Legally Avoid the Data Tax

White Paper

Table of contents

Breaking News (CNN exclusive) – By a narrow margin (52-48), the Palmer administration has just won a highly contested Senate vote that establishes the new "Data Tax". Governments across the world are scrambling to adopt similar legislation. In a time of fiscal conservatism, this new data tax is a severe blow to businesses that struggle to leverage their data assets...

Fiction or reality? Thankfully, no such tax is being levied by governments (what, you don't think you're paying enough taxes already?). However, worse than an "official" tax, if you are using traditional, proprietary data integration tools, your providers are looming, imposing a hefty tax that can severely impact your business.

This White Paper explains how this data tax is being levied, and provides valuable advice from our data tax advisors on how to legally circumvent this tax.

The data tax is a tax on volume

We all know that data volumes are growing exponentially. Data volumes that a typical organization stores and manages double every 18 to 36 months! An IDC study, commissioned by EMC, reveals that between now and 2020, the total amount of information stored will grow by a factor of 44. (Source: The Digital Universe Decade)

Efficient data governance and management requires this data to be processed, cleansed and reconciled between systems. Typical examples of data processing include:

- Loading a data warehouse from multiple production systems
- Synchronizing several applications in real-time to ensure data consistency across the information system

- Replicating data for performance reasons, or to maintain a disaster recovery site

All this processing is heavily impacted by the volumes of data that are being processed. Given a set amount of processing power, increasing data volumes will increase the amount of time it takes to process the data.

In many cases, the time window used for data processing cannot be increased due to business constraints. In this case, the only other option available is to increase the processing power available, typically by adding resources – servers, CPUs, etc. This is the driving force of the first postulate of the Data Tax:

Data Tax Postulate #1

Any organization shall pay a data tax that is based on the number of CPUs (or equivalent) used to process their data.

How does this apply to the facts? Look at the price list of your preferred proprietary data integration vendor (assuming you can get your hands on such a price list…). The runtime price of their tools is based on the number of CPUs, cores, channels, virtual servers, etc. - everyone uses a different terminology but the net result is the same: the more data you need to process, the more additional licenses you need to purchase.

The data tax is a tax on diversity

Information systems become more and more complex as they are consolidated subsequent to M&As, or as new best of breed systems get deployed on-premise, in the Cloud, etc.

This diversity is important, and good business sense would require this to be taken into account in all data integration projects. If you are trying to get a holistic view of the information system for reporting or analytics, but are missing several applications, your view is not holistic.

This is the driving force of the second postulate of the Data Tax:

Data Tax Postulate #2

Any organization shall pay a data tax that is based on the number and type of source and target systems that they wish to connect.

How does this apply? Again, look at your proprietary vendor's price list and find the area where they price their connectors: SAP, Oracle, Salesforce.com, DB2, mainframes… there is a price tag for each of these. You want to transfer your SAP data into the data warehouse? This is how much it will cost you. You want to replicate your DB2/400 data into Oracle? This will be such and such amount.

The data tax is unpredictable

The data tax is especially insidious because it is very difficult to estimate up front. This is the driving force of the third postulate of the Data Tax:

Data Tax Postulate #3

The data tax shall be impossible to estimate upfront and shall only be levied once the organization is locked into its choice.

As a result, the data tax is deeply rooted into the pricing model of proprietary vendors. By "giving away" development seats with "low" initial license fees, they compel organizations to use their solutions. Only when most of the developments have been performed, and are ready to be deployed in production, will the client be able to evaluate their runtime needs. And this is when the tax man shows up…

Examples of data tax victims

The following organizations have recently fallen victim to the data tax.

- Retail company "A" deployed ETL tools from vendor "I" two years ago to load a data warehouse from the data collected at the point of sales. As business grew over time, the data warehouse loading time has increased, and now takes too long. As the holiday season approaches, company "A" has no choice but to pay the data tax to be able to obtain additional CPU power to run its ETL processes.

- Manufacturing company "B" uses data integration tools from vendor "I" to synchronize data every night between its ERP and accounting systems. To increase sales efficiency, company "B" deploys a new CRM in the Cloud and needs to integrate its new CRM with the ERP. This company has to pay the data tax to obtain the CRM connectors from vendor "I".

- Retail bank "C" has been running an ETL from vendor "S" to load its data warehouse. As its small business loan division is growing, branch managers start to request that data about this business be brought into the data warehouse, and that a data mart be created. The ETL servers of bank "C" are already at full capacity, and new servers will need to be deployed to accommodate the extra workload. Bank "C" will have to pay the data tax to vendor "S" in order to deploy the ETL engine on these new servers.

- Technology company “D” uses tools from vendor “A” to replicate data every night to a remote site. Because replication jobs run during the downtime of other systems, they can use the same servers as those running the ELT processes. Accelerating business constraints now require that this replication occur in near-real-time, and company “D” needs to pay the data tax to vendor “A” to be able to deploy dedicated servers to run replication.

An unfair tax

Every tax system has its supporters and opponents, and some would argue that all taxes are unfair. Most taxes, however, have been designed for the common good: to run public services, to deploy transportation infrastructure, etc.

The data tax is very different:

- It benefits a single vendor and does not contribute to the common good of the community.
- It prevents pervasive access to data. Because connecting more data will cost more, IT has no incentive to extend data integration beyond the initial requirements.
- It more heavily impacts short term projects, such as data migration. Regardless of the shelf life of your project, you need to purchase the same perpetual licenses.

It locks capital that could otherwise be used for new projects, such as Cloud Computing.

Avoiding the data tax

The solution to avoiding the data tax is actually surprisingly simple: relocate your data integration to a data tax haven!

If you are not yet using traditional data integration technology, it's easy. Select the right solution using the cheat sheet below.

However, if you are already using traditional data integration technology, the move to a data tax haven is more complicated – but by no means impossible. Here is a step-by-step guide:

- Select new data integration technology, one that does not charge the data tax.
- Deploy the new tax-free technology alongside your existing processes, focusing on the most heavily taxed processes. For example, you may consider using it for:
  - Pre-process data that requires complex transformations to reduce CPU strain.
  - Connecting to new data sources, for which you would need to buy additional connectors under the data tax, and get the data into staging areas for which the data tax has already been paid.
- Over time, when your existing processes require maintenance, consider migrating these processes to the tax-free technology.

Benefits and savings will be immediate, and the more processes you deploy using the tax-free technology, the faster you will decrease your data tax bill.

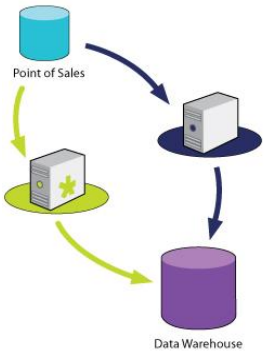
And, of course, it's perfectly legal.

Concrete solutions for the data tax victims

Now, let's review the cases of the organizations mentioned above, and make recommendations on how they can lower their data tax bill.

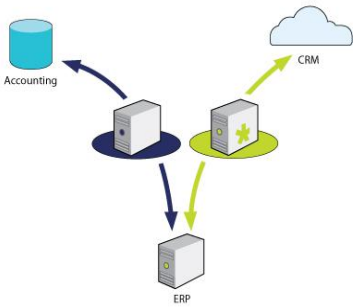
- Retail company "A" deployed ETL tools from vendor "I" two years ago to load a data warehouse from the data collected at the point of sales. As business grew over time, the data warehouse loading time has increased, and now takes too long. As the holiday season

approaches, company "A" has no choice but to pay the data tax to be able to obtain additional CPU power to run its ETL processes.



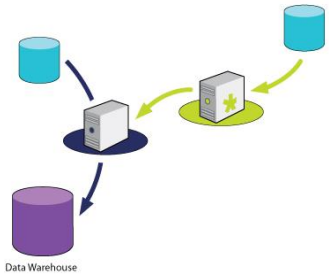
➥ *The quickest save for company "A" is to offload the aggregation of facts data and the refresh of dimensions to a tax-free data integration solution. The current ETL tool would continue to perform the extracts, but all the CPU-intensive heavy lifting would be transferred to the tax-free tool. Once the holiday season is over, Company "A" will be able to start migrating some of the extracts to the new tool.*

- Manufacturing company "B" uses data integration tools from vendor "I" to synchronize data every night between its ERP and accounting systems. To increase sales efficiency, company "B" deploys a new CRM in the Cloud, and needs to integrate its new CRM with the ERP. This company has to pay the data tax to obtain the CRM connectors from vendor "I".



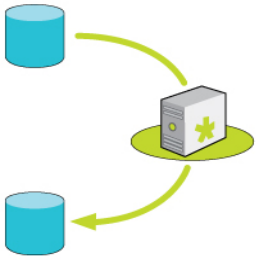
➥ *Instead of buying new connectors from vendor "I", Company "B" should deploy a tax-free data integration tool that has native connectivity to their new CRM, and use it to move data to a staging area, which the traditional tool would then access. Down the road, Company "B" will be able to decrease its data tax even further by using the tax-free tool to connect to all systems.*

- Retail bank "C" has been running an ETL from vendor "S" to load its data warehouse. As its small business loan division is growing, branch managers start to request that data about this business be brought into the data warehouse, and that a data mart be created. The ETL servers of bank "C" are already at full capacity, and new servers will need to be deployed to accommodate the extra workload. Bank "C" will have to pay the data tax to vendor "S" in order to deploy the ETL engine on these new servers.



➥ *Bank "C" should select a tax-free data integration solution to connect to its small business loan data sources, and run it in parallel to the existing ETL. Data can either be loaded directly in the data warehouse, or into a staging area from which the existing ETL tool will load the data warehouse.*

- Technology company "D" uses tools from vendor "A" to replicate data every night to a remote site. Because replication jobs run during the downtime of other systems, they can use the same servers as those running the ETL processes. Accelerating business constraints now require that this replication occur in near-real-time, and company "D" needs to pay the data tax to vendor "A" to be able to deploy dedicated servers to run replication.



➥ *Company "D" should rewrite its replication processes using tax-free technology and stop using tools from vendor "A" (which were not designed for replication anyway). Once they have acquired expertise using the tax-free tools, they will then also be able to tackle the migration of their ETL processes.*

Conclusion

Let's not forget that the data tax has been designed to pour money into the coffers of its initiators! And they have done a good job at maximizing their profits. However, fiscally conservative IT organizations should try to minimize their expenses and leverage their IT budget.

Understanding the data tax and where potential savings reside is a good place to start.

Data tax friendliness calculator

The calculator below will help you estimate the data tax friendliness of your data integration solution. It is simple to use:

- For each category, select the relevant choice. If multiple choices apply, only select the option with the highest score.
- Write the score in the Results column.
- Add up the results for all categories in the Total box at the bottom.
- Look for the total for your tool in the classification under the calculator.

Category	Choice	Score	Results
Pricing model includes fees based on CPUs, servers, etc.	Core-based	+5	.....
	CPU-based	+4	
	Server-based	+3	
	Not applied	0	
Pricing model includes fees based on data volumes or data streams	Applied	+6	.....
	Not applied	0	
Connectors	Priced individually	+5	.....
	Priced in bundles or “packs”	3	
	Included in base license	0	
License type	Perpetual	+5	.....
	Subscription	0	
Total			**.....**

Classification

If the Total is…	Your tool's tax friendliness is…
>10	Nonexistent – you are paying a very high amount of data tax. It is critical that you look for alternatives.
6-10	Low – some users may be worse off than you, but it is critical that you look for alternatives.
1-5	Medium – you should carefully examine if there are places where you can reduce your data tax.
0	High – you already have a tax-optimized solution in place.