

Масштабируемый алгоритм (Scalable Algorithm)

Алгоритм обработки данных называется масштабируемым, если его вычислительные затраты растут прямо пропорционально увеличению объема обрабатываемых данных. Т.е. при увеличении объема данных, например, в два раза количество вычислительных операций также увеличивается вдвое.

Кроме того, в аналитических технологиях масштабируемость подразумевает следующие свойства алгоритма:

- Однопроходность — алгоритм должен использовать не более одного прохода базы данных, при этом желательно, чтобы проход был завершен досрочно.
- Возможность отсечения по времени — алгоритм должен быть способен выдать наилучшее на данный момент решение в любой момент, даже если процесс вычислений не доводится до естественной остановки.
- Прерываемость и продолжаемость — алгоритм должен предусматривать возможность временной приостановки и продолжения работы. Промежуточные результаты должны быть сохраняемы с возможностью дальнейшего продолжения вычислений, а возможно и с использованием новых данных.
- Ограниченность по объему требуемой памяти — алгоритм должен работать в пределах ограниченного объема памяти компьютера, выделяемого пользователем.
- Однонаправленность — при проходе базы данных последовательность считывания записей производится только в одном направлении. Такой режим работы называется режимом однонаправленного указателя, когда указатель движется только вперед и без возможности обратного перемещения.

Особенно важным является свойство масштабируемости для алгоритмов, используемых при анализе Больших данных и в машинном обучении, где объем обучающих данных может быть очень велик.