

Агрегатный взрыв (Aggregate explosion)

Многokратное возрастание объема данных, содержащихся в реляционной, многомерной базе или хранилище за счет хранения вместе с детализированными (атомарными) данными данных различного уровня агрегирования.

Например, пусть в хранилище загружается временной ряд продаж за год, детализированный по дням (т.е. 356 значений за год). Но при этом автоматически вычисляются агрегаты по неделям (52 значения), по декадам (36 значений), по месяцам (12 значений), по полугодиям (2 значения).

Следовательно, вместо ожидаемых 356 значений в хранилище оказываются на 102 значения больше. Если в хранилище поступают данные по десяткам тысяч товаров, продаваемых в тысячах точек продаж, то дополнительный объем новых данных, формирующихся в хранилище при агрегировании, может оказаться очень большим и привести к агрегатному взрыву.

Чтобы снизить перегрузку баз и хранилищ данных агрегатами, можно использовать два подхода.

1. Хранить только агрегаты тех уровней, которые точно потребуются для анализа, а не все подряд.
2. Отказаться от хранения части агрегатов, а вычислять их «на лету», непосредственно в процессе выполнения аналитического запроса.

Применение второго подхода обычно приводит к некоторому снижению скорости отклика на запросы, но позволяет уменьшить вероятность агрегатного взрыва.

Многомерные хранилища больше подвержены риску агрегатного взрыва по сравнению с реляционными, поскольку хранят данные менее компактно.