PRODUCT LATEST NEWS
Sr. Sergio Rodríguez de Guzmán

**CTO**

PUE

www.pue.es
Hadoop & Why Cloudera

Sergio Rodríguez | Systems Engineer | sergio@pue.es
Industry-Leading Consulting and Training

PUE is the first Spanish Cloudera Silver Integrator
PUE is the only Training partner delivering classes across all Europe

Cloudera has trained over 40,000 people on Hadoop since 2009
Common BigData Early Problems in a Project

• Infrastructure investment
• Security and Compliance concerns
• Architecture sizing
• Wide and heterogeneous Hadoop ecosystem
• Support
• Ease of management
Best-In-Class Support

8.9 Overall satisfaction makes Cloudera the industry benchmark for support

95% Customers agree they benefit from Cloudera technical support outreach

#1 Ability to solve technical issues is the top reason to recommend Cloudera for Hadoop
Cloudera Platform
Cloudera Enterprise
Making Hadoop Fast, Easy, and Secure

Hadoop delivers:
• One place for unlimited data
• Unified, multi-framework data access

Cloudera delivers:
• Leading performance
• Easy system management
• Compliance-ready security
From Hadoop to an Enterprise Data Hub

Open Source Scalable Flexible Cost-Effective

Managed

Open Architecture

Secure and Governed

CLOUDERA’S ENTERPRISE DATA HUB

- BATCH PROCESSING (MapReduce)
- ANALYTIC SQL (Impala)
- SEARCH ENGINE (Solr)
- MACHINE LEARNING (Spark)
- STREAM PROCESSING (Spark Streaming)
- 3RD PARTY APPS

WORKLOAD MANAGEMENT (YARN)

STORAGE FOR ANY TYPE OF DATA
UNIFIED, ELASTIC, RESILIENT, SECURE (Sentry)

FILESYSTEM

ONLINE NOSQL
The Only Complete Hadoop Management Suite

Deliver optimum system utilization and meet SLA commitments.

Cloudera Manager

Focus on the solution, not the cluster, with the only complete, zero-downtime administration tool for Apache Hadoop.

Unique Capabilities:
- Unified configuration, management and monitoring across all services
- Online installation and upgrades
- Direct connection to Cloudera Support
- 3rd Party Extensibility
The Only Portable Cloud Experience for Hadoop

Maximize flexibility in Hadoop deployment architectures.

Cloudera Director

The first portable, self-service solution for deploying and managing enterprise-grade Hadoop in the Cloud.

Unique Capabilities:

• Dynamic cluster lifecycle management
• Cloud blueprints
• Multi-cluster health visibility
• Usage reporting for billing models
Why Cloudera is the Leader in Spark Support

- Integrated with other Cloudera Components – Cloudera Manager, Sentry, Navigator, etc.
- Cloudera more customers running Spark today than all our competitors combined. Installations range from a few nodes to 1000 node installs.
- Cloudera has been supporting Spark since early 2014 and first Hadoop vendor
- Between Cloudera and Intel, have over 20 developers working on Spark and 4 Committers
- The first and only Spark Training Class
Apache Kudu
Completes Hadoop's storage layer to enable fast analytics on fast data.

Data Model
Stores tables like relational databases

Low-latency Random Access
Live storage system which supports low-latency millisecond-scale access to individual rows

Built by and for Operators
Advanced in-process tracing capabilities, extensive metrics support, and even watchdog threads
The Only Hadoop Data Governance Solution

Enable compliance and maximize analyst productivity.

Cloudera Navigator

Minimize risk and maintain compliance with the only native end-to-end data governance solution for Apache Hadoop.

Unique Capabilities:

- Auditing
- Lineage
- Metadata Tagging and Discovery
- Lifecycle Management
Adaptive Data Model Management

Improve DBA productivity through continuous optimization.

Navigator Optimizer

Instantly understand data warehouse and Hadoop cluster usage, and drive optimizations to reduce cost and improve performance.

Unique Capabilities:
- Schema and workload profiling
- Data model discovery
- Optimization guidance
- Optimization automation (future)
The Only Comprehensively Secure Hadoop Platform

Meet compliance requirements and reduce risk exposure from storing sensitive data.

Cloudera is the leader in Hadoop security.

Unique Capabilities:
- Comprehensive and Unified
  - Secure at the core
- No Performance Impact
  - Jointly engineered with Intel
- Compliance-Ready
  - Only distribution to pass PCI audit

1. Perimeter Standards-based Authentication

2. Access Unified Role-based Authorization

3. Visibility Auditing & Governance

4. Data Encryption & Key Management

Security and Administration

Unlimited Storage
**Challenge:** All applications, databases, or file systems that have the potential to handle personal account-related data must undergo full PCI certification.

**Solution:** MasterCard’s Cloudera environment fully conforms to the PCI-DSS V 2.0 security standards so it can host PCI datasets and potentially integrate with other internal systems.

“Data privacy and protection is a top priority for MasterCard. As we maximize the most advanced technologies from partners and vendors, they must meet the rigorous security standards we’ve set. With Cloudera’s commitment to the same standards, we now have additional options in how we manage our data center.”

Gary VonderHaar
Chief Technology Officer,
Architecture
MasterCard
# Security and Governance

## Perimeter
Managing access to the cluster

### Access
Protecting access to data

### Visibility
Reporting on data access and lineage

### Data
Protecting data at rest or in transmission

## Cloudera
Unified, Compliance-Ready, Transparent

- **Kerberos with Cloudera Manager**
  - Automated, industry-standard authentication integrated with existing systems

- **Apache Sentry**
  - Working within the community to deliver centralized, granular RBAC across frameworks

- **Cloudera Navigator**
  - Transparent end-to-end data and metadata visibility, including column-level visibility in lineage and audit

## Competitors
Fragmented, Incomplete, Complex

- **Kerberos**
  - Manual configuration and integration

- **Hive ATZ-Ng, Ranger**
  - RBAC configuration silos, GUI “Band-Aid”

- **Falcon, Knox, Ranger, Atlas...**
  - Manual and limited auditing through multiple half-baked tools, with more added each release to fill gaps

- **Cloudera Navigator**
  - Transparent, comprehensive, high-performance, compliance-ready encryption and key management

- N/A
Why Cloudera?

Your trusted partner for getting results with enterprise Hadoop.

- **Open Source Innovation**
  No one knows Hadoop better than Cloudera. Cloudera leads development of enterprise Hadoop and offers the best support, training, and services.

- **Powerful Enterprise Tools**
  Cloudera extends open source Hadoop with capabilities required by the largest enterprises.

- **Ecosystem**
  Cloudera partners with industry leaders to ensure Hadoop works with the platforms, tools, and integrators our customers rely on.

---

- **Enterprise Security**
  Meet compliance requirements and reduce risk exposure from storing sensitive data.

- **Data Governance**
  Enable compliance and maximize analyst productivity.

- **Complete Management**
  Deliver optimum system utilization and meet SLA commitments, on-premises or in the cloud, with minimum effort.
Thank You!

Sergio Rodríguez
sergio@pue.es