Article

Shed Light on Dark Data

Dark data is digital information that is not being used. Consulting and market research company Gartner Inc. describes Dark Data as "information assets that an organization collects, processes and stores in the course of its regular business activity, but generally fails to use for other purposes.

Dark data often comprises most organizations' universe of information assets. Thus, organizations often retain dark data for compliance purposes only. Storing and securing data typically incurs more expense (and sometimes greater risk) than value."

Many times, an organization may leave data dark for practical reasons. The data may be dirty and by the time it can be scrubbed, the information may be too old to be useful. In such a scenario, records may contain incomplete or outdated data, be parsed incorrectly or be stored in file formats or on devices that have become obsolete.

Increasingly, the term dark data is being associated with big data and operational data.

Examples include server log files that could provide clues to website visitor behaviour, customer call detail records that incorporate unstructured consumer sentiment data and mobile geolocation data that could reveal traffic patterns that would help with business planning.

Potentially, this type of dark data can be used to drive new revenue sources, eliminate waste and reduce costs. As a result, many organizations that store dark data for regulatory compliance purposes are using Hadoop to identify useful dark bits and map them to possible business uses.

What is Dark Data?

Organizations gather huge volumes of data which, they believe, will help improve their products and services. For example, a company may collect data on how users use its products, internal statistics about software development processes, and website visits. However, a large portion of the collected data are never even analysed. According to IDC, 90% of the unstructured data are never analysed. Such data is known as dark data. Though the categories of dark data may vary across companies, the following categories of unstructured data usually are considered dark data:

- Customer Information
- Log Files
- Previous Employee Information

- Raw Survey Data
- Financial Statements
- Email Correspondences
- Account Information
- Notes or Presentations
- Old Versions of Relevant Documents

Why Dark Data Is Handled The Way It Is?

At the time of data collection, the companies assume that the data is going to provide value. Companies invest a lot on data collection so both monetarily and otherwise, data should be considered important.

Dark data is a subset of big data, but it constitutes the biggest portion of the total volume of big data collected by organizations in a year. Dark data is not usually analysed or processed because of various reasons by companies but that does not lessen its importance in the context of business value.

There are two ways to view the importance of dark data. One view is that unanalysed data contains undiscovered, important insights and represents an opportunity lost. The other view is that unanalysed data, if not handled well, can result in a lot of problems such as legal and security problems.

Here are a few reasons why there is so much of dark data.

Misplaced priorities

Take the example of a bank analysing online applications for credit cards. The credit card marketing team is focused solely on customer details and eligibility, but no attention is paid to the data on how the customer arrived at the application page. The unattended data could have provided valuable insights on the usability of the bank website and the application page. But there is no priority assigned to this aspect.

Disconnect among departments

In large organizations, departments have their own data collection and storage processes which may not be known to other departments. So, data, even if relevant to other departments, lie unused. This is a process issue obviously.

Technology and tool constraints

If data collection is done by separate technologies and tools in the same organization, there may be cases that these technologies and tools do not interact with each other because of technological constraints. This prevents bringing all the data together and creating a cohesive picture. This happens especially for companies that have different IT systems and formats. For example, it may be difficult to integrate audio file contents from call center with click data from websites. Companies that are at the early stages of a data analytics program face these problems.

Inconsistent or low-quality data

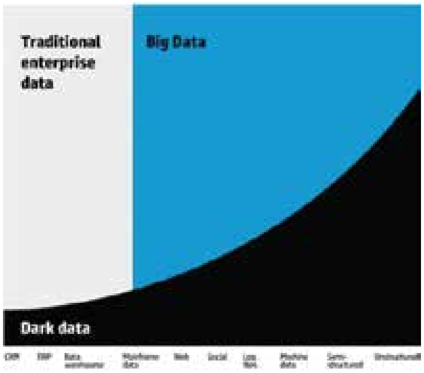
Quality problems prevent you from making use of the data. Data has inconsistencies, missing information, formatting errors and other problems that lower its quality.

GAVS' GAVel For Extracting Business Insights, Perspective & Importance Of Dark Data

There are two ways to view the importance of dark data. GAVS technologies helps organizations to examine their dark data from the different perspectives. Using our flagship solutions like GAVel, which is built on Microsoft's Cortana Intelligent Suite and driven by AzureML and HDInsights, it analyzes deep insights from gamut of IT operations sources and extract predictive insights. These insights are delivered from the intricate algorithms which empower business as well as IT users to innovate new ideas, resonate meaningful decision and resolve constraints at a faster pace and proactively be warned of upcoming issues. The insights provided by GAVel's powerful Social Analytics platform aggregates data from different social sites such as Twitter, Facebook, Glassdoor & LinkedIn and avails instant stats and insights that will aid the management to make key decisions.

Perspective of opportunity not accessed

The area shown in black in the image below indicates dark data. The image illustrates the notional percentage of dark data that is present at any time.



Dark data represents a huge opportunity for companies to gain valuable insights which can drive their business. Take a look at the following examples:

- Server log files can provide website visitor behavior.
- Customer call detail records reveal customer sentiments and feelings.
- Mobile geo-location data can provide traffic patterns.

Companies are missing opportunities by not tapping into dark data. They need better processes, coordination and technologies to appropriately use dark data.

Awareness of problems dark data can cause

Dark data can cause legal, financial and other problems if it is not handled well enough. In fact, companies with piling dark data are already staring into issues. Companies could face the following issues with dark data:

- **Legal and regulatory issues**
  If the data stored is covered by legal regulations such as credit card data, exposure of such data could throw companies into financial and legal liabilities.

- **Intelligence risk**
  Companies could, through deliberate or inadvertent disclosures, lose proprietary or sensitive data on business operations, products, financial status and business plans. This could adversely impact the business.

- **Loss of reputation**
  Customers view their personal information as private and any loss of data, especially sensitive and confidential data by companies can result in a loss of reputation.

- **Opportunity costs**
  If a company decides not to invest in the analysis and processing of dark data but its competitors do, its competitors are more likely to inch ahead in the competition because of the usage of insights from dark data. That is the cost the company is paying because of lost opportunities.

GAVS Technical Expertise For Better Ways To Handle Dark Data

Depending on the perspective, dark data represent an opportunity or a reflection of problems. The ideal way to handle dark data is to utilize it well. But practical difficulties like prior investments that are already made cannot be ignored. Unused data may make some of it redundant over time. Also, it is unlikely that all the dark data will be valuable. So, you should neither discard all of the dark data nor consider all of it valuable.

GAVS helps to bring the best out of dark data through:

- Regularly audit and prune the database. This means that you should be structuring or assigning categories to the old data so that you know what kind of data is stored and where. With storage becoming inexpensive, there is no need to dump data. Since the data is organized and structured, you can find it quickly.

- Apply strong encryption standards on the data. This should be applicable both for data sitting in the in-house servers and the cloud storage. Encryption can prevent a lot of security issues with data.

- Have data retention and safe disposal policies in place. The policies should be aligned with the prescribed regulatory authorities. Carefully formulate policies identifying data for erasure or destruction. Good retention policies will help you retain valuable data for later use.

- Utilizing non-textual data. Most data analytics workflows are built around textual data, which is easier to ingest. However, by using video, audio or other non-textual files that have meta data descriptions, or using speech to text translation you can analyze and gain more insight into the content of the data itself. While it might not be feasible in all cases, it will bring your dark data into focus for better utilization.

Summary

Dark data certainly represents unused opportunities that many companies are letting go of because of process, investment and technology constraints. In a sense, this failure to use dark data also makes big data collection, which is a big exercise, a partial failure. Though the investments needed to tap dark data potential may be costly, the effort is worth the investment. And, even if companies choose to just sit on dark data and do nothing, they are in fact exposing themselves to several risks.

About GAVS

GAVS Technologies (GAVS) is a global IT services & solutions provider enabling digital transformation through automation-led IT infrastructure solutions. Our offerings are powered by Smart Machines, DevOps & Predictive Analytics and aligned to improve user experience by 10X and reduce resource utilization by 40%.

For more information on how GAVS can help solve your
business problems, write to inquiry@gavstech.com
www.gavstech.com

GAVS