

Scale Your Metrics with Elasticsearch

Philipp Krenn

@xeraa
























elasticsearch.

```
$ curl http://localhost:9200
{
  "name": "elasticsearch",
  "cluster_name": "docker-cluster",
  "cluster_uuid": "16wfwv8XSniiI_Fx6qqrcw",
  "version": {
    "number": "5.6.9",
    "build_hash": "877a590",
    "build_date": "2018-04-12T16:25:14.838Z",
    "build_snapshot": false,
    "lucene_version": "6.6.1"
  },
  "tagline": "You Know, for Search"
}
```






















<https://db-engines.com/en/>



346 systems in ranking, October 2018

ranking

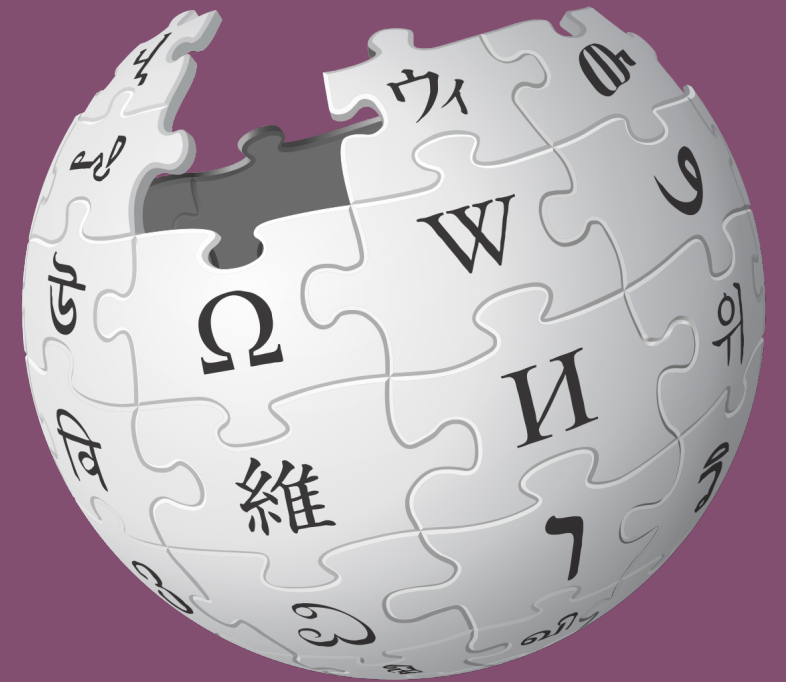
Rank			DBMS	Database Model	Score		
Oct 2018	Sep 2018	Oct 2017			Oct 2018	Sep 2018	Oct 2017
1.	1.	1.	Oracle 	Relational DBMS	1319.27	+10.15	-29.54
2.	2.	2.	MySQL 	Relational DBMS	1178.12	-2.36	-120.71
3.	3.	3.	Microsoft SQL Server 	Relational DBMS	1058.33	+7.05	-151.99
4.	4.	4.	PostgreSQL 	Relational DBMS	419.39	+12.97	+46.12
5.	5.	5.	MongoDB 	Document store	363.19	+4.39	+33.79
6.	6.	6.	DB2 	Relational DBMS	179.69	-1.38	-14.90
7.	 8.	 9.	Redis 	Key-value store	145.29	+4.35	+23.24
8.	 7.	 10.	Elasticsearch 	Search engine	142.33	-0.28	+22.09
9.	9.	 7.	Microsoft Access	Relational DBMS	136.80	+3.41	+7.35
10.	10.	 8.	Cassandra 	Wide column store	123.39	+3.83	-1.40
11.	11.	11.	SQLite 	Relational DBMS	116.74	+1.29	+4.76
12.	12.	12.	Teradata 	Relational DBMS	78.63	+1.25	-1.45
13.	13.	 16.	Splunk	Search engine	76.90	+2.87	+12.54
14.	14.	 18.	MariaDB 	Relational DBMS	73.13	+2.49	+16.73
15.	15.	 13.	Solr	Search engine	61.31	+1.11	-9.82

346 systems in ranking, October 2018

Oct 2018	Rank		DBMS	Database Model	Score		
	Sep 2018	Oct 2017			Oct 2018	Sep 2018	Oct 2017
1.	1.	1.	Oracle 	Relational DBMS	1319.27	+10.15	-29.54
2.	2.	2.	MySQL 	Relational DBMS	1178.12	-2.36	-120.71
3.	3.	3.	Microsoft SQL Server 	Relational DBMS	1058.33	+7.05	-151.99
4.	4.	4.	PostgreSQL 	Relational DBMS	419.39	+12.97	+46.12
5.	5.	5.	MongoDB 	Document store	363.19	+4.39	+33.79
6.	6.	6.	DB2 	Relational DBMS	179.69	-1.38	-14.90
7.	 8.	 9.	Redis 	Key-value store	145.29	+4.35	+23.24
8.	 7.	 10.	Elasticsearch 	Search engine	142.33	-0.28	+22.09
9.	9.	 7.	Microsoft Access	Relational DBMS	136.80	+3.41	+7.35
10.	10.	 8.	Cassandra 	Wide column store	123.39	+3.83	-1.40
11.	11.	11.	SQLite 	Relational DBMS	116.74	+1.29	+4.76
12.	12.	12.	Teradata 	Relational DBMS	78.63	+1.25	-1.45
13.	13.	 16.	Splunk	Search engine	76.90	+2.87	+12.54
14.	14.	 18.	MariaDB 	Relational DBMS	73.13	+2.49	+16.73
15.	15.	 13.	Solr	Search engine	61.31	+1.11	-9.82

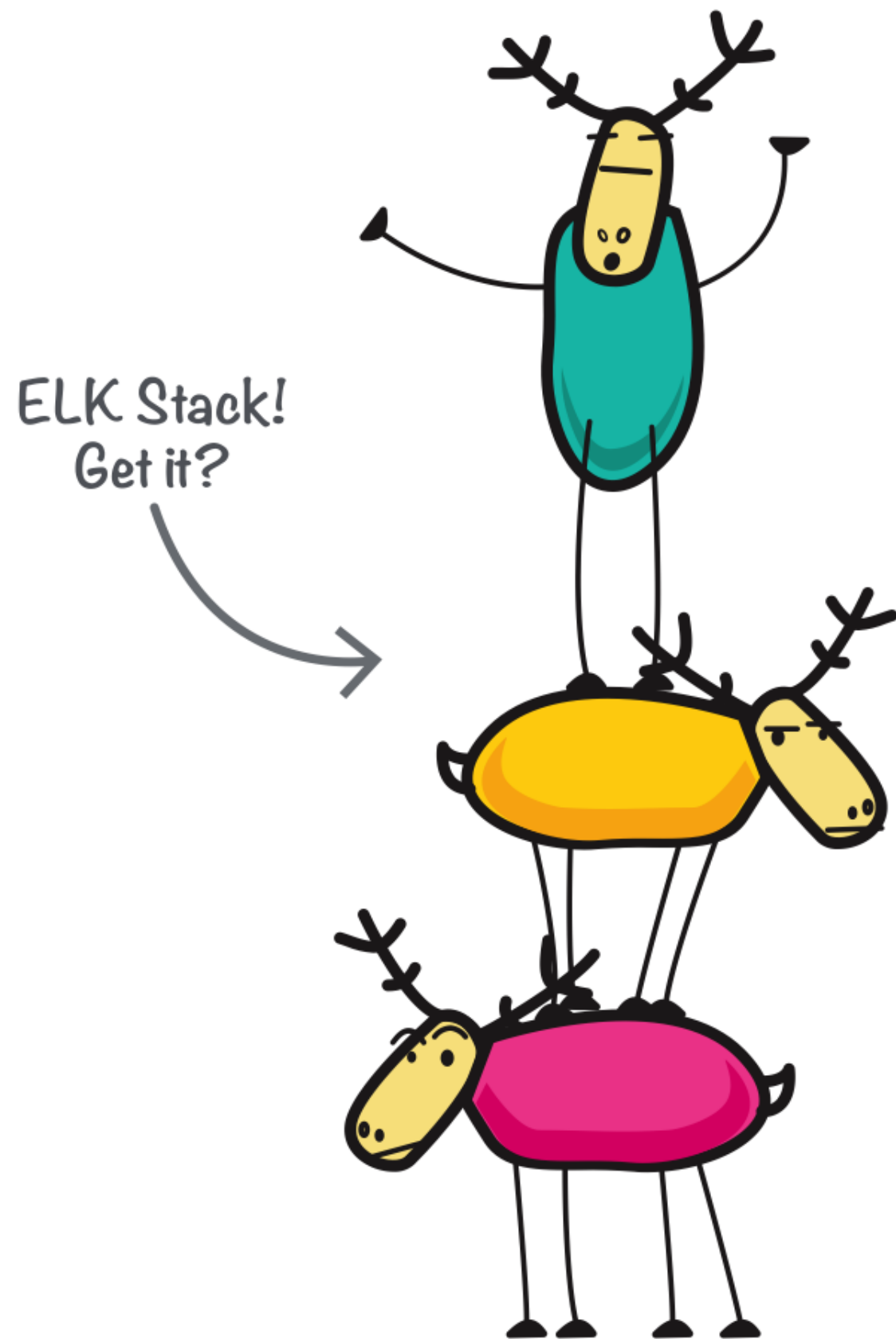
9. Redis 
10. Elasticsearch 
7. Microsoft Access
8. Cassandra 
1. SQLite 

Full-Text Search





logstash



E Elasticsearch

L Logstash

K Kibana

Logs

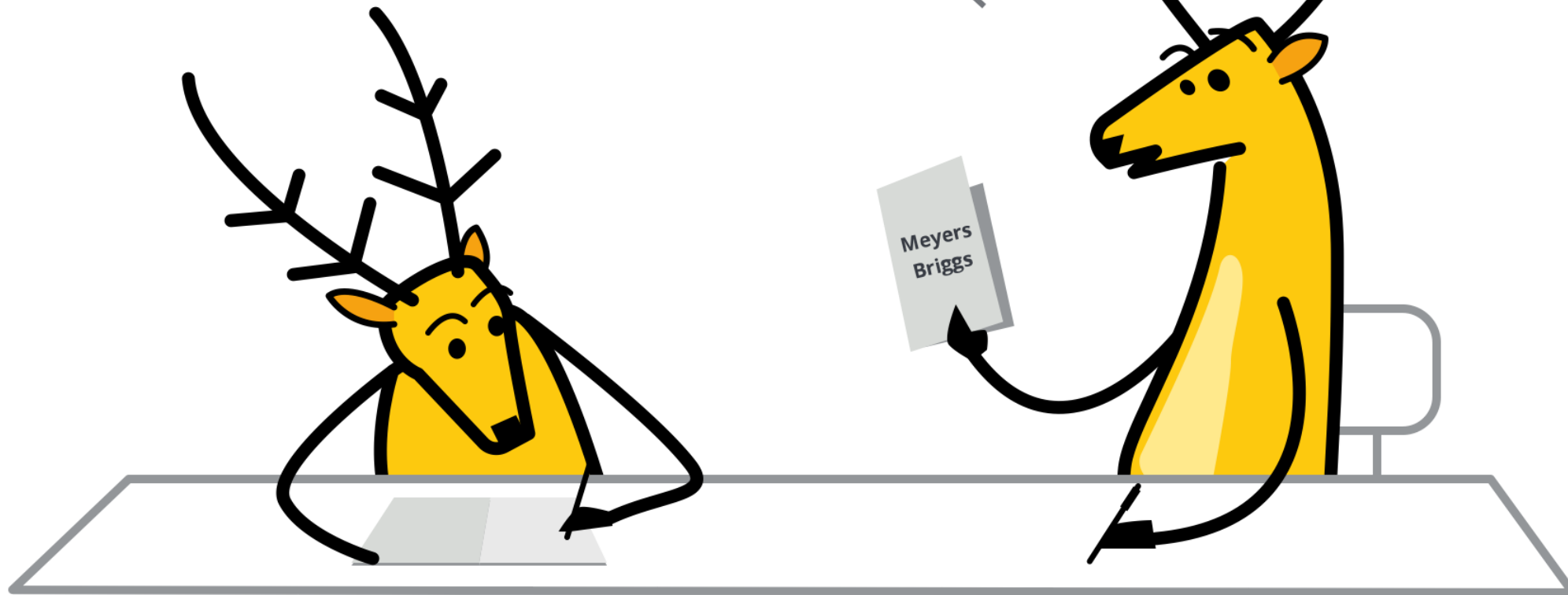


slack

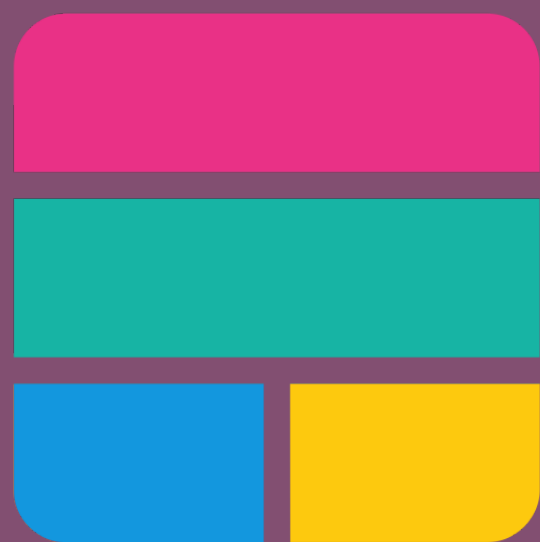


fitbit

*Apparently, I'm an
ELKB personality.*







elastic stack

Metrics



*I'm not going to use a search
engine for metrics.*

— Too often



Developer 🥑

Agenda

Building Blocks

Tuning

Delivering

Building Blocks



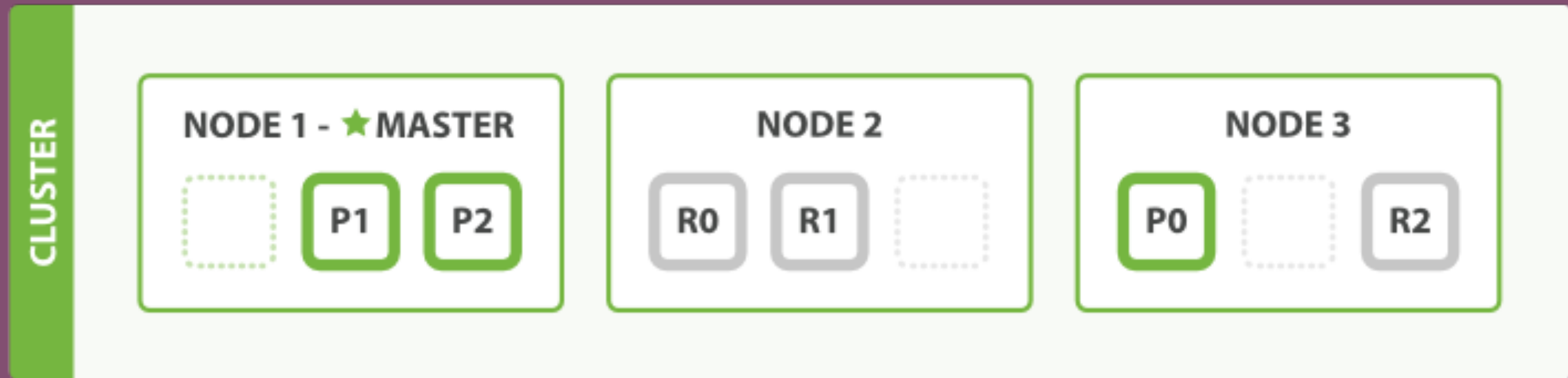
Only accept features that scale.

— https://github.com/elastic/engineering/blob/master/development_constitution.md

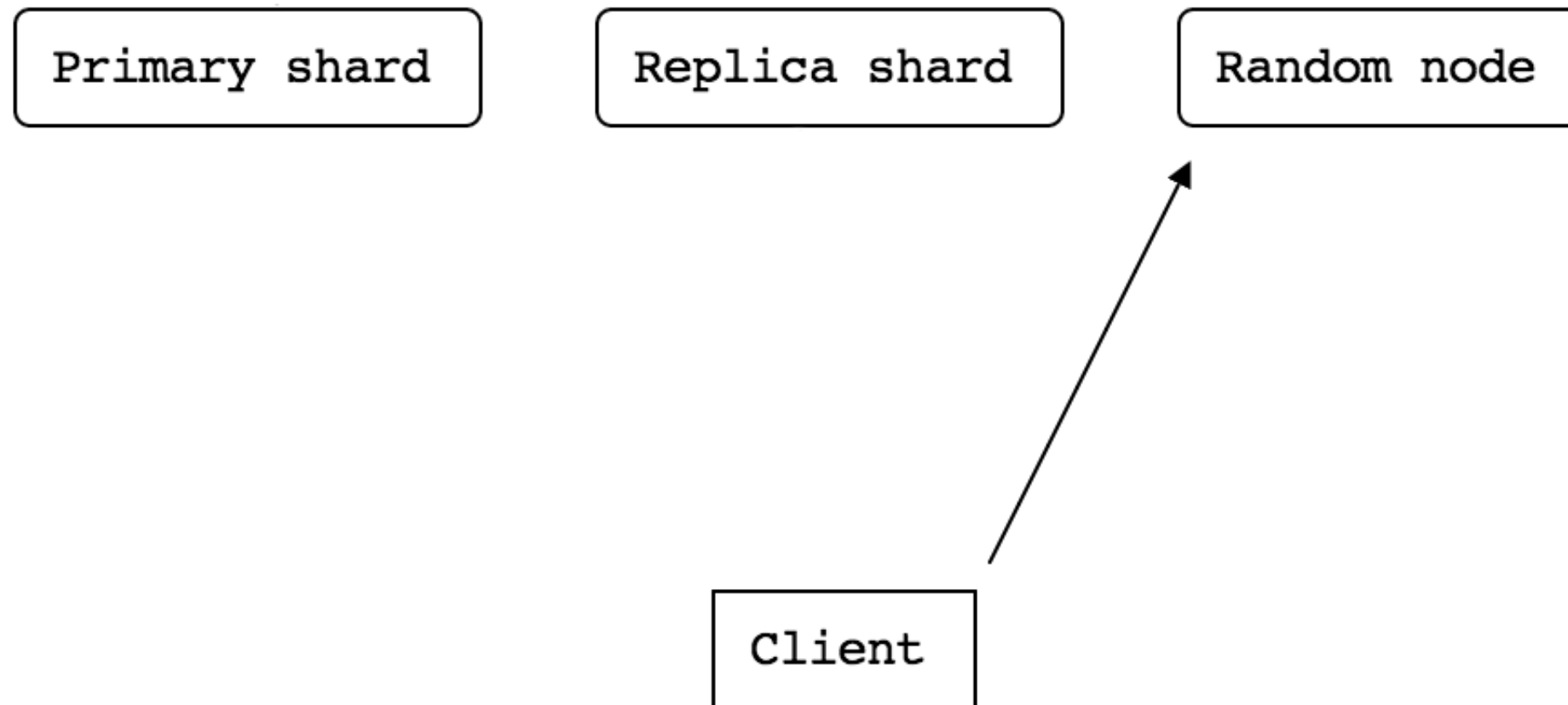
Horizontal Scaling

Shards
Replication
Writes & Reads

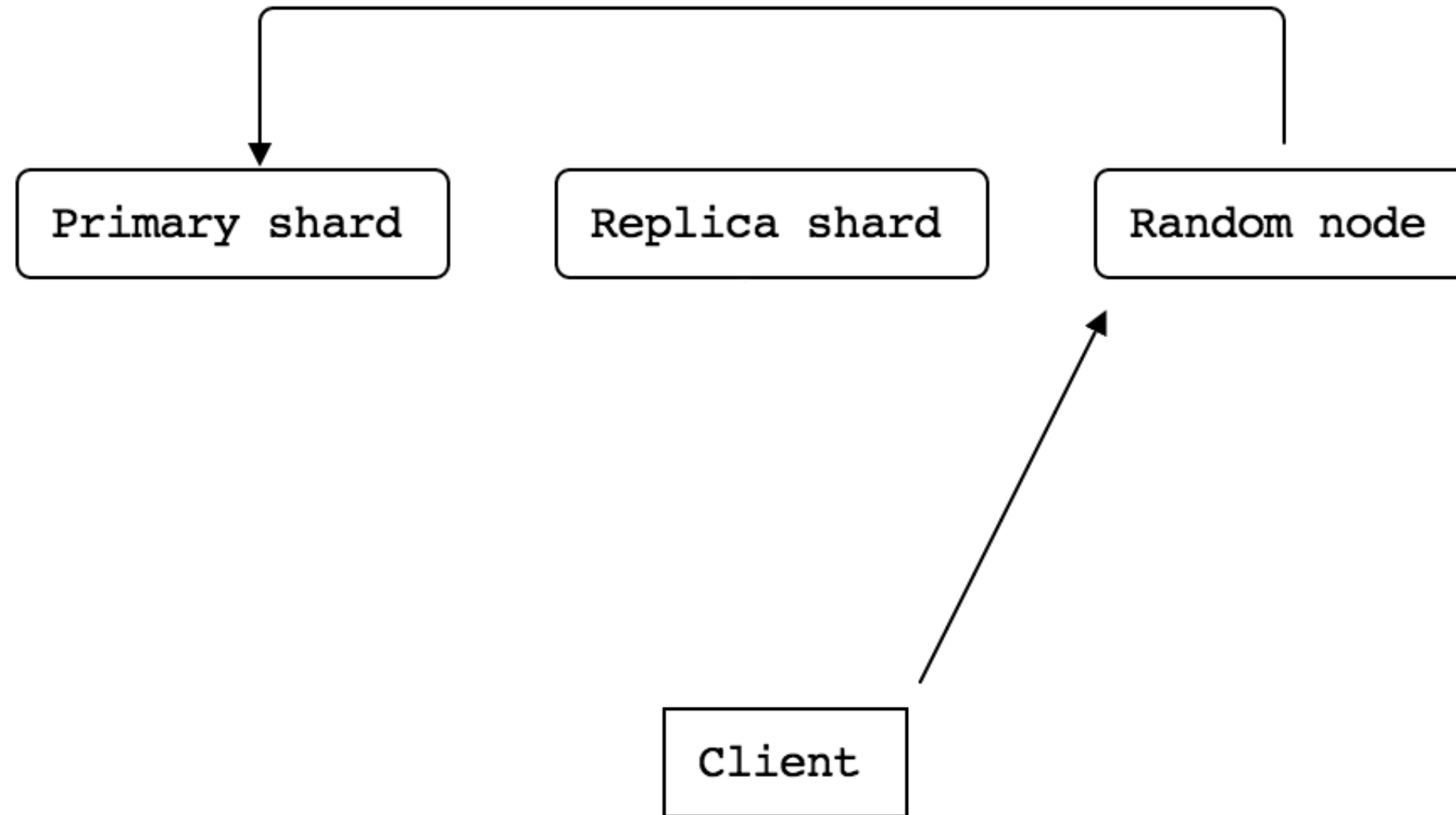
Cluster, Node, Index, Shard



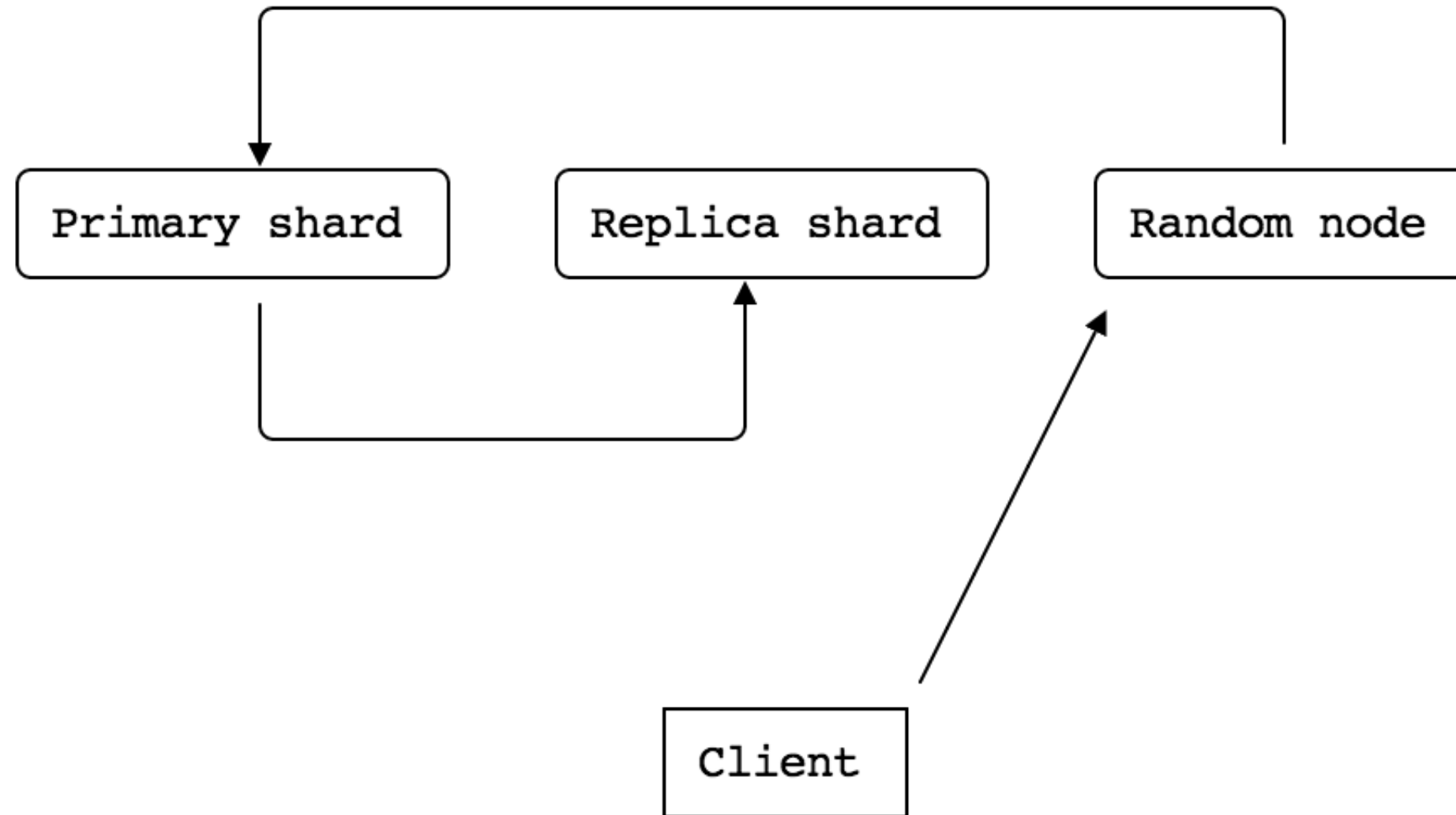
Write



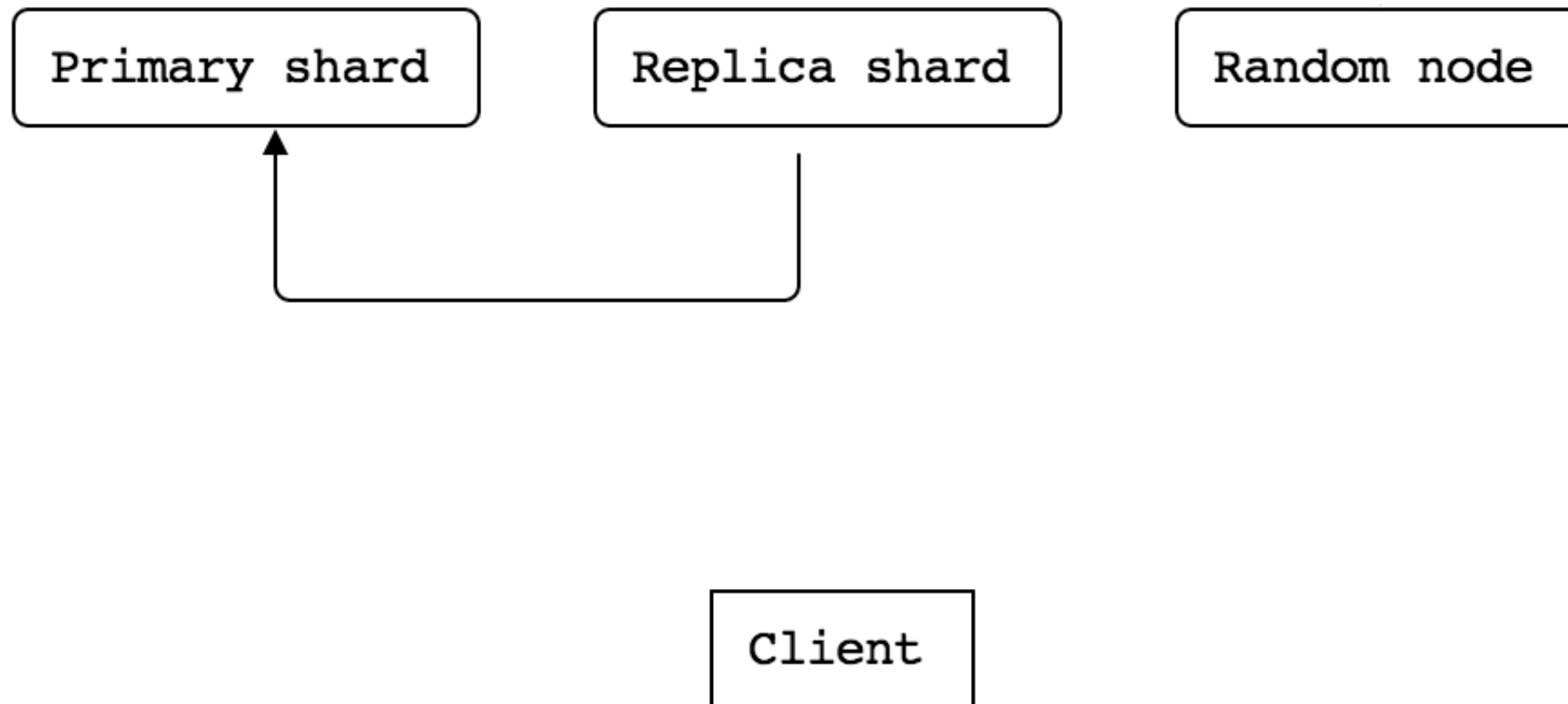
Write



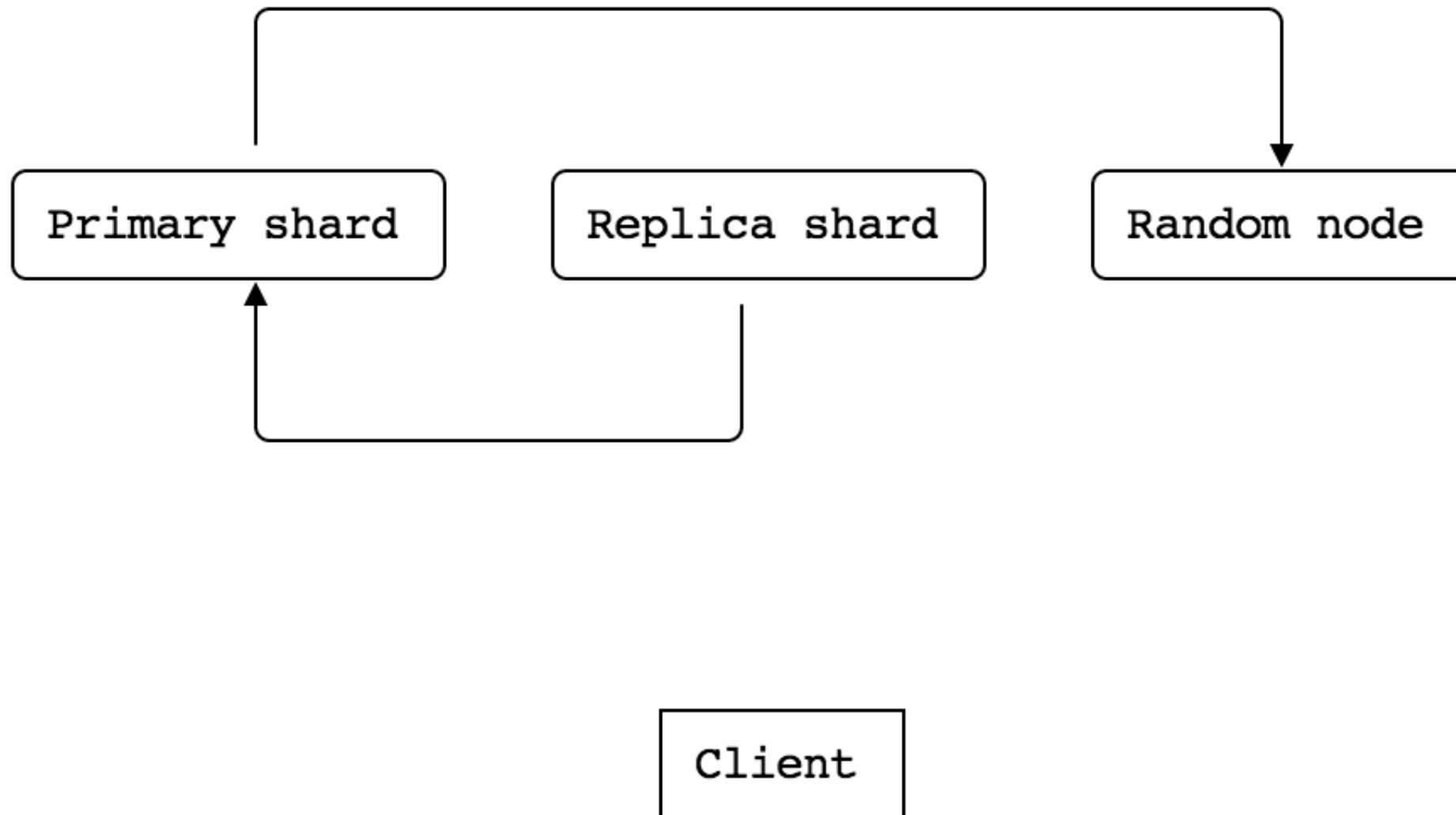
Write



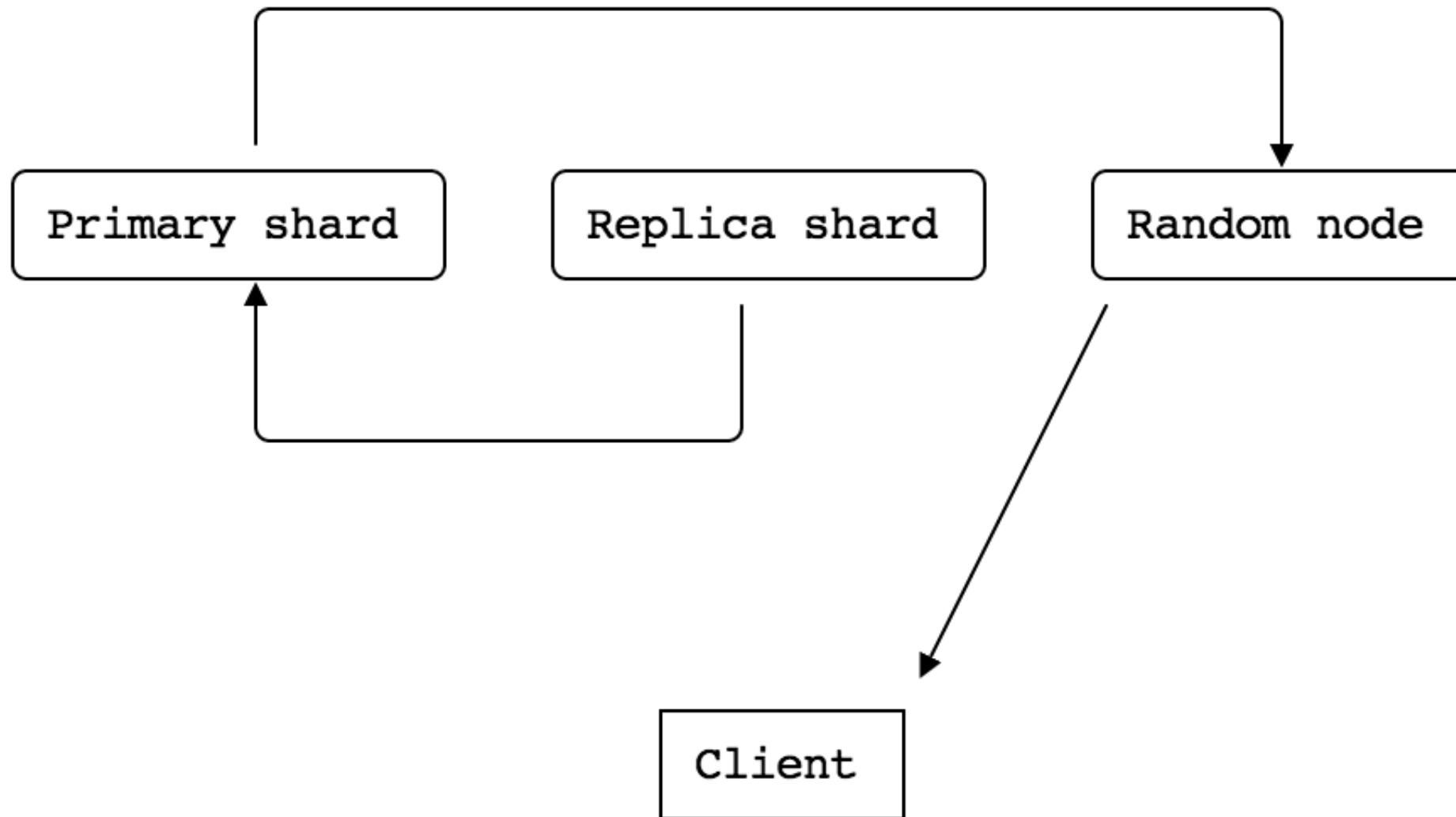
Write



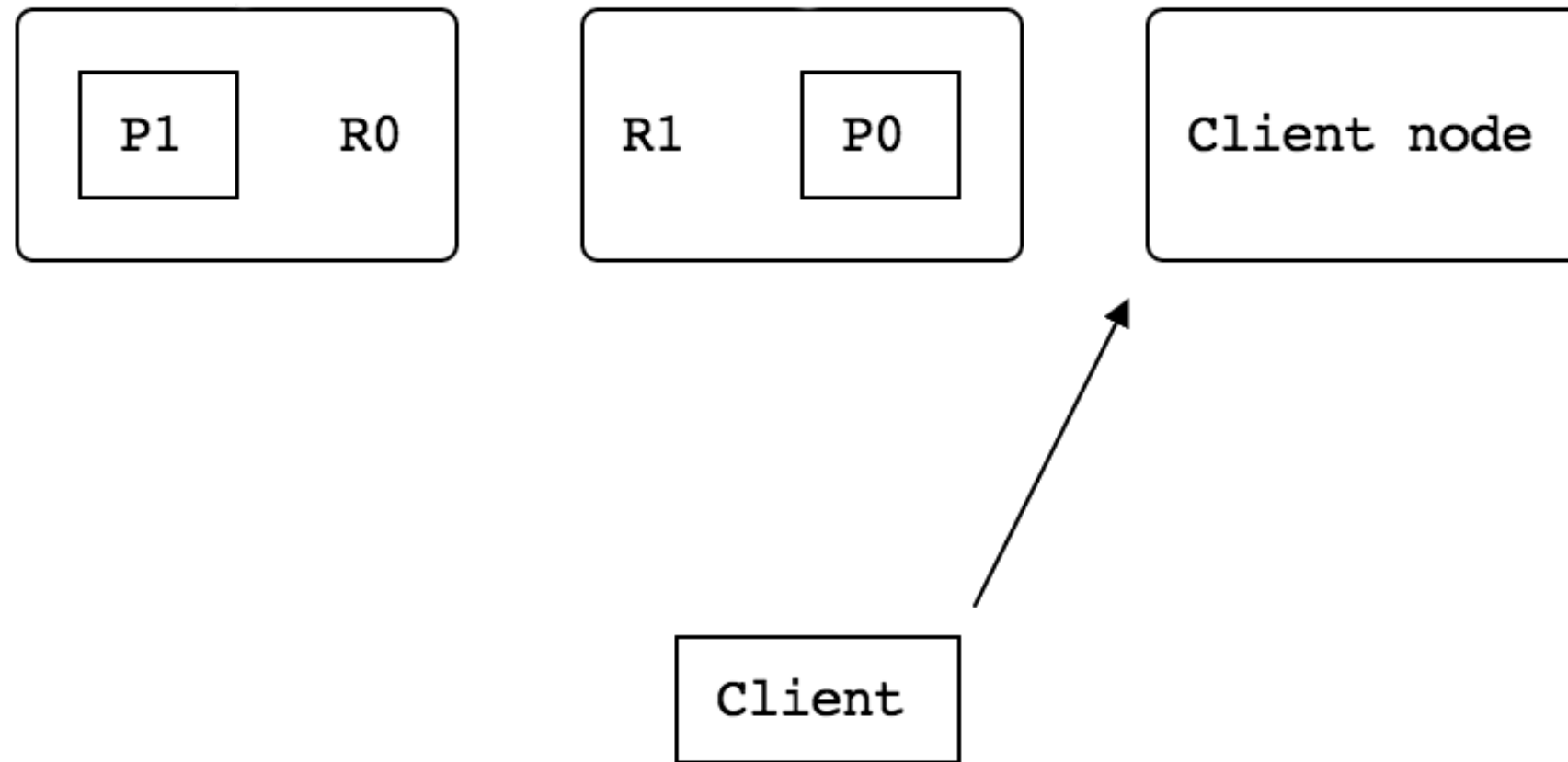
Write



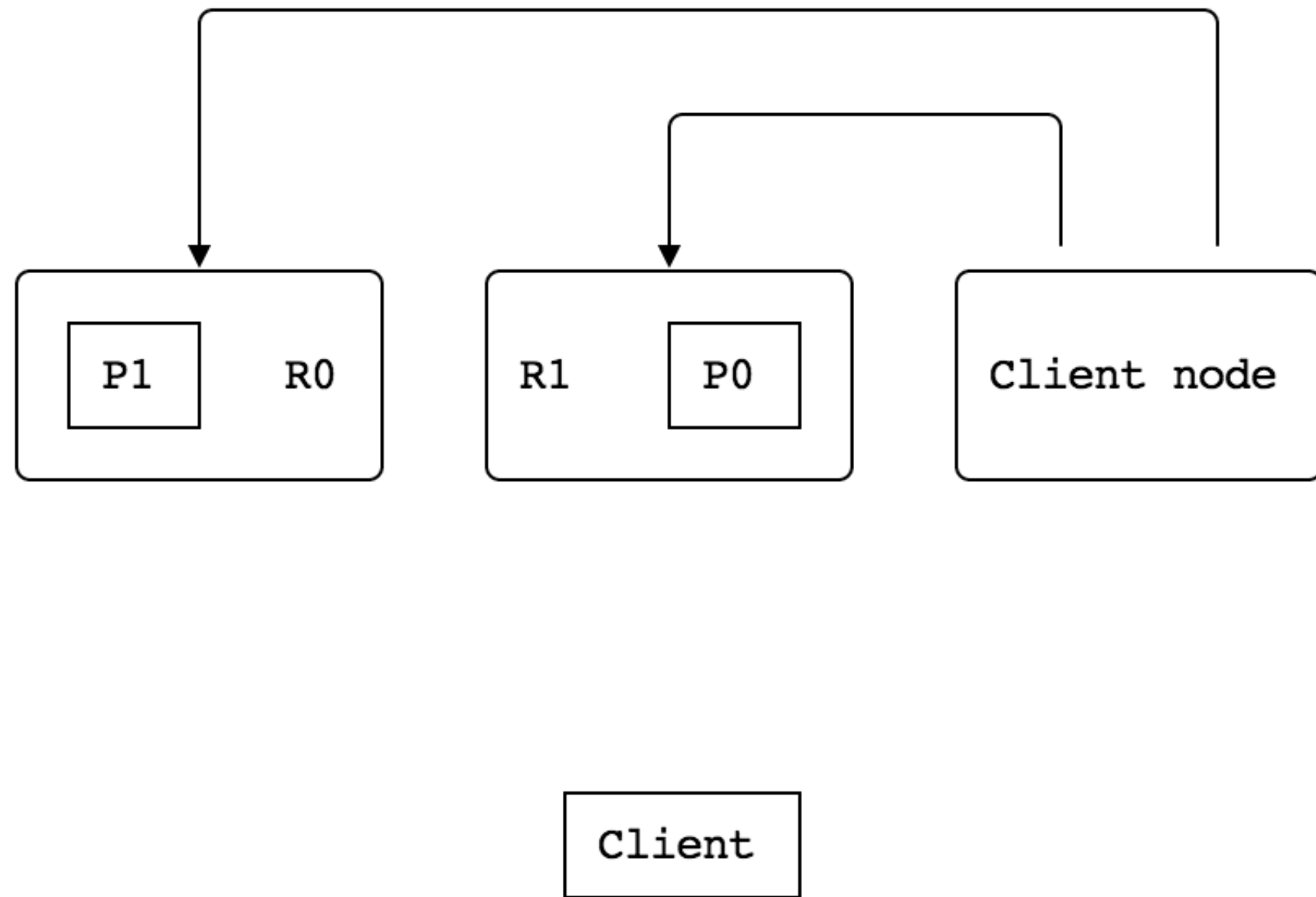
Write



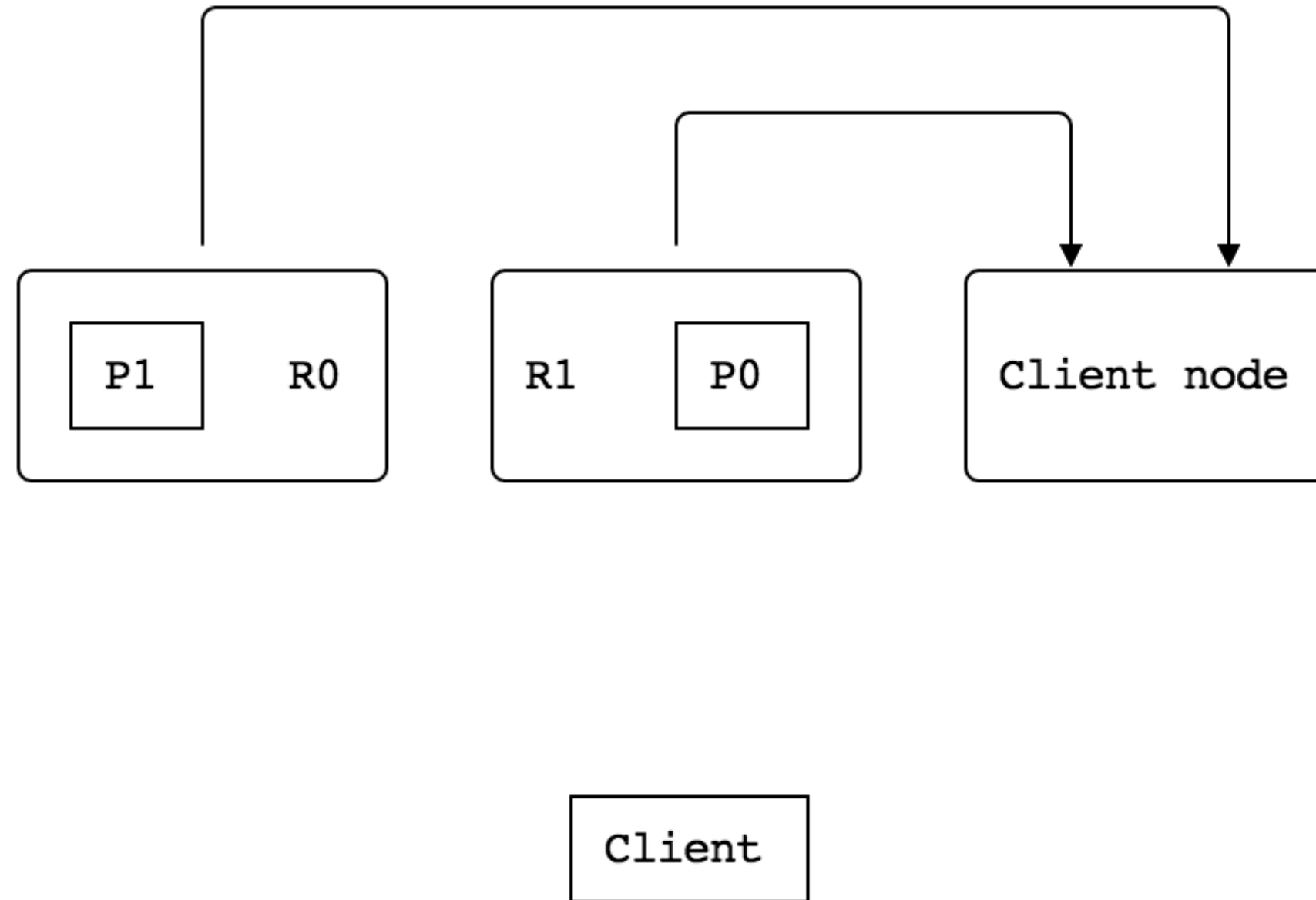
Read



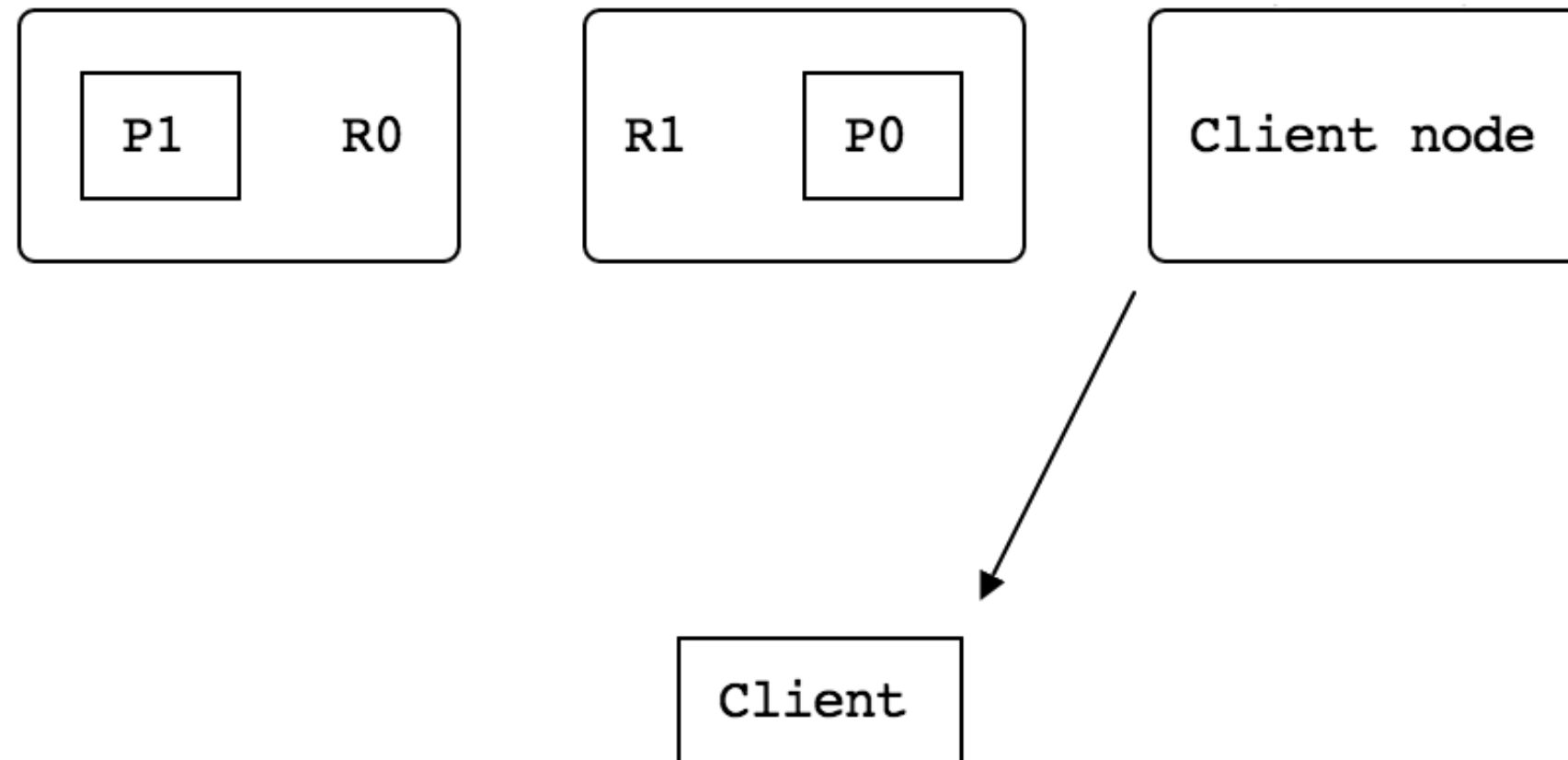
Read



Read



Read



Append-Only Optimization

IDs assigned on coordinating node

Fast add docs instead of the slow
update docs

Storage Compression

LZ4 (default), DEFLATE (best_compression)

BKD Trees

Points in Lucene

Half & Scaled Floats

Pipeline Aggregations

Tuning

_a11 Removal

<https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping-all-field.html>

Doc Values Replaced Fielddata

https://www.elastic.co/guide/en/elasticsearch/guide/current/_deep_dive_on_doc_values.html

Delivering

Architecture

Hot — Warm — Cold

(Frozen)

```
bin/elasticsearch -Enode.attr.rack=rack1 -Enode.attr.size=hot
```

```
PUT today/_settings
```

```
{  
  "index.routing.allocation.include.size": "hot"  
}
```

Time-Based Indices

Rollover Indices

Rollups



think of the bytes

```
PUT _xpack/rollup/job/metrics
{
  "index_pattern": "metrics-*",
  "rollup_index": "metrics_rollup",
  "cron": "*/30 * * * * ?",
  "page_size": 1000,
  "groups": {
    "date_histogram": {
      "field": "timestamp",
      "interval": "1h",
      "delay": "7d"
    },
    "terms": {
      "fields": ["node"]
    }
  },
  "metrics": [
    {
      "field": "cpu",
      "metrics": ["min", "max", "sum"]
    },
    {
      "field": "memory",
      "metrics": ["avg"]
    }
  ]
}
```



Management

Elasticsearch

- License management
- Users
- Roles
- Watches
- Index management
- [Index rollups](#)

Kibana

- Index patterns
- Reporting
- Advanced settings
- Saved objects

Logstash

- Pipelines

Create a new rollup job



Optional: Collect metrics on important fields

You can collect metrics on as many fields as you want.

Field	Min	Max	Avg	Sum	Value count	Cardinality
system.network.out.bytes	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
system.network.out.errors			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
system.network.usage			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>

Add a metric

Continue

Metric to aggregate

Field

system.network.high_fives

Metrics to capture



Min



Sum



Max



Value count



Avg



Cardinality

Add

Index Lifecycle Management

Currently

<https://github.com/elastic/curator>

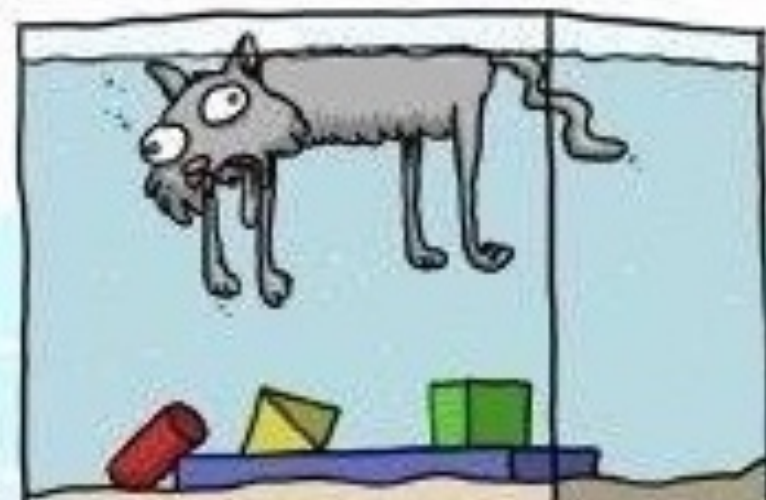
Conclusion

Benchmarks

Fair

Reproducible

Close to Production



Professor Zapinsky proved that the squid is more intelligent than the housecat when posed with puzzles under similar conditions





Thank You

Questions?

Philipp Krenn

@xeraa