Scalable Full-Text Search with DataStax Enterprise

Piotr Kołaczkowski
DataStax

pkolaczk@datastax.com
@pkolaczk

DataStax Enterprise

Apache Cassandra

+

Apache Solr

+

Apache Hadoop

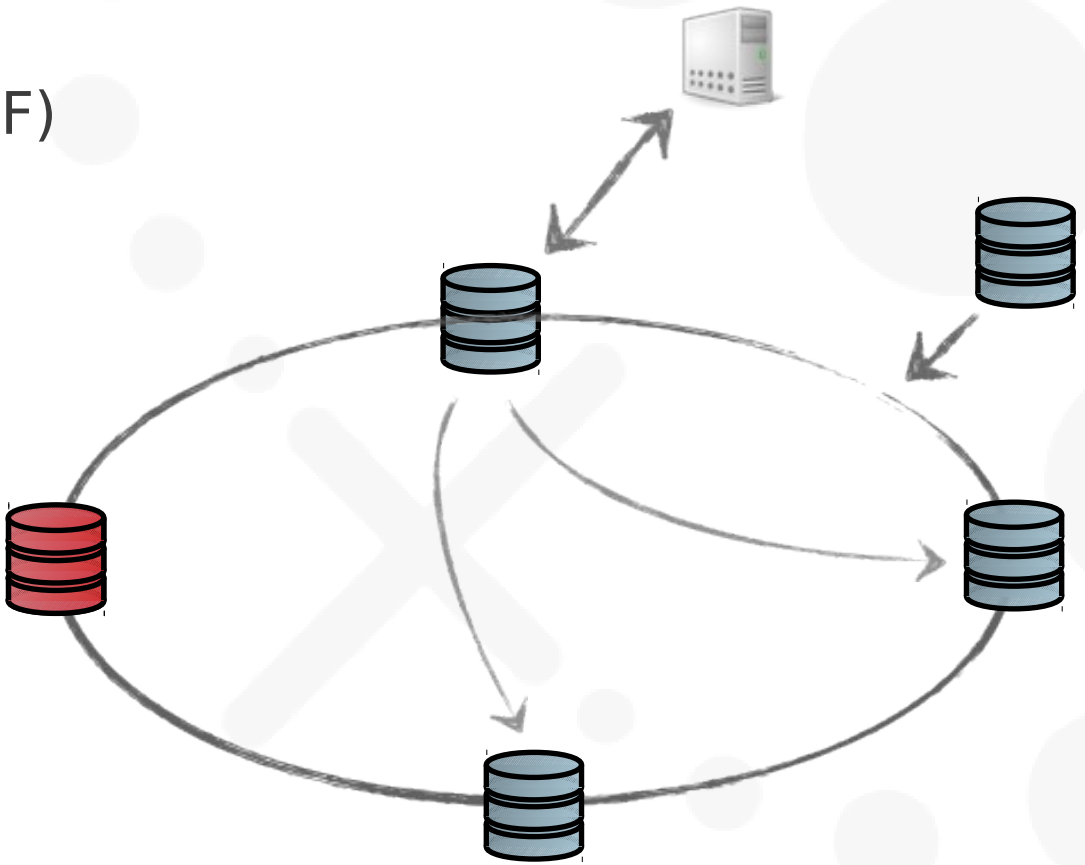
+

...

Apache Cassandra

A database system:

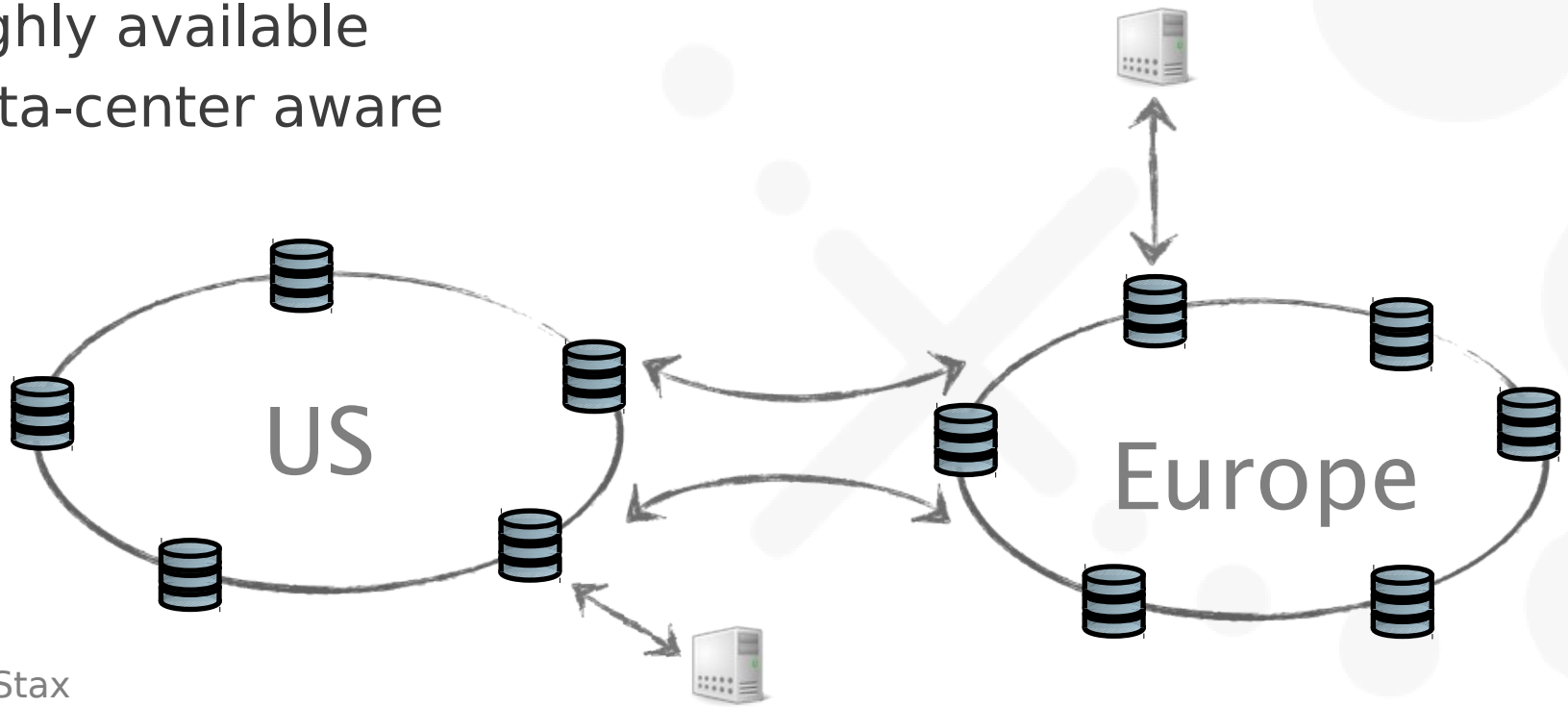
- distributed
- replicated & durable
- scalable
- fault-tolerant (no SPOF)
- highly available



Apache Cassandra

A database:

- distributed
- replicated & durable
- scalable
- fault-tolerant (no SPOF)
- highly available
- data-center aware



Apache Cassandra Storage Engine

- Wide tables, up to 2GB per row
- Fast writes
- Fast primary-key searches
- Durability (commit log)
- Secondary indexes

- No full text search

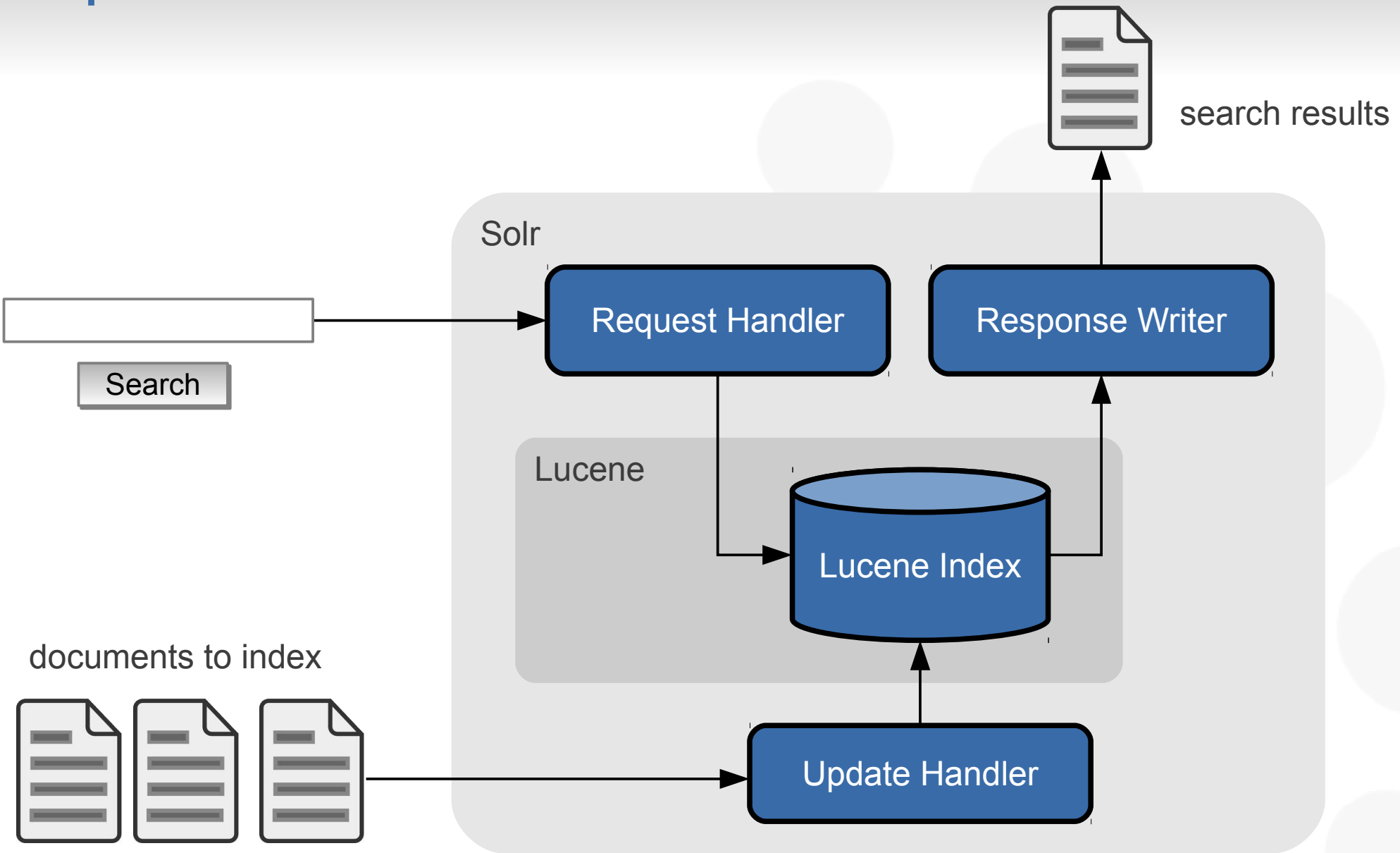
Apache Solr

Database system specialized at searching text:

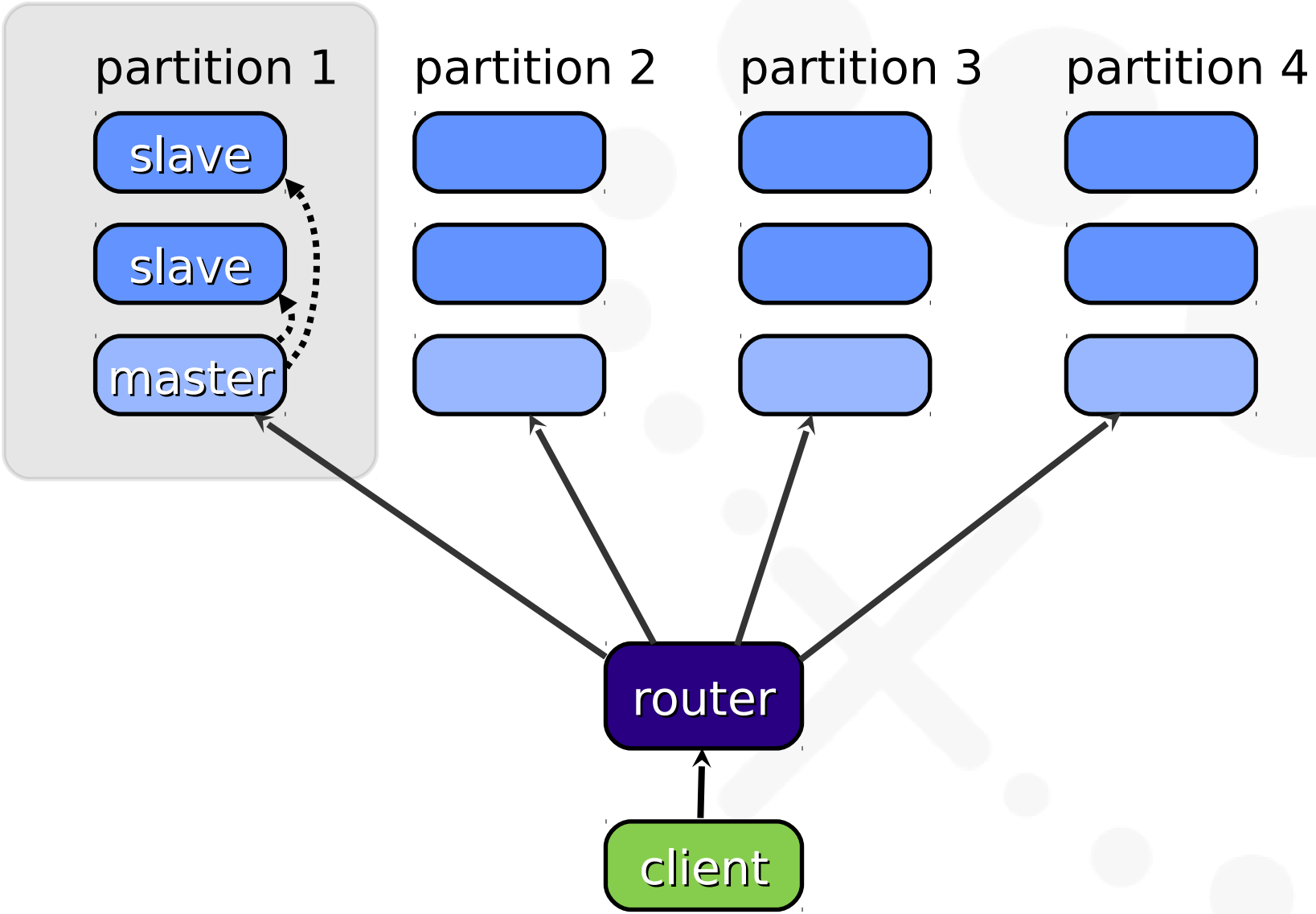
- Language-aware
  - e.g. lowercasing ὈΔΥΣΕΎΣ produces ὀδυσεύς
  - can do stemming / stop-world elimination, etc.
- Supports relevance scoring
- Supports complex range queries

- Centralized

Apache Solr



Classic Partitioning with SPOF



Availability

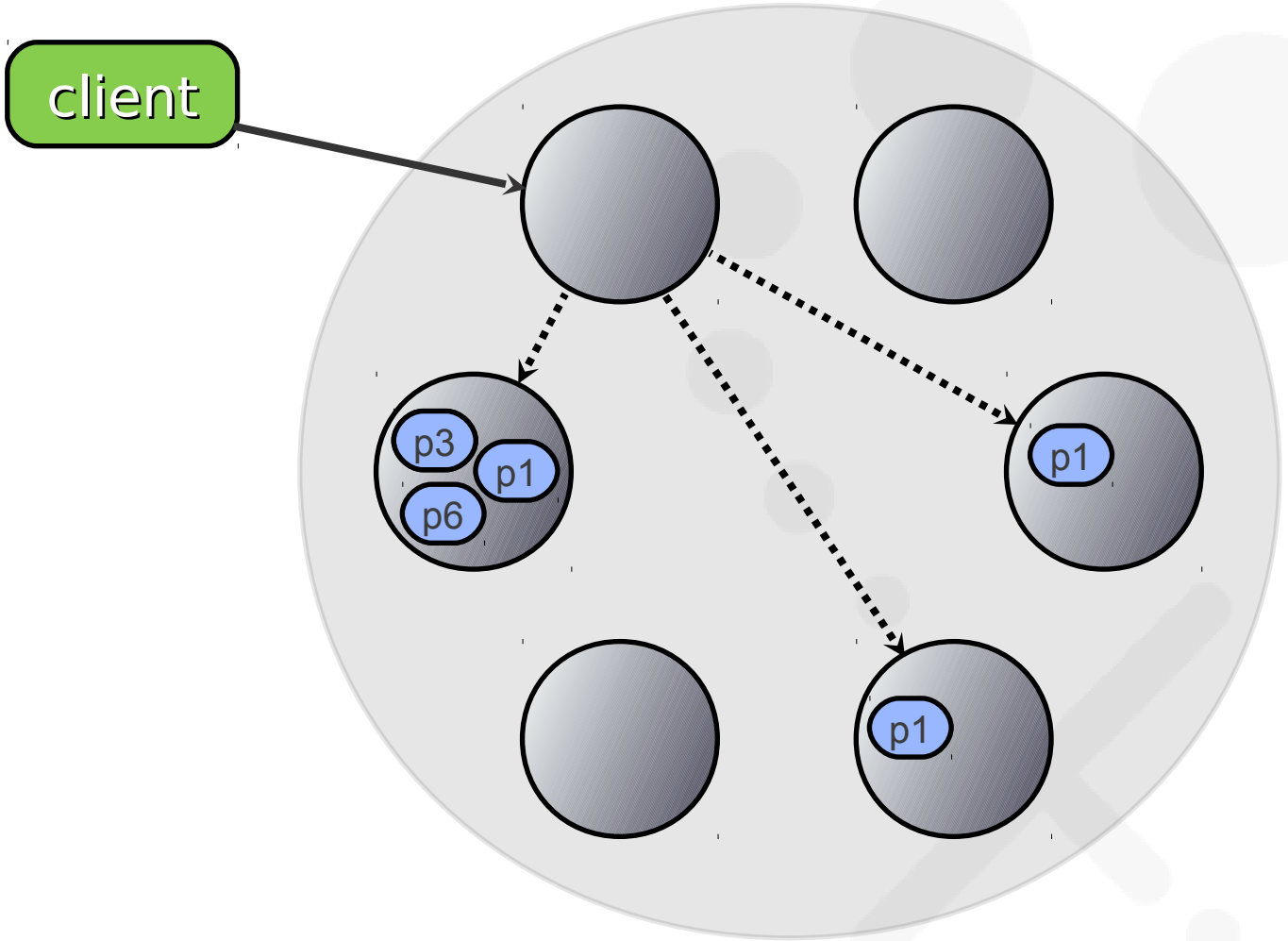
“High availability implies that a single fault will not bring down your system. Not ‘we’ll recover quickly.’”

-- Ben Coverston: DataStax

“The biggest problem with failover is that you're almost never using it until it really hurts. It's like backups that you never test.”

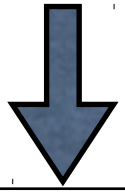
-- Rick Branson: Instagram

Fully Distributed, no SPOF



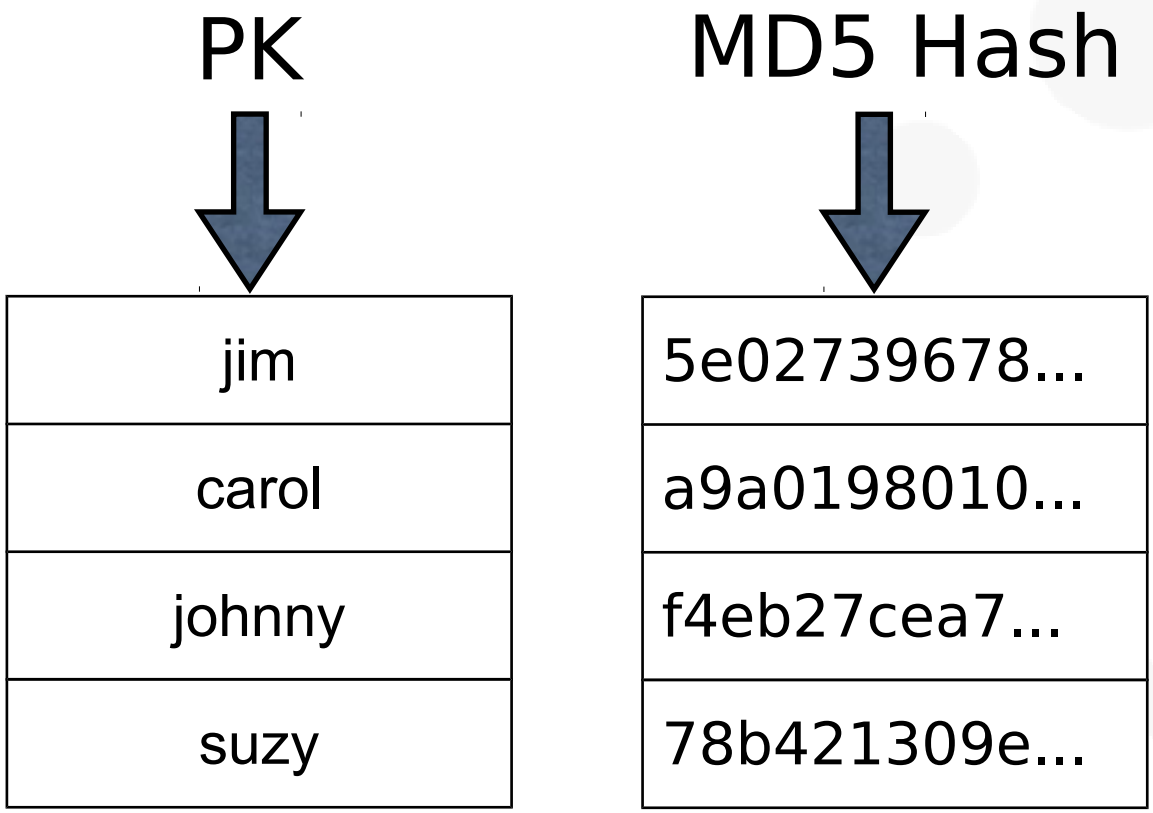
Partitioning

Primary key determines placement*



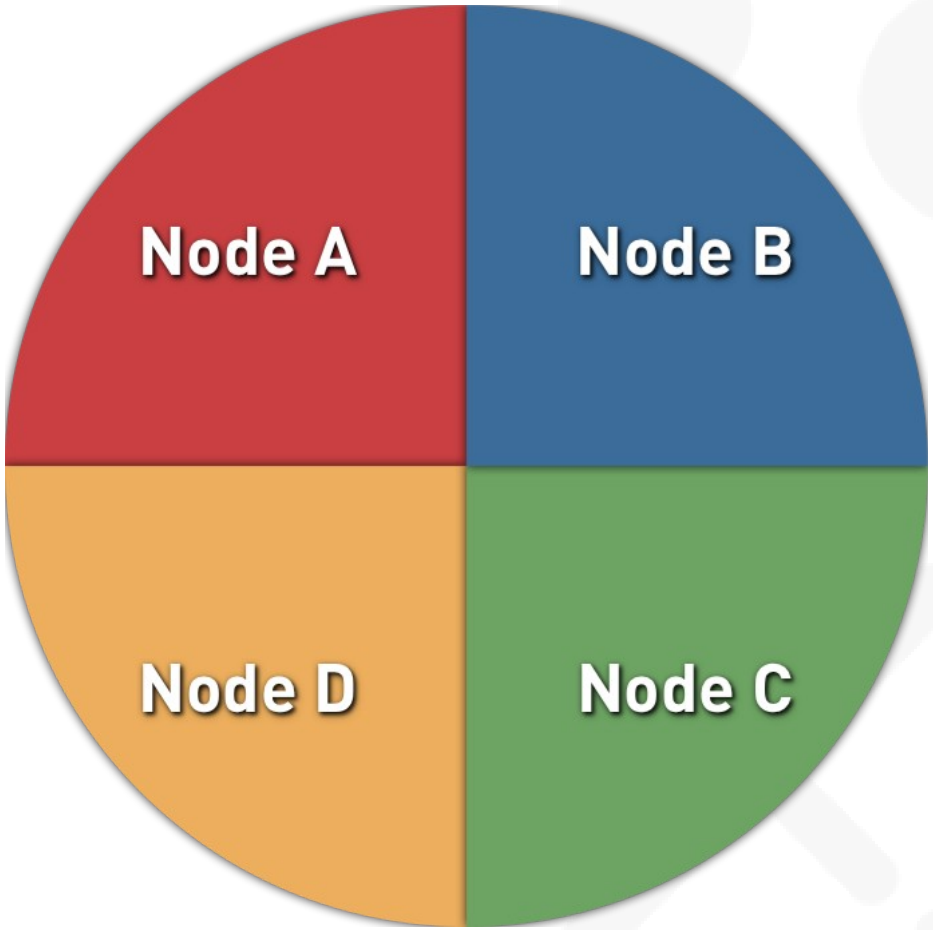
jim	age: 36	car: camaro	gender: M
carol	age: 37	car: subaru	
johnny	age: 12	gender: M	
suzy	age: 10	gender: F	

Partitioning



MD5* hash operation
yields a 128-bit number
for keys
of any size.

The “Token Ring”



	Start	End
A	0xc000000000..1	0x0000000000..0
B	0x0000000000..1	0x4000000000..0
C	0x4000000000..1	0x8000000000..0
D	0x8000000000..1	0xc000000000..0

jim	5e02739678...
carol	a9a0198010...
johnny	f4eb27cea7...
suzy	78b421309e...

	Start	End
A	0xc000000000..1	0x0000000000..0
B	0x0000000000..1	0x4000000000..0
C	0x4000000000..1	0x8000000000..0
D	0x8000000000..1	0xc000000000..0

jim	5e02739678...
carol	a9a0198010...
johnny	f4eb27cea7...
suzy	78b421309e...

	Start	End
A	0xc000000000..1	0x0000000000..0
B	0x0000000000..1	0x4000000000..0
C	0x4000000000..1	0x8000000000..0
D	0x8000000000..1	0xc000000000..0

jim	5e02739678...
carol	a9a0198010...
johnny	f4eb27cea7...
suzy	78b421309e...

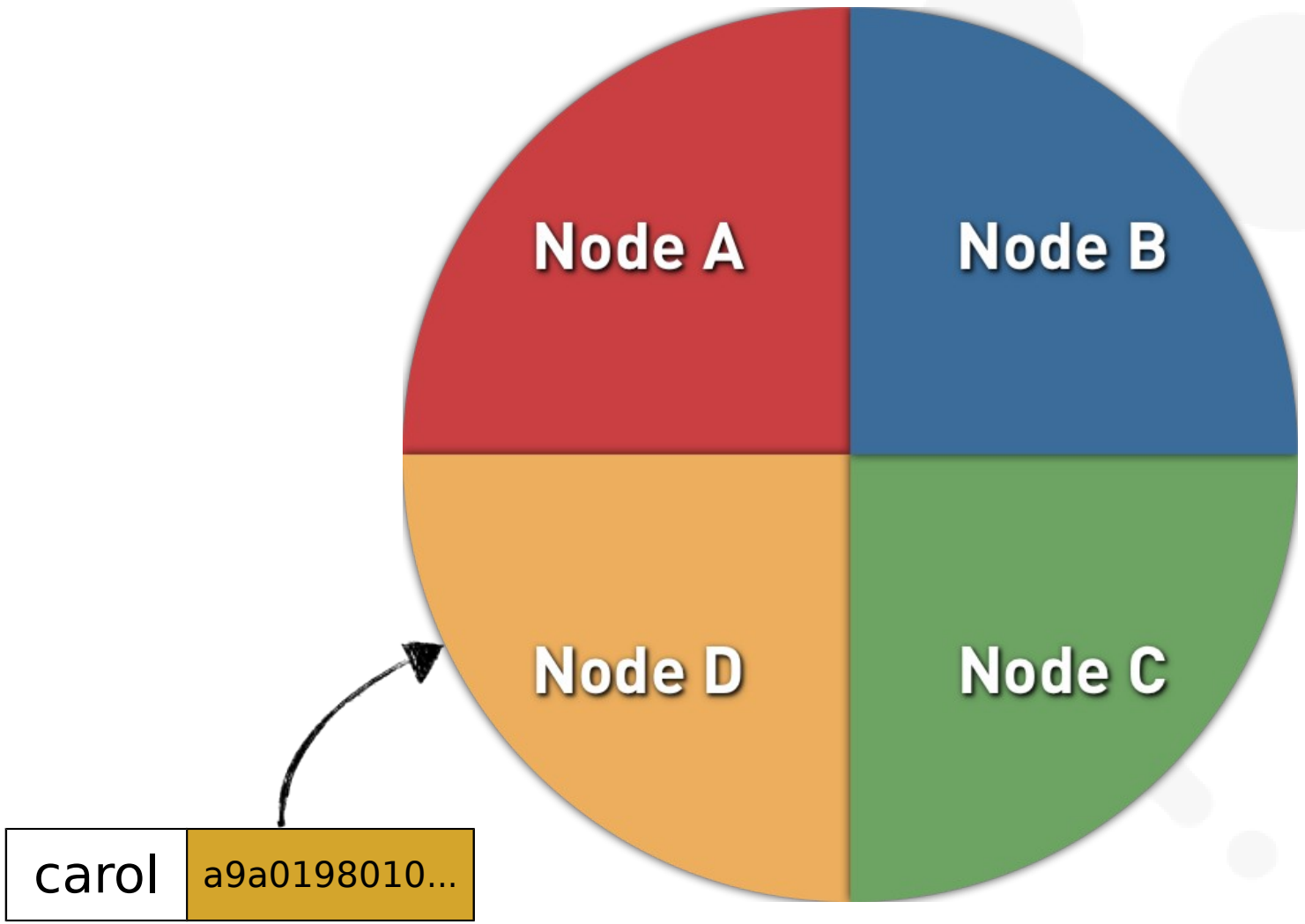
	Start	End
A	0xc000000000..1	0x0000000000..0
B	0x0000000000..1	0x4000000000..0
C	0x4000000000..1	0x8000000000..0
D	0x8000000000..1	0xc000000000..0

jim	5e02739678...
carol	a9a0198010...
johnny	f4eb27cea7...
suzy	78b421309e...

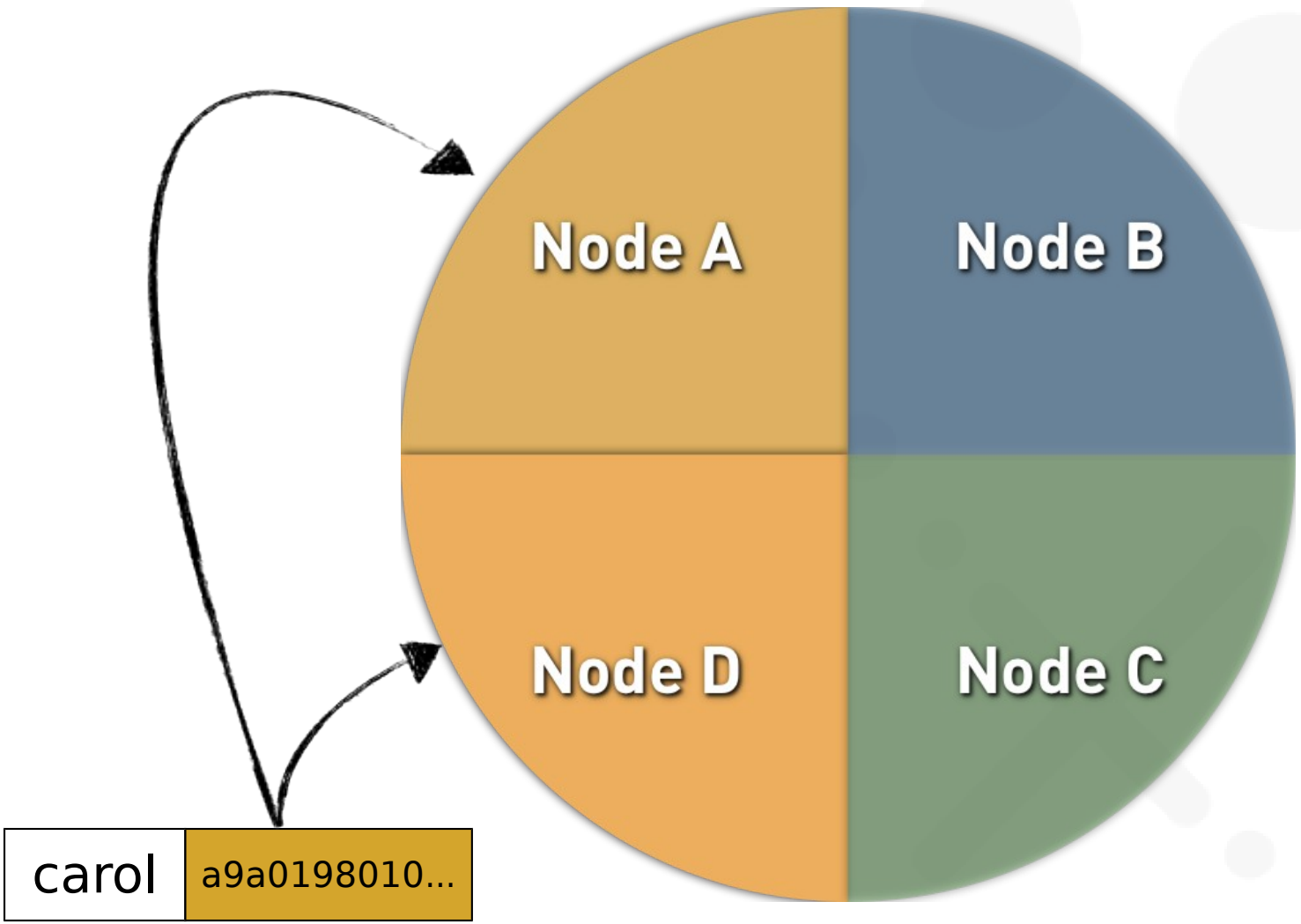
	Start	End
A	0xc000000000..1	0x0000000000..0
B	0x0000000000..1	0x4000000000..0
C	0x4000000000..1	0x8000000000..0
D	0x8000000000..1	0xc000000000..0

jim	5e02739678...
carol	a9a0198010...
johnny	f4eb27cea7...
suzy	78b421309e...

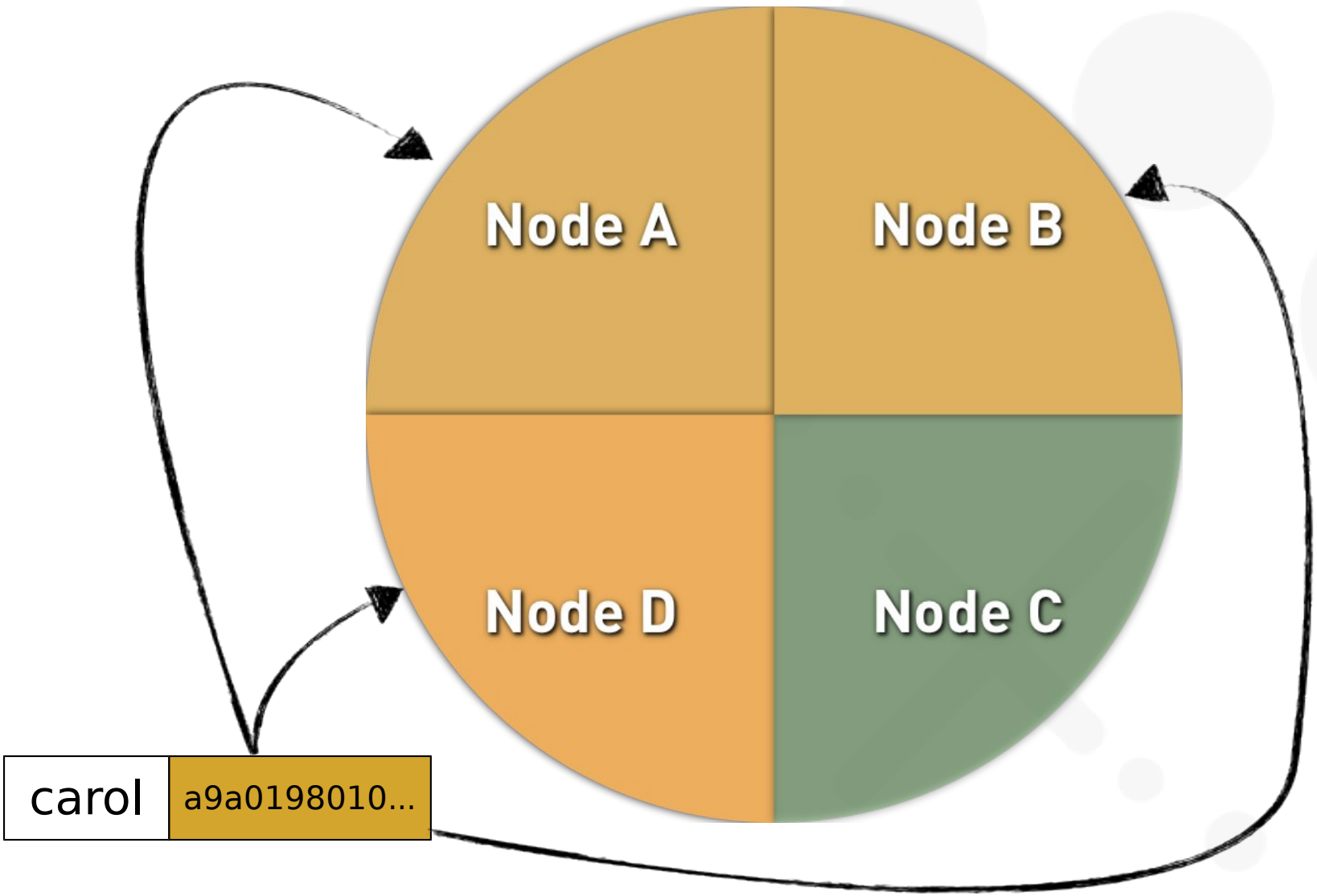
Replication



Replication



Replication



Bringing Cassandra and Solr Together

Cassandra	Solr
Keyspace	Core
Table (Column Family)	
Row	Document
Column	Field

Schema

schema.xml and *solrconfig.xml* stored and distributed by Cassandra

```
<schema name="wikipedia" version="1.1">

  <types>
    <fieldType name="string" class="solr.StrField"/>
    <fieldType name="text" class="solr.TextField"/>
  </types>

  <fields>
    <field name="part_id" type="string" indexed="true"  stored="true"/>
    <field name="description" type="text" indexed="true"  stored="true"/>
  </fields>

  <defaultSearchField>description</defaultSearchField>
  <uniqueKey>part_id</uniqueKey>

</schema>
```

Data Mapping

Cassandra table

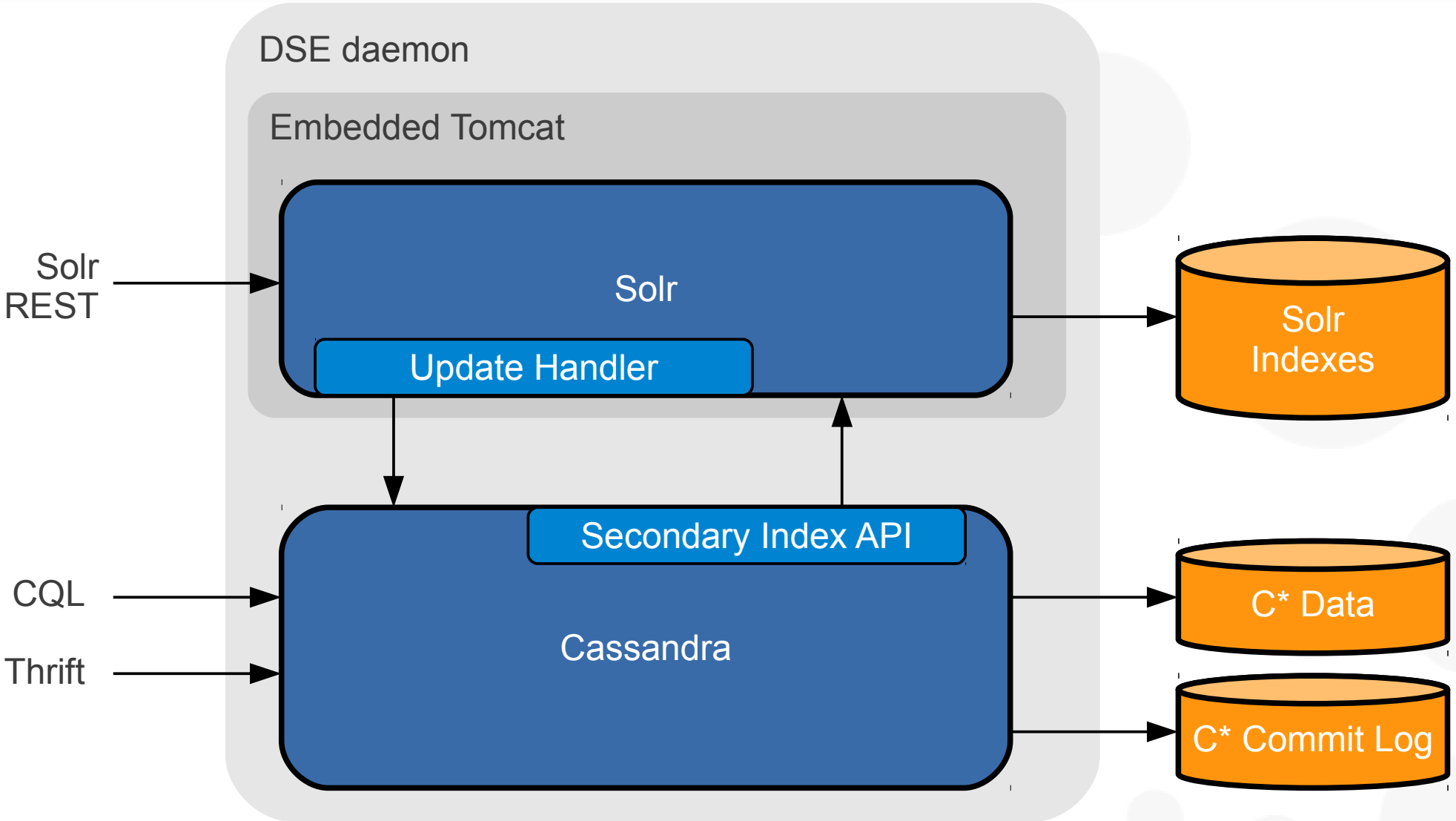
part_id	description
2N2222	Low power bipolar NPN transistor
TL074	Low noise JFET-input operational amplifier
LM3886	High performance integrated audio power amplifier

field	value
part_id	2N2222
description	Low power bipolar NPN transistor

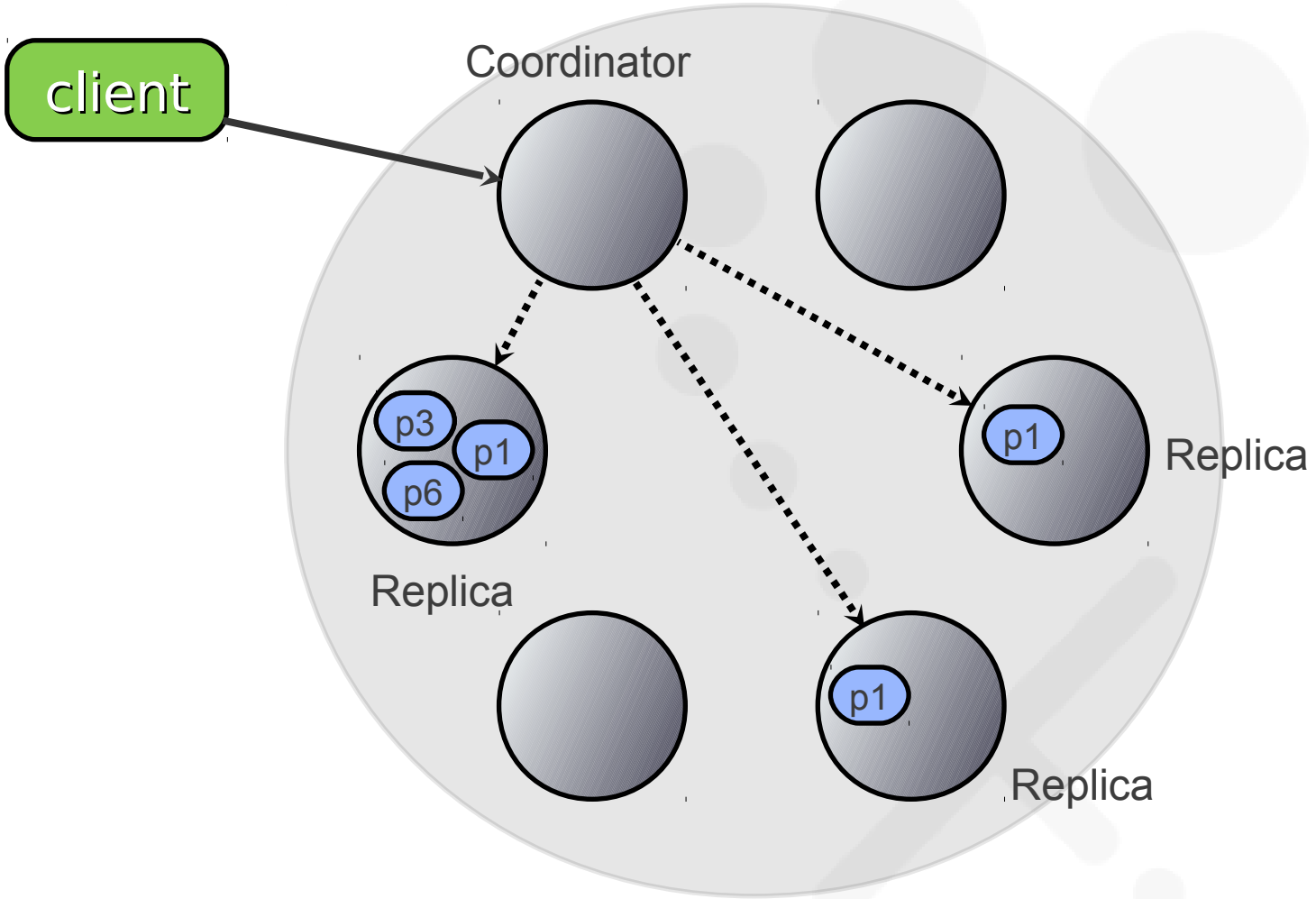
field	value
part_id	TL074
description	Low noise JFET-input operational amplifier

field	value
part_id	LM3886
description	High performance integrated audio power amplifier

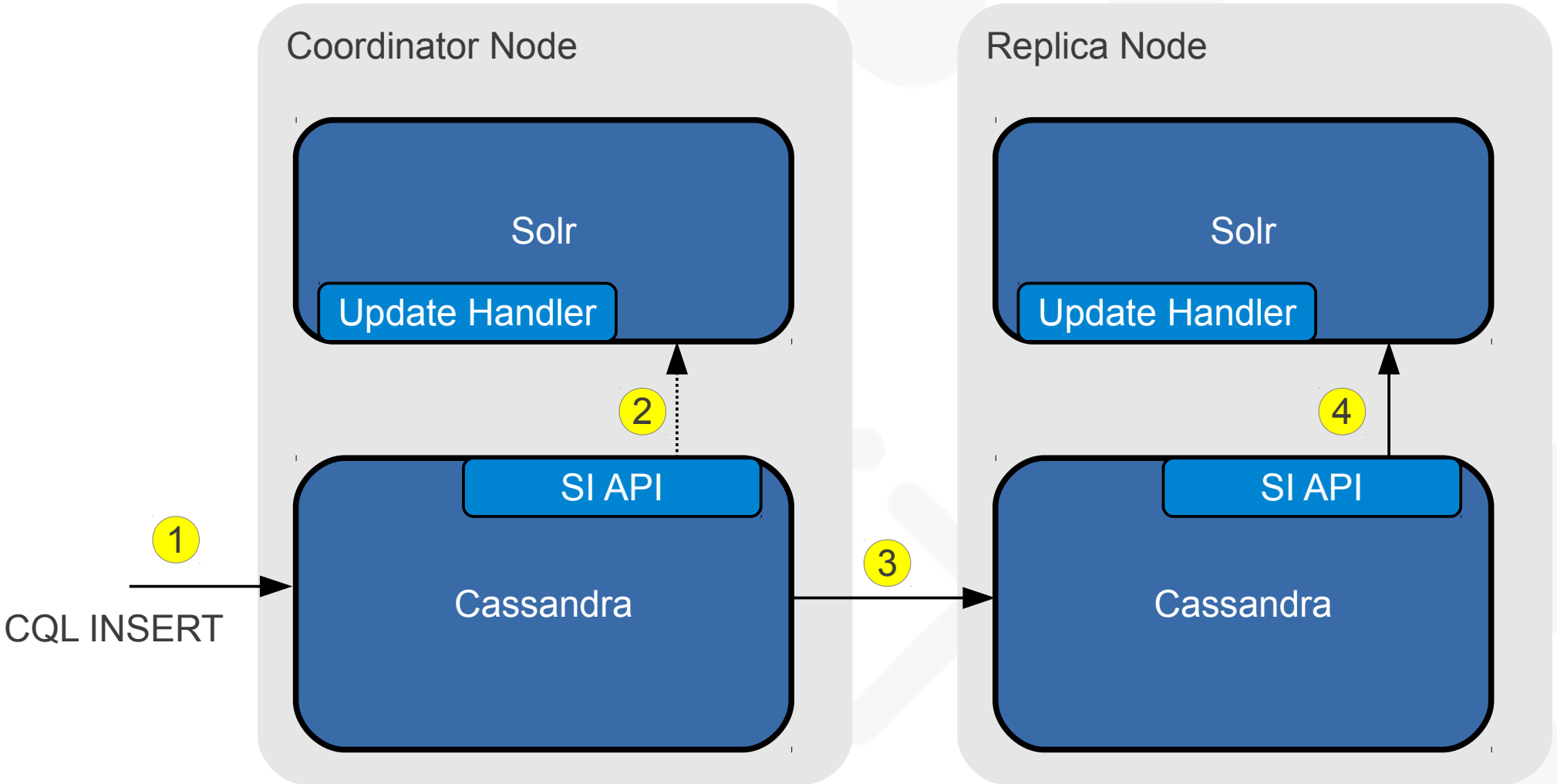
DSE Search Architecture



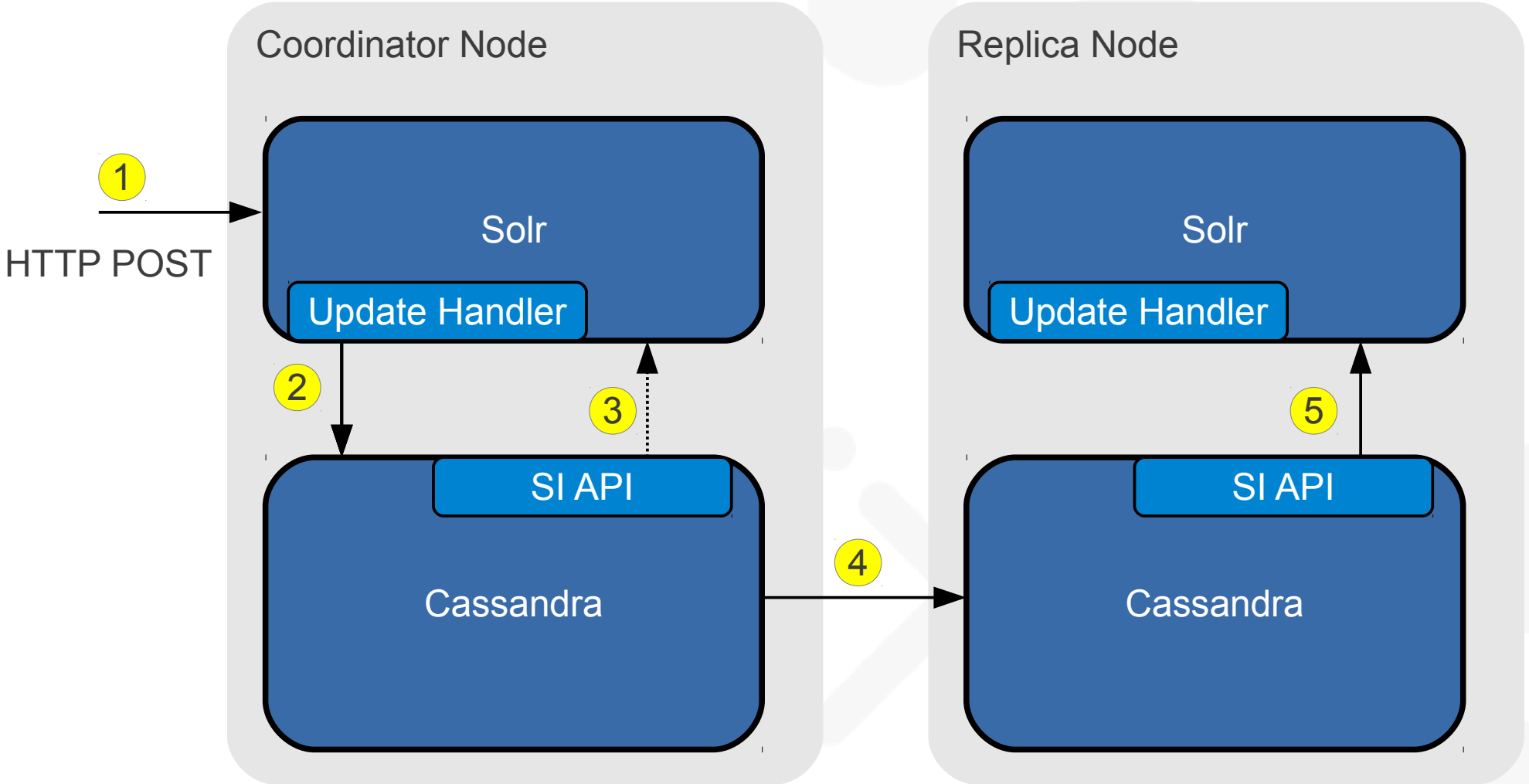
Inserting



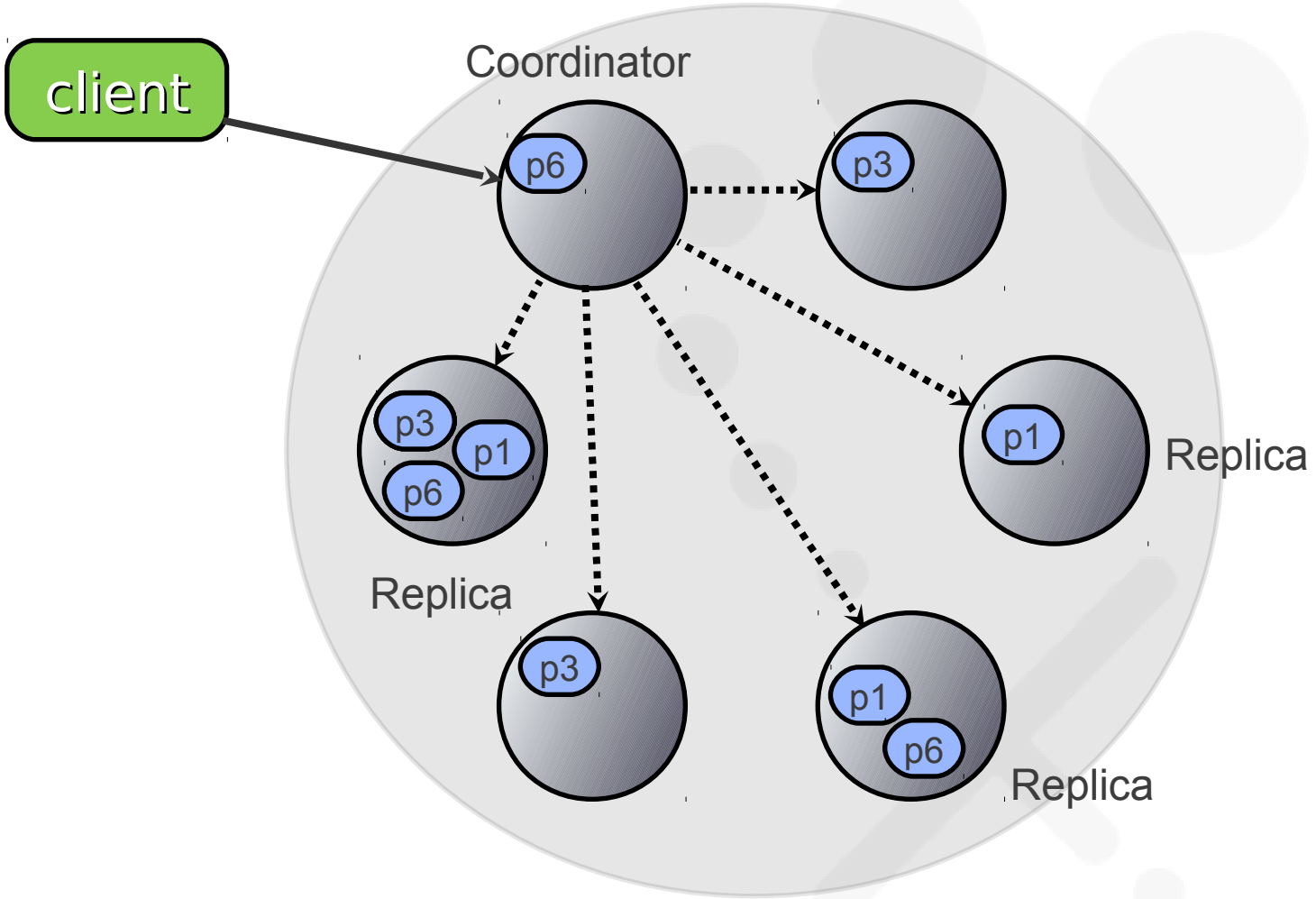
Inserting through Cassandra API



Inserting through Solr API



Querying



How many nodes to contact?

- We don't know the primary key
- Theory:
  contact at least one replica for every token range

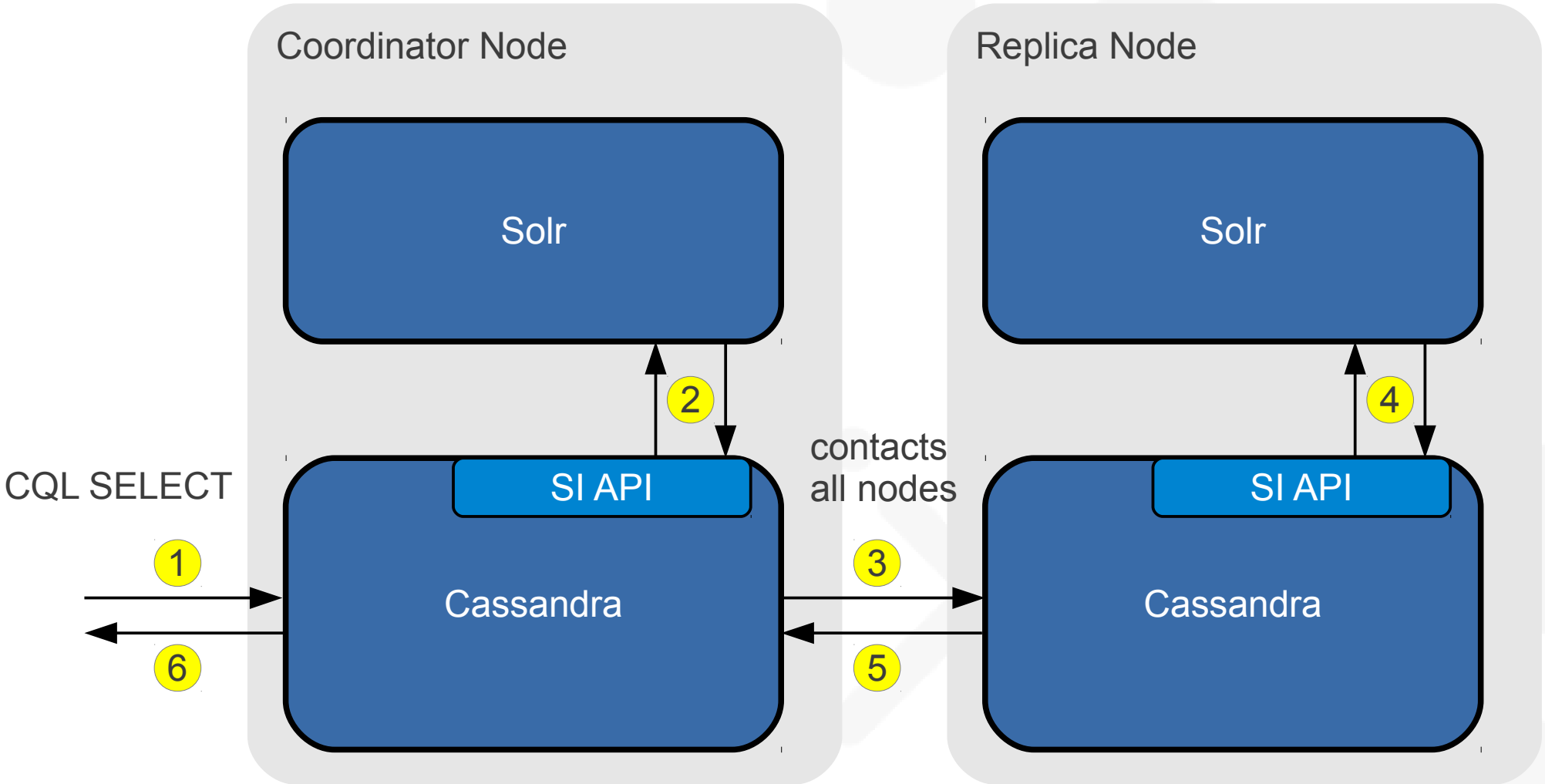
- Cassandra contacts all nodes
- Our custom Solr SearchComponent
  does intelligent shard selection

Querying through CQL

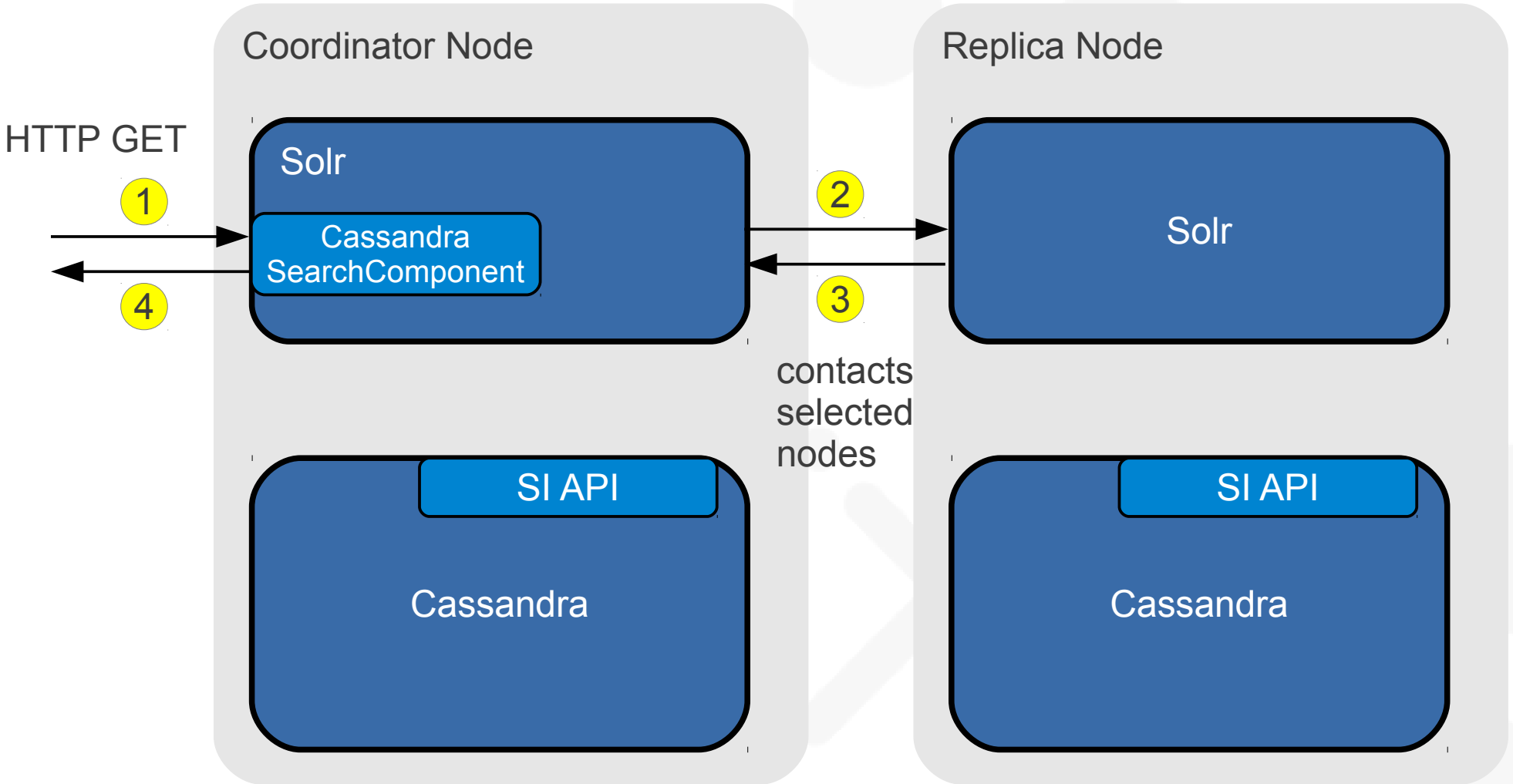
```
SELECT title FROM solr WHERE solr_query='title:natio*';

 title
-------------------------------------------------------------------------
                                  Bolivia national football team 2002
 List of French born footballers who have played for other national teams
                  Lithuania national basketball team at Eurobasket 2009
                                  Bolivia national football team 2000
                                Kenya national under-20 football team
                                  Bolivia national football team 1999
                              Israel men's national inline hockey team
                                  Bolivia national football team 2001
```

Querying through CQL



Querying through Solr API



Shard Selection Algorithm

- Tries to minimize the number of selected shards

  optimum number of shards = $\lceil N / RF \rceil$

- Tries to fetch data from the closest nodes
  - local node
  - nodes on the same rack
  - nodes in the same DC

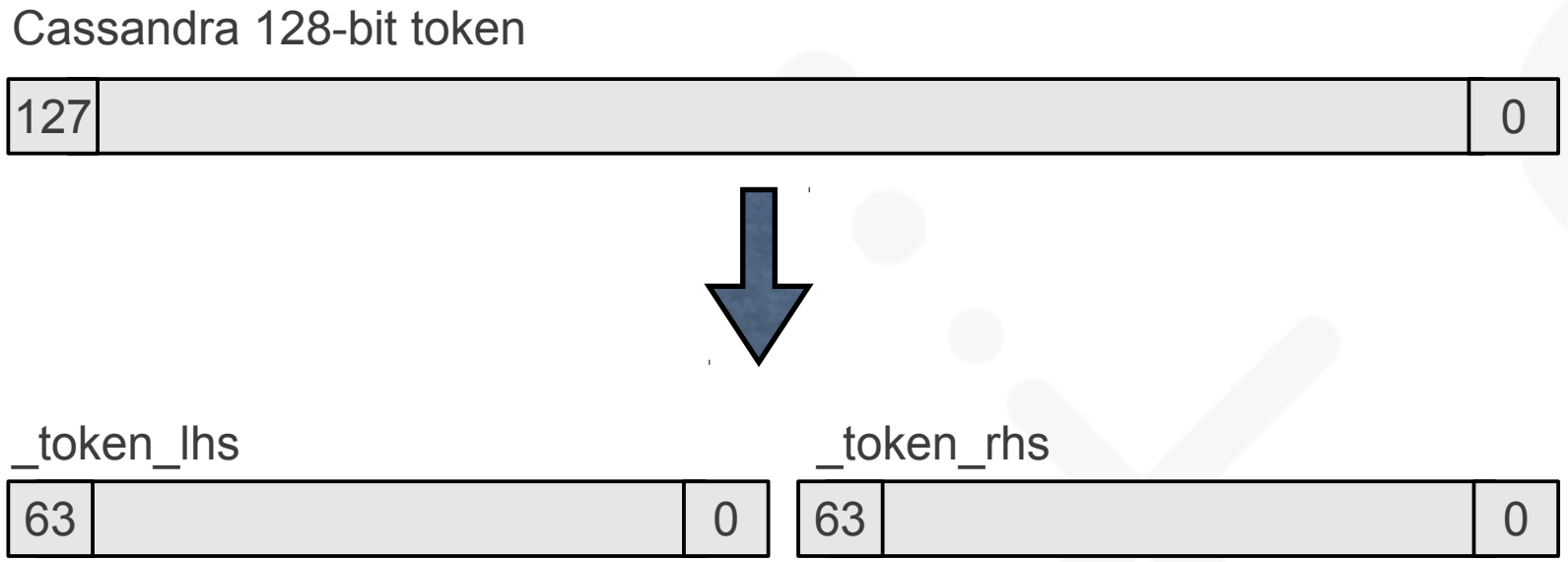
- Balances the load

Shard Selection Algorithm

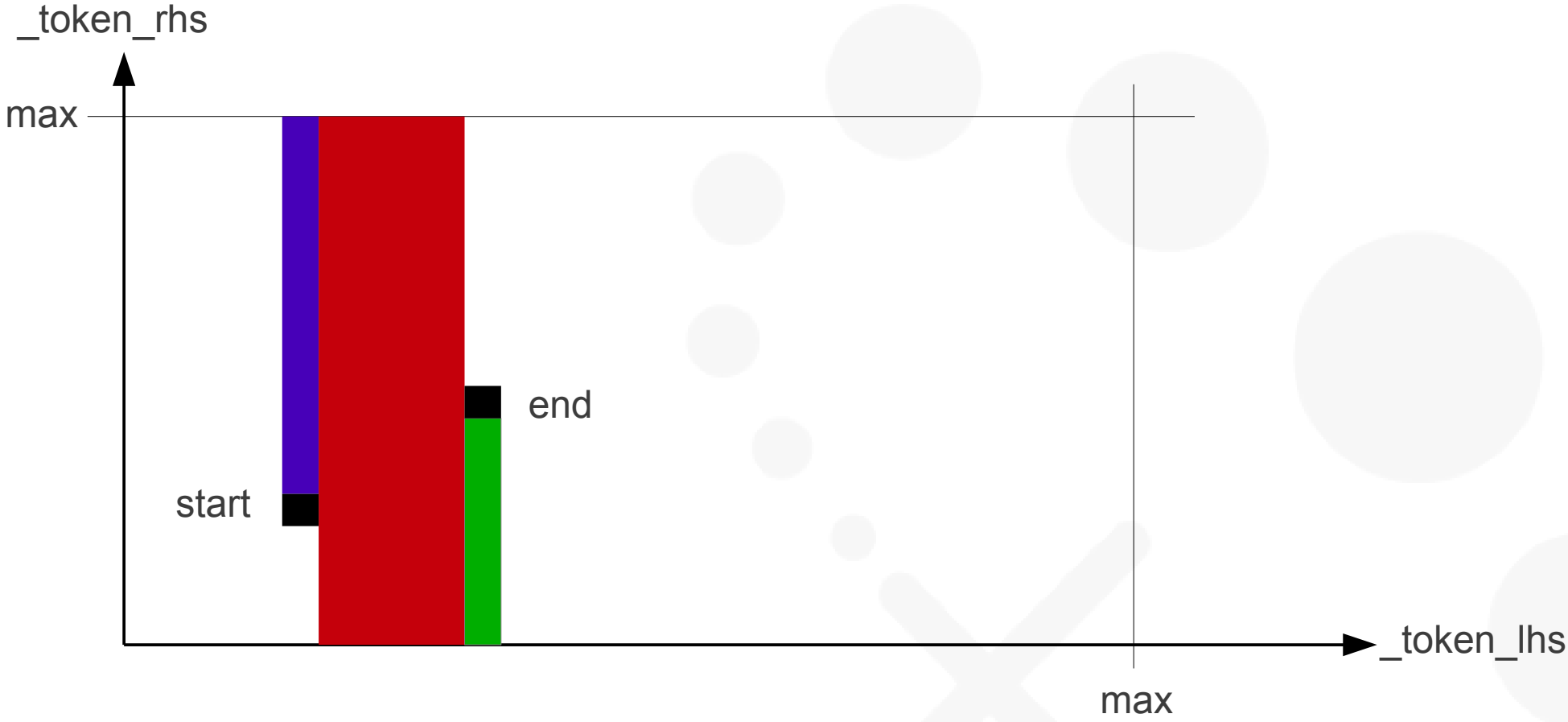
1. Always select the local node first
2. Select the closest node that is covering the highest number of token ranges not yet covered.
3. Repeat the previous step until all ranges are covered.

Querying Shards

Solr does not support indexing 128-bit numbers

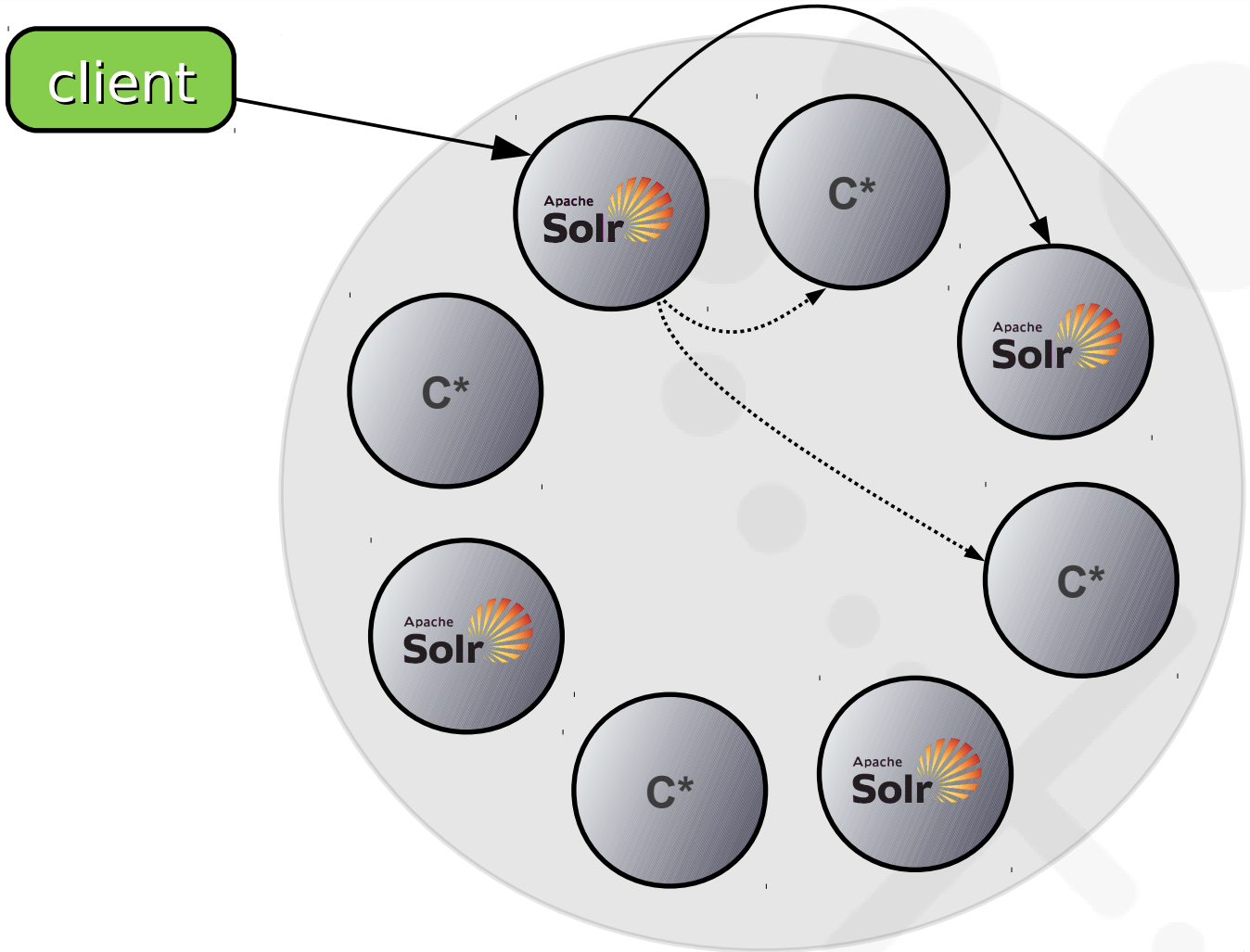


Querying Shards




```
((+_token_lhs:3074457345618258602 +_token_rhs:[3074457345618258604 TO *])
  OR (+_token_lhs:[3074457345618258603 TO 6148914691236517204])
  OR (+_token_lhs:6148914691236517205 +_token_rhs:[* TO -3074457345618258602]))
```

Workload Separation



Replication Factor
Solr: 2
Cassandra: 2

Questions?

- http://www.datastax.com/docs
- http://www.datastax.com/products/enterprise

Cassandra Summit 2013, June 11-12, San Francisco, CA
http://www.datastax.com/company/news-and-events/events/cassandrasummit2013