*М.Р. Лапшин, асп.; рук. С.Т. Лескин, д.т.н, проф.,*

*рук. А.О. Скоморохов д.т.н, проф.*

**(ИАТЭ НИЯУ МИФИ, г. Обнинск)**

Подготовка данных для проведения диагностики состояния ГЦН 3-го блока Калининской АЭС

ГЦН - это сложный насосный агрегат, который состоит из рабочего колеса, вала, корпуса, подшипников, уплотнений, электродвигателя, теплообменника и т.д. За состоянием всех составляющих насоса необходимо вести постоянный контроль, так как останов ГЦН вследствие выхода из строя приведет к большим экономическим потерям. На каждый ГЦН устанавливается свыше 50 датчиков контроля с различных систем. Данные накапливаются, но не анализируются на предмет выявления совокупных зависимостей элементов оборудования друг от друга, выявления неочевидных и скрытых тенденций развития аномалий.

В докладе представлены результаты обработки эксплуатационных данных полученных на ГЦН 3-го блока КАЭС с использованием программ с открытым исходным кодом и утилит системы *LINUX*. Рассматривается первый этап анализа данных - подготовка и приведение снятых с приборов показаний в удобный для последующего детального анализа вид. Около 70% от времени интеллектуального анализа данных тратится на подготовку данных[2].

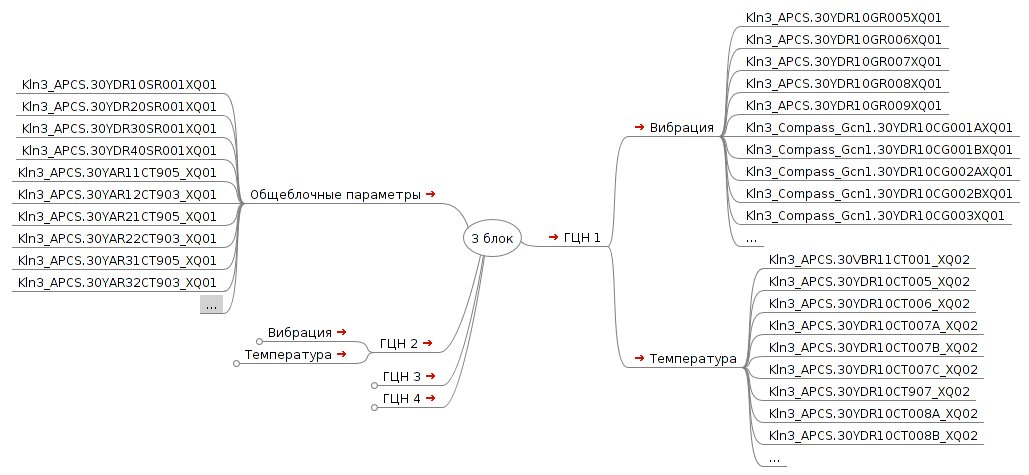
На каждом ГЦН установлено свыше 50 датчиков контроля.

Исходные данные представлены в текстовом файле, структура которого иллюстрируется приведенной ниже таблицей.

**Таблица 1 – Содержание файла**

|  |  |  |  |
| --- | --- | --- | --- |
| **Название** | **дата-время** | **значение** | **служеб.параметры** |
| гцн1\_датчик1 | 2011-01-01 00:00:00 | 285,023 | ... |
| гцн1\_датчик2 | 2011-01-01 00:00:01 | 34,8562 | ... |
| гцн3\_датчик3 | 2011-01-01 00:01:01 | NULL | ... |
| общ\_параметр1 | 2011-01-01 00:01:01 | 2998,84 | ... |
| гцн2\_датчик2 | 2011-01-01 00:01:02 | 34,4822 | ... |
| ... | ... | ... | ... |

Всего таблица содержит 13 столбцов, включая служебные параметры (например, характеристики качества сигнала). Всего в файле представлено 38.5 млн. строк, что соответствует периоду работы энергоблока с 5 марта 2011 по 19 ноября 2013 (три кампании реактора). Общий объем файла составляет 6.2 Гб. Структура датчиков представлена на рис. 1.



**Рис. 1 - Структура датчиков**

Сложность подготовки данных для анализа заключается в следующем:

* извлечение относящихся к одному датчику данных, которые находятся в разным местах таблицы;
* различие в частоте опроса разнообразных датчиков, что приводит к отличным по количеству измерений временным сечениям;
* использование алгоритмов сжатия данных, что приводит к рассинхронизации сигналов однотипных датчиков с разных ГЦН;
* отсутствие в файле сортировки сигналов по времени;
* необходимость преобразования формата времени в непрерывную числовую шкалу (например, в секундную) для упрощения вычислительных операций;
* необходимость устранения из таблицы лишних столбцов (например, служебных параметров);
* использование запятой в качестве десятичного разделителя в виде запятой;
* наличие в таблице пропущенных значений (обозначенных как неопределенные значения *NULL*);
* большое количество потенциально ошибочных значений и выбросов.

Далее приводится набор инструментов *LINUX*, которые использовались для подготовки данных.

Выполнение утилит *cat, head, tail* и программы просмотра текстовых файлов *less* в терминале (командной строке) над исходным файлом дает первичное представление о структуре исходного файла.

С помощью потокового текстового редактора *sed*, который позволяет изменять каждую строку файла по определенному правилу, меняется десятичный разделитель с запятой на точку без перенаправления в другой файл, как показано ниже.

***sed -i 's/,/./g' исходный файл.txt***

Утилита выбора столбца *cut*, в случае файла исходного файла позволяет выбрать 1-3 столбцы, необходимые для дальнейшего анализа.

***cut -f 1,2,3 исходный файл.txt > 1\_3.txt***

Утилита командной строки *grep* действует по принципу "искать и выводить строки, соответствующие регулярному выражению". В работе используется для вырезания сигналов *NULL*.

***grep -v 'NULL' 1\_3.txt > 1\_3\_nonull.txt***

Остались строки, которые не содержат значения *NULL*. Полученный файл *1\_3\_nonull.txt* представляет собой таблицу с тремя столбцами (название датчика, дата-время, значение).

C помощью скрипта (wcount), написанного на языке awk, определяется количество измерений для каждого датчика. Для чего достаточно использовать первый столбец с именами датчика (names.txt). На рис. 2 показано количество измерений для всех датчиков в полулогарифмической шкале. Что подчеркивает проблематику анализа данных, описанную выше.

***awk -f wcount names.txt | cut -f 2 > n-измерений.txt***



**Рис. 2 - Количество измерений с каждого датчика**

Для дальнейшего разбиения файла на отдельные части, в которых будут содержаться сигналы с определенного датчика, используется скрипт, написанный на языке *bash*.

**mkdir -p $3  
for name in $(cat $1); do  
 var1=$(echo $name | cut -d';' -f 1)  
 var2=$(echo $name | cut -d';' -f 3)  
 grep -w $var1 $2 > $3/${var2}.txt  
done**

Пример использования:

***. grep.sh list.txt 1\_3\_nonull.txt gcn/g1/temp/1\_3***

Теперь в каталоге *gcn/g1/temp/1\_3* лежат все датчики температур по 1 - ГЦН отдельно по файлам. Каждый файл имеет удобное название и представляет собой таблицу из трех столбцов: аббревиатура датчика (одинаковая во всех строках), дата-время, значение.

Для дальнейшего вырезания столбцов из каждого файла, перевода времени из формата "datatime" в формат "секунды", объединения файлов в единую таблицу и построения графиков в работе используются скрипты *cut.sh*, *date.sh, paste.sh и gnuplot.*

Как было показано выше, сложные операции с файлами совершаются с помощью определенных команд, утилит и небольших скриптов. Это облегчает процесс подготовки данных и позволяет подвести их к следующему этапу анализа в нужном виде. Дальнейшие действия, такие как синхронизация, устранение выбросов и проведение анализа были реализованы на примере языка программирования APL.

# Библиографический список

1. С.Т. Лескин, Д.Г. Зарюгин. Комплекс программ DINA I для диагностирования главных цир куляционных насосов ВВЭР по данным оперативного технологического контроля. Известия вузов. Ядерная энергетика, 2001
2. Fernandez, George. Statistical data mining using SAS applications. CRC Press, Inc., 2010.
3. Граннеман С. Linux. Необходимый код и команды. Москва. 2010.
4. Philipp K.Janert. Gnuplot in action. 2010.