

О проблеме конвертации типов в современных информационных системах

Кочуров В.А., Касперович В.Л.

Белорусский национальный технический университет

Конвертация типов (*type conversion, casting*) – это преобразования данных из формата, поддерживаемого одним типом данных (форматом-источником), в формат, поддерживаемый целевым типом данных.

Данная задача является источником большого количества ошибок в программных системах, в первую очередь из-за применения программистами неявного (*implicit*) преобразования. Данные ошибки могут быть разделены на 2 класса: вызывающие отказ программы из-за невозможности выполнить преобразование; приводящие к потере точности или искажению результатов вычисления.

Преобразования элементарных типов данных могут быть сужающими (множество значений исходного типа включает множество значений целевого типа, но не наоборот) и расширяющими. Как правило, ошибкам подвержены сужающие преобразования, т.к. формат более общих типов является более вариативным по отношению к более специализированным.

При миграции данных проблема преобразования типов возникает в следующих случаях:

- различные форматы данных в исходном и целевом хранилищах;
- необходимость проверки данных на этапе трансформации ETL;
- наличие ошибок в программном средстве, обрабатывающем данные, а также архитектурных дефектов в исходном хранилище.

Для минимизации ошибок преобразования типов должна быть разработана формальная архитектура приложения и хранилища; все операции конвертации должны быть функционально протестированы.

В процессе формализованной миграции данных, приложение, выполняющее этап трансформации ETL, должно иметь однозначную формальную модель данных; согласно которой должна выполняться конвертация; ошибочные ситуации должны логироваться и обрабатываться вручную.

Литература

1. Davies, J. Formal model-driven engineering of critical information systems – Jim Davies, David Milward, Chen-Wei Wang, James Welch – Science of Computer Programming, 2015 – P. 88-113.