

Агрегирование (Aggregation)

Синонимы: Обобщение, укрупнение, объединение

Процесс преобразования данных с высокой степенью детализации к более обобщенному представлению. Заключается в вычислении так называемых агрегатов — значений, получаемых в результате применения данного преобразования к некоторому набору фактов, связанных с определенным измерением. При этом чаще всего используется простое суммирование, вычисление среднего или медианы, выбор максимального или минимального значений.

Например, факты, отражающие объемы ежедневных продаж, могут быть агрегированы по неделям, декадам, месяцам и т.д. Если в течение недели ежедневные продажи составили 12, 9, 14, 11, 15, 7 и 10, то при использовании суммы мы получим 78, среднего — 11.14, медианы — 11, минимума — 7 и максимума — 15.

Переход к более укрупненному представлению с помощью агрегирования необходим по нескольким причинам.

1. Детализированные данные часто оказываются очень изменчивыми из-за воздействия различных случайных факторов и поэтому слабо отражают общие тенденции и закономерности исследуемого процесса.
2. Масштаб данных может не соответствовать решаемой задаче.

Например, если имеется информация о ежедневном числе клиентов, а на ее основе требуется построить прогноз по неделям, то нужно выполнить соответствующее укрупнение масштаба, чтобы единицей наблюдения стало число посетителей в неделю. И наконец, с помощью агрегирования можно уменьшить число обрабатываемых значений фактов. Например, вместо 365 значений ежедневных продаж за год можно использовать 52 по неделям.

Однако следует помнить, что при агрегировании может быть создано большое количество новых фактов, вычисление которых потребует значительных затрат времени и машинных ресурсов и даже способно привести к агрегатному взрыву хранилища данных.

В Loginot обработчик группировка позволяет объединять записи избранных полей в группы, а для оставшихся полей вычислять агрегаты, а визуализатор куб является одним из распространенных методов многомерного анализа и агрегирования данных. Манипулируя с помощью мыши заголовками измерений куба, пользователь может добиться наиболее информативного представления информации.

