ARCHITECTURAL AND LICENSING REQUIREMENTS FOR A MODERN ANALYTICS PLATFORM

By Neil Raden
Hired Brains Research
May, 2018

"Logic will get you from A to B.

Imagination will take you everywhere."

-Albert Einstein

EXECUTIVE SUMMARY

With so many options today for the placement of data and processing, there is a risk of fracturing analytics with a variety of solutions. On-premises strategies, public, private, hybrid and, managed private cloud alternatives are all robust, proven platforms.

It is safe to say that Data Science, Machine Learning (ML) and Artificial Intelligence (AI) have reached a level of maturity, now that there is sufficient infrastructure to support them, that organizations must learn to implement in their operations, product development, customer interaction, supply chains and a multitude of other applications. The problem facing organizations is the vast assortment of tools and technologies to use, both commercial and open source. This complexity will surely impede agility and speed, and create barriers for analysts to leverage these tools.

Analytics of all types has matured from single-user to single department to enterprise level. The number of operations in analytics today and the range of tools and processes requires a platform for orchestration that provides continuity, agility and scale, and can implement security, governance and abstraction for the users across all of those aspects. Unless business analysts and others interested in analytics can approach an environment that is less forbidding than a mismatched bag of tools, platforms and techniques, the promise of analytics will remain out of reach.

In this paper, I'll discuss the architectural and licensing requirements to empower your organization to develop a robust, scalable analytics platform that meets your business needs.

DISCUSSION

For a long time, a top-end relational database system, humming away somewhere in a data center, provided all of the functions and services you needed from a scalable data warehouse.

From the business analyst perspective, the physical aspects of a data warehouse were transparent. Historically, a data warehouse providing services for analytics and reporting operated on a homogeneous platform, and the whole process of ingesting data, conforming it and optimizing and satisfying queries was under the control of the database management system. New versions of the software came along at a regular cadence offering broader features and performance improvements.

It worked. In fact it worked so well that the basic concept was in practice for over thirty years.

Then everything changed. We were introduced to cloud computing, big data, data science, open source, the Internet of Things (IoT), AI, etc., and with that, a growing desire for more, faster, and better analytics. The single database on premises began to look like just one of a number of options. There were public clouds from Amazon Web Services (AWS) and Microsoft, hybrid clouds (where the data and work were split, albeit cooperatively, between on-premises and public cloud), and managed private cloud (a kind of cloud in your data center), in addition to traditional on-premises, or even a combination of the above.

Analytics today requires much more than a database platform optimized for data warehousing. There's much more data today and new sources of data. Businesses need to analyze all these different kinds of data in order to get a 360-degree view of customers and business. It requires new types of analytics; extended to business analysts in a simple and transparent way so they can do their job quickly and efficiently.

THE NEED FOR AN ANALYTICS PLATFORM

Advanced analytical work today often involves many steps and many tools that are only linked together by manual effort. What is missing is an analytics platform that allows for the performance of analytics, at any level of complexity, to be performed in a single workflow. Some ways this benefits the organization as well as the analyst or data scientists are to:

- Simplifying data access to data warehouses, data lakes and the emerging amounts of edge data from IoT
- Incorporating new types of data into your analytics process
- Applying cutting-edge advanced analytics and AI technologies
- Supporting your preferred data science workbenches and languages such as R, Python, and SQL
- Operationalizing prescriptive analytics to enable autonomous decision making
- Standardizing governance and security across a diverse range of data and platforms
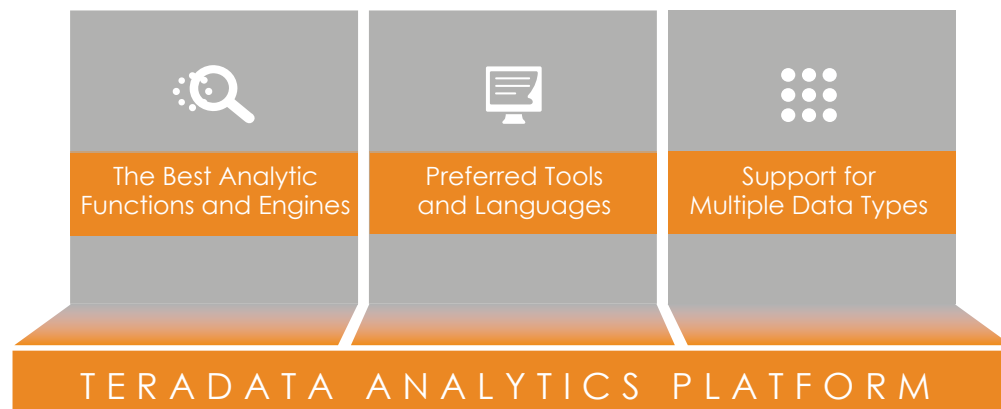
THE NEED FOR AN ANALYTICS PLATFORM

Features of an Analytics Platform include:

1. A wide variety of advanced analytic functions and engines to prepare and analyze data in a single workflow, at speed and scale.
2. A collection of engines such as machine learning and graph, including open-source options such as Apache® Spark™ and TensorFlow™ for access to a full range of algorithms.
3. Access to common commercial and open-source analytic tools and programming languages, like Python, R, SAS, or SQL, so that savvy analysts can analyze data in their language of choice. Analytic users can also leverage their favorite tools such as Jupyter, RStudio, and KNIME.
4. The ability to ingest and analyze multiple data types such as text, spatial, CSV, and JSON formats, including support of Apache Avro™.

A good example of an analytics platform is the Teradata® Analytics Platform, depicted below.

The Best Analytic Functions and Engines	Preferred Tools and Languages	Support for Multiple Data Types

TERADATA ANALYTICS PLATFORM

AN ANALYTICS PLATFORM SUPPORTS THE RAPID EVOLUTION OF DATA SCIENCE

All of this directly affects the work of data scientists. A connected analytics platform is the key to greatly improved efficiency of this scarce resource by providing:
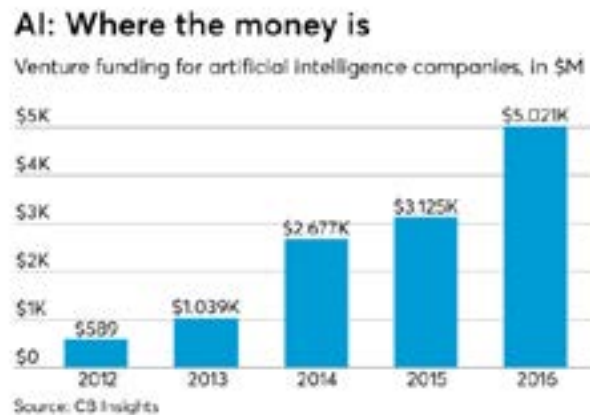
- Single workflow of data blending, through prep, modeling, deployment, and maintenance.
- Availability of cloud resources on-demand so they can apply the power they need episodically
- Single environment to include deep learning for machine learning, text, sentiment, and image analysis.
- Greater levels of automation in both modeling and prep reducing data science labor and increasing speed to solution.

FIRST RECOMMENDATION:

Eliminate, or at least minimize the burden of, extracts and interfaces to all of the components of your analytics infrastructure with an analytics platform that provides the efficiency to acclearate innovation and success in your business.

LOCATION FLEXIBILITY

A year or two ago, no one was thinking about a topology for instrumenting their operations (IoT) or seriously thinking that they would be deploying applications using AI so soon. As the chart below depicts, AI is a runaway train with over $5B in venture capital in just the past year. If decisions about cloud vs. on-premises or other alternatives led to decisions about where and how to implement analytics, there is a good chance those decisions are already wearing out.

AI: Where the money is

Venture funding for artificial intelligence companies, in $M

Year	Funding
2012	$589
2013	$1,039K
2014	$2,677K
2015	$3,125K
2016	$5,021K

Source: CB Insights

Your analytical application platform needs to be a great deal more flexible than you may have imagined just one or two years ago.

SECOND RECOMMENDATION:

Put the analytics and the data where it makes sense for you, and ensure that you analytical environment allows for that choice (i.e. if your data is primarily in the cloud, your analytics should be in the cloud; if your data is primarily on-premises, your analytics should be on-premises). Make the decision that is right for your business wherever your data sits, and refuse to succumb to the limitations of inflexible analytical tools and technologies.

LICENSE FLEXIBILITY

Tools and platforms for analytic technology tend to be inflexible in the market today, and usually require significant upfront investment. The problem is that analytic technologies, such as ML and AI, and even do-it-yourself analytics can stress existing environments. The deficiencies can add critical latency to the progress of your analytics programs.

Most tools for analytics and its supporting infrastructure are not designed to match workload requirements. If you add a server, you need a new license. Subscription-based licenses reduce the need for large capital outlays. **Here are some flexible licensing options you should look for when implementing a modern analytics platform:**

License Tier — Whether your requirements are for a cloud data warehouse or implementing an enterprise-scale production analytic environment, demand right-sized options and price points that meet both your technical and financial requirements.

Simplified Bundles — These should not be separate, incompatible products. All core features should be packaged into each of the bundles. This means that your work is protected. Applications, queries, scripts run without modification across all four versions.

Subscription-Based Licenses — Subscription-based licensing is an approach to purchasing in smaller increments instead of traditional perpetual licensing, where you pay upfront for a license that you own forever. When you choose a subscription-based license for a defined period, you can let it expire when it's no longer needed, or renew the license periodically for continued use. All software maintenance, upgrades, and bundled features are included.

THIRD RECOMMENDATION:

Select analytics tools and technologies that are offered via flexible licensing models, which allow you to focus on managing your business, versus contractual limitations and loopholes that limit the use of the technology and drain your energy and resources.

MOVEMENT FLEXIBILITY

With things running smoothly, why would anyone want to venture into new topologies where issues like security and reliability are not certain? Why complicate things? There can be compelling reasons:

Cost — IT budgets are not unlimited. In fact, so much of the budget is spent keeping the existing systems running there is often little funding left for new initiatives. While operational systems have more or less predictable resource requirements, analytical applications do not. You may want to separate your "bursty" processing, such as training a machine learning model, where the costs are more attractive than burdening your own data center. Or, as your analytical processes evolve, you may find alternative compute and storage models are more compatible with your evolving needs.

Logical location of data and processing — Clearly, much of the subject matter of data warehousing was derived from internal data, so shifting the data to the cloud did not immediately have much appeal. However, access to unimaginable amounts and types of data beyond the firewall piqued organizations' interest, and gathering and processing that data in the cloud made sense.

Freedom — Cloud providers were able to assume much of the ongoing maintenance and day-to-day management of the resources, freeing up valuable in-house resources.

Decisions about the cloud or other novel approaches such as on-premises VMWare and managed private cloud require a careful evaluation. But, there should be the ability to move your licenses around to various combinations without being required to purchase new ones.

FOURTH RECOMMENDATION:

Design your analytical environment with analytic technologies that allow for moving between deployment options without penalty. Changing business dynamics that may cause you to shift from on-premises to cloud, or vice-versa, should not require a complete repurchase or renegotiation of your vendor contracts. In addition, your preferred vendor should provide code compatibility across all of the environments, preserving your development efforts.

A MODEL EXAMPLE

For analytics at scale, Teradata has always stood out as the premier hub for gathering, integrating and serving analytics. Now the Teradata Analytics Platform is the solution for keeping the process whole across any environment. The Teradata Analytics Plaform is as follows:

Analyze Anything – Enables analytic users throughout the organization to use their preferred analytic tools and engines across data sources, at scale. With Analyze Anything, you can:

.

- **Deploy Anywhere** – Provides analytic processing across flexible deployment options, including the Teradata Cloud and public clouds, as well as on-premises on Teradata IntelliFlex® or commodity hardware.
- **Buy Any Way** – Empowers companies to purchase software in more accommodating ways based on specific use cases through simplified pricing bundles, subscription-based licenses, and as-a-service options.
- **Move Anytime** – Future-proofs buying decisions by taking advantage of license portability that provides flexibility to run analytics across deployment options.

CONCLUSION

An architecture for analytics in an organization must support a broad range of activities, from consistent reporting, to ad-hoc analytics to deep learning and AI. It is a broad and deep discipline. Most organizations expend precious time trying to orchestrate all of these activities, supporting a brittle offering with gaps and recurring problems with security, governance, performance and cost.
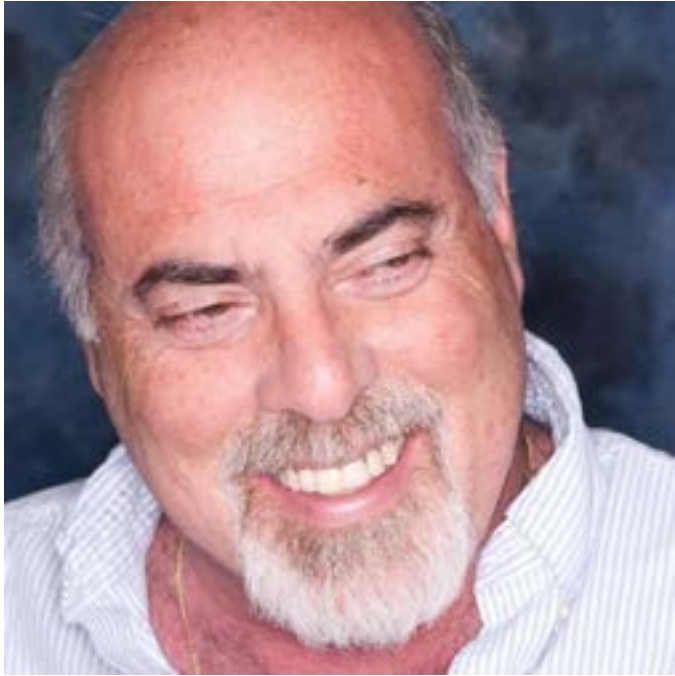
So where does this leave us?

The Teradata Analytics Platform is something to consider. It enables organizations to thrive with deeper analytic insights through operationalizing a wide array of the best analytic functions and engines, preferred tools and languages, and support for multiple data types.

Teradata Everywhere™ allows you to move your license horizontally, across platforms and from one location to another, or vertically, from one version to another; and provides code compatability across deployment options. No other major database vendor for advanced analytics offers this range of opportunities.

ABOUT THE AUTHOR

Neil Raden, based in Santa Fe, NM, is an active consultant and widely published author and speaker and also the founder of Hired Brains Research, Hiredbrains.com. Hired Brains provides thought leadership, context and advisory consulting and implementation services in Information Management, Business Intelligence/Analytics and Data Science for clients worldwide. Hired Brains also provides consulting, market research, product marketing and advisory services to the software industry. Neil was a contributing author to one of the first (1995) books on designing data warehouses and he is more recently the co-author of Smart (Enough) Systems: How to Deliver Competitive Advantage by Automating Hidden Decisions, Prentice-Hall, 2007. He welcomes your comments at nraden@hiredbrains.com or at his blog.