a pdf file and uploaded into EasyChair (by using the previous ID number).
- 6) Procedia Computer Science requires us to submit the LICENSE AGREEMENT FOR OPEN ACCESS form provided in the following URL: http://cdn.elsevier.com/promis_misc/elsevier-cc-by-nc-nd.pdf. The license agreement should be signed by one of the authors, either using a digital signature or hand signature.
- 7) Procedia Computer Science also requires us to submit the manuscript in Ms. Word or Latex format. Since EasyChair only accepts PDF files, please submit both files (license agreement and camera ready paper in Word/Latex) by email to editor-in-chief@isico.info (subject: Camera Ready and License Agreement Paper ID XX).

In addition, please note that in order to publish the article in the proceedings at least one author has to register for the conference. The due date for registration is on October 23, 2017.

Registration details are available at our website: conference registration page (http://www.isico.info/?page_id=147).

If you have more questions, please do not hesitate to contact us at ask@isico.info.

Warm Regards,
ISICO 2017 Organizing Committees

3. Paper resubmission – Final version (24-10-2017)



Leo Willyanto <leo.willyanto@gmail.com>

Camera Ready and License Agreement Paper ID 143

Leo Willyanto <leo.willyanto@gmail.com>
To: Editor In Chief <editor-in-chief@isico.info>

Tue, Oct 24, 2017 at 4:19 PM

Regards,
Leo W. Santoso

On Wed, Oct 18, 2017 at 2:27 PM, ISICO 2017 <isico2017@easychair.org> wrote:
Paper: 143
Authors: Leo Santoso
Title: Academic Data Warehouse with Big Data Technology for Higher Education

Dear Leo,

You have already received the comments by the reviewers in a previous email. Please take them carefully into account when preparing your camera-ready paper for the proceedings. The final paper was due on 10 October 2017.

This is a firm deadline for the production of the proceedings. You should submit your paper using your EasyChair proceedings author role for ISICO 2017.

The page limit is 8 pages and there is an extra fee for each additional page.

We greatly appreciate your cooperation in these matters. Thank you again for your contribution to ISICO 2017.

All questions about papers should be sent to the volume editor Isico Organizing Committee <editor-in-chief@isico.info>.

Sincerely,
Editor in Chief of ISICO 2017

2 attachments

 **Licensed Agreement for Open Access.pdf**
342K

 **PROCS-ISICO_2017 Leo.docx**
2990K



Leo Willyanto <leo.willyanto@gmail.com>

New final version of ISICO 2017 paper 143

ISICO 2017 <isico2017@easychair.org>
To: Leo Santoso <leo.willyanto@gmail.com>

Tue, Oct 24, 2017 at 4:29 PM

Dear authors,

This is to notify you that we received a new version
of your ISICO 2017 paper:

Authors : Leo Santoso

Title : Academic Data Warehouse with Big Data Technology for Higher Education

The new version has been submitted by Leo Santoso.

Best regards,
EasyChair Messenger.

JOURNAL OWNER
(hereinafter the "Journal")
LICENSE AGREEMENT FOR OPEN ACCESS

PLEASE PROVIDE US WITH THE FOLLOWING INFORMATION, REVIEW OUR POLICIES AND THE LICENSE AGREEMENT, AND INDICATE YOUR ACCEPTANCE OF THE TERMS

Manuscript number/article number (to be found in acknowledgement email you have received):	<u>143</u>
Article title:	<u>Data Warehouse with Big Data Technology for Higher Education</u>
Corresponding author:	<u>Leo Willyanto Santoso</u>
To be published in the journal:	<u>Procedia Computer Science</u>
Your organization:	<u>Petra Christian University</u>
Institute:	_____
Department:	<u>Informatics</u>
Address:	<u>Siwalakerto 121-131</u>
Address-contd:	_____
City:	<u>Surabaya</u>
Postcode:	<u>60296</u>
Country:	<u>Indonesia</u>
E-mail	<u>leo@petra.ac.id</u>
Fax:	_____

YOUR STATUS (PLEASE MARK ALL THAT APPLY)

I am the sole author of the manuscript

Please indicate if any of the below also apply to you:

I am a UK Government employee electing to license copyright

I am a UK, Canadian or Australian Government employee and Crown Copyright is claimed

I am a US Government employee and the Article is public domain and therefore the 'License of publishing rights' clause does not apply

I am a contractor of the US Government under contract number: _____ and the US Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US Government purposes

None of the above apply

I am one author signing on behalf of all co-authors of the manuscript

Please indicate if any of the below also apply to you and your co-authors:

- We are all US Government employees and the Article is public domain and therefore the 'License of publishing rights' clause does not apply
- I am a US Government employee but some of my co-authors are not
- I am not a US Government employee but some of my co-authors are
- The work was performed by contractors of the US Government under contract number: _____ and the US Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US Government purposes
- We are all UK Government employees electing to license copyright
- We are all UK, Canadian or Australian Government employees and Crown Copyright is claimed
- I am claiming Crown Copyright but some of my co-authors are not employees of the UK, Canadian or Australian Government
- I am not claiming Crown Copyright but some of my co-authors are employees of the UK, Canadian or Australian Government
- None of the above apply

- The Article is a 'work made for hire' and I am signing as an authorized representative of my employer**
Name and job title of licensee if different from corresponding author: _____

Please indicate if any of the below also apply to you:

- The Article is authored by US Government employees and is public domain and therefore the 'License of publishing rights' clause does not apply
- The work was performed by contractors of the US Government under contract number: _____ and the US Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US Government purposes
- The Article is authored by UK, Canadian or Australian Government employees and Crown Copyright is claimed
- My employer is owner of the copyright in the article organization (Please use the full name of the organization as it should appear in the copyright notice):
- None of the above apply

Signed on behalf of corresponding author

Please complete this section if you are not the corresponding author as listed above. A copy of the agreement will be sent to you and the corresponding author.

I am signing on behalf of the corresponding author

Name: Leo Willyalo Santoso

job title and company (if employer representative): lecturer

E-mail address: leow@peta.ac.id

DATA PROTECTION & PRIVACY

I understand that staff of the Copyright owner and the publisher and its affiliated companies worldwide will be contacting me concerning the publishing of the Article and occasionally for marketing purposes (unless, with respect to such marketing, I tick the box below).

I do not wish to receive news, promotions and special offers about products and services of the Copyright owner and the publisher and its affiliated companies worldwide

Please tick **one or more** of the above boxes (as appropriate), review the License Agreement and then sign and date the License Agreement in black ink.

License of publishing rights

I hereby grant to the Copyright Owner (hereinafter the "Journal") an exclusive publishing and distribution license in the manuscript identified above and any tables, illustrations or other information submitted for publication as part of the manuscript (the "Article") in all forms and media (whether now known or hereafter developed), throughout the world, in all languages, for the full term of copyright, effective when the Article is accepted for publication. This license includes the right to enforce the rights granted by this license against third parties and to sublicense such rights.

Supplemental Materials

With respect to Supplemental Materials that I wish to make accessible either through a link in the Article or on a site or through a service of the Copyright Owner, the Copyright Owner shall be entitled to publish, post, reformat, index, archive, make available and link to such Supplemental Materials on a non-exclusive basis, in all forms and media (whether now known or later developed) and permit others to do so. "Supplemental Materials" shall mean additional materials that are not an intrinsic part of the Article, including but not limited to experimental data, e-components, encodings and software, and enhanced graphical, illustrative, video and audio material.

SCHOLARLY COMMUNICATION RIGHTS

CC BY-NC-ND

I understand that I retain the copyright in the Article and that no rights in patents, trademarks or other intellectual property rights are transferred to the Journal. As the author of the Article, I understand that I shall have: (i) the same rights to reuse the Article as those allowed to third party users of the Article under the CC BY-NC-ND License, as well as (ii) the right to use the Article in a subsequent compilation of my works or to extend the Article to book length form, to include the Article in a thesis or dissertation, or otherwise to use or re-use portions or excerpts in other works, for both commercial and non-commercial purposes. Except for such uses, I understand that the license of publishing rights I have granted to the Journal gives the Journal the exclusive right to make or sub-license commercial use.

USER RIGHTS

CC BY-NC-ND

The publisher will apply the *Creative Commons Attribution-Noncommercial-NoDerivative Works 4.0 International License* (CC BY-NC-ND) to the Article where it publishes the Article in the journal on its online platforms on an Open Access basis. For further information, see <http://www.elsevier.com/about/open-access/open-access-options>.

The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-nd/4.0>.

REVERSION OF RIGHTS

Articles may sometimes be accepted for publication but later rejected in the publication process, even in some cases after public posting in "Articles in Press" form, in which event all rights will revert to the author.
See <http://www.elsevier.com/locate/withdrawalpolicy>.

Revisions and Addenda

I understand that no revisions, additional terms or addenda to this License Agreement can be accepted without the Journal's express written consent. I understand that this License Agreement supersedes any previous agreements I have entered into with the Journal in relation to the Article from the date hereof.

COPYRIGHT NOTICE

The publisher shall publish and distribute the Article with the appropriate copyright notice.

Author Representations / Ethics and Disclosure

I affirm the Author Representations noted below, and confirm that I have reviewed and complied with the relevant Instructions to Authors, Ethics in Publishing policy, and Conflicts of Interest disclosure. Please note that some journals may require that all co-authors sign and submit Conflicts of Interest disclosure forms. I am also aware of the publisher's policies with respect to retractions and withdrawal (<http://www.elsevier.com/locate/withdrawalpolicy>).

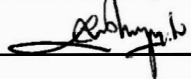
For further information see the publishing ethics page at <http://www.elsevier.com/publishingethics> and the journal home page.

Author representations

- The Article I have submitted to the journal for review is original, has been written by the stated authors and has not been published elsewhere.
The Article was not submitted for review to another journal while under review by this journal and will not be submitted to any other journal.
The Article and the Supplemental Materials contain no libellous or other unlawful statements and do not contain any materials that violate any personal or proprietary rights of any other person or entity.
- I have obtained written permission from copyright owners for any excerpts from copyrighted works that are included and have credited the sources in the Article or the Supplemental Materials.
- Except as expressly set out in this License Agreement, the Article is not subject to any prior rights or licenses, and if my or any of my co-authors' institution has any policy that might restrict my ability to grant exclusive rights under this License Agreement, a written waiver of that policy has been obtained.
- If I am using any personal details or images of patients, research subjects or other individuals, I have obtained all consents required by applicable law and complied with the publisher's policies relating to the use of such images or personal information. See <http://www.elsevier.com/patientphotographs> for further information.
- Any software contained in the Supplemental Materials is free from viruses, contaminants or worms.
- If the Article or any of the Supplemental Materials were prepared jointly with other authors, I have informed the co-author(s) of the terms of this License Agreement and that I am signing on their behalf as their agent, and I am authorized to do so.

For information on the publisher's access policies, please see <http://www.elsevier.com/access>.

I have read and agree to the terms of the License Agreement.

Signed: 

Name printed: Leo Willyanto Santoso

Title and Company (if employer representative): _____

Date: Oct 20, 2017

Please sign this License Agreement and return the completed and signed original of this form by mail or fax, or by e-mailing a scanned copy of the signed original, retaining copy for your files, to the Journal Manager email address given in the email that carried this PDF.



4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia

Data Warehouse with Big Data Technology for Higher Education

Leo Willyanto Santoso*, Yulia

Petra Christian University, Siwalankerto 121-131 Surabaya, 60236, Indonesia

Abstract

Nowadays, data warehouse tools and technologies cannot handle the load and analytic process of data into meaningful information for top management. Big data technology should be implemented to extend the existing data warehouse solutions. Universities already collect vast amounts of data so the academic data of university has been growing significantly and become a big academic data. These datasets are rich and growing. University's top-level management needs tools to produce information from the records. The generated information is expected to support the decision-making process of top-level management. This paper explores how big data technology could be implemented with data warehouse to support decision making process. In this framework, we propose Hadoop as big data analytic tools to be implemented for data ingestion/staging. The paper concludes by outlining future directions relating to the development and implementation of an institutional project on Big Data.

© 2018 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017.

Keywords: data warehouse, big data, academic, hadoop, higher education, analysis

1. Introduction

Higher educations are working in a more and more complex and competitive environment. They have to compete with other institutions to answer to national and global economic, political and social changes. Moreover, different stakeholders are expecting higher education institutions to produce right solution in a timely manner to these demands. To overcome this condition, higher education needs to produce the right decisions required for dealing with these rapid changes by analyzing vast data sources that have been generated. Most of higher education institution invest enormous resources in information technology to implement data warehouse system [1].

* Corresponding author. Tel.: +62 31 2983455.

E-mail address: leow@petra.ac.id

The development of data warehouse is a way to extract the important information from the scattered data in some information systems into a centralized integrated storage and support the need for data history. This integrated data can be utilized for information delivery activities that can be reviewed from various dimensions and can be set the level of detail.

Further utilization of the information contained in the data warehouse is the activity of data analysis using certain techniques and methods. There are several algorithm for knowledge data discovery, like classifying, clustering and mining [2]. The data contained in the data warehouse can used as input for the application system for example like a dashboard. With the existence of this dashboard is expected to be a solution for the learning process to monitor the academic condition and then could take the right decision. However, organizations are recognizing that traditional data warehouse technologies are dying to meet new business requirements, especially around streaming data, real-time analytics, large volumes of unstructured and complex data sets.

To solve this problem, this paper aims to design and implement a modern data warehouse for academic information system to support decision making process. The designed system accommodates Hadoop platform, a powerful analytical tools which is able to produce a graph that displays the student data information statistically. To support parallel and distributed processing of large volumes of data, most solutions involve Hadoop technology. Hadoop is capable to perform analysis of large heterogeneous datasets at unprecedented speeds

As a result, top management will have a dashboard to monitor the existing condition of the academic atmosphere of university. The reporting dashboard itself will cover operational, strategic and analytical dashboard. The operational dashboards will tell us what is happening now, while strategic dashboards will track key performance indicators in academic process. Moreover, analytical dashboards will process data to identify trends.

The main contributions of this paper are as follows: (1) the designed system enables the communication among different platform and datasets, including smart phones, web, and desktop application whether it is structured, semi-structured and unstructured data. 2) The system provides solution to the top level management in order to know the academic condition in their university. 3) the proposed system could be implemented to other university who need a decision support system for big data.

The remaining part of this paper is organized as follows. Section 2 presents the background and the related work. Section 3 presents the design of the system and section 4 present the testing of the proposed system. Finally, the conclusions are drawn in section 5.

2. Traditional Data Warehouse and Modern Data Warehouse

This section describes about traditional data warehouse and modern data warehouse. The differences between them are also discussed.

Data Warehouse is the combination of concepts and technologies that facilitate organizations to manage and maintain historical data obtained from operational and transactional applications [3]. It helps knowledge workers (executives, managers, analysts) to make quicker and more informed decisions. Data Warehouse is a new paradigm in strategic decision making environment. Data Warehouse is not a product but an environment in which users can find strategic information [4]. Data Warehouse is a place to store information that is devoted to help make decisions [5]. The Data Warehouse contains a collection of logical data separate from the operational database and is a summary. Data Warehouse allows the integration of various types of data from a variety of applications or systems. This ensures a one-door access mechanism for management to obtain information and analyze it for decision making. Data Warehouse has several characteristics [5, 6]: subject-oriented, integrated data, nonvolatile, time-variant, and not normalized.

Data Warehouse used data modeling technique called dimensional modeling technique. Dimensional modeling is a call-based model that supports high-level query access. Star Schema is a form of dimensional modeling scheme that contains a fact table at its center and dimensional tables. Fact table contains descriptive attribute that is used for query and foreign key process to connect to dimension table. Decision analysis attributes consist of performance measures, operational metrics, aggregate sizes, and all other metrics needed to analyze organizational performance. Fact table shows what is supported by data warehouse for decision analysis. The dimension table contains attributes that describe the entered data in the fact table.

Extract, Transform, and Load (ETL) is a data integration process that extracts data from outside sources, transforms the data according to business needs, and stores it into data warehouse [4]. The data used in the ETL process can come from a variety of sources including enterprise resource planning (ERP) applications, flat files, and spreadsheets.

Data warehouse support decision support system. Decision Support Systems (DSS) is a computer-based system that helps decision makers use the data and models available to solve problems [7]. DSS functions combine the resources of each individual with the ability of the computer to improve the quality of the decision. DSS requires data coming from various sources to solve the problem. Every problem needs to be solved and every opportunity and strategy requires data. Data is the first component of the DSS architecture. The data relate to a state that can be simulated using a model that is the second component of the DSS architecture. Some systems also have knowledge which is the third component of the DSS architecture. The fourth user interacts with the system through a user interface which is the fifth component in the DSS architecture. In building the DSS, it is necessary to plan a mature system accompanied by the preparation and incorporation of components well.

Data warehouse is widely implemented, including in the education industry. It is possible to implement data warehouse for typical university information system [8]. Academic data warehouse supports the decisional and analytical activities regarding the three major components in the university context: didactics, research, and management [9]. Data warehouse has important role in educational data analysis [10].

With the arriving of big data, traditional data warehouse cannot handle large amount of data [11]. In the past, educational data has been gathered mainly through academic information system and traditional assessments. However, it is increasingly being gathered through online educational systems, educational games, simulations and social media now. Huge workload, concurrent users and data volumes require optimization of both logical and physical design. Therefore, data processing must be in parallel. Moreover, traditional data warehouse cannot extract unstructured data that has varying data structure into information. Traditional data warehouse was design with the purpose of integrating structured data from transactional sources that is supported by OLAP-based analysis. It is the opportunity for big data technology to solve the problem. The integration between big data technology such as Hadoop and data warehouse is very important. To support parallel and distributed processing of large volumes of data, most solutions involve Hadoop technology [12, 13]. Hadoop is capable to perform analysis of large heterogeneous datasets at unprecedented speeds.

The Table 1 summarizes the characteristics of traditional data warehouse and modern data warehouse, from the several point of views like the purpose, data sources, scope, architecture, technology, and end-user.

Table 1. The characteristic of traditional data warehouse and modern data warehouse

Characteristics	Traditional Data Warehouse	Modern Data Warehouse
Purpose	Treatment of collected data for a specific business area that is integrated, non-volatile and time-varying. It supports decision-making process.	Processing of structured, semi-structured, and unstructured data, from diverse sources and the volume of data exceeds the ability of traditional tools to capture, store, manage and analyse them.
Data source	Usually transactional and operational databases.	Various sources and data types (social media, sensors, blog, video, and audio).
Scope	The integrated structured data to support Business Intelligence (BI) and Online Analytical Processing (OLAP).	Analyse and discover knowledge from large volumes of data characterized by the 4Vs (volume, velocity, variety and veracity)
Architecture	Oriented to processes of extraction, transform and load (ETL). Star schema is the appropriate solution for the architecture.	The architecture is depends on the problem. There is still no reference architecture or standardized terminology. They are some proprietary and product-oriented architectures from the vendor.
Technology	The technology is mature and tested tools in large amount applications, both free and licensed software.	The technology is still growing. Hadoop is one of the open-source software framework used for distributed storage and processing of dataset of big data
End-user	Business analysts or top managements who do not require specific knowledge of technologies or data exploration.	Data scientists with knowledge in technologies, algorithms, mathematics and statistics.

3. System Design

ETL is the main process in traditional data warehouse technology which cannot handle unstructured data. In this system, we need a flexible ETL process which can handle several data quality issues, as for instance duplicated data, inconsistency data, and garbage data. The proposed system can be seen in Fig 1. In the system, there is a combination between Hadoop and RDBMS. Hadoop can enhance RDBMS as data ingestion/staging tool, but also as data management and data presentation platform.

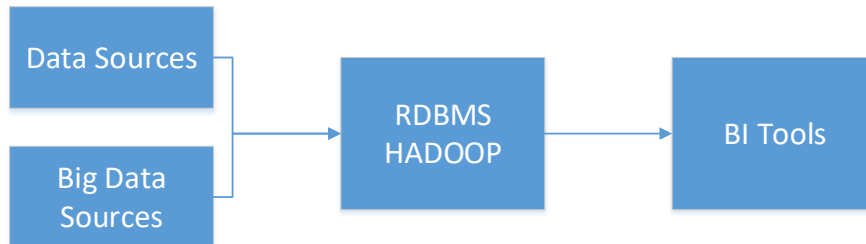


Fig. 1. The proposed system

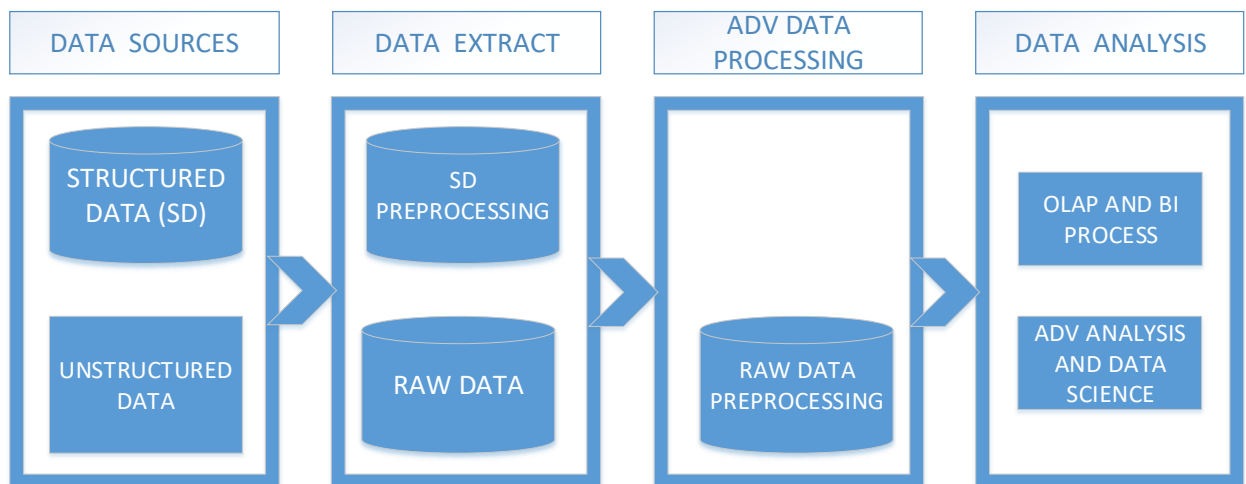


Fig. 2. The Architecture of System

As is shown in Figure 2, the architecture of proposed system was presented. Structured data are aggregated into our schema, while unstructured data is unpredictable data, and usually does not have an easily computer-recognizable format. The examples of unstructured data are free-text, images, videos, webpages, RSS feeds, meta data and web server logs. In our proposed system, the unstructured data will be processed by performing categorization and filtering and then it will store in the contextualized data. The uncategorized data will remain in the raw data. Next, through a process of searching for relationships or patterns, the data in the contextualized data will store into related data. Then, the related data that are already processed and capable to be adapted to predefined structures will be store into explored data. Finally, it is possible to integrate between explored data and aggregate data to be analyzed using OLAP techniques and business intelligence.

4. Implementation and Testing

In this section will be discussed about the implementation of the system in accordance with the analysis and system design. The structured data comes from PostgreSQL databases, while unstructured data comes from social media such as Facebook, twitter and LinkedIn.

Figure 3 shows the analysis page. In this application, users could create new analysis so the report can be customised as they need. In every analysis, it is possible to produce some graphs or charts to support the generated report. Some advanced users need OLAP Navigator and MDX Query Editor to create powerful report.

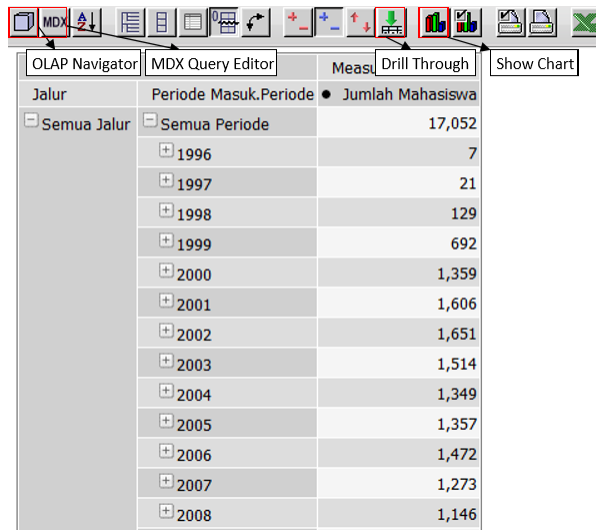


Fig. 3. Analysis page of DSS Application

The sample chart can be seen in Figure 4. User can customised the type of the chart, so the generated report will be more meaningful for the reader.

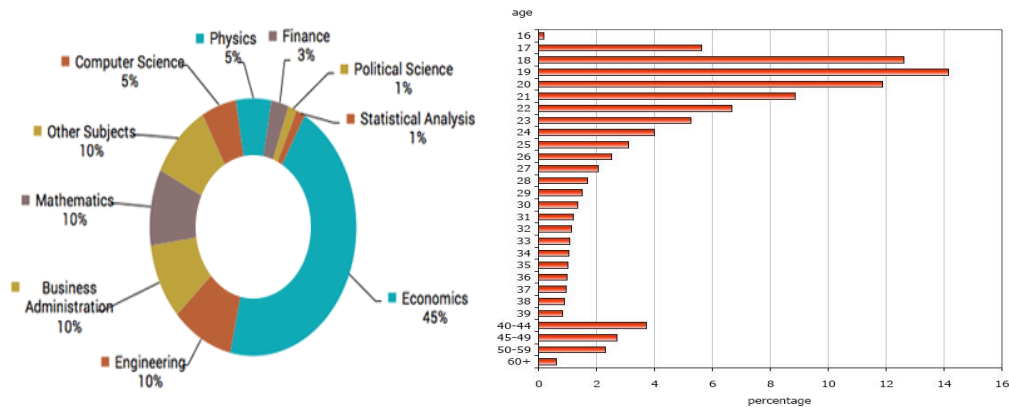


Fig. 4. Chart page of DSS Application

The more advanced interface for analytical feature can be seen in Figure 5. Advance users could drag and drop item dimensions in the left panel, then put into the column, row or filter in the right panel to produce the insightful report.

Angkatan	Jumlah Mahasiswa Laki-Laki	Jumlah Mahasiswa N/A	Jumlah Mahasiswa Perempuan
2008	327		409
2009	374		422
2010	416		472
2011	434	2	510
2012	427		527

Fig. 5. Advanced Analysis of DSS Academic

Questionnaires were distributed among thirty university staffs which cover from top management, middle management and bottom management. Rector and vice rectors are categorized as top management. In the middle management, it contains dean, vice dean and their staffs. Head of departments are grouped into bottom management. The assessed indicators include application interface, graphic customization features, ease of use of the application, ability to meet the user needs, and overall application. Detailed assessment of application usage can be seen in Table 2.

Table 2. Assessment of Application Usage

Indicators	1	2	3	4	5
User Experience			7	15	8
Graphics customization feature			3	20	7
Ease of using			4	16	10
Applications meet the needs			2	23	5
Overall application			1	25	4

Description of the rating scale:

- Value 1: Very bad
- Value 2: Bad
- Value 3: Enough
- Value 4: Good
- Value 5: Very good

From the questionnaire responses, 77% of users has good user experience. Mostly, the respondents said the graphics customization feature is good. According to the users, 87% of application is easy to use. 93% of users said the applications meet the business requirements. Overall, 97% of survey respondents said that the application is good.

5. Conclusions

This paper has explored modern data warehouse which could substitute traditional data warehouse which cannot handle big data in educational system. The big data technology approach to data warehouse will help reduce difficulties associated with traditional data analysis. Moreover, this has the potential of enriching the education system with new learning ways, and making decision making by policy makers more effective and efficient.

References

- [1] Santoso, Leo Willyanto and Yulia. (2014) “Analysis of the impact of information technology investments - a survey of Indonesian universities.” *ARNP JEAS* **9** (12): 2404-2410.
- [2] Santoso, Leo Willyanto (2011) “Classifier Combination for Telegraphese Restoration.” In: *International Conference on Uncertainty Reasoning and Knowledge Engineering*, Denpasar - Indonesia. DOI: 10.1109/URKE.2011.6007844
- [3] Golfarelli, Matte and Stefano Rizzi. (2009) “A Survey on Temporal Data Warehousing.” *Int Journal of data Warehousing & Mining* **5**(1): 1–17.
- [4] Ponniah, Paulraj (2010) “Data Warehousing: a Comprehensive Guide for IT Professional.” 2nd ed., New York: The McGraw-Hill Companies.
- [5] Reddy, G. Satyanarayana, Rallabandi Srinivasu, M. Poorna Rao, and Srikanth R. Rikkula. (2010) “Data Warehousing, Data Mining, OLAP and OLTP Technologies Are Essential Elements To Support Decision-Making Process In Industries”. *Int. Journal on Comp. Sci. and Eng.* **2**(9): 2865-2873.
- [6] Gour, Vishal, S.S. Sarangdevot, G.S. Tanwar, and A. Sharma. (2010) “Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse”. *Int. Journal on Comp. Sci. and Eng.* **2**(3): 786-789.
- [7] Turban, Efraim, Jay E. Aronson, Ting-Peng. Liang, and R. Sharda (2004) “Decision Support and Business Intelligence Systems”, 10th ed., New Jersey: Pearson Education, Inc.
- [8] Youssef, Bassil (2012) “A Data Warehouse Design for A Typical University Information System”, *Journal of Comp. Sci. & Research* **1**(6): 12-17, Dec. 2012.
- [9] Dell’Aquila, Carlo, Francesco Di Tria, Ezio Lefons, and Filippo Tangorra. (2007) “An Academic Data Warehouse”, *Proceedings of the 7th WSEAS Int. Conf. on Applied Informatics and Communications*, Athens, Greece, August 24-26.
- [10] Mirabedini, Shirin (2014) “The Role of Data warehousing in Educational Data Analysis”, *Journal of Novel Applied Sciences* **3**(5): 1439-1445.
- [11] Salinas, Sonia Ordóñez and Alba C.N. Lemus. (2017) “Data Warehouse and Big Data integration” *Int. Journal of Comp. Sci. and Inf. Tech.* **9**(2): 1-17.
- [12] S. G. Manikandan and S. Ravi. (2014) “Big Data Analysis Using Apache Hadoop,” in *International Conference on IT Convergence and Security (ICITCS) 2014*, pp. 1–4.
- [13] J. Nandimath, E. Banerjee, A. Patil, P. Kakade, and S. Vaidya. (2013) “Big data analysis using Apache Hadoop,” *2013 IEEE 14th Int. Conf. Inf. Reuse Integr.*, pp. 700–703, 2013.