



# Cloudera Operational DB

(powered by Apache HBase and  
Apache Phoenix)

December 2019

Timothy Spann



# Welcome to Future of Data - Princeton



<https://www.meetup.com/futureofdata-princeton/>

From Big Data to AI to Streaming to Containers to Cloud to Analytics to Cloud Storage to Fast Data to Machine Learning to Microservices to ...



@PaasDev



---

# Who Am I? Timothy Spann

## Data in Motion Field Engineer



DZone Zone Leader and Big Data MVB;  
Princeton NJ Future of Data Meetup;  
ex-Pivotal Field Engineer;  
Author of Apache Kafka RefCard  
<https://github.com/tspannhw>  
<https://www.datainmotion.dev/>



@PaasDev

# This Meetup Made Possible Thanks To:



Paul Vidal from <https://www.meetup.com/futureofdata-philadelphia/> for CDP HBase Environment and Cloud Magic

Josh Elser and Josiah Goodson for OpDB Slides and HBase Guidance

Milind Pandit from <https://www.meetup.com/TechnologySolutionsHub>

Mehul Shah from <https://www.meetup.com/TechnologySolutionsHub>

Vijay Garg from <https://pga.fund/>

Madhavi from <https://www.nuwaysolutions.com/>



<https://www.slideshare.net/bunkertor/tracking-crime-as-it-occurs-with-apache-phoenix-apache-hbase-and-apache-nifi>

*Operational DB is DBMS used to manage dynamic and changing data in real time and enable applications that drive the business*



## WHAT HAVE Apache HBase 2.0 & Apache Phoenix ENABLED?

- Operationalizing ML / AI to revolutionize healthcare, public utilities, etc
- Serving real-time content at webscale
- Empowering big data analytics for operational and offline uses
- Acting as a resilient store of record

# APACHE HBASE FAST FACTS

---

Largest Database

**14 Petabytes**

Best Known App

**Siri**

Fastest  
Ingestion

**20M Events/s**

Users

**750+**

# COMPANIES THAT USE HBASE

---

**SIEMENS**



**RingCentral®**



**OP@WER**



**ALLEGIS**  
GROUP

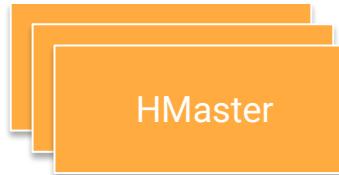


# HBASE ARCHITECTURE

Orchestration layer



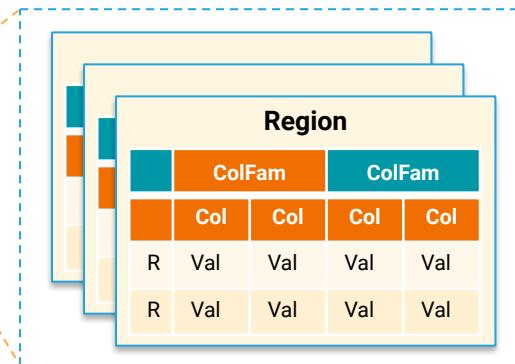
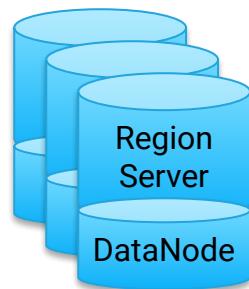
- Heartbeat
- Server state



- DDL operations
- Region assignment
- Recovery orchestration

Data plane

- Services client reads & writes
- Maximizes in-memory operations for low-latency operations
- Provides data resiliency



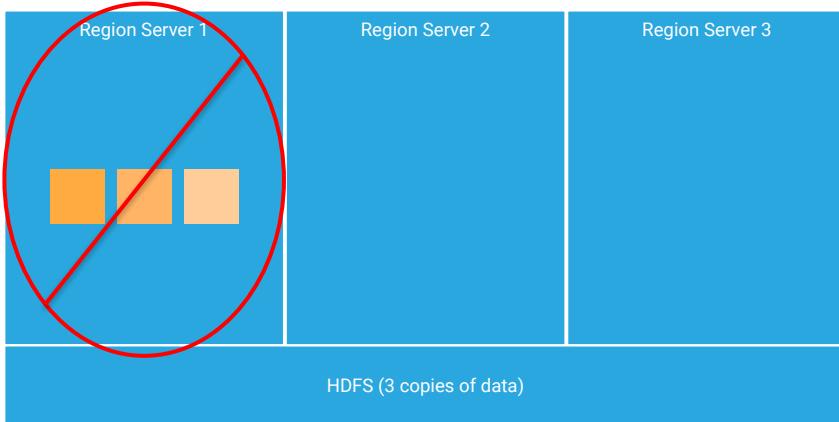
- Regions are table segments
- Read and write path are in the data plane

# SCHEMA-LESS DATA MODEL

	Column Family		Column Family	
	Column	Column	Column	Column
RowKey	Cell	Cell	Cell	Cell
RowKey	Cell	Cell	Cell	Cell

- Column families defined at time of table creation
- Columns created as required (at time of data insertion)
  - No limits to number of columns
- Tables can grow in two dimensions – columns and rows
- Compression & encoding applied at column family level
- No declaration of data types (i.e., a column can contain multiple data types)

# HIGHLY AVAILABLE OUT OF THE BOX (<1 MINUTE RECOVERY)



## What happens when a region crashes

1. Region server crashes
2. Writes and reads time-out for regions in impacted region server
3. Regions are redistributed to other region servers
4. WAL is replayed in other region servers
5. Reads & writes are able to continue to impacted regions

Typical recovery period < 1 minute (for impacted regions only)  
No manual intervention

# SECURITY MODEL



## Authentication

- Kerberos



## Role Based Access control

- Permissions & Scope enable flexible role based access control
- Scope: Global, Namespace, Table, Column Family, Cell



## DB security & encryption

- Transparent encryption of data on the wire and data on disk (HFile for data at rest, secure WAL for data in motion within HBase)
- Logging & auditability: configurable & fixed event

# What Phoenix adds to HBase

## Flexible, scale-out, no-sql database

HBase

```
APP  
Put put = new Put(Bytes.toBytes(rowKey));  
put.addColumn(COLUMN_FAMILY_NAME, COLUMN_NAME,  
Bytes.toBytes(GREETINGS));  
  
table.put(put);
```

## RDBMS-like, scale-out database

Phoenix

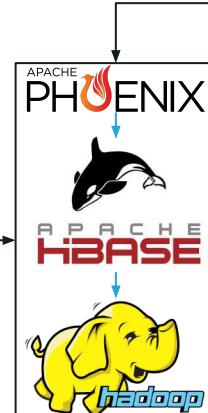
```
APP  
stmt.executeUpdate("UPSERT INTO TABLE_NAME  
VALUES(rowKey, GREETINGS)");  
  
stmt.execute();
```

### Pros:

- Maximally flexible & customizable
- SQL only for data remediation

### Cons:

- Unfamiliar to SQL developers
- Requires non-traditional data architecture



- Programmatic ANSI SQL support
- RDBMS-like data architecture
- Auto-applies performance best practices
- Can co-exist with HBase apps

- Reduced flexibility vis-à-vis vanilla HBase
- Phoenix specific data format means you can't use HBase APIs directly

# Key Phoenix capabilities

- **ANSI SQL** including joins
- Flexible Schemas / Dynamic Columns
- Secondary Indexes
- Aggregation pushdowns
- Cross-language client support
- Query logging
- Security through **Ranger** (supports RBAC, ABAC, etc)
- **JDBC/ODBC** connectivity for operational reporting
  - Plugs in to any JDBC/ODBC-compatible BI tool to enable self-service analytics and insight



Phoenix  
Applications



# ANSI SQL 92 Support

Supported today	Roadmap
Standard SQL Data Types	UNION
SELECT, UPSERT, DELETE	Windowing Functions
JOINS: Inner and Outer	Transactions
Subqueries	Cross Joins
Secondary Indexes	Authorization
GROUP BY, ORDER BY, HAVING	Replication Management
AVG, COUNT, MIN, MAX, SUM	Column Constraints and Defaults
Primary Keys, Constraints	UDFs
CASE, COALESCE	
VIEWS	
Flexible Schema	
UNION ALL	

---

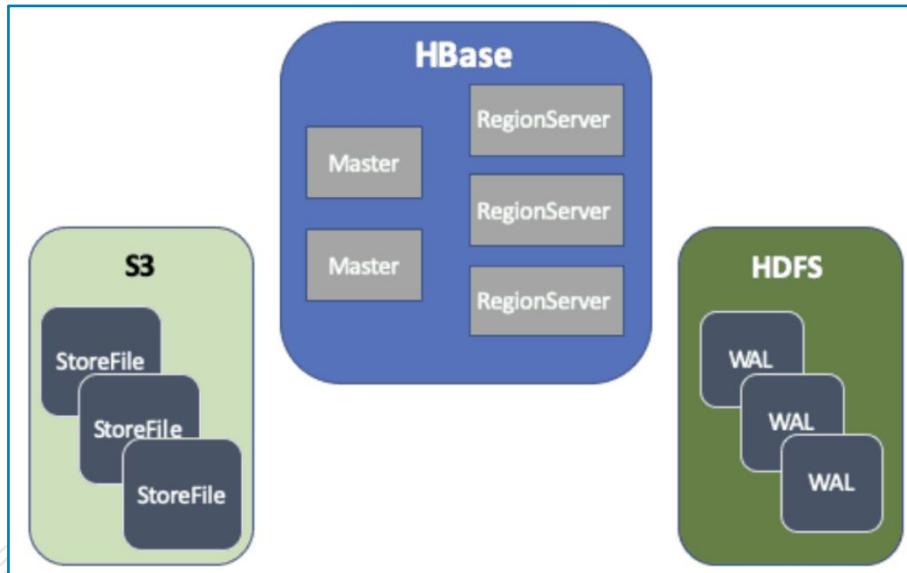
# CLOUD OPTIMIZED : HBASE backed by both HDFS and S3

*Cludera provides HBase backed by Amazon's S3*

- Cludera Data Platform (CDP) provides an out-of-the-box solution that allows Apache HBase deployments to use Amazon Simple Storage Service (S3) as its main persistence layer for saving table data
- Amazon's Simple Storage Service (S3) is an eventually consistent object store, and HBase requires a consistent and atomic filesystem which means that it cannot directly use S3. *Let's look at the topology.*

# CLOUD OPTIMIZED : Cloudera HBASE backed by both HDFS and S3

Cloudera with CDP has built a solution where when you launch an Operational Database (HBase) cluster on CDP, **HBase StoreFiles (the backing files for HBase tables)** are stored in **S3** and **HBase write-ahead-logs (WAL)** are stored in an **HDFS** instance run alongside HBase per usual.



# CLOUD OPTIMIZED : HBASE backed by both HDFS and S3

- Configuring HBase to use S3 for its StoreFiles has many benefits to our users.
- One such benefit is that users can decouple their storage and compute.
- If there are times in which no access to HBase is necessary, HBase can be cleanly shut down and all compute resources reclaimed to eliminate any cost of compute.
- When HBase access is needed again, the HBase cluster can be recreated, pointing to the same data in S3. Upon startup, HBase can re-initialize itself solely from the data in S3.

# WHERE IS APACHE HBASE TODAY



- Large ecosystem (Nifi, Spark, Hive, Impala, SOLR, Ranger, Atlas, etc)
- Supports NoSQL, SQL, Geospatial, Graph, TimeSeries, Key Value and other modes<sup>1</sup>
- Sold by: Cloudera, IBM, Microsoft, Amazon, Teradata and more

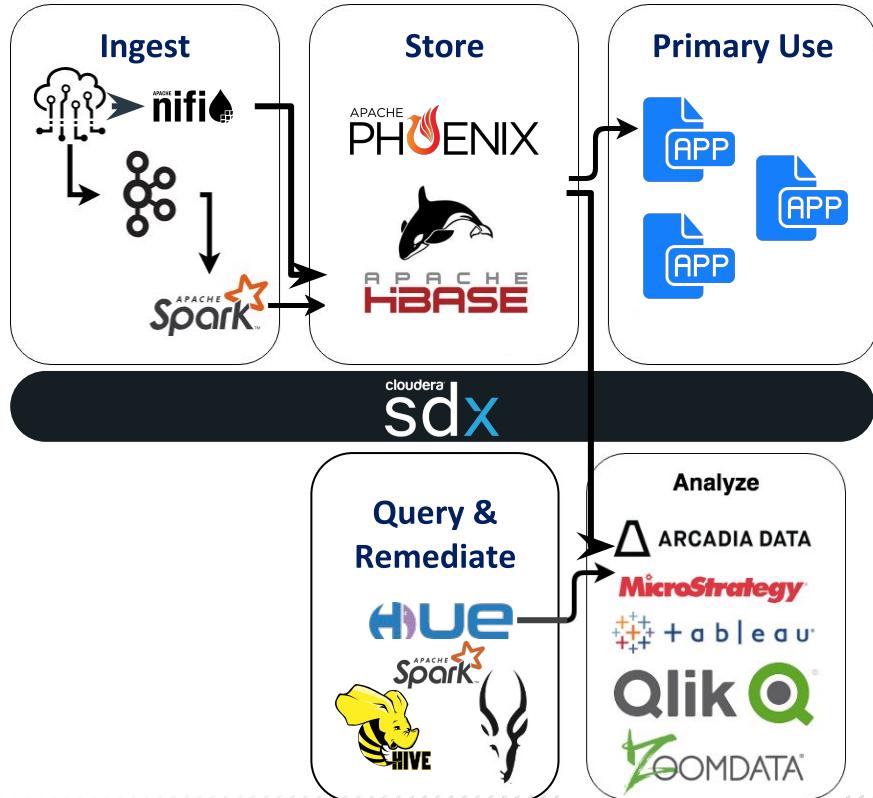
# Cloudera and Apache HBase

- The upstream community is pretty huge and very active with contributions coming from multiple developers from Cloudera, Microsoft, Amazon, Alibaba, Apple Salesforce and Xiaomi etc.
- Cloudera is a very active contributor to upstream HBase along with Apache Phoenix.
  - **Currently > 8 PMCs and > 2 committers.**
- CDP is based off latest HBase v2 and Phoenix v5.

# New features in HBase 2+

- Operational simplicity
  - Assignment Manager V2 (using Procedure Framework 2)
  - Offline compaction tool (outside regionservers to save I/O thrasing)
  - Replication: namespace & serial and for bulk-loads
- Performance
  - Off-heap cache improvements (Uses `DirectByteByffers` to manage buckets outside of the JVM heap to eliminate impact of gc to get better read perf)
- Space Quotas (to support multi tenancy)
- S3 support
- Spark 2 integration
- Async Client

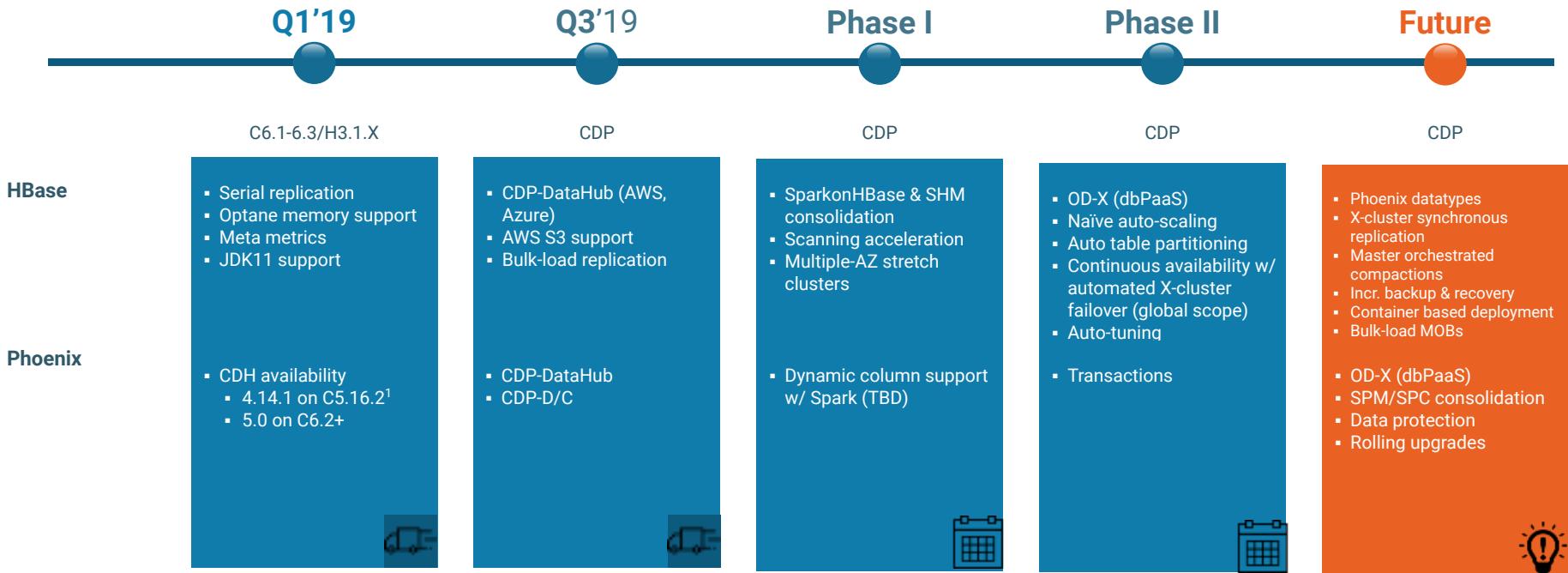
# NO(T ONLY)SQL PHOENIX



## Storage for business apps that require big-data

- Provides familiar & easy interface for developers
- Advanced multi-tenancy capabilities
- Support near 100% availability for mission critical applications & many traditional transactional apps
- Scale to billions of rows and millions of columns
- Easily combine data sources that use a wide variety of different structures and schemas

# Operational DB ROADMAP

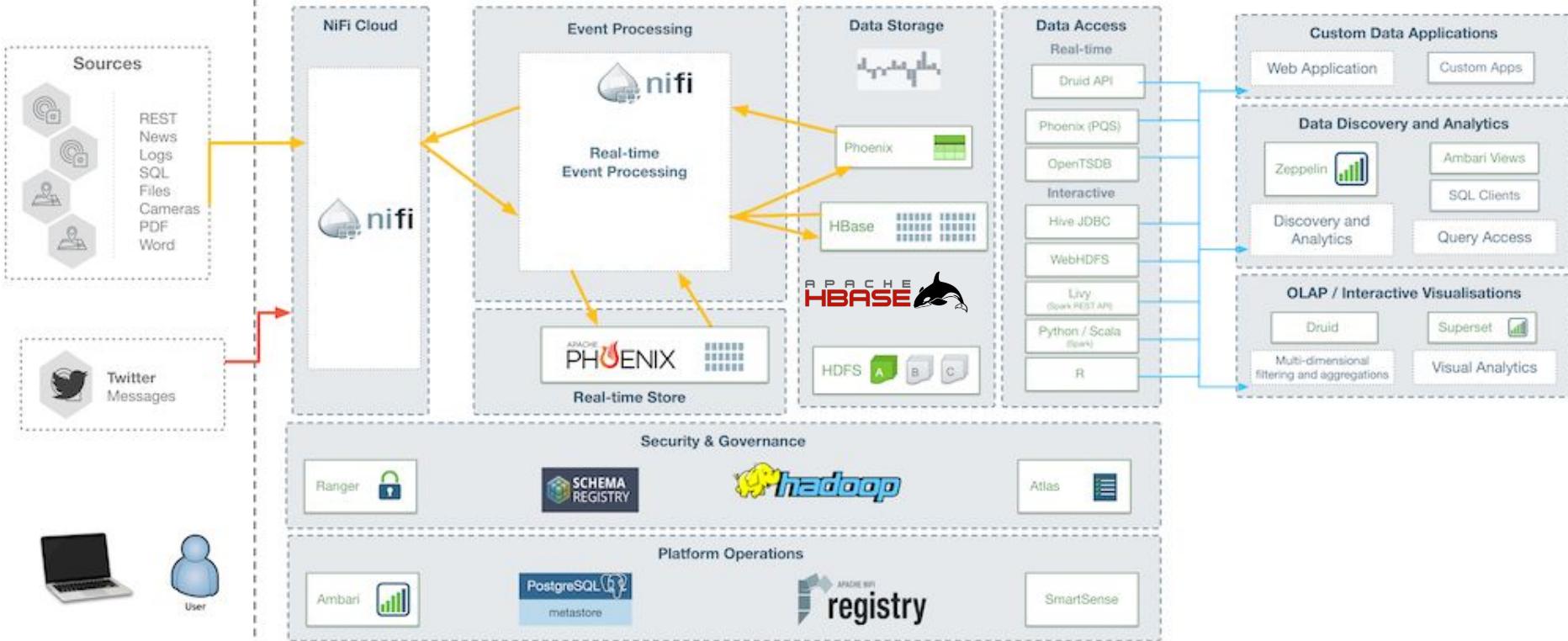


1. Phoenix-Hive connector, Phoenix Query server support are not supported

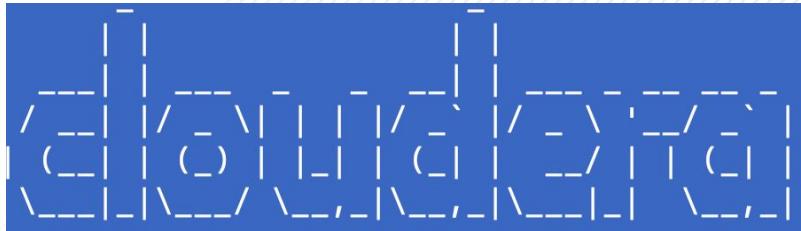
# HBase Applications & Demos

# Tracking Crime

Any Data, Any Time, Any Where, Any Workload



# None of this command line mess:



HBase Shell

Use "help" to get list of supported commands.

Use "exit" to quit this interactive shell.

For Reference, please visit: <http://hbase.apache.org/2.0/book.html#shell>

Version 2.2.0-7.0.2.0-185, rUnknown, Mon Oct 28 09:45:44 UTC 2019

Took 0.0010 seconds

hbase(main):001:0> list

TABLE

# HUE FOR SQL & DATA BROWSING FOR REMEDIATION

SQL interface using Impala or Hive for query processing

```
1 INVALIDATE METADATA;
2| SELECT * FROM my_hbase_table WHERE rowkey='abcd';
3
```

Query 47a70eda3960db:bdbef23600000000: 0% Complete (47a70eda3960db:bdbef23600000000)

Results (1)	
rowkey	value
1 abcd	my value 1

Supports SQL based insert, update,  
delete query for data in HBase



GUI based data browser

Supports search, insert, update,  
delete, DDL for HBase

# Visualization

HUE

Query

Search data and saved documents...

Jobs

</>

Home - HBase / crime

Switch Cluster

row\_key, row\_prefix\*+scan\_len [col1, family:col2, fam3:, col\_prefix\*+3, fam: col2 to col3]  crime

Filter Columns/Families  All Sort By ASC

200314096797

crime: ucr_general	crime: dispatch_date_time	crime: hour_	crime: objectid	crime: dispatch_time	crime: dispatch_date	crime: cartodb_id	crime: shape	crime:
600	2016-02-23T17:29:00Z	1	2553851	17:29:00	2016-02-23	2016294	MapRecord[ {type=Point, coordinates=[Ljava.lang.Object@#ee012b9c]} ]	01010 88E95

200435042534

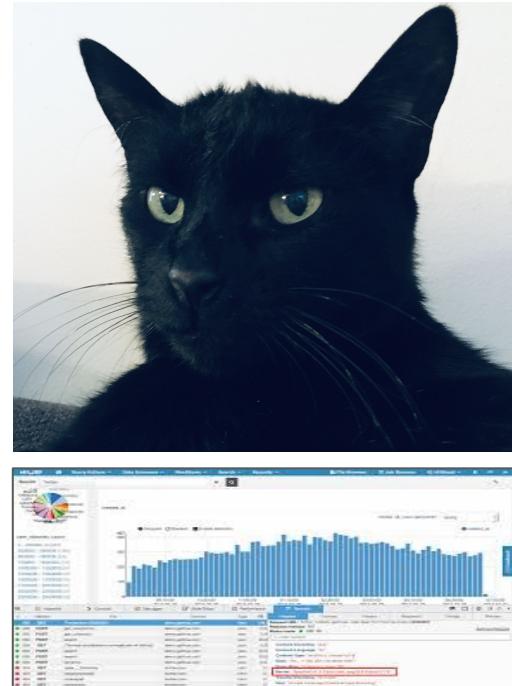
crime: ucr_general	crime: dispatch_date_time	crime: hour_	crime: objectid	crime: dispatch_time	crime: dispatch_date	crime: cartodb_id	crime: shape	crime:
600	2017-10-11T08:51:00Z	14	19811488	08:51:00	2017-10-11	2325199	MapRecord[ {type=Point, coordinates=[Ljava.lang.Object@#353aef1f]} ]	01010 02E85

201501021103

crime: ucr_general	crime: dispatch_date_time	crime: hour_	crime: objectid	crime: dispatch_time	crime: dispatch_date	crime: cartodb_id	crime: shape	crime:
400	2015-06-05T00:41:00Z	0	36262	00:41:00	2015-06-05	34151	MapRecord[ {type=Point, coordinates=[Ljava.lang.Object@#4e77dc67]} ]	01010 22ED5

201501021116

crime: ucr_general	crime: dispatch_date_time	crime: hour_	crime: objectid	crime: dispatch_time	crime: dispatch_date	crime: cartodb_id	crime: shape	crime:
2600	2015-06-05T01:36:00Z	1	66657	01:36:00	2015-06-05	57607	MapRecord[ {type=Point, coordinates=[Ljava.lang.Object@#010109]} ]	01010



# HBase CRUD



Insert New Row X

Row Key

row\_key

family:column\_name cell\_value

[+ Add Field](#)

[Cancel](#) Submit



---

# ENABLING DATA-DRIVEN APPS

---

## Fast

- Real time model serving w/ <5ms latency
- Limitless concurrency (>100M updates/sec)

## Easy

- Stream and bulk ingest & processing
- Process automation
- Consolidate multiple databases
- Schema flexibility
- SQL & NoSQL interface

## Scalable

- Multi-petabyte scale
- Unlimited Tenants

## Highly-available

- Automatic recovery from server failure
- Advanced replication & synchronization topographies
- Multiple backup methodologies

## Multi-tenant capable

- Resource isolation
- Throttling & Quotas

## Secure

- Role based access control
- Fine-grained authorizations (e.g., tenant, table, column family, cell)

# NiFi Integration with HBase

- PutHBaseRecord
- PutHBaseJSON
- PutHBaseCell
- FetchHBaseRow
- GetHBase
- ScanHBase
- DeleteHBaseRow
- DeleteHBaseCells

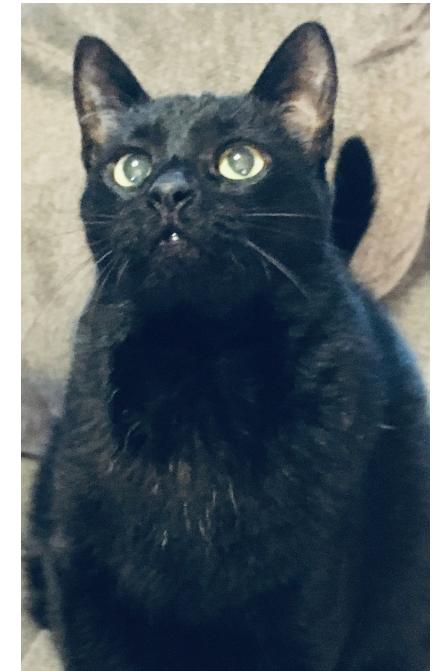
## ELT/ETL Lookup Services

- HBase\_1\_1\_2\_ListLookupService
- HBase\_2\_RecordLookupService
- HBase\_2\_ClientService
- HBase\_2\_ClientMapCacheService



<https://community.cloudera.com/t5/Community-Articles/Reading-OpenData-JSON-and-Storying-into-Phoenix-Tables/ta-p/247323>

<https://www.datainmotion.dev/2019/11/exploring-apache-nifi-110-parameters.html>



## Add Processor

Source

all groups

Displaying 11 of 297

hbase

amazon attributes  
avro aws consume  
csv database  
delete fetch get  
hadoop ingest  
insert json kafka  
logs message  
pubsub put  
record restricted  
send source text  
update

Type	Version	Tags
DeleteHBaseCells	1.10.0.2.0.0.0-85	cells, visibility, cell, delete, hbase
DeleteHBaseRow	1.10.0.2.0.0.0-85	delete, hbase
FetchHBaseRow	1.10.0.2.0.0.0-85	fetch, get, scan, enrich, hbase
GetHBase	1.10.0.2.0.0.0-85	get, hbase, ingest
PutHBaseCell	1.10.0.2.0.0.0-85	hadoop, hbase
PutHBaseJSON	1.10.0.2.0.0.0-85	json, hadoop, hbase, put
PutHBaseRecord	1.10.0.2.0.0.0-85	record, hadoop, hbase, put
ScanHBase	1.10.0.2.0.0.0-85	fetch, get, scan, hbase
StoreInKiteDataset	1.10.0.2.0.0.0-85	hive, hdfs, hadoop, kite, hbase, ...

### DeleteHBaseCells 1.10.0.2.0.0.0-85 org.apache.nifi - nifi-hbase-nar

This processor allows the user to delete individual HBase cells by specifying one or more lines in the flowfile content that are a sequence composed of row ID, column family, column qualifier and associated visibility labels if visibility labels are enabled and in use. A user-defined separator is used to separate each of these pieces of data on each line, with :: being the default separator.

CANCEL

ADD



Type ▾	Version	Tags
HBase_1_1_2_ClientMapCacheService	1.10.0.2.0.0.0-8	cluster, cache, distributed, state...
HBase_1_1_2_ClientService	1.10.0.2.0.0.0-8	client, hbase
HBase_1_1_2_ListLookupService	1.10.0.2.0.0.0-8	
HBase_1_1_2_RecordLookupService	1.10.0.2.0.0.0-8	lookup, service, record, hbase
HBase_2_ClientMapCacheService	1.10.0.2.0.0.0-8	cluster, cache, distributed, state...
HBase_2_ClientService	1.10.0.2.0.0.0-8	client, hbase
HBase_2_RecordLookupService	1.10.0.2.0.0.0-8	lookup, service, record, hbase

# HBase 2.0 on CDP Data Hub on Public Cloud

[Dashboard](#)[Environments](#)[Data Lakes](#)[User Management](#)[Data Hub Clusters](#)[Data Warehouses](#)[ML Workspaces](#)[Classic Clusters](#)

## Environments / Environments

**1 Register Environment**

- Name your environment
- Select a cloud provider
- Provide a credential for your cloud provider account

**2 Data Lake Scaling**

- Provide Data Lake name
- Choose Data Lake scale

**3 Region, Networking, Security and Storage**

- Select a region
- Add your SSH settings
- Select an existing VPC, subnet and S3 bucket
- Define Security Access

**General Information**

Environment Name\*

princeton-opdb-meetup-tspann

?

Description

Princeton Meetup OpDB HBase Demo

?

Select Cloud Provider

aws amazon

**Amazon Web Services Credential**

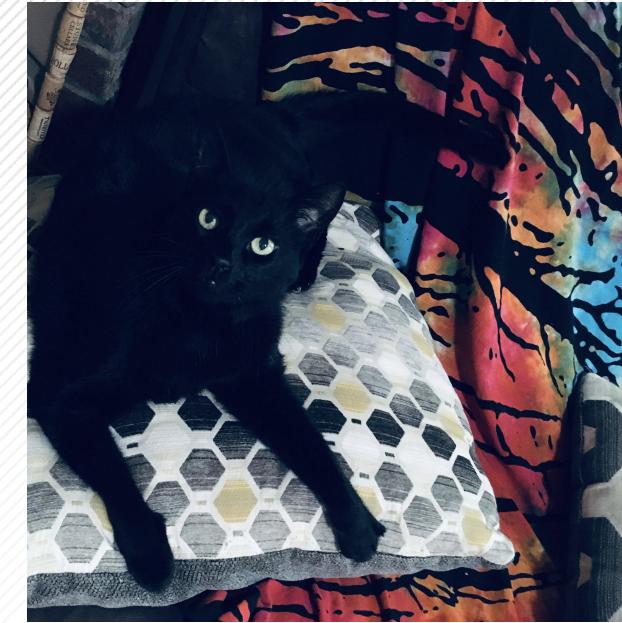
Select Credential

tspann-opdb-demo

?

[Next](#)[Help](#)

Timothy Spann



[Dashboard](#)[Environments](#)[Data Lakes](#)[User Management](#)[Data Hub Clusters](#)[Data Warehouses](#)[ML Workspaces](#)[Classic Clusters](#)

## Environments / workshop01-cdp-env / Data Lake / Event History

DATA LAKE NAME  
 sdxDATA LAKE STATUS  
 Creating Stack...REASON  
Datalake stack creation in progressNODES  
2[Atlas](#) [Ranger](#) [Data Hub Clusters](#) [Data Lake Cluster](#) [Summary](#)[RETRY](#)[REPAIR](#)[SYNC](#)

## Environment Details

NAME  
[workshop01-cdp-env](#)CREDENTIAL  
 sandbox-credREGION  
us-east-1AVAILABILITY ZONE  
us-east-1b

## Services

[Atlas](#) [CM](#) CM-UI [HBase UI](#) [Name Node](#) [Ranger](#) [Solr Server](#)

## Cloudera Manager Info

CM URL

<https://workshop01->

cloudera.site/workshop01-cdp-dl/cdp-proxy/cmf/home/

CM VERSION  
7.0.2PLATFORM VERSION  
7.0.2-1.cdh7.0.2.p0.1617729[Event History](#) [Hardware](#) [Cloud Storage](#) [Tags \(4\)](#) [Gateway](#) [Recipes \(0\)](#) [Attached clusters \(0\)](#) [Repository Details](#) [Image Details](#) [Network](#) [Telemetry](#)

## Events

[DOWNLOAD](#) Building cluster; Cluster manager ip: 

12/9/2019, 1:20:21 PM

Starting cluster services

12/9/2019, 1:15:49 PM

Mounting attached disks

12/9/2019, 1:15:31 PM

Setting up infrastructure metadata

12/9/2019, 1:13:35 PM

[Help](#)

Timothy Spann



- Dashboard
- Environments**
- Data Lakes
- User Management
- Data Hub Clusters
- Data Warehouses
- ML Workspaces
- Classic Clusters

## Cluster Templates

- Credentials
- Recipes
- Image Catalogs

LIST VIEW

RAW VIEW

**gateway** hostgroup with cardinality 1

hbase-GATEWAY-BASE

hdfs-GATEWAY-BASE

knox-KNOX-GATEWAY-BASE

**master** hostgroup with cardinality 2

hbase-GATEWAY-BASE

hbase-MASTER-BASE

hdfs-FAILOVERCONTROLLER-BASE

hdfs-GATEWAY-BASE

hdfs-JOURNALNODE-BASE

hdfs-NAMENODE-BASE

zookeeper-SERVER-BASE

**leader** hostgroup with cardinality 1

hdfs-JOURNALNODE-BASE

zookeeper-SERVER-BASE

**worker** hostgroup with cardinality 3

hbase-GATEWAY-BASE

hbase-REGIONSERVER-BASE

hdfs-DATANODE-BASE

hdfs-GATEWAY-BASE

## -azure-op ⚡

[TERMINATE](#)[STOP](#)[ACTIONS ▾](#)

STATUS

✓ Running

NODES

7

CREATED AT

12/05/19, 06:54 PM EST

CLUSTER TEMPLATE

 CDP 1.2 - Operational Database: Apache HBase Environment Details

NAME

pvidal-azure-env

DATA LAKE

 azure-dl

CREDENTIAL

 az-cred

REGION

East US

AVAILABILITY ZONE

N/A

 Services CM UI HBase UI HBase UI Name Node Name Node Cloudera Manager Info

CM URL

<https://cloudera.site/pvidal-azure-op/cdp-proxy/cmf/home/>

CM VERSION

7.0.2

PLATFORM VERSION

7.0.2-1.cdh7.0.2.p0.1617729

[Event History](#) [Hardware](#) [Cloud Storage](#) [Tags \(4\)](#) [Gateway](#) [Recipes \(0\)](#) [Repository Details](#) [Image Details](#) [Network](#)

Search

[Status](#)   [All Health Issues](#)   [Configuration](#) 14 ▾   [All Recent Commands](#) [Clusters](#) [Hosts](#) [Diagnostics](#) [Charts](#)

Last Updated: Dec 9, 5:24:57 PM UTC



## Compute Clusters

Status ↑	Name	Linked Data Context	Issues	Hosts	Services	Version	⋮
	opdb	pvidal-azure-dl (pvidal-azure-dl (Proxy))	9	7 Hosts	4 Services ▾	7.0.2 (Parcels)	

## Other

Status	Name	Issues
	MGMT	4

1 - 1 of 1

Search

## Clusters

Hosts

Diagnostics

Charts

Status Health Issues Configuration 9 ▾

## Status

Data Context: pvidal-azure-dl

- ranger
- atlas
- knox
- hive

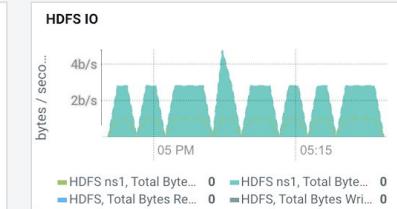
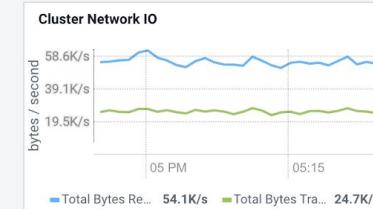
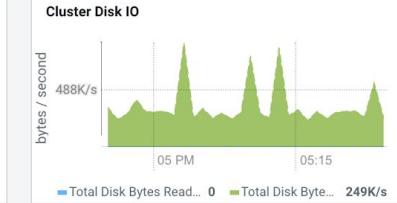
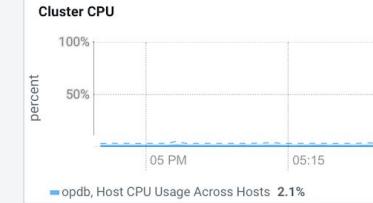
1

Compute Cluster, Cloudera Runtime 7.0.2 (Parcels)

- 7 Hosts 5
- HBase
- HDFS 4
- Knox
- ZooKeeper
- pvidal-azure-dl

## Charts

30m 1h 2h 6h 12h 1d 7d 30d ⚙





Search

Clusters

Hosts

Diagnostics

Audits

Charts

Replication

Administration

Parcels

Recent Commands

Support

admin

operations / second

## Compactions: Average, Max, Min

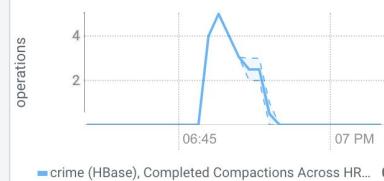


Chart showing the average, maximum, and minimum compaction rate for regions in the table.

## Compactions Distribution

last(moving\_avg(integral(compactions\_completed\_rate), 300.0))

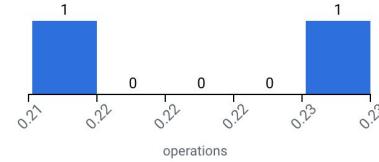


Chart showing the distribution of compactions done across regions in the table. Click the chart to drill down and find outliers.

## Bytes Compacted: Average, Max, Min

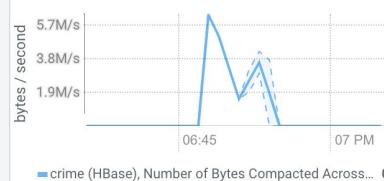


Chart showing the average, maximum, and minimum bytes compacted for regions in the table.

## Bytes Compacted Distribution

last(moving\_avg(num\_bytes\_compacted\_rate, 300.0))

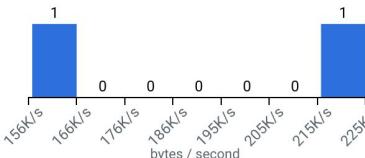


Chart showing the distribution of the amount of bytes compacted across regions in the table. Click the chart to drill down and find outliers.

## Region Size: Average, Max, Min

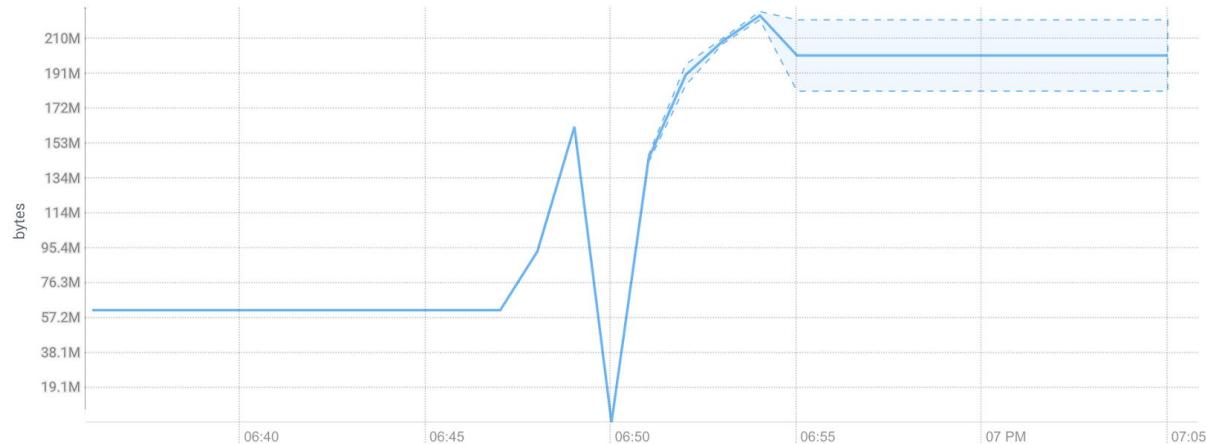
X

Query      SELECT storefiles\_size\_across\_hregions WHERE entityName = "hbase:crime" AND category = HTABLE

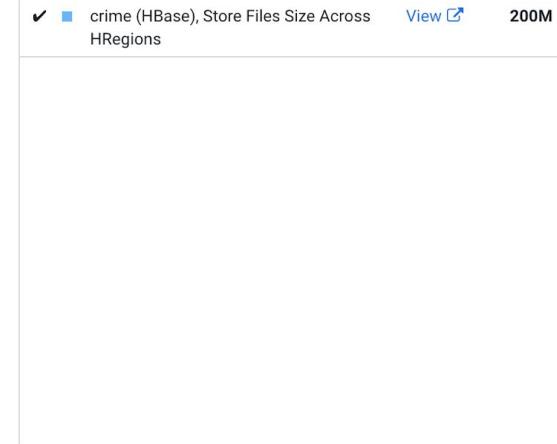
Data Granularity

Auto

30 minutes preceding Dec 9, 7:05 PM UTC
30m 1h 2h 6h 12h 1d 7d 30d



Legend Details Distribution



## Master

-a.a465-9q4k.cloudera.site

## Region Servers

Base Stats	Memory	Requests	Storefiles	Compactions	Replications	ServerName	Start time	Last contact	Version	Requests Per Second	Num. Regions	
						.cloudera.site,16020,1575597009057	Fri Dec 06 01:50:09 UTC 2019	1 s	2.2.0.7.0.2.0-212	0	1	
						.cloudera.site,16020,1575597014543	Fri Dec 06 01:50:14 UTC 2019	0 s	2.2.0.7.0.2.0-212	0	1	
						.cloudera.site,16020,1575597009331	Fri Dec 06 01:50:09 UTC 2019	1 s	2.2.0.7.0.2.0-212	0	1	
Total:3										0	3	

## Backup Masters

ServerName	Port	Start Time
.cloudera.site	16000	Fri Dec 06 01:50:08 UTC 2019
Total:1		

BLOOMFILTER	ROW
CACHE_INDEX_ON_WRITE	false
IN_MEMORY	false
CACHE_BLOOMS_ON_WRITE	false
PREFETCH_BLOCKS_ON_OPEN	false
COMPRESSION	NONE
BLOCKCACHE	true
BLOCKSIZE	65536

## Table Regions

Name(2)	Region Server	ReadRequests (0)	WriteRequests (219,385)	StorefileSize (399 MB)	Num.Storefiles (4)	MemSize (6 MB)	Locality	Start Key	End Key
crime,,1575917310002.7f5fc3b255f84cb2e65d7c5425aaddac.		0	110,659	180 MB	1	2 MB	1.0		201715004509
crime,201715004509,1575917310002.551a2c2c761073a41e57031b2682164b.		0	108,726	219 MB	3	4 MB	1.0	201715004509	

## Regions by Region Server

Region Server	Region Count
16030	2



1 / 1

1

330,790 / 217.69 MB

0

0

7

12

0

0

0

0

0

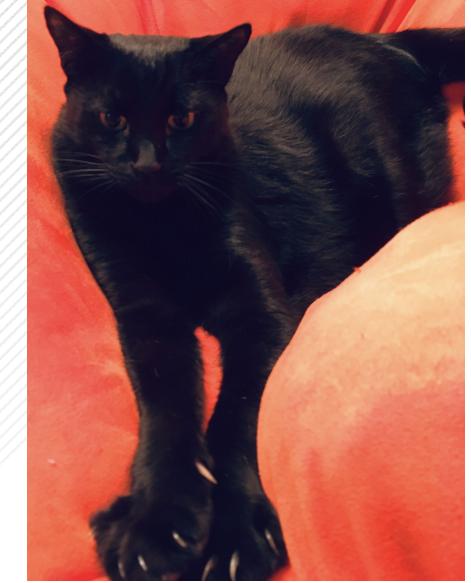
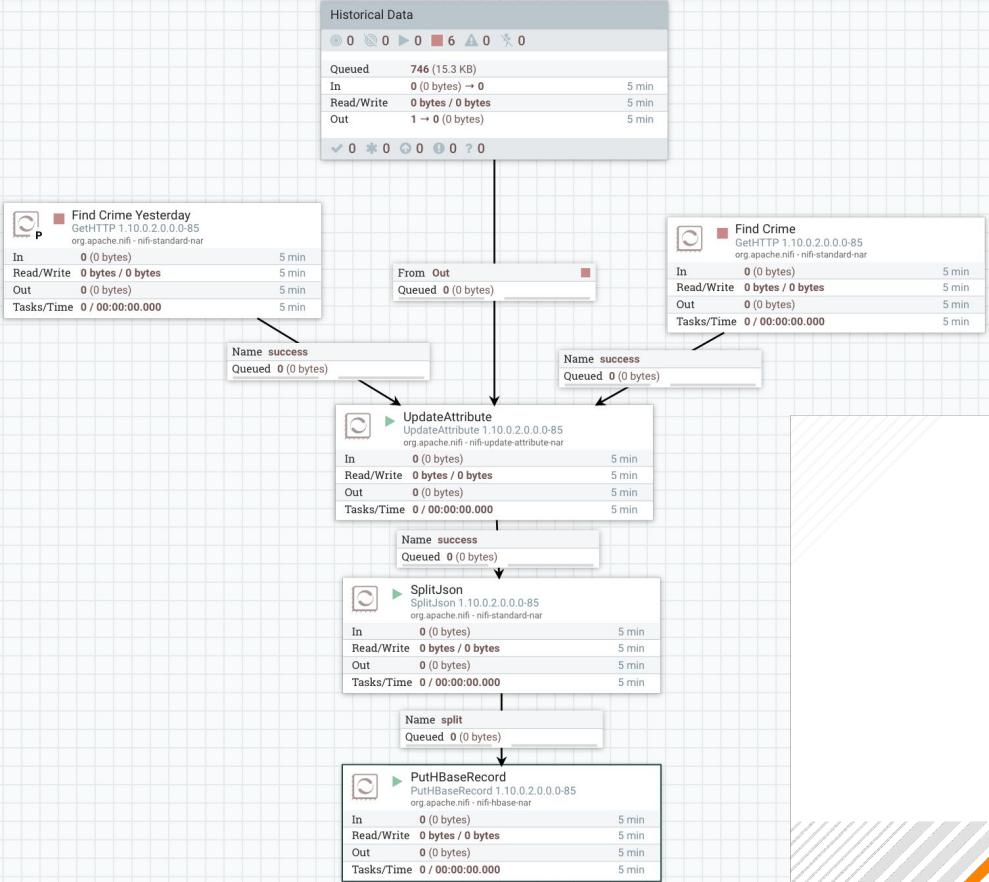
## Navigate



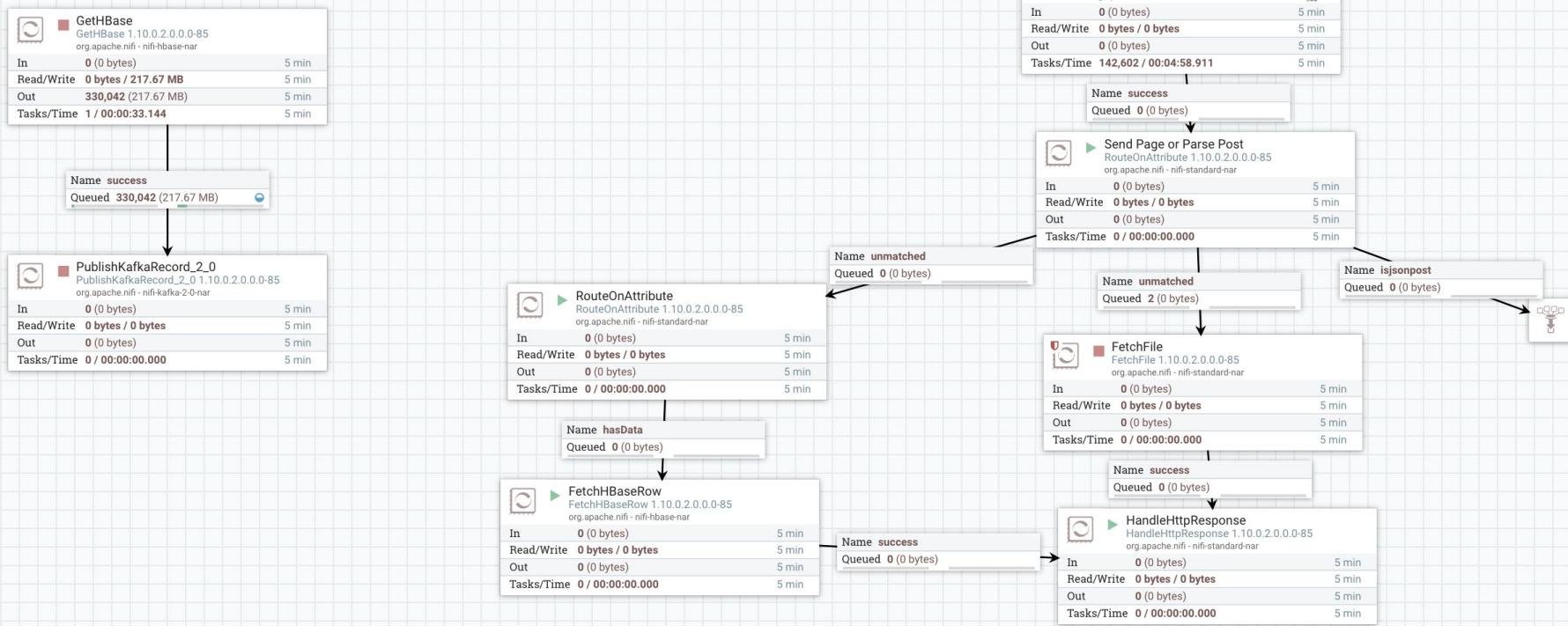
## Operate

PutHBaseRecord  
Processor

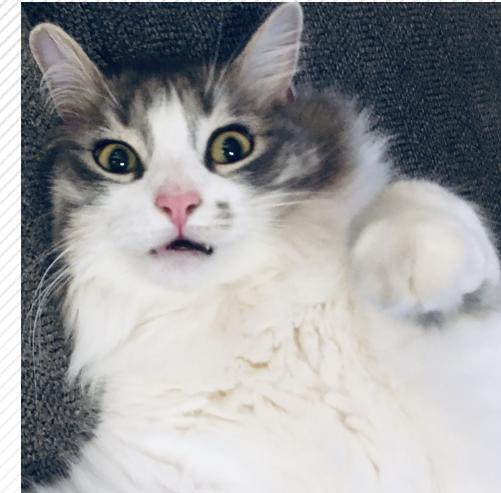
578ea0f7-c4a9-3cba-a8c6-84b4f6fb50a0



## READ HBASE



Property	Value	
HBase Client Service	?	<b>HBase_1_1_2_ClientService</b>
Table Name	?	<b>crime</b>
Row Identifier	?	<b> \${http.param.key} </b>
Columns	?	No value set
Authorizations	?	No value set
Destination	?	<b>flowfile-content</b>
JSON Format	?	<b>col-qual-and-val</b>
JSON Value Encoding	?	<b>none</b>
Encode Character Set	?	<b>UTF-8</b>
Decode Character Set	?	<b>UTF-8</b>



# Configure Processor

■ Stopped

SETTINGS

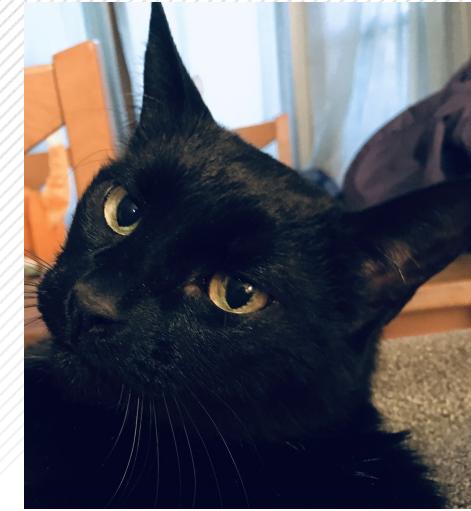
SCHEDULING

PROPERTIES

COMMENTS

Required field

Property	Value	
HBase Client Service	?	<b>HBase_1_1_2_ClientService</b>
Distributed Cache Service	?	No value set
Table Name	?	<b>crime</b>
Columns	?	No value set
Authorizations	?	No value set
Filter Expression	?	No value set
Initial Time Range	?	<b>None</b>
Character Set	?	<b>UTF-8</b>



# Configure Processor

Stopped

SETTINGS	SCHEDULING	PROPERTIES	COMMENTS
----------	------------	------------	----------

Required field

Property	Value
Record Reader	JsonTreeReader
HBase Client Service	HBase_2_ClientService
Table Name	crime
Row Identifier Field Name	dc_key
Row Identifier Encoding Strategy	String
Null Field Strategy	Skip Field
Column Family	crime
Default Visibility String	No value set
Visibility String Record Path Root	No value set
Timestamp Field Name	No value set
Batch Size	100
Complex Field Strategy	Text



<https://github.com/tspannhw/HBase2>



METRICS DATA EXPLORER CONFIGS LATENCY

DESERIALIZER: Keys: String Values: String



FROM OFFSET				TO OFFSET				
Partition 0	3616	0	605	1210	1815	2420	3025	3631
Offset	Timestamp	Key	Value					
3625	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042344", "cells": {"crime:cartodb_id": "2292950", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3624	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042342", "cells": {"crime:cartodb_id": "2254311", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3623	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042335", "cells": {"crime:cartodb_id": "2279410", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3622	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042327", "cells": {"crime:cartodb_id": "2281223", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3621	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042324", "cells": {"crime:cartodb_id": "2284400", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3620	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042314", "cells": {"crime:cartodb_id": "2290592", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3619	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042309", "cells": {"crime:cartodb_id": "2290930", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3618	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042303", "cells": {"crime:cartodb_id": "2276126", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3617	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042276", "cells": {"crime:cartodb_id": "2276191", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					
3616	Sun, Dec 08 2019, 5:43:50	null	{"row": "201739042269", "cells": {"crime:cartodb_id": "2284145", "crime:dc_dist": "39", "crime:dispatch_date": "2017-06-04", "crime:dispatch_date_time": "2017-0... show more					

# SPRING BOOT APPLICATION TO PHOENIX

Recent Philadelphia Police Reports:	
2016-08-28 00:11:00	-0400
Robbery Firearm - 1300 BLOCK S 08TH ST	
2016-08-28 00:41:00	-0400
Robbery Firearm - 6600 BLOCK E ROOSEVELT BLVD	
2016-08-28 00:27:00	-0400
Aggravated Assault Firearm - 1300 BLOCK W ERIE AV	
2016-08-28 00:56:00	-0400
Recent Philadelphia Police Tweets:	
2016-08-29 08:04:30	-0400
Wanted: Suspects for Commercial Robbery in the 2nd District [VIDEO] <a href="https://t.co/yiiU9dyxI">https://t.co/yiiU9dyxI</a>	
<a href="https://t.co/lQ8o9sk0Bd">https://t.co/lQ8o9sk0Bd</a>	
2016-08-26 07:53:05	-0400
FBI/PPD Violent Crimes Task Force Search for Bank Robbery Suspect in the 8th District	
<a href="https://t.co/elyVt4OqW">https://t.co/elyVt4OqW</a>	
<a href="https://t.co/pNHZaTmeXc">https://t.co/pNHZaTmeXc</a>	

Table: PHILLYCRIME  
phoenixcluster//TABLE/PHILLYCRIME

Actions ▾

Info Columns Data Row Count Primary Key Indexes Grants Row Id References

Filter:

*	DC_DIST	DC_KEY	DISPATCH_DATE	DISPATCH_DATE_TIME	DISPATCH_TIME	HOUR	LOCATION_BLOCK	J	PSA	TEXT_GENERAL_CODE
824	39	ZUUV39V11863	ZUUV-U-C-23	ZUUV-U-C-23 23:00:00	Z3:20:UU	23	ZUUV BLOCK W PACIFIC ST	J	I	nets
825	39	200839014429	2008-03-05	2008-03-05T10:17:00.000	10:17:00	10	1400 BLOCK W CLEARFIELD SARFIELD ST	B	All Other Offenses	
826	39	200839018143	2008-03-20	2008-03-20T15:25:00.000	15:25:00	15	3100 BLOCK N PATTON ST	N	Fraud	
827	39	200839020902	2008-04-01	2008-04-01T07:56:00.000	07:56:00	7	2800 BLOCK W SOMERSET STRETT ST	K	Theft from Vehicle	
828	39	200839034061	2008-05-20	2008-05-20T19:40:00.000	19:40:00	19	5200 BLOCK KEYSER	T	Narcotic / Drug Law Violations	
829	39	200839057934	2008-08-13	2008-08-13T06:40:00.000	06:40:00	6	4900 BLOCK KEYSER ST	U	Arson	
830	39	200839059552	2008-08-18	2008-08-18T18:21:00.000	18:21:00	18	2800 BLOCK N LAMBERT ST	G	Thefts	
831	39	200839068846	2008-09-22	2008-09-22T04:07:00.000	04:07:00	4	5400 BLOCK WISSAHICKON AVE	P	Fraud	
832	39	200839070385	2008-09-29	2008-09-29T09:53:00.000	09:53:00	9	2900 BLOCK CAMBRIA ST	K	Burglary Non-Residential	
833	39	200839076821	2008-10-27	2008-10-27T21:24:00.000	21:24:00	21	300 BLOCK ZEPHYRA ST	S	Fraud	
834	39	200839082186	2008-11-21	2008-11-21T11:41:00.000	11:41:00	11	N 2600 SILVER ST	K	All Other Offenses	
835	39	200839084598	2008-12-02	2008-12-02T19:03:00.000	19:03:00	19	N 15TH ST & ST LUKES ST	F	Recovered Stolen Motor Vehicle	
836	39	200839087959	2008-12-17	2008-12-17T11:58:00.000	12:58:00	12	3500 BLOCK INDIAN QUEEN L	Q	Thefts	
837	39	200839088972	2008-12-21	2008-12-21T16:16:00.000	16:16:00	16	200 BLOCK WINONA ST	V	Fraud	
838	39	200839090909	2008-12-30	2008-12-30T10:09:00.000	10:09:00	10	2400 BLOCK W INDIANA AVE	H	Rape	
839	77	20087700049	2008-01-03	2008-01-03T09:48:00.000	09:48:00	9	HERTZ	A	Fraud	
840	77	200877000147	2008-01-06	2008-01-06T13:46:00.000	13:46:00	13	C-18	A	Thefts	
841	77	200877000711	2008-01-24	2008-01-24T12:49:00.000	12:49:00	12	US AIRWAY POS#14	A	Thefts	
842	77	200877000719	2008-01-24	2008-01-24T16:44:00.000	16:44:00	16	NATIONAL RENT-A-CAR	A	Other Assualts	
843	77	200877000750	2008-01-26	2008-01-26T16:58:00.000	16:58:00	16	A-WEST CHECKPOINT	A	Narcotic / Drug Law Violations	
844	77	200877001313	2008-02-12	2008-02-12T11:08:00.000	15:08:00	15	TERMINAL F	A	Thefts	
845	77	200877001413	2008-02-16	2008-02-16T18:46:00.000	18:46:00	18	B TERMINAL DUNKIN DONUTS	A	Other Assualts	
846	77	200877005230	2008-05-03	2008-05-03T15:59:00.000	15:59:00	15	0 BLOCK PIA GATE B-2	A	Thefts	
847	77	20087700549	2008-05-09	2008-05-09T13:51:00.000	13:51:00	13	F-16 BREAK RM	A	Other Assualts	
848	77	200877005711	2008-05-16	2008-05-16T17:05:00.000	16:05:00	16	B/C CONNECTOR	A	Embezzlement	
849	77	200877008014	2008-07-24	2008-07-24T09:59:00.000	09:59:00	9	PIA HERTZ RENTAL	A	Fraud	
850	77	200877008867	2008-08-23	2008-08-23T13:26:00.000	13:26:00	13	EMPLOYEE LOT	A	Other Assualts	
851	77	20087700945	2008-08-31	2008-08-31T17:16:00.000	16:15:00	16	CAR RENTAL RETURN	A	Theft from Vehicle	
852	77	200877009653	2008-09-30	2008-09-30T18:14:00.000	18:14:00	18	PIA CAR RENTAL RD	A	Fraud	

Max Rows: 1000 Max Chars: -1

0.786/1.589 sec | 1000/10 824-852

<https://community.cloudera.com/t5/Community-Articles/Creating-a-Spring-Boot-Java-8-Microservice-To-Read-Apache/ta-p/247379>

<https://github.com/tspannhw/phillycrime-springboot-phoenix>

# Configure Processor

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property	Value
Search Value	?(?s:^.*\$)
Replacement Value	upsert into phillycrime values ('\${'dc_dist'}','\${'dc_key'}','\${'dispatch_d...')
Character Set	UTF-8
Maximum Buffer Si...	1 MB
Replacement Strateg...	Always Replace
Evaluation Mode	Entire text



TH<sup>O</sup>N<sup>G</sup> Y<sup>U</sup>

