Quantifying the Cost of Forced Localization

Leviathan Security Group

leviathan
security group

limitless innovation. no compromise.

Executive Summary

The modern Internet, in many senses, is composed of "the cloud"—the ethereal-seeming collection of computers and electronics that powers the majority of popular websites and Internet-facing services available around the world. Cloud computing, as defined by the US NIST and the Cloud Security Alliance, has five essential qualities: on-demand self-service provisioning of resources, broad network access, resource pooling, rapid elasticity, and measured service. The latter two mean that businesses can think of and account for computing resources in the same way that they account for electricity. After all, modern businesses do not, with a few edge-case exceptions, invest in power generation facilities and high-voltage transmission lines to build products or deliver services to consumers; instead, a utility provides the power and the infrastructure, and businesses pay only for what they use. Similarly, rather than making the capital investment to buy huge amounts of computer hardware to serve a website, businesses can instead use cloud servers to meet their computing needs—and like power, they can use more cloud servers in times of high demand, and pay for less resources when demand has slowed.

Cloud computing allows businesses to think of and account for computing resources in the same way that they account for electricity.

Cloud computing works because for most purposes, it is not relevant to a consumer where their data is stored, as long as it is close to them in network terms; especially with regard to modern "Internet of Things" devices, if consumers' data is close to hand, it does not much matter whether it is stored in Calgary or Calcutta. Indeed, to protect users' data from large-scale natural disasters, it is often helpful to store data on multiple continents at the same time, so that an infrastructure breakdown in one place will not affect data integrity or availability elsewhere.

Data localization laws, however, threaten this ideal of low-capital-investment, high-availability services. These laws, being considered in response to a variety of political pressures, would force companies to keep data within strict geographic bounds. These laws harm data confidentiality, availability, and integrity, as Leviathan has discussed in previous whitepapers on this subject. While there has been some study as to the macroeconomic harms of data localization laws, no such work has been done on the harms to individual businesses of being forced to use only local cloud

resources. Following our Methodology, we find that for many countries that are considering or have considered forced data localization laws, **local companies would be required to pay 30-60% more for their computing needs** than if they could go outside the country's borders. Moreover, many countries considering data localization have no publicly-available cloud computing providers, meaning that local businesses would be forced to use non-public cloud computing resources, or to purchase and maintain their own infrastructure (with the capital investment that entails.)

Our conclusion is that these costs are both significant and avoidable. Forced data localization undermines the distributed design principle of the Internet, and does not achieve its ostensible goals. We recommend that companies work with their national governments to explain the harms of data localization, and to find alternate paths to protect data security and business growth.

THE COSTS OF FORCED DATA LOCALIZATION ARE BOTH SIGNIFICANT AND AVOIDABLE.

Background

Data localization laws are the general category of laws passed by national legislatures with the goal of keeping data about, produced by, or relating to a country's citizenry within the territory of that country. Such laws have significant impacts on businesses and citizens alike, ranging from concerns about a country's domestic legal process for access to data, to the economic and social implications of removing the choice of where data is stored, to impacts on the confidentiality, integrity, and availability of data based on where it is stored. Several authors have explored the general category of these laws from an academic sociopolitical perspective, including Jonah Force Hill's "The Growth of Data Localization Post-Snowden: Analysis and Recommendations for U.S. Policymakers and Industry Leaders." [1]

Forced data localization laws have significant impacts on businesses and citizens alike.

Beyond political and legal inquiries, Leviathan has explored the security implications of forced data localization in three previous papers: "Comparison of Availability Between Local and Cloud Storage," dealing with security data against natural and artificial disasters; "Analysis of Cloud vs. Local Storage: Capabilities, Opportunities, Challenges," dealing with the scarcity problem of hiring sufficient qualified cybersecurity experts to defend increasing numbers of datacenters around the world; and "Value of Cloud Security: Vulnerability," which is a direct comparison of the cost of setting up a modern data storage infrastructure for small, medium, and large enterprises with the cost of cloud storage for each case.[2]

The question of the direct economic impact of forced data localization laws is also of interest; while data security issues will often have a measurable economic impact on businesses in the long term, direct costs of where data is stored make the abstract question of data localization easier to understand. The European Centre for International Political Economy (ECIPE) studied the macroeconomic costs of forced data localization in their paper "The Costs of Data Localisation: A Friendly Fire on Economic Recovery."[3] They found that economy-wide data localization laws drain between 0.7% and 1.1% of GDP from the economy for, in their estimation, no benefit; "Any gains stemming from data localisation are too small to outweigh losses in terms of welfare and output in the general economy."

The question of microeconomic impact, however—the balance-sheet impact of forced data localization laws on individuals and businesses wishing to use cloud computing resources—has not been adequately studied. This whitepaper, therefore, will explore the question: what effect, if any, do forced data localization laws have on individual businesses at a per-hour, per-server level?

Methodology

The National Institute of Standards and Technology, in a standard[4] later adopted by the Cloud Security Alliance,[5] defines cloud computing as having five essential qualities: on-demand self-service provisioning of resources, broad network access, resource pooling, rapid elasticity, and measured service. We expanded the on-demand self-service requirement to include the ability for the public to create an account and provision computers without requiring a business-wide contract, a non-disclosure agreement, or another hurdle making the provider inaccessible to the general public. Our focus was on "Infrastructure as a Service" (IaaS) cloud computing providers; we therefore excluded storage-only, routing-only, and other providers that, while they may well be general-access cloud providers, do not provide general computing instances. Given these requirements, we found the following cloud providers:

Leviathan examined public, Infrastructure as a Service, cloud providers.

- Amazon Web Services
- DigitalOcean
- Google Compute Engine
- HP Helion Public Cloud
- Linode
- Microsoft Azure
- Rackspace Cloud Servers

Cloud providers do not always offer directly-equivalent services; some focus on customer service and their skilled add-on service components, others focus on sheer numbers of datacenters around the world, and others offer special software or hardware. To obtain as much of an "apples to apples" comparison as possible, we are comparing only general-tier virtualized servers (excluding GPU compute servers, high-I/O servers, etc.), running versions of Linux that do not add extra costs (usually Debian, Ubuntu, or CentOS). As CPU resources between platforms are difficult to compare, we instead equate cloud instances based on memory allocated to each instance, which is directly correlated to CPU resources on each platform. RAM may vary slightly between each provider, so we have used Table 1 to equate instances between providers:

Leviathan's analysis focuses on comparing cloud services with the most-equivalent offerings possible.

Two of the providers, Google and Rackspace, have variable pricing models. For Google, the price per hour of an instance decreases as the lifetime of an instance increases over the month.[6] To provide an effective comparison with other providers, our analysis uses the "typical rate" calculated by Google for the average instance lifetime of their customer base.[7] Rackspace, by policy, requires that all instances come with one of three levels of Rackspace-provided management: "Managed Infrastructure," "Managed Operations: SysOps," or "Managed Operations: DevOps Automation." Because the fees for at least one are mandatory, we have added the (cheapest) Managed Infrastructure per-hour fee to the Rackspace pricing we use for comparison, though we have ignored the per-month per-account minimum (which is low enough that most customers will exceed it with normal use).

For the map shown in Figure 1, an interactive visualization of which is available at `http://www.valueofcloudsecurity.com`, we relied on public statements by the cloud providers to locate their underlying datacenters. Where possible, we correlated such statements with other information (e.g., newspaper articles) to find the most accurate locations to place datacenters on the map.

Table 1: Memory Comparison Between Cloud Providers

	1GB Instance	2GB Instance	4GB Instance
Amazon			m3.medium *3.75* GB
DigitalOcean	1gb *1.0* GB	2gb *2.0* GB	4gb *4.0* GB
Google			n1-standard-1 *3.75* GB
HP	standard.xsmall *1.0* GB	standard.small *2.0* GB	standard.medium *4.0* GB
Linode	linode1024 *1.0* GB	linode2048 *2.0* GB	linode4096 *4.0* GB
Microsoft	a0 *0.75* GB	a1 *1.75* GB	a2 *3.5* GB
Rackspace	general1-1 *1.0* GB	general1-2 *2.0* GB	general1-4 *4.0* GB

	8GB Instance	16GB Instance	32GB Instance
Amazon	m3.large *7.5* GB	m3.xlarge *15.0* GB	m3.2xlarge *30.0* GB
DigitalOcean	8gb *8.0* GB	16gb *16.0* GB	
Google	n1-standard-2 *7.5* GB	n1-standard-4 *15.0* GB	n1-standard-8 *30.0* GB
HP	standard.large *8.0* GB	standard.xlarge *15.0* GB	standard.2xlarge *30.0* GB
Linode	linode8192 *8.0* GB	linode16384 *16.0* GB	linode32768 *32.0* GB
Microsoft	a3 *7.0* GB	a4 *14.0* GB	a6 *28.0* GB
Rackspace	general1-8 *8.0* GB		

Results

Several countries and regions have considered or promulgated laws and regulations designed to force all electronic data relating to a country's citizens to be kept within that country. Some of those laws, like the European Union's Data Protection Directive, have been in place for nearly twenty years; others, like Russia's Federal Law 242-FZ, are much more recent. We will consider the effect of each of these laws on the citizens of the country or region that would be affected.

Brazil

Brazil enacted its "Internet Bill of Rights"[8] on April 24, 2014. As part of the legislative process, a forced data localization requirement that was included as a response to revelations about U.S. intelligence activities, was ultimately removed from the bill. Had it stayed, however, what would have been the effect on cloud computing for Brazilian citizens and companies?

Brazil has two cloud providers: Amazon and Microsoft. At the low end, for 1GB-equivalent servers, Microsoft's price is US$0.024/hour; the lowest worldwide price for 1GB-equivalent servers, $0.015/hour, would save a Brazilian customer 37.5% on their server costs over a Brazil-exclusive solution. For a 2GB-equivalent server, a Brazil-located solution would cost $0.08/hour, and the worldwide cheaper price would be $0.03/hour—a 62.5% savings. Averaged across the types of servers, a customer located in Brazil would pay 54.65% less by using cloud servers outside Brazil, rather than requiring only Brazil-located cloud computing resources.

A customer located in Brazil would pay 54.65% less by using cloud servers outside Brazil.

The European Union

The European Union's Data Protection Directive

The European Union promulgated a data protection directive in 1995,[9] holding that, among other requirements, data on EU citizens may only be transferred outside the EU to countries that provide adequate levels of protection for private data—or, in the case of the United States, if the recipient entity agrees to certify that they will comply with certain protections under the Safe Harbor framework.[10] In response to revelations on U.S. intelligence activities, however, as well as the ongoing *In the Matter of a Warrant to Search a Certain E-mail Account Controlled and Maintained by Microsoft Corporation*[11] dispute regarding whether the United States may force a transnational corporation to turn over data it holds on EU citizens inside the EU to the US government without following the Mutual Legal Assistance Treaty (MLAT) procedures, some parties have called for tightening the Data Protection Directive to eliminate extra-EU data entirely.

Most of the lowest-cost datacenters within the European Union lie within the Schengen Area, pricing for which is discussed in the next section. Outside the Schengen Area and within the EU, however (that is, in the United Kingdom and Ireland), pricing is significantly higher. Rackspace, Linode, and DigitalOcean have datacenters in the UK, and Amazon and Microsoft have datacenters in Ireland; while for 1GB and 2GB servers, DigitalOcean's pricing (which Linode matches) is the same as its pricing in Amsterdam (which is the lowest world price for that type of instance), at 4GB and above, the available datacenters are consistently at least 56.5% more expensive than their counterparts elsewhere in the world. Businesses that move their servers outside this region could thus save more than 36% on their server costs.

Businesses that move their servers outside the European Union could save more than 36% on their server costs.

Schengen Routing

In response to the same concerns noted in The European Union's Data Protection Directive, some parties within the European Union have called for data localization within the Schengen Area. This would be

nearly identical to EU-only localization, with the exception that Ireland and the United Kingdom, while part of the European Union, are not part of the Schengen area. (Norway, Iceland, Switzerland, and Liechtenstein are part of the Schengen Area, but are not part of the European Union; this difference, however, is not relevant to this analysis at this time, as there are no datacenters that meet the criteria developed in Methodology in any of these countries.)

Within the Schengen Area, three countries host datacenters for the cloud providers: Belgium hosts a Google datacenter, The Netherlands host DigitalOcean and Microsoft datacenters, and Germany hosts an Amazon datacenter. Price competition within mainland Europe is fierce, and for 1GB and 2GB servers, DigitalOcean's Amsterdam-based datacenters have the cheapest worldwide per-instance pricing of $0.015/hour and $0.030/hour, respectively. At 4GB and above, however, the Google Europe-West1 datacenter's prices are the cheapest in the Area, and are consistently 10.5% more expensive than the lowest-cost alternatives worldwide.

At 4GB and above, the cheapest servers in the Schengen Area are consistently 10.5% more expensive than the lowest-cost alternatives worldwide.

Canada, India, Indonesia, and Russia

Each of these countries has considered a forced data localization law of one sort or another. Canada has two provinces, British Columbia and Nova Scotia, that have strong localization rules that pertain to public bodies; similar rules have been applied to federal bodies. [12]

None of these countries have datacenters from the cloud providers inside their territory.

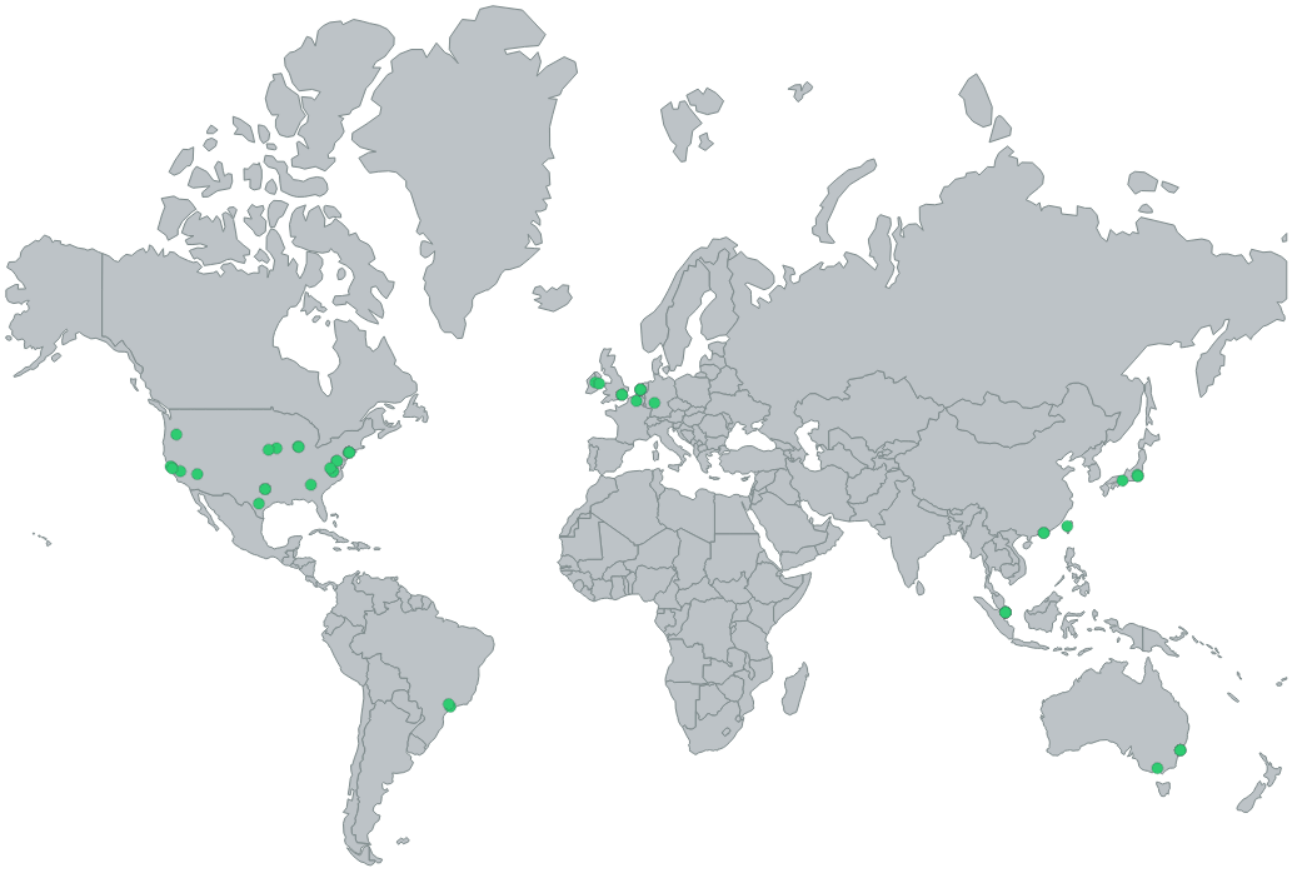
The Indian National Security Council has proposed requiring all email providers to keep all data stored within India,[13] though it should be noted that this is only a proposal, rather than a law or a bill. Indonesia intends to require all electronic systems providers to keep their datacenters—including backups—within Indonesia within the next few years, as part of a 2012 regulation pertaining to a 2008 law.[14] Russia passed a forced data localization law on July 21, 2014.[15] As Russia is currently under significant sanctions related to its behavior in Ukraine, many Western businesses have already diminished their participation in the Russian economy; the effects of this law thus remain fully to be seen.

These countries share one issue: none of the identified public cloud providers have any datacenters inside their territory, meaning that local citizens and companies intending to comply with hypothetical or actual data localization laws must either use traditional datacenters, with the significant capital investment in hardware and periodic upgrades that implies, or non-public cloud providers that require exclusivity, business-wide licensing, non-disclosure agreements, or any of a host of other conditions. A forced data localization law, then, would force companies doing business in these countries to choose among a set of poor choices to protect their data and their livelihood, even as the lack of geographic dispersion in backups makes it difficult to preserve business-critical data in the event of a large-scale disaster.

Conclusion

A look at the map in Figure 1 shows a startling reality for world cloud datacenters: they are distributed in a relatively small number of countries. To wit, the seven cloud providers we examined for this studies have their datacenters in just twelve countries. This result illustrates the significant harm that most countries considering forced localization laws would face; as discussed in Canada, India, Indonesia, and Russia above, many such countries would cut themselves off entirely from the advantages of cloud computing.

Figure 1: A map of world cloud datacenters.



Beyond this, however, and as discussed in Leviathan's previous work on this topic, forced localization laws incur significant harms to the

security and integrity of business data even when countries build out their own cloud-like infrastructures, due to the problems of scarcity of cybersecurity talent, and the simple physical reality that as natural and other disasters grow in size, and thus affect larger chunks of the planet at once, worldwide data replication is necessary in order to preserve valuable data.

A significant design principle of the Internet was, and remains, that the Internet should be able to route around damage in order to ensure that communications between people should never be entirely stopped. It would be a painful irony to allow politics to curtail the resilience of the Internet in pursuit of short-term goals.

Notes

1. http://www.lawfareblog.com/wp-content/uploads/2014/07/Lawfare-Research-Paper-Series-Vol2No3.pdf

2. http://www.valueofcloudsecurity.com

3. http://www.ecipe.org/publications/dataloc/

4. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

5. https://cloudsecurityalliance.org/guidance/csaguide.v2.1.pdf

6. https://cloud.google.com/compute/pricing#sustained_use

7. https://cloud.google.com/compute/pricing#machine-note3

8. "Marco Civil Da Internet," Law No. 12.965, http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm, English version available as a side-by-side translation at https://www.publicknowledge.org/documents/marco-civil-english-version

9. "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

10. http://www.export.gov/safeharbor/

11. *In the Matter of a Warrant to Search a Certain E-mail Account Controlled and Maintained by Microsoft Corporation*, No. 14-2985-CV, (2d Cir.) (ongoing)

12. While one could not consider it a fully unbiased source, the United States Trade Representative has summarized these issues on Page 54 of its 2014 report, available at https://ustr.gov/sites/default/files/2014%20NTE%20Report%20on%20FTB.pdf.

13. http://www.thehindubusinessline.com/features/smartbuy/national-security-council-proposes-3pronged-plan-to-protect-internet-users/article5685794.ece

14. Article 17, Paragraph 2 of Regulation 82 of 2012. Original available at http://www.kemhan.go.id/kemhan/files/74053fdcbba92a4f141234635fe570f0.pdf. English translation available at http://www.flevin.com/id/lgso/translations/JICA%20Mirror/english/4902_PP_82_2012_e.html.

15. "The Federal Law of 21.07.2014 242-FZ 'On Amendments to Certain Legislative Acts of the Russian Federation with regard to the clarification of the processing of personal data in information and telecommunications networks,' " available at http://pravo.gov.ru:8080/page.aspx?112453

Funding for this paper was provided by Google.
All opinions contained in this study reflect the independent views
and analysis of the author(s) alone.

Leviathan Security Group, Inc.
3220 1st Ave S, Suite 100
Seattle, WA 98134

p: 866.452.6997
f: 206.225.2004
e: contact@leviathansecurity.com

www.leviathansecurity.com