



Summer student project report

Implementing database system for LHCb publications page

Supervisor: Marc Oliver Bettler

Fakhriddin Abdullayev

CERN summer student intern from June 19 to August 11, 2017

Switzerland

Acknowledgments

I would like to thank Marc Oliver Bettler for the excellent supervision and important advises provided during my working time at LHCb experiment. Furthermore, I am thankful for Ben Couturier and Marco Clemencic for their valuable suggestions and feedback for the program.

I am very thankful for LHCb secretariat and all people who gave us a chance to have this wonderful experience.

1. Introduction
2. Project details and goals
 - 2.1 Description of the Page
 - 2.2 Information collection on the page
 - 2.3 Choosing Database system
 - 2.4 Implementation process
3. Results and discussion
4. Conclusion
5. Future Work

Abstract

The LHCb is one of the main detectors of Large Hadron Collider, where physicists and scientists work together on high precision measurements of matter-antimatter asymmetries and searches for rare and forbidden decays, with the aim of discovering new and unexpected forces. The work does not only consist of analyzing data collected from experiments but also in publishing the results of those analyses. The LHCb publications are gathered on LHCb publications page to maximize their availability to both LHCb members and to the high energy community.

In this project a new database system was implemented for LHCb publications page. This will help to improve access to research papers for scientists and better integration with current CERN library website and others.

1 Introduction

The screenshot shows the LHCb Public results website. The header includes the LHCb logo and the text "The LHCb Public results". Below this, there is a section titled "LHCb publications" with a link "[to public page]". A sidebar on the left lists various topics: TOOLS, LHCb SLIDES, SB GUIDELINES, PUBLICATIONS PER WORKING GROUP, IONS AND FIXED TARGET, FLAVOUR TAGGING, b-HADRONS AND QUARKONIA, B DECAYS TO CHARMONIUM, DETECTOR PERFORMANCE, CHARMLESS b-HADRON DECAYS, and QCD, ELECTROWEAK AND. The main content area displays a "List of papers (Total of 385 papers and 18691 citations) [plot]". A table lists several papers with their titles, document numbers, journals, submission dates, and citation counts.

TITLE	DOCUMENT NUMBER	JOURNAL	SUBMITTED ON	CITED
Observation of D^0 meson decays to $\pi^+\pi^-\mu^+\mu^-$ and $K^+K^-\mu^+\mu^-$ final states	PAPER-2017-019 arXiv:1707.08377 [PDF]	PRL	26 Jul 2017	
Study of prompt D^0 meson production in pPb collisions at $\sqrt{s} = 5.76$ TeV	PAPER-2017-015 arXiv:1707.02750 [PDF]	JHEP	10 Jul 2017	
Updated branching fraction measurements of $B_{(s)}^0 \rightarrow K_S^0 h^+ h^-$ decays	PAPER-2017-010 arXiv:1707.01665 [PDF]	JHEP	06 Jul 2017	
Observation of the doubly charmed baryon Ξ_{cc}^{++}	PAPER-2017-018 arXiv:1707.01621 [PDF]	PRL	06 Jul 2017	9 [plot]
Prompt and nonprompt J/ψ production and nuclear modification in pPb collisions at $\sqrt{s_{NN}} = 8.16$ TeV	PAPER-2017-014 arXiv:1706.07122 [PDF]	PLB	21 Jun 2017	1
Study of charmonium production in b-hadron decays and first evidence for the	PAPER-2017-007 arXiv:1706.07013	EPJC	21 Jun 2017	

The LHCb publications website has two public pages. First one is public page

http://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCbProjectPublic/Summary_all.html

and second one is restricted-access page

https://lhcbproject.web.cern.ch/lhcbproject/Publications/l/Summary_all.html

One is public page – where people can view published papers and a collection of the most important features, the figures and the tables with their captions. The second is restricted-access page – is the page where the LHCb members have access to more information than the general public. The restricted information consist in the review mailing list and the names of the contacts and reviewers for the example. The scientific papers for the LHCb publications up until now was saved in Python Pickle file which was developed by Marc Oliver Bettler. The goal was to integrate the website with database system to improve the scalability of the system as the number of publication only increases, to reduce the RAM memory requirement, and to allow for a better access for other interested CERN services, such as the CERN Analysis Preservation project.

2 Project details and goals

The aim of this work was to implementing a database system to LHCb publications page and further integrate with other CERN websites. The process was to analyze current Python code, analyze publicly available database systems and choosing the appropriate one for the project.

2.1 Description of the page

LHCb – is one of the main 4 detectors of the CERN and there are many experiments are being performed at the center. The final product of these research projects are published on science related websites. One of them is specifically designed for LHCb publications. The page is open for public and can be used by scientists and anyone interested in the LHCb research papers.

2.2 Information collection from different resources

LHCb publications page is a final page for scientific papers which have been peer reviewed and approved by LHCb members. The information in the articles are collected from following resources:

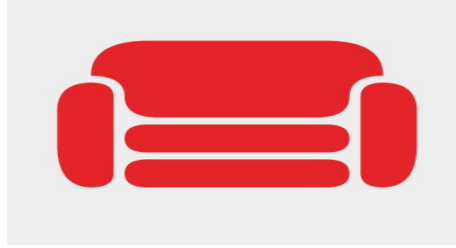
arXiv, INSPIRE, CDS (CERN document server), SVN (LHCb Docs), the LHCb Editorial Board, HEP Data, Rivet. The information is collected in Python Pickle file.

2.3 Choosing a database system

The process of choosing database system was very important in the project. There were many requirements for the database systems that we wanted to implement in the system. They were:

- a) Easy to use, lightweight – so that a developer would not spend much time and would be easy to understand.
- b) Free (Open source and No cost) – an important aspect for the project. Since we did not want use additional resources to implement it.
- c) Document based – because we wanted to implement it for LHCb publications paper we did not need to use relational based system.

d) Interface to Python – easily to integrate with current existing system.



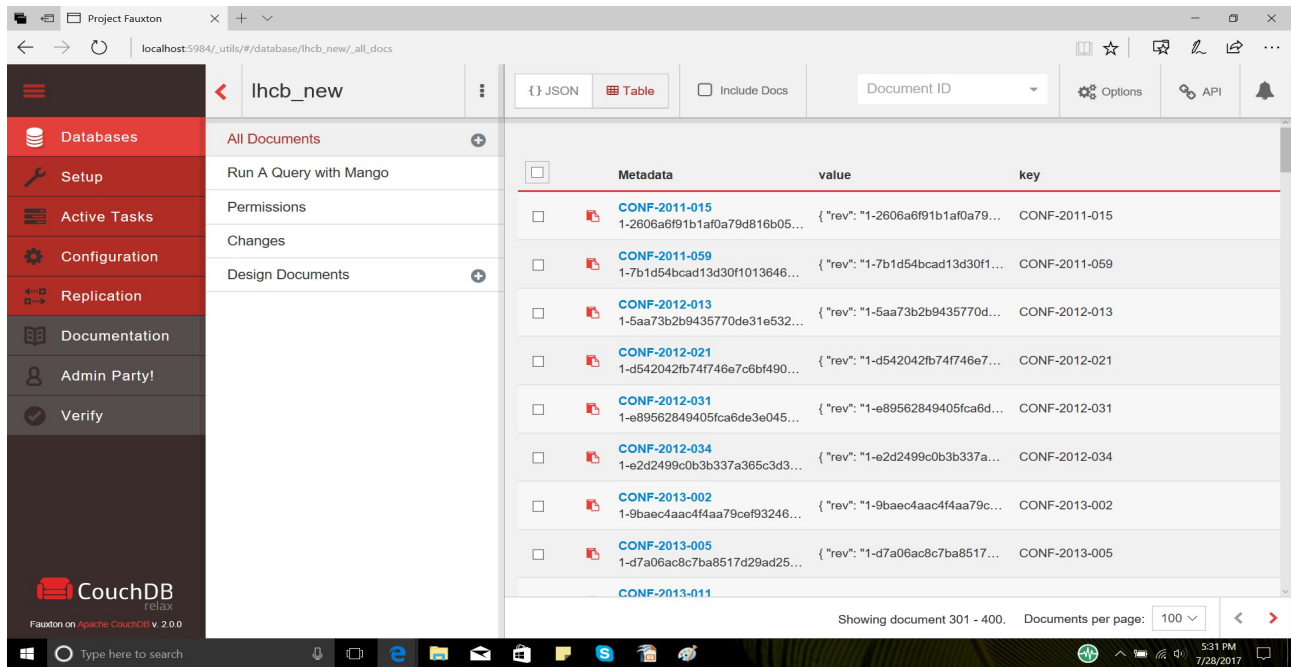
CouchDB – is a database system which is document based, open source and non relational based database system. We decided to make our choice on this and work to integrate it. LHCb software developers team has already used CouchDB in implementing it to internal use Ben Couturier and Marco Clemencic have shared the experience and given valuable advice on it.

2.4 Implementation process

The database system has been successfully integrated with Python, and exported most of the scientific papers from Python Pickle to a new CouchDB server. From CouchDB, the information will be accessed through Python code and it will ease the access to other systems. The importance of the database system is that it shows the last edited paper on the system. It is secure and only administrator can give access to the certain users to edit and change the current papers in the system.

3 Results and discussions

The result of the work has been exportation of current available research papers from LHCb publications page to the database system. A python code has been developed to export scientific papers to the system. The code does not only export to the system but also can make a comparison with other existing files before exporting is done. Furthermore, preparatory work was performed to interface the DB with the main program.



4. Conclusion

In this work, the implementation of the database systems were analyzed, and CouchDB was chosen as the most suitable one. The work also includes writing Python code for integration and exporting research papers to the database system.

5. Future Work

Final steps for the work will be fully integration of the database system with CERN and other publication websites.