HITACHI
Inspire the Next

WHITE PAPER

Architected Blended Big Data With Hitachi Vantara

A Solution Brief

September 2019

Introduction

The value of big data is well recognized, with implementations across every size and type of business today. However, the most powerful analytic insights come not only from big data but also from blending big-data sources with other relevant data from internal and external systems and sources. Further challenging organizations are the changing skill sets, underlying technologies and security requirements that they need in order to be successful in a big-data-blended world. As new data types and sources continue to emerge, and organizations look more and more to big data to drive business decisions, IT departments need to architect a foundation that can future-proof them to the changing data landscape.

In the past, the approach to blending data for analytics focused on transforming and moving data into centralized data marts and warehouses, with all access and analytics targeting those aggregated stores. This approach was consistently time consuming, but it made sense when most data was in structured, largely relational formats at reasonable volumes.

This approach no longer makes sense in the current world of big and specialized data. We have rapidly moved into an era of distributed analytic data architectures, in which data remains housed in the type of store most optimal for its volume and variety. This new ecosystem is called different things by different research organizations: the hybrid data ecosystem by EMA, the logical data warehouse by Gartner and the multiplatform data warehouse environments by TDWI. But all recognize this distributed approach for the future. Why? Because it simply doesn't make sense to move big data into an enterprise data warehouse (EDW) as we did with relational data. The structural variety and volume of big data make it extremely difficult and time consuming, delaying any valuable use of the data. Plus, the economics are not feasible due to the volumes involved. Think for example of the typical volumes major telecoms must manage today with their call data records: One major provider reports that it deals with 140 billion records in 650,000 files per day!

DRIVE VALUE FROM BIG DATA

Drive incremental revenue

- Predict customer behavior across all channels.
- Understand and monetize customer behavior.
- Begin to monetize data as a service.

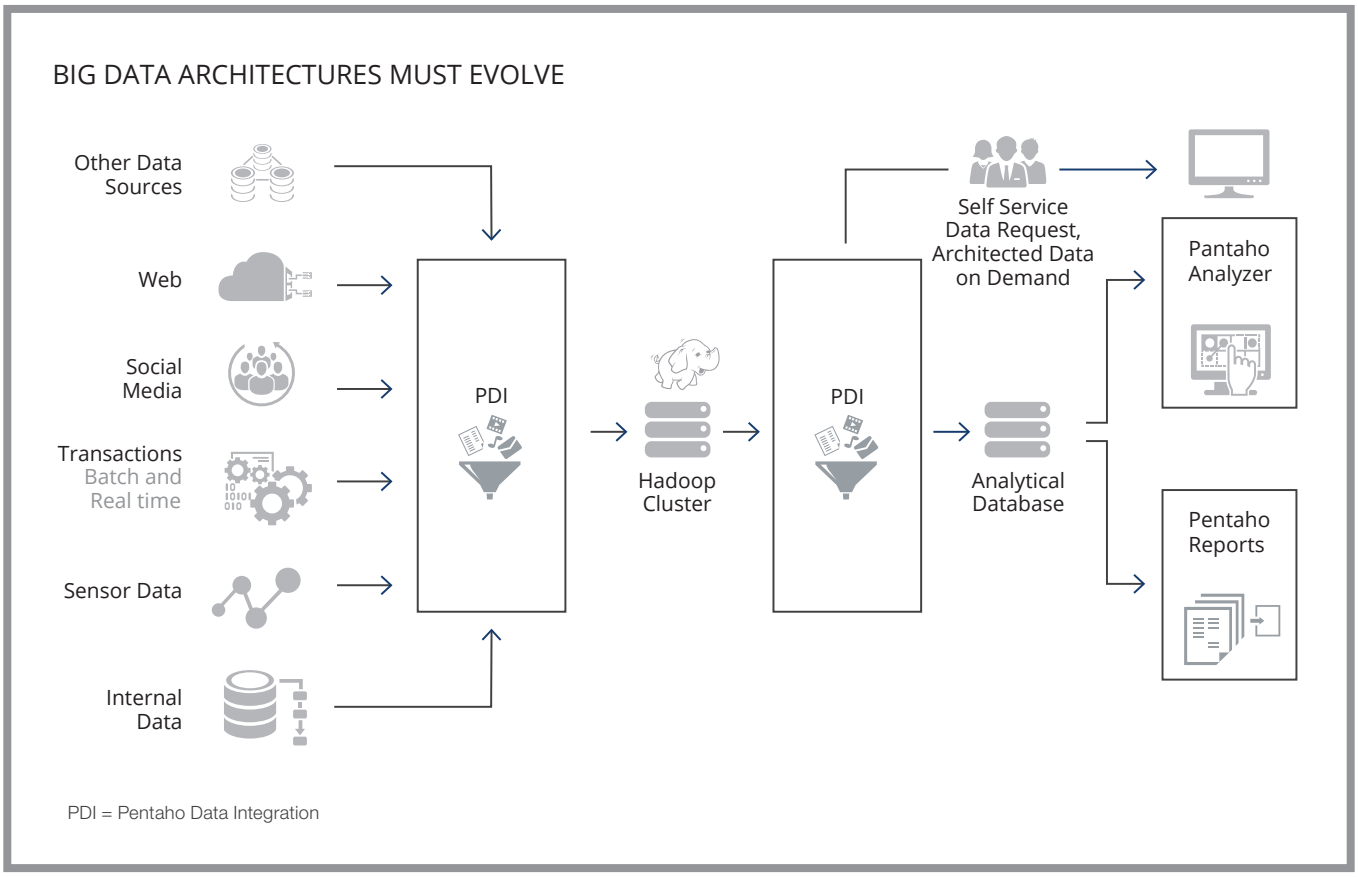
Improve operational effectiveness

- Machines and sensors: Predict failures and network attacks.
- Financial risk management: Reduce fraud and increase security.
- Reduce data warehouse cost.

Customer Experience

- Build a 360° view to fully understand and serve the customer.
- Drive personalized and adjusted interaction.
- Use automated recommendations logic.

Figure 1.Hitachi Vantara's Pentaho platform supports a new age of information architecture.

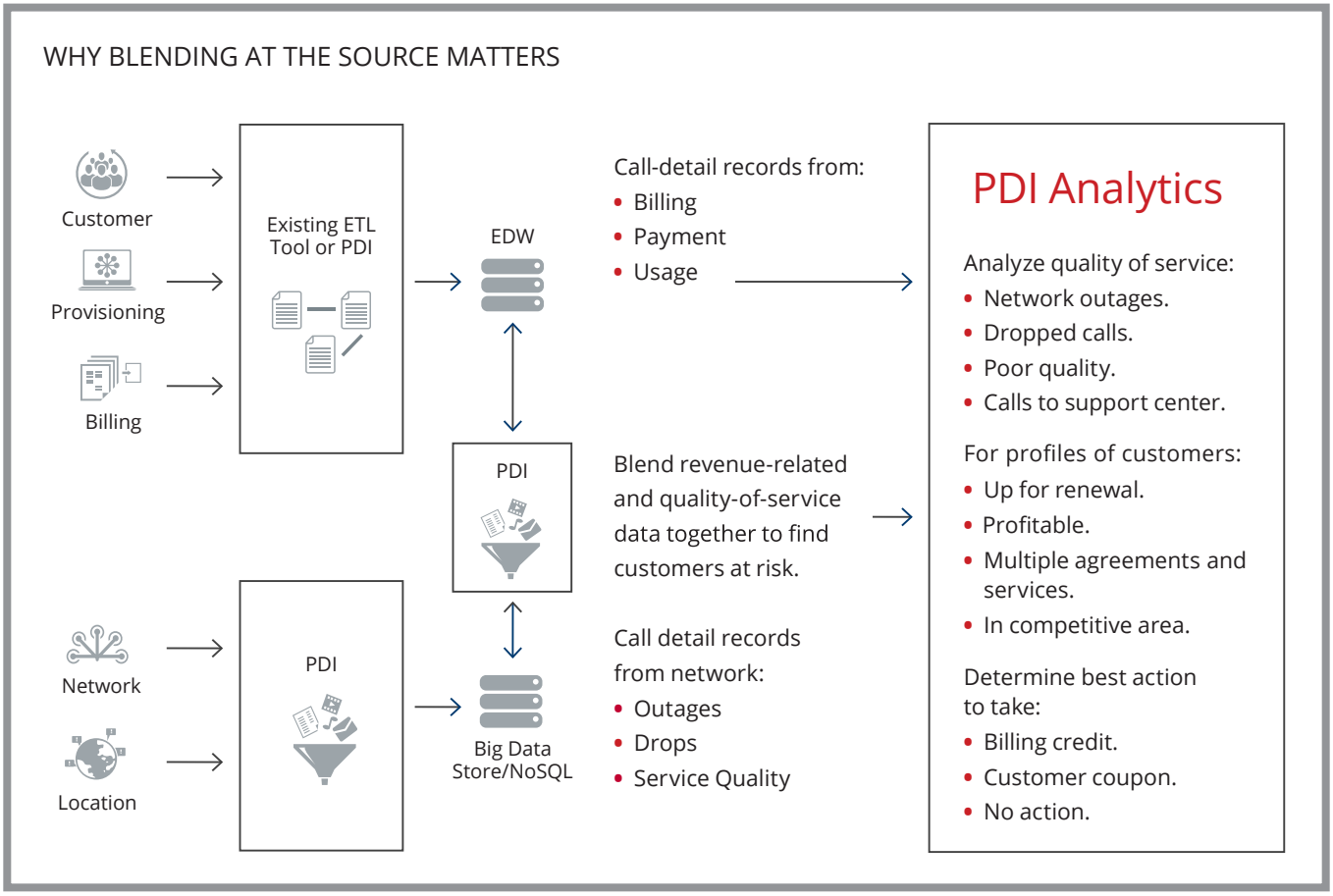


The pace of business and competition has also changed, exacerbating the issues. Competitiveness and success require the ability to respond rapidly to changing conditions in real time. Businesses cannot afford to wait for data to be extracted, merged, cleansed, transformed and stored before it can be analyzed. Agile analytics require the data to be accessible where it resides at the source to ensure it is based on the most up-to-date view possible. Businesses must utilize the most efficient point of processing to deal with the potential volumes being searched and returned.

All of these factors signal a new age of information architecture that must manage the flow of data through analytics differently than before. Data must be blended in place, across disparate stores and structures, while ensuring performance and ease of use across the breadth of analytics, from historical to operational to predictive (see Figure 1).

Let's look at an example of blending at the source to better understand these points. In Figure 2, we have an example of telecom customer experience analytics. Customer experience analytics have the same goal in every industry: Prevent customer churn and create better loyalty to protect and grow revenue. After all, in this age of commoditization, service and fast response to product requests become the new differentiators driving loyalty in most industries. Telco customer allegiance comes mostly from satisfaction with calling plans and the quality and availability of service. Call-detail records have long been created and derived from the operational systems for access to business intelligence and reporting systems via warehousing, but they only make up part of the picture.

Figure 2. Prevent customer churn, create better loyalty and protect revenue.



PDI = Pentaho Data Integration, EDW = enterprise data warehouse, ETL = extract, transform, load

Quality of service changes in real time depending on the network. Ask: the customer able to connect, hear and remain connected without being dropped, and so forth? This network-based data is usually captured in a big data source that is capable of handling the volume and unstructured nature of the data. In addition, it can be blended with the call detail record information to give the complete picture of a customer's experience.

With Hitachi Vantara, you can easily create architected, blended views across both the traditional call-detail records in the warehouse and the network data streaming into any big data or NoSQL store (for example, Cloudera, MongoDB, Cassandra, and so forth). And you can do so without sacrificing the governance or performance you expect. These blended views allow analysts and customer call centers to get accurate, of-the-minute information in near-real time; they can determine the best action to take for each customer to maximize satisfaction and retain them as loyal customers, even when outages or other service quality issues occur.

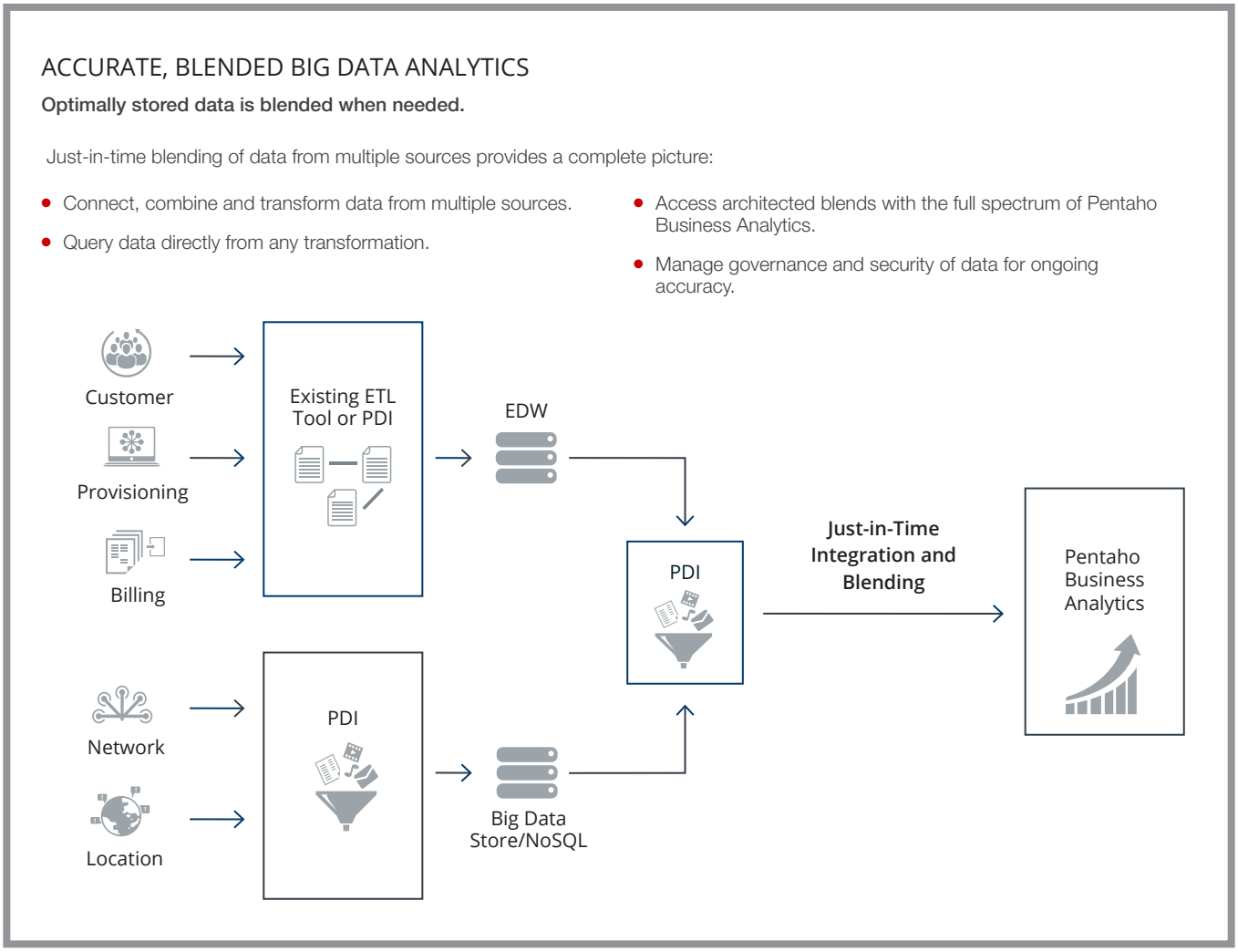
Just-in-time, architected blending delivers accurate big data analytics based on blended data (see Figure 3). You can connect to, combine and even transform data from any of the multiple data stores in your hybrid data ecosystem into blended views. Then, query the data directly via that view using the full spectrum of analytics in Pentaho's analytics platform, including predictive analytics (see FIgure 4).

Most importantly, since these blends are architected on the source data, you maintain all the rules of governance and security over the data while providing the ease of use and real time access needed for today's agile analytics requirements. The key governance and security benefit is that sensitive data is kept from those who are not allowed to use or view it.

- Full life-cycle maintenance, change management and control ensure the blends being used meet changing requirements.
- Auditability is preserved.
- Blends are designed with full knowledge of the underlying data volumes and source system capabilities and constraints. This approach throughput and performance during analytic access and prevents the "query from hell" and "runaway query" problems prevalent in many data federation tools.

We combine the power of drag-and-drop design across all data sources, including schemas generated on read from big data sources, with knowledge of the full data semantics (the real meaning, cardinality, and match of fields and values in the data). This approach ensures that your business gets accurate results. Decisions become optimized and actions can be taken to significantly impact business positively and improve performance results.

Figure 3. Architected blending delivers accurate big data analytics.

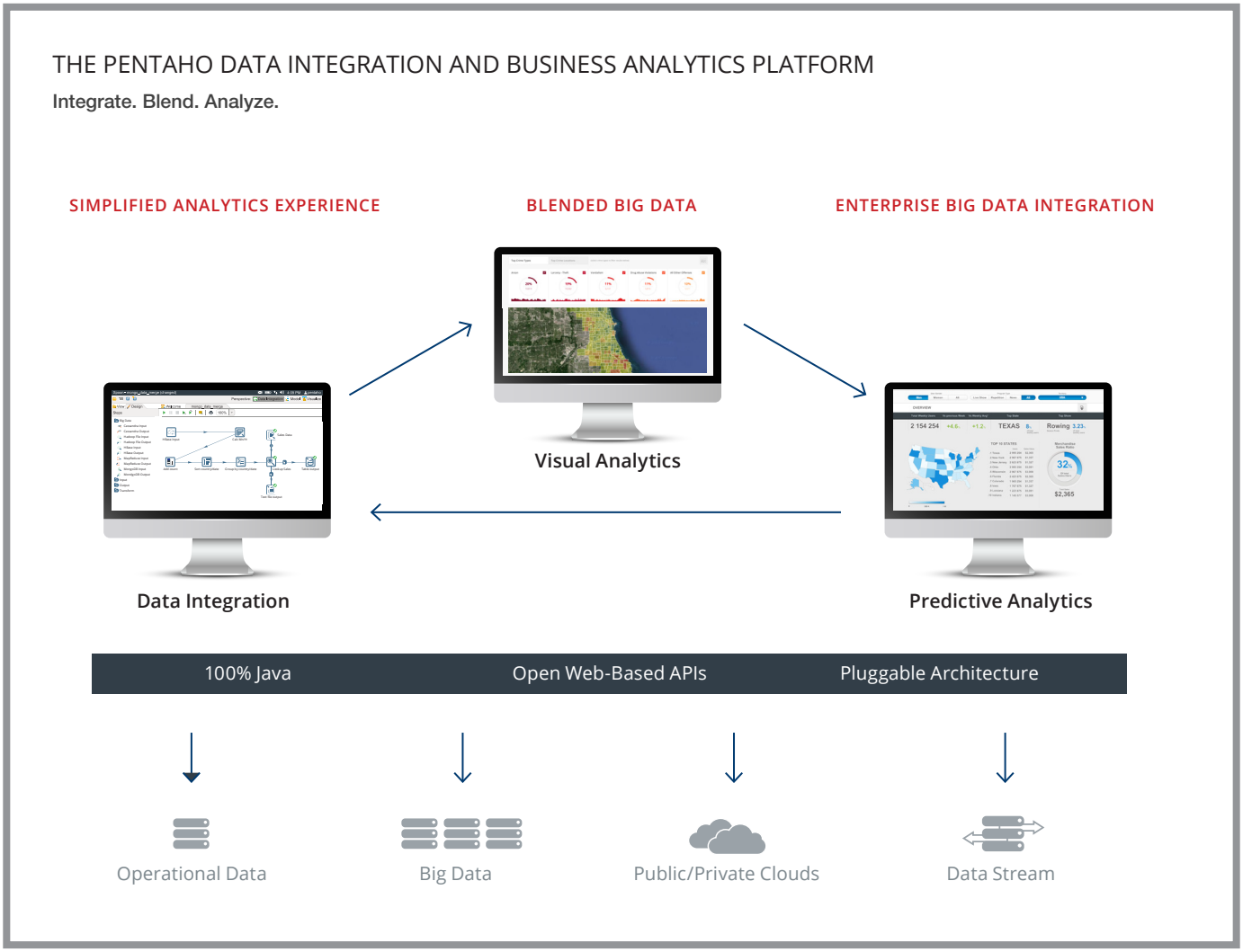


PDI = Pentaho Data Integration, EDW = enterprise data warehouse, ETL = extract, transform, load

Other solutions in the market talk about blending, but they may be referencing a different type of blending. Blending "at the glass" (blending done by end users or analysts away from the source with no knowledge of the underlying semantics) often delivers inaccurate or even completely incorrect results. There is no way to ensure that the chosen fields being matched truly do match.

For instance, think about what happens when someone matches two fields both named "revenue" in records for a single customer, but one is a monthly sum total and the other is a daily total. This difference won't be apparent to an analyst since the blending is done based on similar names. The analyst then runs a summation that adds the two together as the day's total revenue from that customer. Unwittingly, the monthly figure is added into each day's total, distorting the actual revenue generated from that customer dramatically. The business then identifies that customer as highly profitable and offers significant discounts to maintain their interest. Not only have you targeted the wrong customer and potentially ignored the real profitable customers, but you've also now given undeserved discounts. The net result lowers your revenue from this customer and potentially causes a loss of other profitable customers who were more deserving but left in favor of competitors offering them discounts. You've made the wrong decision because the analytics themselves were inaccurate and incorrect.

Figure 4. Use Hitachi Vantara's Pentaho platform solution to integrate, blend and analyze big data.



The only way to avoid this situation without tools that blend at the source is to train every user and analyst on the semantics of the data to ensure reliable results. The solution is of course largely infeasible for most organizations as it would take far too much time and expense while impacting productivity.

Even if you can take on this level of investment in training, you still face issues with the timeliness of the data, since other tools do not pull from the source systems. It is impossible to know if the data pulled is indeed the latest and therefore the most accurate on that level.

Appearance isn't everything, and your decisions should not be based on interpretation. You need to be sure the analytics lead to accurate information so you can make the right decisions. You need architected data blending at the source. You need Hitachi Vantara.

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact

P-002-B BTD September 2019