



# Digital Disruption...

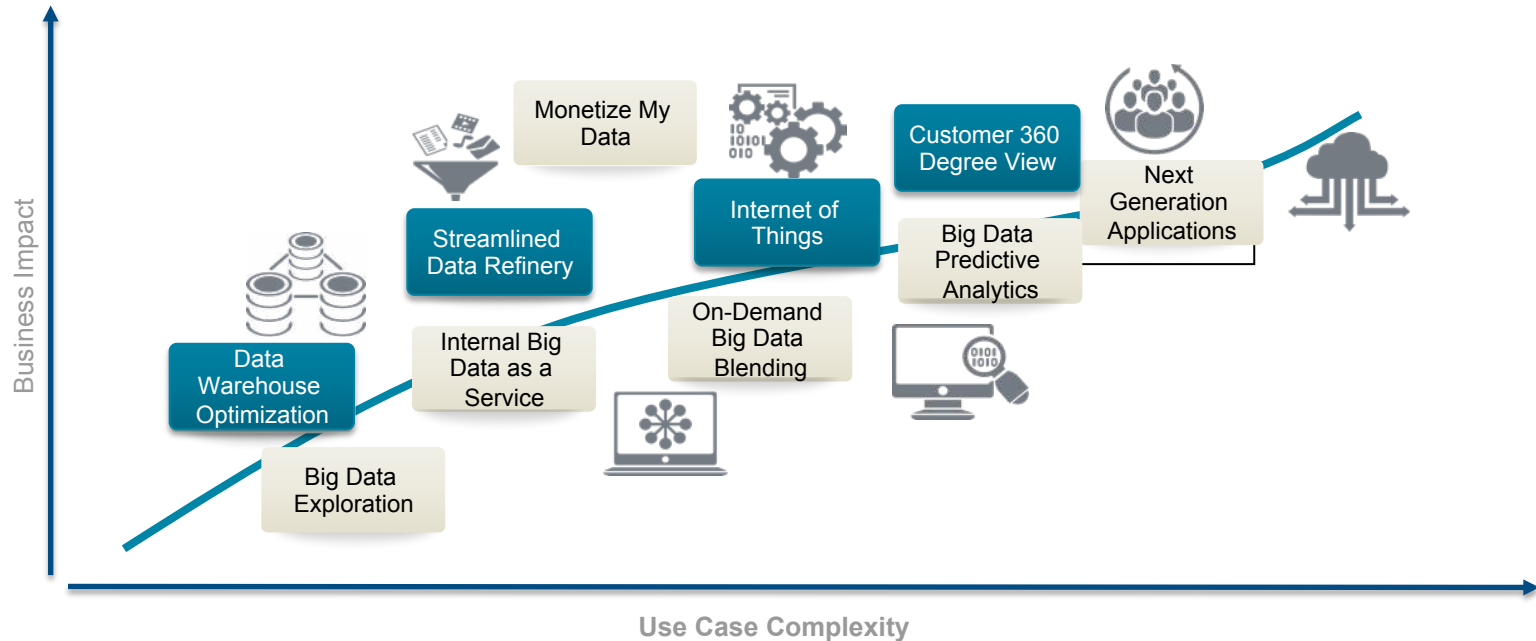
## Are you prepared?

Daniel Schafer, AE  
Pentaho  
713-480-6608  
[dschafer@pentaho.com](mailto:dschafer@pentaho.com)

- The Worlds Largest Retailer - Has No Inventory
  - The Largest Movie House - Has No Cinemas
  - The Most Popular Media Company in the World - Creates No Content
  - The Largest Company in the World That Rents Rooms to You and I - Doesn't Own Any Property
  - The Largest Phone Company in the World - Doesn't Own any Equipment, Switches or Even Land Lines
  - The Largest Transportation Company in the World - Doesn't Own a Single Vehicle
- Alibaba
  - NetFlix
  - Facebook
  - Air BnB
  - Skype
  - Uber

# A Spectrum of Big Data Use Cases

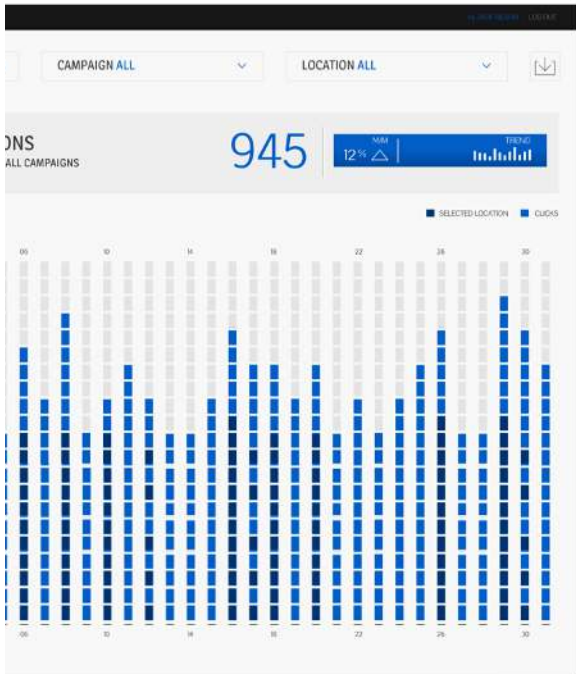
What The Market Is Deploying Today And Planning For Tomorrow



- ① New skills necessary
- ② High effort and risk
- ③ Continuous change

"Through 2018, 70% of Hadoop deployments will not meet cost savings and revenue generation objectives due to skills and integration challenges." -- Gartner<sup>1</sup>

1) Gartner Analyst, Nick Heudecker; infoworld.com, Sept 2015



## Modern, cohesive data integration and business analytics platform (Established in 2004)

- Full spectrum of advanced analytics for all key roles  
Embeddable, cloud-ready analytics
- Big data blending for analytics in real-time environments
- Broadest and deepest big data integration

## Critical mass achieved

- Over 2,000 commercial customers
- Over 20,000 production deployments
- Acquired in 2015 by Hitachi Group



**HITACHI**  
Inspire the Next

Social Innovation, IoT, Smart Cities, & Vertical Solutions

## Hitachi Data Systems

Turnkey BI and Big Data Solutions



Embedded Solutions for  
Enterprise Data Governance and SaaS

Big Data – since 2009

Traditional DI & BI  
Since 2004



- SSO & Java Spring
- DB per Group/Tenant
- Row-level Multi-tenancy
- Object Multi-tenancy
- UI Multi-tenancy
- Scale-out Architecture

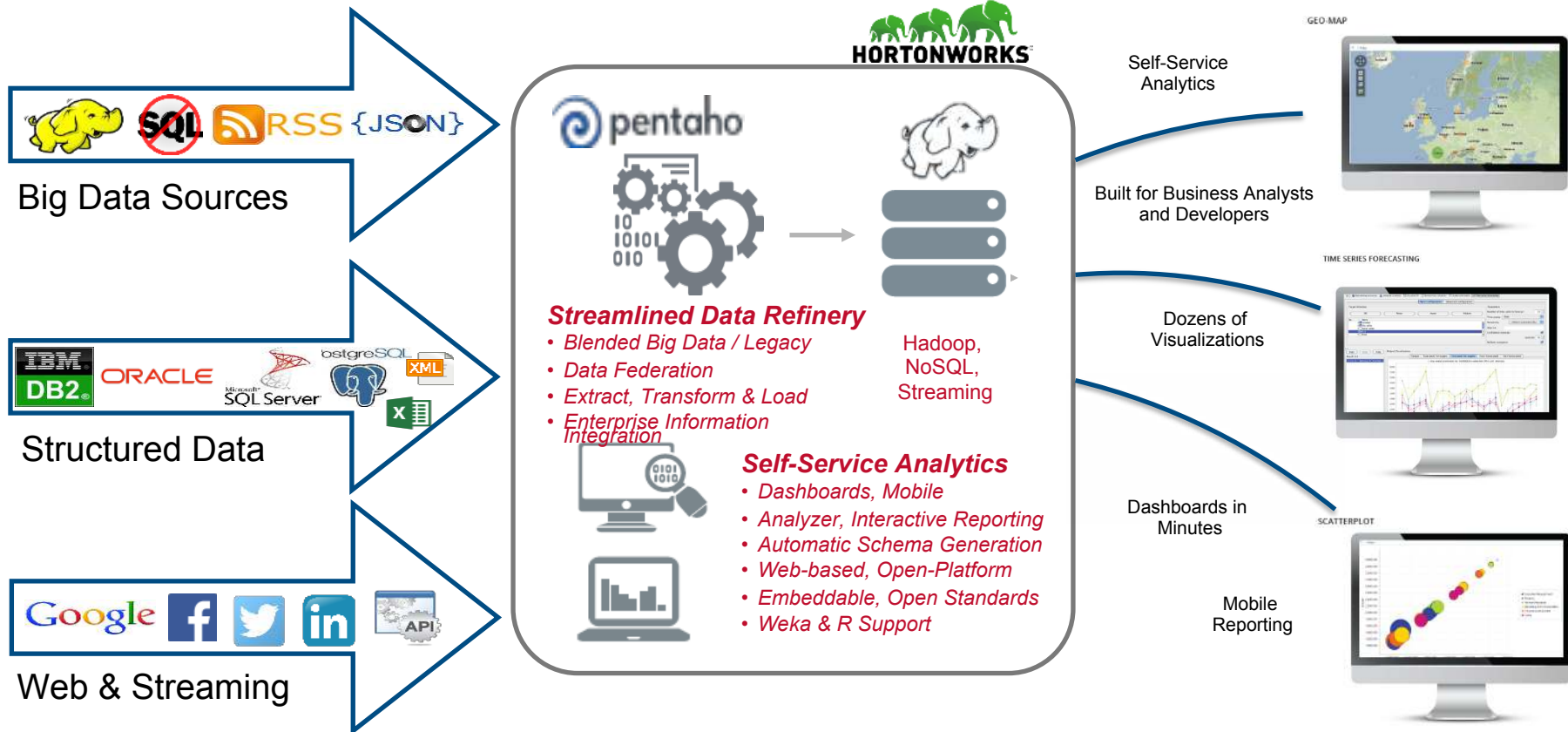
**CLOUD**

**Unified Compute Platform (UCP)**

**Hyper-Scale-out Platform (HSP)**



# End-to-End, Self-Service Data Preparation



# Why Customer's Use Pentaho on Hortonworks pentaho<sup>®</sup> A Hitachi Group Company

1. Metadata Injection. By utilizing dynamic vs. static bindings, we can reduce the number of data transformations by 80-90 percent. [YouTube Demo](#) by [Matt Casters](#), Chief Architect of Data Integration and Kettle Project Founder at Pentaho
2. Visual MapReduce. Graphically build Hadoop data transformations without coding. This enables you to reduce your Hadoop development time by over 85 percent. Additionally, Pentaho automatically manages the deployment and execution of Hadoop transformations with YARN. [YouTube Demo](#) by [Doug Moran](#), Product Manager for Big Data Technologies and Co-Founder of Pentaho
3. Embedded Analytics. “White label” our reports, visualizations and dashboards directly into your web applications. Use Java and REST APIs to access Pentaho data transformations and reports. [YouTube Demo](#) by [Anthony de Shazor](#), SVP of Customer Care and Principal Architect
4. Beyond ETL. Pentaho supports Enterprise Information Integration (EII), also known as data federation. Now you can create ETL jobs that blend data from structured and big data sources and invoke it via JDBC with [Teiid](#)
5. Weka Scoring, Forecasting and R script execution with our [Data Science Pack](#). This package helps Data Scientists reduce data preparation times by 60-80 percent
6. [Deep Integration](#) with your Hortonworks Hadoop ecosystem. Pentaho offers real control over YARN jobs, Spark execution, Oozie, Sqoop and more



## ■ Business Case

- Provide value for faster and better decision making – ranging from cost optimization, increasing ROI, personnel safety, automated drilling

## ■ Challenges

- IoT applications can get very complex very quickly due to the extensive breadth and diversity of data sources and standards
- Real time support from existing and emerging data sources

## ■ Benefits

- Improved pump safety and prevents spills by predicting failure rates
- 60-80% less cost
- 2x to 4x faster development



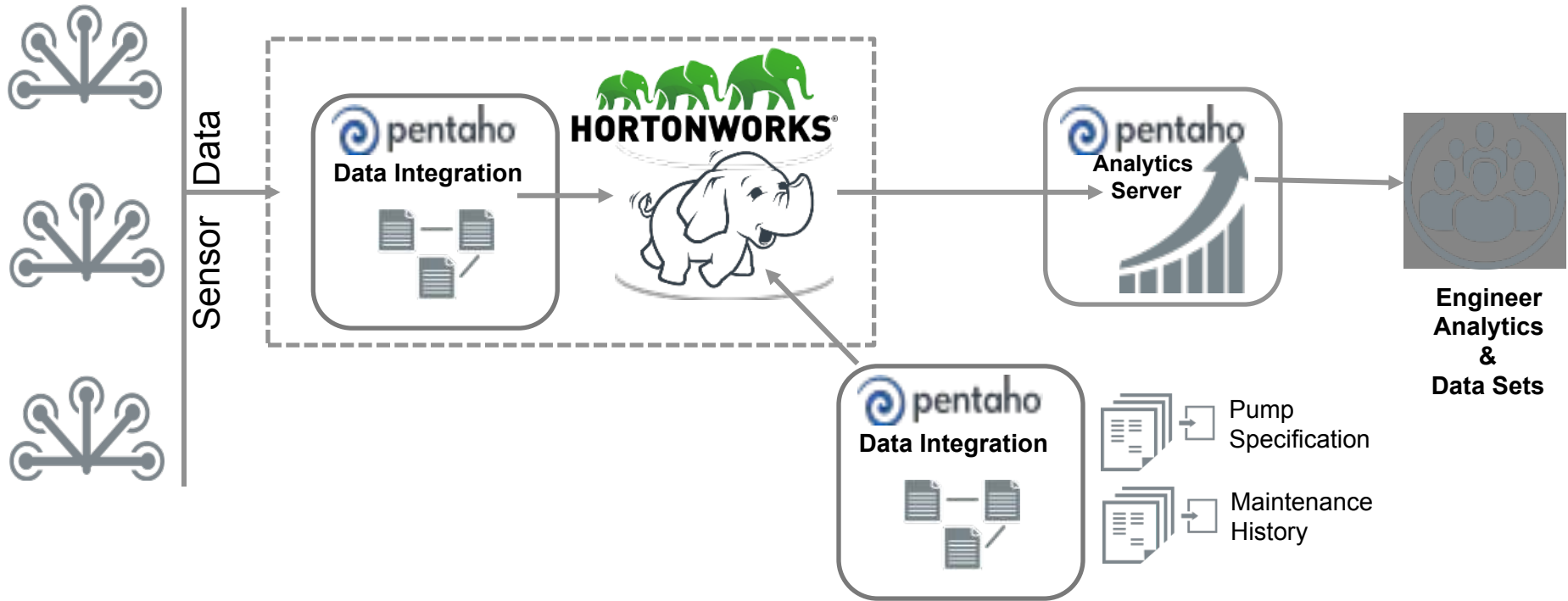
**HALLIBURTON**

Landmark



## Why Pentaho?

- Seamless integration and connectivity to all data sources
- Embedded business analytics for fast time to market
- Enterprise class platform that provides data-driven insight



## Big Data is Hard. Pentaho makes it easier.

```

// Java Implementation of Bubble Sort (Eloquent Code)
public static void bubbleSort(int[] data)
{
    boolean isSorted;
    int temp;
    int numberOfTimesLooped = 0;
    do
    {
        isSorted = true;
        for (int i = 1; i < data.length - numberOfTimesLooped; i++)
        {
            if (data[i] > data[i + 1])
            {
                temp = data[i];
                data[i] = data[i + 1];
                data[i + 1] = temp;
                isSorted = false;
            }
        }
        numberOfTimesLooped++;
    } while (!isSorted);
}
    
```

```

// Perl Implementation of bubble sort (unoptimized code)
sub swap {
    ($J($1, $2) = ($J($2, $1)));
}

sub bubble_sort {
    for ($i = 0; $i < $#_ @_; $i++) {
        for ($j = $i + 1; $j < $#_ @_; $j++) {
            swap($i, $j) if ($_[ $i ] > $_[ $j ]);
        }
    }
}
    
```



### ETL Framework for Big Data



Traditional & Big Data Solutions

Key Big Data Solutions

Social Innovation, Vertical Solutions & IoT

We make it happen. Better.



**Questions ?**