Pentaho Update

Bob Kemper
EVP Engineering, Pentaho

Agenda

- Year in review

- Pentaho Strategy

- Hitachi and Pentaho

- Roadmap(s) - abbreviated

Pentaho 2014/2015 - A year of growth

- Company awareness - Pentaho widely recognized as leader in Big Data
  - Number 1 or 2 in most "Share of Voice" metrics

- Major new large customer wins
  - Enterprise deployments and Big Data represent significant percent of business

- Continued staff growth in both US and EMEA teams
  - Across all areas of the business – Sales, marketing, services, support, engineering

Pentaho 2014/2015

- Major product releases

  - Several large product release during the year

    - 5.3 in Q1, 5.4 in Q2

  - 6.0 Platform release in October

- And … Acquired by Hitachi Data Systems in May 2015

  - More on that to come….

Product Releases – 5.3 (February)

CE/EE Features

- Support for CDH 5.2 and MapR 4.0.1

- Named Clusters

- Carte Clustering

- PDI SDK as an "Official" release package

- BA Server / Plugin Improvements (Data Access for example - 0 bad builds during ALL 5.3.0 RC builds)

- PDI SDK documentation

- Carte Clustering documentation

- MT Documentation

- Over 300 Closed and Validated jiras! (Majority common code)

EE Only Features

- Using an existing model for publish

- Using user-defined annotations to assist auto-model generation

- Validated to work with Amazon Redshift and Impala

- Vizor as a package

- Analyzer JS API and Documentation

- PIR Improvements

- Analyzer API documentation

Product Releases – 5.4 (June)

CE/EE Features

- New PDI Look and Feel (CE)
- BA Server Utilities
- C*Tools support for RequireJS
- Improved Named Clusters support
- AWS EMR Shim Support
- Spark Integration
- User Console localizations
- Mongo 3.0 Support
- Google Analytics step upgrades

EE Only Features

- Analyzer Extensibility APIs
- DI Server REST APIs
- SAP Hana support in PDI
- SDR Modeling improvements

Product Releases – 6.0 (October)

CE/EE Features

- PDI Data Services – Opened and major advances

- OSGI Support

- New PDI Marketplace

- Testing framework for Named Clusters

- Library and back end upgrades (Java8, Tomcat8, Drools, VFS, etc, etc.)

- REST Apis for User Management

- Groundwork for a "Merged BA and DI Server"

- Major step improvements/defect resolution in PDI

- Lineage capture in PDI (see blogpost by Pedro on displaying through yEd)

EE Only Features

- Inline Modeling in Analyzer

- Annotation improvements in PDI

- System monitoring via SNMP

- Lineage information through MetaIntegration

- Improved import/export

Pentaho enables the Modern Data-driven Business

pentaho®
A Hitachi Group Company

Modern, Cohesive Business Analytics and Data Integration Platform

- Full spectrum of analytics for all key roles
- Embeddable, cloud-ready analytics
- Broadest and deepest big data integration

Innovation Through an Open Heritage

- Open, pluggable, purpose-built for the future
- Sustained leadership in big data ecosystem

Business Momentum

- Over 1,500 commercial customers
- Over 10,000 production deployments

Pentaho is Architected for the Future

Purpose-built to support an evolving data & analytics landscape

Data Integration

Data Orchestration		
Big Data	Data Warehouse	Other

Business Analytics

Embedded Analytics		
Analyzer	Dashboards	Reporting

Open

Pluggable

100% Java

Bringing it All Together
Pentaho Bridges Big Data and Traditional Architectures

Network

Location

Web

Social Media

Customer

Provisioning

Billing

pentaho
Existing Process or **Pentaho**

pentaho
Existing ETL Tool or **Pentaho**

NoSQL
mongoDB

Hadoop Cluster

pentaho
Data Integration

pentaho
Data Integration

On-Demand Integration & Blending

EDW

pentaho
Existing ETL Tool or **PDI**

Analytic DB

Data Marts

pentaho
Analytics

BI Tools
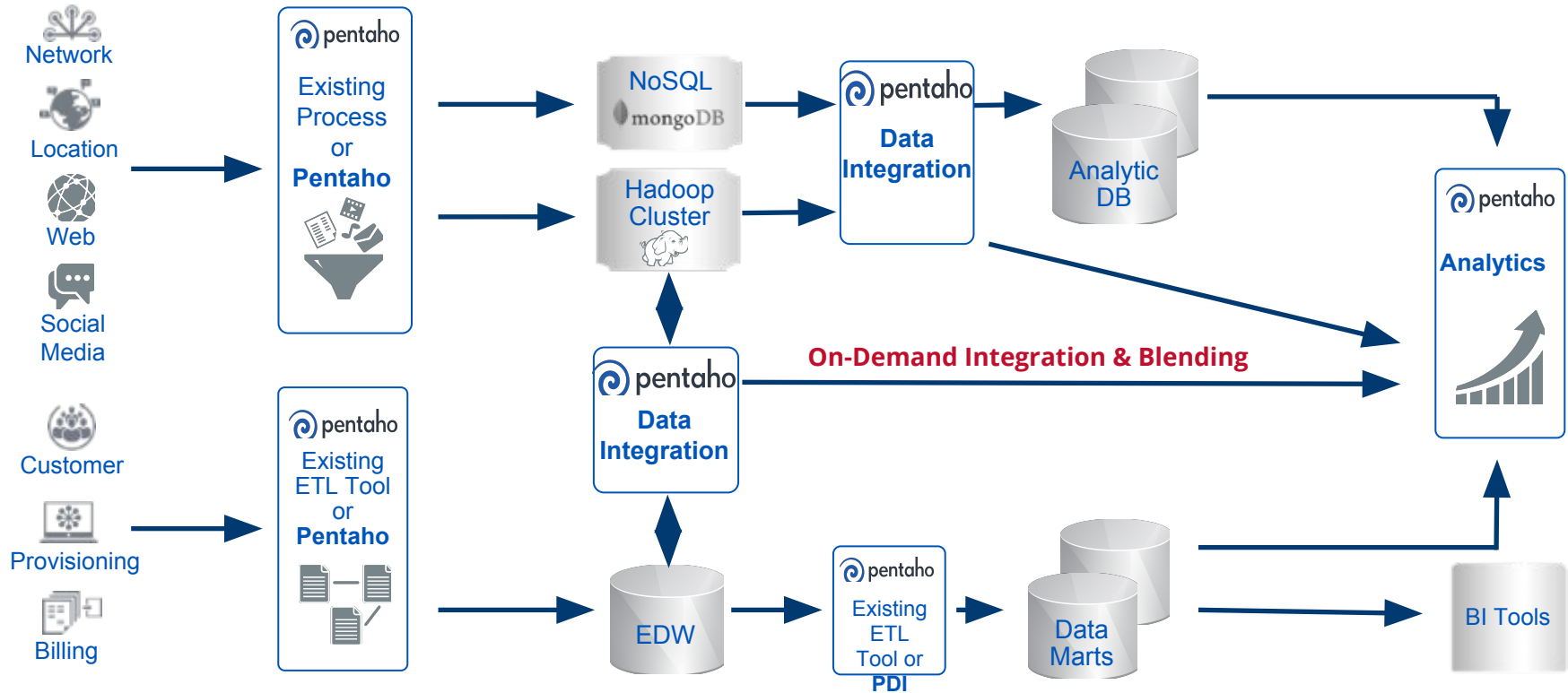
End-to-End Self-Service Delivery of Big Data Analytics
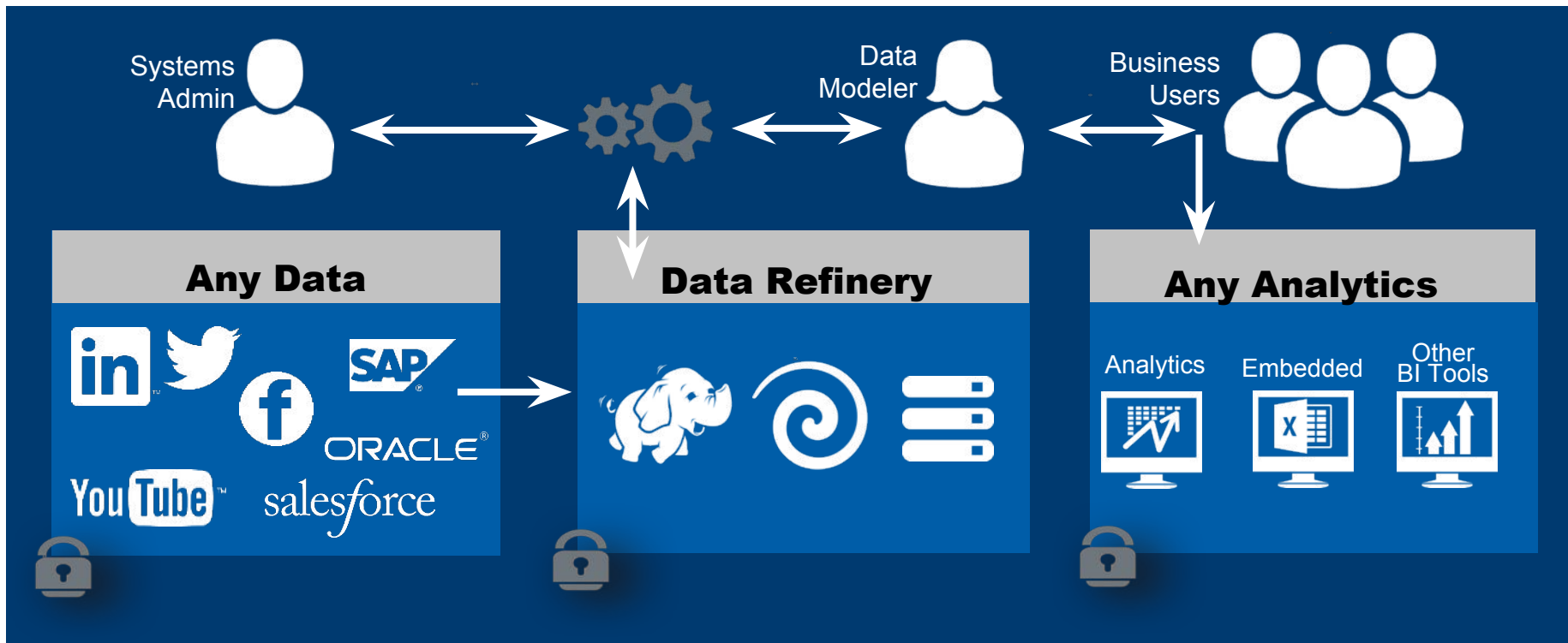
BIG DATA LAKE: STRATEGIC CORPORATE ASSET

BIG DATA SOURCES

HADOOP/ DATA LAKE

PDI

TRADITIONAL DATA

PDI

DATA WAREHOUSE

ANALYTIC DATASETS

CLOUD

VIRTUALIZED DATA

VISUALIZE

REPORT

PREDICT

EMBED

DATA AS A SERVICE
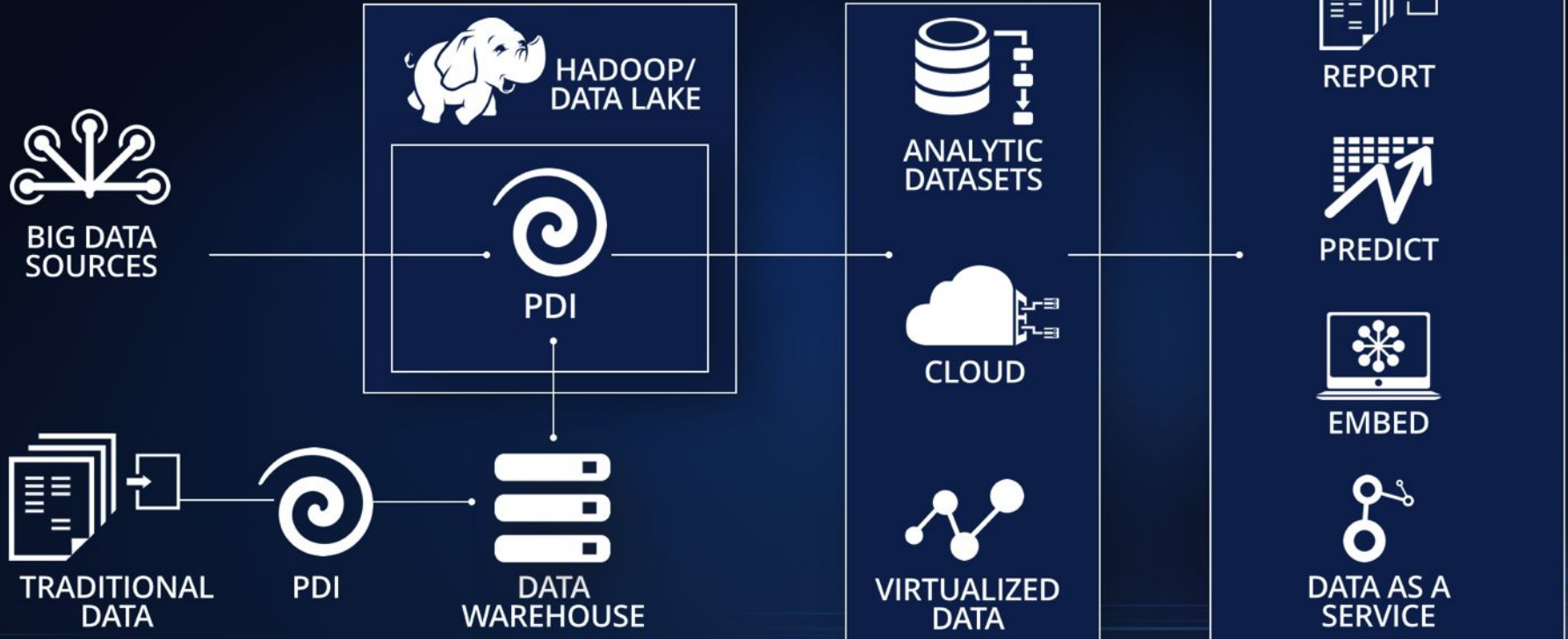
Why Did Hitachi Acquire Pentaho?

Leader in the **big data & business analytics** market

The Next Revolution

INDUSTRIAL
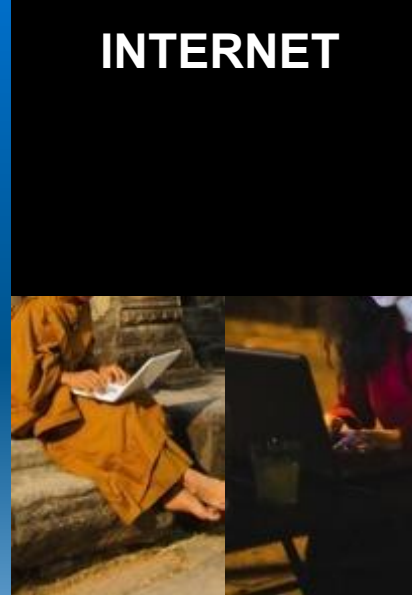
DIGITAL

INTERNET

INTERNET OF THINGS

Internet of Things in 2020

pentaho®
A Hitachi Group Company

152M
connected cars (IHS Automotive)

1.1B
smart meters installed (Navigant Research)

200B
connected objects (Intel)

$19T
economic value created by Internet of Things (Cisco)

$2.5T
annual impact of IoT in healthcare (by 2025, McKinsey)

>90%
IoT data to be hosted on cloud platforms (IDC)

The Opportunity in 2020

Internet of Things
$7.1 trillion

Social Innovation
$2 trillion

Our IoT Strategy

Integrating IT and Infrastructure Technologies for a Safer, Smarter, Healthier World

HITACHI'S SOCIAL INNOVATION BUSINESS

The Internet of Things *That Matter*

The Internet of Things *That Matter*

ENERGY

MICROGRIDS & GRID STABILITY	WIND TURBINES	INFORMATION & CONTROL SYSTEMS
ENERGY SAVINGS AS A SERVICE	SOLAR POWER	NUCLEAR POWER PLANTS
BATTERY SYSTEMS		

HEALTHCARE

CLINICAL DATA MANAGEMENT	NURSING CARE SOLUTIONS	PARTICLE BEAM THERAPY
PRECISION MEDICINE	PHARMACEUTICAL SOLUTIONS	BIOINFORMATICS
TESTING AND DIAGNOSTIC SYSTEMS	MEDICAL IMAGING AND DISTRIBUTION SYSTEMS	

Hitachi Social Innovation Investment

PLATFORM AND TECHNOLOGY

Pentaho
Common analytics framework
Converged Platforms
Content Platform

HITACHI PATENTS

A leader in data analytics
technology patents

2800+ pending

PEOPLE AND EXPERIENCE

$1B
IoT and big data R&D

500+ data scientists and
researchers worldwide

600+ employees in the
Social Innovation Business Unit

75+ resources in Hitachi Live Insight
Center of Excellence

Hitachi R&D

Robotics

Traffic Analysis

Image Search
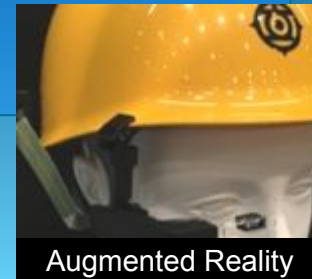
Power Grid Optimization
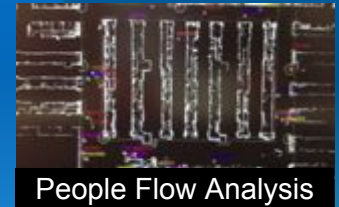
Human Behavior Analysis

Oil Field Analysis

People Flow Analysis

2100+ Labs Researchers Worldwide

Object Recognition

Mining Optimization

Facial Recognition

Augmented Reality

Delivering Real IoT Solutions

SMART CITY
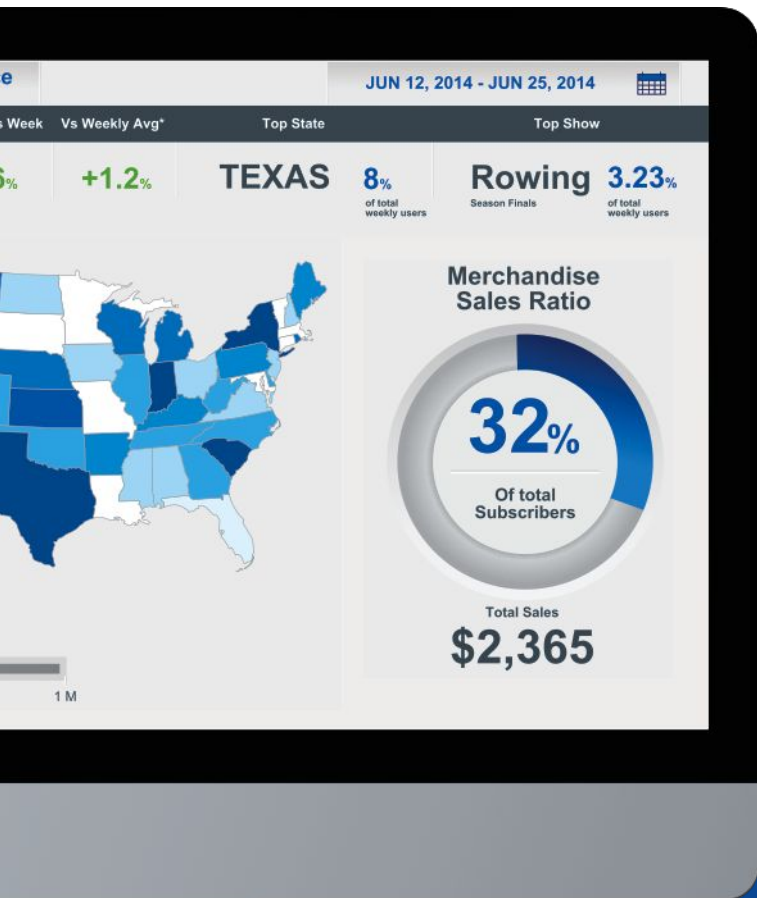
TELECOM

HEALTHCARE

I.T. ANALYTICS

ENERGY

INTELLIGENT VEHICLE

PENTAHO

IoT ANALYTICS PLATFORM

Safe Harbor Statement

- The forward-looking statements contained in this document represent an outline of our current intended product direction. It is provided for information purposes only and is not a commitment to deliver any new or enhanced product or functionality, or that we will pursue the product direction described. Facts and circumstances may occur which may impact current plans, resulting in changes to the information in this presentation. This information is current only as of the date it is made and should not be relied upon in making purchasing decisions. The development, release (if at all), and timing of any features or functionality described for the Pentaho products remains at the sole discretion of Pentaho.

Point of View: Pentaho Data Integration

The challenge:
- More users have access and need access to data blending
- Security, governance and audit are recognized as critical for ALL Data
- Data from the lake needs to be blended with the EDW and external sources
- You can't always know what the data will look like until it arrives

The direction:
- Continue work on security & governance integration
- Improve ease of use and performance of data blending
- Bring data blending closer to the user
- Extend existing capabilities for dynamic blending

Roadmap

Continue security, governance & operations management

2015	▪ Kerberos support ▪ Lineage	▪ SNMP monitoring ▪ OPS Mart improvements
2016	▪ Flexibility in delegation of authority and action via Kerberos ▪ Focus on audit, logging, debugging, resource management	
Futures	▪ Interoperability with security & governance providers	

Improve Ease of Use and Performance of Data Blending

2015
- Sqoop step updates
- Spark executor step
- ElasticSearch updates

2016
- JSON and HBASE step updates for MapReduce
- Investigating broader support for emerging SQL on Hadoop technologies – Presto, Phoenix, Spark-SQL…

Futures
- Constant investigation into opportunities for performance and scaling
- Continue labs investigation of emerging Big Data technologies and productize as appropriate

Roadmap

Bring Powerful Big Data Blending Closer to the Analyst

2015	
	▪ Annotations stream step ▪ Auto-model build step ▪ Data Services

2016	
	▪ Easy to create prompts for dynamic transformations ▪ Investigations into faceted search workflows ▪ Best practices and samples for SDR use cases

Futures	
	▪ Investigation into web-based interface for blending & analytics together

Roadmap

Extend existing capabilities for dynamic & model driven transformations

2015
- Dynamic Blending (Metadata injection) exists and supports a variety of IOT, dynamic and model driven uses cases

2016
- Extend support & documentation of metadata injection
- Investigate IOT device and machine log formats to add

Futures
- Continue documentation and best practices for dynamic transformation with metadata injection
- Investigate input steps to add in support of IOT use case

Point of View: Pentaho Platform

The Challenge:
- Deployment automation & configuration management are increasingly popular – but there are many technologies at play
- Security, governance and audit are increasing concerns
- Use of analytics is exploding & increasing demand for easier application maintenance

The Direction:
- Improve deployment, configuration & upgrade
- Continue expanding capabilities for governance & operational management
- Standardize & simplify common role based administration tasks from a central location

Improve deployment, configuration & upgrade

2015
- Simplification of configuration with Hadoop & Data Services

2016
- Platform deployment, configuration & upgrades on local & remote servers
- Lifecycle management/versioning

Futures
- Investigating tenant administration & configuration

Roadmap

Continue expanding capabilities for governance & operational management

2015	▪ SNMP Monitoring ▪ OPS Mart improvements ▪ Data Lineage in PDI
2016	▪ Investigating next steps on lineage ▪ Focus on audit, logging, debugging, resource management
Futures	▪ Investigating specific industry regulatory requirements

Standardize & simplify common role based administration tasks from a central location

| 2015 | ▪ REST APIs for automated user management |

| 2016 | ▪ **Phase 1:** Standardize & simplify common role-based administration tasks
▪ Investigate scheduling interoperability improvements |

| Futures | ▪ **Phase 2:** Centralize administration & management
▪ Investigating improvements to tenant administration & management |

Point of View : Embedded Analytics

The Challenge:

- The expectation of agile, tailored experiences is increasing
- Security, governance and audit are increasing concerns
- Data from the lake needs to be blended with the warehouse and other external sources

The Direction:

- Continue to expand and harden Analyzer APIs
- Invest in built-in capabilities
- Bring Big Data Blending closer to the user
- Extend and harden the Pentaho Platform Infrastructure

Roadmap

Expanding and hardening APIs

2015	▪ REST API for User Management
2016	▪ Visualization APIs ▪ Geo-mapping APIs
Futures	▪ Investigating componentization for lighter web embedding

Built-in Functionality

| 2015 | ▪ Inline modeling v1 |

| 2016 | ▪ Updates to visualizations
▪ Inline modeling for user defined groups and hierarchies
▪ Export formatting |

| Futures | ▪ Investigating technologies for higher performance analytics |

Roadmap

Bring powerful big data blending closer to the analyst

2015
- Annotations stream step
- Auto-model build step
- Data Services

2016
- Easy to create prompts for dynamic transformations
- Investigations into faceted search workflows

Futures
- Investigation into web-based interface for blending & analytics together

Extend and harden the Pentaho Platform Infrastructure

2015
- Spring 4 enables SSO and Third Party Authentication
- Java, Tomcat and JCR updates for security & performance

2016
- Lineage available to enterprise metadata management tools
- Phase 1: simplified administration & management

Futures
- Investigating improvements to tenant administration & management

C*Tools and overall Roadmap

- Will be provided by Pedro –

  – STAY TUNED!

General Community roadmap additions

- Near term -   Open the Annotations and Model build/publish steps

- New processes for community contributions
    - See Marc Batchelor's session later today

- High level agreement to open PIR

- Discussions with Hitachi and HDS about opening some of their tech
    - Candidates – streaming engine, search and content indexing, IoT platform components, etc.

Pentaho Commitment
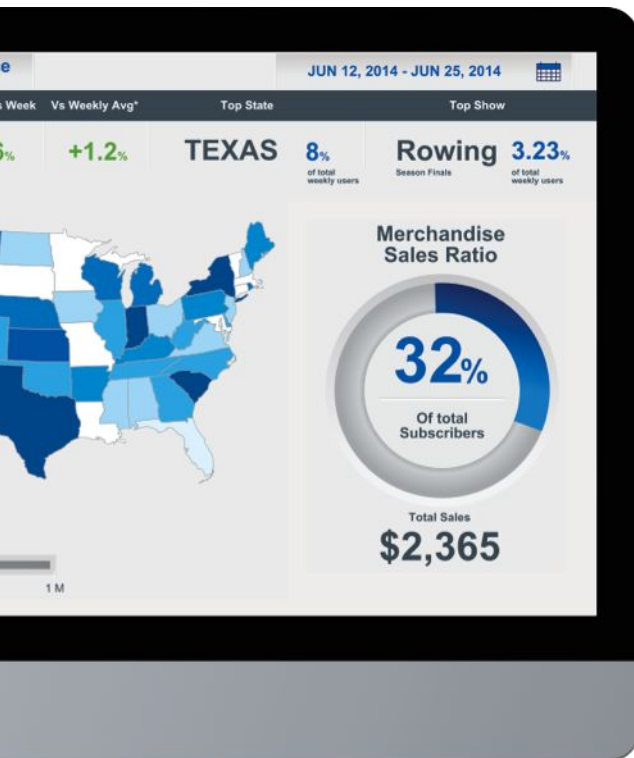
- **A future-proof data foundation**

- **Innovation in the big data ecosystem**

- **Enterprise class platform**

- **Become MORE open – with full Hitachi support**

- **Embrace our roots of Open Source**

THANK
YOU

pentaho®
A Hitachi Data Systems Company