РЕЦЕНЗИЯ

на выпускную квалификационную работу обучающегося СПбГУ

Сдобниковой Анны Михайловны

«Проектирование структуры базы данных для проекта NICA»

В своей выпускной работе Анна Михайловна рассмотрела структуры базы данных для лаборатории магнитных измерений в рамках разработки Проекта NICA–российского ускорительного комплекса. Следует отметить, что, строящийся комплекс NICA в Объединенном институте ядерных исследований (ОИЯИ) в городе Дубна Московской области, представляет собой очень сложный объект, в рамках которого необходимо решать самые различные подзадачи. Именно в рамках этого проекта в данной работе разработана структура хранилища для результатов магнитных измерений. Особое внимание в работе уделяется практическим применениям.

В Первой Главе автор особое внимание уделила исследованию методам и инструментам существующих систем сбора и обработки данных соответствующих данных. В представленной работе используется технологии и методы, которые апробированы в Лаборатории магнитных измерений (ОИЯИ) для проведения тестирования сверхпроводящих магнитных элементов как бустера (предускорителя), так и коллайдера. Кроме того, в рамках проводимых исследований приводится актуализация необходимости дополнительной разработки баз данных, необходимых как для структуризации хранения данных, так и для «удобства» обработки получаемых данных.

Вторая Глава посвящена как выбору и созданию СУБД, так и разработке архитектуры хранилища. В своей работе Анна Михайловна, использует реальные «потребности» оптимального распределения данных по трем слоям структуры в зависимости от степени обработки данных:

– слой первичных данных (содержит «сырые» данные, собранные с датчиков измерительной системы);

– слой обработанных данных (содержит все искомые параметры магнитов, полученные после вычислительной обработки);

– слой витрины данных (содержит выборки и агрегированные данные, собранные из слоя обработанных данных).

Предложенная в работе модель архитектуры хранилища может быть использована с учетом реальных (практических) данных. В работе детально описаны необходимая последовательность распределения используемых данных по слоям. Кроме того, в рамках необходимых исследований в работе предложена схема распределения данных по таблицам в слое уже «обработанных данных» с целью повышения эффективности их использования. Для реализации выбрана колоночная распределенная СУБД Vertica.

В Третьей Главе рассмотрены проблемы загрузки данных в хранилище и соответствующего распределения потоков данных по слоям. Приведены конкретные решения по оптимизации ETL-процессов. Для этой цели в работе используется инструмент (workflow менеджер) Apache Airflow. Кроме того, в дополнение к работе повышения эффективности использования хранилища, Анна Михайловна предлагает осуществить процедуры доработки используемых инструментов обработки магнитных измерений с целью реализации вычислительной обработки «внутри» используемой базы данных. Необходимо отметить, что отметить, что Анна Михайловна опубликовала статьи с исследованием системы магнитных измерений, что подтверждает квалификацию магистра.

Представленная работа написана хорошим языком, содержит необходимые сведения в рамках проводимых исследований. Само изложение написано связным и лаконичным языком с учетом достаточно сложных задач, решаемых в рамках изучаемой проблемы. В целом, предлагаемая структура работы и выбранные используемые инструменты обладают необходимым потенциалом как для дальнейшего развития, так и интеграции в более сложные системы обработки и хранения данных, например, для сбора данных при проведении столкновения пучков в коллайдере NICA. В результате проведенного анализа представленной работы и учитывая очень серьезное отношение к достаточно сложным объектам и технологиям представленная к защите Выпускная Квалификационная работа Сдобниковой Анны Михайловны несомненно представляет собой законченное исследование и заслуживает оценки «отлично».

Учитывая востребованность как полученных Анной Михайловной результатов, так и подобных исследований в дальнейшем, считаю, что как отношение Анны Михайловны к работе, так и полученные результаты позволяет оценить представленную работу оценкой «отлично» и рекомендовать Анну Михайловну к поступлению в аспирантуре факультета ПМ-ПУ.

Научный руководитель, доктор физико-математических наук, профессор, заведующий кафедрой КМиМС Андрианов С.Н.

REVIEW

for the final qualification work of the student of St. Petersburg State University Anna Sdobnikova

“Designing the database structure for the NICA project”

In her graduation work, Anna Mikhailovna examined the database structures for the magnetic measurement laboratory as part of the development of the NICA – Russian Accelerator Complex Project. It should be noted that the NICA complex under construction at the Joint Institute for Nuclear Research (JINR) in the city of Dubna, Moscow Region, is a very complex object, within which it is necessary to solve a variety of sub-tasks. It was within the framework of this project that the structure of the storage for the results of magnetic measurements was developed in this work. Particular attention is paid to practical applications.

In the First Chapter, the author paid special attention to the study of the methods and tools of existing systems for collecting and processing data of relevant data. In this work, we use technologies and methods that were tested at the Laboratory of Magnetic Measurements (JINR) for testing superconducting magnetic elements of both a booster (pre-accelerator) and a collider. In addition, as part of the research, the need for additional development of the databases necessary for both structuring data storage and for the “convenience” of processing the data obtained is updated.

The second Chapter is devoted to both the selection and creation of a DBMS, and the development of a storage architecture. In her work, Anna Mikhailovna uses the real "needs" of the optimal distribution of data across three layers of the structure, depending on the degree of data processing:

- primary data layer (contains “raw” data collected from sensors of the measuring system);

- a layer of processed data (contains all the sought-after magnet parameters obtained after computational processing);

- data showcase layer (contains samples and aggregated data collected from the processed data layer).

The storage architecture model proposed in this work can be used taking into account real (practical) data. The paper describes in detail the necessary sequence of distribution of the data used by layers. In addition, as part of the necessary research, a scheme was proposed for distributing data among tables in a layer of already “processed data” in order to increase the efficiency of their use. For the implementation, the selected distributed distributed DBMS Vertica.

The Third Chapter discusses the problems of loading data into the storage and the corresponding distribution of data flows across layers. Specific solutions for optimizing ETL processes are presented. For this purpose, the tool (workflow manager) Apache Airflow is used in work. In addition, in addition to working to increase the efficiency of storage utilization, Anna Mikhailovna proposes to carry out procedures for finalizing the tools used for processing magnetic measurements in order to implement computational processing “inside” the database used. It should be noted that it should be noted that Anna Mikhailovna published articles examining the system of magnetic measurements, which confirms the qualification of the master.

The presented work is written in good language, contains the necessary information in the framework of ongoing research. The presentation itself is written in a coherent and concise language, taking into account rather complex problems that are solved within the framework of the problem being studied. In general, the proposed work structure and the selected tools used have the necessary potential for both further development and integration into more complex data processing and storage systems, for example, for data collection during beam collisions in the NICA collider. As a result of the analysis of the presented work and given the very serious attitude to quite complex objects and technologies, the final qualification work of Anna Mikhailovna Sdobnikova, presented to the defense, undoubtedly represents a completed study and deserves an excellent rating.

Considering the relevance of both the results obtained by Anna Mikhailovna and similar studies in the future, I believe that both the attitude of Anna Mikhailovna to the work and the results obtained make it possible to evaluate the presented work with an “excellent” mark and recommend Anna Mikhailovna for admission to postgraduate studies at the faculty of PM-PU.

Supervisor, Doctor of Physics and Mathematics, Professor, Head of the Department of CMeMS

Andrianov S.N.