Tracking Crime as It Occurs with Apache Phoenix, Apache HBase and Apache NiFi

TIMOTHY SPANN
Field Engineer, Data in Motion
Cloudera

Introduction

Tim Spann has been running meetups in Princeton on Big Data technologies since 2015.

Tim has spoken at many international conferences on Apache NiFi, Deep Learning and Streaming.

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

https://community.hortonworks.com/users/9304/tspann.html

https://dzone.com/users/297029/bunkertor.html

https://dzone.com/articles/integrating-keras-tensorflow-yolov3-into-apache-ni

Introduction

Using Apache NiFi we can ingest various sources of criminal data real-time as activities happen as well as monitor live traffic cameras (Source: TrafficLand).

We can do a lot of alerting, routing and react to crime data as it arrives, but we need more. We need to update totals, store this data for future machine learning analytics and make it available for instant update dashboards and reports.

The best destination for this data is Apache HBase and Apache Phoenix. We'll populate tables with ease and speed!

Resources:

https://community.hortonworks.com/articles/54947/reading-opendata-json-and-storing-into-phoenix-tab.html

https://community.hortonworks.com/articles/56642/creating-a-spring-boot-java-8-microservice-to-read.html

https://community.hortonworks.com/articles/64122/incrementally-streaming-rdbms-data-to-your-hadoop.html

DATAFLOW

Cloudera DataFlow Data-in-Motion Platform

EDGE DATA MANAGEMENT

Edge data collection, routing and monitoring

- Apache MiNiFi
- Edge Flow Manager

FLOW MANAGEMENT

Enterprise data ingestion, transformation and enrichment

- Apache NiFi
- NiFi Registry

STREAM PROCESSING

Real-time streams processing at IoT scale

- Apache Kafka
- Streams Messaging Manager

STREAMING ANALYTICS

Predictive analytics and real-time insights

- Apache Storm
- Streaming Analytics Manager
- Kafka Streams

ENTERPRISE SERVICES

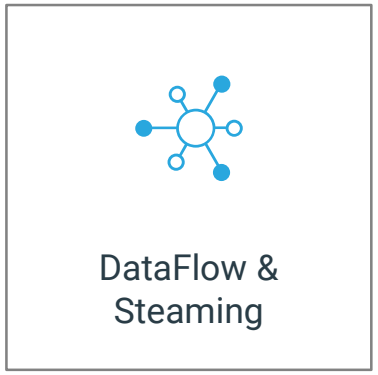
Provisioning, Management and Monitoring

- Unified Security
- Edge-to-Enterprise Governance
- Single Sign-on

- Schema Registry

CONTROL DATA-IN-MOTION FROM EDGE-TO-ENTERPRISE

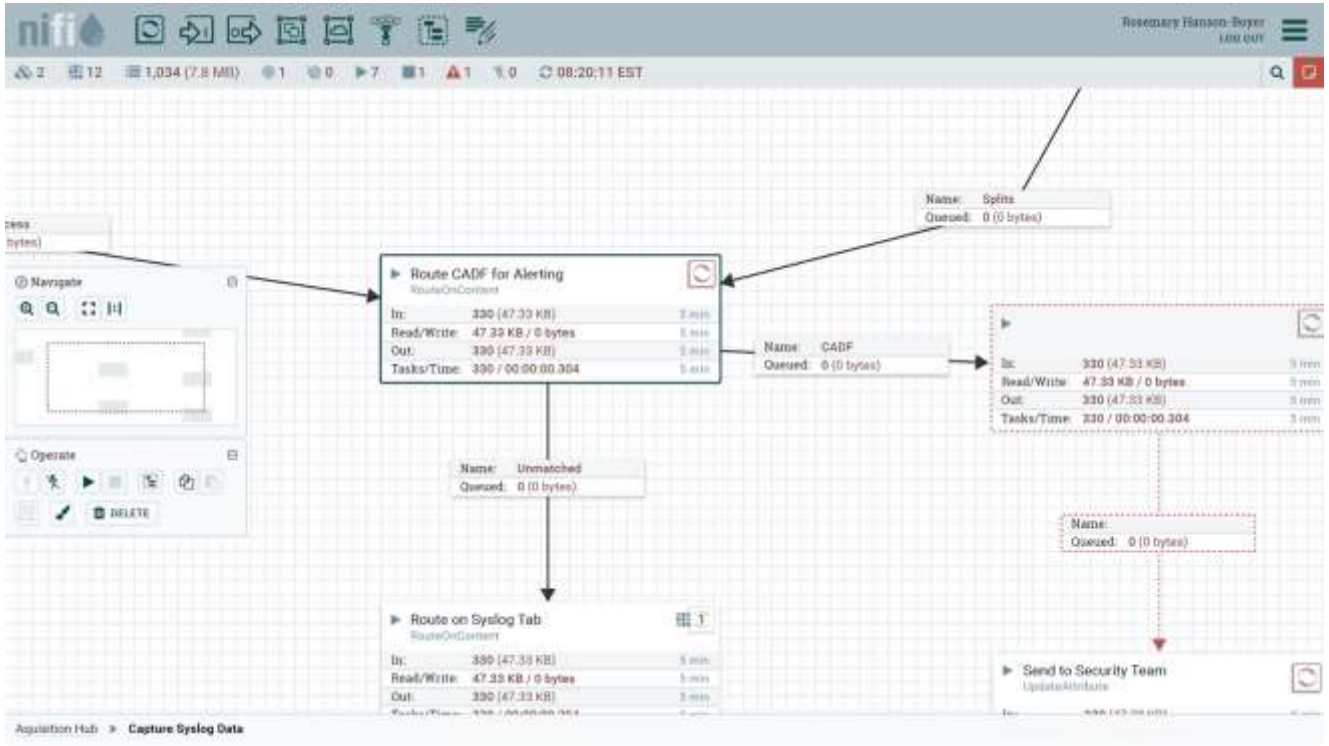
Cloudera DataFlow - Collect, Curate and Analyze Data-in-Motion



- Edge-to-enterprise streaming data platform for management, security and governance of real-time streaming data
- Edge data collection, processing and content routing of sensor data from edge devices
- Continuous data ingestion from any streaming source or IoT device
- Ease-of-use in building sophisticated data flows with drag-and-drop user interface
- Real-time stream processing and content syndication at the scale of millions of messages per second
- Predictive and prescriptive analytics from streaming analytics engines to gain actionable intelligence

CLOUDERA FLOW MANAGEMENT

- Web-based user interface
- Highly configurable
- Out-of-the-box data provenance
- Designed for extensibility
- Secure
- NiFi Registry
  - DevOps support
  - FDLC
  - Versioning
  - Deployment



300+ PROCESSORS FOR DEEPER ECOSYSTEM INTEGRATION

- FTP
- SFTP
- HL7
- UDP
- XML
- HTTP
- WebSocket
- Email
- HTML
- Image
- Syslog
- AMQP

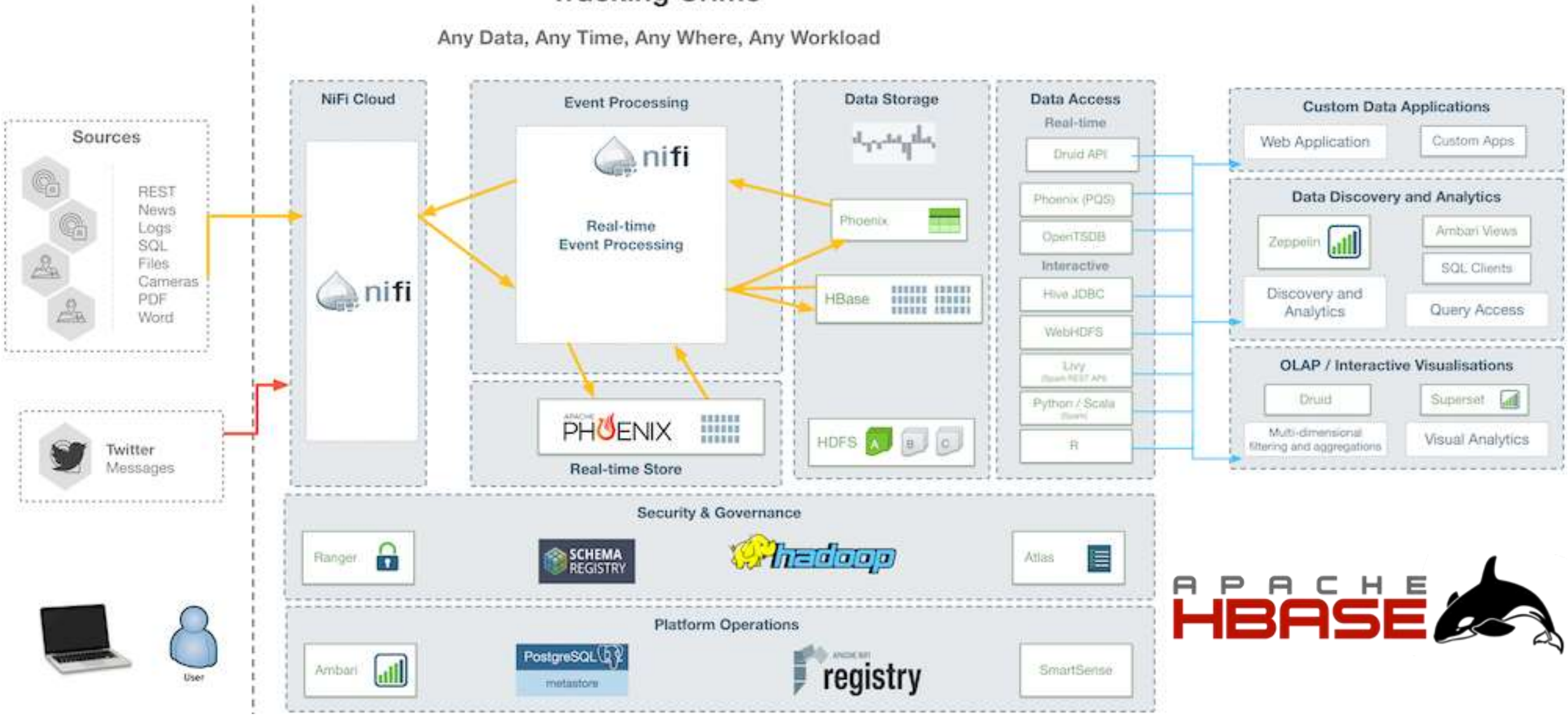
- Hash
- Encrypt
- GeoEnrich
- Merge
- Tail
- Scan
- Extract
- Evaluate
- Replace
- Duplicate
- Execute
- Translate
- Split
- Fetch
- Convert

- Route Text
- Distribute Load
- Route Content
- Generate Table Fetch
- Route Context
- Jolt Transform JSON
- Control Rate
- Prioritized Delivery

ARCHITECTURE

Tracking Crime

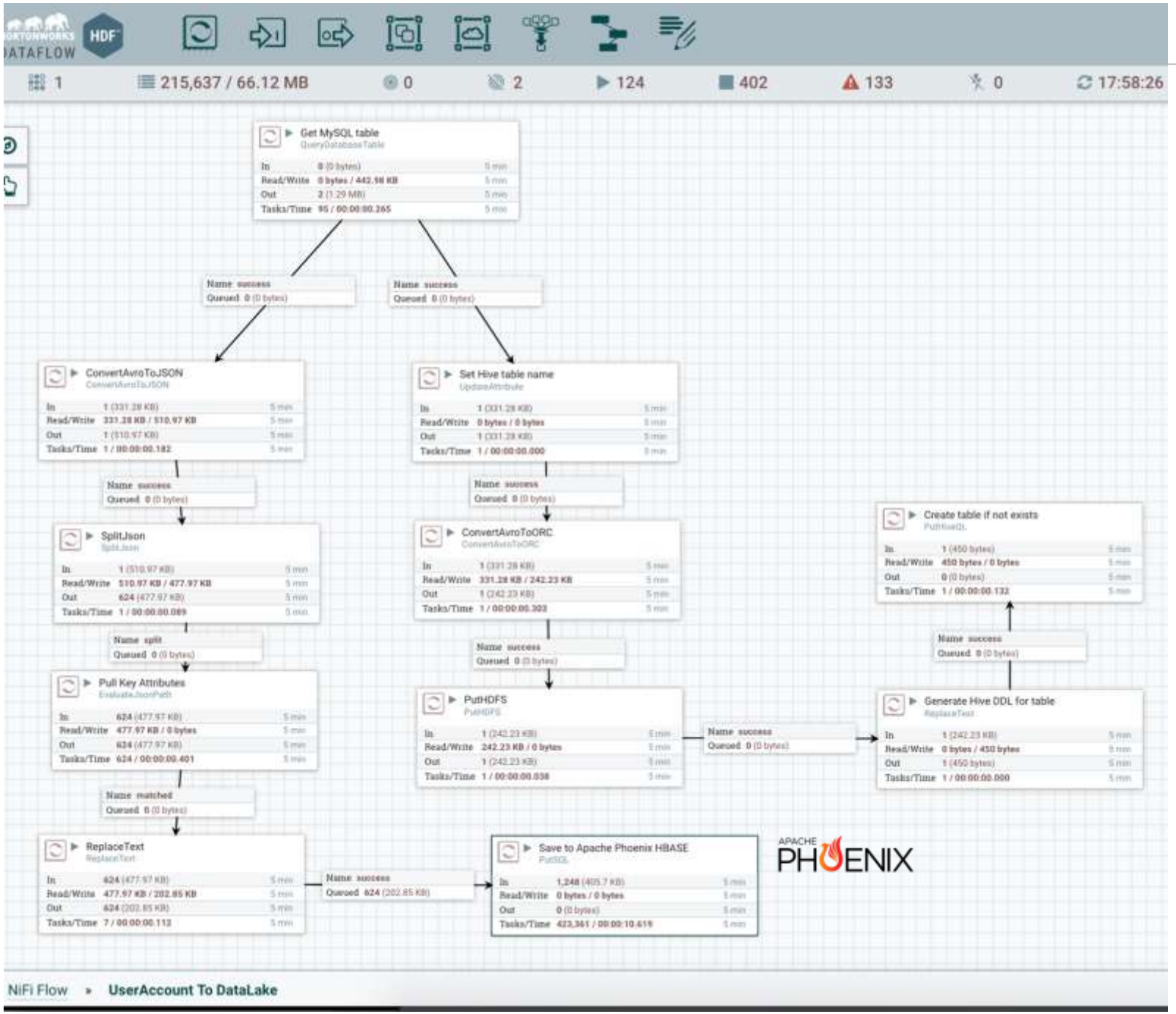
Any Data, Any Time, Any Where, Any Workload

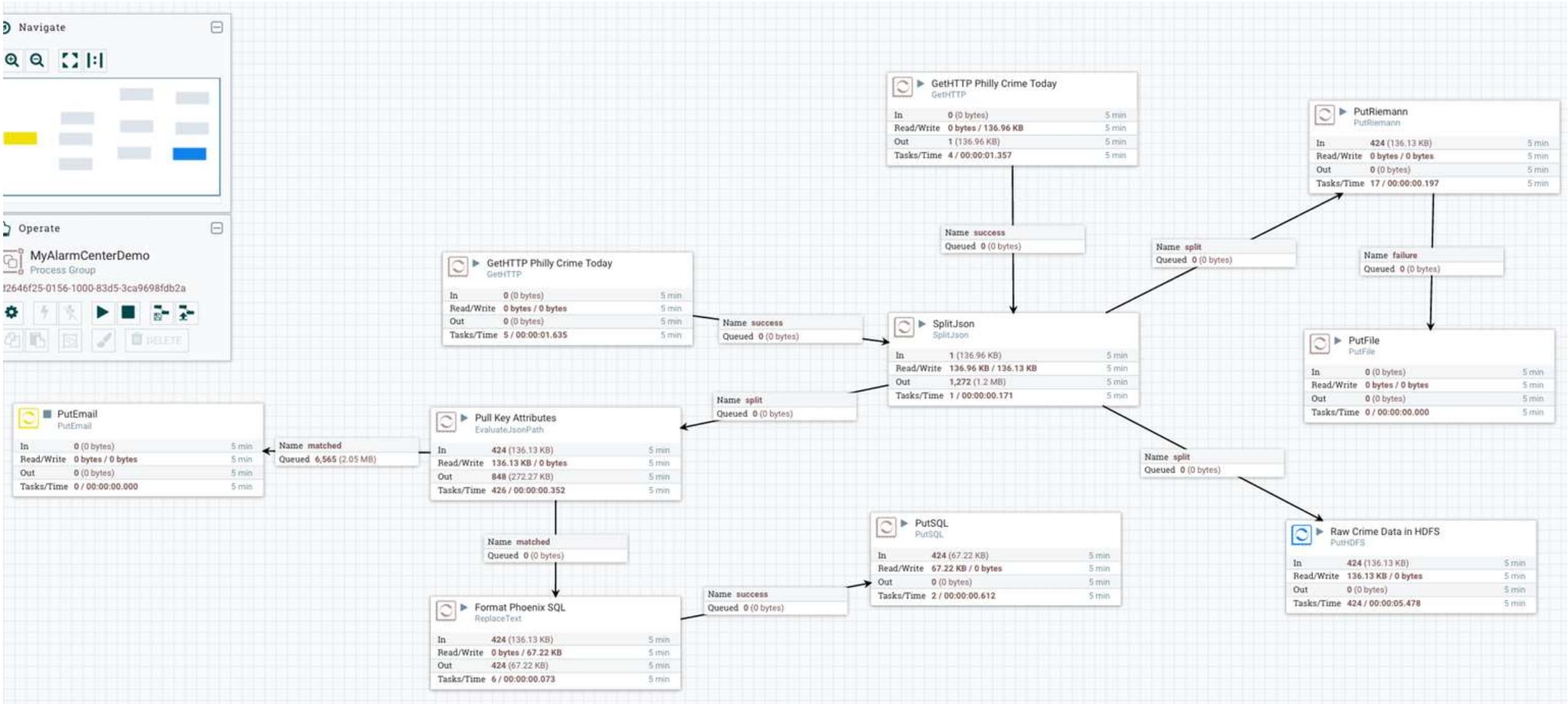


Apache Phoenix-5.0

- Expect similar timeframe for Phoenix-5.0
- We are working for HBase-2.0 support
- Re-write internals using **Apache Calcite**
- SQL-parser, planner and optimizer
- Cost based Optimizer used by Hive, Drill, etc
- Pluggable rules with default rules, and Phoenix specific ones
- SQL-92 support
- Apache NiFi calls Apache Calcite Avatica JDBC

DEMO





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 Strate...	Always Replace
Evaluation Mode	Entire text

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

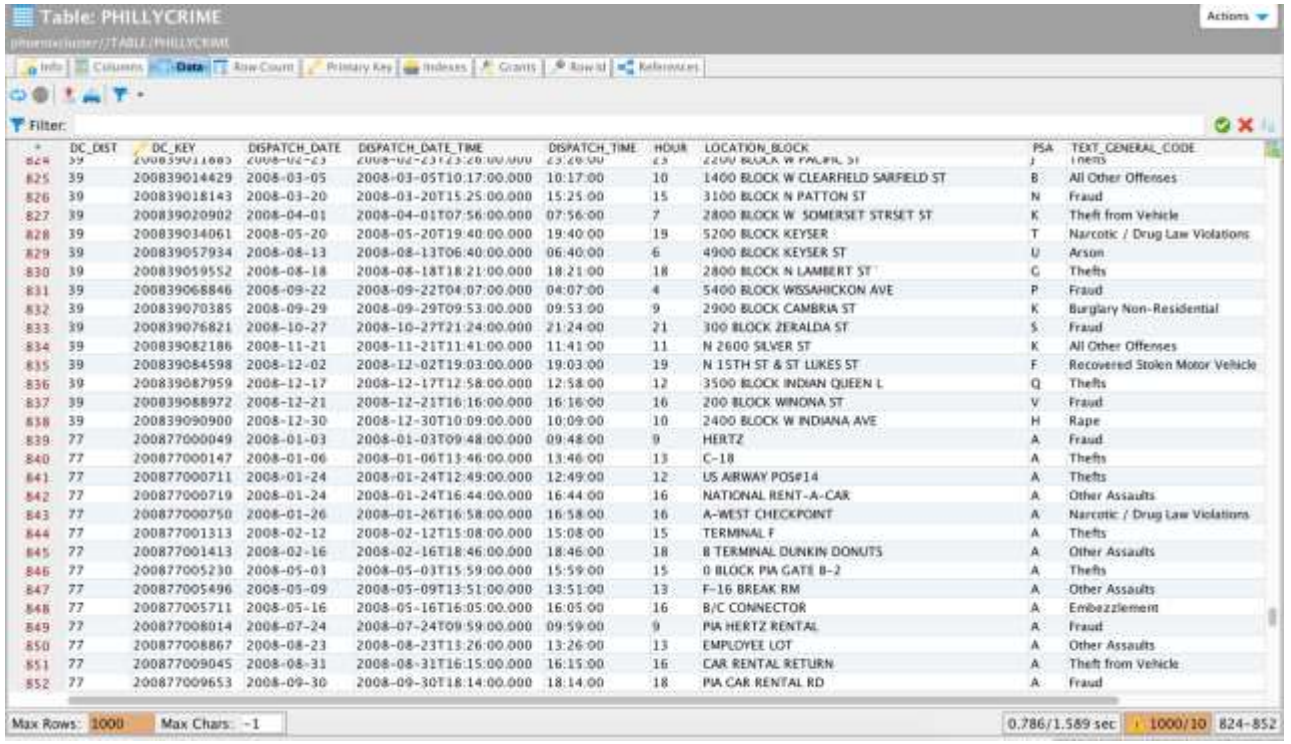
2016-08-28 00:56:00 -0400

Recent Philadelphia Police Tweets:
2016-08-29 08:04:30 -0400

Wanted: Suspects for Commercial Robbery in the 2nd District [VIDEO] https://t.co/yiilU9dyxl https://t.co/5Q8o9sk0Bd

2016-08-26 07:53:05 -0400

FBI/PPD Violent Crimes Task Force Search for Bank Robbery Suspect in the 8th District https://t.co/elyVVT4OqW https://t.co/pNHZaTmeXc



	DC_DIST	DC_KEY	DISPATCH_DATE	DISPATCH_DATE_TIME	DISPATCH_TIME	HOUR	LOCATION_BLOCK	PSA	TEXT_GENERAL_CODE
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 STRSET 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 ZERALDA 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	2008-12-17T12:58:00.000	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	200839090900	2008-12-30	2008-12-30T10:09:00.000	10:09:00	10	2400 BLOCK W INDIANA AVE	H	Rape
839	77	200877000049	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 Assaults
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	2008-02-12T15:08:00.000	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 Assaults
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	200877005496	2008-05-09	2008-05-09T13:51:00.000	13:51:00	13	F-16 BREAK RM	A	Other Assaults
848	77	200877005711	2008-05-16	2008-05-16T16:05:00.000	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 Assaults
851	77	200877009045	2008-08-31	2008-08-31T16:15:00.000	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

https://community.hortonworks.com/articles/56642/creating-a-spring-boot-java-8-microservice-to-read.html

https://github.com/tspannhw/phoenix

SPRING BOOT APPLICATION TO PHOENIX TABLE

```
CREATE TABLE phillycrime (dc_dist varchar,
dc_key varchar not null primary key,dispatch_date
  varchar,dispatch_date_time varchar,dispatch_time varchar,hour
  varchar,location_block varchar,psa varchar,
text_general_code varchar,ucr_general varchar);
```

```
@RequestMapping("/query/{query}")
```

```
java -Xms512m -Xmx2048m -Dhdp.version=3.1 -
Djava.net.preferIPv4Stack=true -jar target/phoenix-0.0.1-SNAPSHOT.jar
```

```
1 row selected (0.153 seconds)
0: jdbc:phoenix:localhost:2181:/hbase-unsecur> delete from phillycrime;
1 row affected (0.011 seconds)
0: jdbc:phoenix:localhost:2181:/hbase-unsecur> select * from phillycrime;
+-----------+--------------+----------------+---------------------+----------+
| DC_DIST   |    DC_KEY    | DISPATCH_DATE  | DISPATCH_DATE_TIME  |    D     |
+-----------+--------------+----------------+---------------------+----------+
|           | 201601036811 |                |                     |          |
|           | 201603048626 |                |                     |          |
|           | 201603048633 |                |                     |          |
|           | 201614069127 |                |                     |          |
|           | 201614069141 |                |                     |          |
|           | 201615083144 |                |                     |          |
|           | 201624083193 |                |                     |          |
|           | 201624083194 |                |                     |          |
|           | 201625072218 |                |                     |          |
|           | 201625072227 |                |                     |          |
|           | 201626043171 |                |                     |          |
|           | 201639069963 |                |                     |          |
|           | 201639069964 |                |                     |          |
+-----------+--------------+----------------+---------------------+----------+
13 rows selected (0.161 seconds)
0: jdbc:phoenix:localhost:2181:/hbase-unsecur> delete from phillycrime;
13 rows affected (0.014 seconds)
0: jdbc:phoenix:localhost:2181:/hbase-unsecur> select * from phillycrime;
+-----------+--------------+----------------+---------------------+----------+
| DC_DIST   |    DC_KEY    | DISPATCH_DATE  | DISPATCH_DATE_TIME  |    D     |
+-----------+--------------+----------------+---------------------+----------+
| 01        | 201601036811 | 2016-08-28     | 2016-08-28T00:42:00.000 | 00:42:00 |
| 03        | 201603048626 | 2016-08-28     | 2016-08-28T00:11:00.000 | 00:11:00 |
| 03        | 201603048633 | 2016-08-28     | 2016-08-28T00:55:00.000 | 00:55:00 |
| 14        | 201614069127 | 2016-08-28     | 2016-08-28T00:57:00.000 | 00:57:00 |
| 14        | 201614069141 | 2016-08-28     | 2016-08-28T01:25:00.000 | 01:25:00 |
| 15        | 201615083144 | 2016-08-28     | 2016-08-28T00:41:00.000 | 00:41:00 |
| 24        | 201624083193 | 2016-08-28     | 2016-08-28T01:02:00.000 | 01:02:00 |
| 24        | 201624083194 | 2016-08-28     | 2016-08-28T00:44:00.000 | 00:44:00 |
| 25        | 201625072218 | 2016-08-28     | 2016-08-28T00:27:00.000 | 00:27:00 |
| 25        | 201625072227 | 2016-08-28     | 2016-08-28T00:56:00.000 | 00:56:00 |
| 26        | 201626043171 | 2016-08-28     | 2016-08-28T01:13:00.000 | 01:13:00 |
| 39        | 201639069963 | 2016-08-28     | 2016-08-28T00:54:00.000 | 00:54:00 |
| 39        | 201639069964 | 2016-08-28     | 2016-08-28T00:54:00.000 | 00:54:00 |
+-----------+--------------+----------------+---------------------+----------+
13 rows selected (0.143 seconds)
0: jdbc:phoenix:localhost:2181:/hbase-unsecur>
```

DEMONSTRATION