Whitepaper

Data environments get cloudy as big data reigns

Using the cloud, companies add new capabilities in days to support big data programs, augment data warehouses and quickly deploy new projects. Read about five emerging deployment styles for leveraging the cloud in your existing data management environment.

Summary

Big data usually starts out messy. You have some idea what you might do with it, but usually you don't know quite how you'll do it. Big data usually doesn't fit your existing infrastructure.

Your first concern might be where to put it all while you tease out the value. Whether you're a line-of-business pro or a data analyst or an IT tech, you need room to work with it all, comb it out, try out new strategies to get inside it. It might not even sit still for you, such as when data keeps streaming in.

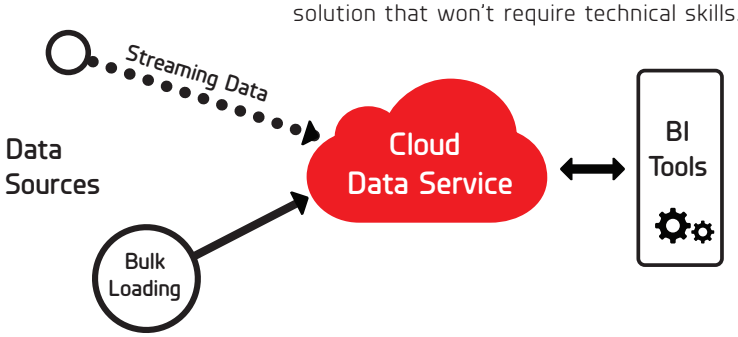
The more you think about it, the messier you realize your big data really is. It's too fast and too raw for the data warehouse, yet you know it will yield the most value when it relates to the older data stored there. Will the big data ever make it into the data warehouse, or will it live somewhere else? Or will it go away once you've tried out a few hypotheses? All the time, the question lingers: Will this big data open doors of insight or die on the doorstep? What do you do?

Many people look to the cloud. There's a cloud for every purpose, budget, and level of skill. All of them come up to speed fast, give you lots of elbow room, keep a low profile, and still let you use the analysis tools you're used to.

Big Data Cloud Solution

If you're in line-of-business, you have questions that big data might answer. But you can't wait for someone to set up new systems. First you try doing it yourself.

For you, the ideal solution offers lightweight data acquisition and processing in the cloud. You're also looking for a managed solution that won't require technical skills.

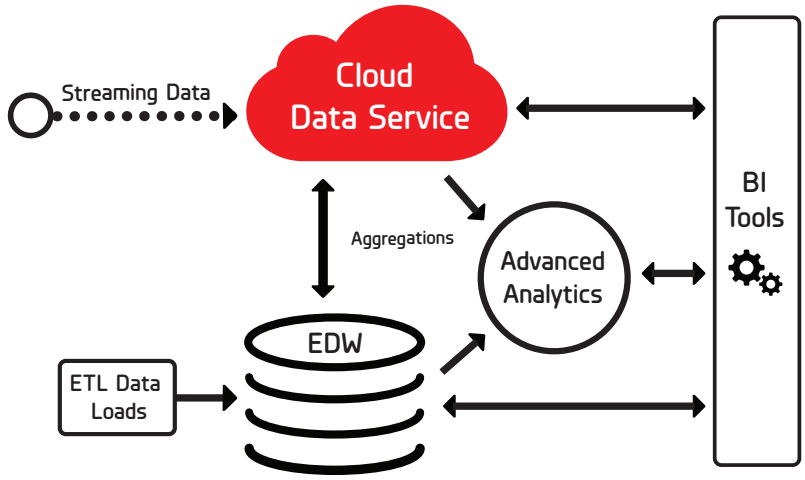


If you're in online-gaming, for example, you have a game that that's exploding with new signups. The past few days have brought a flood, but you're not sure why. The data gushes in: multi-structured event data – coming in fast as players move through the game, make purchases, and progress through levels. But who are these new players and why have they come here? Are they the type of player we want? If they are, how should marketing respond? What makes them play? What makes them stop?

On the cloud, particularly a managed service, the underlying big-data technology never distracts the line-of-business person from the real job: analyzing data for value while it's hot.

Complementary Processing Environment

Complementary Processing Environment
You've got tons of new data, and you've got tons of old. Big data's detail doesn't fit into the data warehouse – yet the data in the data warehouse gives your big data the greatest value when the two can be jointly analyzed.

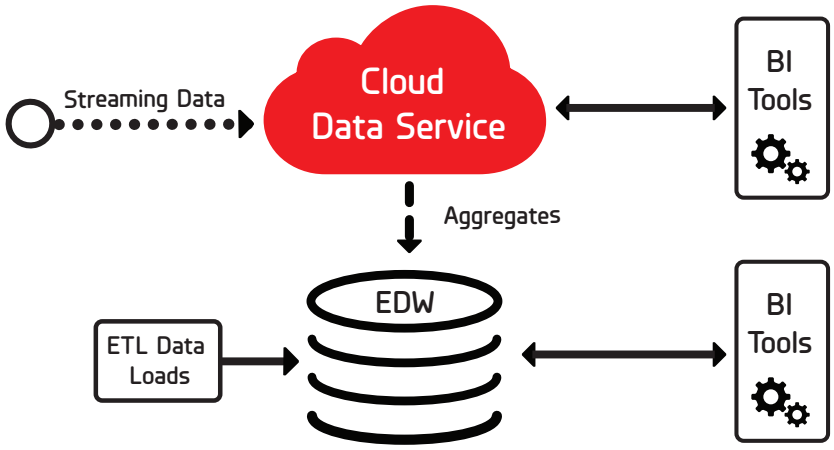


If you are a specialist managing a telecom infrastructure, for example, you have a heavy stream coming from in from hubs, routers, super-routers, sensors, and other machine sources. But the stream is too fast, and storing it in the data warehouse is too expensive. All of your customer data from CRM and marketing applications is stored in the data warehouse – and that's the data that gives your big data meaning. With that, you could segment customers better and know better how network issues effect specific customer types.

You need a platform to process it all. You turn to a cloud service that absorbs the stream as it comes, exports aggregated data to the data warehouse, joins it with business application data, and serves it all to an advanced analytics database and to BI tools. Now you have big data capabilities supporting your existing data warehouse – and at a fraction of the time and cost of building on-premises big data infrastructure.

Staged Integration of Big Data – Fast

You may have seen the future of your big data project in Hadoop, but you know you won't get there in one step. You take it in stages, starting with the cloud. That way, you can show value up front.



A cloud service lets discovery begin within days. Within weeks, the big data's value follows. All the time, you keep costs low, a fraction of your enterprise data warehouse. You can try out ideas and flip through hypotheses – failing fast, over and over, as your ideas keep maturing and improving. Finally, you've got a winner.

Say you're a big data pro, and a marketing exec has you testing clickstream data. No on-premises server can hold it, so you go to the cloud. You try combining it with CRM data, but you get nothing. You try other combinations, but still nothing. Then you try it with marketing automation data, where you see marketing's effects on web traffic, the common paths through the site, and the common exit points.

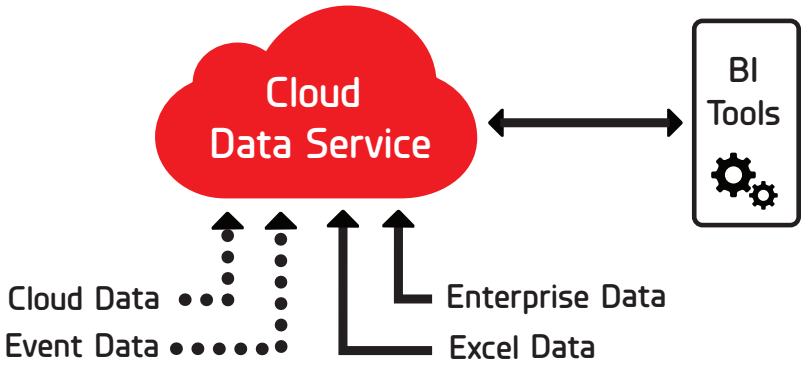
The ending gets even better: You get to choose whether to maintain the cloud service, build your own on-premises system, or keep some of both.

Cloud-based Sandbox Solution

If you're a data analyst, you may have big, disparate data in CSV or Excel files. But you've got nowhere to work with it – certainly no easy method to store, combine and analyze such a large mass. You don't care what happens to it later, but for now you itch to try out your hypotheses. Your boss, his boss, and lots of others around the organization are itching for those results, too.

You need a simple, do-it-yourself environment to set up, load the data, integrate the sources, and run the analysis without help or delay. For example, you might have data from CRM, marketing automation, web analytics, and marketing campaign management, maybe others. Where can you join it all?

You join it on the cloud, of course. You load it all up, join it, analyze it, and test the hypotheses. Repeat as needed.

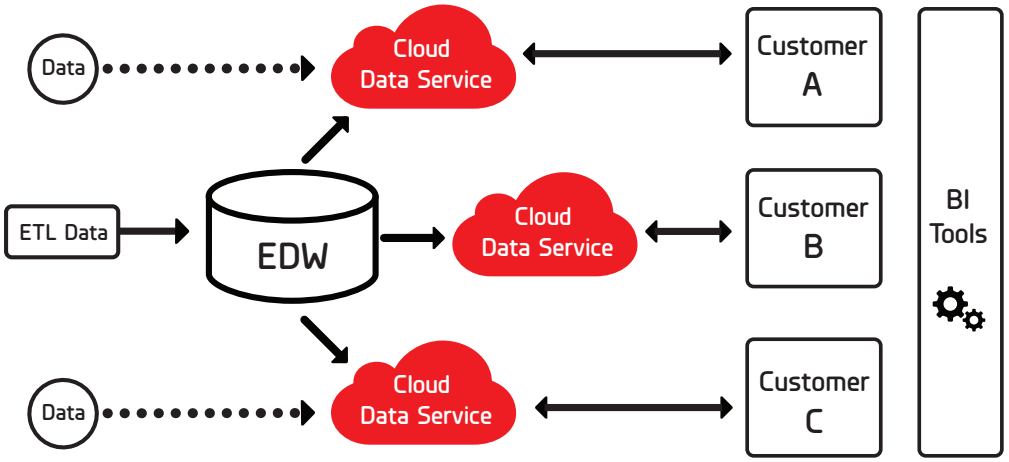


Cloud-based Data Marts

Your customers need analysis, too, but you can't let them behind the firewall. If you're a clever data analyst, you send them to the cloud. There you've created isolated pools of data in which your customers and other external stakeholders can look at anything and ask any questions whenever they want to in a secure and flexible environment.

Your solution never compromises the firewall. And it's so easy to set up, the everyday business user can do it.

In fact, secure and isolated data marts might be valuable enough – such as if they enabled easily shared results – that customers might pay for it.



Whitepaper

Data environments get cloudy as big data reigns

Clouds for any purpose

The cloud comes in all sizes and shapes, from managed services to base plat-forms. One or more can ease your big data hurdle, and quickly. Clouds go up fast. They also stay up at minimal cost, and they are more flexible than many on-premises options.

Within every cloud, there are more options. Some platforms require in-house technical skills. Others, known as managed cloud services, can be run by most business people.

Whichever cloud you choose, they all offer economy and flexibility. They stay around if you need them, or blow away if you don't. Perhaps best of all, they seem to operate just the same as any platform on premises. Users even get to stick with their favorite tools.

Treasure Data: A cloud service to deliver value from big data in days.

Treasure Data provides the first end-to-end cloud service for the entire data pipeline. By simplifying the process of data collection, storage and analysis, the Treasure Data Service helps businesses get rapid value from big data in days. Now organizations can easily and economically process enormous data volumes in near real-time with no infrastructure to manage.

Treasure Data supports structured and multi-structured data types such as web, application, mobile, log, sensor and machine data. Treasure Data is trusted by businesses for Advertising, Digital Marketing, Social, Mobile, Gaming, Automotive and other industry solutions. Customers span startups to Global Fortune 500 companies

Featured Customers

TOYOTA InfoTechnology Center, U.S.A., Inc

developerforce

MONEX

GETJAR

MobFox Mobile Advertising. Simplified.

COOKPAD http://cookpad.com

GREE

viki

About the Author:
Ted Cuzzillo is a longtime marketing writer, researcher, and journalist in the business intelligence industry. Other clients include Tableau Software and Tibco Spotfire. He also contributes to Information Management, TDWI's BI This Week, and Radiant Advisors, and his weblog, Datadoodle. He can be reached at TCTD@datadoodle.com.

TREASURE