

# Pentaho Big Data Test Drive Workshop

Hands-On Exercises To Solve Big Data Use Cases

## WORKSHOP SUMMARY

Many companies are turning to Hadoop, NoSQL and analytic databases to tackle challenges related to the increased variety, velocity and volume of information. The Pentaho Big Data Test Drive Workshop teaches you how to use Hitachi Vantara's Pentaho platform with these new big data technologies. You'll implement four high-impact big data use cases through a combination of instructor-led presentations, demonstrations and hands-on exercises. Using Pentaho, you'll extract, prepare and blend data from disparate sources in order to derive insights using visualizations and analytics.

## Typical Workshop Agenda

### Pentaho Overview and Discussion

You will review overall Pentaho platform and Hadoop architecture to set the foundation for a Pentaho-based big data solution. The session covers blueprints on how to use the Pentaho platform to address four big data use cases with platform components including Pentaho Data Integration (PDI) to integrate big data sources. Use PDI's analyzer feature to analyze big data with Impala, and its report designer and dashboard designer features to build and design reports and dashboards.

## Test Drive

**You will gain hands-on experience for the following four big data use cases.**

**Fill the Data Lake: Use Pentaho to onboard new data sources into Hadoop.** As organizations scale data onboarding from just a few sources to hundreds or more, data engineering time and resources can be monopolized. The process typically involves creating hundreds of hard-coded data movement procedures in a practice that is often highly manual and error-prone. Practice ways to fill the data lake by ingesting multiple data sources. Simplify onboarding this data with the Pentaho platform's proprietary metadata injection methodology.

**Create a Data Refinery: Use Pentaho to process data at scale in Hadoop.** Create a data refinery using MapReduce, Impala and Spark to streamline the data processing and delivery. The data refinery becomes the landing and processing zone for data from many diverse sources before it is pushed downstream to an analytic database for rapid queries. When this is done, extract, transform and load (ETL) and data management cost savings are scaled up, and big data becomes an essential part of the analytics process. Engineer data on Hadoop and Spark by processing, blending and aggregating data for the business.

### Workshop Requirements and Skills

Overall vision of potential solution at a high level

Experience in database design and development, SQL programming and reporting

Prior experience with ETL and reporting tools recommended

Basic understanding of different big data technologies, such as MapReduce, Spark, Kafka, PostGres and HBase recommended

"Much more hands-on than expected. Really good considering this is just a one-day workshop! Met all my expectations and more."

– Tom Simon, Capgemini

**Self-Service Data Preparation: Use Pentaho to prepare data for the business using Impala.** Too many analysts are stuck with data preparation processes that rely on coding or scripting by a data engineer. Let teams do more with existing resources and skills by empowering a broad set of analysts to prepare the data they need in a self service fashion without waiting on IT — but within the boundaries defined by IT. Prepare data with Impala to blend, bin and visualize data from multiple sources.

**Self-Service Analytics: Use Pentaho to analyze and report a 360° view with HBase, RDBMS and Impala.** The 360° view blends a variety of operational and transactional data sources to create an on-demand analytical view across customer touch points. It also includes providing customer-facing employees and partners with information, which is made available inside everyday line-of-business applications. Develop a 360° view using Impala on a variety of data from HBase and RDBMS and visualize within the Pentaho platform.

## Benefits

- Experience governed data delivery with reference architectures and best practices.
- Build an end-to-end data processing and visualization application with Pentaho, PostgreSQL and Hadoop.
- Learn four high-value big data use cases: Fill the Data Lake, Create a Data Refinery, Self-Service Data Preparation and Self-Service Analytics.

## Duration

- One day.

## Questions This Workshop Answers

- What are the popular high-impact use cases available out there?
- What big data technologies are utilized in the solution? Do I have to possess them?
- How does the Pentaho platform fit into the solution?
- What training do I need for the project team?
- How can Hitachi Vantara help make me successful?

## Customize

Do you want to have a workshop tailored to your use cases and hosted at your location? Hitachi Vantara offers tailored workshops for organizations looking to tailor this group workshop for their specific use cases and requirements. We provide extensive upfront scoping and planning and can even use a sample of your approved data in the design of your workshop.

### About Hitachi Vantara

Hitachi Vantara, a wholly owned subsidiary of Hitachi, Ltd., helps data-driven leaders find and use the value in their data to innovate intelligently and reach outcomes that matter for business and society. We combine technology, intellectual property and industry knowledge to deliver data-managing solutions that help enterprises improve their customers' experiences, develop new revenue streams, and lower business costs. Only Hitachi Vantara elevates your innovation advantage by combining deep information technology (IT), operational technology (OT) and domain expertise. We work with organizations everywhere to drive data to meaningful outcomes. Visit us at [HitachiVantara.com](http://HitachiVantara.com).

### Who Should Attend?

Technical architects

Technology leaders and stakeholders

Developers

Project team leaders and members

“Great to be hands-on with the technology! More ‘impartial’ than other vendors.”

– Toby Sykes, Deloitte

“Great to have various people (consultants) to talk to who are very knowledgeable and showed great expertise. I found the session very useful.”

– Olivier Sarma, Specsavers

“It was a very useful one-day quick start on big data.”

– Suresh Kurnootala, NFCU

## Hitachi Vantara



Corporate Headquarters  
2535 Augustine Drive  
Santa Clara, CA 95054 USA  
[HitachiVantara.com](http://HitachiVantara.com) | [community.HitachiVantara.com](http://community.HitachiVantara.com)

Contact Information  
USA: 1-800-446-0744  
Global: 1-858-547-4526  
[HitachiVantara.com/contact](http://HitachiVantara.com/contact)

HITACHI is a trademark or registered trademark of Hitachi, Ltd. Pentaho is a trademark or registered trademark of Hitachi Vantara Corporation. All other trademarks, service marks, and company names are properties of their respective owners.

P-010-B BTD June 2019