

INFLUXDB

SENG8080 - CASE STUDIES BIG DATA

SIHINT JIWANT SINGH (25%)

RAVAL PRATIKSHA MAHENDRABHAI (25%)

SINGH PADAMDEEP (25%)

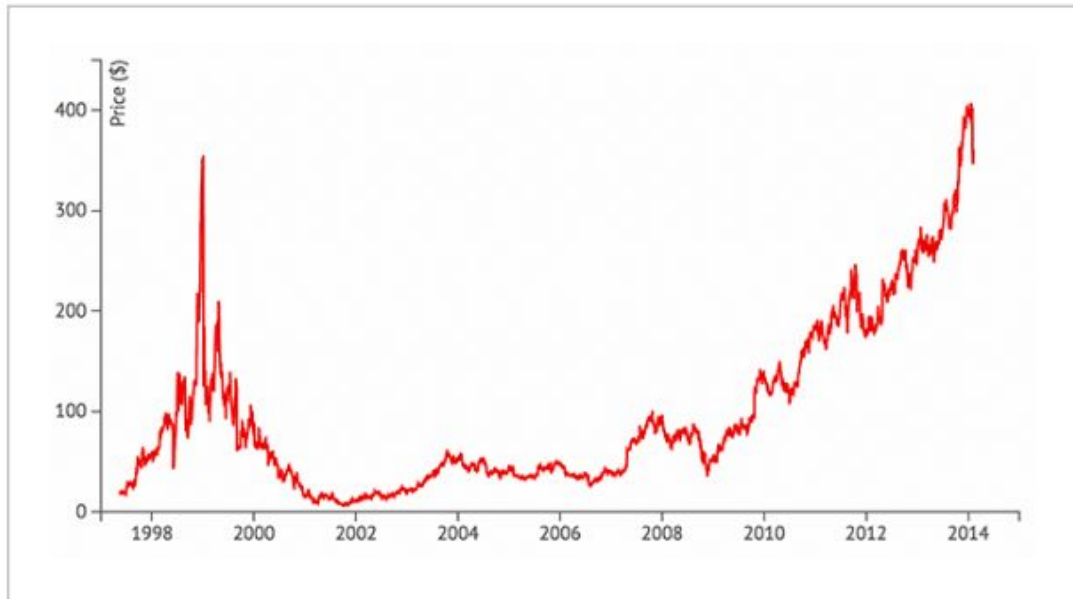
IDODE IGBESKI (25%)



AGENDA

- Introduction to Time Series Databases
- Influx DB
- Influx DB vs SQL server
- Influx DB vs Cassandra
- Influx DB vs Elasticsearch
- Influx DB vs Open TSDB
- Computer Processor Analysis Project

TIME SERIES DATABASE



- A time series database (TSDB) is a database optimized for time-stamped or time series data.
- Time series data are simply measurements or events that are tracked, monitored, collected and aggregated over time.
- Examples : Weather Data, Stock Prices

Difference between Relational Database and Time series Database

- Time series databases work differently.
- Data are still stored in 'collections' but those collections share a common denominator : **they are aggregated over time.**

Classic Relational Databases

Name ▲	Age ⇅	Nickname	Employee ▼
Giacomo Guilizzoni Founder & CEO	40	Peldi	<input type="radio"/>
Marco Botton Tuttofare	38		<input checked="" type="checkbox"/>
Mariah MacLachlan Better Half	41	Patata	<input type="checkbox"/>
Valerie Liberty Head Chef	46	Val	<input checked="" type="checkbox"/>

Data are multidimensional

Time Series Databases

Sensor Temperature ▲	Time	
39.5	12/04/19 @ 14:12	
41.2	12/04/19 @ 14:13	
12.4	14/04/19 @ 12:15	
18.5	16/04/19 @ 10:05	

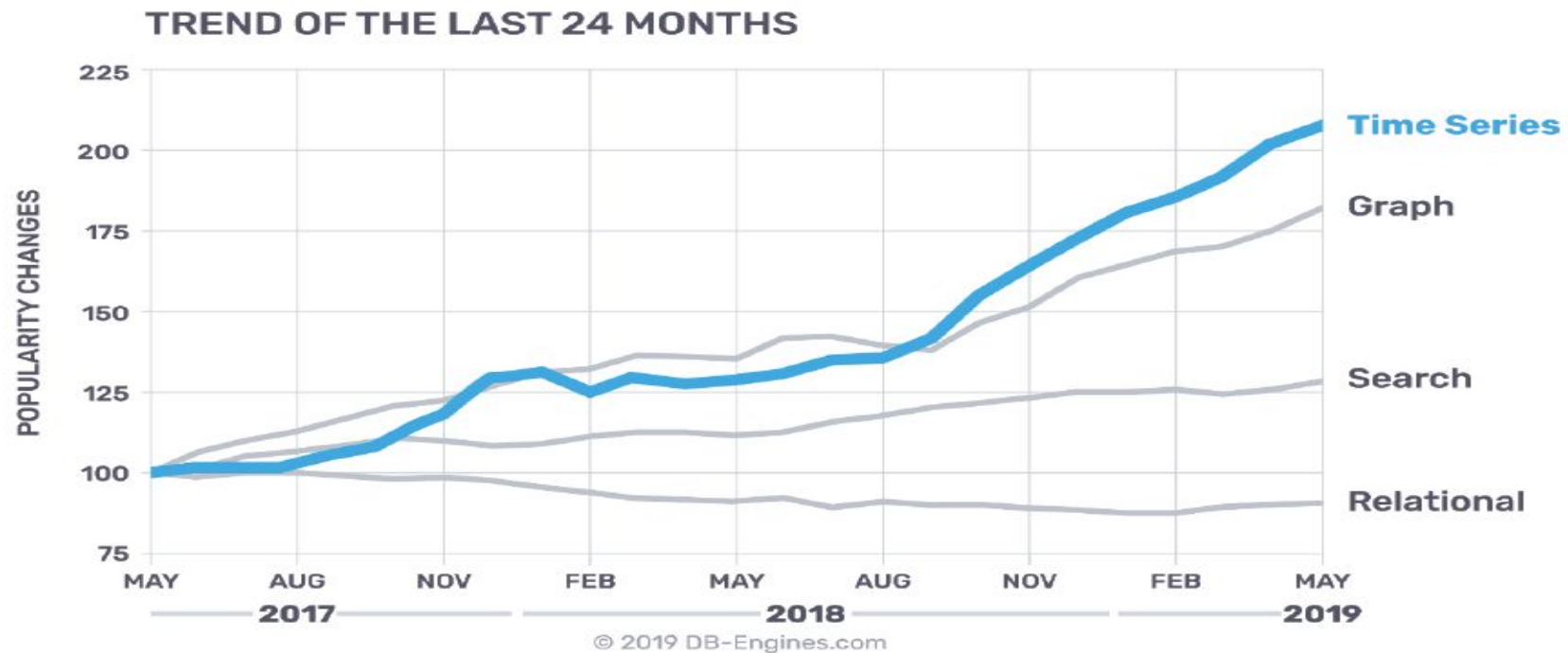
Data are aggregated over time



PROPERTIES OF TSDBS

- High write performance
- Scalability
- Fast Queries
- Data Location
- Usability

POPULARITY OF TSDBS



(a) Popularity Trend (Source:<https://www.influxdata.com/time-series-database/>)

BENEFITS AND USES OF TSDBS

- Low Cost
- Reduced Downtime
- Improved Business Decisions
- Scalability and Performance improvement

INFLUXDB

- InfluxDB is a time series database which was created and designed by InfluxData, the database was written in a programming language called GO programming Language. InfluxDB is designed for metrics and events using its time series platform to handle high workload in a fast, high availability and retrieval of time series data.

USE CASES



IOT (INTERNET OF
THINGS)


































REAL-TIME
ANALYTICS



DEVOPS

INFLUXDB RATING

Rank			DBMS	Database Model	Score		
May 2019	Apr 2019	May 2018			May 2019	Apr 2019	May 2018
1.	1.	1.	InfluxDB 	Time Series	18.08	+0.86	+7.08
2.	2.	2.	Kdb+ 	Time Series, Multi-model 	5.60	-0.25	+2.52
3.	3.	 4.	Graphite	Time Series	3.23	+0.10	+0.96
4.	4.	 6.	Prometheus	Time Series	3.11	+0.20	+1.99
5.	5.	 3.	RRDtool	Time Series	2.90	+0.19	+0.21
6.	6.	 5.	OpenTSDB	Time Series	2.47	+0.10	+0.85
7.	7.	7.	Druid	Multi-model 	1.69	+0.04	+0.67
8.	8.	 18.	TimescaleDB 	Time Series, Multi-model 	1.16	+0.21	+1.12
9.	9.	 8.	KairosDB	Time Series	0.54	-0.09	+0.12
10.	10.	 9.	eXtremeDB 	Multi-model 	0.38	-0.02	+0.07
11.	11.	11.	FaunaDB 	Multi-model 	0.38	+0.01	+0.27
12.	 13.	 14.	GridDB 	Multi-model 	0.34	+0.02	+0.24
13.	 14.	 21.	Heroic	Time Series	0.34	+0.08	+0.34
14.	 12.		Amazon Timestream	Time Series	0.27	-0.06	
15.	15.		IBM Db2 Event Store	Multi-model 	0.26	+0.01	
16.	16.	 10.	Riak TS	Time Series	0.22	-0.02	-0.05
17.	17.	 15.	Axibase	Time Series	0.21	-0.02	+0.15
18.	18.	 17.	Warp 10	Time Series	0.16	-0.02	+0.12
19.	19.	 12.	Hawkular Metrics	Time Series	0.15	-0.02	+0.04
20.	20.	 16.	Quasardb 	Time Series	0.11	0.00	+0.06

<https://db-engines.com/en/ranking/time+series+dbms>

INFLUXDB VERSIONS



OPENSOURCE TICK STACK



INFLUXDB CLOUD









INFLUXDB ENTERPRISE

PACKAGES SUPPORTED BY VERSIONS

	InfluxDB Open Source	InfluxDB Cloud	InfluxDB Enterprise
Open Source Core	✓	✓	✓
Extensible	✓	✓	✓
Support for Regular and Irregular Time Series Data	✓	✓	✓
High Availability (Clustering)	✗	✓	✓
Scalability (Clustering)	✗	✓	✓
Advanced Backup and Restore	✗	✓	✓
Complete Platform Support	Optional	✓	✓
Managed by InfluxData	✗	✓	✗
Runs on Any Cloud	✓	AWS Only	✓
Runs On Premises	✓	✗	✓
Preset Configurations	✓	✓	✓
Custom Configurations		Contact Sales	Contact Sales

<https://www.influxdata.com/products/editions/>

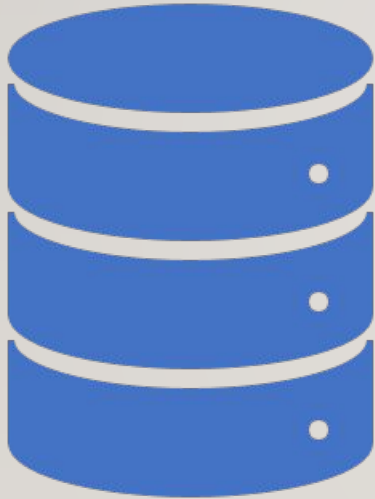
INFLUXDB VS SQL SERVER

Name	InfluxDB 	Microsoft SQL Server 
Description	DBMS for storing time series, events and metrics	Microsofts relational DBMS
Primary database model	Time Series DBMS	Relational DBMS
Secondary database models		Document store Graph DBMS
DB-Engines Ranking 	Score 18.00 Rank #34 Overall #1 Time Series DBMS	Score 1090.83 Rank #3 Overall #3 Relational DBMS
Website	www.influxdata.com/time-series-platform/-influxdb	www.microsoft.com/en-us/sql-server
Technical documentation	docs.influxdata.com/influxdb	docs.microsoft.com/en-ie/sql/sql-server/sql-server-technical-documentation
Developer		Microsoft
Initial release	2013	1989
Current release	1.7.6, April 2019	SQL Server 2017, October 2017
License 	Open Source 	commercial 

System Properties Comparison InfluxDB vs. Microsoft SQL Server. (n.d.). Retrieved from

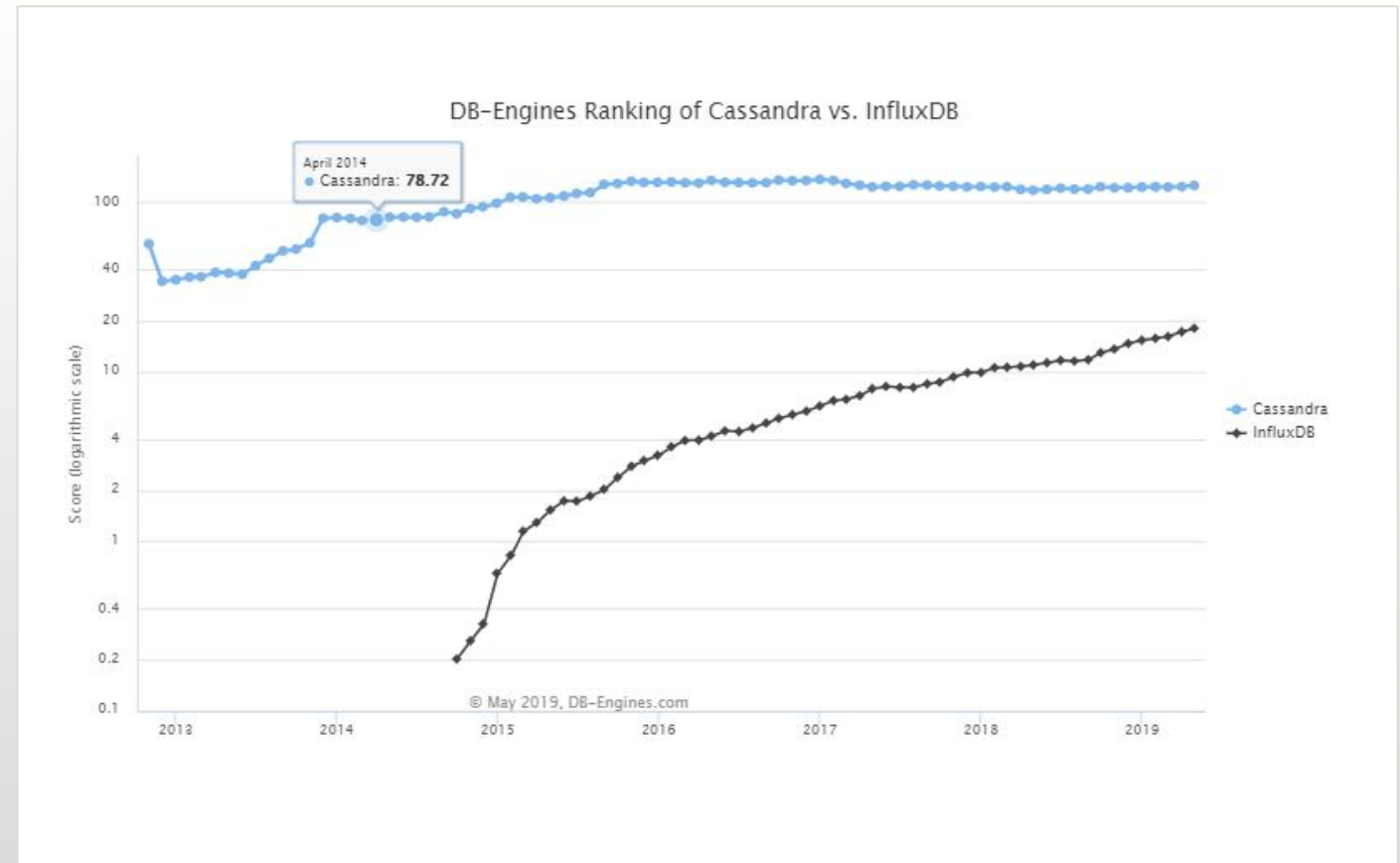
[https://db-engines.com/en/system/InfluxDB;Microsoft SQL Server](https://db-engines.com/en/system/InfluxDB;Microsoft%20SQL%20Server)

CONTINUE...



- 1. provides prediction about schema preferences that they may vary during time which is beneficial while dealing with large datasets.
- 2. Datapoints in InfluxDB have one or all field measurements, which helps in adding new fields without extra burden.

TREND CHART



CONTINUE...



Both Cassandra and Influx DB are open source but the implementation languages are different in both as Cassandra has Java and Influx DB has Go.



The Server operating systems are BSD, Linux, OS X, Windows for Cassandra, where as only Linux and OS X supports Influx DB.

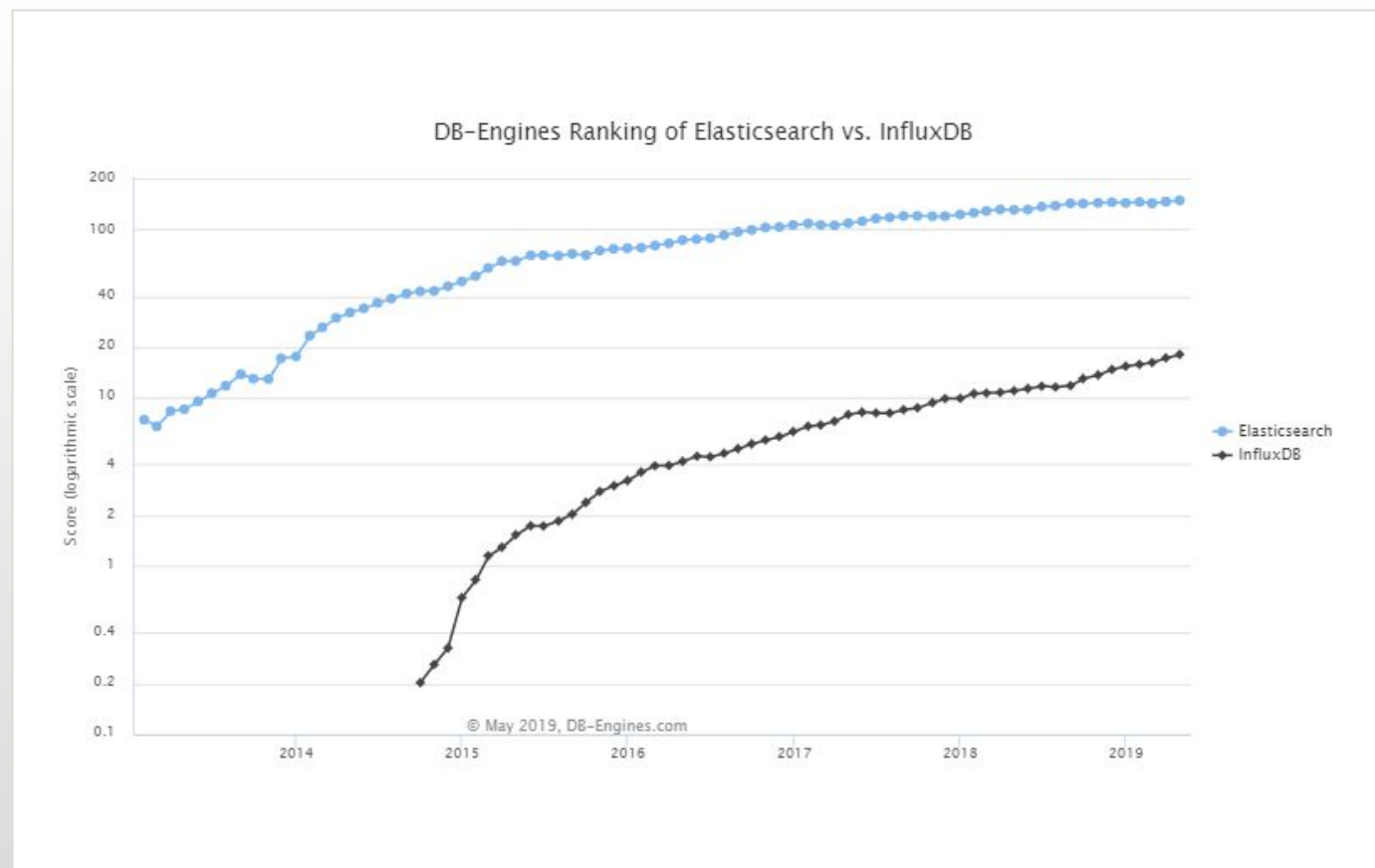


The typing in Cassandra is enables and can be of ay type where as in Influx DB it is only Numeric Data and Strings.



Triggers, which plays main part in some of the databases are introduced in Cassandra but are not used or are programmed in Influx DB.

TREND CHART



INFLUXDB VS ELASTICSEARCH













Name	Elasticsearch 	InfluxDB 
Description	A distributed, RESTful modern search and analytics engine based on Apache Lucene 	DBMS for storing time series, events and metrics
Primary database model	Search engine	Time Series DBMS
Secondary database models	Document store	
DB-Engines Ranking  Trend Chart 	Score 148.81 Rank #7 Overall #1 Search engines	Score 18.00 Rank #34 Overall #1 Time Series DBMS
Website	www.elastic.co/products/elasticsearch	www.influxdata.com/time-series-platform/influxdb
Technical documentation	www.elastic.co/guide/en/elasticsearch/reference/current/index.html	docs.influxdata.com/influxdb
Developer	Elastic	
Initial release	2010	2013
Current release	7.2.0, June 2019	1.7.6, April 2019
License 	Open Source 	Open Source 
Cloud-based only 	no	no
DBaaS offerings (sponsored links) 	Elasticsearch Service on Elastic Cloud : Try out the official hosted Elasticsearch and Kibana offering available on AWS and GCP that's powered by the creators of Elasticsearch.	
Implementation language	Java	Go
Server operating systems	All OS with a Java VM	Linux OS X 
Data scheme	schema-free 	schema-free
Typing 	yes	Numeric data and Strings
XML support 	no	no
Secondary indexes	yes 	no
SQL 	SQL-like query language	SQL-like query language
APIs and other access methods	Java API RESTful HTTP/JSON API	HTTP API JSON over UDP

- System Properties Comparison Elasticsearch vs. InfluxDB. (n.d.). Retrieved from <https://db-engines.com/en/system/Elasticsearch;InfluxDB>

CONTINUE...

- In DB Engines ranking, the Elasticsearch stands at overall rank of 7 with score 148.62 whereas same as above, Influx DB ranks at 34th position with score of 18.08.
- The initial release of Elasticsearch was 2010 and that of Influx DB is 2013 as same described above.
- The Influx DB is having the same release with version 1.7.5 and the Elasticsearch is having the current release in April,2019 with version 7.0.0.

INFLUXDB VS OPENTSDDB

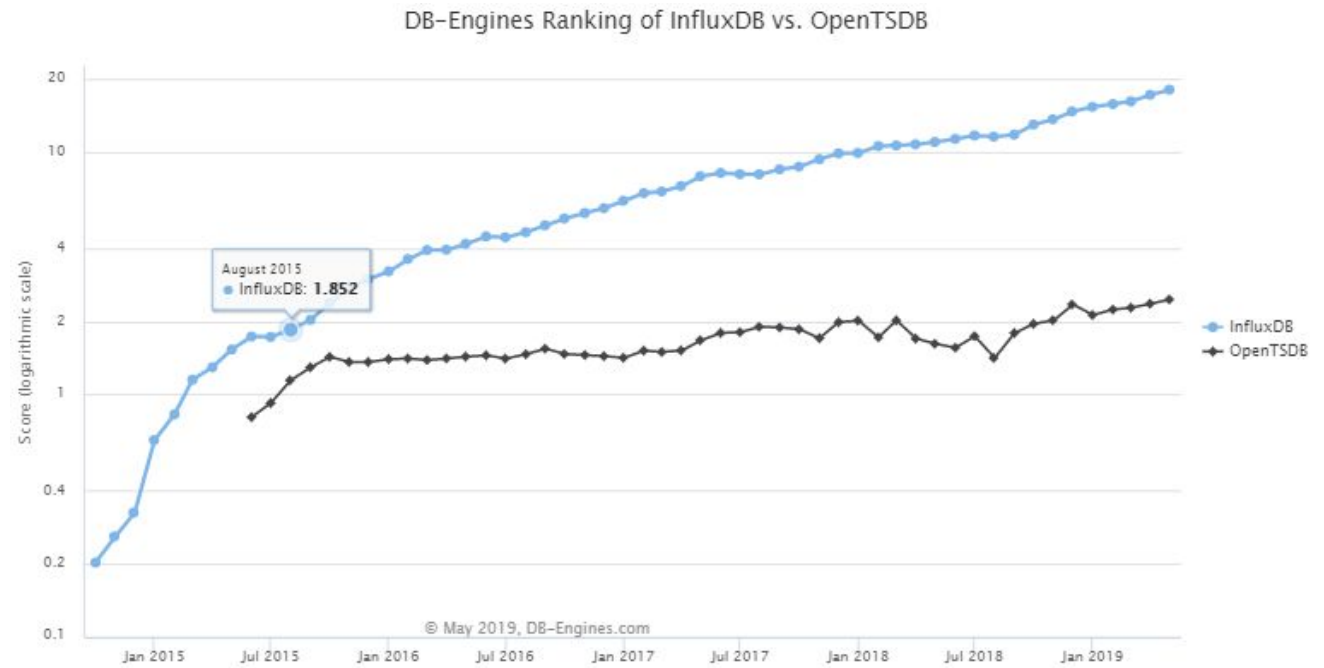
Name	InfluxDB 	OpenTSDB 
Description	DBMS for storing time series, events and metrics	Scalable Time Series DBMS based on HBase
Primary database model	Time Series DBMS	Time Series DBMS
DB-Engines Ranking  	Score 18.00 Rank #34 Overall #1 Time Series DBMS	Score 2.30 Rank #105 Overall #6 Time Series DBMS
Website	www.influxdata.com/time-series-platform/influxdb	opentsdb.net
Technical documentation	docs.influxdata.com/influxdb	opentsdb.net/docs/build/html/index.html
Developer		currently maintained by Yahoo and other contributors
Initial release	2013	2011
Current release	1.7.6, April 2019	
License 	Open Source 	Open Source 
Cloud-based only 	no	no
DBaaS offerings (sponsored links) 		
Implementation language	Go	Java
Server operating systems	Linux OS X 	Linux Windows
Data scheme	schema-free	schema-free
Typing 	Numeric data and Strings	numeric data for metrics, strings for tags
XML support 	no	no
Secondary indexes	no	no
SQL 	SQL-like query language	no
APIs and other access methods	HTTP API JSON over UDP	HTTP API Telnet API
Supported programming languages	.Net Clojure Erlang Go Haskell Java	Erlang Go Java Python R Ruby

- System Properties Comparison InfluxDB vs. OpenTSDB. (n.d.). Retrieved from <https://db-engines.com/en/system/InfluxDB;OpenTSDB>

CONTINUE...

- In DB-Engines ranking, the Influx DB stands high at number 34 with 18.08 score as OpenTSDB stands at far away in the rating at number 104 with score of just 2.47. Also, in time series DBMS influx DB stands at number 1st position whereas OpenTSDB stands at 6th position.
- The server operating systems in Influx DB is Linux, OS X whereas in OpenTSDB, is Linux and Windows.
- Data schemes in both the DB-Engines are schema free.
- In influx DB, the typing is Numeric data and Strings, which is almost same for OpenTSDB, the only addition is of tags in it.

TREND CHART



InfluxDB Concepts

```
name: records-presentation-demo5
time                cpu_percent_utilization free_memory index number_processes virtual_memory
-----
1565623914005156600 13.8                8667684864 1      248                8365965312
> _
```

Point

Measurement

Time Stamp

Tag

Field

Point

4 Elements of a Point:

1. Measurement
2. Time Stamp
3. Tag
4. Field

Point can be uniquely identified with:

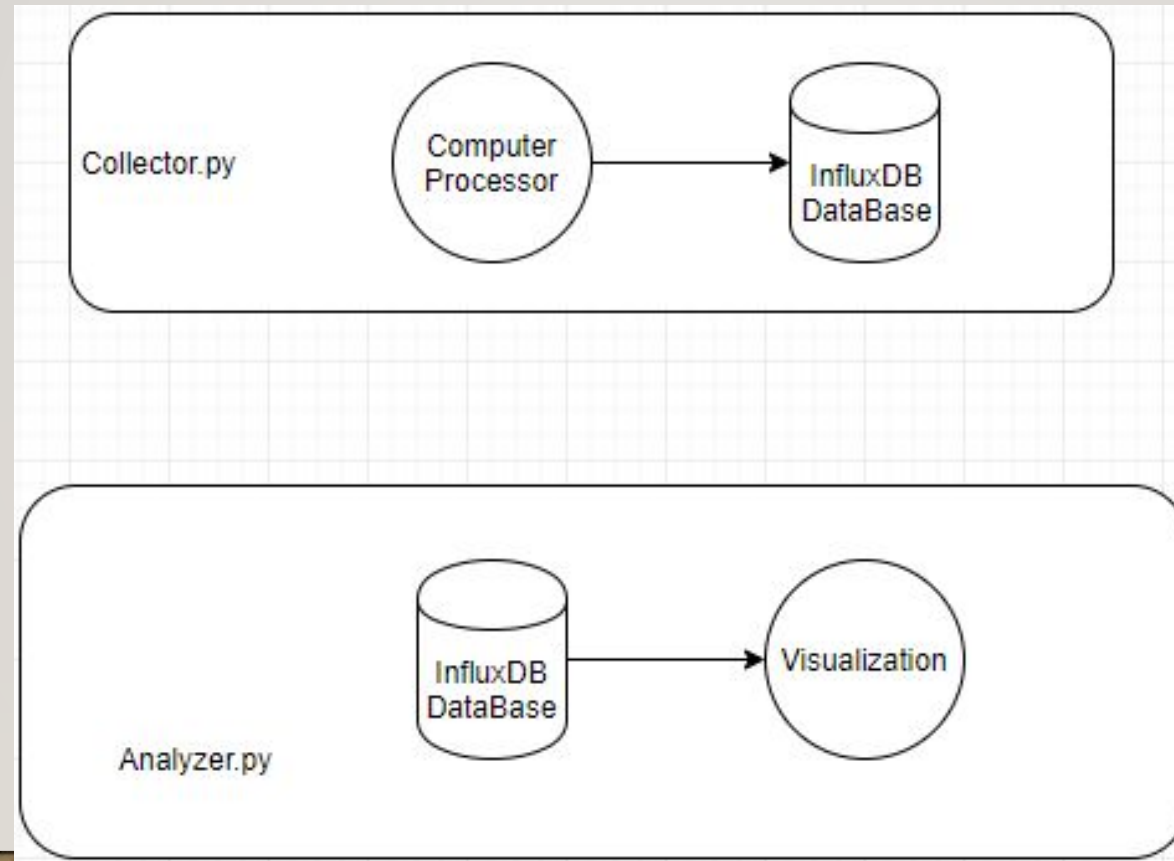
1. Series
2. Timestamp

Series

Following is shared in Series:

1. Retention Policy
2. Measurement
3. Tag

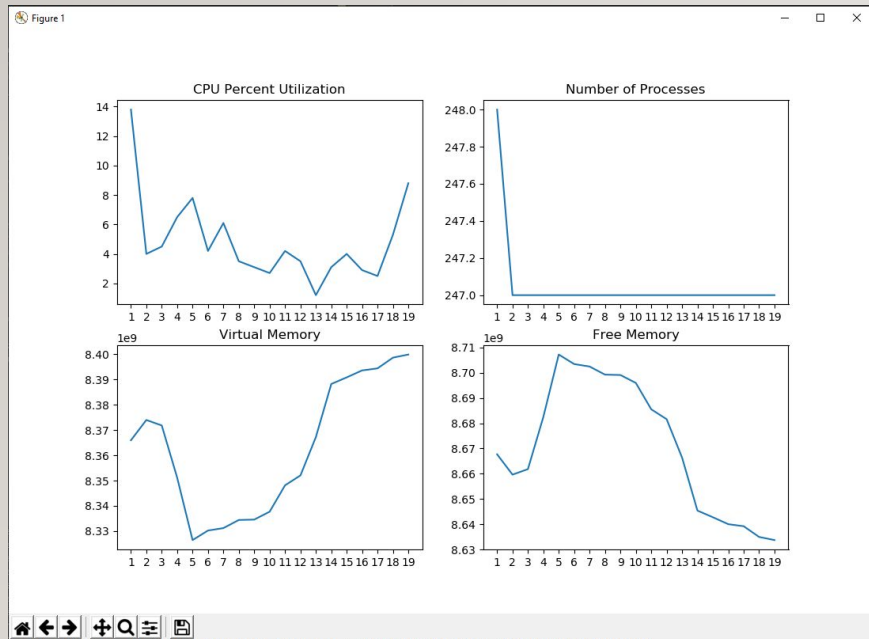
InfluxDB Application



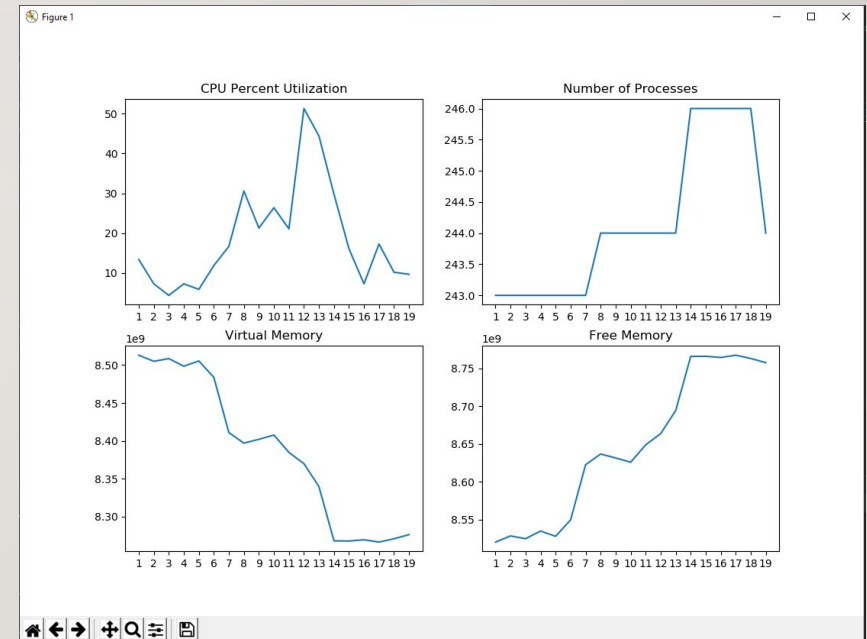
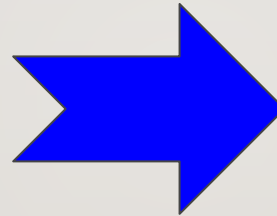
InfluxDB ServerTowards Application

```
D:\Conestoga_2\Case_Studies\Project-InfluxDB\influxdb-1.7.7-1\influxd.exe
2019-08-12T15:26:45.588478Z info Reading file {"log_id": "0HDbrx8W000", "engine": "tsm1", "service": "cacheloader", "path": "C:\\Users\\SINGH PC\\.influxdb\\wal\\testdb\\autogen\\2\\_00001.wal", "size": 40}
2019-08-12T15:26:45.590473Z info Opened shard {"log_id": "0HDbrx8W000", "service": "store", "trace_id": "0HDbrx_G000", "op_name": "tsdb_open", "index_version": "inmem", "path": "C:\\Users\\SINGH PC\\.influxdb\\data\\_int
ernal\\monitor\\1", "duration": "61.837ms"}
2019-08-12T15:26:45.630365Z info Opened shard {"log_id": "0HDbrx8W000", "service": "store", "trace_id": "0HDbrx_G000", "op_name": "tsdb_open", "index_version": "inmem", "path": "C:\\Users\\SINGH PC\\.influxdb\\data\\test
db\\autogen\\2", "duration": "61.835ms"}
2019-08-12T15:26:45.909626Z info Opened shard {"log_id": "0HDbrx8W000", "service": "store", "trace_id": "0HDbrx_G000", "op_name": "tsdb_open", "index_version": "inmem", "path": "C:\\Users\\SINGH PC\\.influxdb\\data\\_int
ernal\\monitor\\6", "duration": "380.990ms"}
2019-08-12T15:26:45.932556Z info Opened shard {"log_id": "0HDbrx8W000", "service": "store", "trace_id": "0HDbrx_G000", "op_name": "tsdb_open", "index_version": "inmem", "path": "C:\\Users\\SINGH PC\\.influxdb\\data\\Proc
essorInfo\\autogen\\7", "duration": "431.846ms"}
2019-08-12T15:26:46.156989Z info Reading file {"log_id": "0HDbrx8W000", "engine": "tsm1", "service": "cacheloader", "path": "C:\\Users\\SINGH PC\\.influxdb\\wal\\_internal\\monitor\\8\\_00006.wal", "size": 7592557}
2019-08-12T15:26:46.471116Z info Opened shard {"log_id": "0HDbrx8W000", "service": "store", "trace_id": "0HDbrx_G000", "op_name": "tsdb_open", "index_version": "inmem", "path": "C:\\Users\\SINGH PC\\.influxdb\\data\\_int
ernal\\monitor\\8", "duration": "942.480ms"}
2019-08-12T15:26:46.471116Z info Open store (end) {"log_id": "0HDbrx8W000", "service": "store", "trace_id": "0HDbrx_G000", "op_name": "tsdb_open", "op_event": "end", "op_elapsed": "1078.081ms"}
2019-08-12T15:26:46.474108Z info Opened service {"log_id": "0HDbrx8W000", "service": "subscriber"}
2019-08-12T15:26:46.477101Z info Starting monitor service {"log_id": "0HDbrx8W000", "service": "monitor"}
2019-08-12T15:26:46.478098Z info Registered diagnostics client {"log_id": "0HDbrx8W000", "service": "monitor", "name": "build"}
2019-08-12T15:26:46.479095Z info Registered diagnostics client {"log_id": "0HDbrx8W000", "service": "monitor", "name": "runtime"}
2019-08-12T15:26:46.479095Z info Registered diagnostics client {"log_id": "0HDbrx8W000", "service": "monitor", "name": "network"}
2019-08-12T15:26:46.480092Z info Registered diagnostics client {"log_id": "0HDbrx8W000", "service": "monitor", "name": "system"}
2019-08-12T15:26:46.481089Z info Starting precreation service {"log_id": "0HDbrx8W000", "service": "shard-precreation", "check_interval": "10m", "advance_period": "30m"}
2019-08-12T15:26:46.481089Z info Storing statistics {"log_id": "0HDbrx8W000", "service": "monitor", "db_instance": "_internal", "db_rp": "monitor", "interval": "10s"}
2019-08-12T15:26:46.484082Z info Starting snapshot service {"log_id": "0HDbrx8W000", "service": "snapshot"}
2019-08-12T15:26:46.487074Z info Starting continuous query service {"log_id": "0HDbrx8W000", "service": "continuous_querier"}
2019-08-12T15:26:46.488071Z info Starting HTTP service {"log_id": "0HDbrx8W000", "service": "httpd", "authentication": false}
2019-08-12T15:26:46.489067Z info opened HTTP access log {"log_id": "0HDbrx8W000", "service": "httpd", "path": "stderr"}
2019-08-12T15:26:46.491063Z info Listening on HTTP {"log_id": "0HDbrx8W000", "service": "httpd", "addr": "[::]:8086", "https": false}
2019-08-12T15:26:46.492060Z info Starting retention policy enforcement service {"log_id": "0HDbrx8W000", "service": "retention", "check_interval": "30m"}
2019-08-12T15:26:46.495051Z info Listening for signals {"log_id": "0HDbrx8W000"}
2019-08-12T15:26:46.496052Z info Sending usage statistics to usage.influxdata.com {"log_id": "0HDbrx8W000"}
```

Project - Statistics Analysis



Visualization - Without Stimulus



Visualization - With Stimulus

REFERENCES

- [1] "InfluxDB 1.7 release notes". Retrieved 2 April 2019.
- [2] InfluxData, "Time series database (TSDB) explained" [Online]. Available:
<https://www.influxdata.com/time-series-database/>
- [3] InfluxDB Website to analyze different editions:
<https://www.influxdata.com/products/editions/>
- [4] InfluxDB IOT Data Platform: <https://www.influxdata.com/customers/iot-data-platform/>
- [5] InfluxDB Real Time Analytics: <https://www.influxdata.com/customers/real-time-analytics/>
- [6] <https://en.wikipedia.org/wiki/InfluxDB>
- [7] <https://www.influxdata.com/products/compare/>
- [8] Using psutil to get Data from computer processor: <https://pypi.org/project/psutil/>
- [9] Using influxdb to interact with Influx DB Database:
<https://github.com/influxdata/influx-python>
- [10] Using influxdb connections with Influx DB Database:
<https://www.influxdata.com/blog/getting-started-python-influxdb/>
- [11] Using Pandas for dataframes: <https://www.geeksforgeeks.org/python-pandas-dataframe/>
- [12] Using matplotlib for plotting real time data: <https://matplotlib.org>
- [13] Using seaborn for visualizing Data:
- [14] <https://jakevdp.github.io/PythonDataScienceHandbook/04.14-visualization-with-seaborn.html>