

Что такое Dark Data?

Большинство компаний собирают, обрабатывают и хранят большие объемы данных, которые могут помочь им улучшить свои товары и услуги в будущем. Есть новый телефон Samsung? Samsung, скорее всего, будет собирать ваши данные об использовании. У вас есть аккаунт в Facebook? Facebook собирает ваши данные о просмотре, списки друзей и т. д. И это верно практически для всех компаний. После того, как эти данные собраны, тогда Data Analytics входит в картину!

Но есть большая часть собранных данных, которые нельзя проанализировать с помощью обычной аналитики данных. Эти данные известны как Темные Данные, которые обладают огромным количеством неиспользованного потенциала. Несмотря на то, что Dark Data может предоставить огромную информацию для компании, которая может привести к увеличению прибыли и росту бизнеса, в основном она хранится в архивах компании и не подвергается анализу. Это потому, что очень трудно собирать, идентифицировать и точно анализировать Dark Data.

Согласно определению Gartner, **Dark Data** - это информационные активы, которые организация собирает, обрабатывает и хранит в ходе текущей операционной деятельности, но, как правило, не умеет использовать для других целей (для анализа, совершенствования взаимодействия с партнерами и клиентами или непосредственно монетизировать их).

Некоторые общие примеры Dark Data приведены здесь:

- Электронные таблицы
- Вложения электронной почты и файлы .zip , которые игнорируются после просмотра
- Неактивные и старые базы данных
- Данные предыдущего сотрудника
- Лог-файлы
- Аналитические отчеты и данные опросов
- Старые версии документов все еще доступны
- Личные данные, такие как заметки проекта

Все эти примеры Dark Data являются предметами, которые больше не считаются важными. Таким образом, эти Темные Данные не учитываются, когда они могут быть фактически добыты для очень ценной информации.

Каковы различные измерения темных данных?

Темные данные в основном делятся на 2 измерения, которые составляют различные типы темных данных. Итак, теперь давайте посмотрим, что это:

1. Традиционные неструктурированные данные

Почти 80% всех доступных данных в мире — это традиционные неструктурированные данные. И это часть Dark Data. Очевидно, что в мире много Темных Данных !!! Традиционные неструктурированные данные — это в основном данные в текстовой форме, которые не организованы определенным образом. Это может включать в себя все виды данных в организации, такие как электронные письма, офисные документы, сообщения сотрудников и т. Д., Которые не имеют единой структуры. Поэтому анализ этих данных для получения действенного понимания является очень сложной задачей для организаций.

2. Нетрадиционные неструктурированные данные

В то время как традиционные неструктурированные данные в основном представлены в текстовом виде, нетрадиционные неструктурированные данные еще сложнее! Эти данные в основном состоят из приложений реального времени, таких как аудио и видео файлы . Эту форму Dark Data еще сложнее анализировать, поскольку значение данных в реальном времени может со временем меняться. И если эти данные не анализируются своевременно, то они могут даже потерять свою ценность и устареть.

Источник: <http://espressocode.top/does-dark-data-have-any-worth-in-the-big-data-world/>