Big Data

**Big Data** – это серия подходов, инструментов и методов обработки данных огромного объема и большого многообразий для получения эффективных результатов при постоянном приростке данных. Таким образом Big Data это не какие-то конкретные данные, а способы их обработки.

Принципы работы с Big Data:

1. Горизонтальная масштабируемость

Поскольку данных может быть сколь угодно много, любая система должна иметь возможность при увеличении объема данных просто увеличивать количество железа, быть расширяемой без изменений самой системы.

1. Отказоустойчивость

При работе с большими данными машин в кластере бывает очень много, что означает что, машины будут гарантированно выходить из строя и методы работы должны учитывать подобные сбои и переживать их без каких-либо значимых последствий.

1. Локальность данных

Так передавать большие объемы данных очень сложно и затратно, обработка данных должно быть проходить на той же машине, где они хранятся.

MapReduce

MapReduce – это модель распределенной обработки данных, предложенная компанией Google для обработки больших объёмов данных на компьютерных кластерах.

Ссылки на статьи для продолжения разбора

[**https://habr.com/ru/post/267361/**](https://habr.com/ru/post/267361/)

[**https://habr.com/ru/company/dca/blog/268277/**](https://habr.com/ru/company/dca/blog/268277/)

[**https://habr.com/ru/post/270453/**](https://habr.com/ru/post/270453/)

[**https://habr.com/ru/company/dca/blog/280700/**](https://habr.com/ru/company/dca/blog/280700/)