Introduction

Providing business analytics solutions in today’s environment of big and diverse data can be a challenge. Technologies are evolving every day and solutions require computing competencies in addition to traditional DW/BI skills.

Hitachi Vantara’s Pentaho platform provides an ideal solution for companies looking to expand their analytics capabilities to include these new big data types and sources, and gain from our experts who have successfully deployed many production solutions for financial, healthcare, advertising, publishing, and technology industries.

The Partner Workshop is designed to guide Hitachi Vantara Partners in acquiring requisite knowledge and skills in implementing a Pentaho Solution following Professional Services VANTAGE guidelines.

Audience

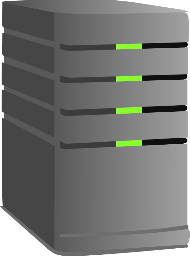
To complete the Workshop, each of the Partner organizations will require a person with the following role / activities:

|  |  |
| --- | --- |
| Role | Activities |
| Solution Architect | * Data Discovery: |
|  | * Data Warehouse Techniques: Slowly Changing Dimensions (CDC) |
|  | * Data Modelling: Using MySQL Workbench or ER Tool |
|  |  |
|  | * GitHub: Repository for Jb and ktr |
|  | * SQL: scripts to create / join tables / schemas |
|  | * Unix: Archive install of Pentaho server |
|  |  |
| ETL Developer | * Data Discovery: |
|  | * Data Warehouse Techniques: Slowly Changing Dimensions (CDC) |
|  | * Pentaho Data Integration: ETL workflows |
|  |  |
|  | * GitHub: Repository for Jb and ktr |
|  | * Rapid Prototyping: Data Source Wizard or DET |
|  | * SQL: scripts to create / join tables / schemas |
|  |  |
| Analytics Developer | * Reporting Requirements: |
|  | * Rapid Prototyping: Data Source Wizard or DET |
|  | * SQL: Join tables |
|  |  |
|  | * Pentaho Business Analytics Client Tools |

Lab Environment

Each team member will have their own server. The servers will be networked and have access to the internet.

The Pentaho server is hardware-independent and runs on server-class computers that comply with these specifications for minimum hardware and required operating systems:



Pentaho Server

Processor: 4 cores

RAM: 16 GB

Disk Space: 500 GB

Networked

Internet Access

Ubuntu Server 16.04 LTS

*Pentaho Server*

*Repository*

MySQL CE 5.7



Pentaho Clients 1 to 3

Processor: 4 cores

RAM: 16 GB

Disk Space: 500 GB

Networked

Internet Access

Microsoft Windows 10

*Pentaho Aggregation Designer*

*Pentaho Data Integration*

*Pentaho Metadata Editor*

*Pentaho Report Designer*

*Pentaho Schema Workbench*

Partner Workshop: Overview & Installation

|  |  |  |
| --- | --- | --- |
| Day | Topic |  |
| 1 | Introduction | * Agenda * VANTAGE methodology * Lab Environment |
|  | Story | * Secondhand Lens * Requirements:   + Non-functional – GitHub Repository   Database Management Tool  MySQL - Relational  MySQL - Star   * + Business – Analysis Reports   Interactive Reports  Report Designer  Dashboard |
|  | Installation & Configuration | * Pentaho Business Analytics Suite v8.0   + Archive Installation   + Client Tools (Wizard)   + Security (Roles/Users) |
| HALF DAY BREAK AND CHECKPOINT | | |
|  | Solution Architecture | * Assess customer’s requirements and design the solution architecture |
| Deliverables |  | * Working Pentaho Servers on windows and Linux servers * Working client and development tools   + Git clients   + DB Explorers   + Etc. |
| END OF DAY CHECKPOINT | | |

Partner Workshop: Solution Design, Models / Schemas

|  |  |  |
| --- | --- | --- |
| Day |  | Milestones |
| 2 | Planning | * Team definition * Data Discovery * Solution Design Template review * User stories and task list template review |
|  | Discover | * Data Discovery * Discover Infrastructure * Setup local environments + Git setup |
| HALF DAY BREAK AND CHECKPOINT | | |
|  | Design | * Design the schemas * Design dim and facts * Define reports * Define implementation tasks * Solution Design Document |
| Deliverables |  | * Solution Design Document |
| END OF DAY CHECKPOINT | | |

Partner Workshop: ETL Scripts, Metadata Models & Schemas, Reports

|  |  |  |
| --- | --- | --- |
| Day |  | Milestones |
| 3 | Execute | * Create scripts * Implement ETLs to create the DWH and Datamarts and populate them * Create Metadata Model * Create Mondrian Schema * Create PRD reports |
| HALF DAY BREAK AND CHECKPOINT | | |
|  | Execute | * Continue to work on   + Metadata Model   + Mondrian Schema   + PRD Reports   + Analyzer Reports   + Interactive Reports   + Dashboards |
| Deliverables |  | * Demo the application |
| END OF DAY CHECKPOINT | | |

Partner Workshop: Deployment, Testing

|  |  |  |
| --- | --- | --- |
| Day |  | Milestones |
| 4 | Execute | * Test jobs and transformations * Schedule jobs * Test reports * Deploy end-2-end |
| HALF DAY BREAK AND CHECKPOINT | | |
|  | Enable | * Document Deployment * Prepare Presentation to client * Clean up |
| END OF DAY CHECKPOINT | | |

Partner Workshop: Presentations

|  |  |  |
| --- | --- | --- |
| Day |  | Milestones |
| 5 | Enable | * Present to client * Demo * Deploy to production server |

Story

Secondhand Lens (SL) have a number of Camera stores throughout the United Kingdom and United States, which primarily sell secondhand camera lenses. Along with their store sales, they also have a website which customers can purchase these lenses online.  
Up to now, SL’s business analysts and management team had to reach out to the I.T operations team to run queries on the transactional database to generate basic reports. Data Analysts need to be empowered to 'slice and dice' the data on an ad-hoc basis, i.e. by time, geography, or lens type. Managers wish to updated in a timely manner with operational and financial reports.

A new approach is required to access their data, and Pentaho Analytics Suite is a perfect solution. The company has provided the following high-level requirements:

* Users login to a portal and execute / create reports.
* Data Analysts require an OLAP solution, so they can measure their sales:
  + By sales channel
  + By store
  + By time
  + By city
  + By product
* Business needs an ad hoc way to generate reports on transaction details.
* End of month reports should be generated as pdf to be distributed to the management team for monthly strategic meetings, and investors.
* Sales Manager would like to have a high-level view of the overall sales.

Phase I

For Phase I of the project, SL can use any security scheme that comes with the product, but all users must have accounts with proper roles and permissions. For future phases, the platform needs to be integrated with Active Directory.

* SL’s IT environment, has a development and a production environment. Code or artifacts need to be promoted from a source control repository.
* The platform will not have direct network access to the production database. Data file extracts can be placed in any location for the platform to process them.
* The data files are extracted several times during day to keep the reporting data as up to date as possible.
* Development and production environments are on virtual servers running Ubuntu Linux.

Phase II

* SL would like to know how the application can be extended so it can support reports for store managers such that they can only see the performance of their own stores.
* SL is in the process of implementing a portal for their employees, store personnel and customers which goes live next year. They would like to know how they can leverage the Pentaho platform for reporting and analytics within their web application.