



**adaptTo()**

EUROPE'S LEADING AEM DEVELOPER CONFERENCE

28<sup>th</sup> – 30<sup>th</sup> SEPTEMBER 2020

# Analysing Components and Assets usage

M. Goszczyński, Adobe Consulting Services

- How many components are used on pages
- How many times each components is used
- ... and where
- What assets are used
- Group results by brand/channel/section etc.

# Custom AEM reporting tool

- Full dev/test/release cycle
- Answers only specific questions
- Gets complicated quickly
- Re-inventing the wheel

# Better tools available



- **Spark** – Analytics Engine
- **Zeppelin** – Interactive analytics notebook
- Allow to interactively query unstructured data with SQL

- Traverse JCR and collect eg.
  - resourceType
  - ReplicationStatus
  - fileReference
  - lastModified/Created
- Export as JSON-L or CSV file

## ■ AbstractResourceVisitor

```
@Override
protected void visit(Resource r) {
    //...
    String rs = resource.getValueMap().get("sling:resourceType");
    JsonObject outputJson = new JsonObject();
    if (isNotBlank(rs)
        outputJson.addProperty("path", resource.getPath());
        outputJson.addProperty("resourceType", rs)
    //... - add further properties to output and write to Response object
}
```

# Traverse content path (2/2)

- Call from servlet
- Pass response object, to stream data
- Provide a starting point

```
@Override
protected void doGet(SlingHttpServletRequest request,
SlingHttpServletResponse response) throws IOException {
    MyVisitor visitor = new MyVisitor(response);
    visitor.accept("/content/wknd");
}
```

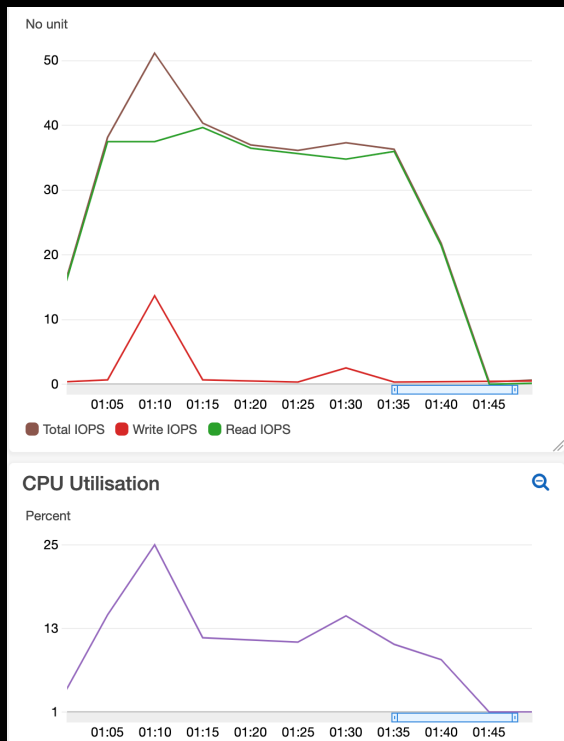
# Example output - Component

```
{
  "path": "/content/wknd/ca/fr/jcr:content/root/container/container_1679842506/separator",
  "resourceType": "wknd/components/separator",
  "page": {
    "path": "/content/wknd/ca/fr",
    "template": "/conf/wknd/settings/wcm/templates/landing-page-template",
    "pageLastModified": "2020-07-09 17:06:13.0644",
    "isActivated": false,
    "lastReplicationAction": "",
    "lastPublished": ""
  },
  "component": {
    "path": "/apps/wknd/components/separator",
    "superType": "/apps/core/wcm/components/separator/v1/separator",
    "group": "WKND.Content",
    "title": "Separator"
  }
}
```



# Example output - Asset

```
{
  "path": "/content/wknd/it/it/jcr:content/root/container/item_1571954853062",
  "fileReference": "/content/dam/wknd/en/adventures/riverside-camping-australia/AdobeStock_216674449.jpeg",
  "page": {
    "path": "/content/wknd/it/it",
    "template": "/conf/wknd/settings/wcm/templates/landing-page-template",
    "pageLastModified": "2020-07-09 16:56:55.0586",
    "isActivated": true,
    "lastReplicationAction": "Activate",
    "lastPublished": "2020-07-09 16:59:49.1629"
  }
}
```



- m5a.xlarge EC2 instance
- gp2 ssd volume (1000 IOPS)
- 10m Pages (30GB)
- 38 minutes extraction time
- No significant impact

## ■ Run with docker

```
docker run -p 8080:8080 \           # expose port 8080
-v $PWD/notebook:/zeppelin/notebook \ # share folder with notebooks
-v $PWD/local-repo:/zeppelin/local-repo \ # share folder with data
--rm \                               # remove container when stopped
apache/zeppelin:0.7.3               # name of image to use
```

## ■ Accessible on <http://localhost:8080>

# DEMO

# Benefits

- No need to go through SDLC every time
- Very low AEM impact
- Allows to dive deeper into data
- Compare data over time
- Only considers data from repository

# Q&A



# Get in touch

- [linkedin.com/in/michalgoszczynski/](https://www.linkedin.com/in/michalgoszczynski/)
- Email: [goszczyn@adobe.com](mailto:goszczyn@adobe.com)
  
- Demo code:
  - <https://github.com/mgoszczynski/adaptto-2020>

# Appendix



# Demo outline

# Loading data to Zeppelin

- `Spark.read.json` – loads the data as `DataFrame` (table)
- `DF.union(otherDF)` – joins two dataframes
- `DF.createOrReplaceTempView("table_name")` – creates a table that can be used by Spark SQL

```
val wkndDF = spark.read.json("file:///zeppelin/local-repo/wknd-all-assets.json")
val weRetailDF = spark.read.json("file:///zeppelin/local-repo/we-retail-all-assets.json")
val mainDF = wkndDF.union(weRetailDF)
mainDF.createOrReplaceTempView("components")
```

FINISHED ▶ ⌘ ⌘ ⌘ ⌘

```
wkndDF: org.apache.spark.sql.DataFrame = [component: struct<group: string, path: string ... 2 more fields>, fileReference: string ... 3 more fields]
weRetailDF: org.apache.spark.sql.DataFrame = [component: struct<group: string, path: string ... 2 more fields>, fileReference: string ... 3 more fields]
mainDF: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [component: struct<group: string, path: string ... 2 more fields>, fileReference: string ... 3 more fields]
```

Took 1 sec. Last updated by anonymous at September 14 2020, 6:29:24 PM.

Zeppelin Notebook - Job Search your Notes anonymous

AdaptTo-2020

```
%sql
select count(distinct page.path) as page_count from components
```

FINISHED

page\_count

584

Took 2 sec. Last updated by anonymous at September 15 2020, 1:13:43 PM. (outdated)

```
%sql
select ifnull(component.path,resourceType) as component, count(ifnull(component.path,resourceType)) as usage from components group by
component.path,resourceType order by usage desc
```

FINISHED

component	usage
/libs/commerce/components/product/image	1275
/apps/weretail/components/structure/product	1273
weretail/components/structure/product/assets	1271
/libs/wcm/foundation/components/responsivegrid	1060
/apps/weretail/components/content/title	637
/apps/weretail/components/structure/page	400
/apps/wknd/components/container	393

1. Count pages
2. Count component occurrences

```
var compDF = mainDF
  .withColumn("brand", regexp_extract(col("path"), "A-Za-z0-9\\-\\_+\\|\\.\\*", 1))
  .withColumn("region", regexp_extract(col("path"), "A-Za-z0-9\\-\\_+\\|\\.\\*", 2))
  .withColumn("language", regexp_extract(col("path"), "A-Za-z0-9\\-\\_+\\|\\.\\*", 3))
compDF.createOrReplaceTempView("components");
```

```
%sql
select brand, region, language, count(distinct page.path) as no_of_pages from components
group by brand, region, language order by no_of_pages desc
```

brand	region	language	no_of_pages
we-retail	us	en	216
we-retail	ca	en	126
we-retail	language-masters	en	126
wknd	ca	en	24

1. Add brand, region and language as a second, third and fourth level page name
2. Count pages per above

Title	Path	Type	Size
 wknd-events.jpg	/content/dam/wknd-events/wknd-events.jpg	IMAGE	96.1 KB
 italy.png	/content/dam/we-retail-screens/italy.png	IMAGE	30.6 KB
 we-retail-instore-logo.png	/content/dam/we-retail-screens/we-retail-instore-logo.png	IMAGE	16.2 KB
 usa.png	/content/dam/we-retail-screens/usa.png	IMAGE	59.2 KB

- Out of the box
- Lists all the assets under path (eg. /content/dam)
- Exportable as CSV

```
spark.read.format("csv")  
  .option("header", "true")  
  .option("inferSchema", "true")  
  .load("file:///zeppelin/local-repo/adapitto-assets.csv")  
  .createOrReplaceTempView("assets")
```

FINISHED ▶ ✖ 📖 ⚙️

Took 0 sec. Last updated by anonymous at September 14 2020, 6:31:05 PM.

```
%sql  
select count(distinct PATH) from assets where type is not  
  null and PATH not in (select distinct fileReference  
  from components where fileReference is not null)
```

FINISHED ▶ ✖ 📖 ⚙️



count(DISTINCT PATH) ▼

143

- Load CSV
- Create Assets Table
- Find assets that are not referenced