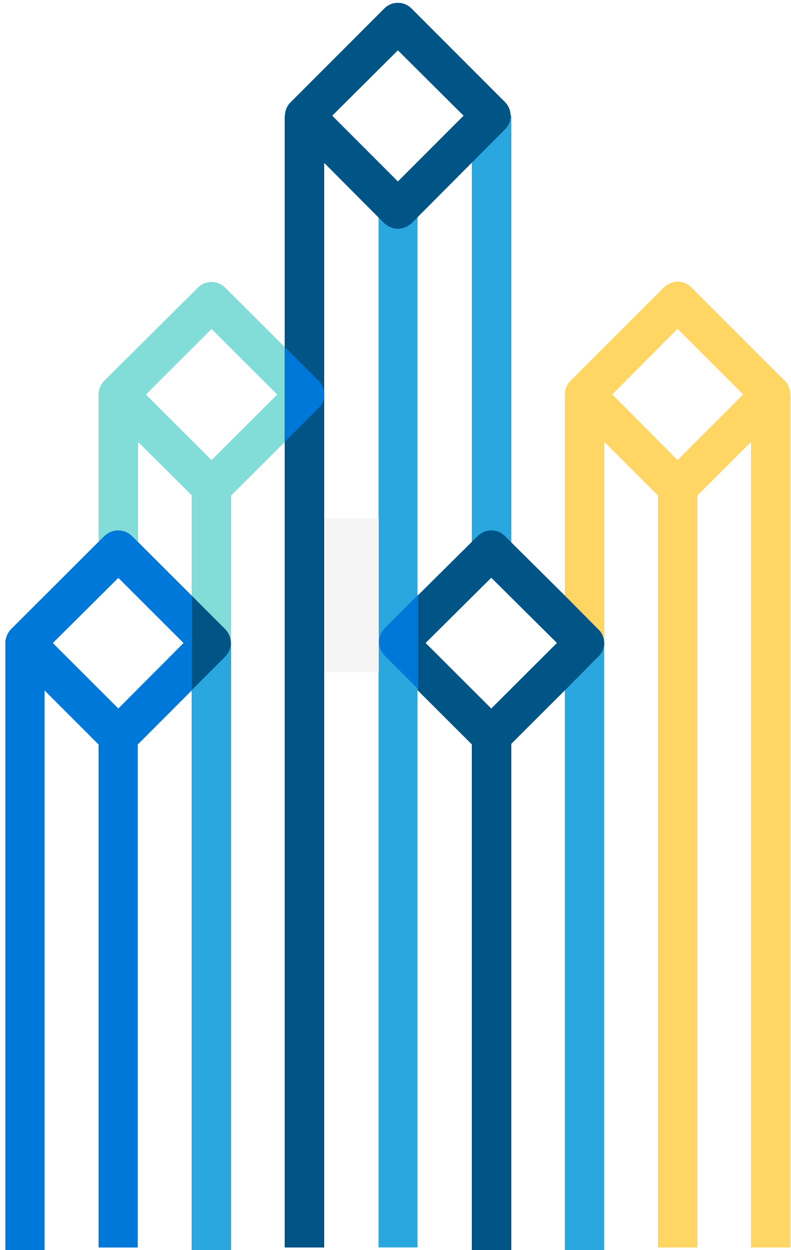
cloudera®



Leveraging Predictive Tools to Decrease Resolution Time

Angus Klein – Vice President, Global Support

Adam Warrington – Director, Engineering

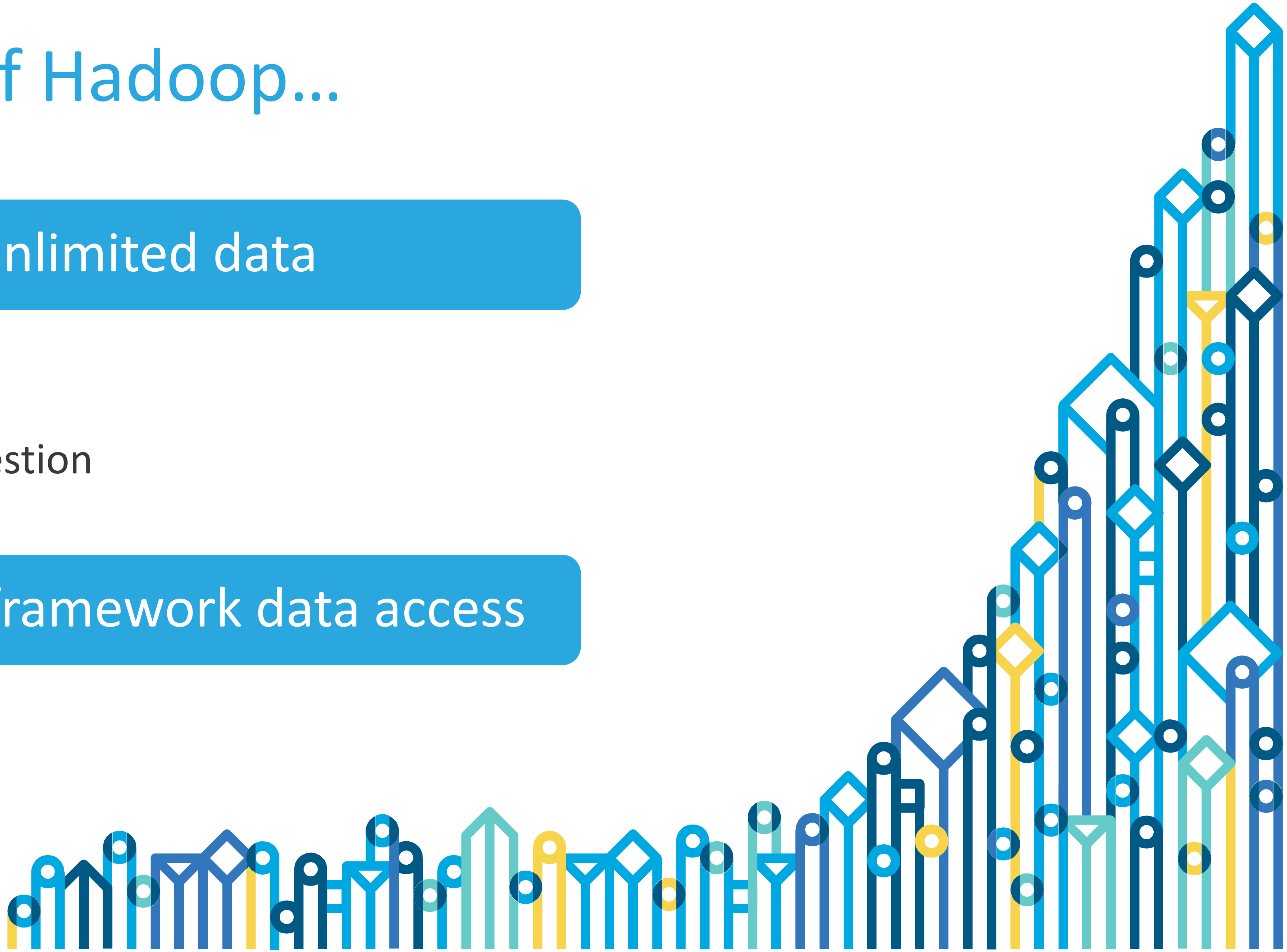
The Value of Hadoop…

One place for unlimited data

- All types
- More sources
- Faster, larger ingestion

Unified, multi-framework data access

- More users
- More tools
- Faster changes



cloudera

The Cloudera Value Chain

DEVELOP & PACKAGE OPEN SOURCE PROJECTS

FORM A STABLE, RELIABLE PLATFORM

THAT SUPPORTS POPULAR APPLICATIONS

TO SOLVE CUSTOMER PROBLEMS

Problem Statement

Supporting our product is **complex**

Issues **can be related** or root cause **might not be the same**

Looking for ways to work globally **at scale** as company continues to grow

cloudera

Support Team History

Training Introduced

2009

Services Offered

2010

Support Subscription Introduced

2011

- COTT
- Support systems
- Proactive

2012

- Frontline COE
- Backline

2013

Quality Team

2015

Enterprise grade customer support

1st

and only Hadoop vendor to have support certified

1yr

renewal cycle for certification

12

identified criteria areas

100+

service factors audited

Where are we?

Global time zone coverage

- EST (Raleigh)
- MST (Denver)
- CST (Austin)
- PST (Phoenix & Palo Alto)
- AEDT (Melbourne)
- IST (Chennai)
- JST (Tokyo)
- CST (Shanghai)
- GMT (London)
- CET (Budapest)

cloudera

Structuring Predictive Support

How we integrated predictive support capabilities into our support organization

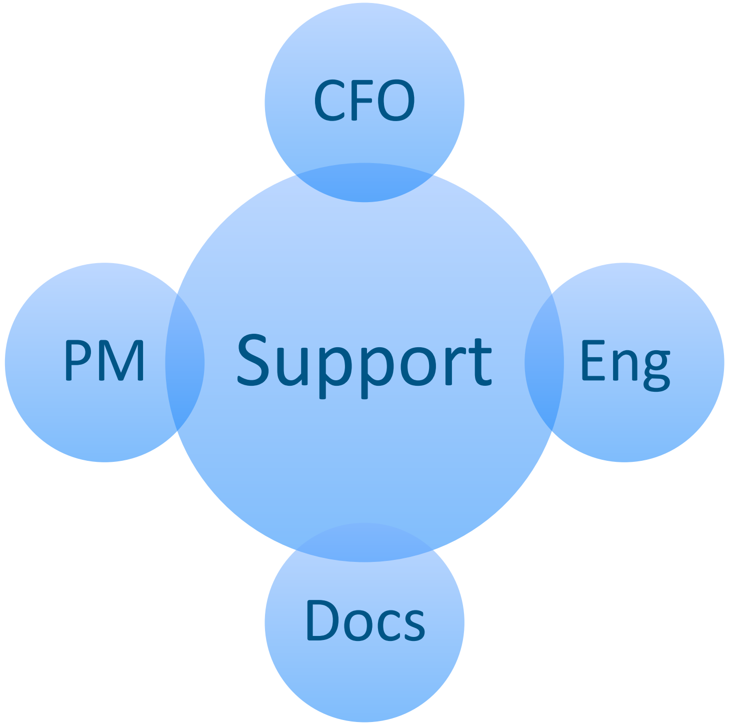
The Support Tools Team

Mission Statement: Build data driven software that reduces our time to solve on support cases while increasing customer satisfaction

Key Metrics

- Decrease Time to Solve
- Increase Support Rep Throughput
- Increase Case Deflection
- Improve Support Margins

Data Driven Support Changed the Game



Support organizations are one of the largest data drivers in any organization

Support becoming data-driven at Cloudera has been critical to establishing internal credibility at the exec table

cloudera

We dogfood

Diagnostic Tooling – What Does This Buy Us?

Lowers Time to Resolution

Improves our relationship with Engineering

Improves moral of Support Engineers

cloudera

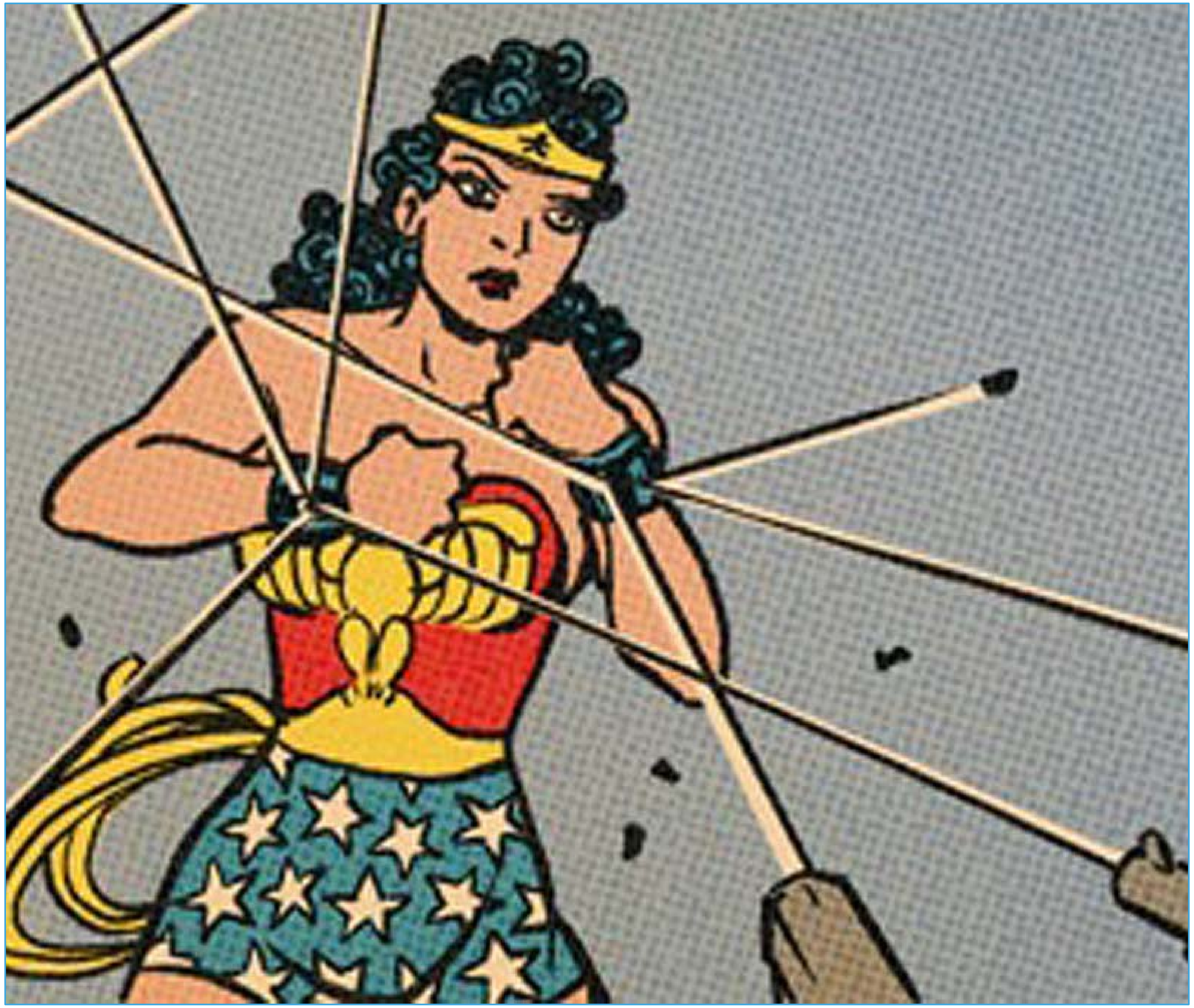
Customer cases leveraging the Cloudera Diagnostic Tools demonstrated an approximate 35% drop in time to resolution.

Predictive Support

Larger gains through case deflection possible with predictive support

Identifying known issues from diagnostic data

Notifying and working with the customer towards a solution to their problem



cloudera

Proactive Support

Onboard Process

We start our partnership at the very beginning by walking you through how things work

Predictive Validations

Powerful predictive alert system creates support tickets on behalf of our customers to help avoid known issues before they happen

Health Check

An early warning system which looks at key indicators that represents the health of our relationship with each customer

Over 15% of support cases are deflected by Cloudera Support's predictive support system.

Building Predictive Support

Step 1: Team Building

Team building

Create a clear and measurable mission statement

Hire 2-3 qualified engineers to prove the concept

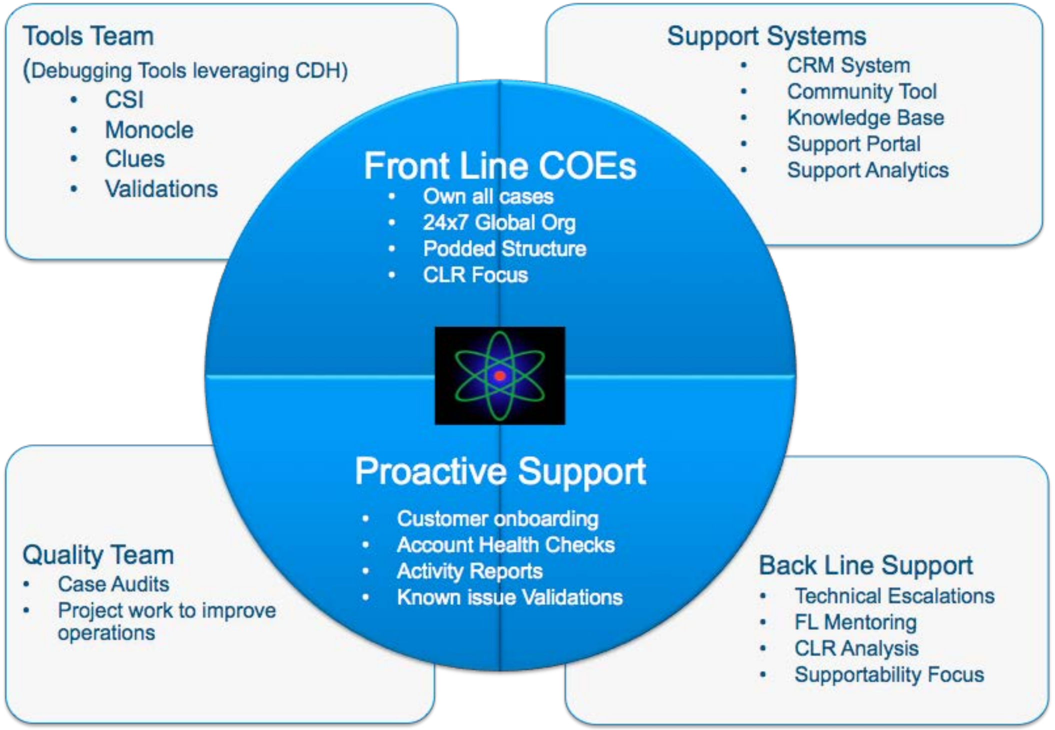
Understand your customer – in this case, that's the supporters

Keeping the Tools Team close

The tools team was kept close to the **full time support engineers**

Support engineers provide the **feedback loop** that allows us to build these applications

Looking for ways to work **at scale** as company continues to grow



Tools Team
(Debugging Tools leveraging CDH)
- CSI
- Monocle
- Clues
- Validations

Support Systems
- CRM System
- Community Tool
- Knowledge Base
- Support Portal
- Support Analytics

Front Line COEs
- Own all cases
- 24x7 Global Org
- Podded Structure
- CLR Focus

Proactive Support
- Customer onboarding
- Account Health Checks
- Activity Reports
- Known issue Validations

Quality Team
- Case Audits
- Project work to improve operations

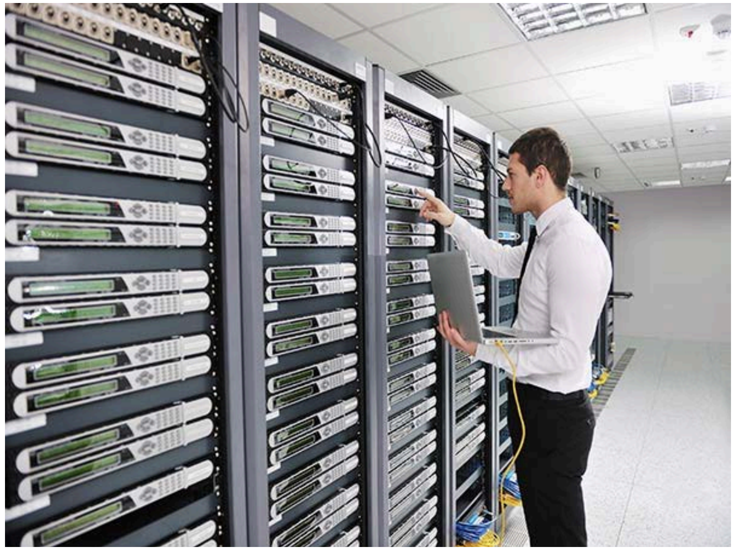
Back Line Support
- Technical Escalations
- FL Mentoring
- CLR Analysis
- Supportability Focus

Building Predictive Support

Step 2: Data Collection

Collect all the data

Support case interactions generates valuable support information



Troubleshooting sessions generate information about data relevant to solving a specific issue

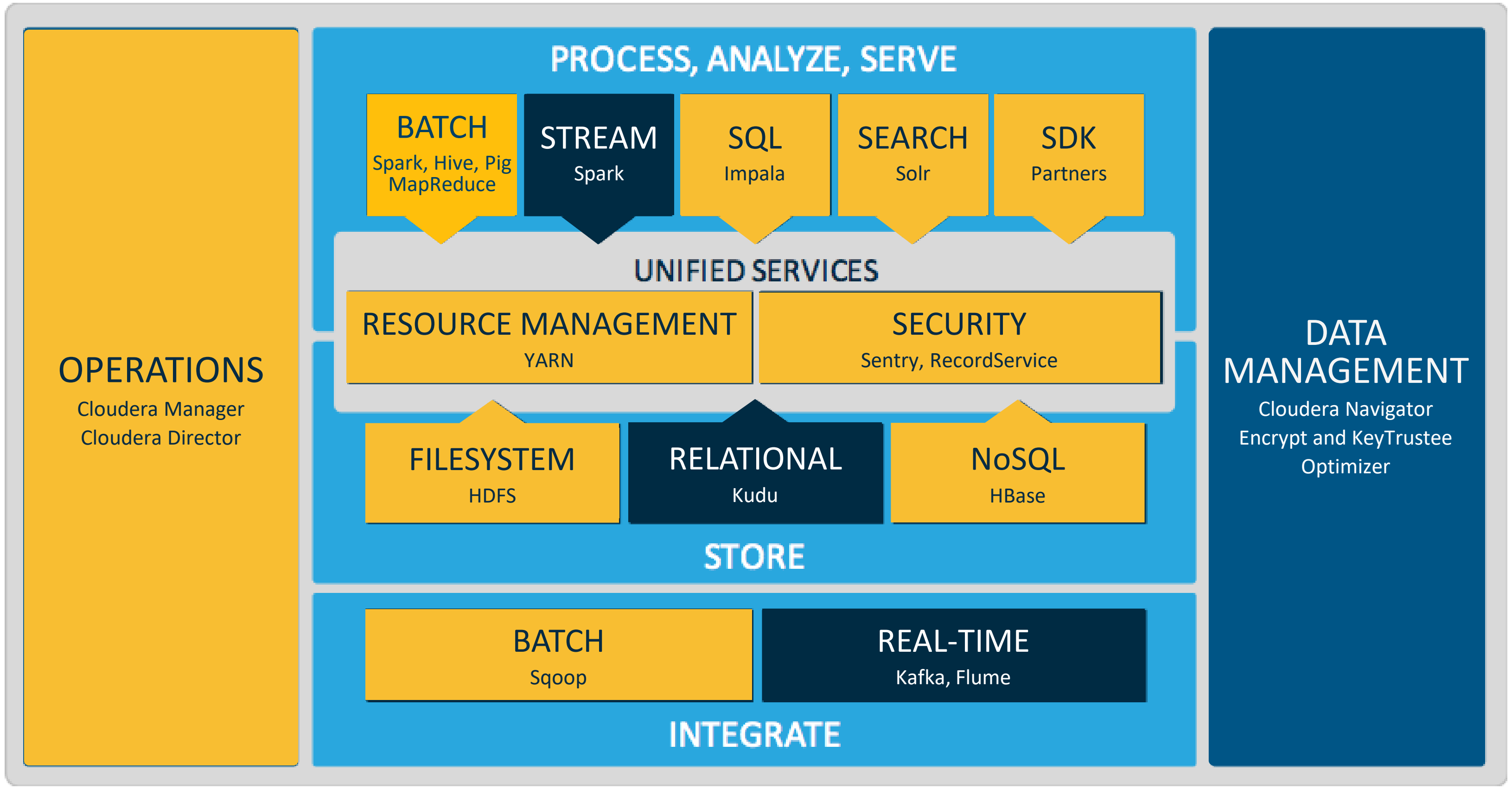
Customer installs generate diagnostic data critical to support

cloudera

Data sources we collect

Ingest & Consolidate

- Knowledge Base
- Internal CRM Data
- Support Records
- Apache Mailing Lists
- Community Forums
- Diagnostic bundles



OPERATIONS
Cloudera Manager
Cloudera Director

PROCESS, ANALYZE, SERVE

BATCH
Spark, Hive, Pig
MapReduce

STREAM
Spark

SQL
Impala

SEARCH
Solr

SDK
Partners

UNIFIED SERVICES

RESOURCE MANAGEMENT
YARN

SECURITY
Sentry, RecordService

FILESYSTEM
HDFS

RELATIONAL
Kudu

NoSQL
HBase

STORE

BATCH
Sqoop

REAL-TIME
Kafka, Flume

INTEGRATE

DATA MANAGEMENT
Cloudera Navigator
Encrypt and KeyTrustee
Optimizer

Data Collection Best Practices

Shadow Support Engineers to identify data and information they regularly use in a case lifecycle

Use support systems that are easy to extract data from

Store that data in a central data repository

cloudera

Game of Nodes

50
Nodes in cluster

550TB
Data size

3TB
New data per day

100K
Queries per day

cloudera

Building Predictive Support

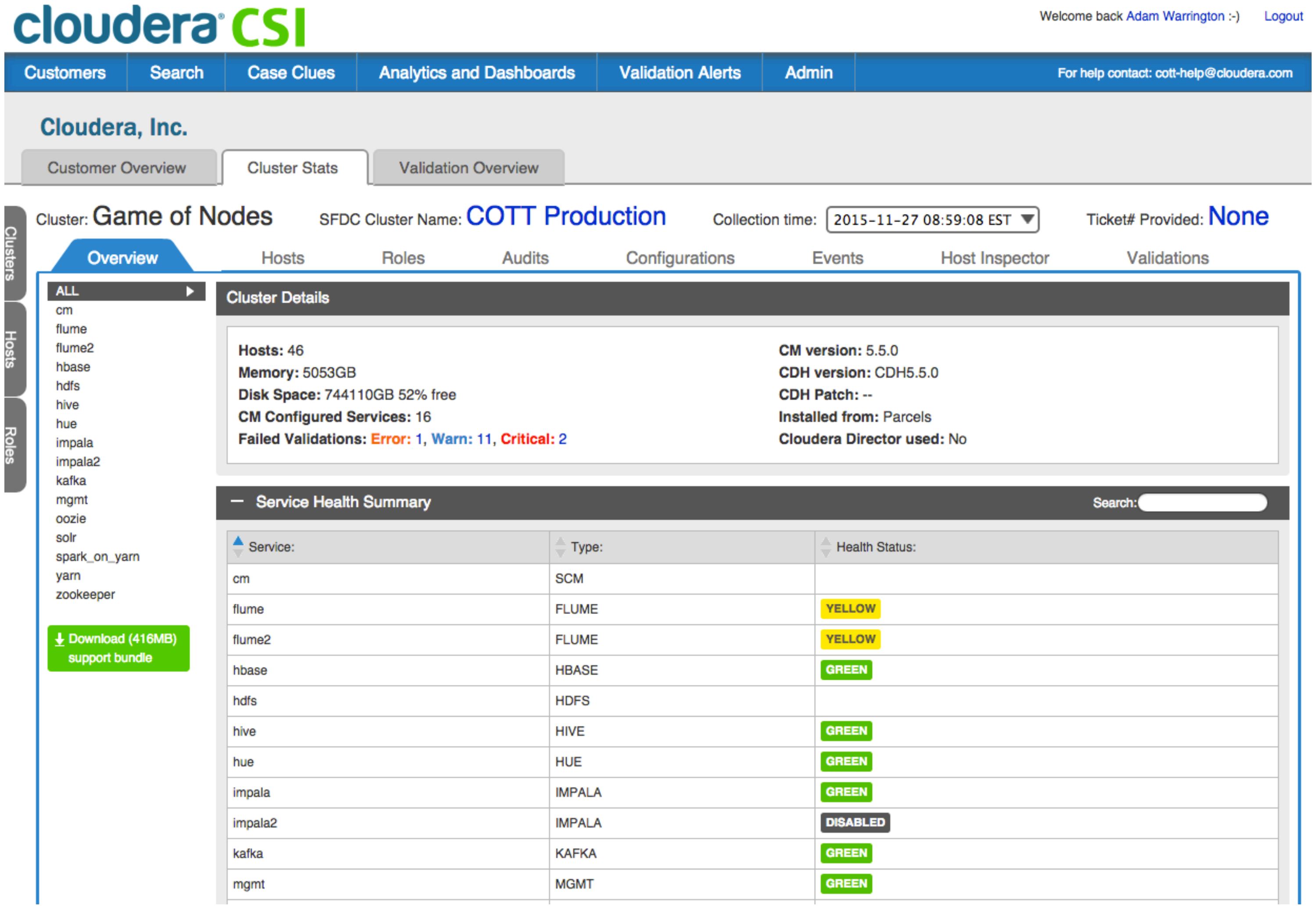
Step 3: Visualize

Customer Support Interface (CSI)

Data Ingestion
Our internal EDH ingests 10 support specific data sources. We have access to over 500TB of data and it is growing each month.

Data Visualization
Our goal is to visualize all data that is useful to a support engineer in a useful way. CSI is a java web application that sits on top of the EDH

Tools exist within CSI
All support tools exist as a function or feature within CSI. This includes all the tools we are about to go over (e.g. Diagnostic Bundles, *Validations*, *Monocle*, and *Clues)*



Cloudera, Inc.

Cluster: Game of Nodes SFDC Cluster Name: COTT Production Collection time: 2015-11-27 08:59:08 EST Ticket# Provided: None

Cluster Details

Hosts: 46
Memory: 5053GB
Disk Space: 744110GB 52% free
CM Configured Services: 16
Failed Validations: Error: 1, Warn: 11, Critical: 2

CM version: 5.5.0
CDH version: CDH5.5.0
CDH Patch: --
Installed from: Parcels
Cloudera Director used: No

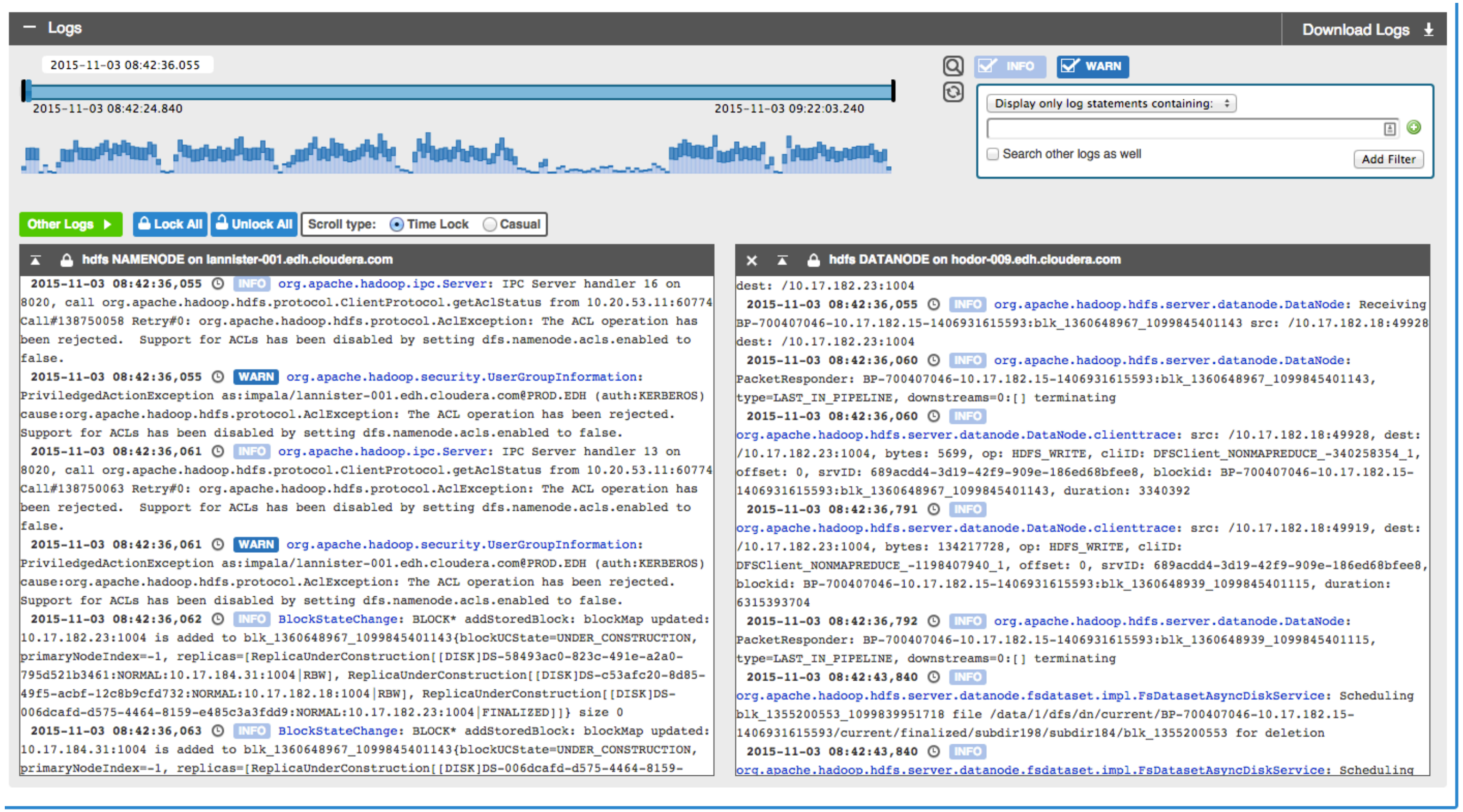
Service Health Summary

Service:	Type:	Health Status:
cm	SCM	
flume	FLUME	YELLOW
flume2	FLUME	YELLOW
hbase	HBASE	GREEN
hdfs	HDFS	
hive	HIVE	GREEN
hue	HUE	GREEN
impala	IMPALA	GREEN
impala2	IMPALA	DISABLED
kafka	KAFKA	GREEN
mgmt	MGMT	GREEN

Diagnostic Bundles

Cloudera Manager
One of the original problems in supporting Hadoop was seeing into the customer environment. Cloudera Manager has the ability to send a snapshot using diagnostic bundles.

What are they used for
Support engineers are able to dive into these bundles to get a granular view of the scenario and quickly solve issues using our tools.



hdfs NAMENODE on lannister-001.edh.cloudera.com

```
 2015-11-03 08:42:36,055 INFO org.apache.hadoop.ipc.Server: IPC Server handler 16 on
8020, call org.apache.hadoop.hdfs.protocol.ClientProtocol.getAclStatus from 10.20.53.11:60774
Call#138750058 Retry#0: org.apache.hadoop.hdfs.protocol.AclException: The ACL operation has
been rejected.  Support for ACLs has been disabled by setting dfs.namenode.acls.enabled to
false.
 2015-11-03 08:42:36,055 WARN org.apache.hadoop.security.UserGroupInformation:
PriviledgedActionException as:impala/lannister-001.edh.cloudera.com@PROD.EDH (auth:KERBEROS)
cause:org.apache.hadoop.hdfs.protocol.AclException: The ACL operation has been rejected.
Support for ACLs has been disabled by setting dfs.namenode.acls.enabled to false.
 2015-11-03 08:42:36,061 INFO org.apache.hadoop.ipc.Server: IPC Server handler 13 on
8020, call org.apache.hadoop.hdfs.protocol.ClientProtocol.getAclStatus from 10.20.53.11:60774
Call#138750063 Retry#0: org.apache.hadoop.hdfs.protocol.AclException: The ACL operation has
been rejected.  Support for ACLs has been disabled by setting dfs.namenode.acls.enabled to
false.
 2015-11-03 08:42:36,061 WARN org.apache.hadoop.security.UserGroupInformation:
PriviledgedActionException as:impala/lannister-001.edh.cloudera.com@PROD.EDH (auth:KERBEROS)
cause:org.apache.hadoop.hdfs.protocol.AclException: The ACL operation has been rejected.
Support for ACLs has been disabled by setting dfs.namenode.acls.enabled to false.
 2015-11-03 08:42:36,062 INFO BlockStateChange: BLOCK* addStoredBlock: blockMap updated:
10.17.182.23:1004 is added to blk_1360648967_1099845401143{blockUCState=UNDER_CONSTRUCTION,
primaryNodeIndex=-1, replicas=[ReplicaUnderConstruction[[DISK]DS-58493ac0-823c-491e-a2a0-
795d521b3461:NORMAL:10.17.184.31:1004|RBW], ReplicaUnderConstruction[[DISK]DS-c53afc20-8d85-
49f5-acbf-12c8b9cfd732:NORMAL:10.17.182.18:1004|RBW], ReplicaUnderConstruction[[DISK]DS-
006dcafd-d575-4464-8159-e485c3a3fdd9:NORMAL:10.17.182.23:1004|FINALIZED]]} size 0
 2015-11-03 08:42:36,063 INFO BlockStateChange: BLOCK* addStoredBlock: blockMap updated:
10.17.184.31:1004 is added to blk_1360648967_1099845401143{blockUCState=UNDER_CONSTRUCTION,
primaryNodeIndex=-1, replicas=[ReplicaUnderConstruction[[DISK]DS-006dcafd-d575-4464-8159-
```

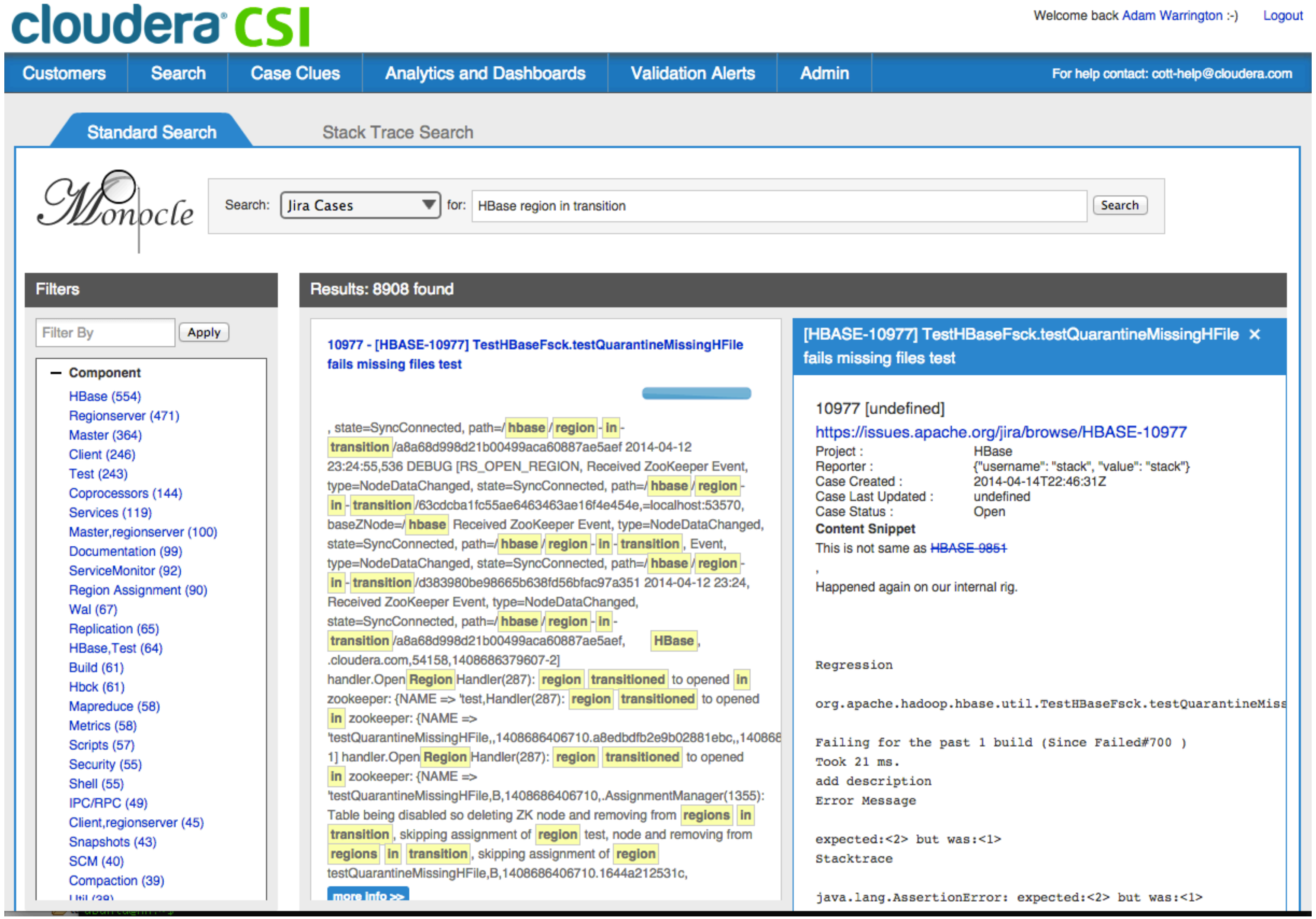
hdfs DATANODE on hodor-009.edh.cloudera.com

```
dest: /10.17.182.23:1004
 2015-11-03 08:42:36,055 INFO org.apache.hadoop.hdfs.server.datanode.DataNode: Receiving
BP-700407046-10.17.182.15-1406931615593:blk_1360648967_1099845401143 src: /10.17.182.18:49928
dest: /10.17.182.23:1004
 2015-11-03 08:42:36,060 INFO org.apache.hadoop.hdfs.server.datanode.DataNode:
PacketResponder: BP-700407046-10.17.182.15-1406931615593:blk_1360648967_1099845401143,
type=LAST_IN_PIPELINE, downstreams=0:[] terminating
 2015-11-03 08:42:36,060 INFO
org.apache.hadoop.hdfs.server.datanode.DataNode.clienttrace: src: /10.17.182.18:49928, dest:
/10.17.182.23:1004, bytes: 5699, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_-340258354_1,
offset: 0, srvID: 689acdd4-3d19-42f9-909e-186ed68bfee8, blockid: BP-700407046-10.17.182.15-
1406931615593:blk_1360648967_1099845401143, duration: 3340392
 2015-11-03 08:42:36,791 INFO
org.apache.hadoop.hdfs.server.datanode.DataNode.clienttrace: src: /10.17.182.18:49919, dest:
/10.17.182.23:1004, bytes: 134217728, op: HDFS_WRITE, cliID:
DFSClient_NONMAPREDUCE_-1198407940_1, offset: 0, srvID: 689acdd4-3d19-42f9-909e-186ed68bfee8,
blockid: BP-700407046-10.17.182.15-1406931615593:blk_1360648939_1099845401115, duration:
6315393704
 2015-11-03 08:42:36,792 INFO org.apache.hadoop.hdfs.server.datanode.DataNode:
PacketResponder: BP-700407046-10.17.182.15-1406931615593:blk_1360648939_1099845401115,
type=LAST_IN_PIPELINE, downstreams=0:[] terminating
 2015-11-03 08:42:43,840 INFO
org.apache.hadoop.hdfs.server.datanode.fsdataset.impl.FsDatasetAsyncDiskService: Scheduling
blk_1355200553_1099839951718 file /data/1/dfs/dn/current/BP-700407046-10.17.182.15-
1406931615593/current/finalized/subdir198/subdir184/blk_1355200553 for deletion
 2015-11-03 08:42:43,840 INFO
org.apache.hadoop.hdfs.server.datanode.fsdataset.impl.FsDatasetAsyncDiskService: Scheduling
```

Monocle

Making specialized knowledge searchable
Searching all of the data sources within CSI we are able to create a single index of both internal and open source knowledge for a one stop Hadoop engine.

What is it used for
No longer making support engineers have to “Google” for information. Our internal search platform is the most powerful Hadoop support engine for all their needs.



Welcome back Adam Warrington :-) Logout

For help contact: cott-help@cloudera.com

Search: Jira Cases for: HBase region in transition

Filters

Filter By

- Component
  - HBase (554)
  - Regionserver (471)
  - Master (364)
  - Client (246)
  - Test (243)
  - Coprocessors (144)
  - Services (119)
  - Master,regionserver (100)
  - Documentation (99)
  - ServiceMonitor (92)
  - Region Assignment (90)
  - Wal (67)
  - Replication (65)
  - HBase,Test (64)
  - Build (61)
  - Hbck (61)
  - Mapreduce (58)
  - Metrics (58)
  - Scripts (57)
  - Security (55)
  - Shell (55)
  - IPC/RPC (49)
  - Client,regionserver (45)
  - Snapshots (43)
  - SCM (40)
  - Compaction (39)
  - Util (39)

Results: 8908 found

10977 - [HBASE-10977] TestHBaseFsck.testQuarantineMissingHFile fails missing files test

, state=SyncConnected, path=/hbase/region-in-transition/a8a68d998d21b00499aca60887ae5aef 2014-04-12 23:24:55,536 DEBUG [RS_OPEN_REGION, Received ZooKeeper Event, type=NodeDataChanged, state=SyncConnected, path=/hbase/region-in-transition/63cdcba1fc55ae6463463ae16f4e454e,=localhost:53570, baseZNode=/hbase Received ZooKeeper Event, type=NodeDataChanged, state=SyncConnected, path=/hbase/region-in-transition, Event, type=NodeDataChanged, state=SyncConnected, path=/hbase/region-in-transition/d383980be98665b638fd56bfac97a351 2014-04-12 23:24, Received ZooKeeper Event, type=NodeDataChanged, state=SyncConnected, path=/hbase/region-in-transition/a8a68d998d21b00499aca60887ae5aef, HBase, .cloudera.com,54158,1408686379607-2] handler.OpenRegionHandler(287): region transitioned to opened in zookeeper: {NAME => 'test,Handler(287): region transitioned to opened in zookeeper: {NAME => 'testQuarantineMissingHFile,,1408686406710.a8edbdfb2e9b02881ebc,,140868 1] handler.OpenRegionHandler(287): region transitioned to opened in zookeeper: {NAME => 'testQuarantineMissingHFile,B,1408686406710,.AssignmentManager(1355): Table being disabled so deleting ZK node and removing from regions in transition, skipping assignment of region test, node and removing from regions in transition, skipping assignment of region testQuarantineMissingHFile,B,1408686406710.1644a212531c,

[HBASE-10977] TestHBaseFsck.testQuarantineMissingHFile fails missing files test ×

10977 [undefined]

https://issues.apache.org/jira/browse/HBASE-10977

Project :	HBase
Reporter :	{"username": "stack", "value": "stack"}
Case Created :	2014-04-14T22:46:31Z
Case Last Updated :	undefined
Case Status :	Open

Content Snippet

This is not same as ~~HBASE-9851~~

,

Happened again on our internal rig.

```
Regression

org.apache.hadoop.hbase.util.TestHBaseFsck.testQuarantineMiss

Failing for the past 1 build (Since Failed#700 )
Took 21 ms.
add description
Error Message

expected:<2> but was:<1>
Stacktrace

java.lang.AssertionError: expected:<2> but was:<1>
```

cloudera

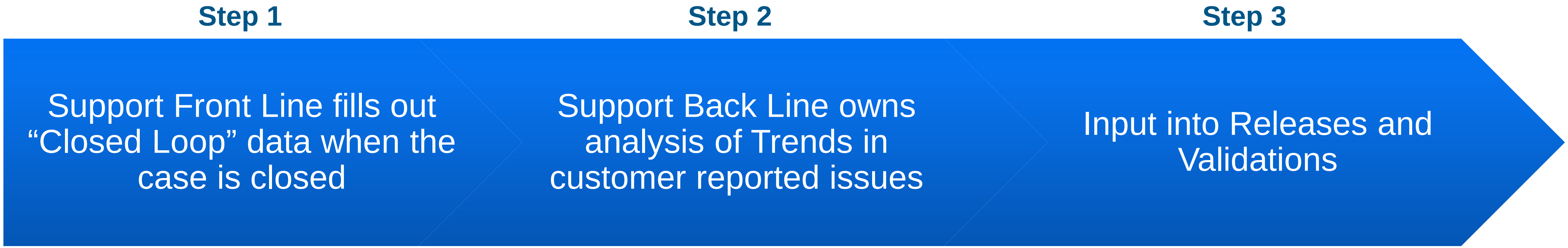
Building Predictive Support

Step 4: Signature Identification

cloudera

Closed Loop Review - Linear Process Flow

Goal: To drive supportability in the Cloudera Platform to improve the customer experience



Step 1

Support Front Line fills out "Closed Loop" data when the case is closed

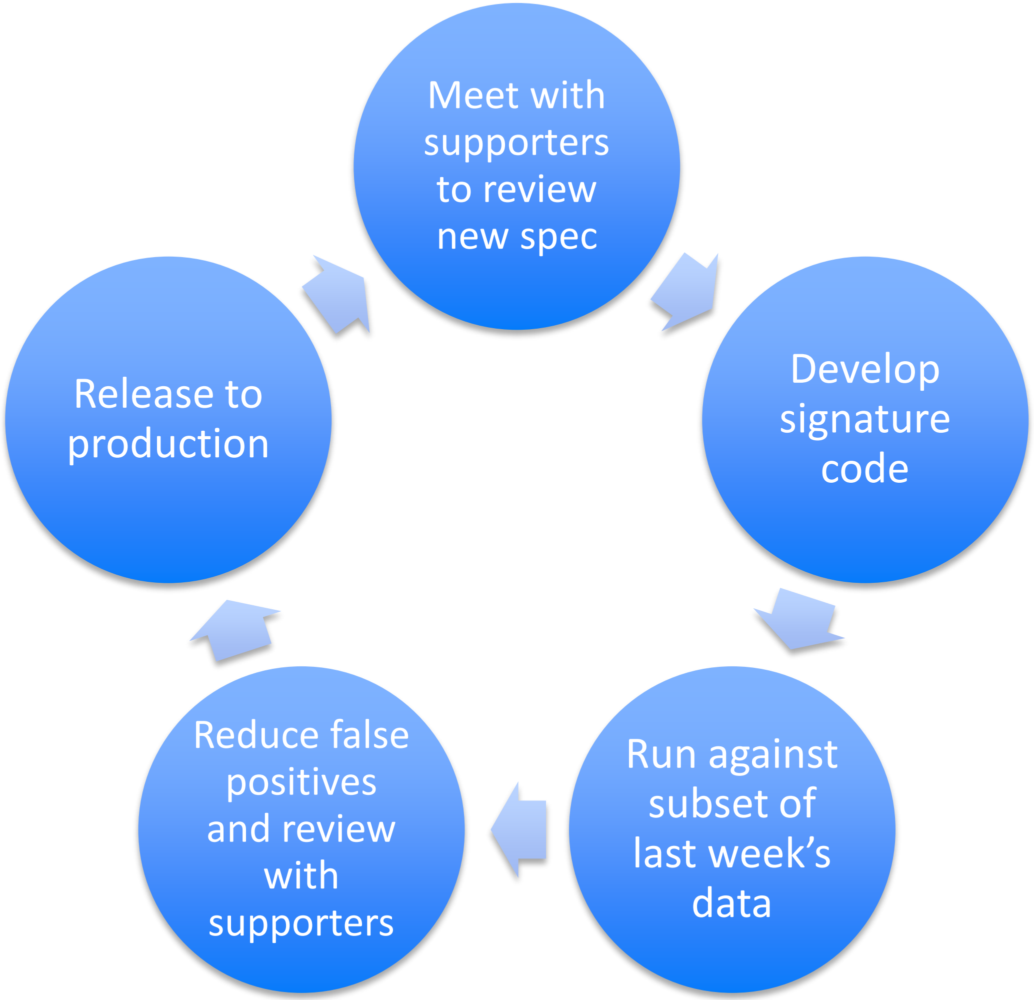
Step 2

Support Back Line owns analysis of Trends in customer reported issues

Step 3

Input into Releases and Validations

Issue signature creation – a collaborative process



Meet with supporters to review new spec

Develop signature code

Run against subset of last week's data

Reduce false positives and review with supporters

Release to production

cloudera

Building Predictive Support

Step 5: Delivering Predictive Support

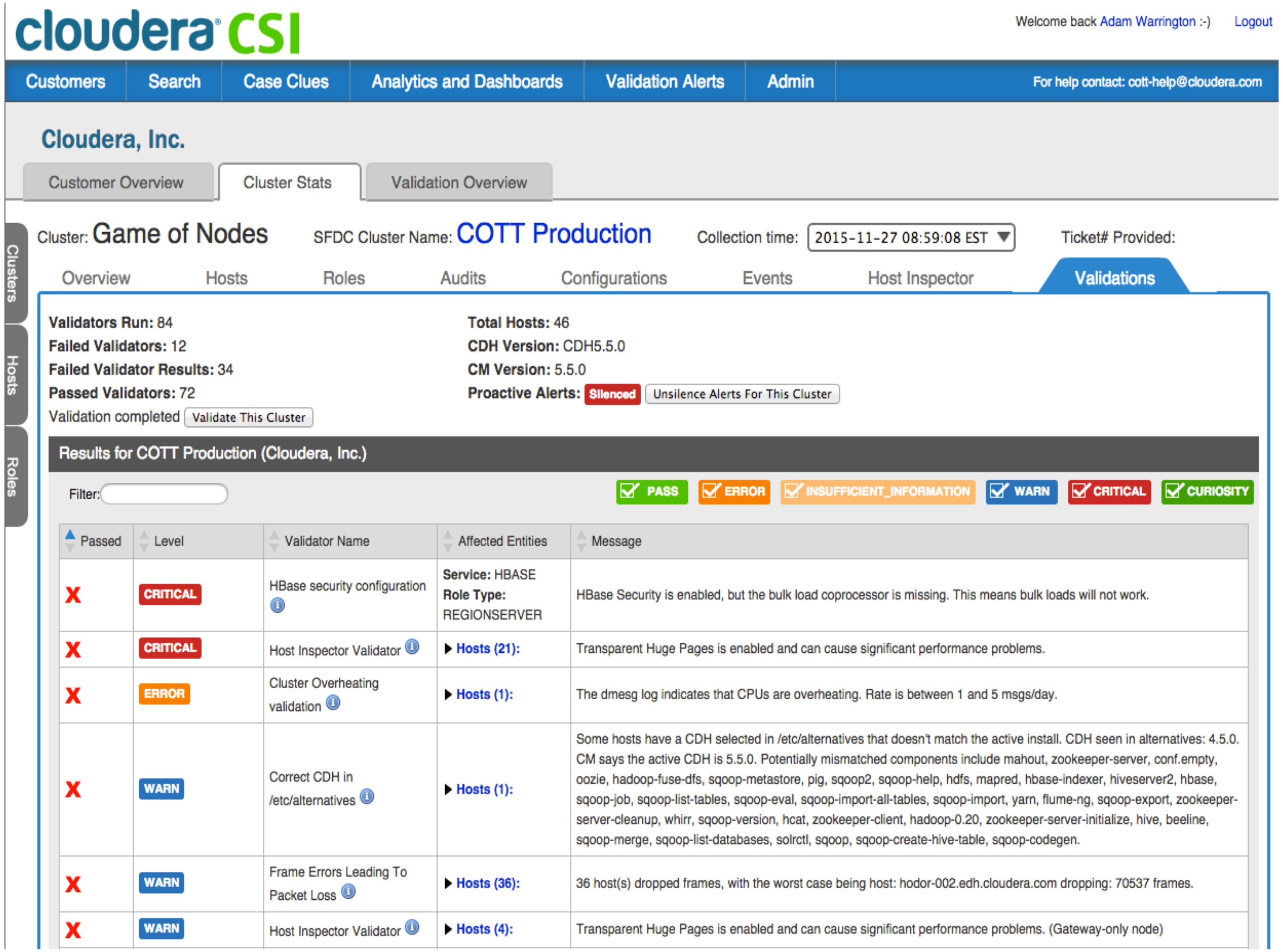
cloudera

Proactive Validations

Automatic support case creation on critical issue detection

Drive engagement with customer through known issue resolution channel

Leverage known troubleshooting mechanism and best practices



Cluster: Game of Nodes SFDC Cluster Name: COTT Production Collection time: 2015-11-27 08:59:08 EST ▼ Ticket# Provided:

Validators Run: 84
Failed Validators: 12
Failed Validator Results: 34
Passed Validators: 72
Validation completed Validate This Cluster

Total Hosts: 46
CDH Version: CDH5.5.0
CM Version: 5.5.0
Proactive Alerts: Silenced Unsilence Alerts For This Cluster

Results for COTT Production (Cloudera, Inc.)

Filter:

☑ PASS ☑ ERROR ☑ INSUFFICIENT_INFORMATION ☑ WARN ☑ CRITICAL ☑ CURIOSITY

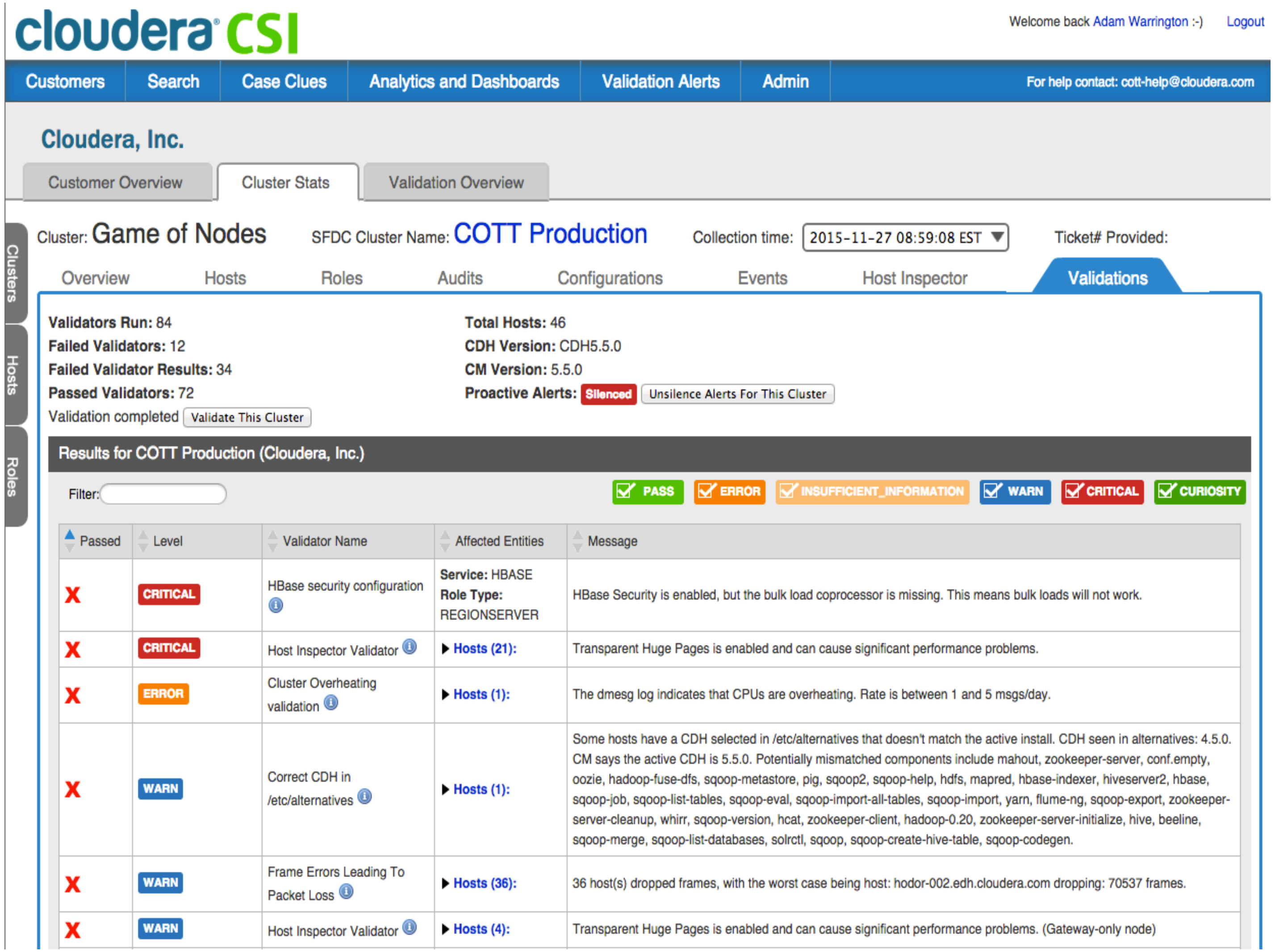
Passed	Level	Validator Name	Affected Entities	Message
X	CRITICAL	HBase security configuration ⓘ	**Service:** HBASE **Role Type:** REGIONSERVER	HBase Security is enabled, but the bulk load coprocessor is missing. This means bulk loads will not work.
X	CRITICAL	Host Inspector Validator ⓘ	▶ Hosts (21):	Transparent Huge Pages is enabled and can cause significant performance problems.
X	ERROR	Cluster Overheating validation ⓘ	▶ Hosts (1):	The dmesg log indicates that CPUs are overheating. Rate is between 1 and 5 msgs/day.
X	WARN	Correct CDH in /etc/alternatives ⓘ	▶ Hosts (1):	Some hosts have a CDH selected in /etc/alternatives that doesn't match the active install. CDH seen in alternatives: 4.5.0. CM says the active CDH is 5.5.0. Potentially mismatched components include mahout, zookeeper-server, conf.empty, oozie, hadoop-fuse-dfs, sqoop-metastore, pig, sqoop2, sqoop-help, hdfs, mapred, hbase-indexer, hiveserver2, hbase, sqoop-job, sqoop-list-tables, sqoop-eval, sqoop-import-all-tables, sqoop-import, yarn, flume-ng, sqoop-export, zookeeper-server-cleanup, whirr, sqoop-version, hcat, zookeeper-client, hadoop-0.20, zookeeper-server-initialize, hive, beeline, sqoop-merge, sqoop-list-databases, solrctl, sqoop, sqoop-create-hive-table, sqoop-codegen.
X	WARN	Frame Errors Leading To Packet Loss ⓘ	▶ Hosts (36):	36 host(s) dropped frames, with the worst case being host: hodor-002.edh.cloudera.com dropping: 70537 frames.
X	WARN	Host Inspector Validator ⓘ	▶ Hosts (4):	Transparent Huge Pages is enabled and can cause significant performance problems. (Gateway-only node)

Reactive Validations

Reactive support greatly benefits from validations

Able to show validations that might have a higher false positive rate

Able to show validations that have lower criticality, but might relate to ongoing support cases



Cluster: Game of Nodes SFDC Cluster Name: COTT Production Collection time: 2015-11-27 08:59:08 EST Ticket# Provided:

Validators Run: 84
Failed Validators: 12
Failed Validator Results: 34
Passed Validators: 72
Validation completed

Total Hosts: 46
CDH Version: CDH5.5.0
CM Version: 5.5.0
Proactive Alerts: Silenced

Results for COTT Production (Cloudera, Inc.)

Passed	Level	Validator Name	Affected Entities	Message
X	CRITICAL	HBase security configuration	**Service:** HBASE **Role Type:** REGIONSERVER	HBase Security is enabled, but the bulk load coprocessor is missing. This means bulk loads will not work.
X	CRITICAL	Host Inspector Validator	▶ Hosts (21):	Transparent Huge Pages is enabled and can cause significant performance problems.
X	ERROR	Cluster Overheating validation	▶ Hosts (1):	The dmesg log indicates that CPUs are overheating. Rate is between 1 and 5 msgs/day.
X	WARN	Correct CDH in /etc/alternatives	▶ Hosts (1):	Some hosts have a CDH selected in /etc/alternatives that doesn't match the active install. CDH seen in alternatives: 4.5.0. CM says the active CDH is 5.5.0. Potentially mismatched components include mahout, zookeeper-server, conf.empty, oozie, hadoop-fuse-dfs, sqoop-metastore, pig, sqoop2, sqoop-help, hdfs, mapred, hbase-indexer, hiveserver2, hbase, sqoop-job, sqoop-list-tables, sqoop-eval, sqoop-import-all-tables, sqoop-import, yarn, flume-ng, sqoop-export, zookeeper-server-cleanup, whirr, sqoop-version, hcat, zookeeper-client, hadoop-0.20, zookeeper-server-initialize, hive, beeline, sqoop-merge, sqoop-list-databases, solrctl, sqoop, sqoop-create-hive-table, sqoop-codegen.
X	WARN	Frame Errors Leading To Packet Loss	▶ Hosts (36):	36 host(s) dropped frames, with the worst case being host: hodor-002.edh.cloudera.com dropping: 70537 frames.
X	WARN	Host Inspector Validator	▶ Hosts (4):	Transparent Huge Pages is enabled and can cause significant performance problems. (Gateway-only node)

Basic Cluster Checklist

Run the predictive support validations at the start of a customer's contract

Getting in front of issues early saves money and increases customer satisfaction

Targeting types customers or environments that are high cost to support can improve chances of success

cloudera

Thank You