DataStax Enterprise Search

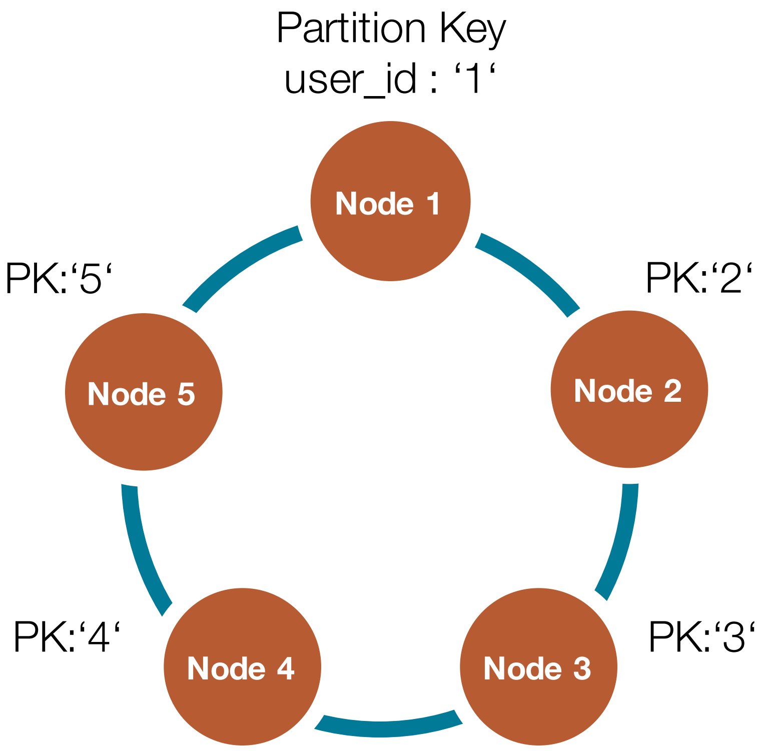
Negib Marhoul, Solution Engineer, DataStax

27. Februar 2017

Agenda

1	Introduction DSE Search
2	Hands-On DSE Search

The data model is important



```
create table users (
    user_id int,
    name text,
    age int,
    gender boolean,
    PRIMARY KEY(user_id))
```

```
SELECT name FROM users WHERE user_id=1
```

```
SELECT * FROM users WHERE name=„Thomas“

SELECT * FROM users WHERE age>45 and name=„Thomas“
```

Not possible out of the box without secondary Idx or further extra tables

No problem with DSE Search

What is the value of DSE Search?

Multi-criteria WHERE Constraints

- WHERE constraints with multiple columns
- No extra tables needed

Full Text Search

- **Wildcards** ? *, like or Lemmatisation
- **Faceting**, Slice and Dice

Live Indexing

- Real Time Search, High Index throughput

Geospatial queries

- Queries with coordinates and distances search

Integrated with CQL

1. Create core, schema.xml and solrconfig.xml.
2. [optional] customize schema.xml, solrconfig.xml
3. Start indexing, re-indexing

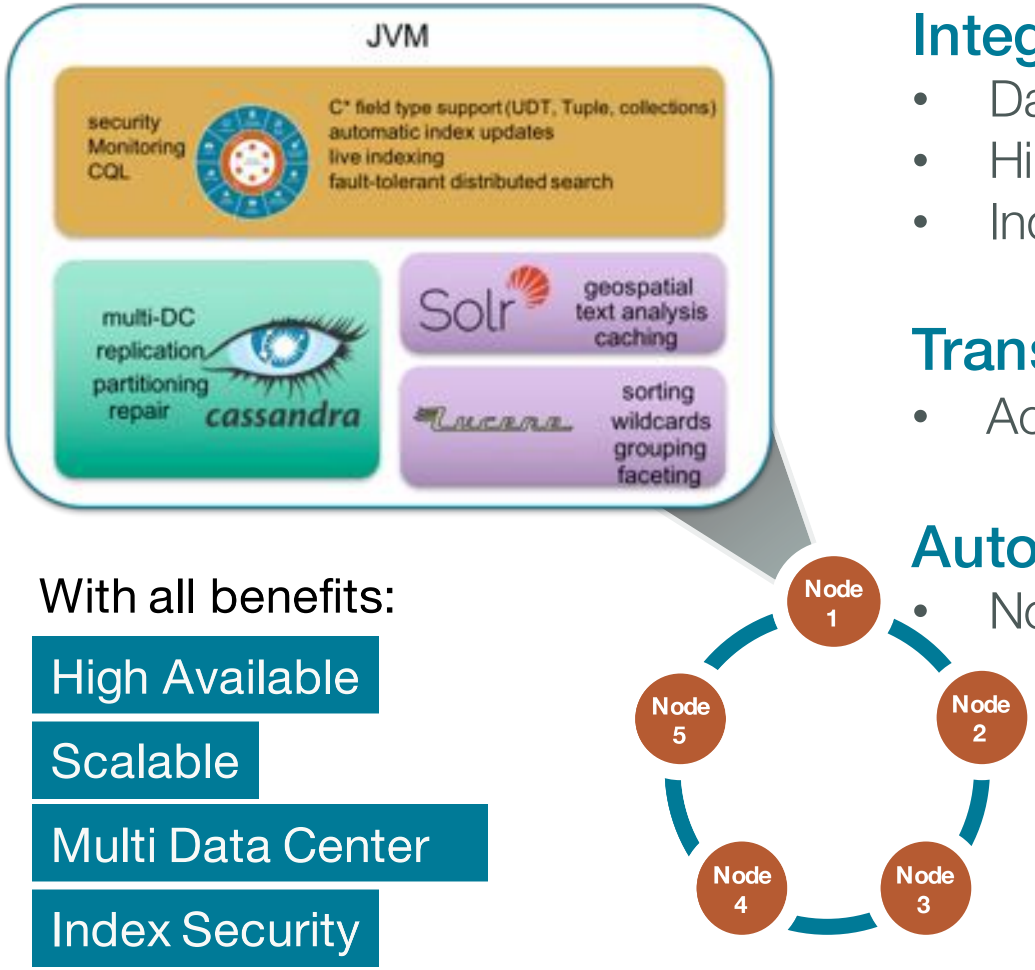
```
Wildcard search
cqlsh> select * from sales where solr_cql = {`fq`:`name=gre*`}

Facet query
cqlsh> SELECT * FROM sales WHERE solr_query='{"q":”name:*","facet":{"field":”item"}}';

Range query
Cqlsh> SELECT * FROM sales WHERE solr_query='{"q":’dt:[2017-01-01 TO 2017-01-10]’}';

Geo search
cqlsh> select * from sales where solr_cql = {`fq`:`name=Hein*`}
```

DSE Search Architektur



With all benefits:

- High Available
- Scalable
- Multi Data Center
- Index Security

Integrated in the the same JVM

- Data Locality (TokenRanges) and Shared Memory
- High Available, no master needed
- Index is sharded to all nodes

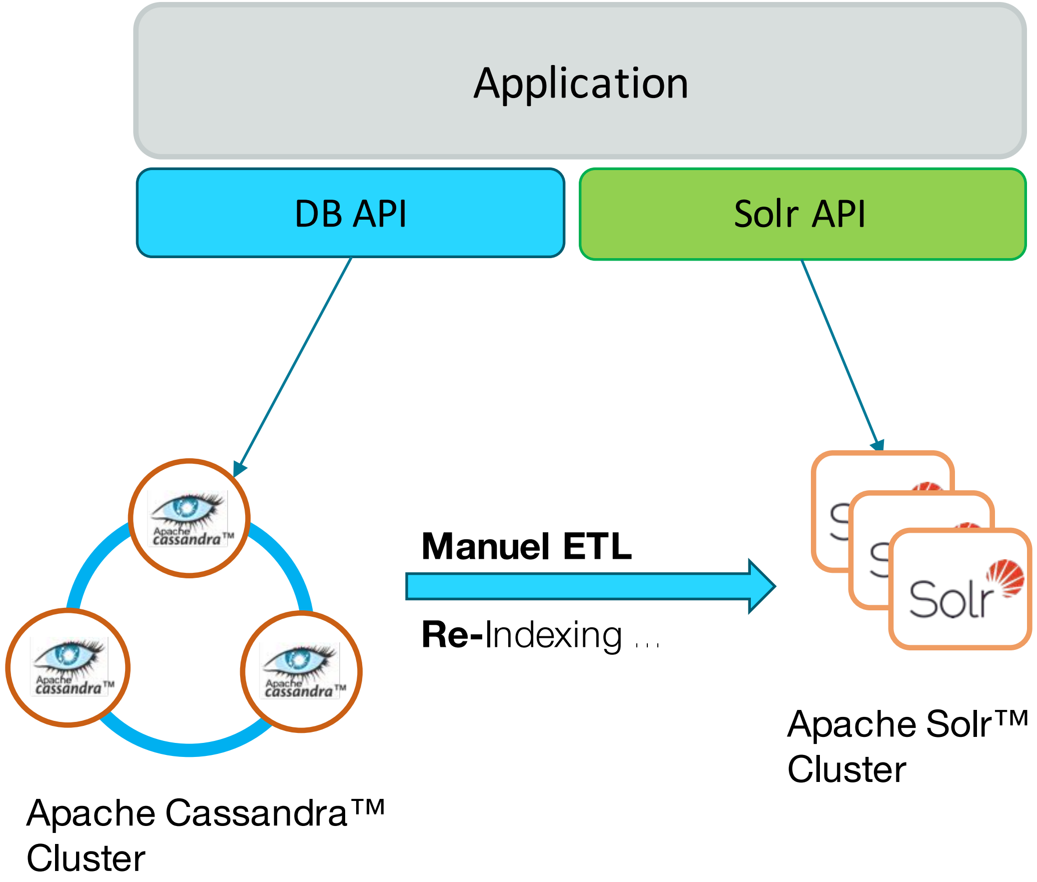
Transparency

- Access via CQL or REST API

Automated indexing with INSERT / UPDATE

- No extra ETL Process needed

The Open Source Way



Separated Search Cluster

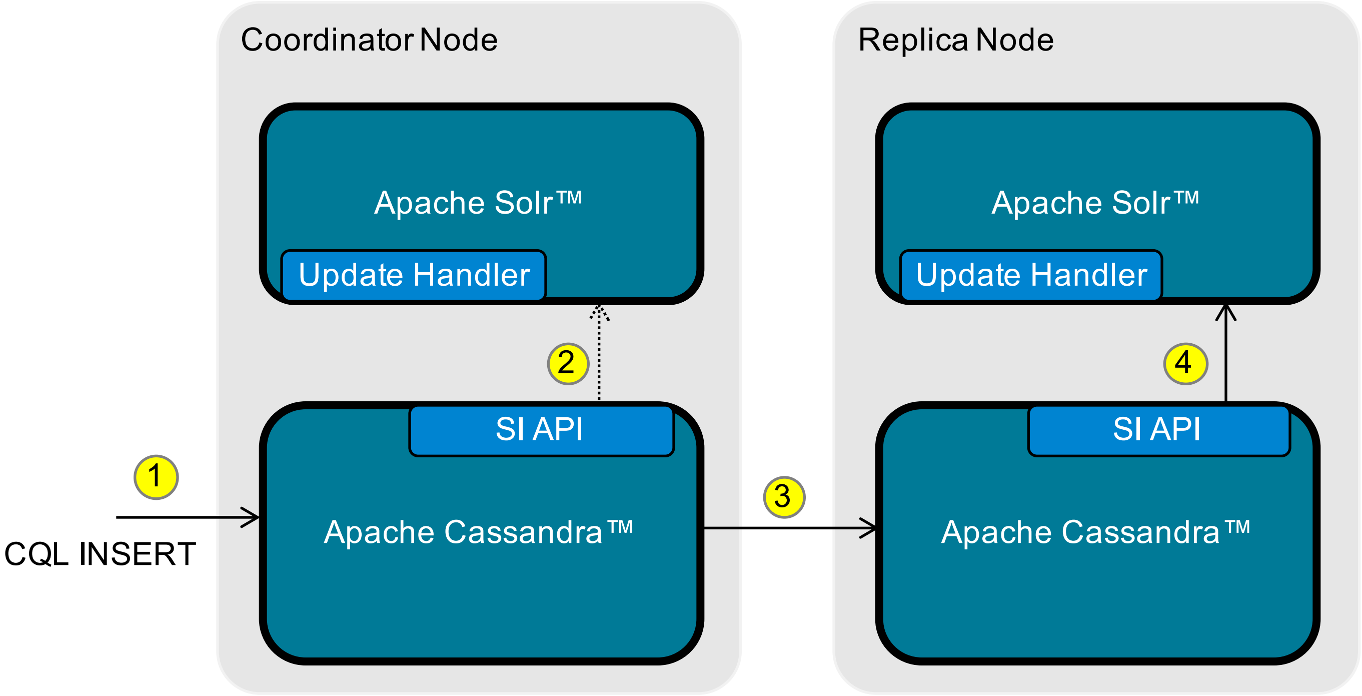
- “Split Brain” risk, data inconsistency
- ETL to generate, update and re-create index

Complex application

Two seperated APIs, driver, read pathes

https://docs.datastax.com/en/datastax_enterprise/5.0/datastax_enterprise/srch/searchOssSolrDiff.html

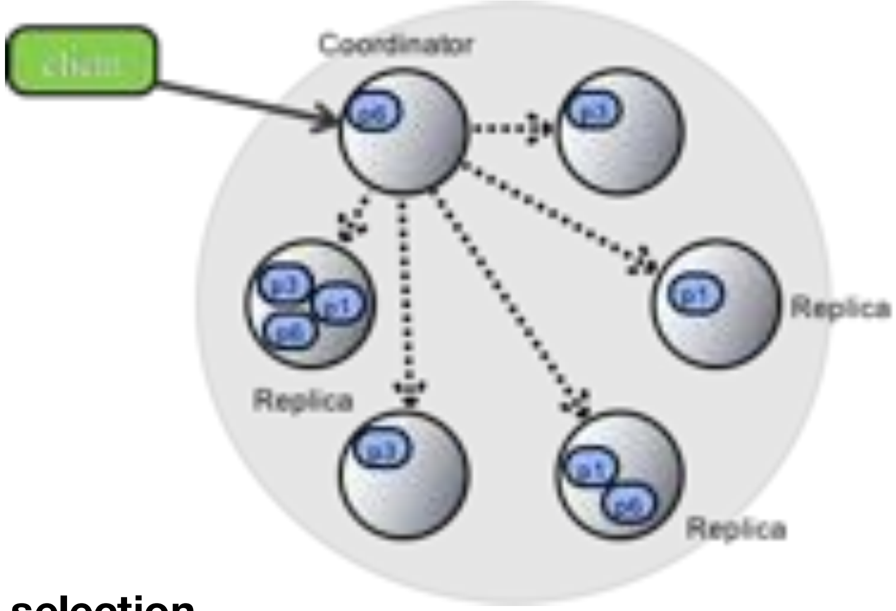
Inserting through Apache Cassandra™ CQL



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
```

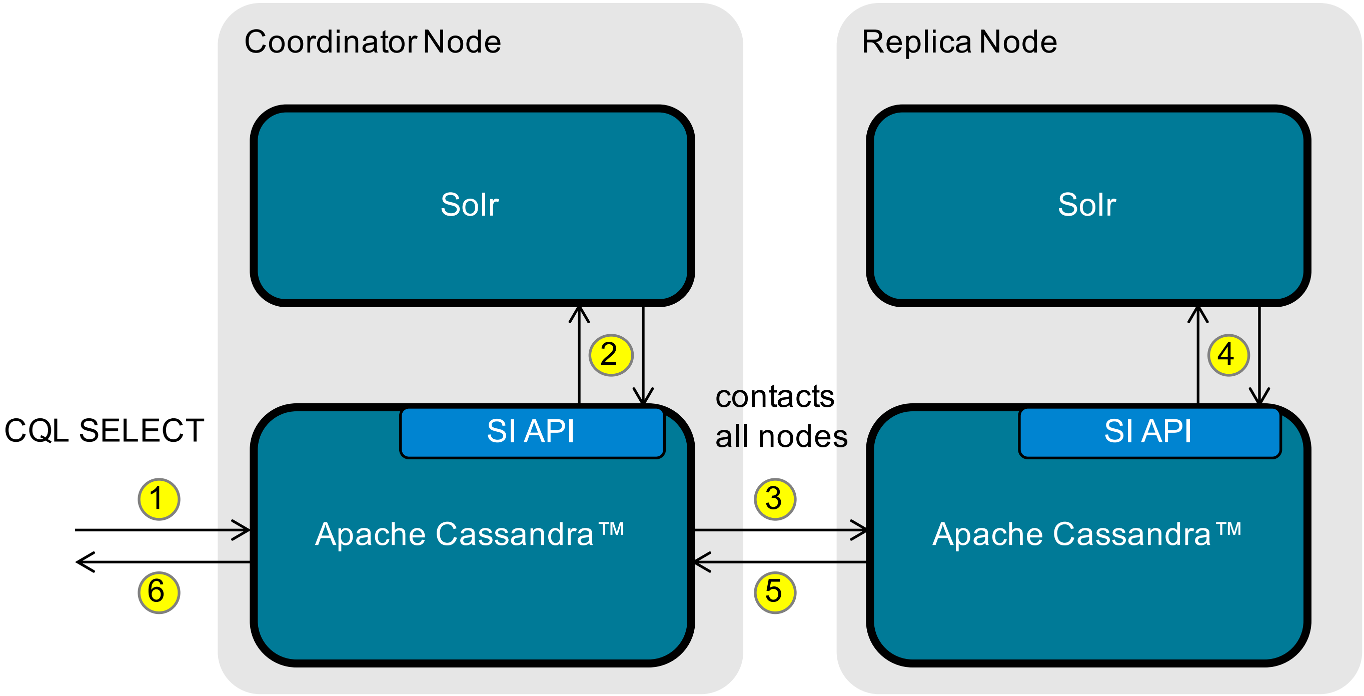


How many nodes to contact?

- 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



Lab 5 : Hands-on DSE Search

Vielen Dank!