06-03T18:42:18.018.-
.orgmanager.handlers
:"5022"
o.chart-data_new.json
"144o2n620jm9trnd3s3n7wg0k
Millis":"0", "level":"INFO"
:"789d89cb-bfa8-4e7d-8047-49845af885d"
llis":"7"}{"timestamp":"2017-06-03T18:46:921.000"
.orgmanager.handlers.RequestHandler", "method"."handle"
:"10190", "message":"Duration Log", "durationMillis":"10"}{
app/rest/json/file", "webParams":"file=chartdata_new.json", "class"
:"7ac6ce95-19e2-4a60-88d7-6ead86e273d1", "sessionID": "144o2n620jm9
llis":"23"}{"timestamp":"2017-06-03T18:42:18.018", "delta:startmillis
.orgmanager.handlers.RequestHandler", "method": "handle" }{"timestamp
:"5022", "message", "webParams":"null", "class": "com.orgmanager.handle
app/page/analyze"}, "level":"INFO", "sessionID": "144o2n620jm9trnd
"8249868e-afd8-46ac-9745-839146a20f09"}{"timestamp": "2017-06-03T18
llis":"36"}{"timestamp": "2017-06-03T18:46 921.000", "class":"sizeChars
:"file=chartdata_new.json", "webURL":"/app/page/report", "level":"INFO
:"144o2n620jm9trnd3s3n7wg0k", "level": "INFO", "sessionID": "144o2n620
Millis":"0", "level": "INFO", "webURL":/app/report_"2017-06
:"789d89cb-bfa8-4e7d-8047-49845af885d"}{"timestamp": Reque
"V" "timestamp":"2017-

Article

Big Data Set to Transform Businesses in 2018

What's filling the Data Lake?

The central promise of Big Data, is that it provides a new way to gain insights into the challenges businesses face every day. In the past, collecting and interpreting vast quantities of data was not feasible because the technologies that automate that process did not yet exist – plus, organizations did not generate much digital data because the computer systems they used were much smaller.

Next generation technologies are transforming the way business is done, compelling organizations to adopt and optimize their IT infrastructure to stay competitive and meet customer demands. IT professionals working with Big Data get a real-world perspective on the opportunities and challenges facing enterprises as we enter 2018. At present, businesses are generating diverse data – from machine data associated with your servers and networks, to emails, to customer purchasing trends. If enterprises want to stay competitive, they need to leverage that data to help move your business forward.

GAVS' services include automation led infrastructure services, enabled by smart machines, DevOps & predictive analytics. Our focus is to reduce incidents through automation to improve user experience by 10X.

Plus, the flagship GAVel solution is a predictive analytics platform that uses intelligent & insightful data aggregation to provide proactive and predictive risk management.

Built on Microsoft's Cortana Intelligent Suite and driven by AzureML, HDInsights & Open source technology, it aggregates information from multiple sources (Big Data) to deliver predictive insights.

The predictive algorithms correlates events and hidden patterns within this data to empower IT users and the business to make insightful decisions to manage operations efficiently.
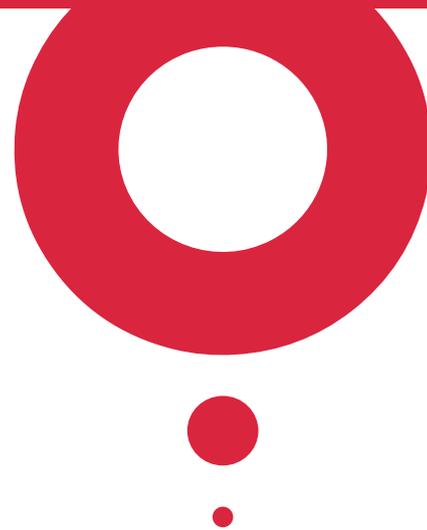
A wide variety of data is filling data lakes, including data from mainframes, relational databases and enterprise data warehouses, along with relatively newer streaming and NoSQL data.

In addition, as more organizations leverage cloud as a deployment platform, it is also gaining importance as a data source.

Collating all this data from across the enterprise is key to accessing the benefits of this technology. Leaving any data source behind means a gap in knowledge, a blind spot in corporate insight. Putting enterprise data together, blending and integrating it, then analyzing that holistic view provides the greatest benefits from Big Data implementations.

Typical sources of big data are:

- RDBMS – 69.3%

- Enterprise Data warehouse – 62.5%

- NoSQL databases – 46.4%

- Data from third party providers/partners – 45.3%

- Cloud repositories – 40.6%

- Mainframes – 31.8%

The rest of the data sources include Web/mobile/social media, IBM/AIX power systems, Machines, IoT sensors etc.

Benefits of Big Data Implementations

An implementation of Big Data technology provides a wide variety of business benefits. Data Lakes are helping businesses cut costs through operational efficiencies, cost effective storage strategies, and avoiding compliance fines.

This technology is also boosting revenue by increasing productivity, extending the capabilities of existing infrastructure, increasing agility, and providing better business insights with advanced analytics.

Another benefit of Big Data projects is getting a better understanding of capacity management value from Data Lake projects.

Top Challenges of Implementing a Data Lake

While Big Data projects help organizations overcome challenges and open up new opportunities in their business, they also throw up their own hurdles. Challenges around data quality and data governance top the list this year, followed by the skills shortage.

The difficulty keeping up with rapid change in technology, as well as keeping the data lake fresh are the other challenges.

The more Big Data workloads become productionized, the need for ensuring data quality and including the data lake in data governance initiatives. Organizations will also use more tools that leverage the skills of the staff they already have to keep the cost of implementation and maintenance down.

- Data quality: Assessing and improving the quality of data as it enters and/or resides in the data lake

- Skills/Staff: Need to learn a new set of skills and tools and Hadoop programmers are difficult to find and/or too expensive

- Data Governance: Including the data lake in data governance initiatives and meeting regulatory compliance mandates

- Rapid change: Compute frameworks and tools continue to evolve and it's difficult to keep up with the latest technologies

- Fresh data: Difficult to keep data lake up-to date with changes made on other platforms

- Mainframe: Difficult to move mainframe data in and out of Hadoop/Spark

- Data movement: Difficult to move data in and out of Hadoop/Spark

- Cost: Hadoop/Spark may be more expensive

- Connectivity: Ability to integrate with existing data sources/applications

- Uncertainty: I don't know what I don't know

Keeping the Data Lake fresh

The only way to keep insights accurate over time is to continually refresh data in the lake with up-to-date source data. Filling the Data Lake isn't a one-time job. Streaming data sources pour in all the time. Transactional sources change with each transaction. That data needs to go in and get integrated with the rest of the data at a reasonable speed, or the Data Lake becomes a stagnant, stale source of old news.

Legacy systems meet Data Lake

Data from the legacy platforms – including the mainframe and IBM I (AS/400) play a significant role in the data lake. Organizations that have relied on these systems for decades understand that leaving this valuable data stranded would severely limit their analytics.

Many of the organizations with mainframes believe its valuable to access and integrate that data in the data lake for real-time analytics.

Leveraging Hadoop and Spark to take the load off traditional systems proves valuable not only to drive bigger insights, but also to fund the new Big Data technology itself.

Source: Syncsort

Data Lake Use Cases

The distributed architectures of Hadoop and Spark are especially adept at data integration and manipulation, whether that takes the form of traditional ETL, offloading data and workloads from legacy systems, filling an active archive, or simply blending disparate data types.

The insights from analytics that drive Big Data projects have the high-tech plumbing of distributed integration powering them.

Some use cases for Data lakes are:

- ETL – 70.8%
- Advanced/Predictive Analytics – 63.5%
- Real Time Analytics – 60.4%
- Data Discovery & Visualization – 53.1%
- Operational Analytics – 45.3%
- Data Blending – 39.5%
- Active Archive – 30.2%

The other use cases include Offload Data and/or Workloads from Legacy Systems – such as EDW or Mainframe, Clickstream analytics, IoT etc.

5 Big Data Trends for 2018

More Enterprise Data Flows into the Data Lake

Demands for data across the organization will continue to rise as business users enjoy the downstream benefits of early Hadoop deployments. Data silos will continue to be broken as data from traditional systems blends with newer sources like sensors.

Data Quality Improves

Organizations need to trust the data they are analyzing in order to trust the resulting insights for decision-making. Expanded use of data lakes will drive an increased emphasis on improving data quality – whether that's when the data is entering, or once it's within the data lake itself.

Data Governance Expands

As regulatory compliance becomes a top concern in the board room, the scope of data governance will expand to include the data that is stored and analyzed in the data lake. Organizations will initiate processes and employ technology that allows them to understand what their data is, where it has been, and who has been touching it.

Data Lakes Will be Fresher

Coinciding with the influx of enterprise data and the emphasis on quality, organizations will implement solutions to ensure their data lakes stay up to date – even if the source of the data is hard to access.

Big Data Will be Stronger than Ever

The names and technologies might change (e.g., Hadoop is giving way to Spark), but initiatives involving the processing of massive data volumes for greater insights are not going away. Cost savings from optimizing expensive traditional platforms like mainframes and EDWs, as well as revenue increases driven by better decision-making, will be funneled back to fund more Big Data projects

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

GΛVS