



Занятие №39

Работа с большими
данными - pandas, numpy,
matplotlib





1. Что такое big-data?

1. Что такое big-data?

Big Data или **большие данные** — это структурированные или неструктурированные массивы данных большого объема. Их обрабатывают при помощи специальных автоматизированных инструментов, чтобы использовать для статистики, анализа, прогнозов и принятия решений.

Сам термин «большие данные» предложил редактор журнала Nature Клиффорд Линч в спецвыпуске 2008 года [1]. Он говорил о взрывном росте объемов информации в мире. К большим данным Линч отнес любые массивы неоднородных данных более 150 Гб в сутки, однако единого критерия до сих пор не существует.

1. Что такое big-data?

До 2011 года анализом больших данных занимались только в рамках научных и статистических исследований. Но к началу 2012-го объемы данных выросли до огромных масштабов, и возникла потребность в их систематизации и практическом применении.

С 2014 на Big Data обратили внимание ведущие мировые вузы, где обучают прикладным инженерным и ИТ-специальностям. Затем к сбору и анализу подключились ИТ-корпорации — такие, как Microsoft, IBM, Oracle, EMC, а затем и Google, Apple, Facebook и Amazon. Сегодня большие данные используют крупные компании во всех отраслях, а также — госорганы.

