

## ИСПОЛЬЗОВАНИЕ СТАТИСТИЧЕСКИХ АЛГОРИТМОВ БАЗ ДАННЫХ, В ИНСТРУМЕНТАЛЬНОМ КОНТРОЛЕ, ПОСРЕДСТВОМ ЦИФРОВОГО ПРЕОБРАЗОВАНИЯ СИГНАЛА

Галахов В.В., Скачко Ю.В.

Московский институт электроники и математики  
1090288, Москва, Б. Трехсвятительский пер., 3/12, кафедра МиС

Усиливающаяся тенденция широкого распространения использования ПК, как универсального и недорогого средства обработки измерительной информации, остро ставит вопрос о выборе программного инструментария. С одной стороны, можно воспользоваться, например, широко распространенным продуктом MS Excel, экспортировав или внося данные в плоскую таблицу вручную и проведя анализ доступным инструментарием. Однако, данный подход имеет следующие недостатки: отсутствие глубокой автоматизации, невозможность создать зависимые таблицы, высокая вероятность «человеческого фактора», необходимость платить за лицензию (от данного недостатка возможно избавиться используя электронные таблицы Star Office). С другой стороны, по мнению авторов, более гибким инструментом в задачах автоматизированного статистического контроля параметров, полученных в цифровом виде, являются базы данных. Одной из немаловажной оценкой такого похода является стоимость создания и владения. По приблизительным подсчетам, исходя из экспериментальных данных, в среднем, современная Система Управления Базами Данных (СУБД) способна выполнять от 2 до 4 транзакций на каждый мегагерц тактовой частоты одного процессора. Этого вполне достаточно, что бы используя платформы невысокопроизводительных персональных компьютеров обеспечить статистический контроль масштаба малого предприятия. Стоит отметить, что данное оборудование достаточно дешево, надежно и не требует больших вложений денежных средств. Другой немаловажной особенностью является выбор СУБД. По мнению авторов, в целях сокращения издержек задач статистического контроля, в масштабах малого предприятия и не критичности такого контроля, следует остановиться на выборе свободно-распространяемых СУБД и СУБД с открытым кодом. Последние нужны там, где необходима сертификация ПО (недокументированные возможности, несанкционированный доступ), например в области ВПК.

Под задачами статистического контроля будем понимать:

- Регистрацию случайной величины (запись в таблицу БД);
  - Идентификация закона распределения случайной величины;
  - Проверка статистической гипотезы о том, принадлежит ли случайная величина генеральной совокупности закона распределения или является «грубой погрешностью» (например, если измерение или регистрацию проводит оператор);
  - Построение оперативной характеристики и прогнозирование ошибок первого и второго рода (риска поставщика и риска заказчика);
  - Построение графиков, диаграмм, контрольных карт;
  - Оперативный мониторинг (например, состояния технологического процесса);
- Практически все задачи, перечисленные выше, требуют вычисления средних значений и среднеквадратического отклонения.

Были выбраны следующие СУБД:

- Oracle XE (свободно-распространяемая)
  - PostgreSQL (свободно-распространяемая, open source)
- Первая является урезанной версией широко распространенной коммерческой СУБД Oracle. Ограничения касаются величины хранимого пространства и отсутствие поддержки нескольких языков программирования.

При необходимости Oracle XE трансформируется в коммерческую версию. Полная версия имеет широкие возможности и мощный функционал. Обе СУБД имеют версии как под платформу Windows, так и под Unix. Подсчет среднего арифметического значения средствами СУБД не вызывает особых проблем, так как функ-

ция SUM является встроенной во всех базах данных использующих SQL. Соответственно  $\bar{X} = \frac{\sum_{j=1}^n X_j}{n}$ , где

$\bar{X}$  – среднее арифметическое случайной величины X, n – размер выборки.

В ИСЧЕСКИХ СИСТЕМАХ среднее арифметическое будет выглядеть следующим образом:

```
SELECT AVG(a1) FROM tab1;
```

Функция AVG принимает имя столбца и возвращает его среднее арифметическое значение, игнорируя пустые значения.

a1 – название столбца таблицы tab.

Поскольку AVG функция агрегатная, она имеет высокую скорость выполнения.

Алгоритм вычисления среднего квадратического отклонения (СКО) необходимо реализовывать самостоятельно, т.к. он не включен в встроенные агрегатные функции SQL. Для этого используется свой процедурный язык (диалект SQL или его расширение), который может быть отличный в каждой СУБД. Для Oracle XE – это PL/SQL, для PostgreSQL – PL/pgSQL.

СКО определяется формулой: 
$$\sigma = \sqrt{\frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n-1}}$$

Код вычисления среднего квадратического отклонения для PL/SQL будет выглядеть следующим образом:

```
DELETE FROM tab2;
DECLARE
sred NUMBER(10,2);
sko NUMBER(10,2);
CURSOR ff_cur IS
SELECT a1 FROM tab1;
BEGIN
    SELECT AVG(a1) INTO sred FROM tab1;
    FOR ff_rec IN ff_cur
    LOOP
        INSERT INTO tab2(a2) VALUES (POWER(ff_rec.a1-sred,2));
    END LOOP;
SELECT SQRT(SUM(a2)/n-1) INTO sko FROM tab2;
    DBMS_OUTPUT.PUT_LINE (sko);
END;
```

В данном примере округление происходит с двумя знаками после запятой, целая часть числа содержит восемь знаков (число задается десятью разрядами). Обычным шрифтом показаны переменные примера. Таблица данных tab1 содержит единственный столбец a1, tab2 содержит единственный столбец a2. Первый оператор полностью удаляет данные таблицы tab2 при каждом запуске программного кода.

Алгоритм вычисления СКО для PostgreSQL (PL/pgSQL) такой же, несколько различается только синтаксис языка:

```
DELETE FROM tab2;
DECLARE
sred NUMERIC(10,2);
sko NUMERIC(10,2);
ff_rec RECORD;
BEGIN
    SELECT AVG(a1) INTO sred FROM tab1;
    FOR ff_rec IN SELECT a1 FROM tab1
    LOOP
        INSERT INTO tab2(a2) VALUES (POWER(ff_rec.a1-sred,2));
    END LOOP;
SELECT SQRT(SUM(a2)/n-1) INTO sko FROM tab2;
    RETURN sko;
END;
```

Данный алгоритм можно оформить в виде процедуры и обращаться к ней при решении статистических задач. Oracle XE и PostgreSQL поддерживают механизм «хранимых процедур». Это означает, что роль вычисления отдается серверу, а клиенты только посылают запросы и исходные данные. Таким образом ускоряется процесс вычисления, снижается трафик обмена «клиент-сервер», алгоритмы вычисления создаются, хранятся и управляются централизованно.

#### Литература

1. Стоунс Р., Мэтью Н. Основы PostgreSQL. –СПб.: Символ-Плюс, 2002.
2. Цейтлин Я.М., Скачко Ю.В., Капырин В.В. Модифицированные струнные преобразователи для измерения геометрических величин. –М.: изд-во стандартов, 1989.
3. Аллен К. Oracle PL/SQL. –М.: Лори, 2001.
4. Фейерштейн С., Прибыл Б. Oracle PL/SQL для профессионалов 3-е изд. –СПб.: Питер, 2004.

### THE USE OF STATISTICAL ALGORITHMS OF DATABASES, IN THE TOOL CONTROL, BY MEANS OF DSP

Galahov V., Skachko Yu.

Moscow Institute of Electronics and Mathematics

109028, Moscow, B. Tryohsvyatitsky pereulok, 3/12, Metrology and Certification chair

Databases are powerful tool of accumulation and information processing. Their use will allow to leave to the enterprise on high level of automation of the control of the parametres presented in digital form. In the given report algorithms of use of databases for calculation of statistical parametres are resulted: average of arithmetic and average quadratic deviation. They are intended for Oracle XE and PostgreSQL. Both databases is free for develop and deploy. The average arithmetic and average quadratic deviation is used for calculation of many statistical parametres. The given decision cheaply enough, reliably also doesn't demand the big investments of money resources.