



Congratulations, you have a data lake, and you even have some successes using it

Are you protecting Customer's data from hackers?  
If it's installed in your data lake you probably are not!!!

2016 was a record year for data breaches, the number of exposed records exceeded 4.2 billion, nearly four times the previous high.

- Step 1:** The Intelligent Catalog identifies Personally Identifiable Information (PII) and other sensitive data
- Step 2:** Files containing PII data are sequestered, and copies with encrypted columns are written to the data lake (A)
- Step 3:** The sequestered file is reviewed, encryption and masking is applied as directed (B)
- Step 4:** The sequestered file after release (B) replaces the surrogate encrypted file (A)
- Step 5:** Management statistics are collected and made available for managing the overall process

The digital economy depends on you protecting customer data as a key to conducting commerce. If you use a data lake and store customer information in the data lake, it is of paramount importance you prioritize the protection of PII data in your data lake. In this short video, we would like to present our approach to protecting PII data in your data lake from those intent on using your data for their benefit.



We developed a Personal Identifying Information (PII) detection and sequestering capability to augment our **one of a kind Hadoop** intelligent catalog because of recent issues reported by our partnering organizations and their concerns about the security of their Hadoop environments.

We believe it is imperative this capability lives in the big data environment, and that any information identified as being a candidate for sequestering be parked into an area with more strenuous security attributes than is normally associated with big data. We invite you to challenge our approach, we believe that you will be very happy to invest a small amount of your time to learn about this very critical capability.

Cyber-Security is no longer something that just your CIO worries about, it is a risk that deserves the attention of your risk organization, particularly if you market services through digital interfaces like the web or mobile platforms or have sensors like proximity beacons integrated into your network.

You do not want to learn that you were exposed in one of the national newspapers or through a visit from the Federal Trade Commission or Securities and Exchange Commission.

# BDR Architecture - Powered with Apache™ Hadoop®

( For the technicians, how we reside in the data lake )



- Teradata
- Databases (Oracle, DB2, SQL)
- MySQL
- CLOUD

Hadoop Components

Hive  
Spark  
Pig  
Drill  
Analytic Libraries  
Impala  
TIKA  
Hbase  
Spark Stream

API FRAMEWORK

KERBEROS SECURITY FRAMEWORK

WIZARD

Intelligent Catalog



- D3.JS Interactive GUI
- BigDataRevealed Callable Java Modules
- Map/Reduce, Spark, NLP, Deep Learning Externalized Callable Modules
- Departmentalized BDR-Apache-VMWare

Native Hadoop File System

EXECUTABLES

BDR

API FRAMEWORK

KERBEROS SECURITY FRAMEWORK

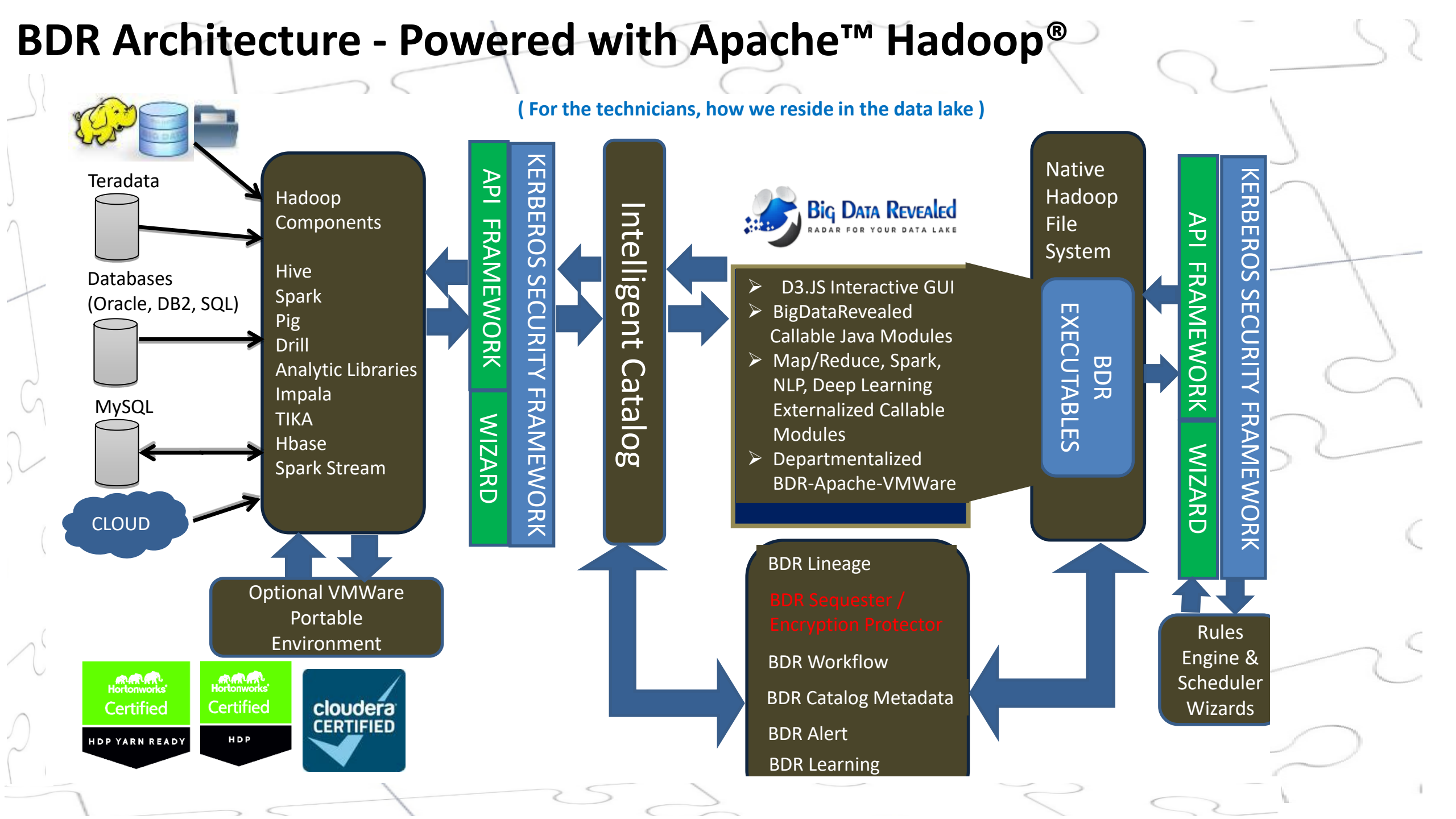
WIZARD

Rules Engine & Scheduler Wizards

Optional VMWare Portable Environment



- BDR Lineage
- BDR Sequester / Encryption Protector
- BDR Workflow
- BDR Catalog Metadata
- BDR Alert
- BDR Learning





**Big DATA REVEALED**  
RADAR FOR YOUR DATA LAKE



# BigDataRevealed Data Discovery for Big Data Hadoop

**BigDataRevealed Discovers and Isolates Personally Identifiable & Potentially Risky(Outlier/Anomaly) information/data in your Hadoop ecosystem!**

Steven Meister  
CTO & Founder, BigDataRevealed  
[steven.meister@bigdatarevealed.com](mailto:steven.meister@bigdatarevealed.com)  
847-791-7838  
[www.bigdatarevealed.com](http://www.bigdatarevealed.com)



Q & A

Thank You