

Обобщенная модель построения системы обработки данных

Селезнёв Александр Игоревич, ассистент;

Селезнёв Игорь Львович, кандидат технических наук, доцент

Белорусский государственный университет информатики и радиоэлектроники (г. Минск, Беларусь)

В статье рассматривается построение системы обработки данных. Проведен анализ этапов обработки данных, спроектирована обобщенная модель системы обработки данных.

Ключевые слова: обработка данных, система обработки данных.

В настоящее время для организации эффективной обработки возрастающих массивов данных и проектирования высокопроизводительной системы обработки данных (СОД) с использованием современных технологий необходимо понимать базовые принципы взаимодействия элементов СОД, что предполагает собой построение обобщенной модели системы обработки данных.

Обработка данных — преобразование данных в требуемый вид информации. Как правило, она включает в себя сбор, организацию, очистку, проверку, анализ и преобразование данных в другие форматы, такие, например, как графики или документы. Обработка данных

может осуществляться ручным, механическим и электронным методом.

Целью обработки данных является повышение ценности информации. Это позволяет предприятиям совершенствовать свою деятельность и принимать своевременные стратегические решения. Программные решения для автоматизированной обработки данных играют в этом значительную роль, что особенно актуально для «больших данных» — крупных и сложных наборов данных, объем которых постоянно растет. На рисунке 1 представлена структурная схема основных этапов обработки данных.

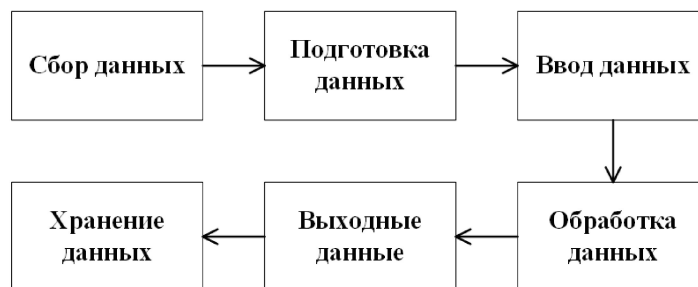


Рис. 1. Структура основных этапов обработки данных

На этапе сбора данные собираются из различных источников, специализированных баз данных, информационных хранилищ. Крайне важно, чтобы источники данных были точными, надежными и хорошо структурированными для гарантии надлежащего качества и функциональности собранных данных.

На этапе подготовки данных собранные данные подготавливаются и очищаются. На этом этапе, также называемом «предварительной обработкой», необработанные данные компонуется и структурируются для облегчения реализации дальнейших этапов. Очистка или подготовка данных включает в себя устранение ошибок, удаление «шума» и неверных данных (неточных или недостоверных данных) для их сортировки по требуемым категориям.

На этапе ввода данных необработанные данные преобразуются в информационную форму. «Чистые» данные вводятся в систему или место назначения (например, хранилище данных, такое как Astera построитель хранилищ данных [1], или система управления взаимоотношениями с клиентами, например Salesforce [2]). Ввод данных осуществляется путем перевода их на язык, понятный системе, вручную или через устройства ввода, настроенные для сбора данных.

Главной процедурой этапа обработки данных является непосредственно обработка данных, например для интерпретации с использованием алгоритмов машинного обучения и искусственного интеллекта. Фактический процесс может отличаться в зависимости от источника данных (хранилища данных, социальные сети, различные сетевые устройства) и его предполагаемого использования (выявление закономерностей и тенденций, определение решений, стратегий и оптимизация).

На этапе выходных данных, также часто называемом этапом интерпретации данных, система обработки данных переводит и представляет их в других форматах, например таких как документы, графики и изображения.

Заключительным этапом обработки данных является этап хранения данных, который включает в себя сохранение обработанных данных для использования в будущем [3–5].

Система обработки данных — это комплекс взаимодействующих методов и средств сбора и обработки электронной информации, необходимых для управления объектами с помощью электронно-вычислительных машин (ЭВМ) и других технических средств. Благодаря применению ЭВМ системы обработки данных представляют

собой не отдельные информационно-вычислительные операции, а именно взаимосвязанный комплекс действий, реализуемых на основе единого технологического процесса. В силу того, что основными функциями таких систем, помимо обработки данных, является также их хранение в специальных базах, такие системы называют системами обработки и хранения данных.

В крупных IT-компаниях системы обработки данных используют мощности центров обработки данных (ЦОД), вынесенных в отдельные структурные подразделения предприятий.

Главным предназначением системы обработки данных является отбор и автоматизация трудоемких и регулярно повторяющихся рутинных операций над большими массивами данных, что позволяет многократно повысить эффективность управления и производительность [6].

Основные задачи системы обработки данных:

- сбор данных в автоматическом или ручном режиме;
- обработка данных (классификация, сортировка, маркировка и архивация);
- передача и обмен данными (например, управление по удаленному доступу);
- контроль доступа к данным;
- визуальное представление и статистическая обработка данных (в виде таблиц, графиков, картинок, текста и т.д.).

ЦОД является наиболее эффективной структурой, использующей системы обработки данных. Благодаря высокотехнологичному подходу к вышеперечисленным функциям, ЦОД позволяют достичь необходимого уровня автоматизации, а также обеспечить сохранность данных и другой информации, имеющей большую ценность. Как правило, ЦОД включает в себя следующие компоненты:

- специальные помещения с отдельной системой кондиционирования и бесперебойного электроснабжения;
- комплекс программно-аппаратных средств (собственно, процессоры обработки баз данных);
- система охраны ЦОД (системы контроля и управления доступом, видеонаблюдения, пожаротушения);
- специально обученный IT-персонал, обслуживающий ЦОД.

На основании вышеизложенного сформирована структурная схема системы обобщенной обработки данных, изображенная на рисунке 2.

Входные данные могут поступать от пользователя напрямую или из отдельных массивов данных. Поступа-



Рис. 2. Обобщенная структура системы обработки данных

ющие данные могут быть различны, поэтому в блоке «Подсистема получения (сбора) данных» происходит их фильтрация и исключение неподходящих данных. «Подсистема управления» предназначена для управления всем процессом обработки в ручном/автоматическом режимах и имеет прямой доступ ко всем другим блокам. В блоке «Подсистема хранения» обеспечивается хранение необходимых данных как для последующей обработки, так и для передачи обработанных данных в блок «Подсистема вывода (передачи) данных», из которого данные передаются конечному потребителю.

Выводы

Проектирование любой системы начинается с построения базовых элементов и описания взаимодействия между ними. В ходе анализа и исследования СОД был предложен один из вариантов проектирования системы путем детализации структуры этапов обработки данных.

Особенностью спроектированной обобщенной системы обработки данных является её гибкость, что позволяет использовать её в качестве базовой структуры в зависимости от заданных параметров и требований к системе.

Литература:

1. Создание хранилища данных: пошаговое руководство — Текст: электронный // Astera: [сайт].— URL: <https://www.astera.com/ru/type/blog/building-data-warehouse/> (дата обращения: 14.07.2024)
2. Salesforce CRM: Everything you need to know // Salesforce: [сайт].— URL: <https://www.salesforce.com/crm/> (дата обращения: 14.07.2024)
3. Селезнёв, А.И. Контейнеризация в системах обработки данных / А.И. Селезнёв, И.Л. Селезнёв.— Текст: непосредственный // Молодой учёный.— 2023.— № 43.— С. 7–11.
4. Селезнёв, А.И. Актуальность применения микросервисной архитектуры в системах обработки данных / А.И. Селезнёв, И.Л. Селезнёв.— Текст: непосредственный // Молодой учёный.— 2023.— № 48.— С. 22–32.
5. Системы обработки и хранения данных // Flylink: [сайт].— URL: <https://www.flylink.ru/info/articles/553/1059> (дата обращения: 14.07.2024)
6. Что такое обработка данных? Определение и этапы // Astera: [сайт].— URL: <https://www.astera.com/ru/knowledge-center/what-is-data-processing-definition-and-stages/> (дата обращения: 14.07.2024)