BLUEPRINT FOR BIG DATA SUCCESS:

Optimize the Data Warehouse
Reduce Costs & Improve Operational Performance

WHAT IS IT?

> Existing Data Warehouse implementations are struggling with rising data volumes, causing performance concerns and impacting a businesses ability to meet SLA's. And at the same time, the costs involved in adding additional infrastructure has increased rapidly.

> The optimization solution is intended to relieve pressure on existing infrastructure by offloading less frequently used data, and improve performance by moving data transformation workloads to Hadoop.

WHY DO IT?

> Reduce data warehouse costs by reducing the need for expensive DW infrastructure, and save management costs through the use of low latency data store.

> Meet SLAs to deliver data on time, & empower business users to meet goals on time.

> Satisfy compliance requirements by providing ready access to historical data (such archived financial transaction records).

VALUE OF PENTAHO

> **Staff savings & productivity:** Pentaho Data Integration for big data, along with Pentaho MapReduce means existing data warehouse developers can more easily move and process data between the data warehouse and Hadoop without coding.

> **Time to value:** Organizations can speed development time with Pentaho MapReduce by up to 15 times versus hand-coding and scripting.

> **Faster job execution:** Pentaho MapReduce runs in parallel in cluster, delivering faster performance when compared to other scripting tools.

Example: Network Storage Company

Challenge:
> Management of rapidly expanding storage device data not possible with existing Oracle RDBMS-based Data Warehouse.
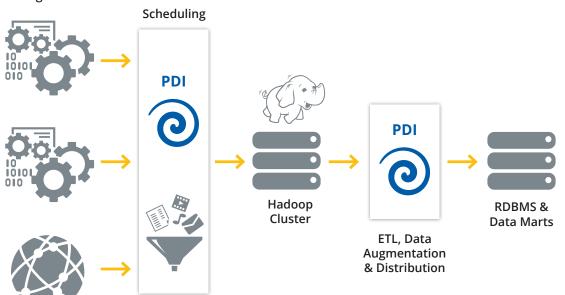> Meeting data processing SLAs.

Solution:
> Implemented Hadoop to offload data from existing Oracle data warehouse.
> Implemented Pentaho Data Integration for scheduling, ETL, data augmentation, & distribution.

Benefits:
> Achieved a 15x data cost improvement, through reduced storage costs, more efficient use of personnel and streamlined mapReduce processing in Hadoop.
> Easy to use ETL for Hadoop & existing data sources
> Met SLAs to deliver data to business users, engineering, and customer support on time.

Diagnostic Data from Storage Devices

Scheduling

PDI

Hadoop Cluster

PDI

ETL, Data Augmentation & Distribution

RDBMS & Data Marts

One million automated performance messages per week

Business User
Aggregate device performance analysis

Engineer
Identify device failure risks

Customer Support
Reporting to resolve cases

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