

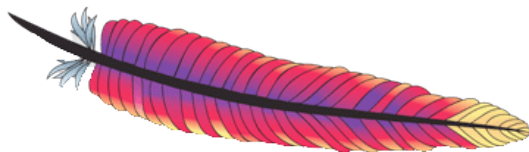
HDInsight Deep Dive: Storm, HBase and Hive

Architecting a Solution with HDInsight



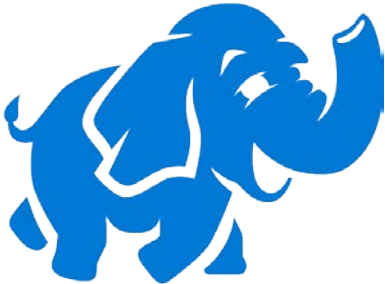
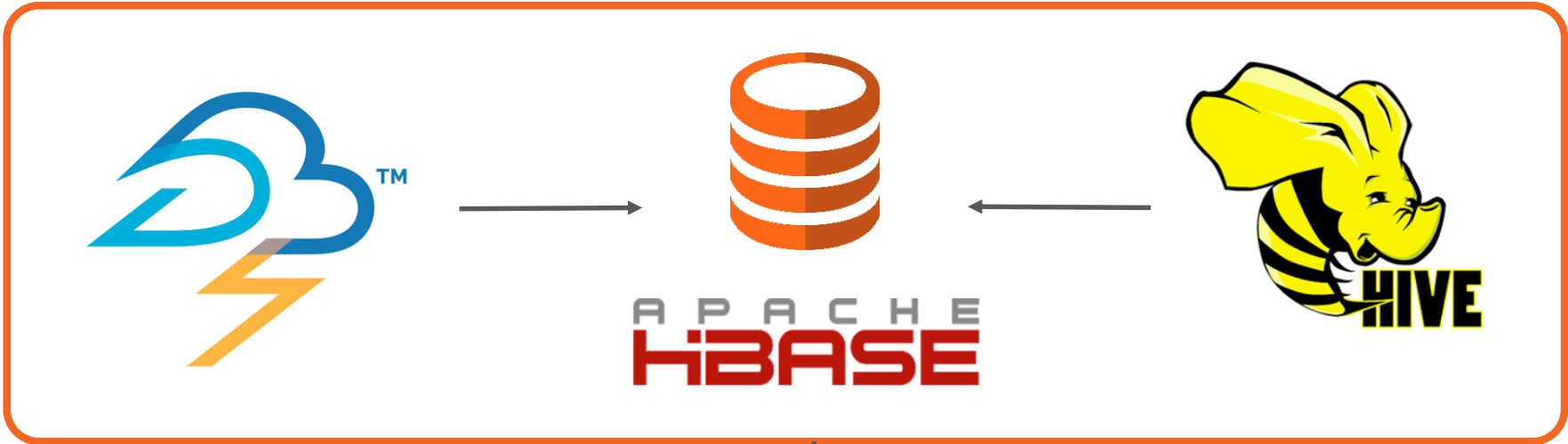
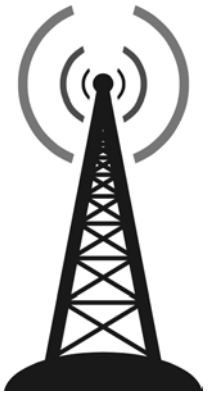
Elton Stoneman

@EltonStoneman | blog.sixeyed.com



APACHE
HBASE





HBase



APACHE
HBASE

No-SQL Database

Fast & Resilient

Terabytes of Data

Real-time Read & Write

Storm



Event Processor

Handles 1,000s msg/sec

Loosely-coupled Workflows

Real-time Processing

Hive

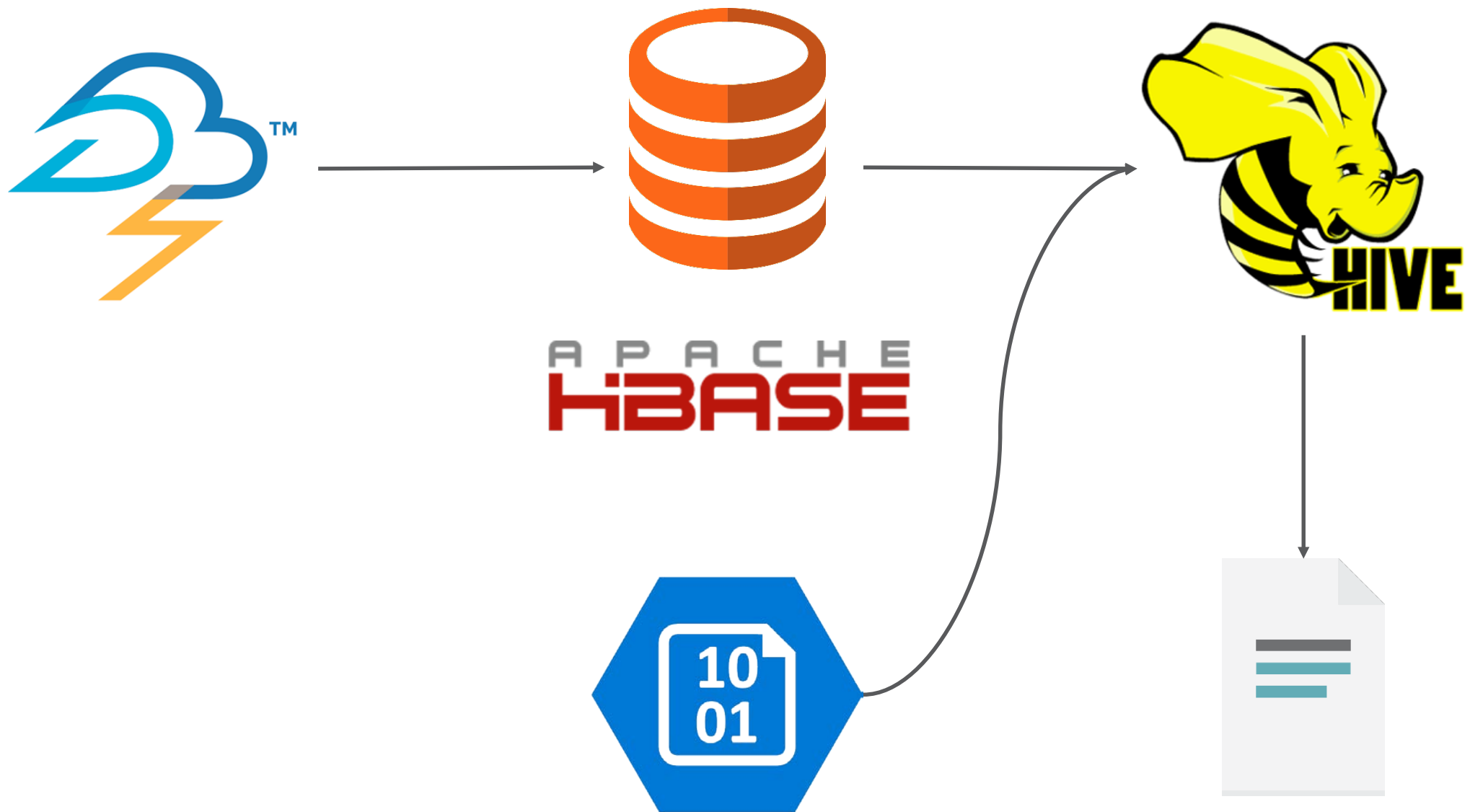


Data Warehouse

Hadoop & HBase Sources

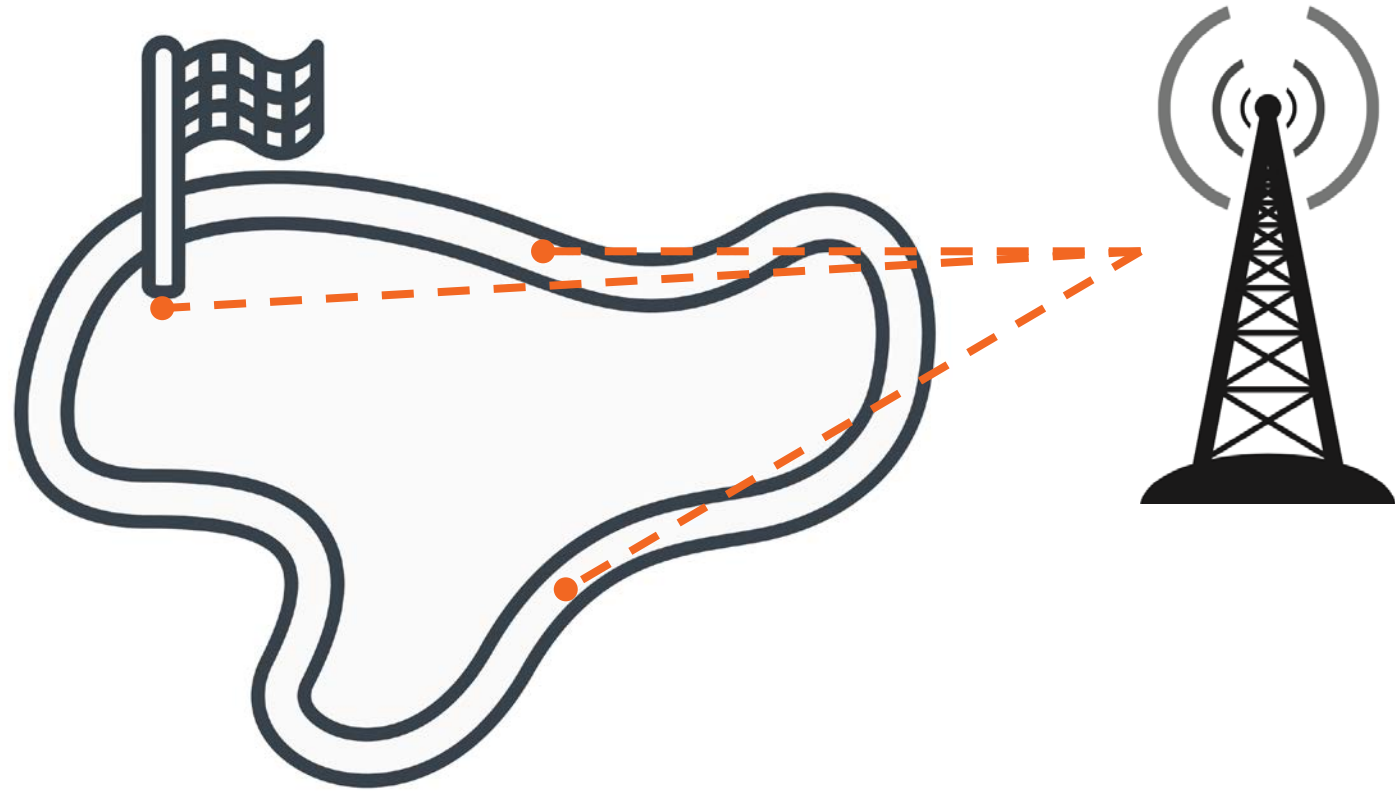
SQL-like Querying

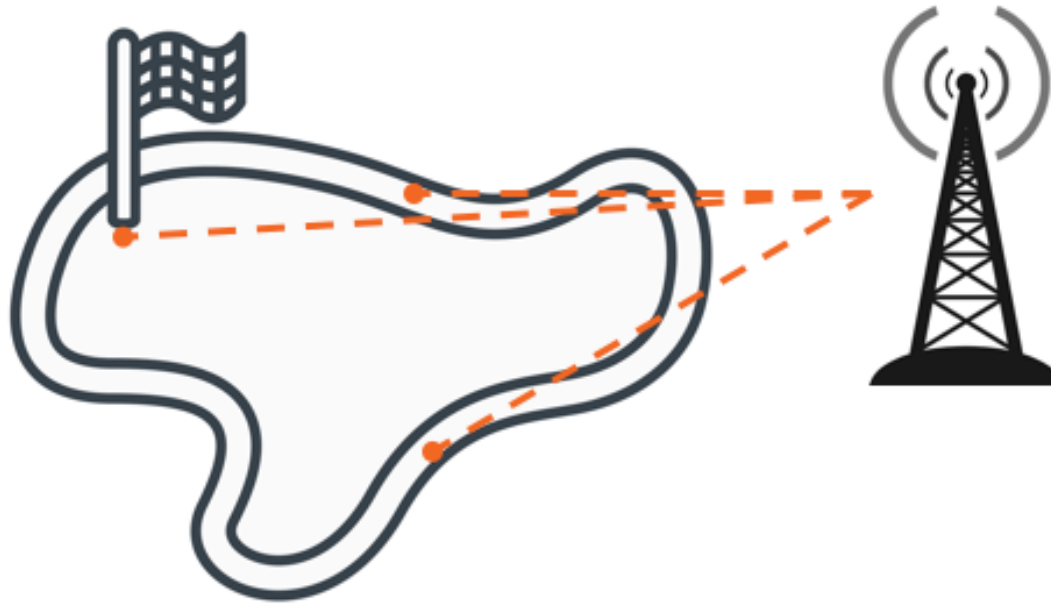
Batch Processing



ULYSSES by James Joyce-- I --Stately, plump Buck Mulligan came from the stairhead, bearing a bowl of lather on which a mirror and a razor lay crossed. A yellow dressing gown, ungirdled, was sustained gently behind him on the mild morning air. He held the bowl aloft and intoned an Introit: *Dei Gratia Amen*. Halted, he peered down the winding stair and called out coarsely:--Come up, you fearful Jesuit! Solemnly he came up and mounted the round gunrest. He blessed and blessed gravely thrice the tower and the air and the awaking boys. Then, on a sight of Stephen Dedalus bent toward him and made rapid crosses in the air, gurgling in his throat and shaking his head. Stephen Dedalus, displeased and sleepy, leaned his arms on the top of the staircase and looked coldly at the shaking gurgling face that blessed him, equine in its length, and at the light untanned hair, grained and hued like pale oak.

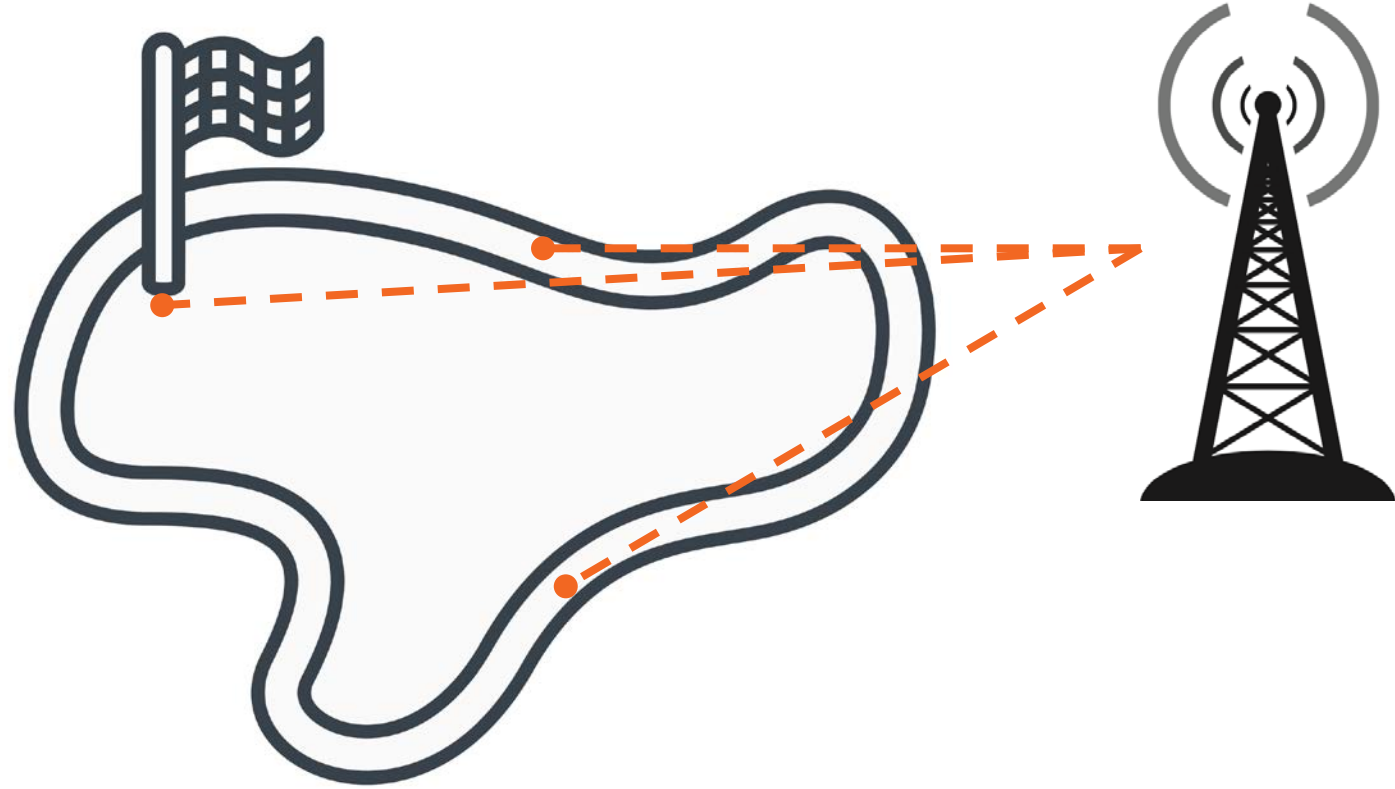
```
#Software: Microsoft Internet Information
Services 7.5 #Version: 1.0 #Date: 2012-01-06
09:09:05 #Fields: date time s-sitename s-
computername s-ip cs-method cs-uri-stem cs-
uri-query s-status cs-username c... cs-version
cs(User-Agent) (Cookie) cs(host) cs-host
sc-status s-status sc-status sc-
bytes cs-bytes ... 2012-01-06 09:09:05
W3SVC1273337584 10.211.146.27
GET /cuisine-francaise/huitres - 94.245.127.11
HTTP/1.1 Mozilla/5.0 (compatible;
+MSIE+9.0;+Windows NT+6.1;+WOW64;+Trident/5.0)
- http://site.supersimple.fr/ cuisine-francaise
site.supersimple.fr 200 0 0 4922 346 890 2012-01-06
2012-01-06 09:12 W3SVC1273337584
RD00155D360168 10.211.146.27 GET /cuisine-
francaise/huitres - 80 - 94.245.127.11 HTTP/1.
Mozilla/5.0+ (compatible;+MSIE+9.0;
+Windows+NT+6.1;+WOW64;+Trident/5.0) -
http://site.supersimple.fr/ cuisine-francaise
site.supersimple.fr 200 0 0 4922 346 890 2012-
01-06
```

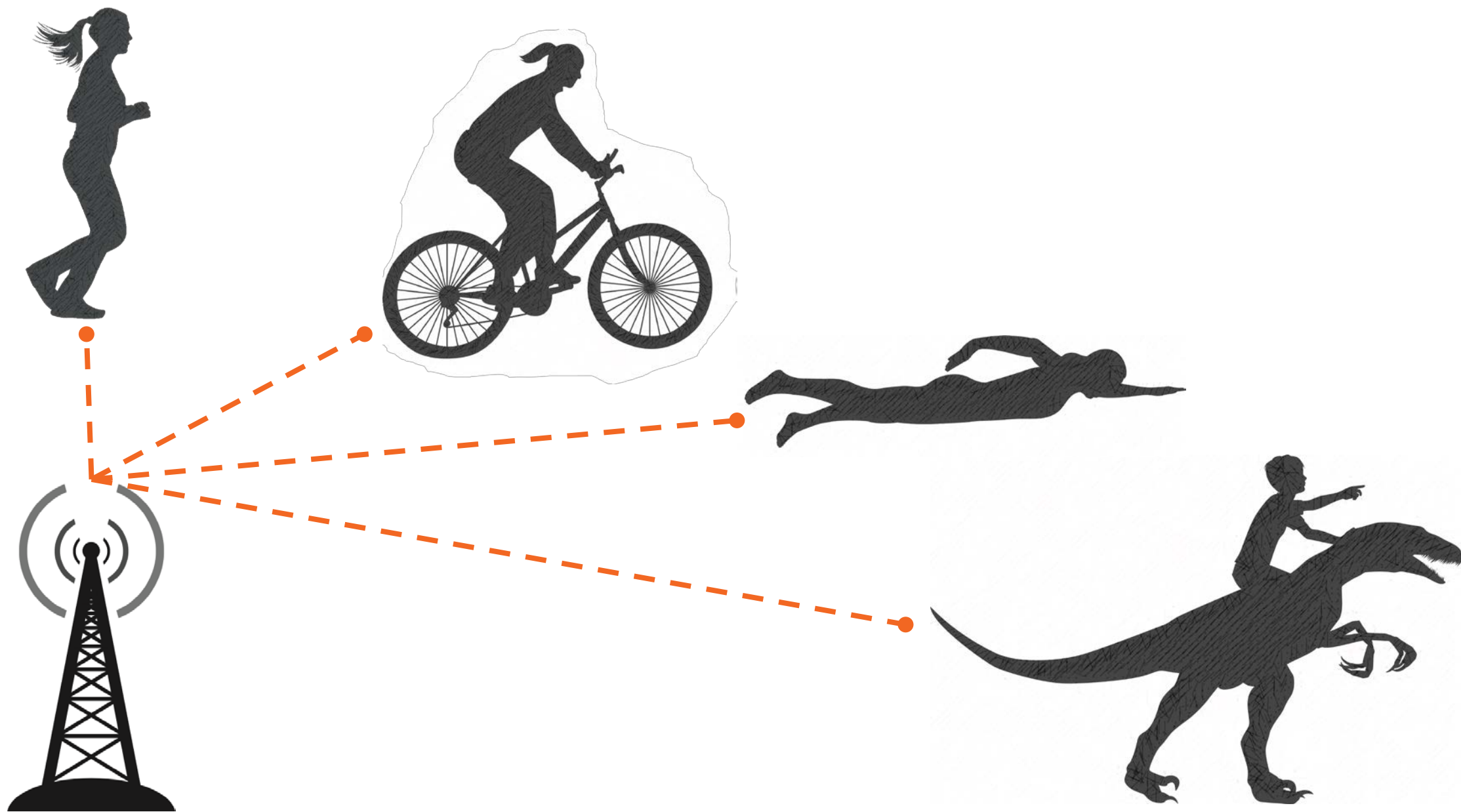





APACHE
HBASE







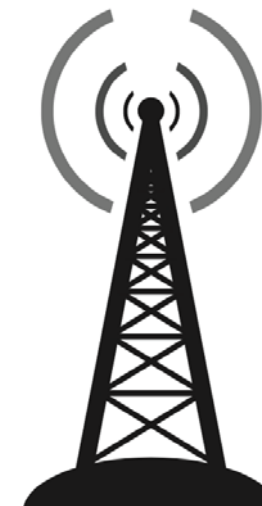
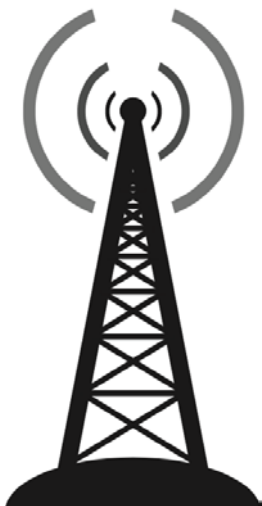


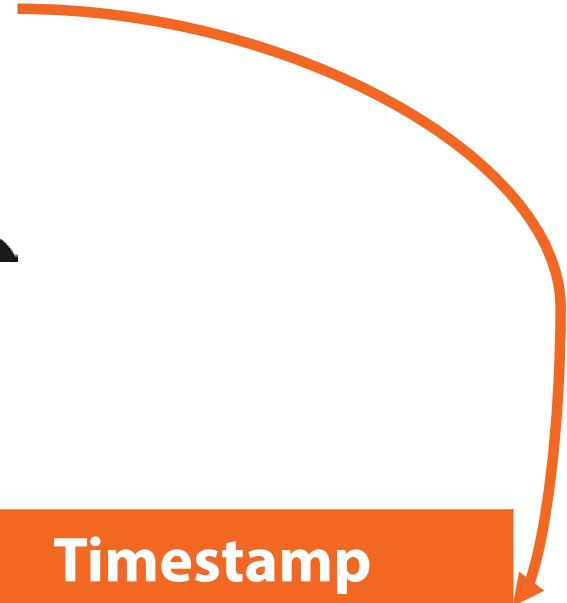
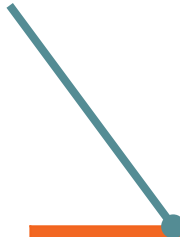
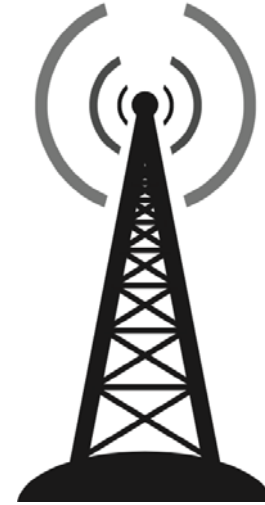
Racer ID

bb65ff



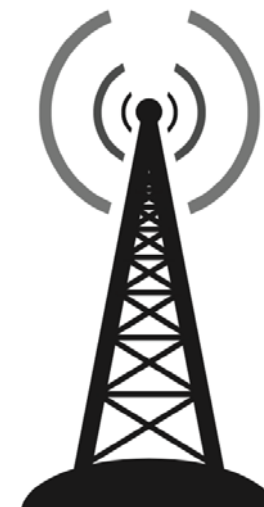
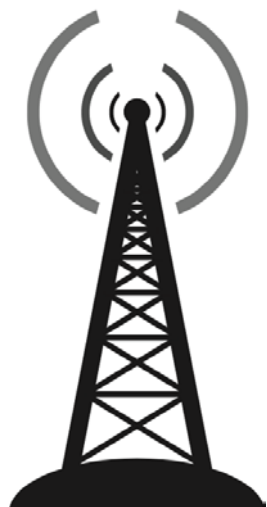
5Km





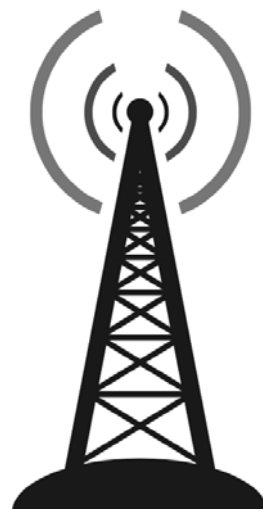
Racer ID	Timer ID	Timestamp
2e6869	a3b	1441880529000

Started At	Finished At	Total Time
1441880477000	1441881530000	00:17:32





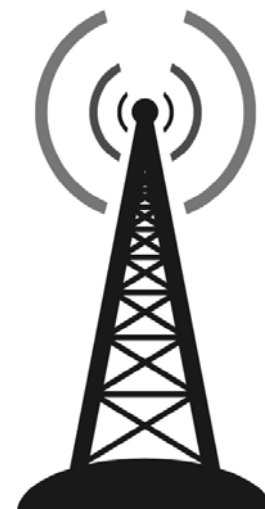
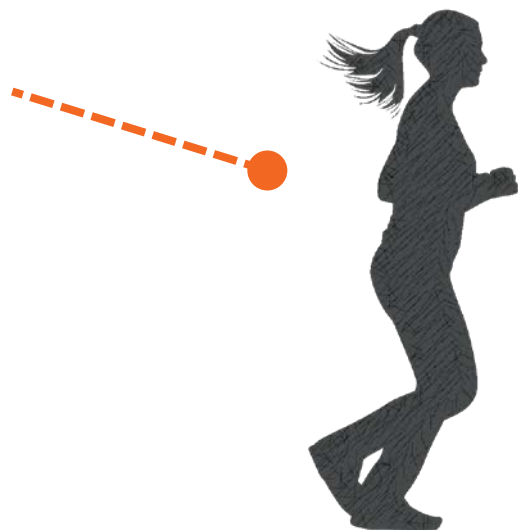
Racer ID	Total Duration (ms)	Position
e4324	754,230	1
c31dc	756,606	2
...		

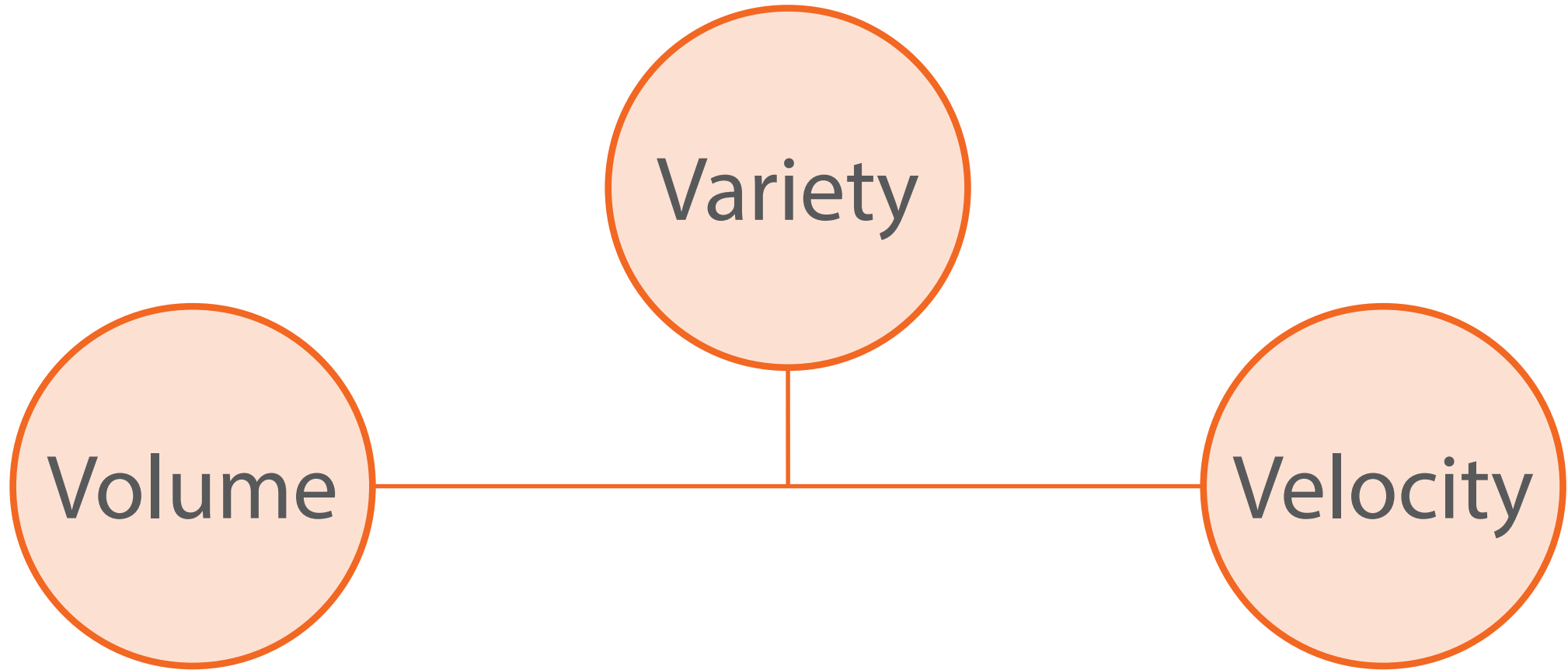


APACHE
HBASE


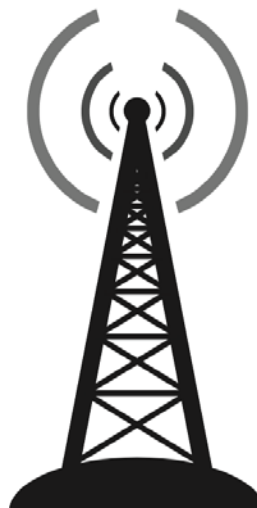


RFID
Bluetooth
Wearables

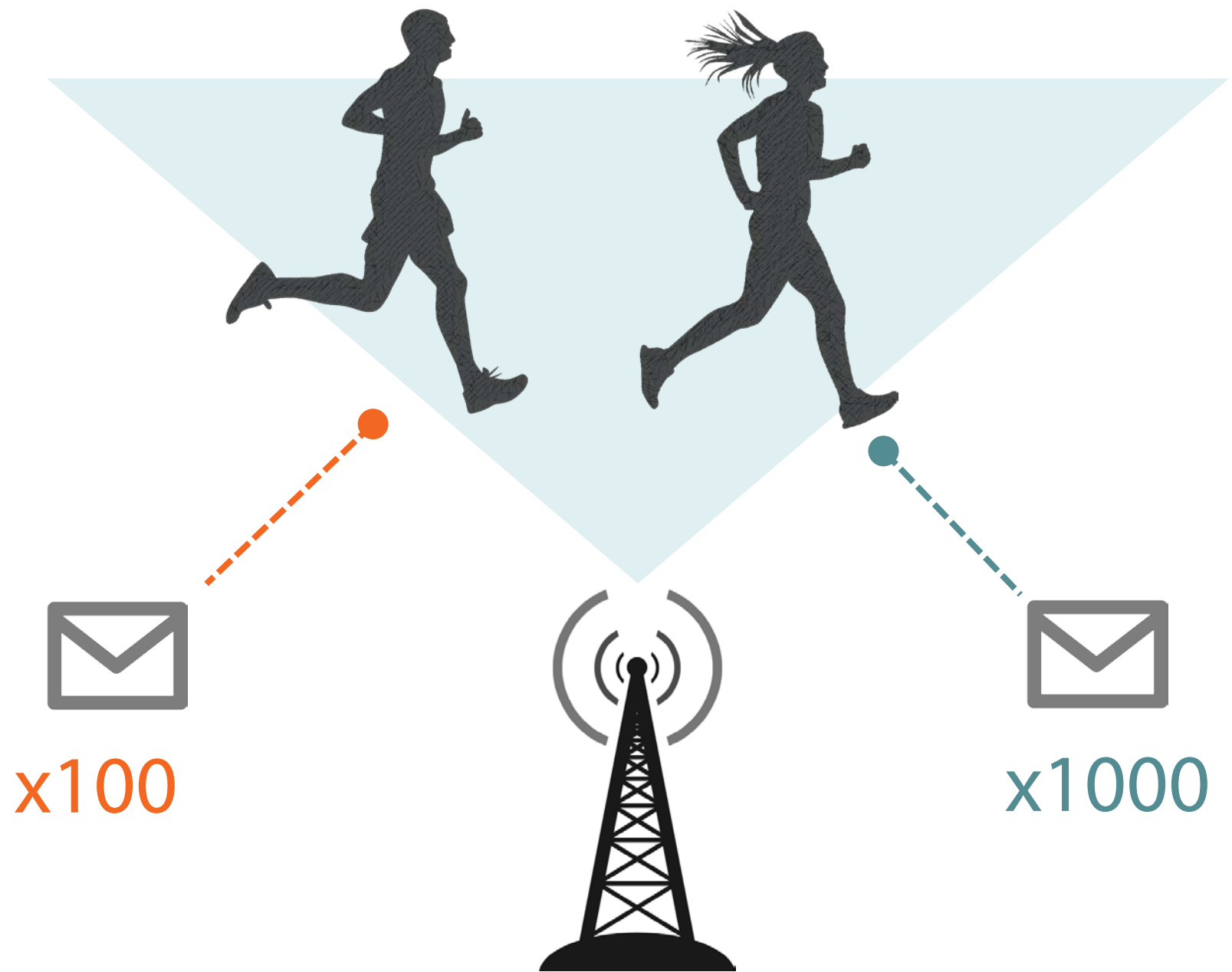




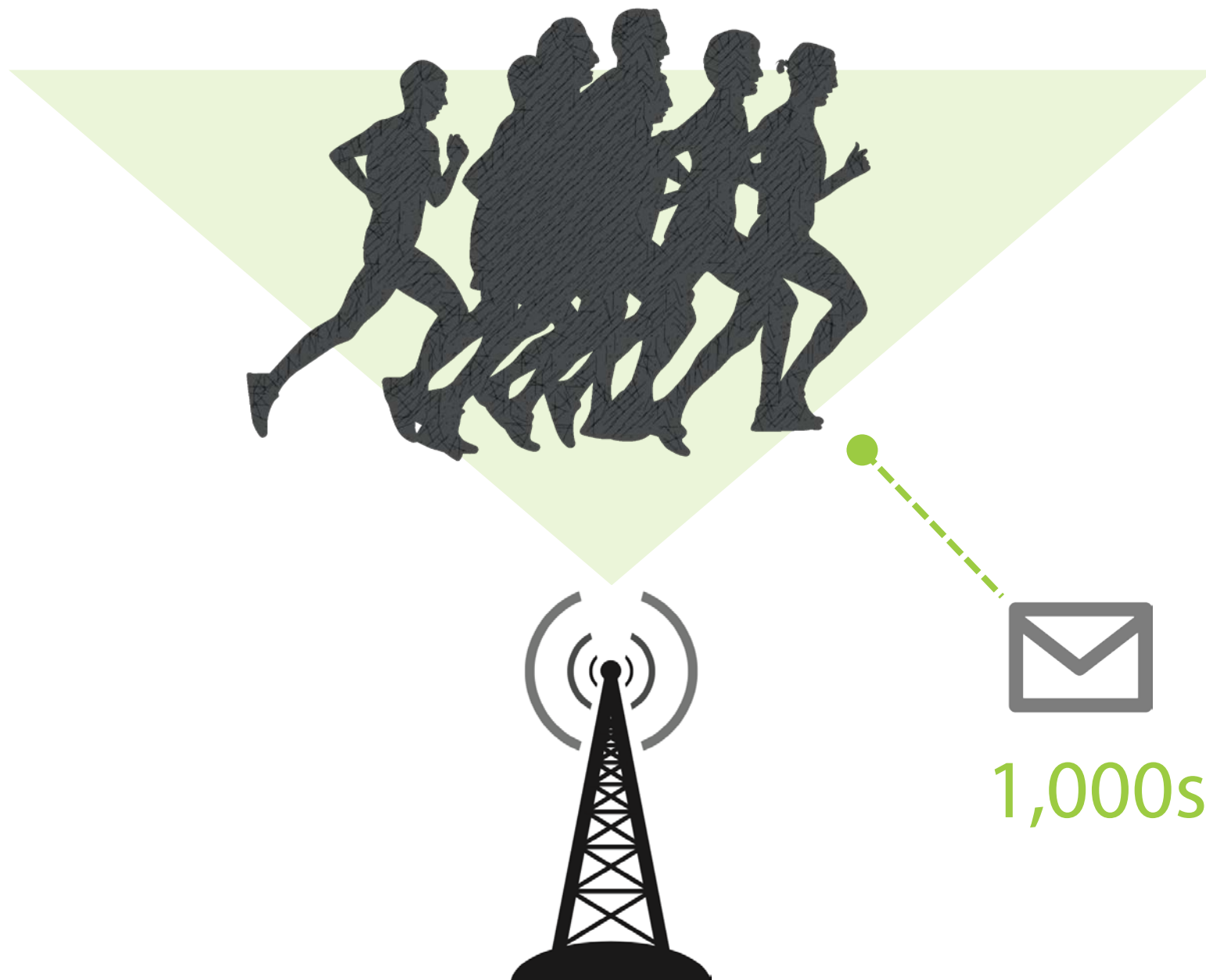
Volume

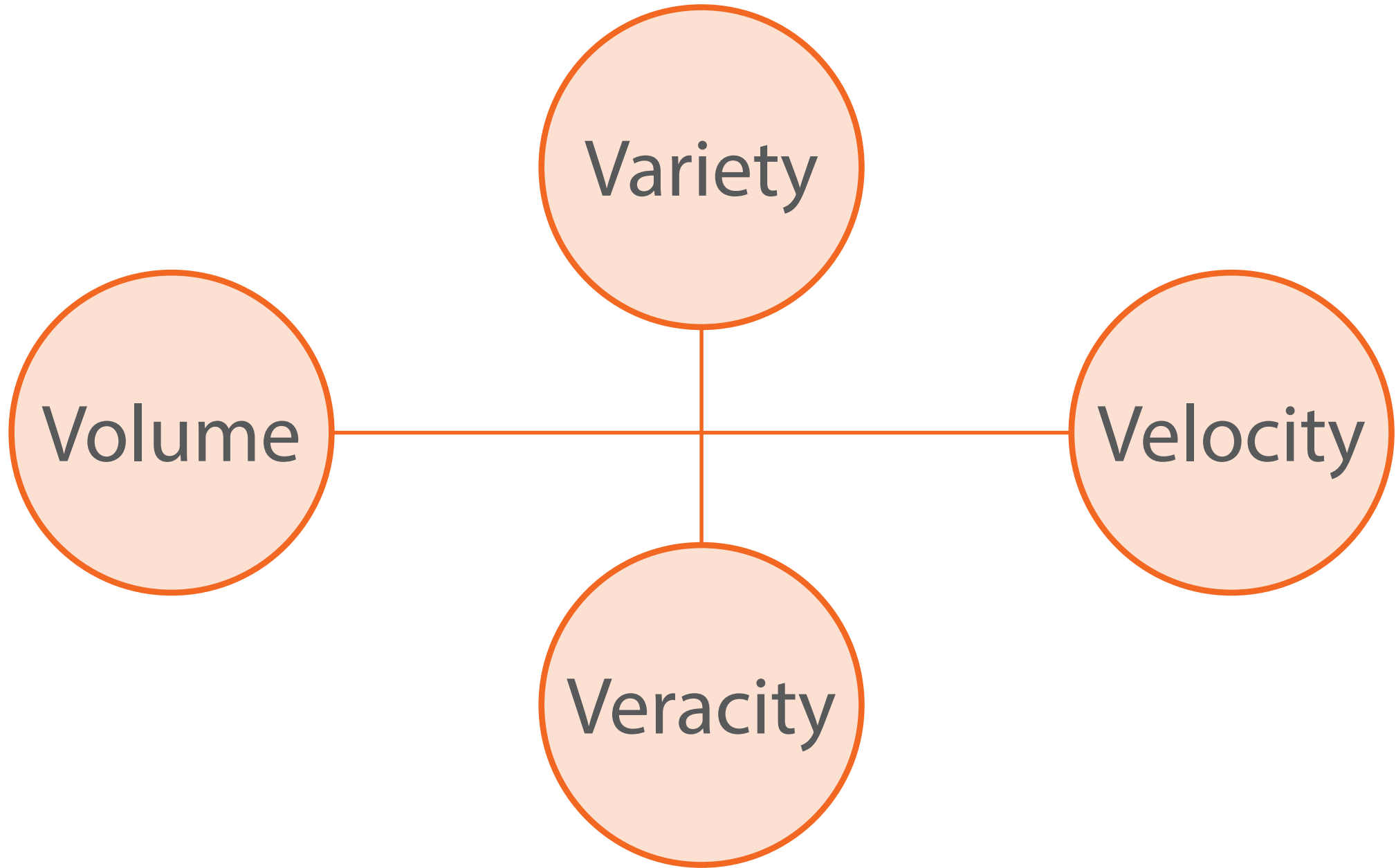


Racer ID	Timer ID	Timestamp
e4324	bc3b2	1231412509

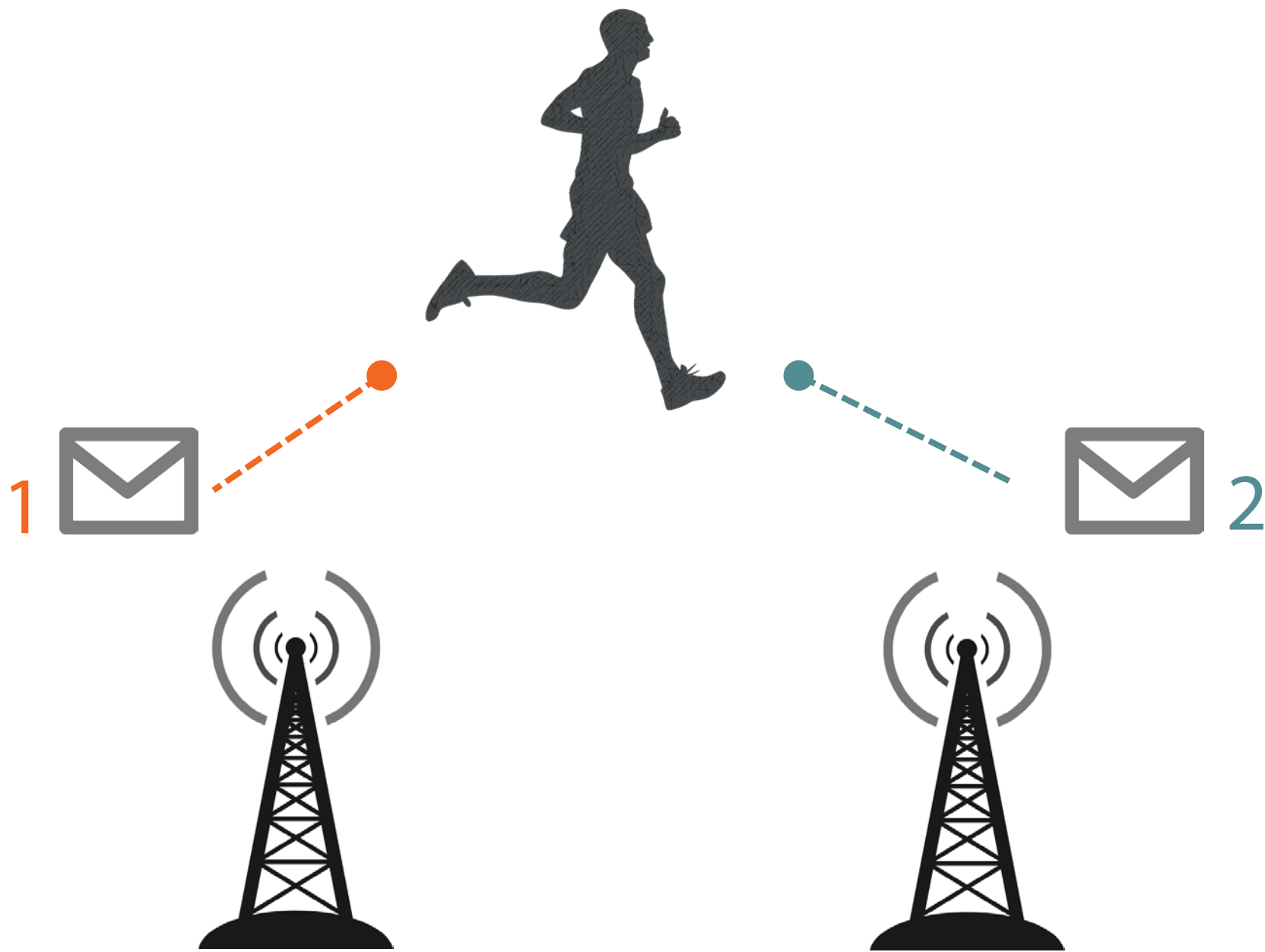


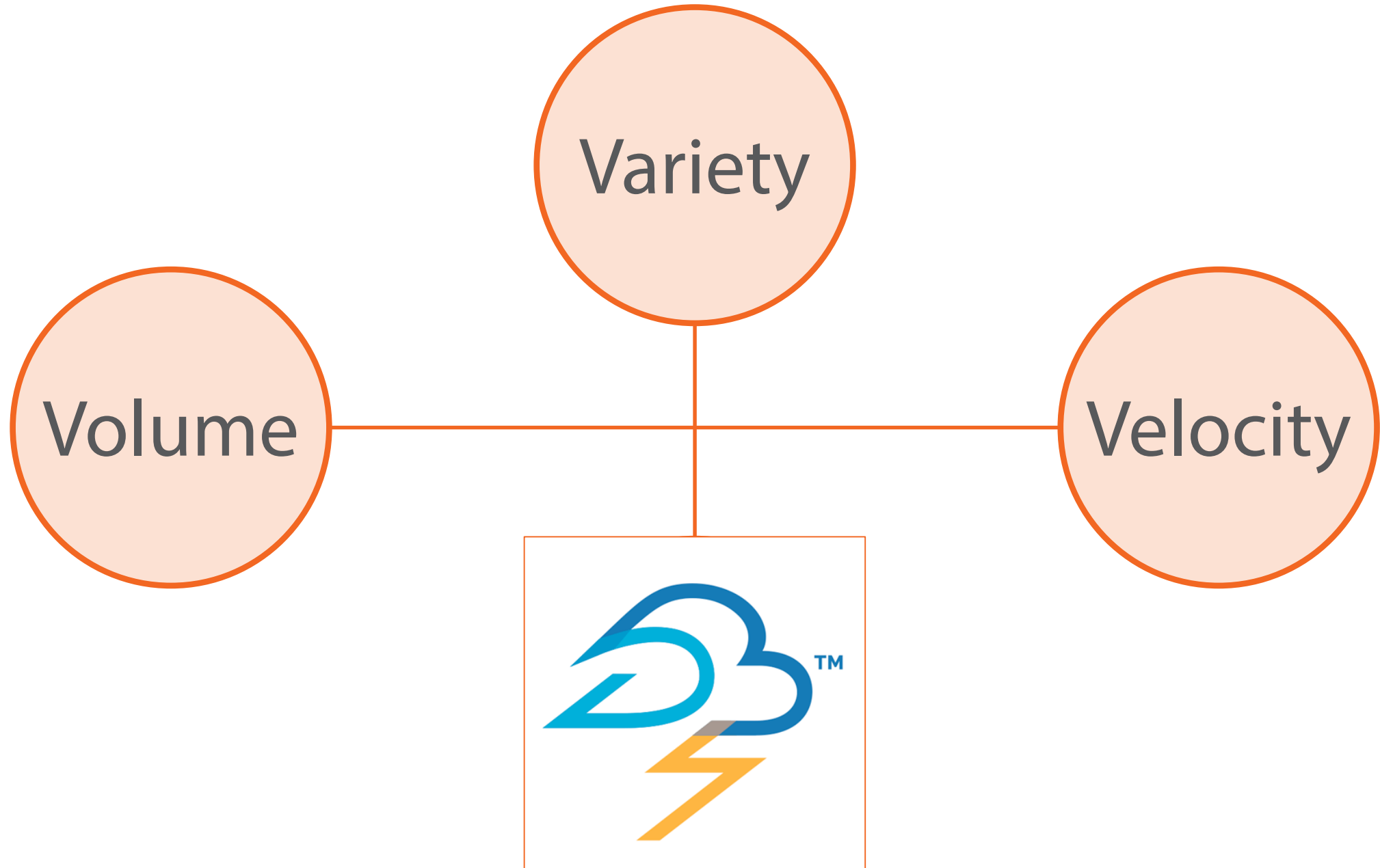
Velocity

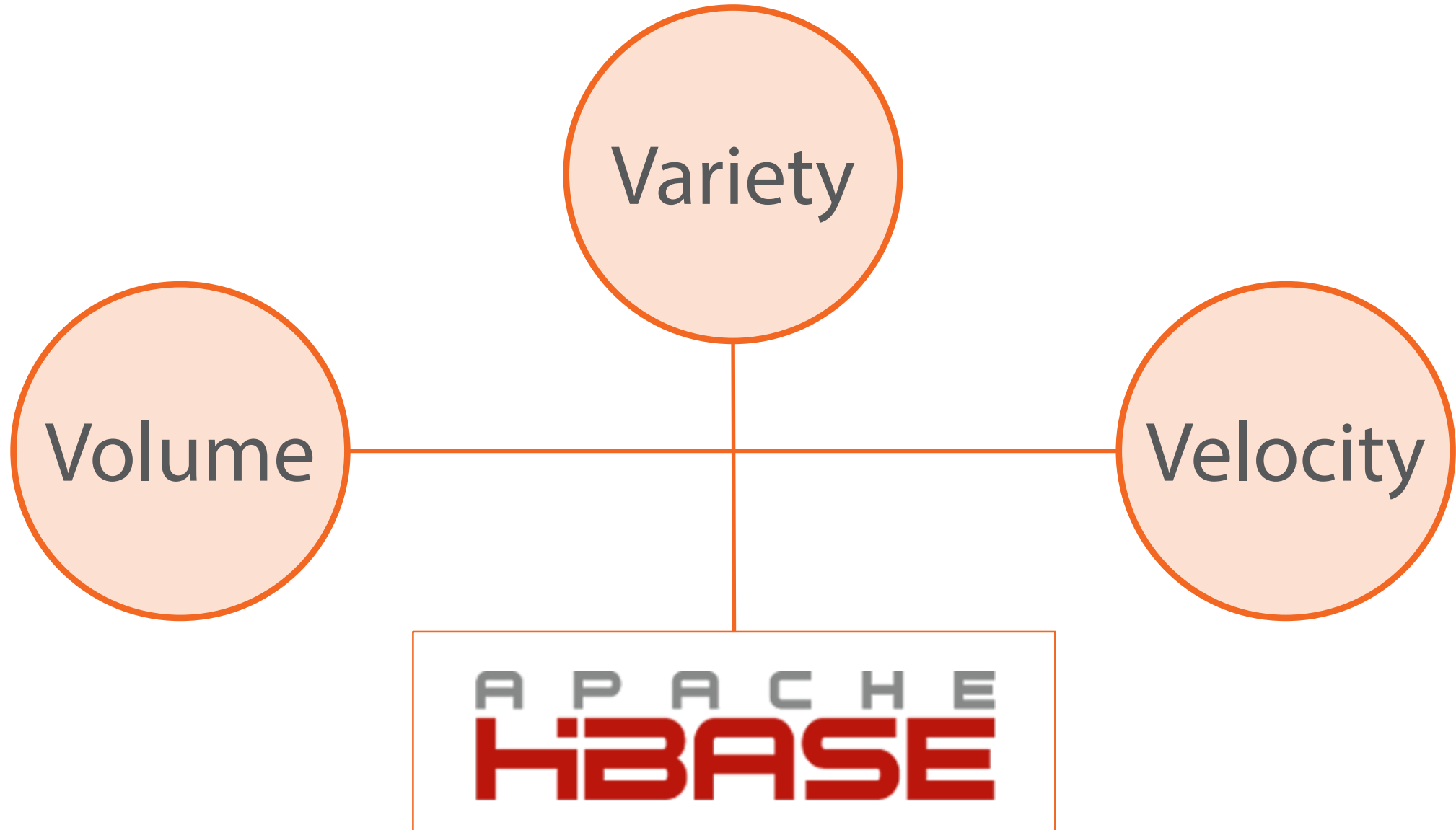


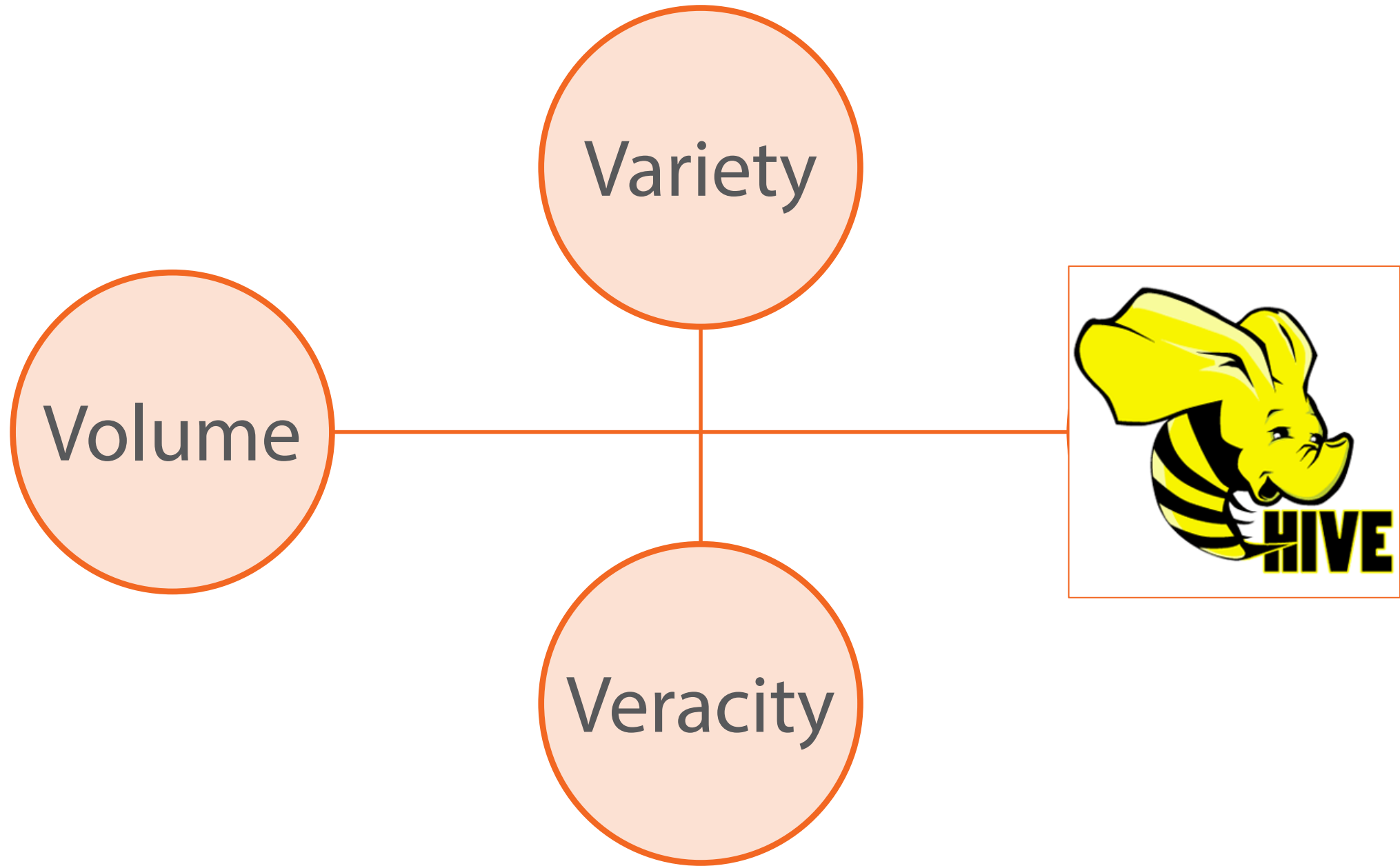


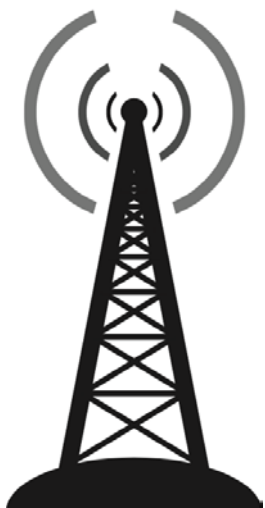
Veracity











Start events **x100**

Finish events **x100**

Duration **15-30 minutes**



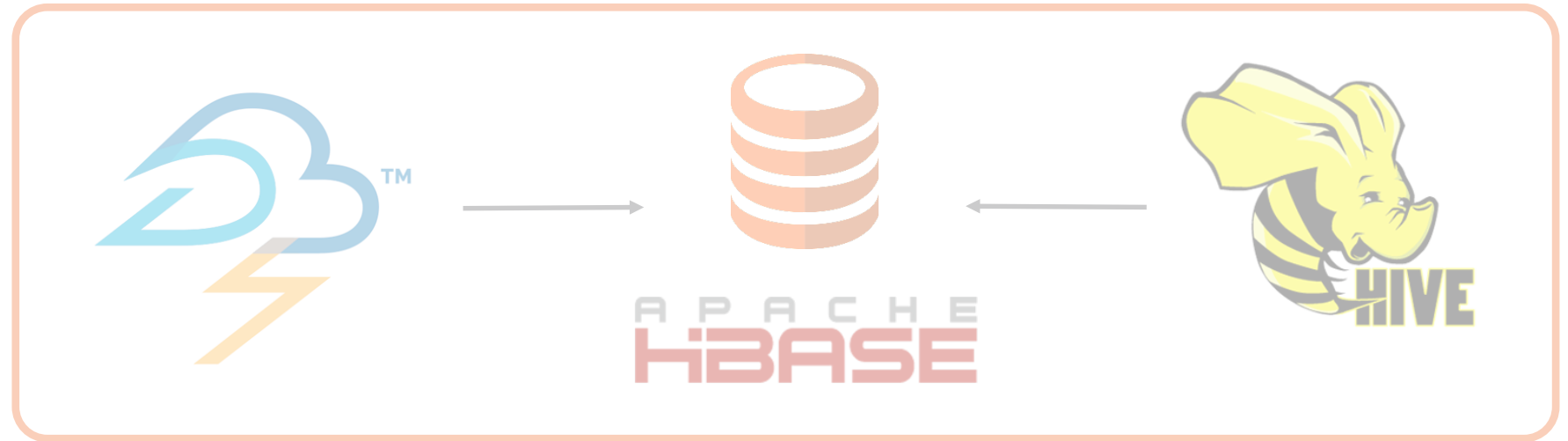
Frictionless **scale**

Low volume: **small clusters**

High volume: **big clusters**



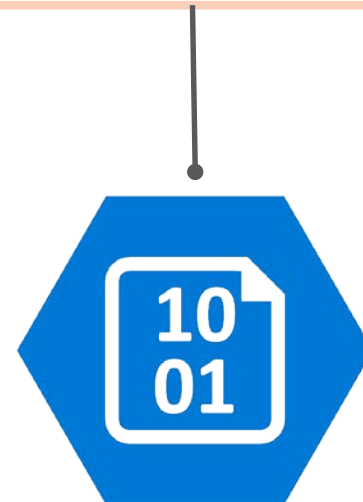
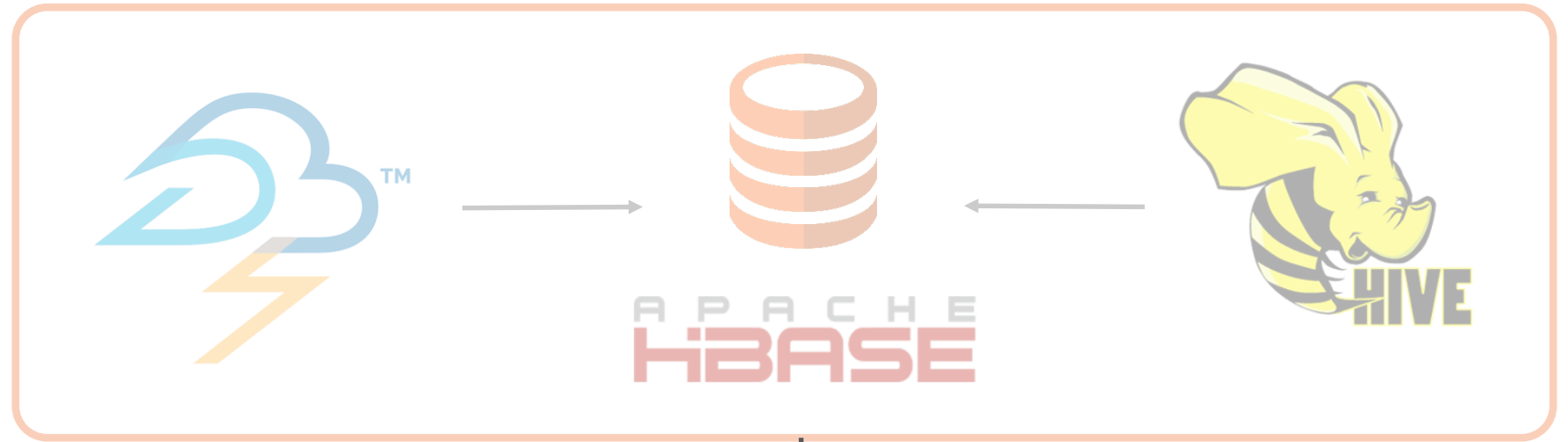
Fexible **cost model**
Pay-for-use compute

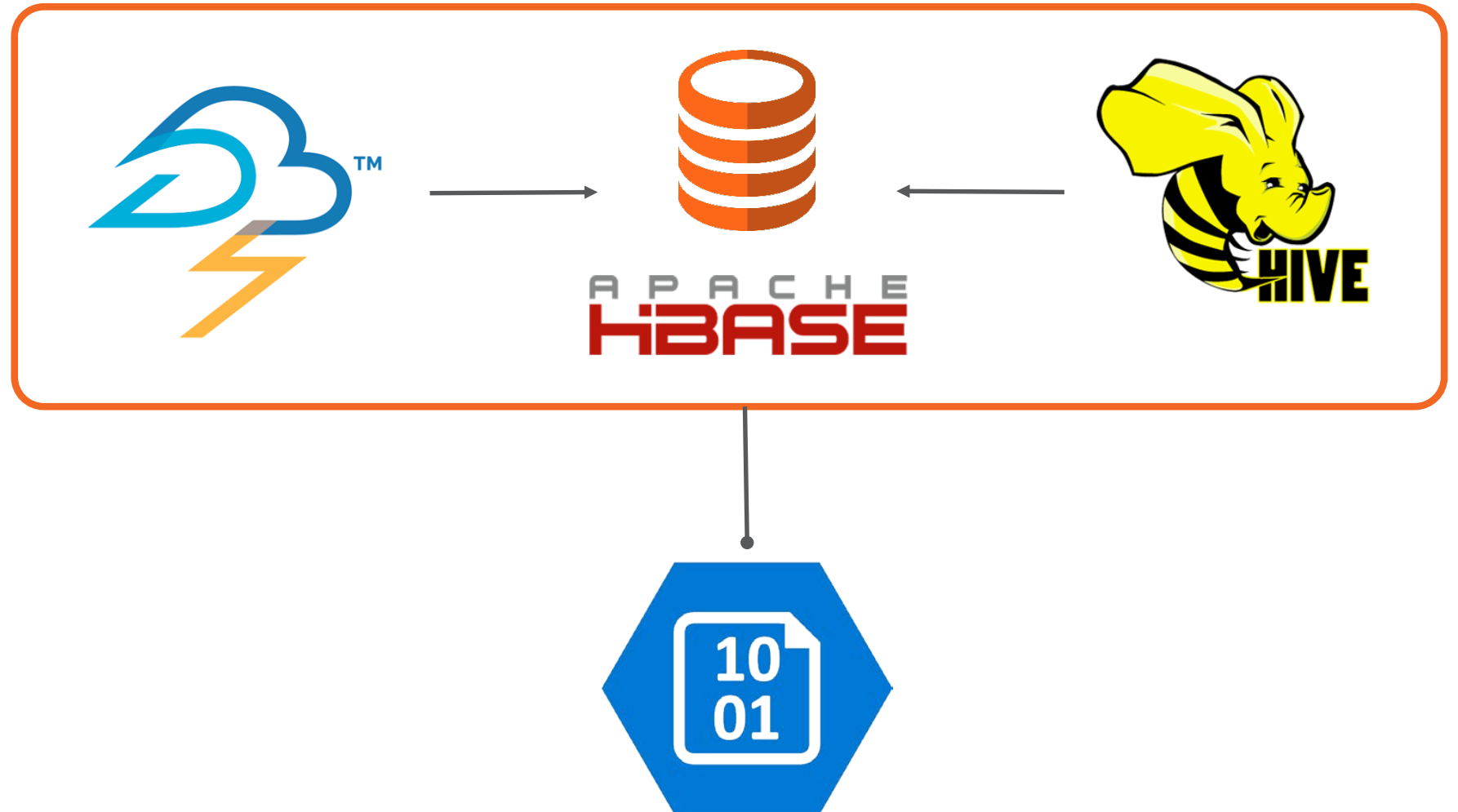


Flexible **cost model**

Pay-for-use compute

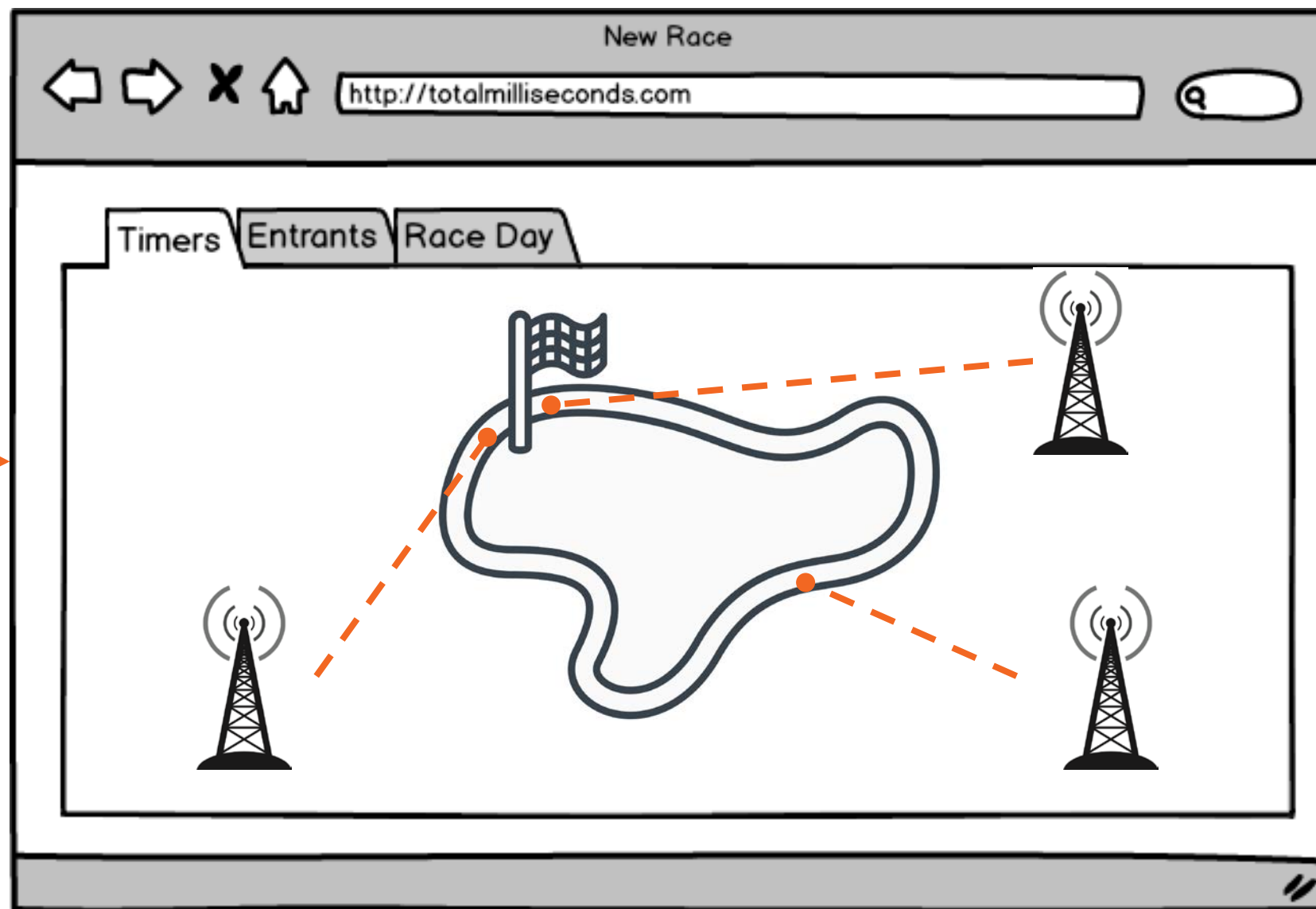
No races: **no clusters** = \$0







Minimal **custom dev**
Performance from stack
Scale from stack



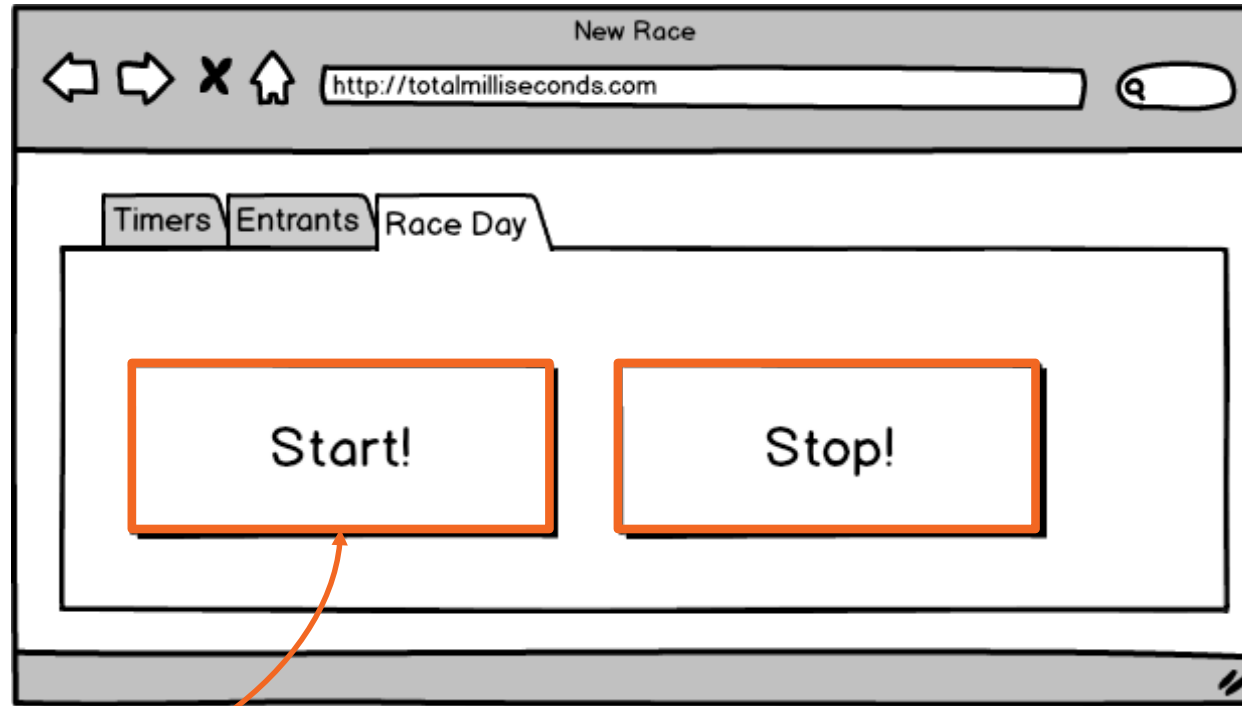


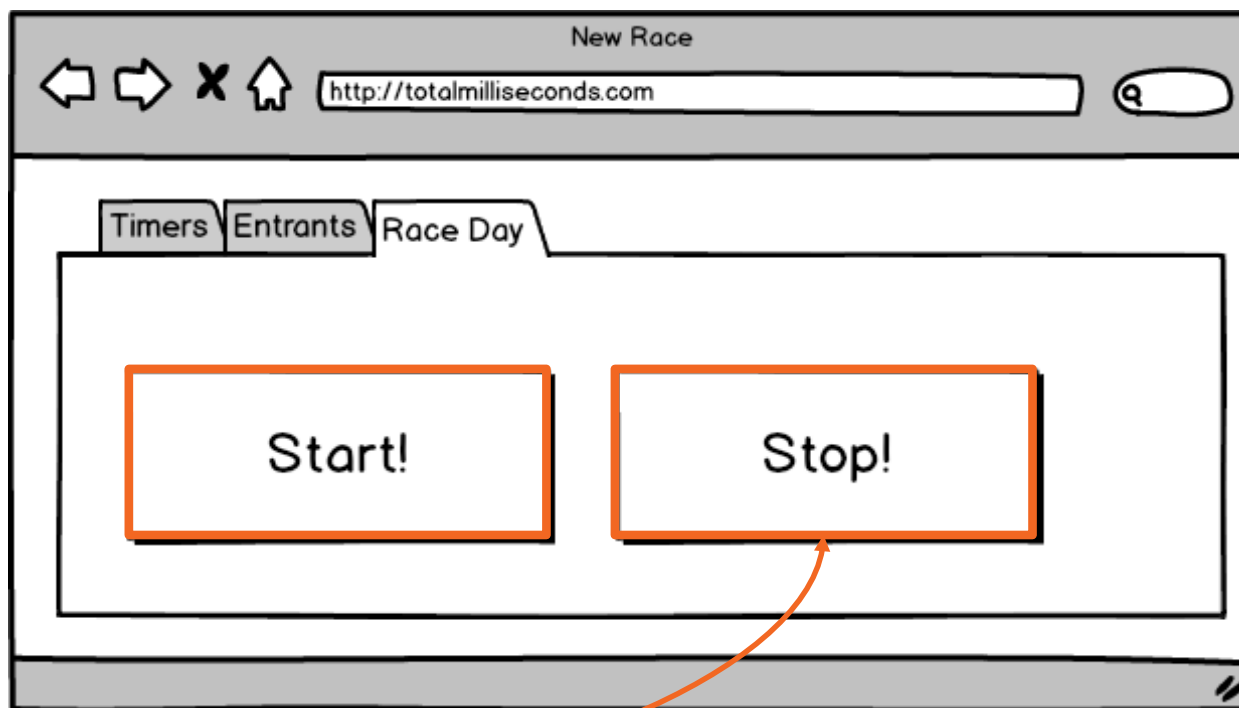
New Race

http://totalmilliseconds.com

Timers Entrants Race Day

First Name	Last Name	Racer ID
Vernice	Sperazza	e35761
Vicente	Rawicki	15bb9c
Craig	Cua	ee03e7
Jenise	Dulle	8c78ca
Marylin	Cornelison	e08b91
Marget	Gunst	6a1e37
Annett	Bunselmeyer	c61f39





```
c:\RaceSimulator.exe
```





Azure **Event Hub**
Receives **Timing Events**
Queued until read



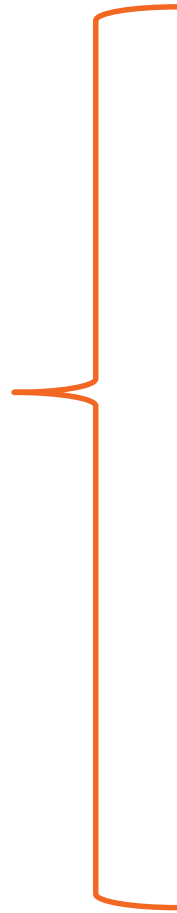
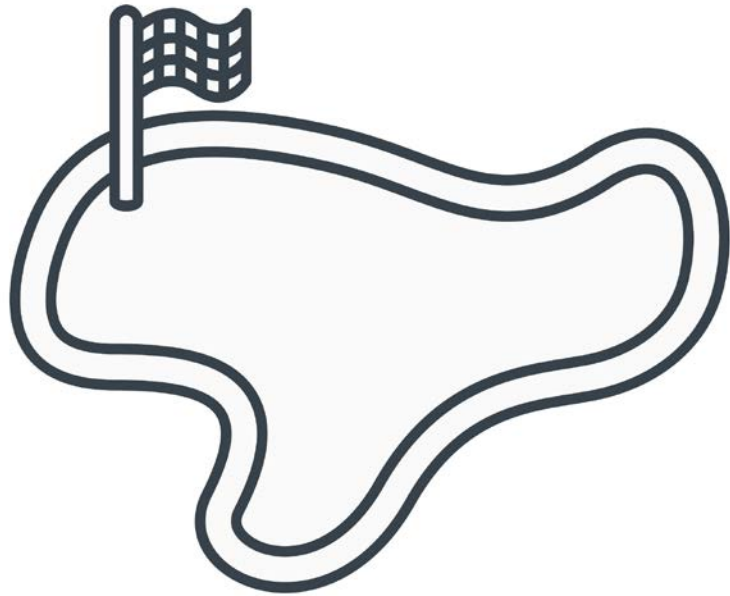
Storm on HDInsight
Consumes **Timing Events**
Processes and **Stores**

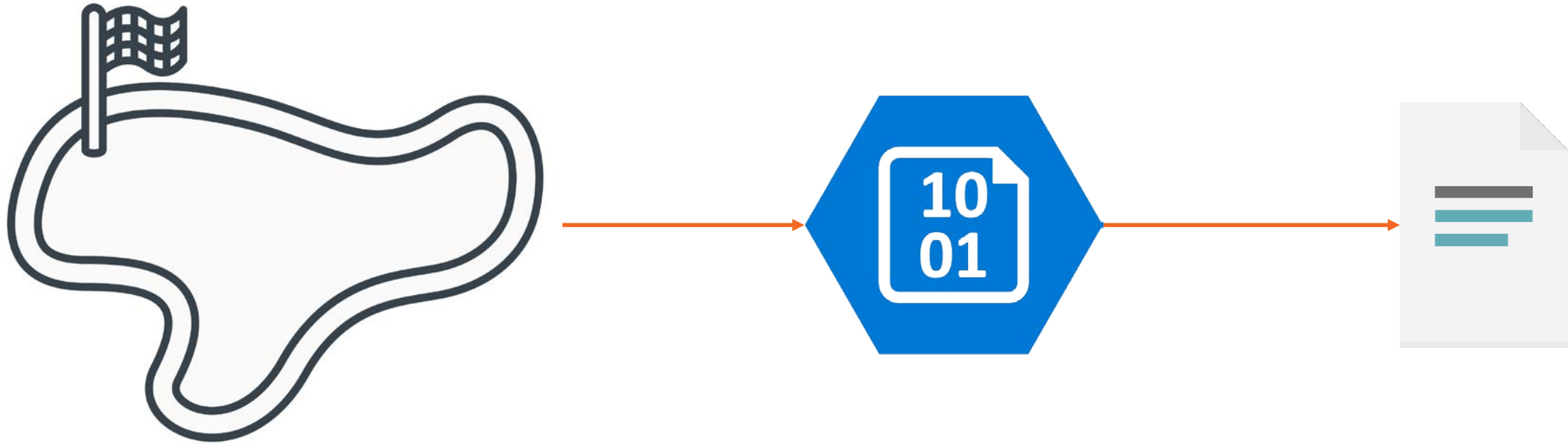


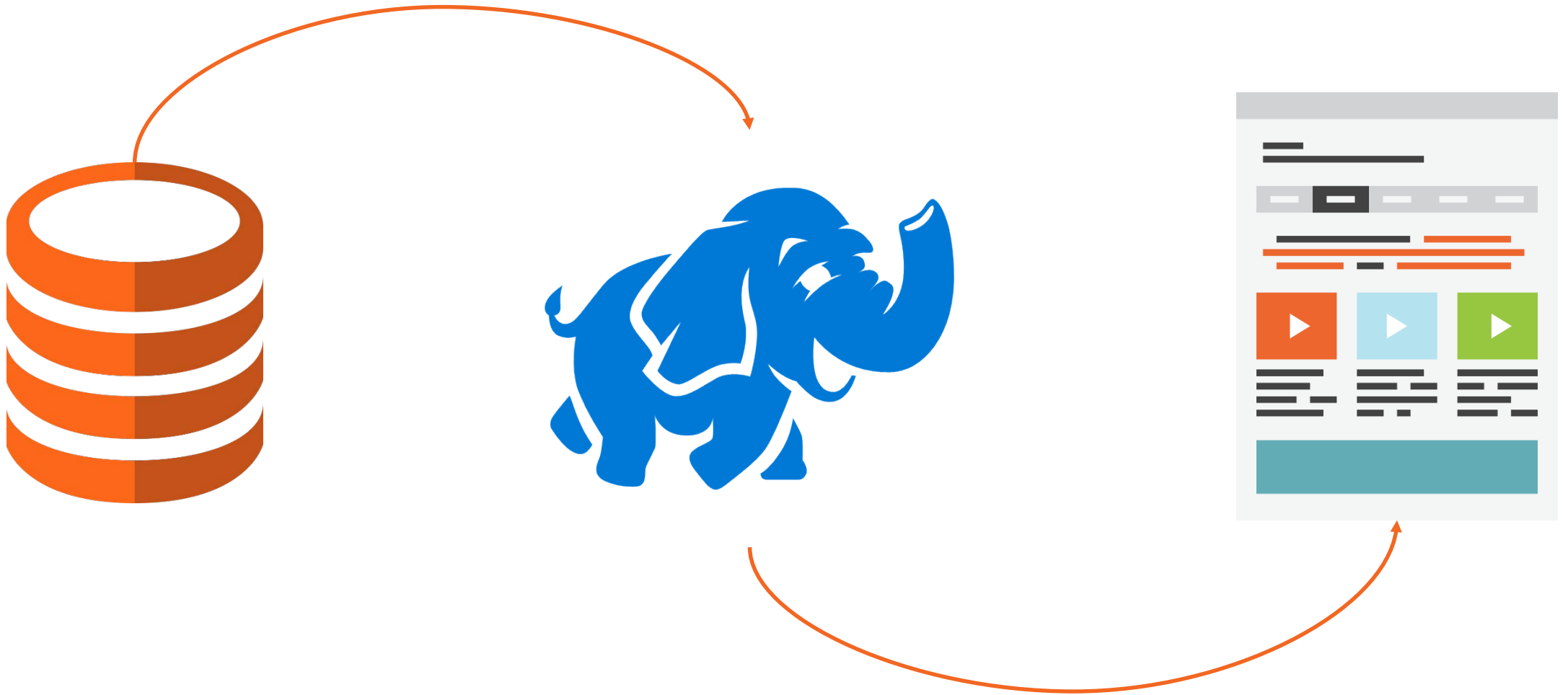
HBase on HDInsight
Stores raw **Timing Events**
Stores **Provisional Results**

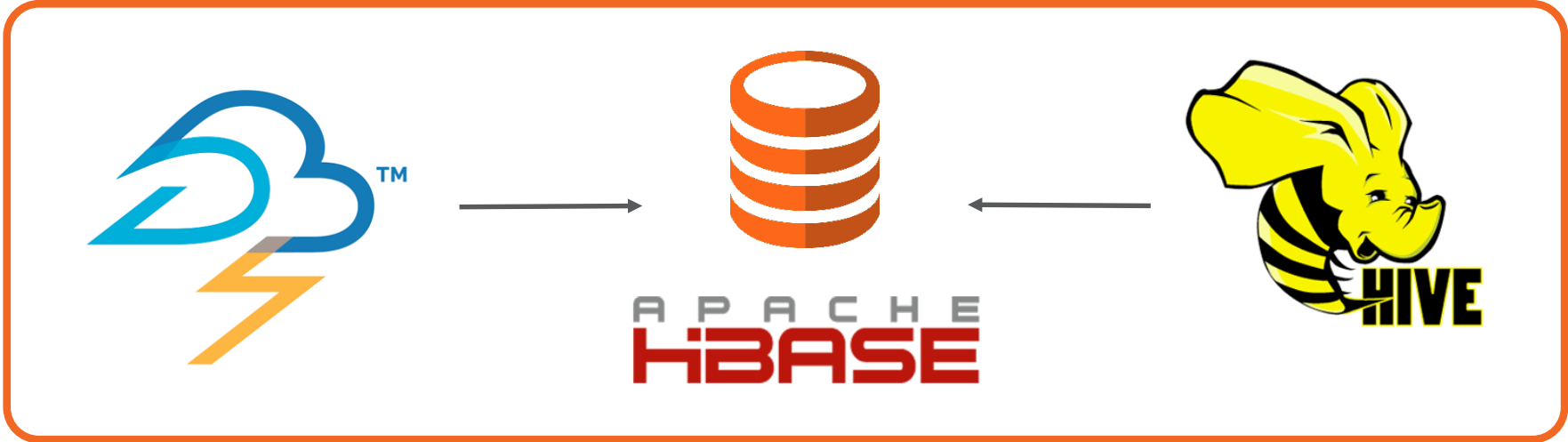


Hive on HDInsight
Generates **Final Results**
Joins **HBase** and **CSV** data









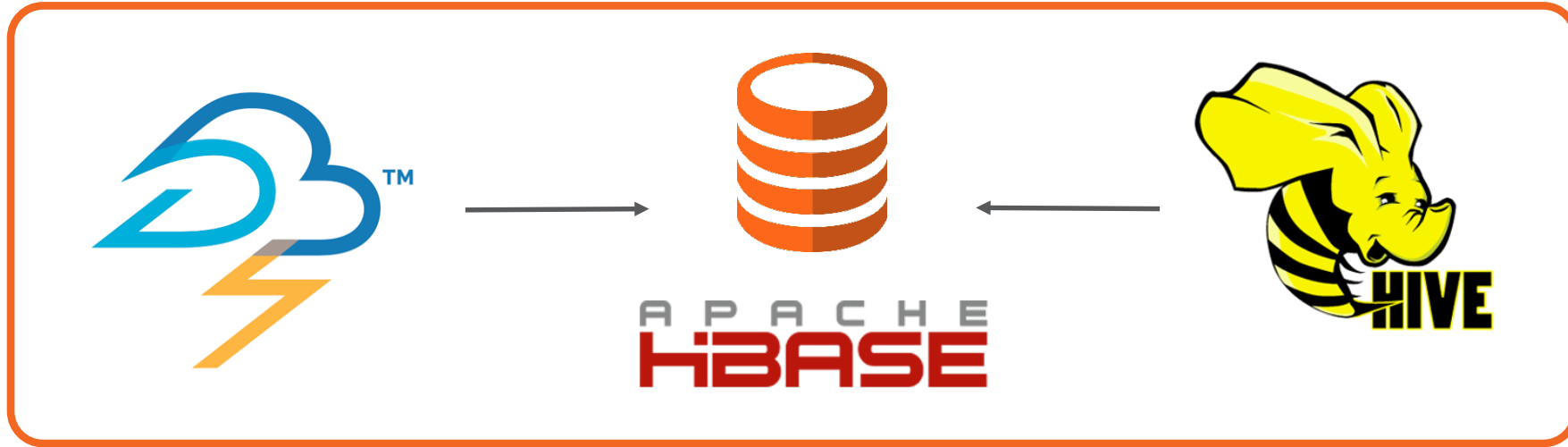


Real World Big Data in Azure

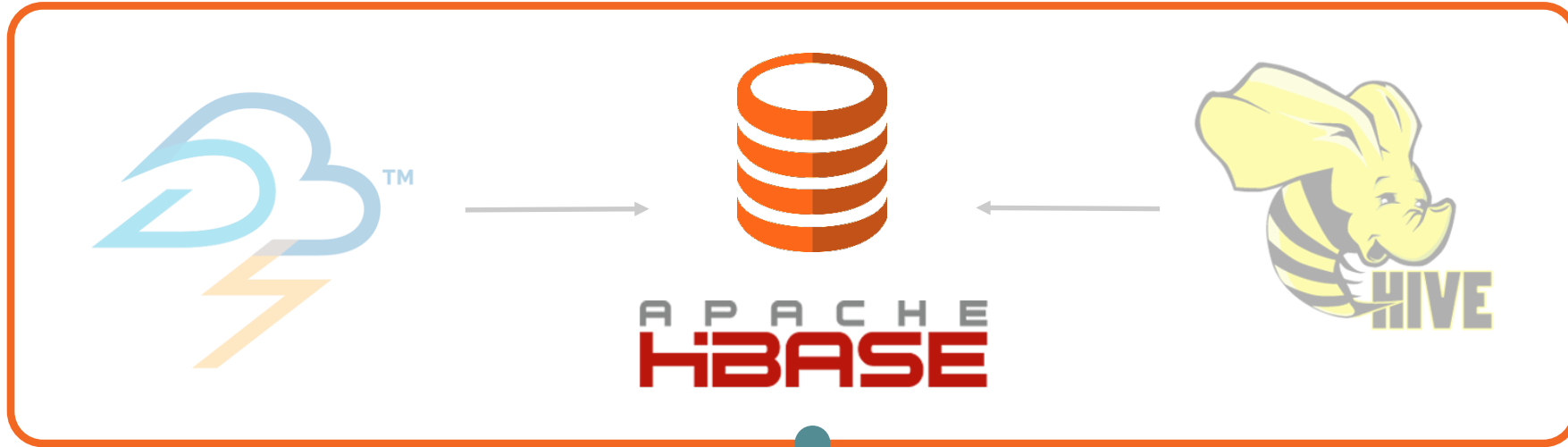


Elton Stoneman

@EltonStoneman | blog.sixeyed.com



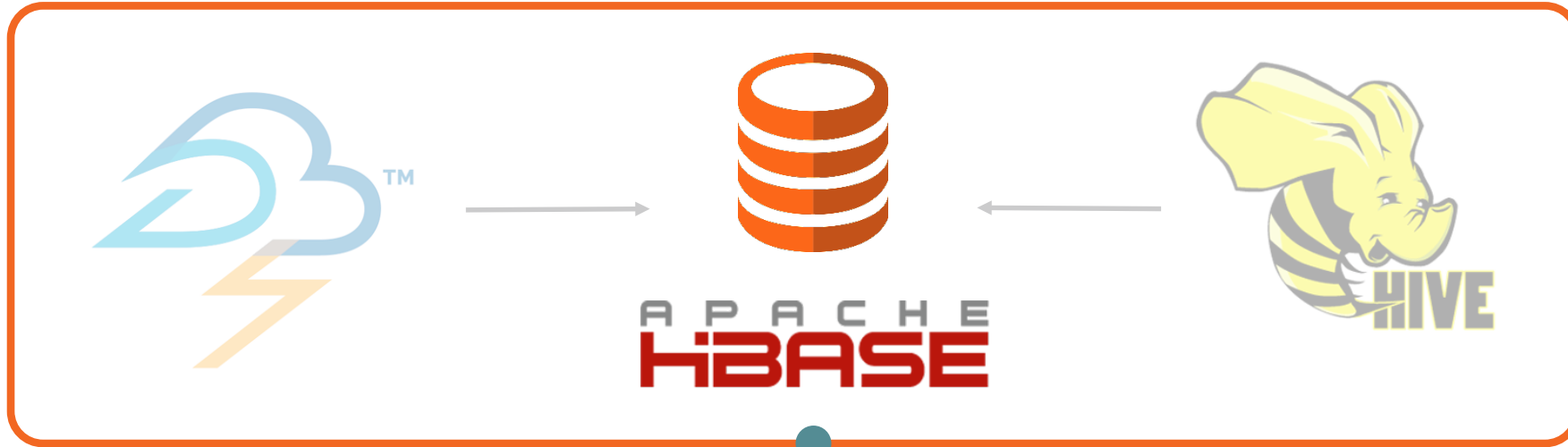
1x Functional Module
1x Deep Dive



Storing Race Data in HBase

.NET Data Access Layer

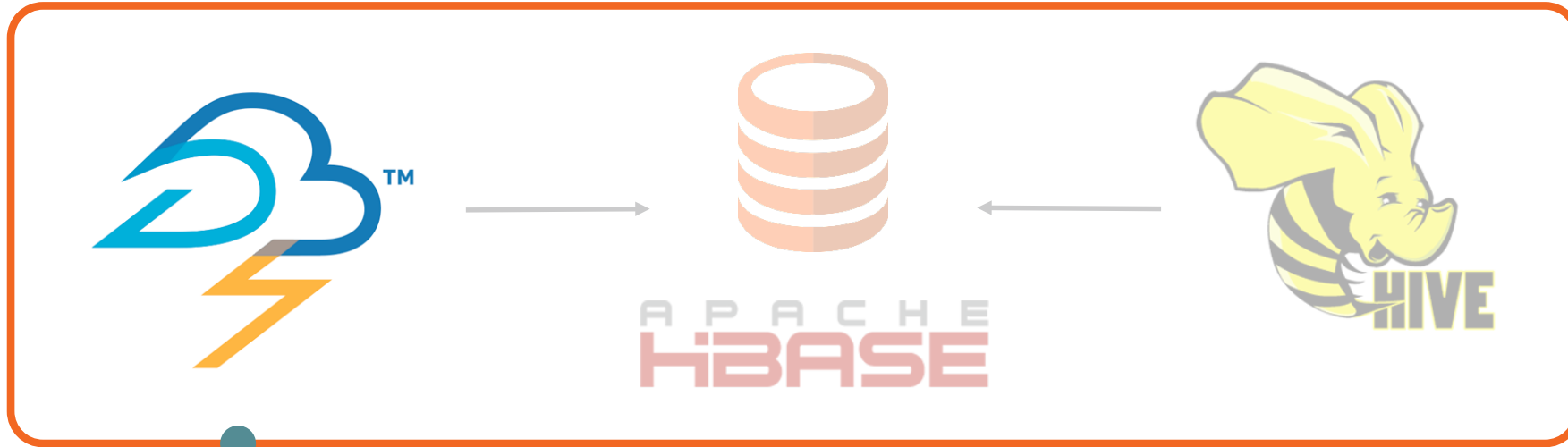
HBase Table Design



HBase Deep Dive

Cluster Administration

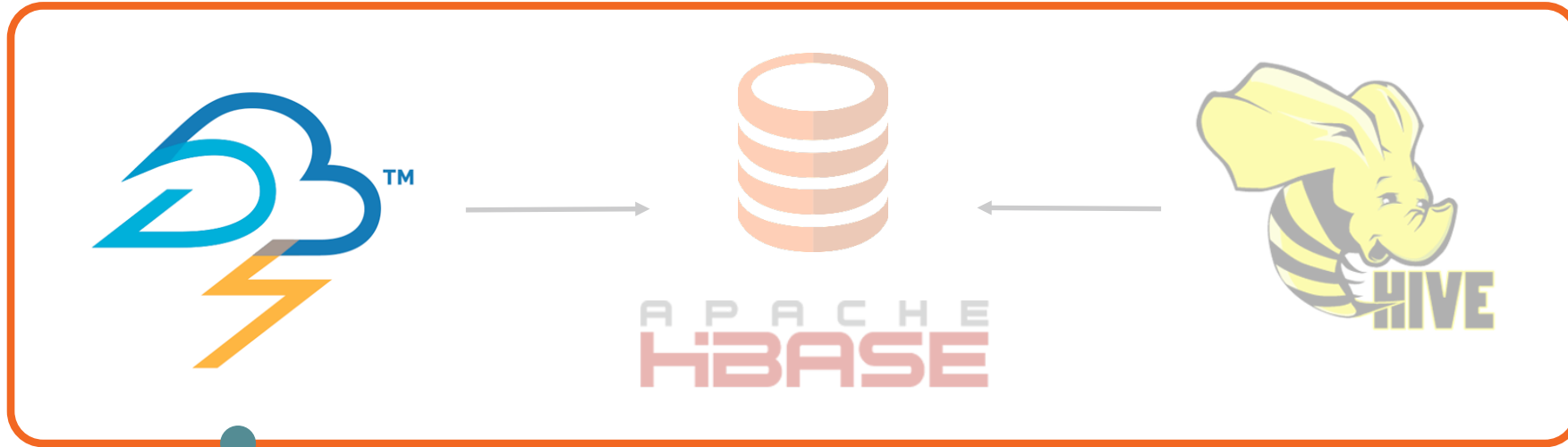
Read & Write Performance



Processing Timing Events with Storm

Consume from Event Hub

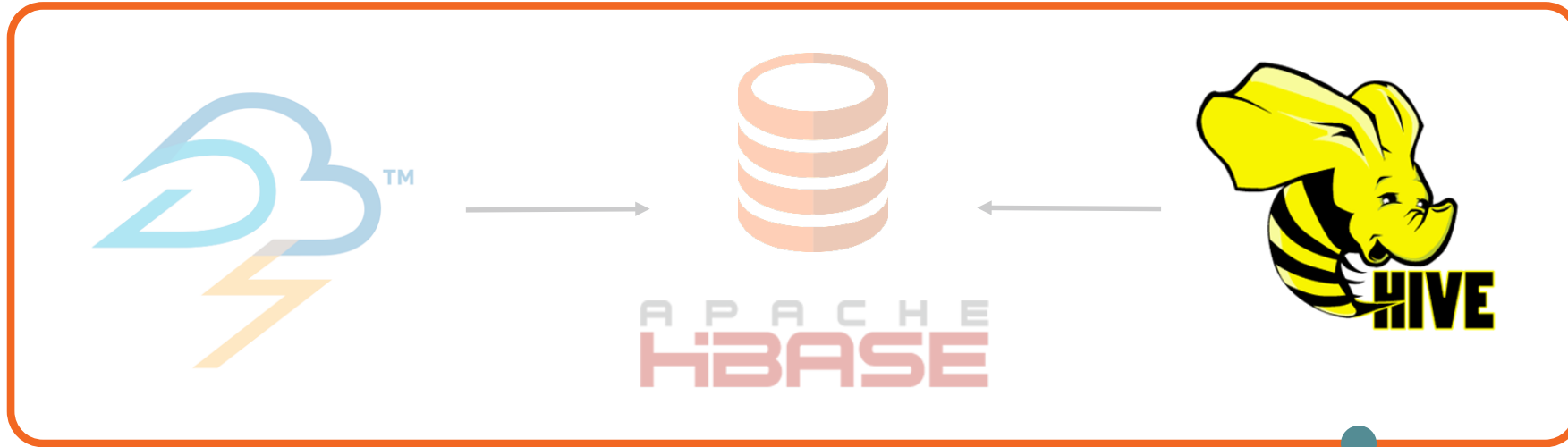
Process & Store in HBase



Storm Deep Dive

Distributed Architecture

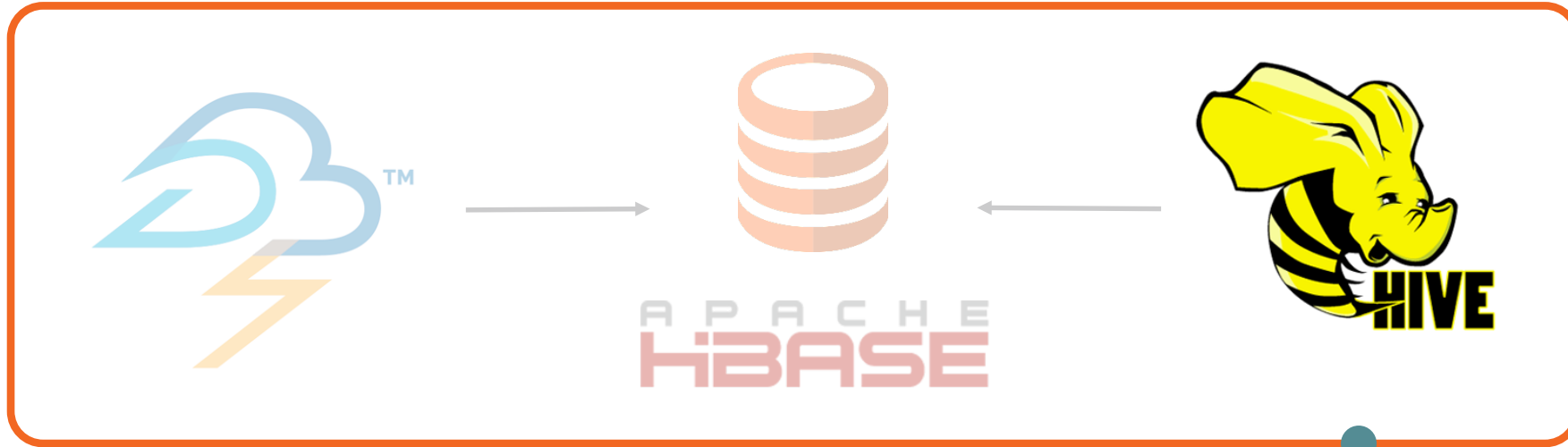
Performance & Reliability



Querying Race Data with Hive

Read from HBase & CSV

Write Results to Azure



Hive Deep Dive

Query Performance

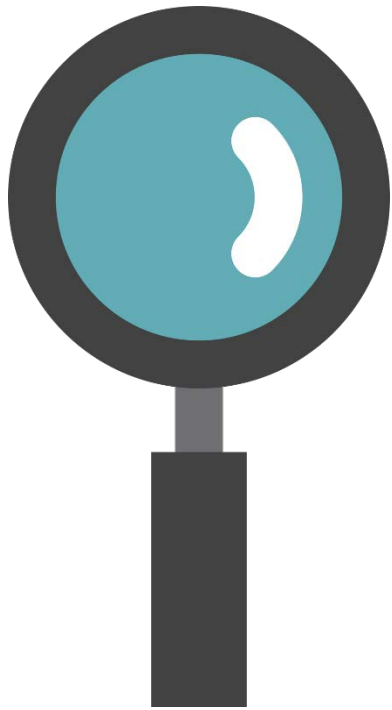
PowerShell & .NET



APACHE
HBASE



Coming Next



Structuring data in HBase

Data access patterns

Accessing HBase from .NET

Running HBase on HDInsight