Cloud or On-Premise for Data Analytics?

4 Factors to Consider

LEXALYTICS
Read ● Between The Lines

KEY TAKE-AWAYS

→ **Data analytics companies and enterprises are returning to private cloud or on-premise infrastructure.**

→ **4 factors are driving these companies out of the cloud.**

→ **Do your research and choose a flexible, scalable data analytics provider.**

EXECUTIVE SUMMARY

INTRODUCTION

78% of small businesses will be adapted to the cloud by 2020, according to Intuit. But larger firms are beginning to buck the trend.

Following rapid growth in the cloud, GitLab returned to dedicated hardware in 2016. Meanwhile, Dropbox's 2018 IPO filings reported huge success from their infrastructure move. Even WhatsApp is planning to leave IBM's public cloud in favor of a Facebook data center.

Of course, some companies say they're never going back. But many more enterprises, tech companies, and data analytics firms are returning to private cloud or on-premise infrastructure.

What is driving these reverse-migrations? Our research uncovered 4 factors driving companies out of the cloud: **cost savings**, **data control**, **power**, and **performance**.

Cost Savings
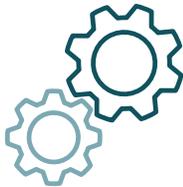
Power

Data Control

Performance

IN THE FOLLOWING WHITE PAPER, WE WILL:

- **Explore each factor in depth, and the implications for you**
- **Feature success stories from companies who have moved back on-premise**
- **Suggest further reading and resources**

The goal of this paper is to aid technical and business executives as they decide whether or not to move their data analytics projects and products out of the cloud.

Bear in mind that we approach this from the perspective of natural language processing and text analytics, since that's our focus. The lessons within, however, are widely applicable to all forms of data analytics.

For this paper, we use these Microsoft definitions of public, private and hybrid clouds, and consider private cloud and on-premise to be roughly synonymous.

And finally, some disclosure: We sell both cloud and on-premise text analytics solutions, so it is immaterial to us which you choose.

KEY TAKE-AWAYS

→ **Data analytics companies and enterprises handling large volumes of data may see big cost savings by moving out of the cloud.**

→ **Calculate the break-even point between cloud and on-premise for your data analytics needs.**

→ **Minimize risk by choosing a solution provider with demonstrated migration success.**

According to our research, the most common incentive for companies to get out of the cloud is money.

For example Moz saved $4.35 million by moving out of the cloud; MemSQL cut their three-year server spend by 80%. And DropBox, the popular file-hosting service, reduced their operational expenses by $74.6 million over 2 years.

Cost factors for data analytics

How did these companies achieve such great success by getting out of the cloud? For most, it came down to cost factors and scale.

First, cloud-based storage providers base their fees on how much data you store with them. Second, data analytics services charge based on volumes processed each month. Third, most natural language processing (NLP) providers add variables for document size (because a tweet and a research paper take very different amounts of processing power), as well as the number of languages you're processing.

Finally, if you're changing data analytics providers, you should factor in the risks involved in the migration. These are the cost factors you must consider whenever you choose a data analytics provider:

▪ **Volume of documents stored and processed**

▪ **Average size of these documents**

▪ **Number of languages/dialects (NLP only)**

▪ **Risks of switching, compared to reduced costs**

Outgrowing the cloud

Cloud computing gained popularity through the mid-2000's based on claims of cost savings and convenience. But as many companies are finding, cloud services become far more expensive than on-premise solutions as they scale up.

As CNBC points out, "Relying on a third-party cloud provider is common for young companies, which have to choose their priorities. But cloud costs can add up as companies grow."

LiquidSky, a cloud service for gamers, is a great example. "We started the company on a public cloud infrastructure," CEO Ian McLoughlin said, "but over time as we grew we needed to move towards a private cloud solution." LiquidSky's transition, he continued, has resulted in "much better economics".
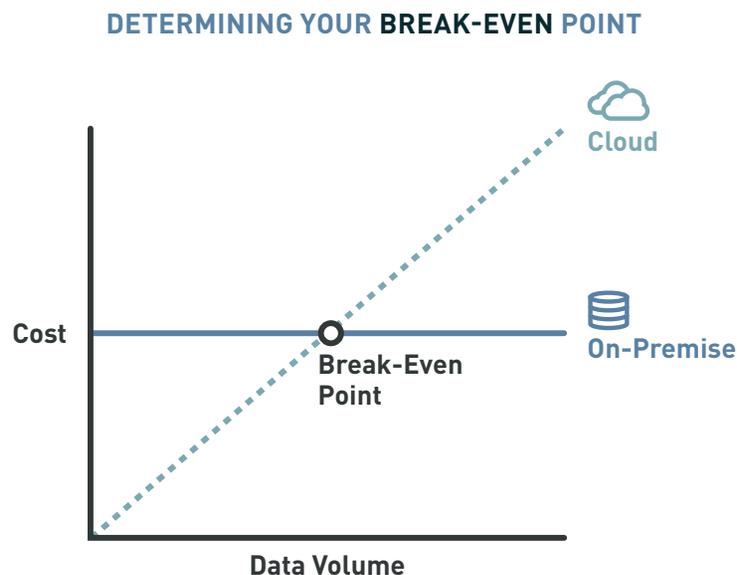
Calculate your break even

How can you tell when your company has "outgrown" the cloud, like LiquidSky or DropBox? The first clue will come when you do a double-take at your cloud services bill.

At this point, we suggest you consult your engineering team and calculate your company's break-even point between cloud and on-premise infrastructure. This calculation should consider:

- **The type of data you're analyzing**
- **Size of each individual datum**
- **Total volume of data**
- **Complexity of analytics**

DETERMINING YOUR BREAK-EVEN POINT

For example, when analyzing large volumes of tweets through the Lexalytics Intelligence Platform, we've calculated the break-even point to come around 10 million documents (tweets) per month. However, the break-even point for larger text documents can be substantially lower.

In some cases, we've found that **processing in the cloud can be 46 times more expensive than on-premise**. With larger text documents and more complex analytics, that number can rise even higher.

Don't be afraid to move

Lastly, we understand that moving out of the cloud may seem risky. But don't stay with a cloud solution just because you're worried about migrating.

All of the companies mentioned above started out in the cloud, and all of them have successfully moved to on-premise. Just think, DropBox has saved $74.6 million in just two years! Even if you're handling just 10% of the data DropBox stores, the potential rewards could outweigh the risks by far.

That said, we'd be remiss to totally play down the risks of moving out of the cloud. Any platform migration comes with some potential for disaster.

As you consider making the switch, consult with an on-premise solution provider who already has experience transitioning clients out of the cloud. Make sure they can alleviate your fears by demonstrating prior success, well before you commit to anything.

KEY TAKE-AWAYS

→ **Any company handling sensitive data should get out of the public cloud.**

→ **Hybrid cloud solutions may offer a "best-of-both-worlds" approach.**

→ **Look for a provider that can provide consistent results across any access method and transition you between them without any inconsistencies.**

Money isn't everything

Or rather, up-front cost isn't everything.

Because if you're doing data analytics for a heavily-regulated industry, losing control of your data can spell disaster. Anyone working with healthcare (HIPAA) or financial (PCI-DSS) data already knows this.

But everyone else should be worried about data control, too. Just look at some recent high-profile data breaches. Hackers have hit industries ranging from entertainment to hospitality to the Pentagon and beyond.

Don't take our word for it: read through this 17-item listicle covering the biggest data breaches of the 21st century.

The security risks of public clouds are well-documented. For example, CISO Magazine's "12 top cloud security threats for 2018" lists no fewer than 13 cloud security concerns that organizations must understand in order to make educated decisions. And last year, Gartner described "a world of unknowns" when reporting on "5 Trends in Cybersecurity for 2017 and 2018".

Consult your security team

We encourage you to consult your information security team or legal counsel to better understand your own data security needs. Ask your team these questions:

▪ **Are we legally required to limit external exposure of our clients' data?**

▪ **Do we have other reasons to be particularly concerned about data security and privacy?**

Bear those questions (and their answers) in mind as you re-evaluate your existing data analytics infrastructure.

Private cloud or hybrid cloud?

Our research into the state of cloud security suggests that any company handling sensitive data should get out of the public cloud. But where should you go instead?

Consider that even now, most cloud infrastructure providers (think Amazon Web Services or Microsoft Azure) still "lease" you access to a public cloud instance. Of course, most analytics providers can arrange private clouds when you need them, but this often adds more cost. If you're not ready leave the cloud entirely, you're left with two options: private cloud or hybrid cloud.

What are private clouds?

As we said, most analytics providers can handle your data within private clouds, roughly synonymous with "on-premise". In these instances, data services and infrastructure are dedicated to your organization. This usually adds to your already-great financial burden, however.
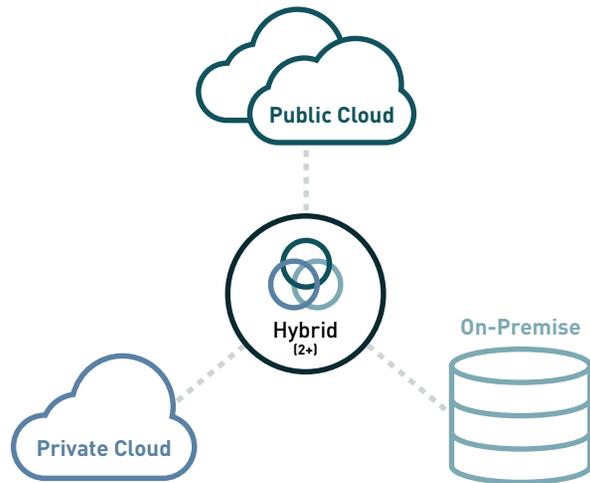
For more about the difference between private cloud and on-premise, refer to this LiquidWeb article.

DATA ANALYTICS INFRASTRUCTURE OPTIONS

What are hybrid clouds?

Your second option, hybrid clouds, combine a mix of on-premise infrastructure and public cloud services. In a hybrid cloud setup, you can use the public cloud for handling lower-security items while keeping sensitive data within your private cloud.

The hybrid approach is rapidly gaining popularity amongst mid-sized and larger analytics companies and enterprises. For example, TapJoy has found great success with their hybrid infrastructure.

Choosing your next cloud provider

Most analytics providers can arrange private clouds when you need them, and some can deliver hybrid cloud solutions. But few companies can offer a coherent data analytics infrastructure spanning public cloud, private cloud and on-premise.

As you choose a new data analytics platform, look for one that can provide consistent results across any access method and transition you between all three states without inconsistencies.

KEY TAKE-AWAYS

→ **Data analytics companies should consider moving their technology out of the cloud to improve competitive differentiation and reduce customer churn by customizing analyses and results for individual end-users.**

→ **Small firms and data analysts will probably do just fine with a cloud solution.**

The great cloud trade-off
Cloud-based data analytics providers promise to make things simpler.

Instead of managing complex technology in-house, you let them handle the details. This is a great option for small- or medium-sized teams who need to integrate solid data analytics with a minimum of upkeep. But there's a trade-off.

To serve the most users with the most convenience as possible, the majority of cloud providers optimize for simplicity. Their goal is to deliver a solution that gets you up-and-running as quickly as possible, with the minimum required analytics features.

Problems down the line
Indeed, choosing a cloud analytics provider may get your minimum-viable product off the ground sooner. But it will inevitably create problems down the line, because choosing the cloud means accepting some level

of "one-size-fits-all". And when your product or project is built on one-size-fits-all technology, what competitive differentiation can you show? Remember: Outsourcing technology means outsourcing ownership of those features.

Just ask Meg Whitman, CEO at Hewlett Packard Enterprises. Speaking at HPE's 2017 Global Partner Summit, Whitman described why many enterprises are deciding the public cloud is no longer the best option for them:

"They want to scale to a hybrid environment that is developer-friendly and gives their business more control and better total cost of ownership."
- Meg Whitman, CEO, HPE

Of course, most cloud analytics solutions offer at least some level of customization. Lexalytics, for one, offers pre-built configurations for specific industries, and our users can tune them even further. But we're the first to admit that the scope of customization in the cloud just can't compete with the deep access offered by on-premise solutions.

On-premise means power
With the right tools and technical know-how, on-premise solutions let you dive deep into the nuts and bolts of the technology. This level of access gives you clear visibility into the decisions being made. And with that visibility, you gain full ownership of the analytics performed, and can fit the results exactly to each end-user.

For example, consider the various features of natural language processing features (since this is our focus). Does one of our clients need to tweak sentiment scoring for individual customers? Easy. Do they want to develop sets of custom categories for their own clients? Simple enough.

All these tasks are (at least relatively) straightforward to accomplish with on-premise text analytics. But with a cloud solution, that level of customization is more difficult to accomplish.

What does power mean for you?
Of course, not every company needs the power of an on-premise solution. Indeed, not every company or team has the technical ability to utilize that level of access. Most small firms and data analysts within larger enterprises will probably do just fine with a cheaper cloud solution.

Deeply technical data analytics companies, however, should seriously consider the potential benefits of on-premise technology. With full control over each analytics feature, your team can customize for individual customers.

Just imagine the resulting level of competitive differentiation, positive press coverage, and happy customers.

KEY TAKE-AWAYS

→ **Some data analytics applications are more latency- and bandwidth sensitive than others.**

→ **On-premise (5-10 milliseconds latency) is always faster than cloud (200 milliseconds latency).**

→ **Consult with your engineering and product teams to determine whether this is a concern for your products.**


High **Bandwidth**
High **Latency**

Low **Bandwidth**
Low **Latency**

What are latency and bandwidth?

Finally, let's touch on performance, as measured by latency and bandwidth.

In data analytics, latency is the total time it takes to send, process, and receive the results for a single datum (such as an individual text document). Bandwidth measures the volume of data you can process within a period (for example, number of tweets per second).

Note that latency (time required to process a single datum) and bandwidth (volume of data you can process per second) aren't synonymous. For example, a truck filled with high-density hard drives is higher-bandwidth than an optical network connecting New York to San Francisco.

However, latency is substantially lower on the optical network, because the truck travels much more slowly.
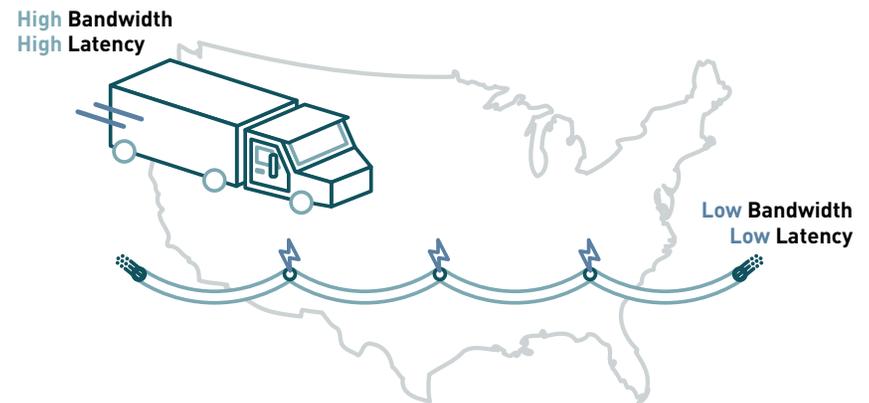
Latency benchmarks

Performance can be a key competitive differentiator for data analytics companies. But even the fastest cloud analytics services have some delay between submission, analysis, and reporting. In the end, no cloud solution can beat the performance of an on-premise setup.

In the Lexalytics Intelligence Platform, cloud latency runs at approximately 200 milliseconds. On-premise latency, on the other hand, is around 5-10 milliseconds. These numbers are comparable with other data analytics services and show the vast latency gap.

Consult your engineering team

There are many more complex technical details involved in these performance metrics. Much like with Factor 3: Power, we suggest you consult your engineering team to determine whether performance should impact your "cloud or on-premise" decision. If you'd like to learn more in the context of natural language processing/text analytics, drop us a line.
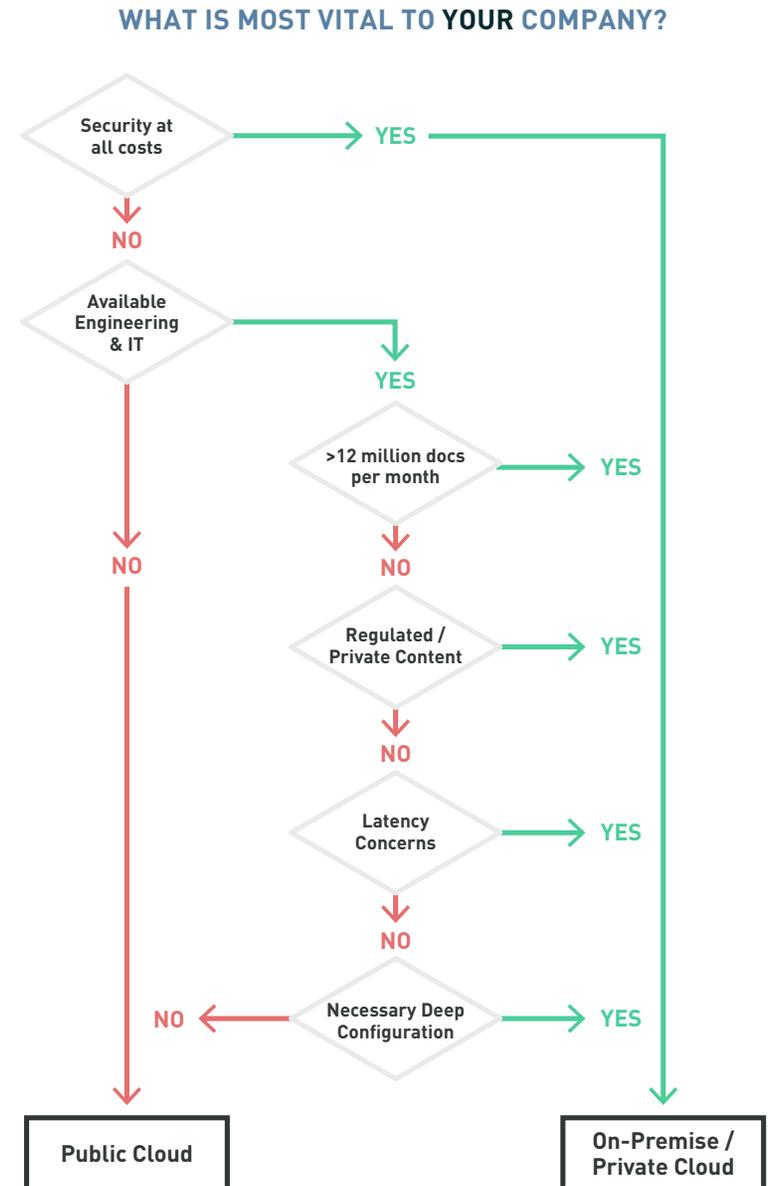
Basic guidelines

Now that you understand the factors driving others out of the cloud, what should your company do? The answer is complicated (of course). But here are some basic guidelines:

- **Small-medium businesses**: You're probably better off staying in the cloud, unless you handle sensitive data, in which case you may need to get out of the public cloud.
- **High-volume data analytics firms**: Consider moving to on-premise or hybrid cloud infrastructure to reduce costs, improve competitive differentiation, and lower customer churn. If you work with text data, consult with our experts to review your options.
- **Data analysts at larger enterprises**: Our recommendation will depend on your industry, budget and technical expertise. Drop us a line with some details so we can help you decide.

Alternatively, here's how it looks in a flowchart:

WHAT IS MOST VITAL TO YOUR COMPANY?

Security at all costs → YES

NO ↓

Available Engineering & IT → YES

NO ↓

>12 million docs per month → YES

NO ↓

Regulated / Private Content → YES

NO ↓

Latency Concerns → YES

NO ↓

NO ← Necessary Deep Configuration → YES

Public Cloud

On-Premise / Private Cloud

In addition to the sources cited in the article above, we recommend these items for follow-up reading:

- Solution Profile: Lexalytics Intelligence Platform
- 4 Tech Companies Decided to Move Out of the Cloud – Data Foundry
- What Are Public, Private, and Hybrid Clouds? - Microsoft
- Why Businesses Are Exiting the Public Cloud – TechNative
- Case Studies in Cloud Decisions – Data Foundry

LEXALYTICS
Read ● Between The Lines

Address	**Sales**
320 Congress St	sales@lexalytics.com
Boston, MA 02210	1-800-377-8036 x1

General Inquiries	**International**
1-800-377-8036	1-617-249-1049