OPEN SOURCE BACKUP CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

R&B Elasticsearch Data

Curate the holy Art!

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Who am I...

- Senior Open Source Consultant
- Data Nerd
- Beekeeper

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Where am I working..

@NETWAYS GmbH

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Elasticsearch

A Vernissage of distributed information storage

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Elasticsearch is a...

Apache Lucene based distributed data search engine.

Data Format

- Entity is a DOCUMENT
- DOCUMENT consist of FIELDS

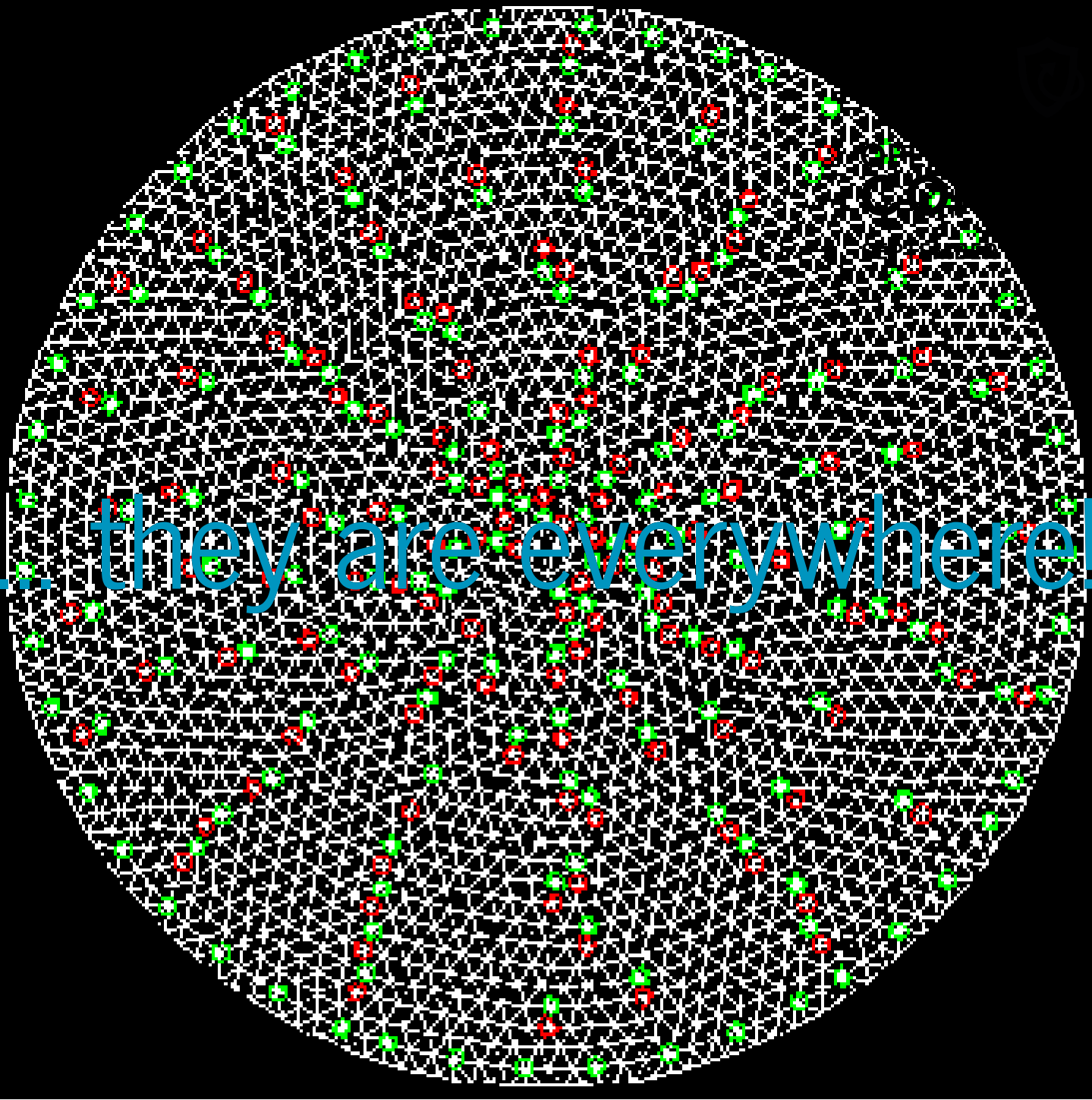
Storage Mechanism

- DOCUMENTS are stored in INDICES
- INDICES are divided into SHARDS
- SHARDS are divided into SEGMENTS
- SHARDS are distributed over NODES

OPEN SOURCE BACKUP
CONFERENCE
SEPTEMBER 26, 2018 | COLOGNE

Cluster function

- Indices and their Shards are distributed stored
- The shard are balanced
- Shards have replicas which are never on the same node as their priamary


...they are everywhere!

Why should i do backups then..?

- Total loss prevention
- Historical Archive or Audit/ISO purposes
- Housekeeping

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Backup Method

Explain the Art

Function

- Snapshot with merge function (no incremental backup is needed)
- Repository based
- Distributed

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Conclusion 

- File based backups from a single node are not valid
- All nodes must be reach the repository
- You need storage with shared write/read access

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

The Depot aka Repository

Repository Types

- S3 -> AWS Only!
- HDFS (Hadoop)
- AZURE
- GCS Google Cloud Storage

The Winner 

- fs -> file shared system like nfs
- Needs to be mounted on each node
- Needs to be configured on each node

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Building the base

OPEN SOURCE BACKUP
CONFERENCE
SEPTEMBER 26, 2018 | COLOGNE

Configure the repository path

```
path.repo: ["/elasticbackup"]
```

The path in elasticsearch.yml must be set on all nodes

OPEN SOURCE BACKUP CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Create the Repository

```
curl -X PUT "localhost:9200/_snapshot/fs_repository" -H 'Content-T
{
    "type": "fs",
    "settings": {
        "location": "/elasticbackup",
        "compress": true
    }
}
'
```

This register the repository cluster wide on all nodes

Different Setup purposes

- Can be used for read and write
- Can be used with read only for restoring data into different clusters

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Working with Snapshots

OPEN SOURCE BACKUP
CONFERENCE
SEPTEMBER 26, 2018 | COLOGNE

Tool Tips

- You can use the Elasticsearch HTTP API on 9200 with curl
- You can use the Elasticsearch Curator python library as singletones
- You can use Curator -> Your focus!

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Curator

he runs the museum...

Skills

- Based on python and uses the Elasticsearch client API
- Provides automation with YML config
- Can be used for a lot of housekeeping topics

OPEN SOURCE BACKUP CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Considerations for Snapshots

- All open and active indices will be backuped
- The global cluster state will be included: include_global_state = true (default)
- If shards are not available the backup fails -> partial = false (default)

Considerations for Restore

- Existing index can only be restored if it is closed
- Creates a new index if the index does not exist in cluster
- Restore cluster states: include_global_state = false (default)
- You can restore a index with renaming it and you can alter minimal settings

OPEN SOURCE BACKUP
CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

3 Restore Scenarios

OPEN SOURCE BACKUP CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Disaster Recovery

- Indices will be newly created with same number of shards and replicas
- You should consider to set include_global_state = true (defaults to false)
- The cluster should be able to handle the indices

Restore for Historical Search

- Restore specific index
- Set replica shards to zero
- Do not include cluster state

OPEN SOURCE BACKUP CONFERENCE

SEPTEMBER 26, 2018 | COLOGNE

Restore due to Upgrade/Migration

- Restore a for example a Elaticsearch Version 5 based index into Elasticsearch Version 6
- This for example can be helpful on your way from a Elasticsearch 2.x to Elasticsearch Version 6


OPEN SOURCE BACKUP
CONFERENCE
SEPTEMBER 26, 2018 | COLOGNE
lets leave the theory...

OPEN SOURCE BACKUP
CONFERENCE
SEPTEMBER 26, 2018 | COLOGNE

Questions?

Thank you for your attention