

We've long relied on our canine friends to provide boozy sustenance in our times of greatest need, ever since 16th-century St. Bernards revived snowbound Alpine travelers with mini-kegs of brandy. Actually, that's [almost entirely made up](#). But hey, this dog can fetch beer from the fridge on command!

Name	Path	Method	Status Text	Type	Initiator	Size Content	Time Latency	Timeline	3.3 min	5.0 min
pixel?google_nid=crimtan&google_push=AHNF13KjgCbRr69Yheh_PkPU...	cm.doubleclick.net	GET	200 OK	image/png	http://i.ctnsnet.com/int... Redirect	506 B 170 B	115 ms 113 ms			
?h.key=8HB75-GTLRZ-D5A8S-LFUSZ-UUNZF&rt.si=5aefad33-feab-45...	36e4f0de.mpstat.us	GET	204 No Content	application/...	Other	233 B 0 B	118 ms 117 ms			
pixel.gif?source=smarttag&fired=user_data_timeout&confid=JMWgLQa...	beacon.krxd.net	GET	204	image/gif	Other	599 B	120 ms			
pixel.gif?e=11&i=GAWKERV4&k=1%3D1x1%3Ahttp%3A%	v4.moatads.com									
track.gif?&objectType=permalinkTag&objectType=perm...	inja.com/api/analytics/stats									
t?pid=52e9531be79548ced6000008&title=This%20Beer...	edge.simplereach.com									
t?pid=52e9531be79548ced6000008&title=This%20Beer...	edge.simplereach.com									
t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D...	edge.simplereach.com	GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	125 ms 122 ms			
t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D...	edge.simplereach.com	GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	125 ms 122 ms			
refresh?callback=jQuery211012063115951605141_1414617120253&...	inja.com/api/analytics/t...	GET	200 OK	text/javascript	Main-en-US-4159edb7... Script	799 B 190 B	125 ms 123 ms			
t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D...	edge.simplereach.com	GET	200 OK	application/j...	reach.js:8 Script	263 B 103 B	125 ms 123 ms			
t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D...	edge.simplereach.com	GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	125 ms 123 ms			
t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D...	edge.simplereach.com	GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	125 ms 123 ms			
t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D...	edge.simplereach.com	GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	126 ms 124 ms			
t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D...	edge.simplereach.com	GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	126 ms 124 ms			

Behavioural Tracking

Architecture for data pipelines

Outline

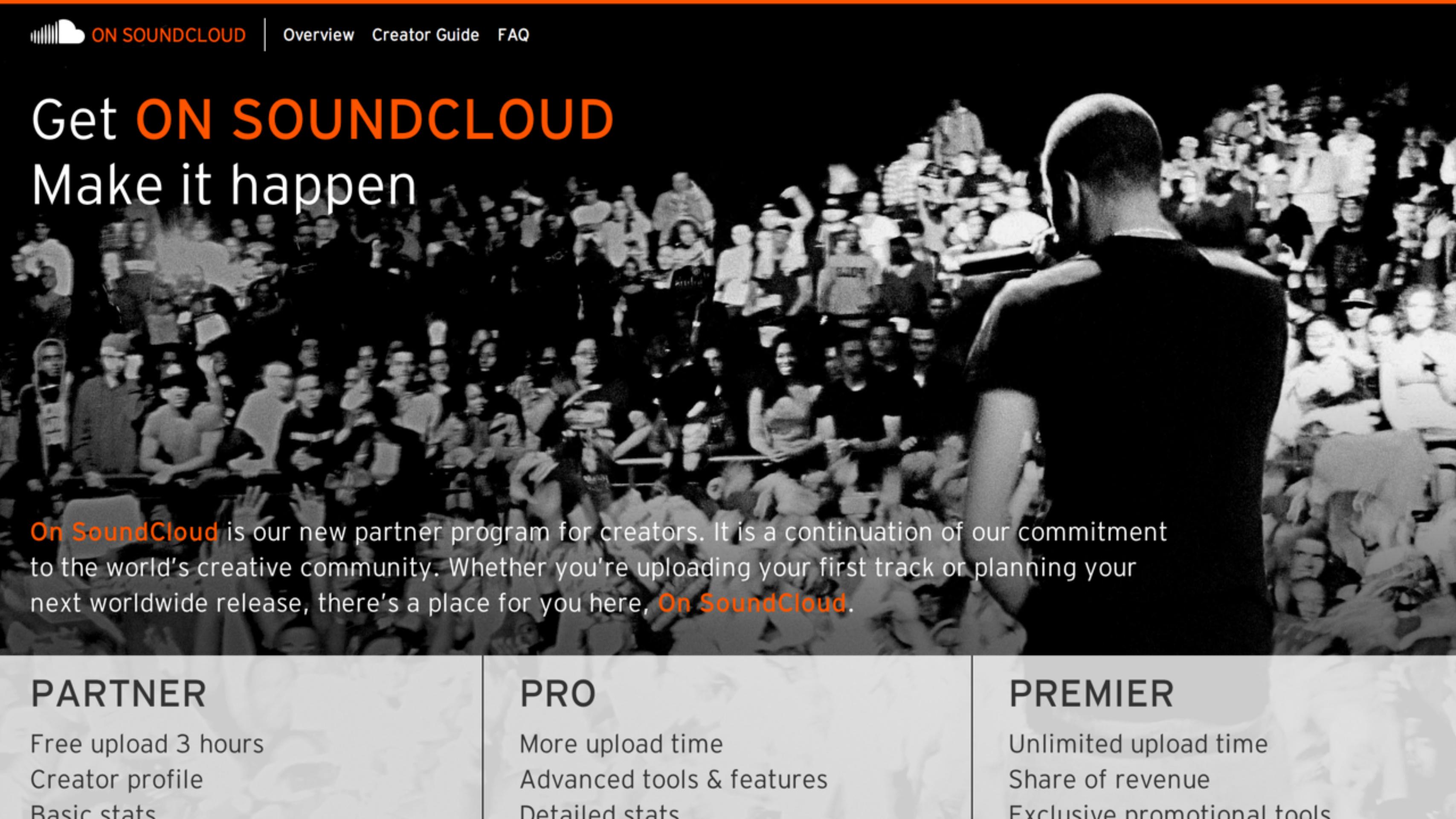
- Motivation
- Technical requirements
- Architecture
- Example: Stitch
- Event sourcing
- Alternative implementation
- Cloud computing
- Open problems
- Q and A



SOUNDCLOUD

Get ON SOUND CLOUD

Make it happen



On SoundCloud is our new partner program for creators. It is a continuation of our commitment to the world's creative community. Whether you're uploading your first track or planning your next worldwide release, there's a place for you here, On SoundCloud.

PARTNER

Free upload 3 hours
Creator profile
Basic stats

PRO

More upload time
Advanced tools & features
Detailed stats

PREMIER

Unlimited upload time
Share of revenue
Exclusive promotional tools

We've long relied on our canine friends to provide boozy sustenance in our times of greatest need, ever since 16th-century St. Bernards revived snowbound Alpine travelers with mini-kegs of brandy. Actually, that's **almost entirely made up**. But hey, this dog can fetch beer from the fridge on command!

Name	Path	Method	Status Text	Type	Initiator	Size Content	Time Latency	Timeline	3.3 min	5.0 min
 pixel?google_nid=crimtan&google_push=AHNF13KjgCbRr69Yheh_PkPU... cm.doubleclick.net		GET	200 OK	image/png	http://i.ctnsnet.com/int... Redirect	506 B 170 B	115 ms 113 ms			
 ?h.key=8HB75-GTLRZ-D5A8S-LFUSZ-UUNZF&rt.si=5aefad33-feab-45... 36e4f0de.mpstat.us		GET	204	application/	Other	233 B	118 ms			
 pixel.gif?source=smarttag&fired=user_data_timeout&confid=JMWgLC beacon.krxd.net										
 pixel.gif?e=11&i=GAWKERV4&k=1%3D1x1%3Ahttp%3A%2F%2Fgawke... v4.moatads.com										
 track.gif?&objectType=permalinkTag&objectType=permalinkTag&obj... kinja.com/api/analytics/stats										
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%2... edge.simplereach.com										
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%2... edge.simplereach.com										
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%2... edge.simplereach.com										
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D... edge.simplereach.com		GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	125 ms 122 ms			
 refresh?callback=jQuery211012063115951605141_1414617120253&... kinja.com/api/analytics/t		GET	200 OK	text/javascript	Main-en-US-4159edb7... Script	799 B 190 B	125 ms 123 ms			
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D... edge.simplereach.com		GET	200 OK	application/j...	reach.js:8 Script	263 B 103 B	125 ms 123 ms			
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D... edge.simplereach.com		GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	125 ms 123 ms			
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D... edge.simplereach.com		GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	125 ms 123 ms			
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D... edge.simplereach.com		GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	126 ms 124 ms			
 t?pid=52e9531be79548ced6000008&title=This%20Beer-Fetching%20D... edge.simplereach.com		GET	200 OK	application/j...	reach.js:8 Script	264 B 104 B	126 ms 124 ms			

Tracking Pixels

- Sending data from browser
- Backbone of Ad-tech
- Ubiquitous throughout the web

Motivation for tracking

Operational

- Logs
- Payment reports

Product

- Creator analytics
- Recommendation

Insight

- Analytics
- Research

Technical Requirements

- Collect and transmit data
- Partition, process, filter
- Make available to people

Architecture Overview



```

85
86     where = document.getElementsByTagName('script')[0];
87     where.parentNode.insertBefore(iframe, where);
88
89     try {
90         doc = iframe.contentWindow.document;
91     } catch(e) {
92         dom = document.domain;
93         iframe.src="javascript:var d=document.open();d.domain='"+dom+"'";void(0);
94         doc = iframe.contentWindow.document;
95     }
96     doc.open()._l = function() {
97         var js = this.createElement("script");
98         if(dom) this.domain = dom;
99         js.id = "boomer-if-as";
100        js.src = '//c.go-mpulse.net/
101        '8HB75-GTLRZ-D5A8S-LFUSZ-UUN
102        BOOMR_lstart=new Date().getTime();
103        this.body.appendChild(js);
104    };
105    doc.write('<body onload="document.close()">');
106    doc.close();
107 })();
108 </script><script>window.kinja.script
109 type='text/javascript'>
110     try {
111         amznads.getAds('3076');
112     } catch(e) { console.log(e.message);
113     </script>      <script>
114     window.Krux || ((Krux = function () {
115         Krux.q.push(arguments);
116     }).q = []);
117     (function () {
118         function retrieve(n) {
119             var m, k = 'kx' + n;
120             if (window.localStorage) {
121                 return window.localStorage[k] || "";
122             } else if (navigator.cookieEnabled) {
123                 m = document.cookie.match(k + '=([^\n]*?)');
124                 return (m && unescape(m[1])) || "";
125             } else {
126                 return '';
127             }
128         }
129         Krux.user = retrieve('user');
130         Krux.segments = retrieve('segs') && retrieve('segs').split(',') || [];
131     })();
132     </script><!-- Included CSS Files --><!--[if IE 9]><link rel="stylesheet" href="http://c.kinja-static.com/assets/stylesheets/tiger-split-7ca07e9d79a959e3ed448de1c35bd9cc.css"><![endif]--><!--[if !IE]>--><link rel="stylesheet" href="http://c.kinja-static.com/assets/stylesheets/tiger-eldeb8fb2alf708836b8a6aalcbfcf07.css"><!--<![endif]--><meta name="dynamic-

```

Client side implementation

- Instrument application logic/state
- Schedule + plan for HTTP Calls
- Handle Failure

```

85
86     where = document.getElementsByTagName('script')[0];
87     where.parentNode.insertBefore(iframe, where);
88
89     try {
90         doc = iframe.contentWindow.document;
91     } catch(e) {
92         dom = document.domain;
93         iframe.src="javascript:var d=document.open();d.domain='"+dom+"'";void(0);";
94         doc = iframe.contentWindow.document;
95     }
96     doc.open()._l = function() {
97         var js = this.createElement("script");
98         if(dom) this.domain = dom;
99         js.id = "boomer-if-as";
100        js.src = '//c.go-mpulse.net/';
101        '8HB75-GTLRZ-D5A8S-LFUSZ-UUN
102        BOOMR_lstart=new Date().getTime();
103        this.body.appendChild(js);
104    };
105    doc.write('<body onload="document.close();">');
106    doc.close();
107 })();
108 </script><script>window.kinja.script
type='text/javascript'>
109     try {
110         amznads.getAds('3076');
111     } catch(e) { console.log(e.message);
112     </script>      <script>
113     window.Krux || ((Krux = function () {
114         Krux.q.push(arguments);
115     }).q = []);
116     (function () {
117         function retrieve(n) {
118             var m, k = 'kx' + n;
119             if (window.localStorage) {
120                 return window.localStorage[k] || "";
121             } else if (navigator.cookieEnabled) {
122                 m = document.cookie.match(k + '=([^\;]*)');
123                 return (m && unescape(m[1])) || "";
124             } else {
125                 return '';
126             }
127         }
128         Krux.user = retrieve('user');
129         Krux.segments = retrieve('segs') && retrieve('segs').split(',') || [];
130     })();
131     </script><!-- Included CSS Files --><!---[if IE 9]><link rel="stylesheet" href="http://c.kinja-static.com/assets/stylesheets/tiger-split-7ca07e9d79a959e3ed448de1c35bd9cc.css"><![endif]--><!---[if !IE]>--><link rel="stylesheet" href="http://c.kinja-static.com/assets/stylesheets/tiger-eldeb8fb2alf708836b8a6a1cbfcf07.css"><!--<![endif]--><meta name="dynamic-
132     link" content="1" /><meta name="fragment" content="!ID=101010-0660-0056016-16300510105" /><meta name="dynamic-link" content="1" />
```

Platforms Matter

- Battery life on mobile is an issue
- Keep alive, Scheduling, Batching
- Device ID and local storage

Collection

- Front door of the pipeline
- Highly available
- Consistent application protocol

Application Protocol

- Schemas matter
- Must communicate failure
- Evolutionary path

HA HTTP services

- DNS offers high level routing
- Load Balancers
- Circuit Breakers

Production

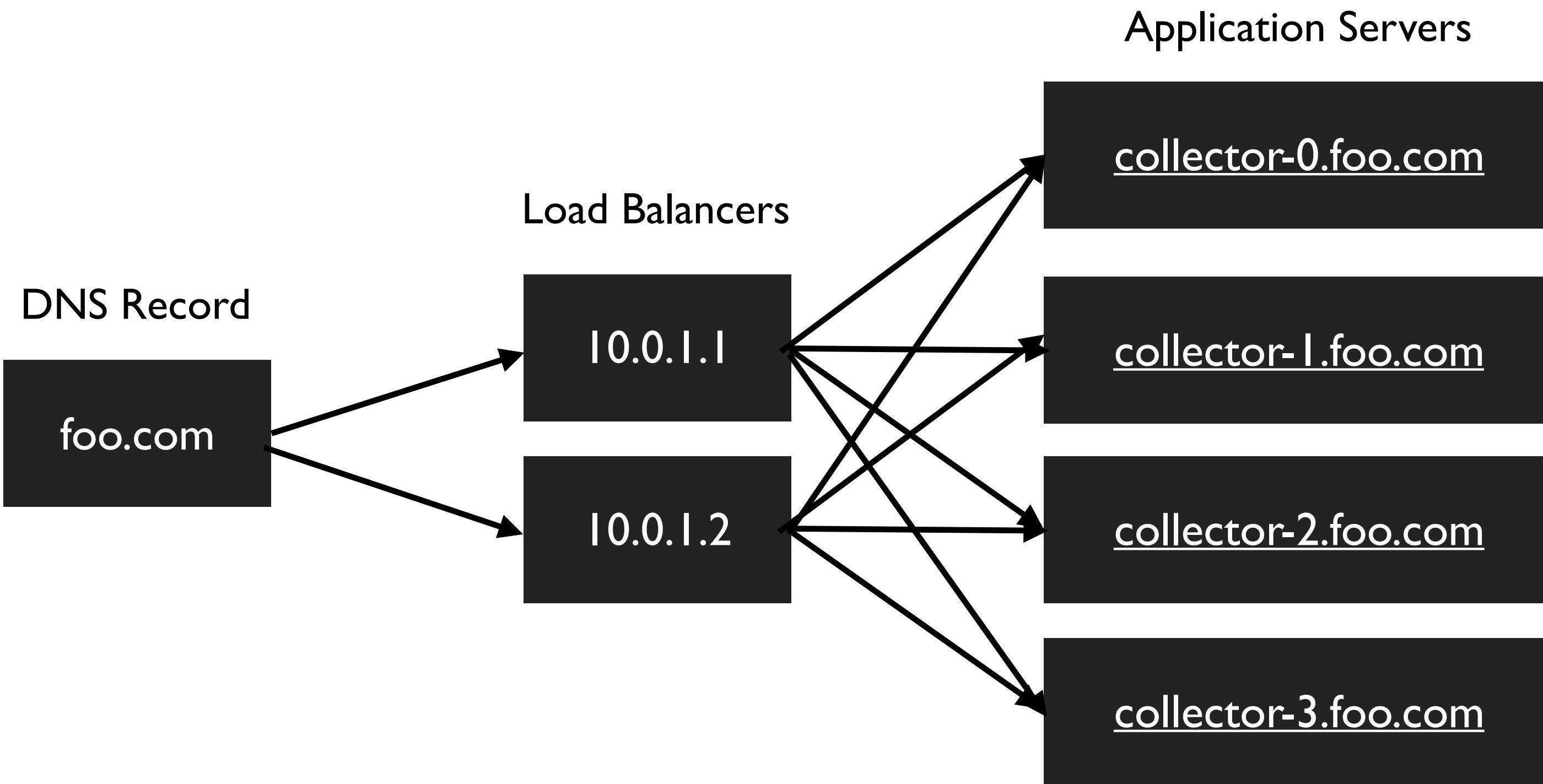
Collection

Transmission

Augmentation

Storage

Query



Transmission

- Delivery guarantees
- Highly available
- Queuing behaviour

Production

Collection

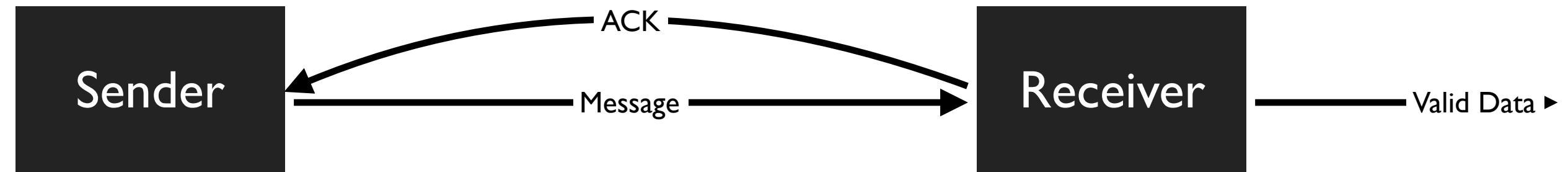
Transmission

Augmentation

Storage

Query

Ideal Case



Production

Collection

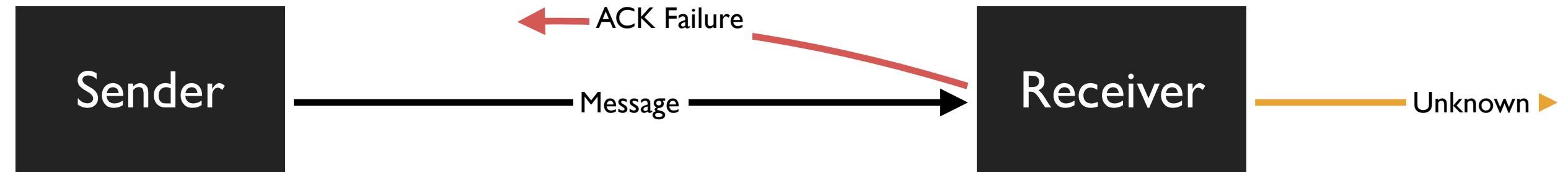
Transmission

Augmentation

Storage

Query

At most once



Production

Collection

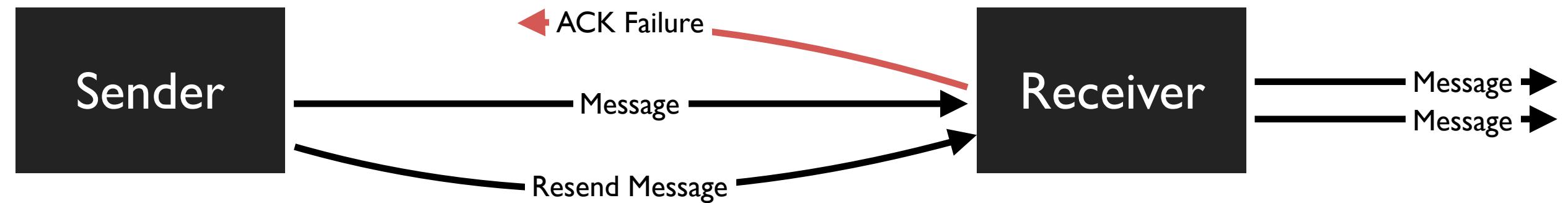
Transmission

Augmentation

Storage

Query

Duplicate Case: at least once



Production

Collection

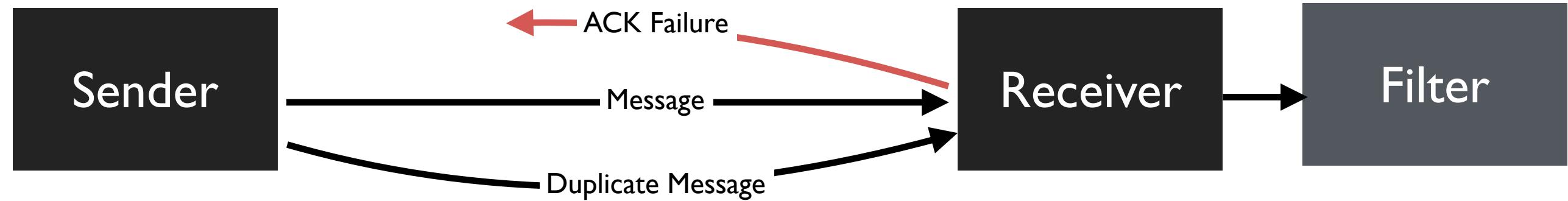
Transmission

Augmentation

Storage

Query

Exactly Once*: De-duplication

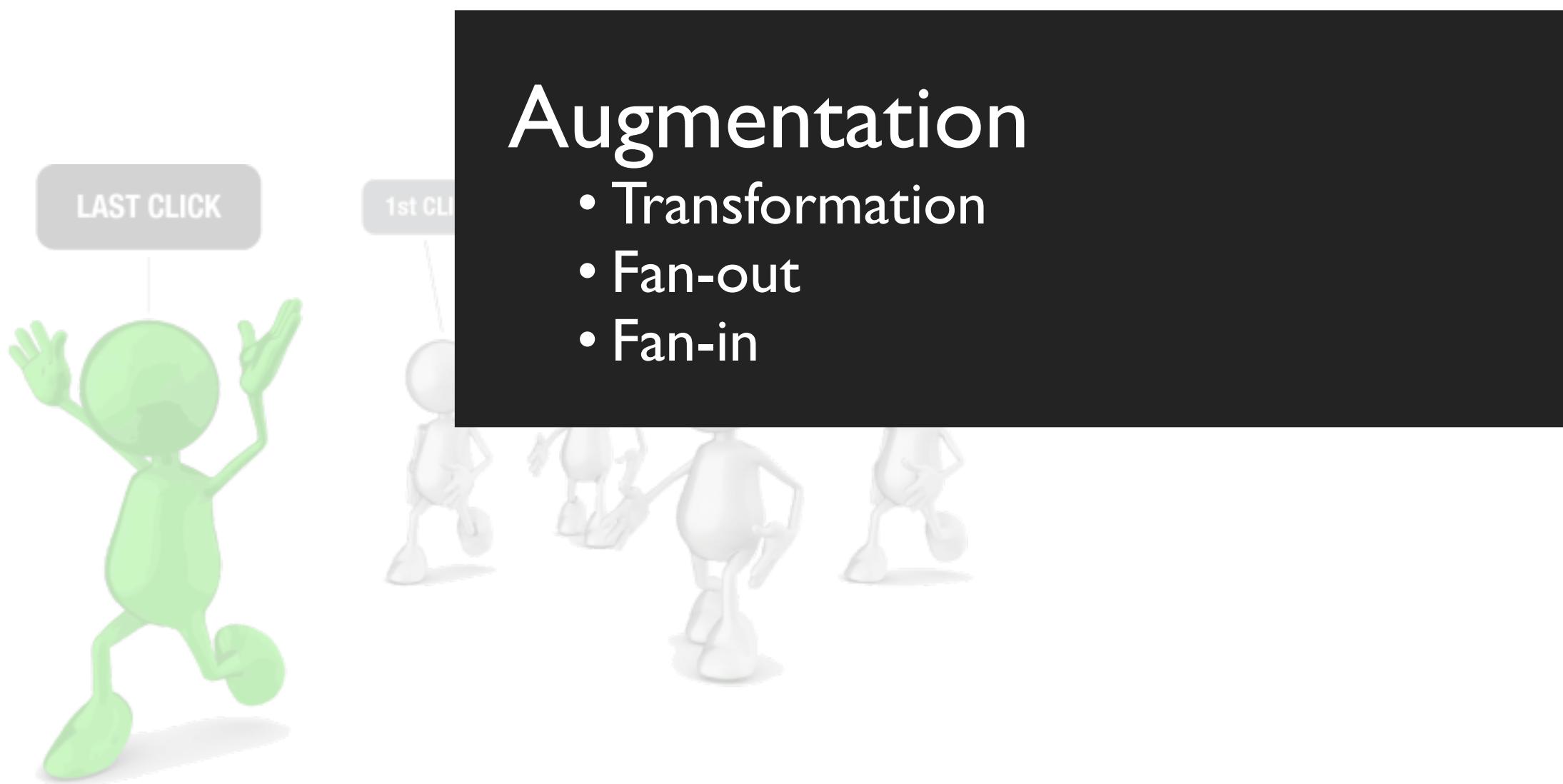


RabbitMQ

- Master/Slave
- Complex Topology
- Short lived queues
- At least once

Kafka

- Masterless
- Zookeeper
- Long lived logs
- At least once



Production

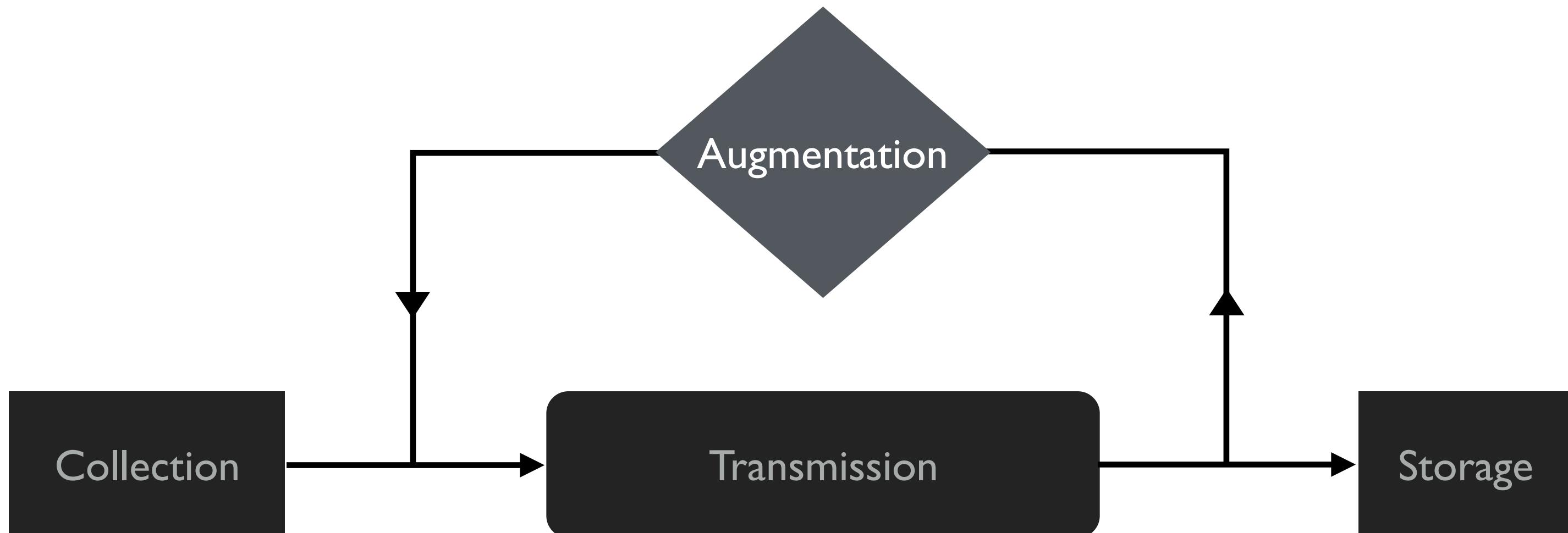
Collection

Transmission

Augmentation

Storage

Query



Production

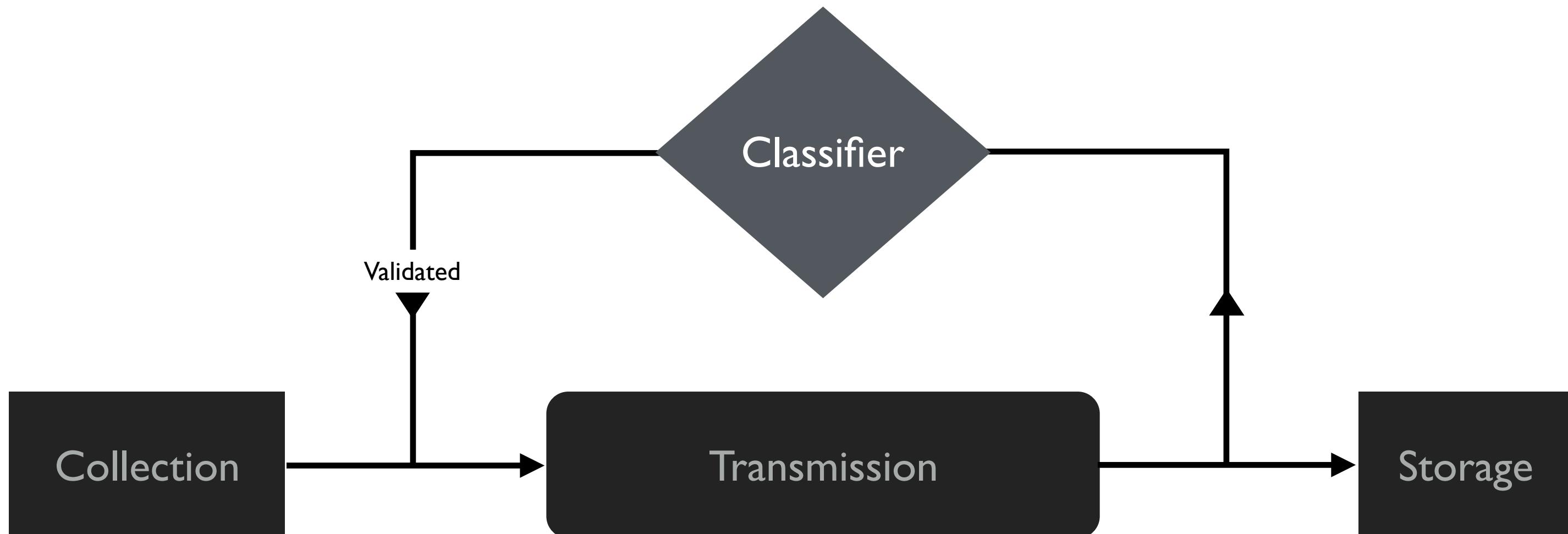
Collection

Transmission

Augmentation

Storage

Query



Production

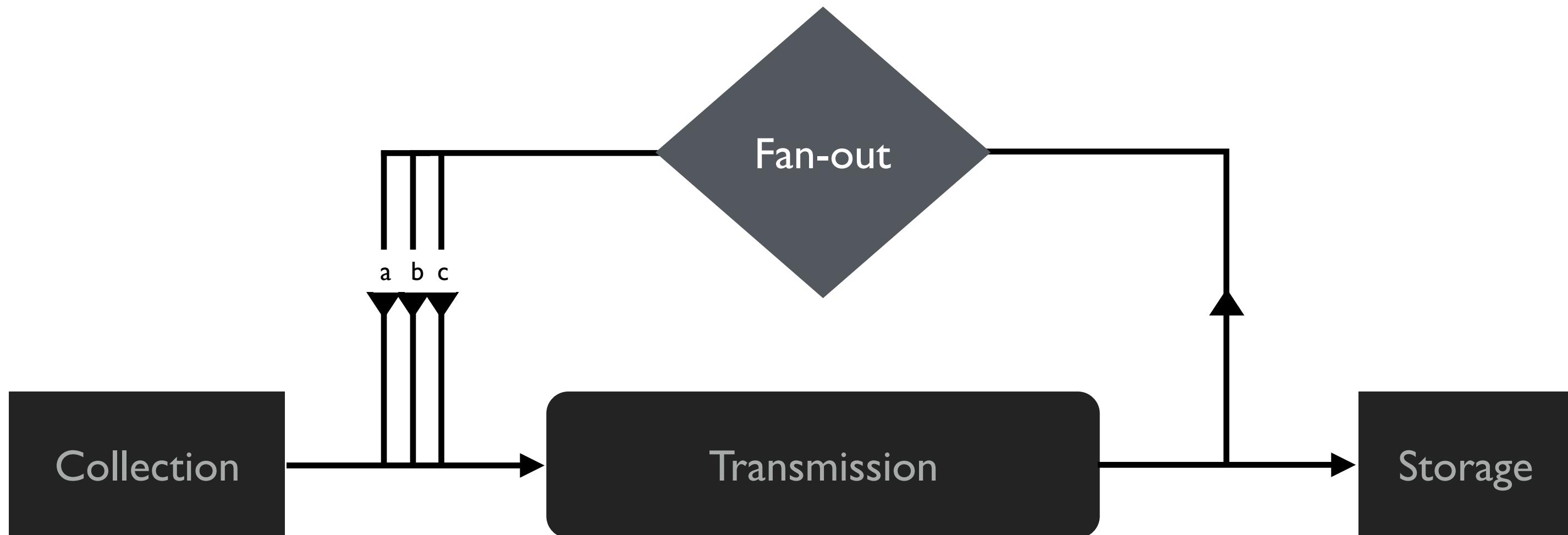
Collection

Transmission

Augmentation

Storage

Query



Production

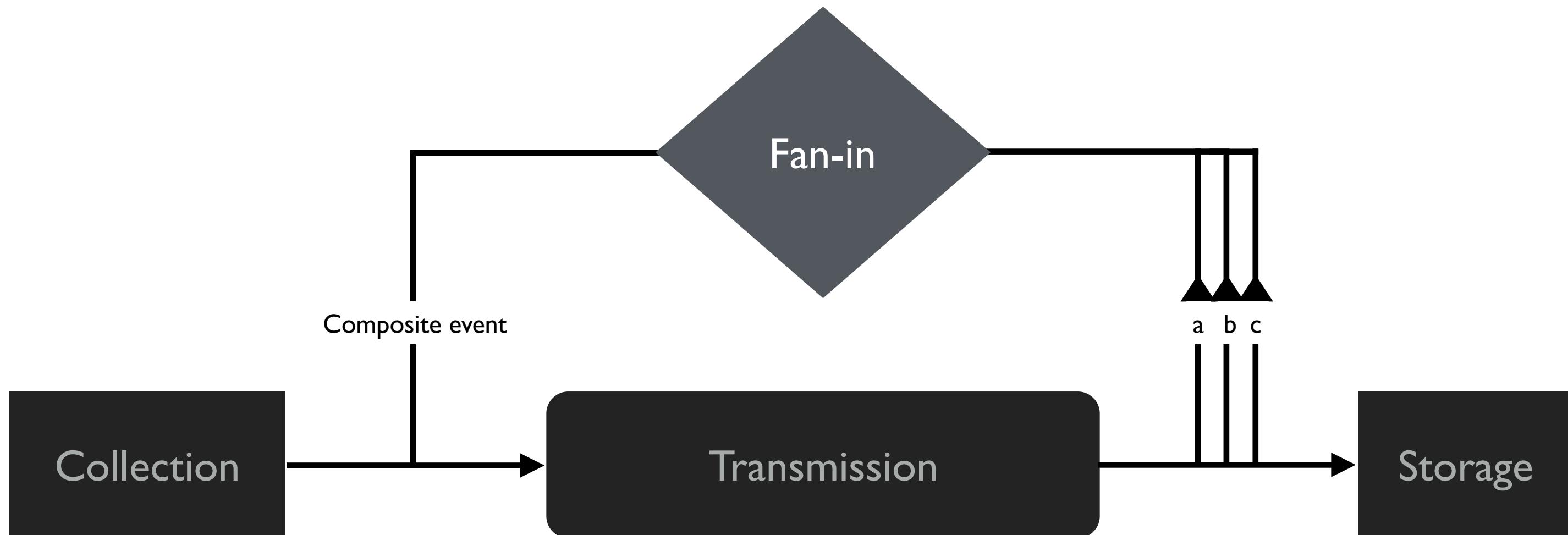
Collection

Transmission

Augmentation

Storage

Query



Production

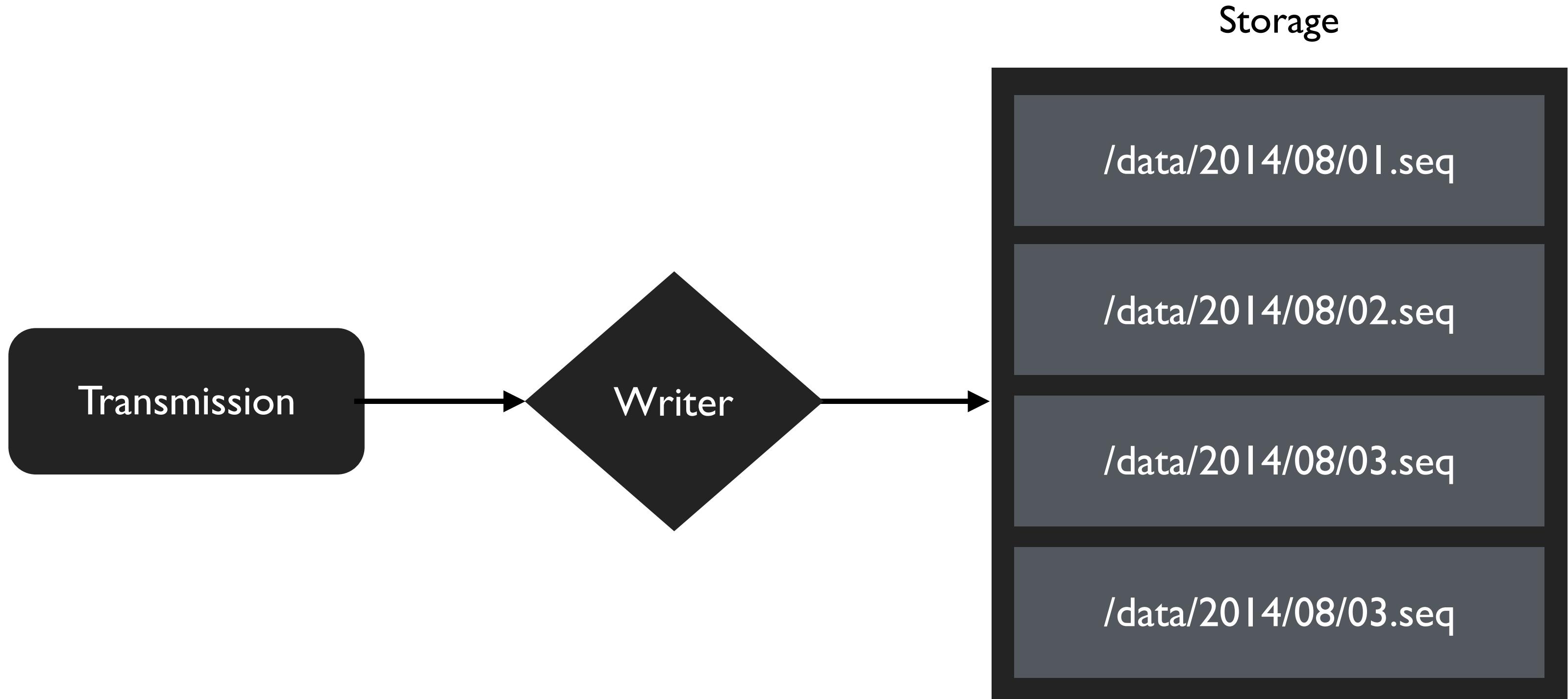
Collection

Transmission

Augmentation

Storage

Query



Production

Collection

Transmission

Augmentation

Storage

Query

Storage

- Source of truth
- Scalable
- Replicated
- Available

HDFS

- Self hosted
- Cost effective at scale
- Can run multi tenant
- Non trivial operational cost

S3

- Managed
- Cost prohibitive at scale
- Network to Map Reduce
- EMR cost

Query

- Database
- Low vs. high latency
- Common vs custom operations



Hadoop Based

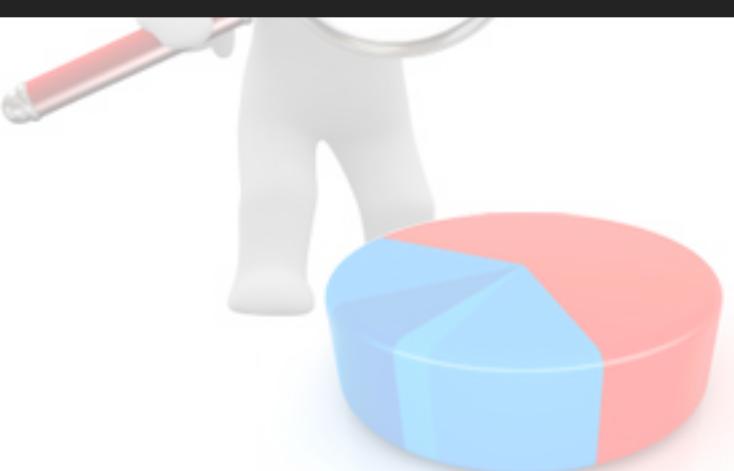
- Pig/Hive/Spark
- Shared tenancy
- Highly Scalable

Columnar Store

- Redshift/Vertica
- SQL
- Expensive

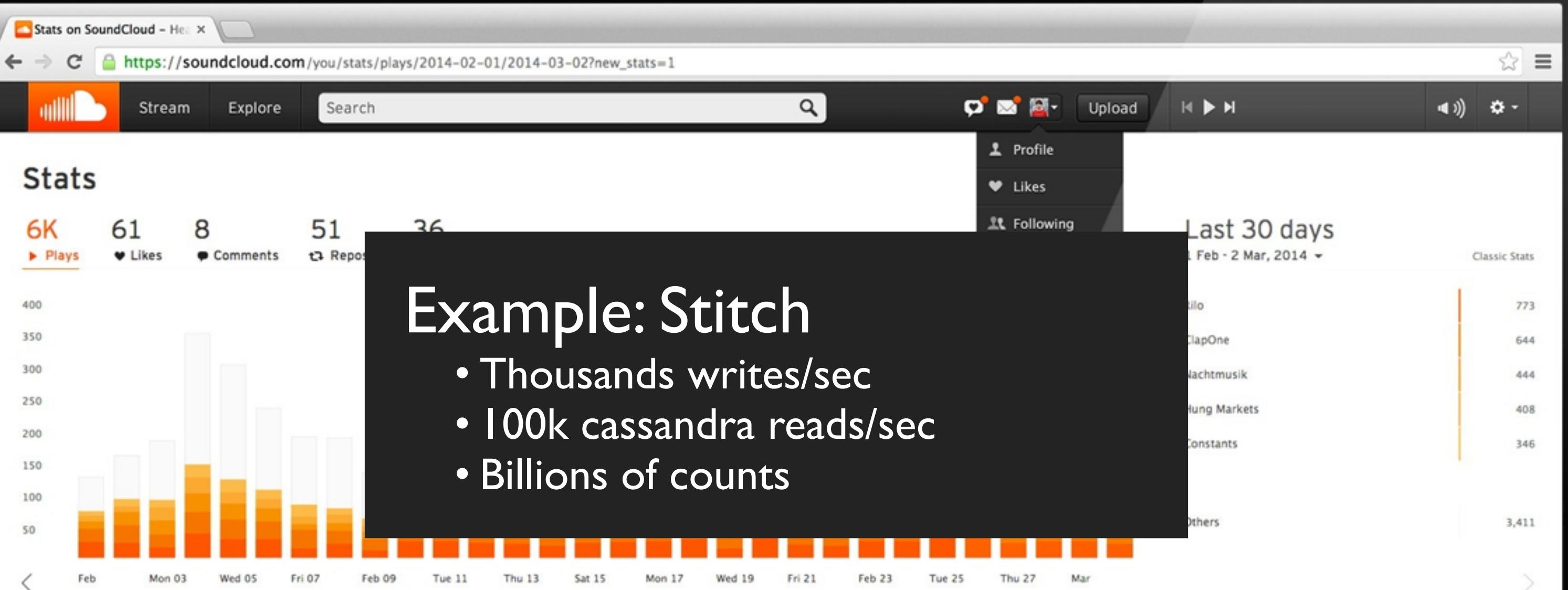
Key Value

- Cassandra/Riak
- Simple queries
- Complex



Architecture Summary

- Connected set of components
- Different failure modes
- Devils in the details



Most played tracks



Rilo
773

Top countries



United States
1,252

Who played the most



detailloop
122

Websites



soundcloud.com/seams
1,696

2 ClapOne

3 Nachtmusik

4 Hung Markets

644

444

408

1,238

737

274

94

71

39

soundcloud.com/seams/sets/quarters

206

soundcloud.com/stream

134

soundcloud.com/cecile9/likes

95

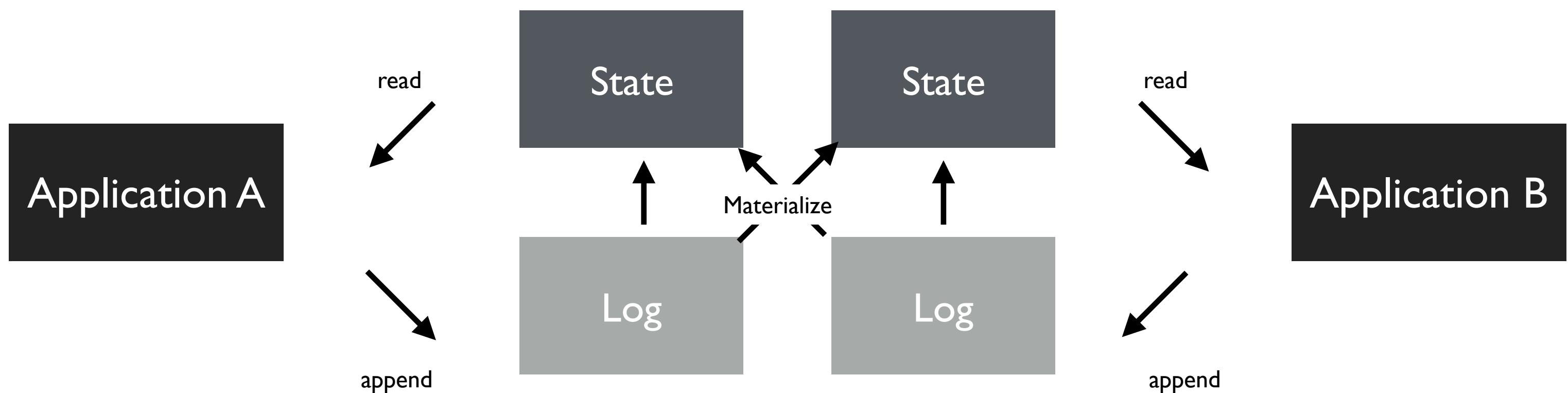
Stitch

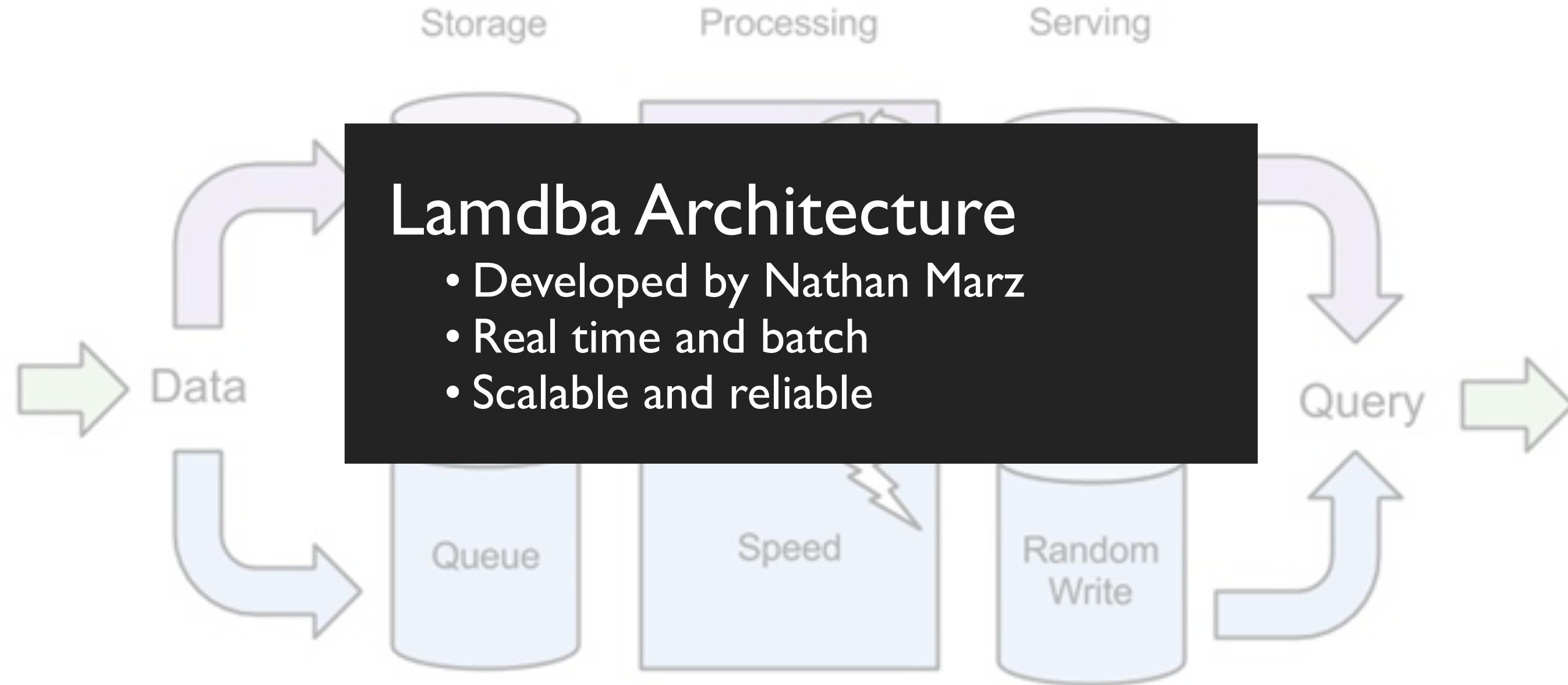


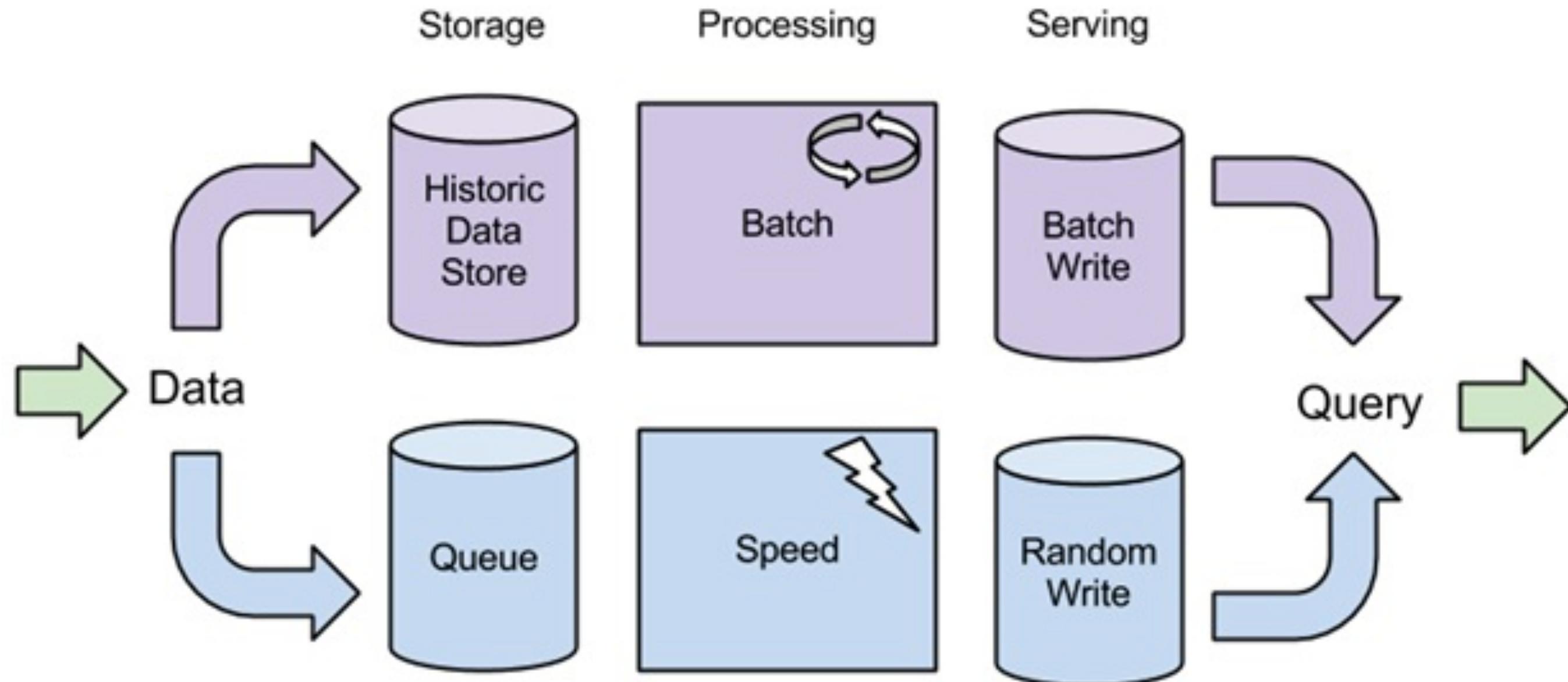
Event Sourcing

- Share data not state
- Decoupling producers/consumers
- Materialize State
- Scale with volume / complexity

Immutable State









Cloud Computing

- Commodity Computation
- Outsourced infrastructure
- Scales down / up

Conclusion

- Motivated tracking
- Stream of immutable data
- Decouple producers / consumers
- Mature ideas
- Commoditized in the cloud

Coordinates: (x,y,z)

Velocity Components: (u,v,w)

Time : t Pressure: p

Density: ρ Stress: τ

Total Energy: Et

Heat Flux: q

Reynolds Number: Re

Prandtl Number: Pr

continuity:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0$$

- Momentum:

$$\frac{\partial(\rho u)}{\partial t}$$

Open Problems

- Ad Blockers
- Device Identification

- Momentum:

$$\frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho u v)}{\partial x} + \frac{\partial(\rho v^2)}{\partial y} + \frac{\partial(\rho v w)}{\partial z} = - \frac{\partial p}{\partial y} + \frac{1}{Re_r} \left[\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yy}}{\partial y} + \frac{\partial \tau_{yz}}{\partial z} \right]$$

$$\frac{1}{Re_r} \left[\frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{xz}}{\partial z} \right]$$

- Momentum

$$\frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho u w)}{\partial x} + \frac{\partial(\rho v w)}{\partial y} + \frac{\partial(\rho w^2)}{\partial z} = - \frac{\partial p}{\partial z} + \frac{1}{Re_r} \left[\frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \tau_{zz}}{\partial z} \right]$$

$$\frac{1}{Re_r} \left[\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yy}}{\partial y} + \frac{\partial \tau_{yx}}{\partial z} \right]$$

Energy:

$$\frac{\partial(E_T)}{\partial t} + \frac{\partial(uE_T)}{\partial x} + \frac{\partial(vE_T)}{\partial y} + \frac{\partial(wE_T)}{\partial z} = - \frac{\partial(up)}{\partial x} - \frac{\partial(vp)}{\partial y} - \frac{\partial(wp)}{\partial z} - \frac{1}{Re_r} \left[\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} \right]$$



Thank you