

Endless Data Piling up Becomes a Serious Cost Burden

Companies: AMZN, DELL, GOOG/GOOGL, IBM, MDB, MSFT, NTAP, ORCL, SPLK, VMW

May 27, 2021

“Heard, tracked, understood, witnessed, confirmed, and you should really think about paying attention to this stuff.”

Research Question:

How does the global explosion of app and mobile phone-driven “everything winds up in a database somewhere” affect the entire IT ecosystem?

Key Findings

“We have crossed the threshold of [data retention] burden. What was once seen as high value, one’s overall collected data, has now become an endless cost sink for the entire information technology universe.” That is how one key Blueshift Tech Trends data science and management source summed up the inability of virtually everyone and every organization in the world to delete data—instead, allowing it to continuously pile up in databases and storage that become less relevant by the hour in an app-driven world of immediacy. Ultimately, the big cloud operators are the only entities that will financially benefit from the crush of data hosted on their infrastructure because they get paid by the organizations hosting and storing it there.

- To a lesser extent, companies like IBM Corp. (IBM) and Oracle Corp. (ORCL) that have customers hosting data on Microsoft Corp. (MSFT) Azure are able to charge for data analytics and database operations off of those deployments, enabling them to claim to customers that they have their own clouds in operation. Sources said Oracle, in particular, is making that work for customers, which has helped the company retain business. IBM, they said, has not been as fortunate.
- From there you move to data capture and targeting companies like Facebook Inc. (FB) and Alphabet Inc.’s (GOOG/GOOGL) Google search and media businesses that self-host, self-network transmit, and self-mine data for financial purposes as beneficiaries of the endless building of databases across the IT ecosystem. For everyone else, data has become an expensive burden that can only be dealt with if prices to simply hold it—or look at it via some type of search capability—become increasingly cheaper to the point of being nearly free.
- Holding data in on-premise legacy networks, sources all agreed, is rapidly becoming a crisis issue on many fronts. For openers, succeeding generations of on-premise IT workers have no idea what data they have, making it almost certain that vast amounts of information will be lost forever. Beyond that, key sources estimated, less than one to three percent of all data created ever gets actively erased, making it all but certain the vast clouds will eventually take over the bulk of it as organizations push their data storage to the cloud.
- Tech vendors still heavily pegged to on-prem business—like MongoDB Inc. (MDB), VMware Inc. (VMW), NetApp Inc. (NTAP), Dell Technologies Inc.’s (DELL) EMC, and other storage divisions—are desperately trying to make themselves relevant in a cloud world where Google, Amazon.com Inc. (AMZN) Web Services (AWS), and Microsoft are building in tools that are far less costly for customers to run than on-prem options, considering the costs of holding onto and analyzing data inside expensive DIY networks and storage arrays.
- Many other small companies and open source initiatives also threaten the norms that any organization must spend to hang onto the mountains of data they create. Use of almost freeware like [Apache Cassandra](#) for non-relational database queries, [Graylog](#) as a free or very low-cost alternative for Splunk Inc.’s (SPLK) policy logging software, and many other options that operate on databases either in-house or in the cloud are upending the entire business structure of data-saturated IT.

Positive: AMZN, GOOG/GOOGL, MSFT, ORCL
Negative: DELL, IBM, MDB, NTAP, SPLK, VMW

Further Information

The major expansion of global undersea fiber optic cables has shifted away from being financed by carriers and onto the major cloud operators, according to sources inside the fiber optic networking community. Google, Facebook, Microsoft, and Amazon are financing cable laying operations that position them at the forefront of next-generation data transmission to connect their mega data center infrastructure in major regions of the world. The concentration of traffic destined for the big clouds is creating

Tech Trends You Need to Know

a stranglehold on the entire data ecosystem that is fed by billions of mobile devices that have become increasingly important to business and organizations since the onset of the pandemic.

Several sources alluded to a type of decision-making shift that is a major threat to all IT vendors still trying to convince customers they should handle the bulk of their IT networking on their own. In the relatively recent past, several factors still fell in favor of DIY IT networking. A lack of wide area connectivity; a lack of trust that an enterprise's data could be safer in the cloud, as opposed to behind a firewall in-house; a lack of application resources in the cloud; and several other issues added up to a robust DIY networking industry led by the likes of Cisco Systems Inc. (CSCO), IBM, and many more. Decisions on spending would remain in-house and be based on the network architecture around switches, routers, and servers and the software licensing needed to run on those architectures. Now, if the chief financial officer and chief technical officer agree to move resources to the cloud, all the spending decisions that used to surround DIY networking are handed off to the people at the cloud companies. That, sources said, is the key decision-making shift in the decline of DIY networking. The cloud companies handle all the architecture and you simply pay for the results. The losers are traditional IT vendors.

In the big DIY days, companies like then-storage leaders EMC (now Dell) and NetApp made billions selling data storage to on-prem customers. Then along came the first generation of AWS, which began selling cheap cloud storage on its overbuilt data center capacity. Mobile computing came along. Bandwidth became cheaper and more plentiful and, sources hastened to point out, the mounting spate of hacking and ransomware is proving that data is actually safer within the confines of Microsoft's clouds, or on AWS, as noted in the [May 14 Tech Trends report](#). With the cost of retaining data under the old DIY sales models becoming unsustainable, we are past the breaking point, sources said.

There are several "in between" business models in the crosshairs, according to sources. MongoDB continues to be cited as extremely vulnerable to the trends. Regardless of the fact that it is in the data analysis business, the additional costs of using MongoDB in clouds where similar tools already exist means the bulk of MongoDB's business still comes from an on-premise mindset among a handful of its large customers. As that changes, the company's hopes to turn to profitability become more challenged. VMware and NetApp were also cited as being caught within the transition zone, where converting to a cloud-focused model makes less sense based on what the two companies have historically had as their core businesses.

As IBM focuses more on a cloud services approach tied back to legacy networks under its heavily promoted "hybrid" data networking campaign, all sources agreed that there is an either/or problem building for the company where customers that want to run in the cloud will turn to the major cloud operators for infrastructure and applications, slowly whittling away what they buy from IBM. As one source put it, "The more hybrid their customers become, the worse it will be for them."

As databases explode everywhere, the bottom line will be how to deal with them as cost-effectively as possible—and that is a business model that does not lend itself to the last-gen way of running data networks.

Background

Blueshift Research's Senior Technology Editor John Harrington tracks the movement and amount of data across global undersea fiber routes, through terrestrial networks, out to the edge to detect changes in critical patterns that will eventually affect how IT operates in the future and what effect that will have on publicly traded companies in the tech networking hardware and software industries. He has tracked how the expansion of fiber networks enabled the rise of the mega clouds for Blueshift Research since 2014. For this report, he interviewed seven long-time repeat executive sources in the data management and data center industries to determine how the proliferation of data is changing the IT sales landscape. Six sources are based in the United States, one is in the UK. Interviews were conducted throughout May.

Key Quotes

- "It's nuts what data people are holding onto. In almost every case we encounter, the simple truth is that they don't even know what they have, except for a fraction of it. Unless it is tied up in the immediate operation of their business, [data] drifts off almost as fast as it is created into a type of limbo. It has to fit into something like an active CRM database, or active workloads, in order to be relevant. Otherwise, it does actually accrue into a very weird realm that may or may not have any focus given to it. This is where these things like MongoDB and Cassandra found this void they can operate in—data that does not fit a class or category. When everyone was doing everything on their own, there was always this niche for the non-relational query world. There are arguments that this is a growing area still but, I can

Tech Trends You Need to Know

assure you, it is only growing if any organization is still 100% committed to doing everything themselves. If you decide to go to the cloud, the set of tasks that accrue to Cassandra, Mongo, or any other no-SQL type of databases are moved behind the development veil in the clouds and they become invisible to the end user to the extent where even complex queries can be run much the same way you'd run a Google search. Cost is the other factor. Hanging onto every bit of data is costly and these costs have to be brought down continually or, after a period of time, a ridiculous amount of money is devoted to just maintaining data, forget the added cost of constantly running queries. The cloud companies are the only players with this limitless economy of scale where they can offer almost free per-gig storage; put tools on top of that for recall, analysis, and security; and blend prices so that you can't afford to not get in deeper with them on your applications and bandwidth usage. If you are in the business of trying to sell some of that as your core proposition, then the sheer volume of data that accrues around the clock will cause customers to question what they are spending unless tapping the data is immediately transactional—like selling you a seat to a ball game or pulling up your medical records at the clinic. What we tell our clients—and they really do not listen—is that, if you're not readily always looking at it, you likely do not need it, in which case, erase it. They never do. Then they complain about how much storing it and protecting it costs. This is why the clouds are building out all of their own infrastructure: because they know they will eventually be holding almost all the information humanity has ever produced—or will.” — CEO of a data management and cloud migration firm on the West Coast

The cloud companies are the only players with this limitless economy of scale where they can offer almost free per-gig storage; put tools on top of that for recall, analysis, and security; and blend prices so that you can't afford to not get in deeper with them... If you are in the business of trying to sell some of that as your core proposition, then the sheer volume of data that accrues around the clock will cause customers to question what they are spending unless tapping the data is immediately transactional.

CEO of a data management and cloud migration firm on the West Coast

- “Every single time you tap an icon on a smart phone, you are interacting with numerous databases. All of the data that are part of your actions are saved. You are tracked, categorized, and targeted. Think of how much data that is. Only an infinitely scalable system can deal with that. If you look at what organizations are now taking over the financing of global undersea cables, you'll find that it is Google, Microsoft, Facebook, and Amazon. The telecoms are not taking that lead anymore. This is only now being recognized by governments, if at all. What it means is that there is a concentration of traffic to the cloud data centers and inside those buildings are the servers that are hosting all of the things that have allowed the world to function during this pandemic. Now, do you really think, as we emerge from these circumstances, things are going to return to some former state?” — CEO of a cloud data management and access firm based in the UK

About the Author

John Harrington is an award-winning investigative reporter and veteran Wall Street researcher. John previously served as senior editor and senior researcher at OTR Global and was a three-time Emmy Award-winning TV journalist.

John brings expertise and relationships in internet networking, network security, fiber optic communications, and data center computing to Blueshift Research. John will contribute regularly, sharing deep insight into tech and communications trends, often before they are recognized by Wall Street.

Report Coverage Areas and Companies

Blueshift Research has been reporting on the following technology areas since Feb. 14, 2014, covering these public companies:

- Cloud Computing/On-Demand Hosted IT (AMZN, BABA, CRM, GOOG/GOOGL, IBM, MSFT, ORCL, WDAY)
- Enterprise IT Networking (ANET, CSCO, CTXS, DELL, FFIV, HPE, IBM, JNPR, MSFT, ORCL, RHT)
- Data Security (CHKP, FEYE, FTNT, INTC, JNPR, MSFT, PANW, SYMC)

Tech Trends You Need to Know

- Data Storage/Management/Analysis (AMZN, BRCD, CSCO, GOOG/GOOGL, HPE, IBM, INTC, MSFT, NTAP, ORCL, PSTG, RHT, TDC, WDC)
- Data Centers and Fiber Optic Networking (AMZN, CONE, DFT, DLR, EQIX, GOOG/GOOGL, IBM, INTC, MSFT, NVDA, QTS, ZAYO)
- Fiber Network Construction and Implementation (ALU, CIEN, CSCO, DY, GLW, IESC, JNPR, NOK)

To access these reports, please contact your Blueshift Research sales representative or [John Harrington](#).

The Author(s) of this research report certify that the information gathered and presented in this report was obtained in accordance with Blueshift Research's compliance protocols as outlined in the company handbook. All Blueshift reporters identified themselves as reporters/researchers from Blueshift Research and articulated the purpose of the research. To the best of our knowledge and efforts, Blueshift confirmed that the underlying source(s) lawfully obtained the information shared with Blueshift and were entitled to provide such information to Blueshift without breaching a duty to another party. The data in this report has undergone review from Blueshift Research's Compliance Officer and has been approved for distribution to Blueshift Research's clients.

© 2021 Blueshift Research LLC. All rights reserved. This transmission was produced for the exclusive use of Blueshift Research LLC, and may not be reproduced or relied upon, in whole or in part, without Blueshift's written consent. The information herein is not intended to be a complete analysis of every material fact in respect to any company or industry discussed. Blueshift Research is a trademark owned by Blueshift Research LLC.