





Pós Graduação em Business Intelligence

Integração de Dados – ETL

Encontro 01: Dinâmica do Curso, Cronograma, Conceitos ETL, Introdução ao Pentaho, Instalação Pentaho, PostgreSQL e Repositório Pentaho





Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.

- Clive Humby

CURSO:

73114 - Especialização em Business Intelligence

DISCIPLINA:

HP192 - Integração de Dados ETL (36ha)

TURMA:

RMTBIB-2022/4

GRUPO WHATSAPP:

<https://chat.whatsapp.com/GwHL7sLCmwfLeRNggAQqLh>

PERÍODO:

De 11/05 a 13/06/2023, às 3as. e 5as.feiras, das 19:15 às 22:45

LINK PARA OS ENCONTROS:

<meet.google.com/wry-onin-xnh>

Assitam às aulas síncronas. Embora o conteúdo seja gravado, a interação em tempo real irá ajudar na resolução de dúvidas.

Evitem distrações como serviços de **mensageria**, **web surfing**, e conversas paralelas. Whatsapp só nos intervalos!

Procurem ampliar o conhecimento com conteúdo **além** do exposto em sala.

Será utilizada apenas a plataforma Windows para este curso.

Henrique Gomes da Silva
Oracle ACS Solution Architect



**Entrem em contato por
email ou whatsapp:
61 - 998690018
hgsilva@gmail.com**



Qual a dinâmica das aulas? (faltas, exercícios, avisos, horários, comprometimento, etc)

Data	Conteúdo
11/05/2023	Encontro 01: Dinâmica do Curso, Cronograma, Conceitos ETL, Introdução ao Pentaho, Instalações Softwares
16/05/2023	Encontro 02: Steps de Input, Output e Transformação de Dados
18/05/2023	Encontro 03: Steps Junção de Dados, Lookups e DataWarehouse
23/05/2023	Encontro 04: Scripts e Jobs
25/05/2023	Encontro 05: Projetos
30/05/2023	Encontro 06: Cargas Stage e DataWarehouse
01/06/2023	Encontro 07: Jobs de Projeto e Automação ETL
06/06/2023	Encontro 08: Grandes volumes e Performance; Revisão de Conteúdo
13/06/2023	Encontro 09: Avaliação Final



Obs:

Material de apoio aos LABS estará disponível em: <https://github.com/hgsilva/posbi-etl>

INTRODUÇÃO AOS PROCESSOS DE TRANSFORMAÇÃO DE DADOS

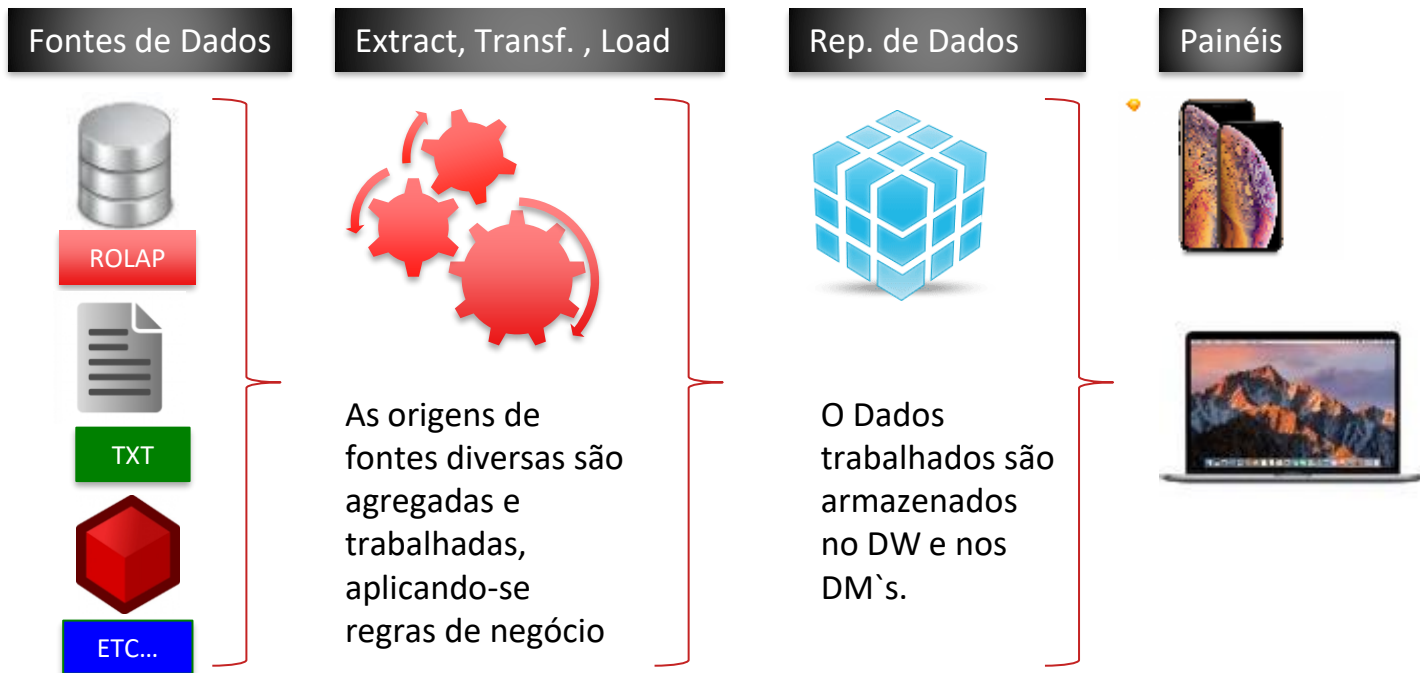


Processo de **movimentação** dos dados a partir de suas origens para seus destinos, aplicando-se as **regras definidas pelo negócio**, além de técnicas de limpeza, padronização, manutenção e por fim, entregando estes **dados tratados** para os **sistemas consumidores**.

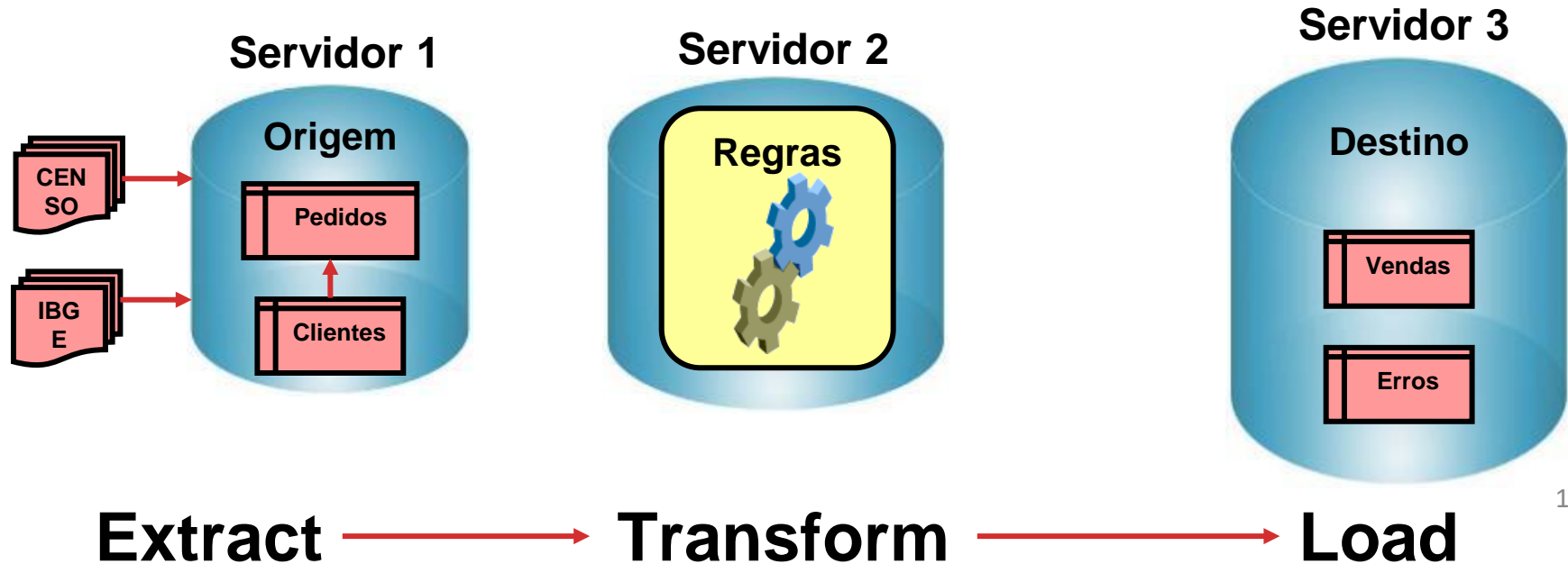
- Corrige erros e dados faltantes;
- Agrega novas origens;
- Entrega dados com confiança;
- Padroniza dados de múltiplas origens;
- Estrutura dados para serem utilizados por ferramentas de usuário.

Não são utilizados apenas para carregar DW ou DM!!

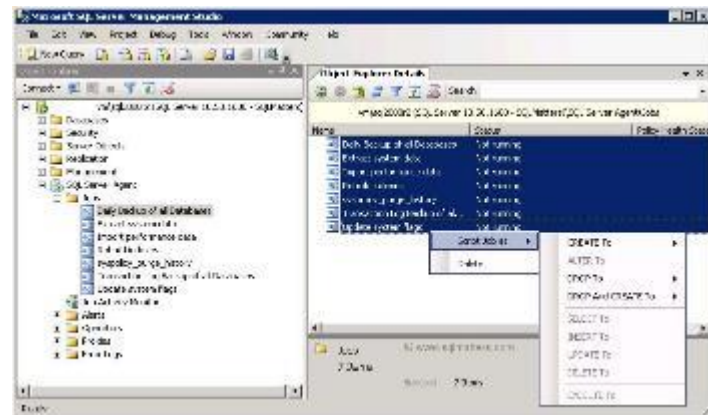
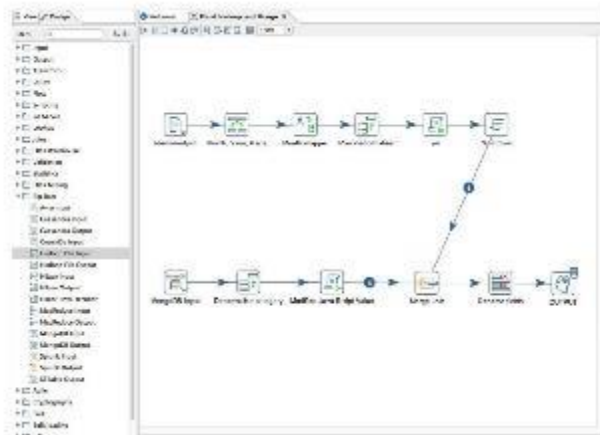
- Outras utilidades:
 - Transferência de dados entre gerenciadores de diferentes tecnologias (Postgre -> Mysql ou HIVE -> Oracle);
 - Rotinas de Backup;
 - Preparação de dados;
 - Qualidade de dados;
 - Aposentadoria de dados (dbms -> fita);
 - Integração entre sistemas.



Processo de Integração: Extract - Transform (check) - Load



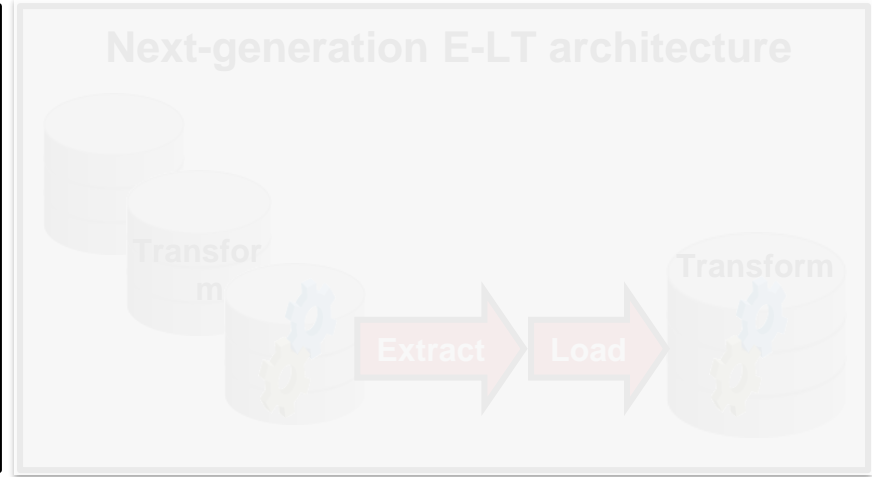
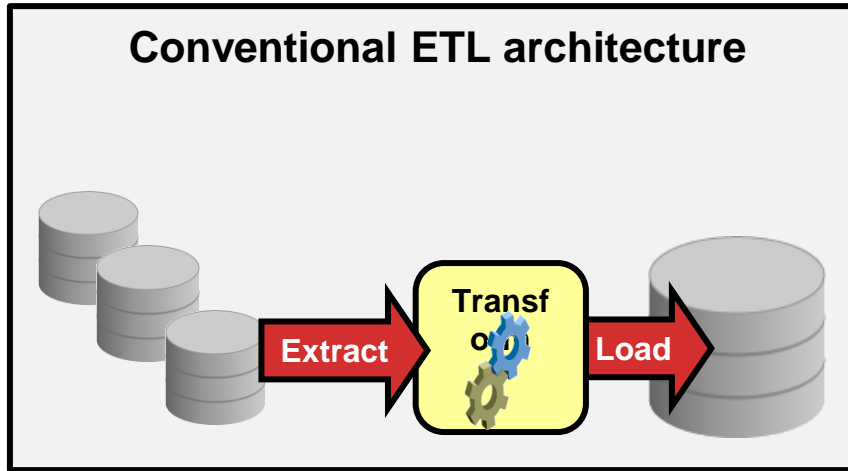
12



- Blocos pré codificados
- Menor necessidade de conhecimento de banco de dados
- Manutenção mais fácil
- Maior facilidade de se compreender fluxos de carga

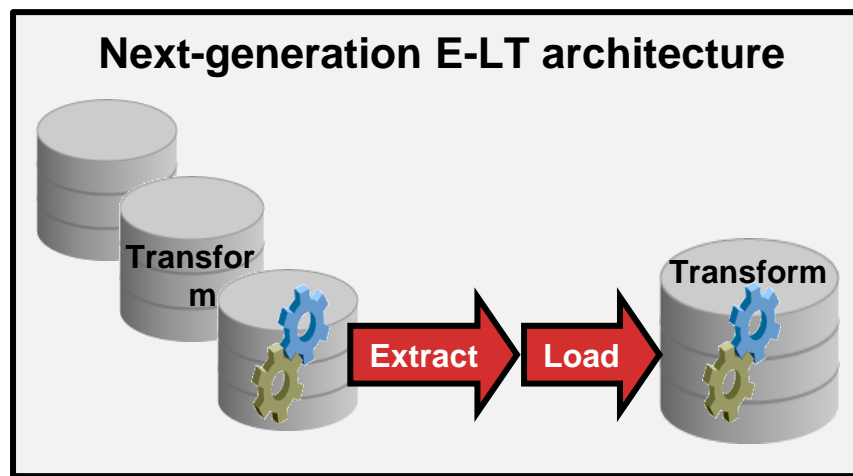
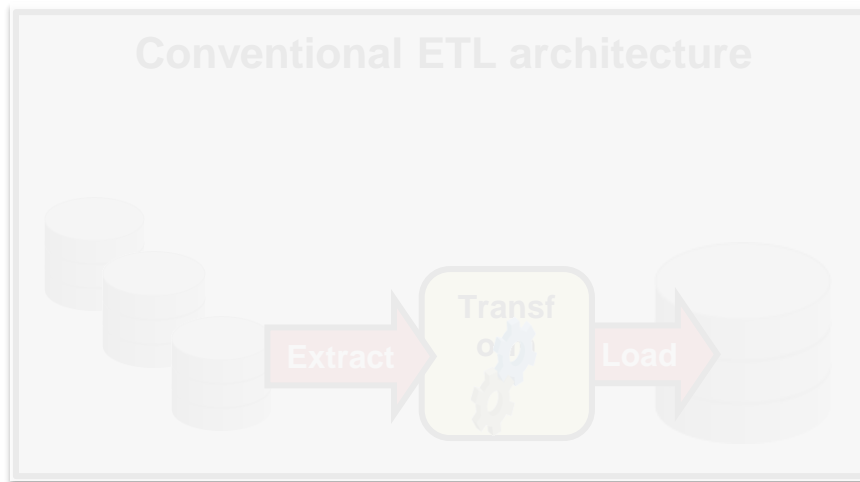
ETL

1. **Extract:** Buscar dados em diversas origens
2. **Transform:** Processar transformações
3. **Load:** Carregar no destino



E-LT

1. **Extract:** Buscar dados em diversas origens
2. **Load:** Levar os dados para o Destino
3. **Transform:** No destino, transformar os dados





Extract → **Load and Transform** ¹⁶

Usuários

Painéis e Dashboards

DBA
Desenv. OLAP
Desenv. ETL
Analista Requisitos
Analista Testes

Gestão do DW
Construção de Painéis
Construção ETL
Levantamento de
Requisitos *
Testes de Navegação
Documentação

70% do tempo de projeto
é gasto em ETL



Pentaho Data Integration



Heritage of innovation

119,000 global patents
11,000 IoT patents
\$2.9B annual R&D*
\$927M 3-year IoT R&D*

\$87.4B
consolidated
revenue*

Thomson Reuters
Top 100 Global
Innovators

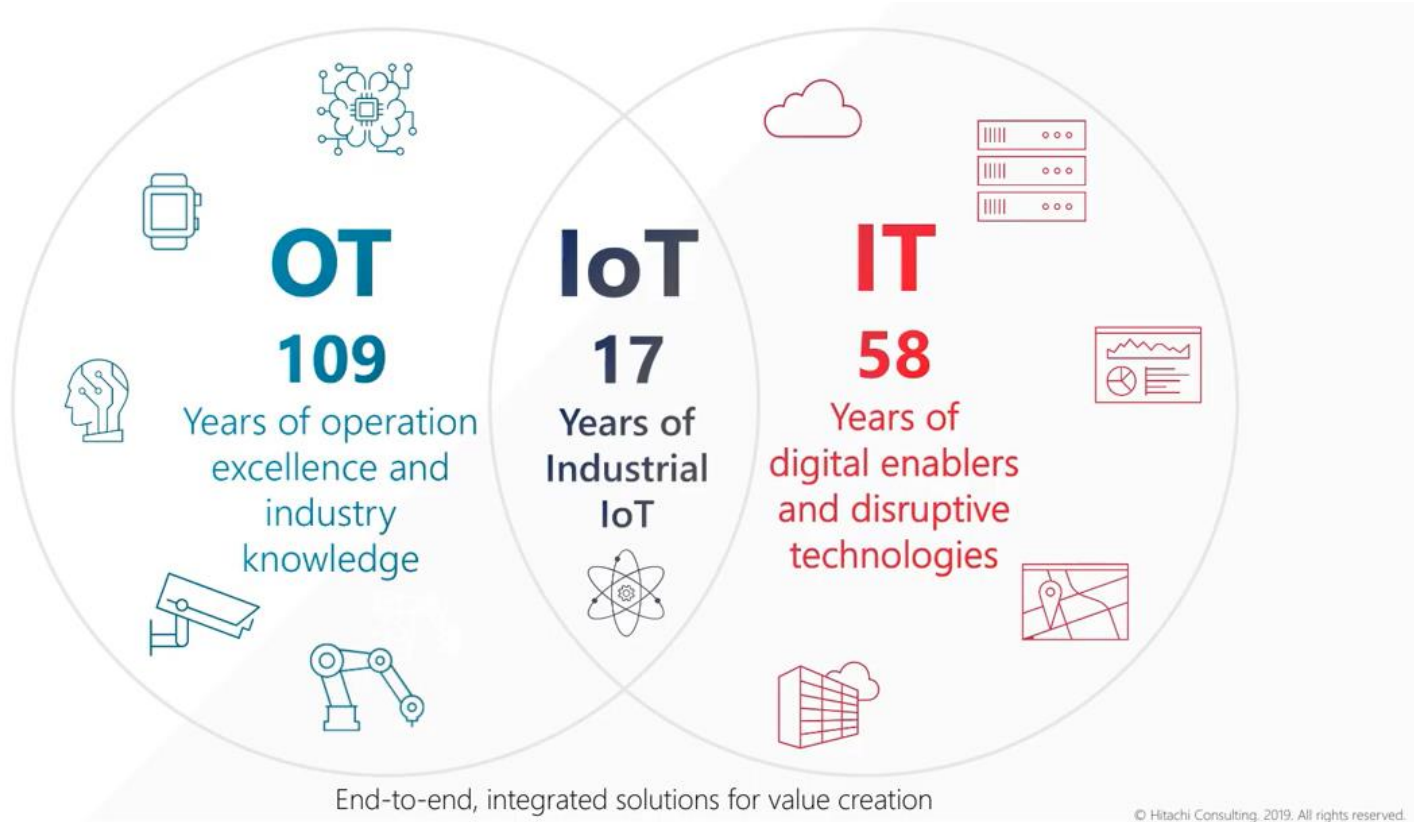
**FORTUNE
500**
Top 10 global
technology
company by
revenue

150+ factories
16 labs

operating in
70+
countries

295,900+
employees

*As of June 2019



© Hitachi Consulting. 2019. All rights reserved.

Hitachi Core Businesses

Mobility
Smart Life
Industrial
Energy
Information Technology

Hitachi Group Companies



© Hitachi Vantara Corporation 2019. All Rights Reserved.



HITACHI
Inspire the Next

Hitachi Vantara

**THE NEWEST DATA COMPANY WITH
MORE THAN 100 YEARS OF EXPERIENCE**

- Informatica PowerCenter
- IBM Data Manager
- SAP DI
- SAS DI
- Oracle DI
- Microsoft Integration Services
- Pentaho Data Integrator

Hitachi Vantara



- Plataforma de Business Analytics
 - Integração de Dados e ETL
 - OLAP
 - Relatórios
 - Dashboards
 - Data Mining
- Desenvolvida em Java
- Licenças Enterprise e Community
- Atualmente, versão 9.0



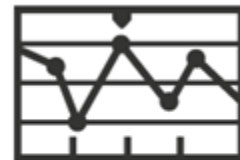
Business Analytics



Internet of Things Analytics

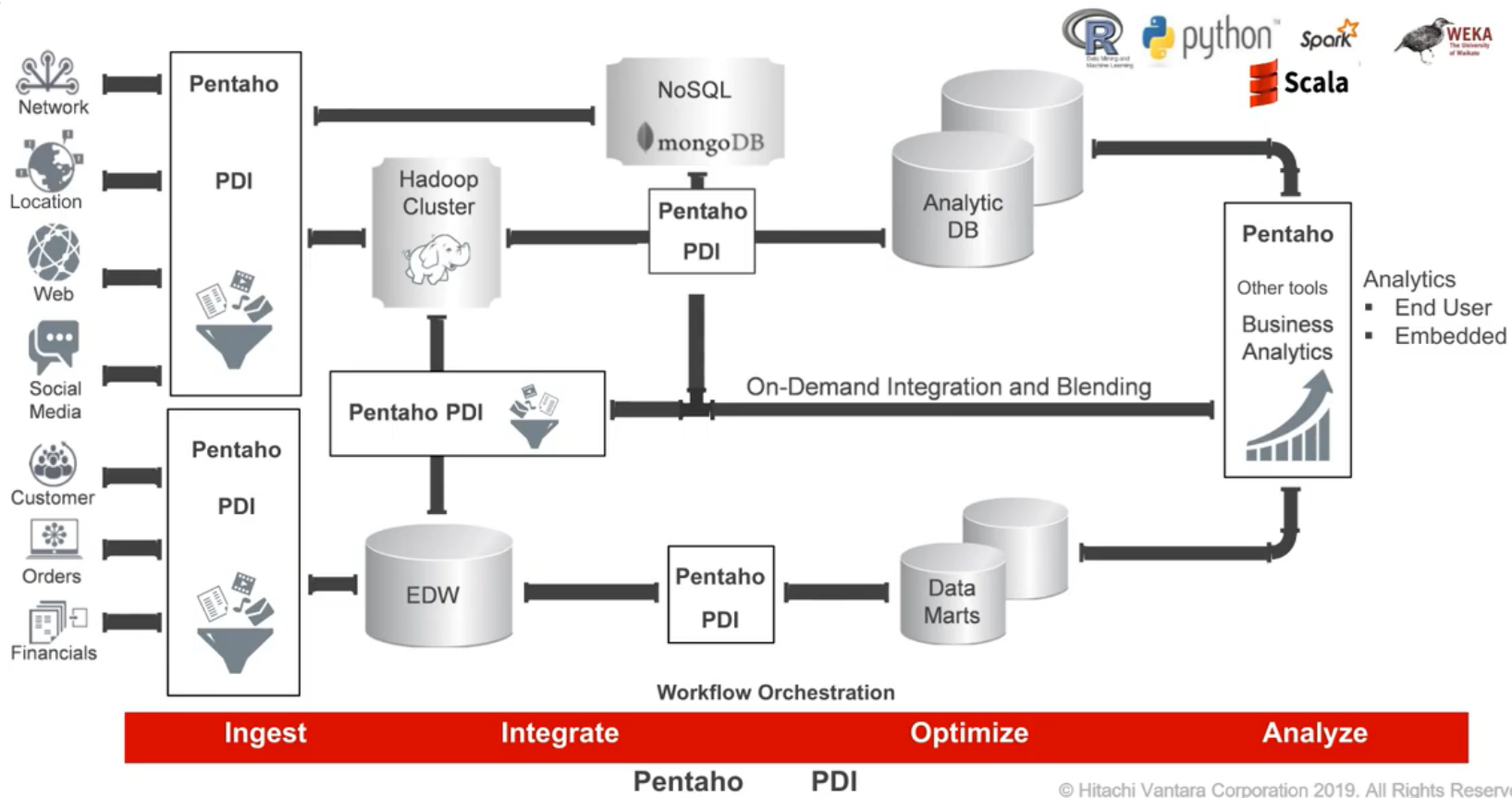


Pentaho Data Integration



Big Data Integration and Analytics

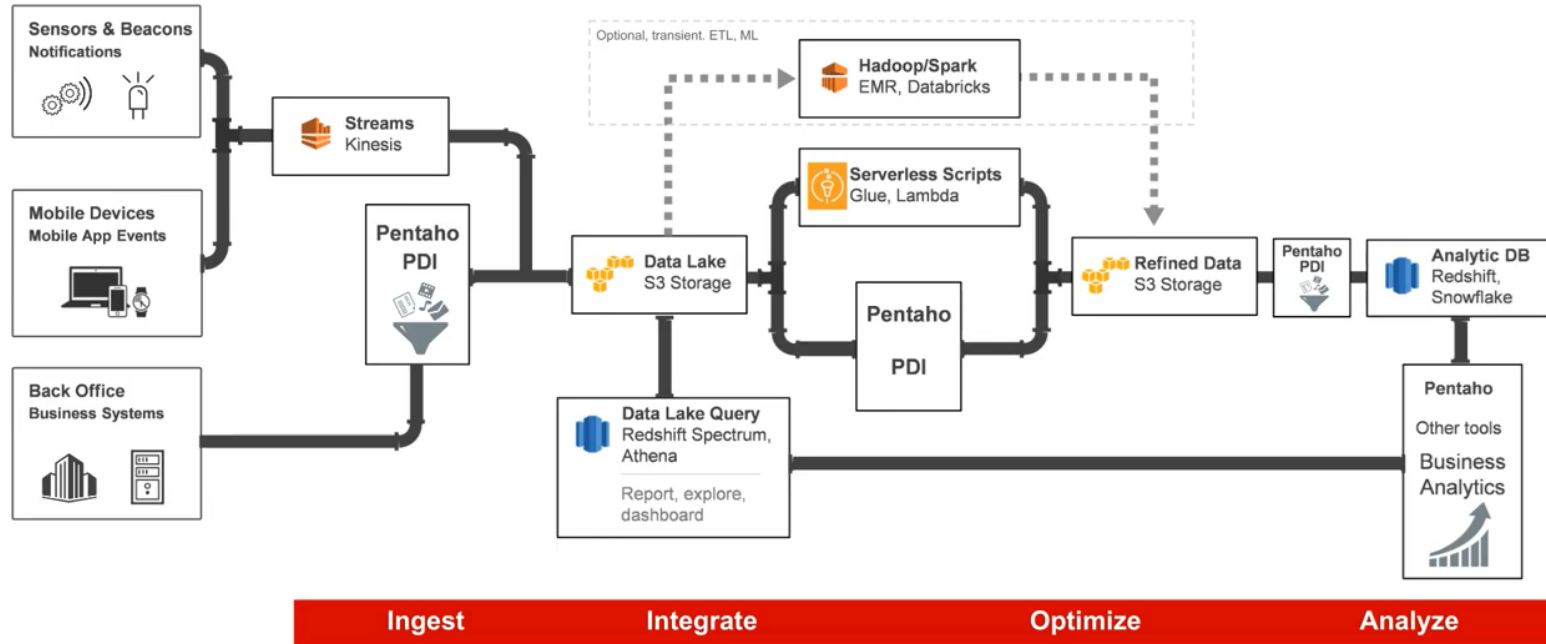
- Comprada em 2015 pela Hitachi Data Systems, posteriormente chamada de **Hitachi Vantara**, consolidando 3 empresas: Pentaho, Data Systems e Hitachi Insight Group



Cloud Analytics Architecture

AWS Cloud Example

HITACHI
Inspire the Next



© Hitachi Vantara Corporation 2019. All Rights Reserved.



End-to-End Analytics Platform

Data Integration

Data Orchestration

Big Data

Data
Warehouse

Other

Business Analytics

Embedded Analytics

Analyzer

Dashboards

Reports

Open



Pluggable



100% Java

Scalability

Monitoring

Data Lineage

Connectivity

Security

Interoperability

Deployment

Demonstração Pentaho

Exercícios Pentaho

Estes dois laboratórios tem por objetivos:

- 1. LAB00** - Entender o processo de instalação do Pentaho Data Integrator e seus softwares auxiliares, para que seja possível ao aluno reproduzir de forma independente o ambiente deste curso (ou outros ambientes com outros propósitos).
- 1. LAB01** - Criar o repositório de metadados, simulando em uma máquina local uma instalação com elementos de ambiente de produção.

OBRIGADO!

Henrique.gomes@iesb.br







Finalizando