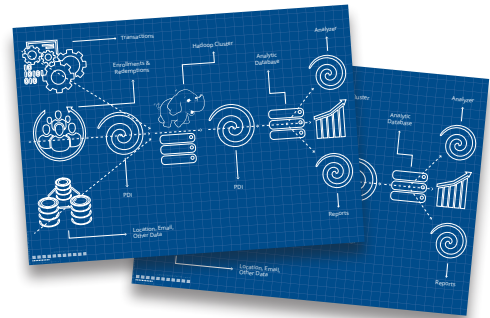


BLUEPRINT FOR BIG DATA SUCCESS:

Streamlined Data Refinery

Drive a Sustainable Big Data Analytics Strategy



WHAT IS IT?

- > Traditional extract, transform and load (ETL) processes can no longer cope with the variety and volume of data that need to be incorporated in analytics because they are designed for top-down development processes.
- > An agile data integration process that includes transformation steps and tools along with enable data processing on a low latency data store, such as Hadoop.
- > Includes self-service analytics along with an analytic database for high speed data querying and visualization.

WHY DO IT?

- > Provide business users insight into all data in a timely fashion, including diverse sources at high volume.
- > Engineer new data sets for predictive analytics more quickly thanks to rapid ingestion and powerful processing.
- > Scale ETL and data management cost savings by utilizing the right technology for the most appropriate purpose.
- > Logical next step from Data Warehouse Optimization use case.

VALUE OF PENTAHO

- > **Data Collaboration:** Using Pentaho Data Integration (PDI), a data developer and business analyst can collaborate in real-time on the creation of new data sets using one product for both data engineering and data visualization.
- > **Staff savings & productivity:** Pentaho provides a simple mechanism for creating visual mapReduce jobs along with Big Data integration which means data developers can move and process data between Hadoop and a variety of other sources and systems easily and efficiently.
- > **Complete Solution:** Broad data integration to accommodate existing architectures plus a powerful array of self-service analytics and visualizations for all end users, whether they be business users, analysts, and/or data scientists!
- > **Time to value:** Significantly reduce development time for mapReduce by 15x over typical handcoding and scripting processes.

Example: Online Shopping Service

Challenge:

- > Analyze customer data in real time to improve targeted advertising and offers.
- > Reduce time and cost of analyzing 5 billion consumer transactions.

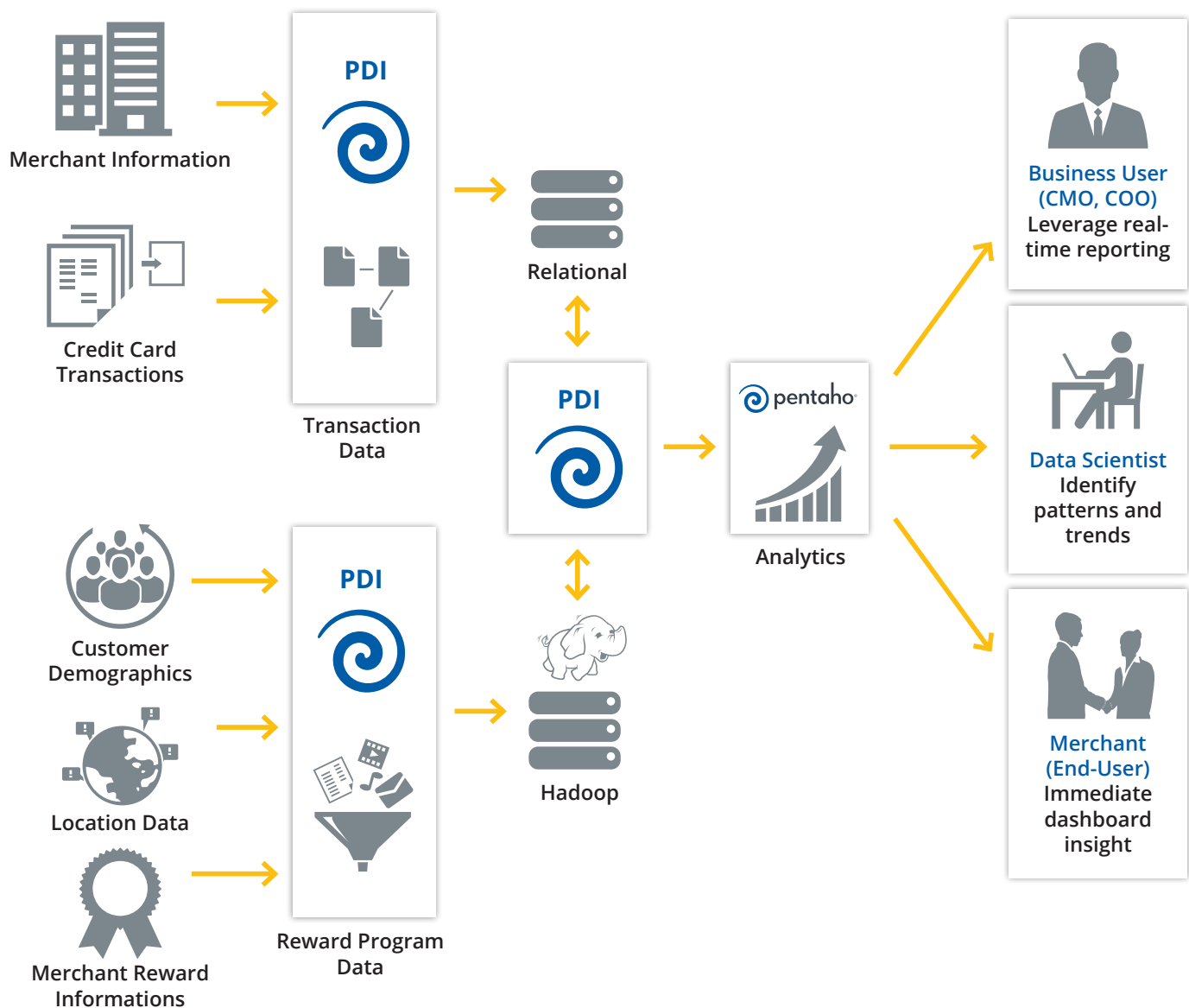
Solution:

- > Pentaho Business Analytics and Data Integration platform was utilized to extract, integrate, and analyze nearly 5TB of historical data from a variety of sources, including merchant info, transactions, demographics, location data, and more.

- > Data ingested via Hadoop then processed and loaded into an analytic DB for highperformance reporting and analysis.

Benefits:

- > Reduction in overall IT costs and no need for highly specialized, and hard to find Hadoop and Java programming skills.
- > Improved analytical capabilities to better understand and serve customers.
- > Reduced processing time by 70%, enabling IT to meet the customer SLAs.



To learn more about Pentaho software and services, contact Pentaho at pentaho.com/contact or +1 (866) 660-7555