
# Das SCAPE Projekt: Langzeitarchivierung und Skalierbarkeit Teil 2: Anwendungsfälle an der Nationalbibliothek

Dr. Sven Schlarb

Österreichische Nationalbibliothek

SCAPE ½ Informationstag
05. Mai 2014, Österreichische Nationalbibliothek.

# Relationale Datenbanksysteme (RDBMS, z.B. MySQL)

- Vorteil: Gute Lese-Performanz
  Nachteil: Schlechte Performanz im Schreiben zufälliger Werte
- Schlechte Performanz bei große Tabellen mit vielen Spalten
- Effiziente Lastverteilung bei voller Verfügbarkeit bedeutet hohen Verwaltungsaufwand
  → Sharding: Aufwändig und wartungsintensiv
- Vertikale Skalierung → Teuere Hardware
- ÖNB: 2011 Versuch der Web-Archiv-Indexerstellung mit MySQL an unzureichender Performanz gescheitert!

*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Hadoop/HBase

- Elastizität → Einfache Erweiterbarkeit
- Hoher Schreibdurchsatz
- Konsistenz → Prüfsummen in HBase
- Effiziente zufällige Lesevorgänge
- Hohe Verfügbarkeit und Fehlertoleranz
- Atomare Lese- und Schreibmethoden → Parallele Datenmodifikation ohne Sperrungen
- Effiziente Bereichsanfragen → Gutes Leseverhalten bei sequentiellen Daten

*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Kosten für Speicher

- Kostenvorteil von Hadoop, da
  - meist auf preiswerter Hardware
  - Ohne Bindung an einen bestimmten Hersteller
  - Basierend auf Open-Source-Software

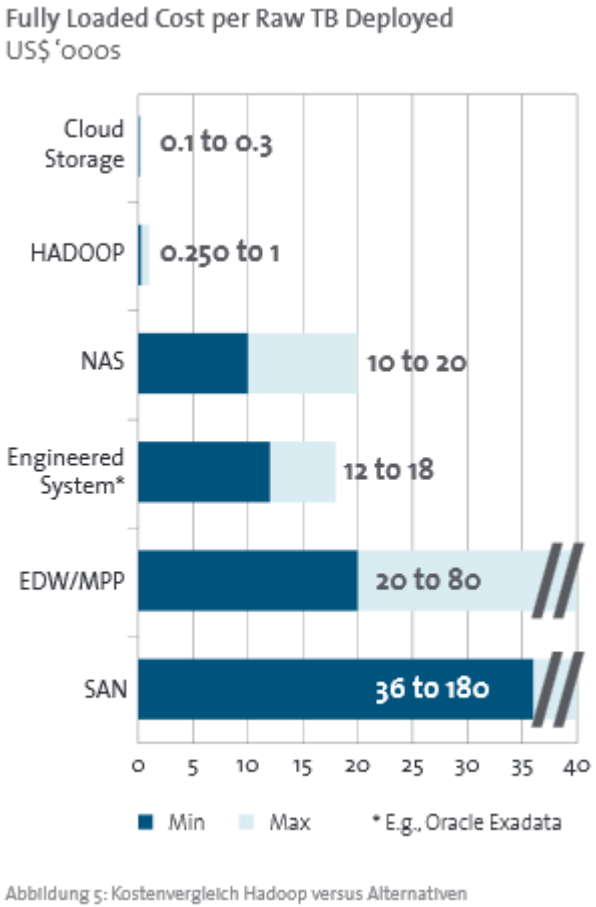



Abbildung 5: Kostenvergleich Hadoop versus Alternativen

Quelle: BITKOM Leitfaden Big-Data-Technologien-Wissen für Entscheider 2014, S. 39

*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Speichern und Verarbeiten getrennt

- Daten auf NAS-Speicher
- Müssen zur Verarbeitung in andere Server-Umgebung geladen werden
- Multi-Terabyte-Szenarien?

# Speichern und Verarbeiten zugleich

- Prozessoren bei den Daten
- Datenverarbeitung kann unmittelbar erfolgen

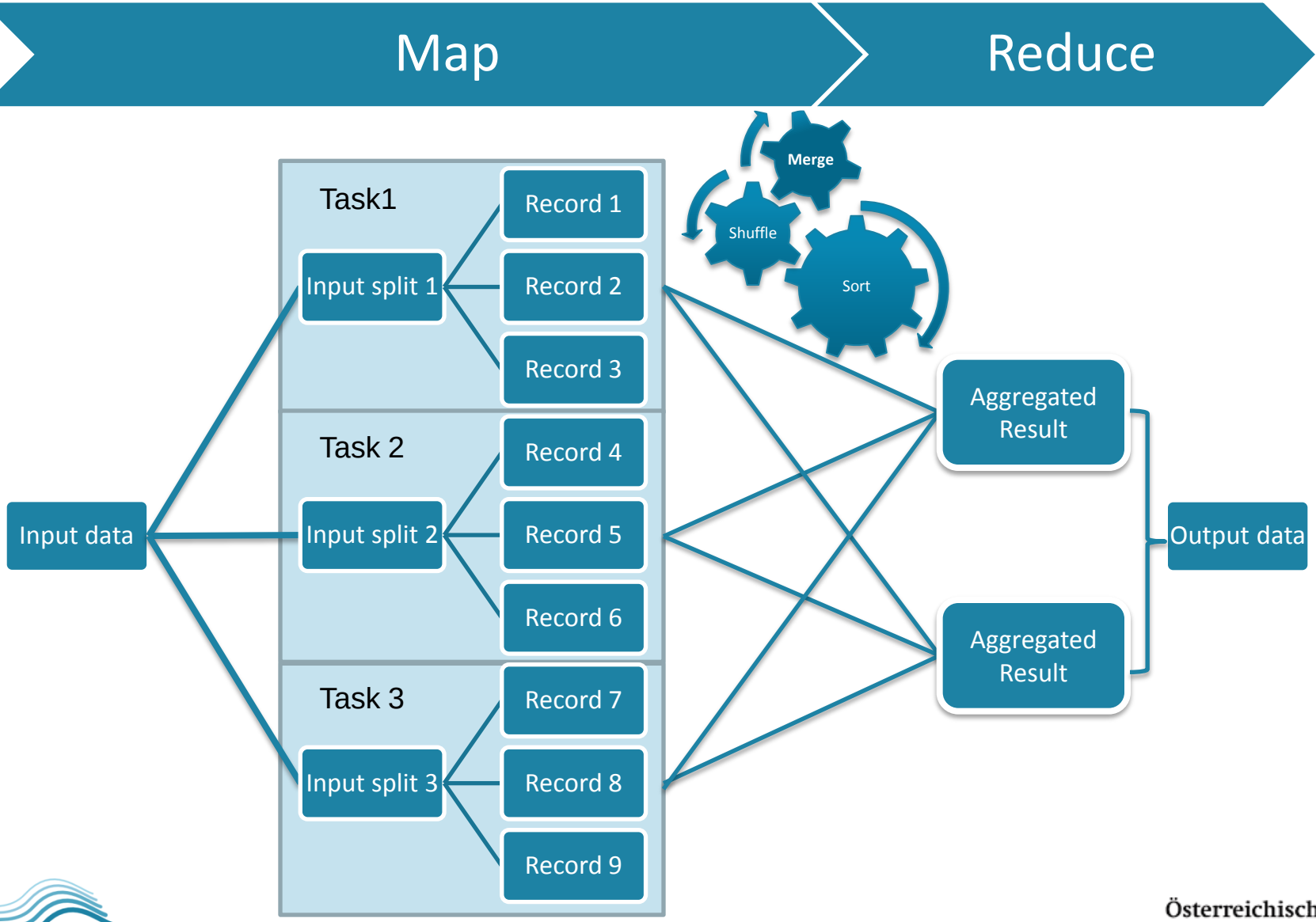
SCAPE

Österreichische Nationalbibliothek

# Einige Hadoop Grundprinzipien

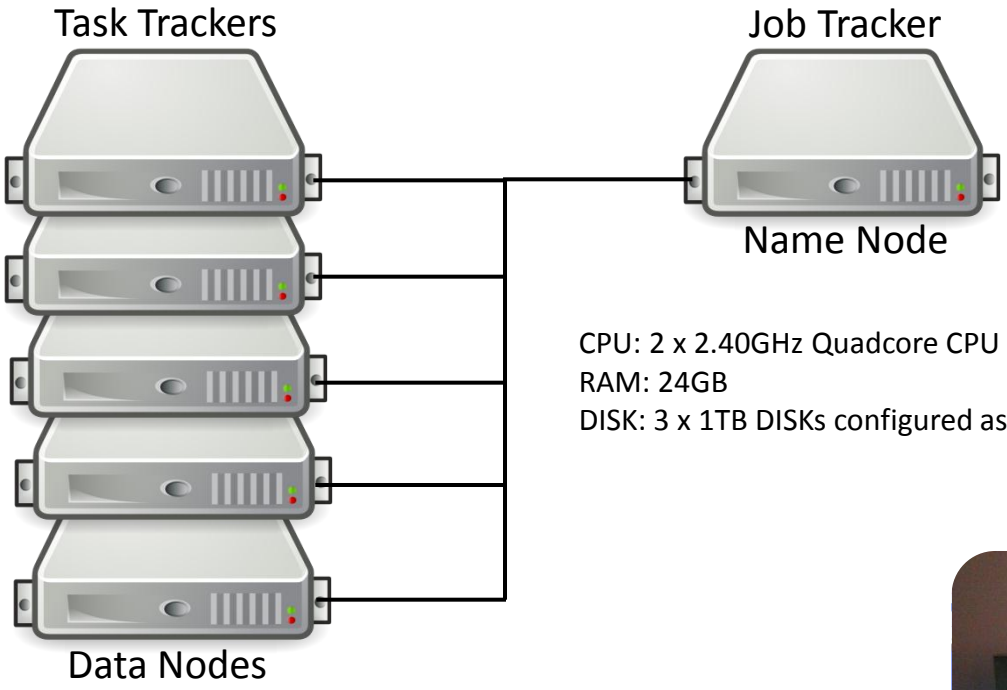
- Bei großen Datenmengen ist es meist einfacher die **verarbeitende Komponente zu den Daten** zu bringen als umgekehrt, die Daten zur verarbeitenden Komponente
- **Feingranulare Parallelisierung**: Die Ausführung der Datenverarbeitung findet auf den zur Verfügung stehenden Prozessorkernen statt
- **Hardware-bedingte Ausfälle** sind keine Besonderheit, sondern es gibt spezielle Vorkehrungen dafür
- **Redundanz**: Datenblöcke werden redundant gespeichert (Default: 3x) → Ausfallsicherheit, Flexibler Zugriff auf Daten
- **Daten-Lokalität**: Freier Knoten mit möglichst direktem Zugang zu Datenblock übernimmt die Verarbeitung

# MapReduce/Hadoop auf einen Blick




*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Experimenteller Computer-Cluster




CPU: 2 x 2.40GHz Quadcore CPU (16 HyperThreading cores)
RAM: 24GB
DISK: 3 x 1TB DISKs configured as RAID5 (Redundanz) – 2 TB effektiv

CPU: 1 x 2.53GHz Quadcore CPU (8 HyperThreading)
RAM: 16GB
DISK: 2 x 1TB DISKs konfiguriert als RAID0 (Performance) – 2 TB effektiv

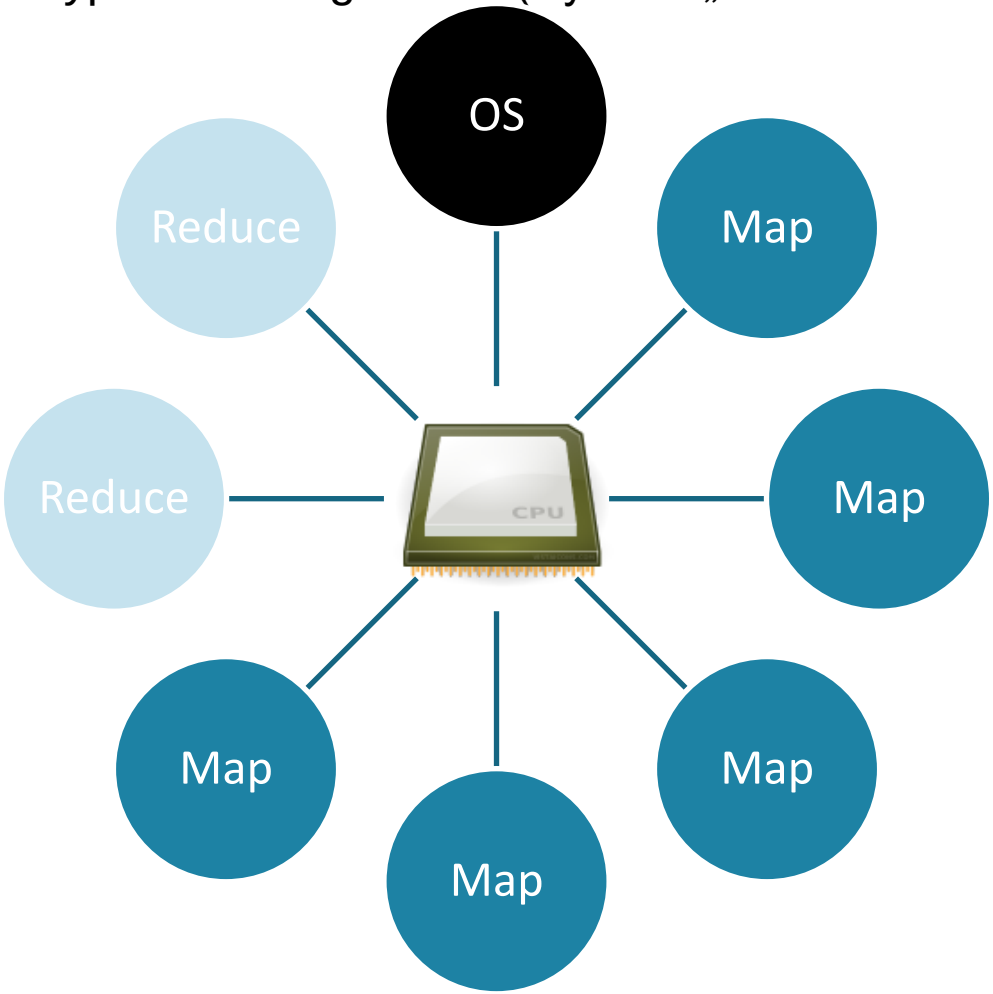
- Of 16 HT cores: 5 for Map; 2 for Reduce; 1 für Betriebssystem.

→ **25** Prozessorkerne für **Map** tasks
→ **10** Prozessorkerne für **Reduce** tasks

*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Die physische Sicht: Prozessor (CPU)

Beispielkonfiguration einer Quad-Core-CPU Pro Cluster-Knoten
4 physische Kerne
8 Hyperthreading-Kerne (System „sieht“ 8 Kerne)

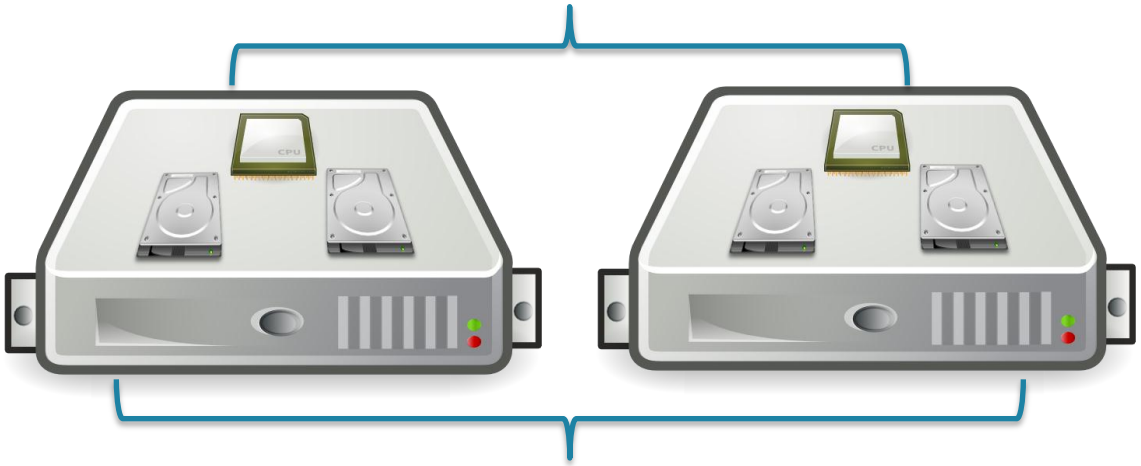



*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Die physische Sicht: Cluster-Knoten

## Verteilte Datenverarbeitung (MapReduce)

2 x Quad-Core-CPUs:
**10 Map(Parallelisierung)**
**4 Reduce (Aggregation)**



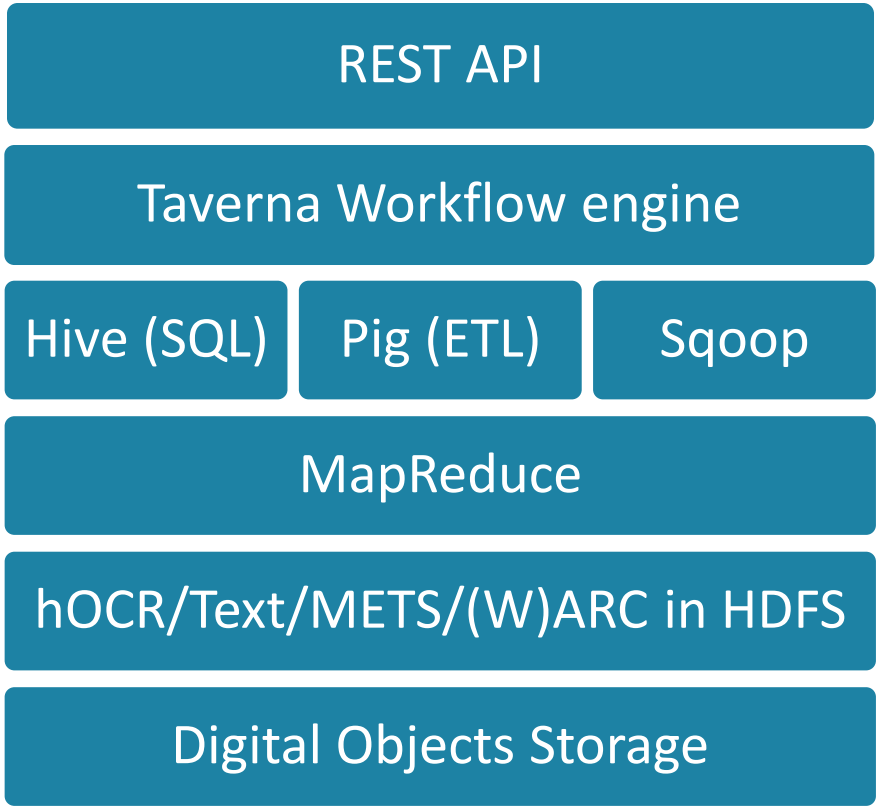
4 x 1 TB Festplatten bei Redundanz 3:
**1,33 TB effektiv** (rein rechnerisch)

## Verteilter Datenspeicher (HDFS)

## Hadoop = MapReduce + HDFS

*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Architektur




- Steuerung via REST API
- Workflow Engine für komplexe Jobausführung
- Hive als Abfrageschnittstelle – an (My)SQL-Sytax angelehnt
- MapReduce/Pig zum Laden und Transformieren v. Daten
- Sqoop für die DBMS-Integration
- „Kleine“ Objekte in HDFS
- „Große“ Objekte auf NAS

*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

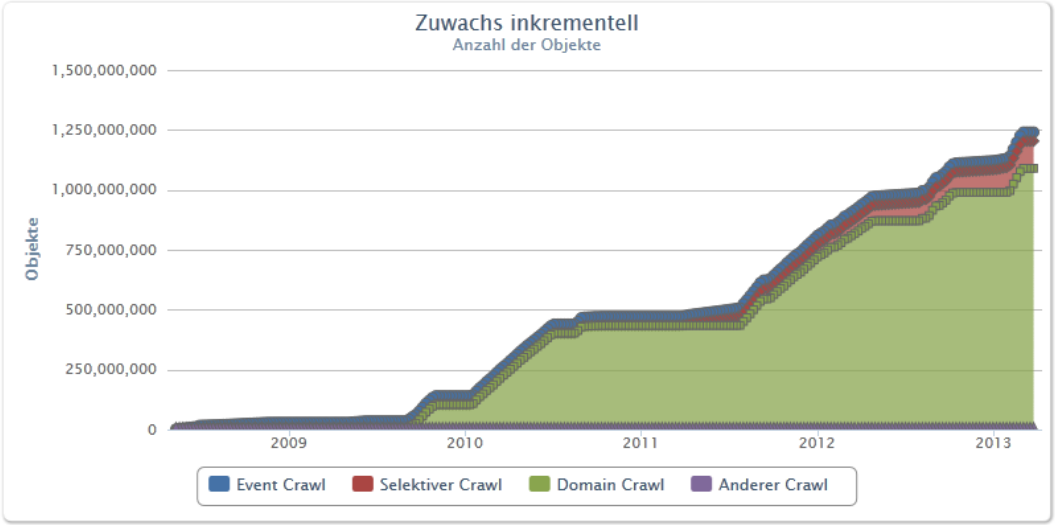
# Anwendungsfälle

- Web-Archivierung
  - Dateiformat-Identifikation im Web-Archiv
- Austrian Books Online
  - Bild-Datei-Format-Migration
  - Vergleich verschiedener Buch-Derivate
  - MapReduce in der Qualitätssicherung digitaler Bücher

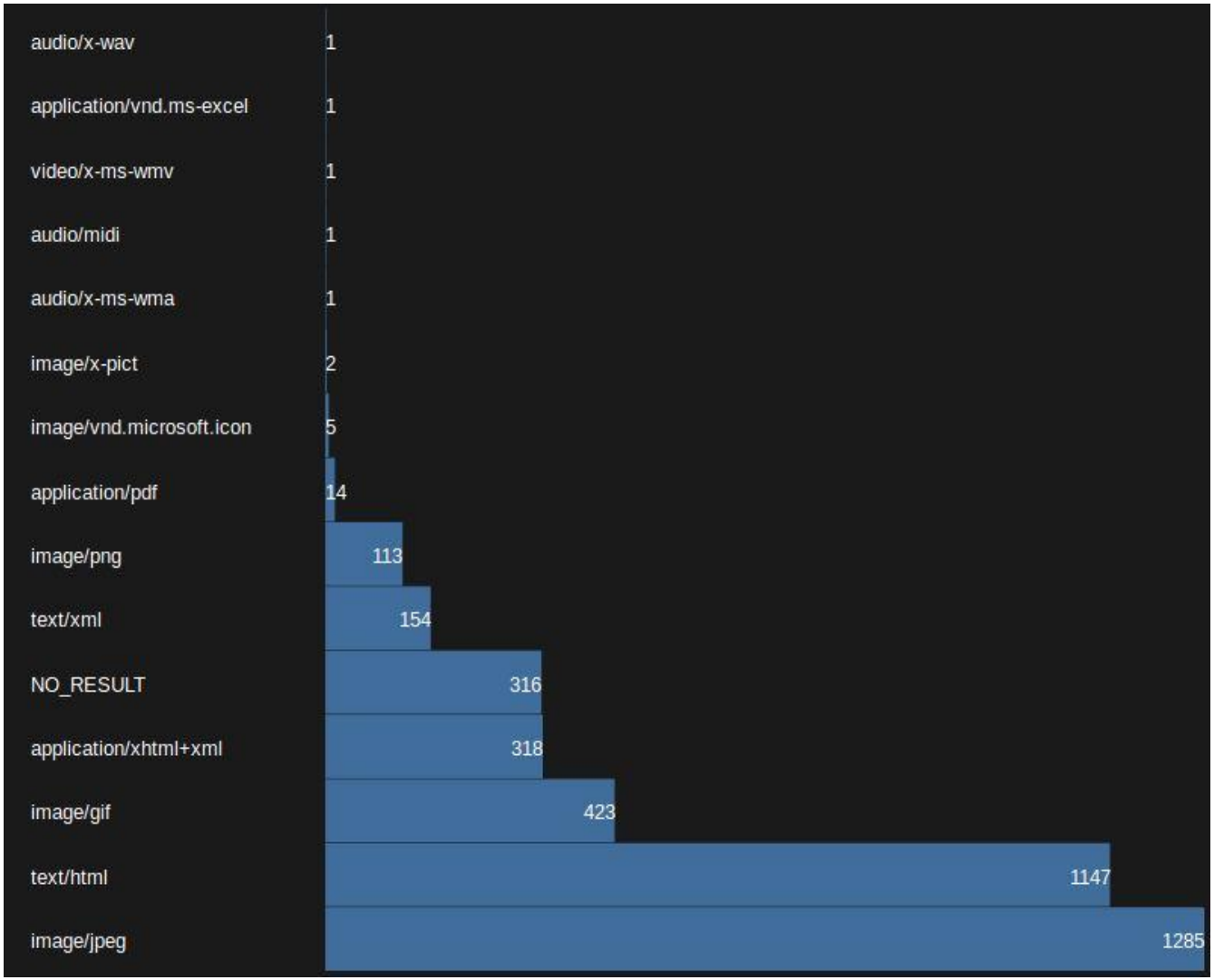
# Web-Archivierung

- Domain Harvesting
  - Gesamte **top-level-domain .at** alle 2 Jahre
- Selektives Harvesting
  - „Wichtige“ Websites die sich häufig ändern
- Event-basiertes Harvesting
  - Besondere Veranstaltungen und Events (z.B. Wahlen oder Sport-Events)

- Speicher: ca. 45TB
- ca. 1,7 Milliarden Objekte



# Dateiformat-Identifikation im Web-Archiv

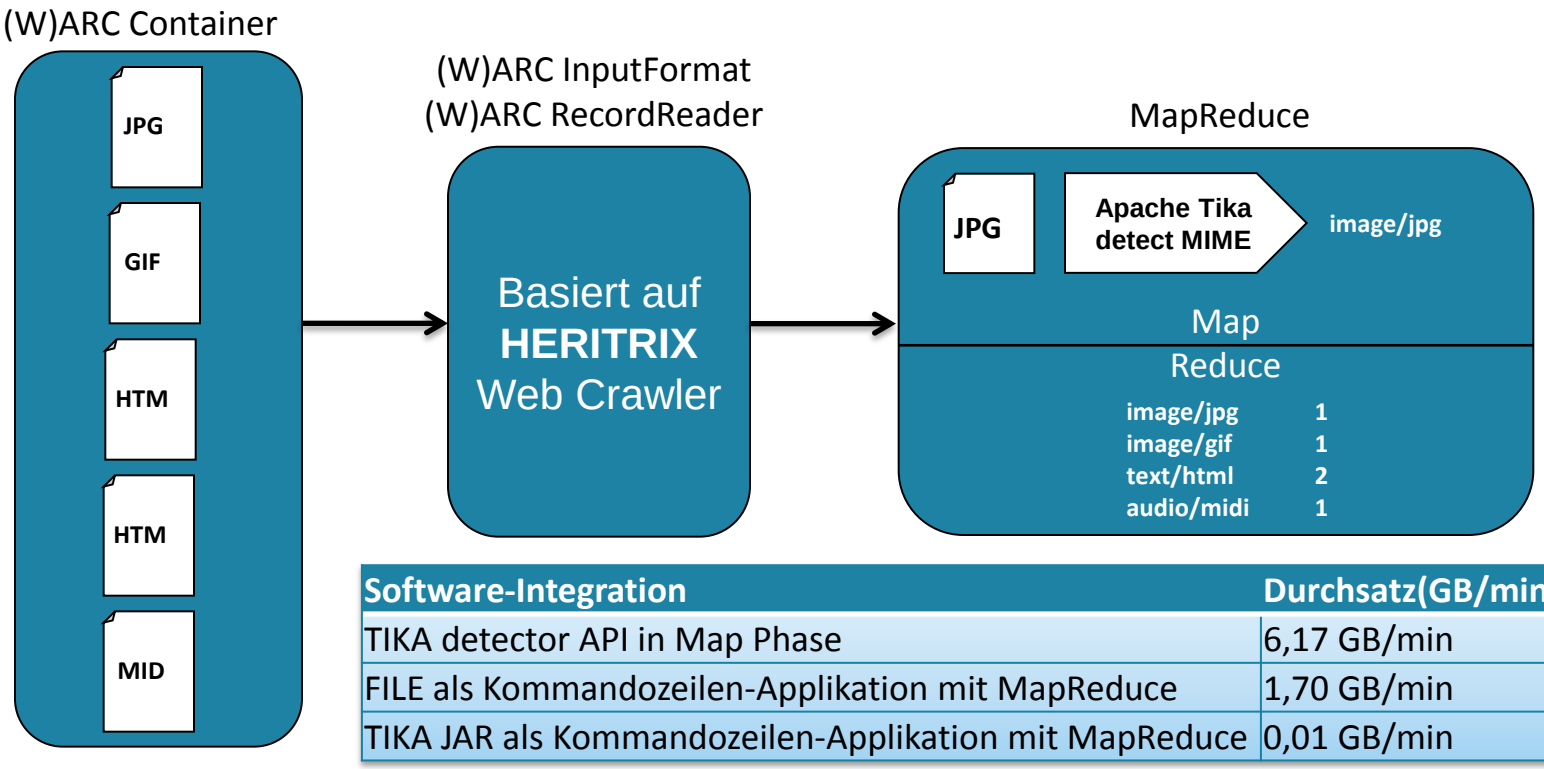



SEVENTH FRAMEWORK PROGRAMME

SCAPE

Österreichische Nationalbibliothek

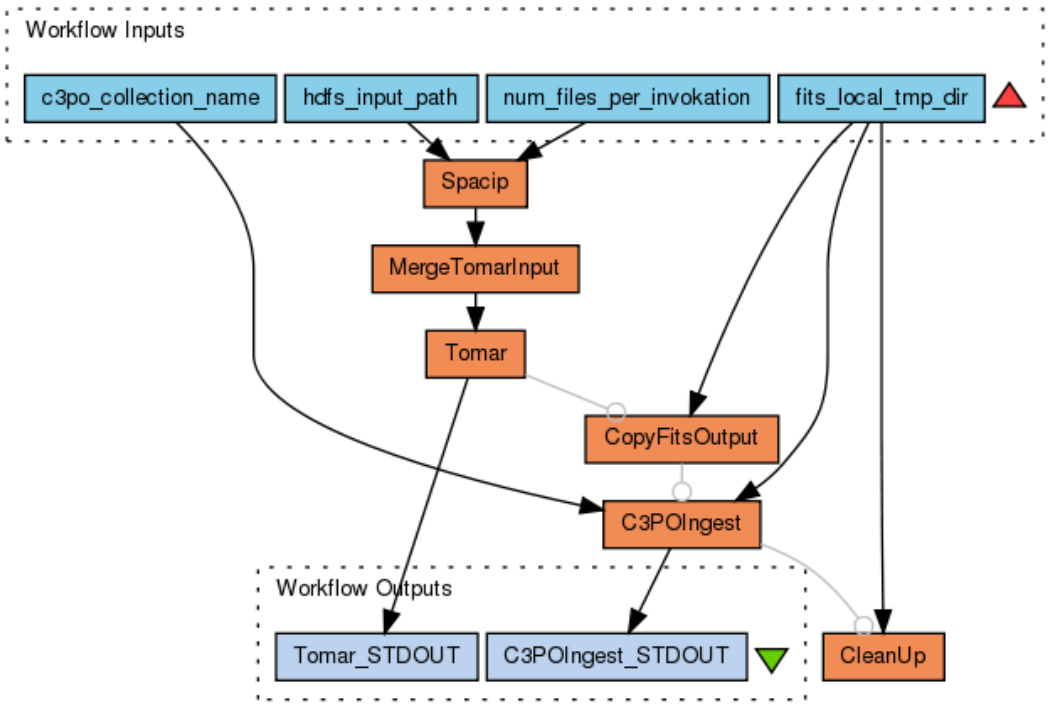
# Dateiformat-Identifikation im Web-Archiv

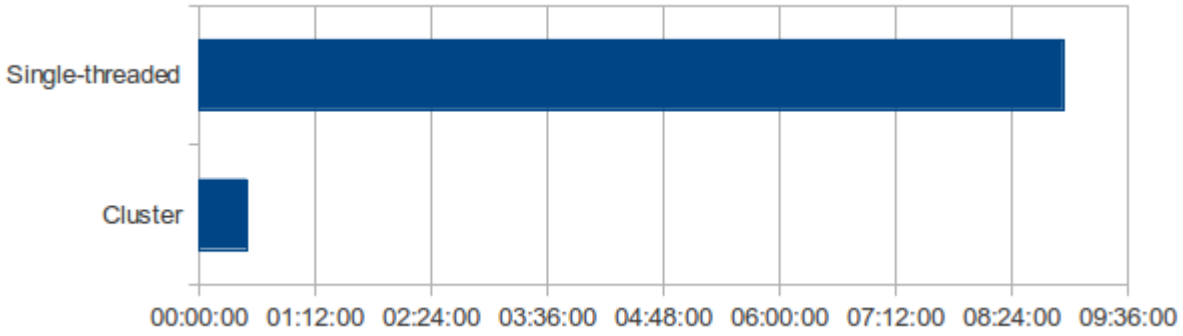



| Software-Integration | Durchsatz(GB/min) |
|---|---|
| TIKA detector API in Map Phase | 6,17 GB/min |
| FILE als Kommandozeilen-Applikation mit MapReduce | 1,70 GB/min |
| TIKA JAR als Kommandozeilen-Applikation mit MapReduce | 0,01 GB/min |

| Datenmenge | Anzahl der ARC-Dateien | Durchsatz(GB/min) |
|---|---|---|
| 1 GB | 10 x 100 MB | 1,57 GB/min |
| 2 GB | 20 x 100 MB | 2,5 GB/min |
| 10 GB | 100 x 100 MB | 3,06 GB/min |
| 20 GB | 200 x 100 MB | 3,40 GB/min |
| 100 GB | 1000 x 100 MB | 3,71 GB/min |

# Extraktion von Datei-Eigenschaften im Web-Archiv



# Extraktion von Datei-Eigenschaften im Web-Archiv

Screencast

SCAPE

Österreichische Nationalbibliothek

# Austrian Books Online

- Partnerschaft mit Google
- Ausschließlich gemeinfreie Werke
- Zielsetzung ~ 600.000 Bände
  - ~ 200 Mio. Seiten
- ~ 70 Projekt-Teilnehmer
  - 20+ im Kernteam
- Aktuell ~ 170.000 Bände verfügbar
  - ~ 50 Mio pages
  - ~ 18 Milliarden Token im Volltext, der mit Hilfe automatisierter Texterkennung (OCR) generiert wurde

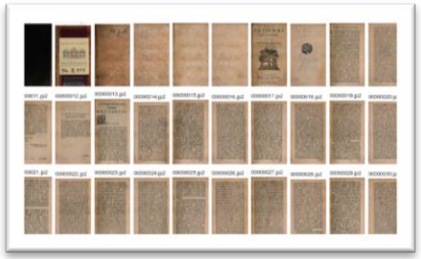
SCAPE

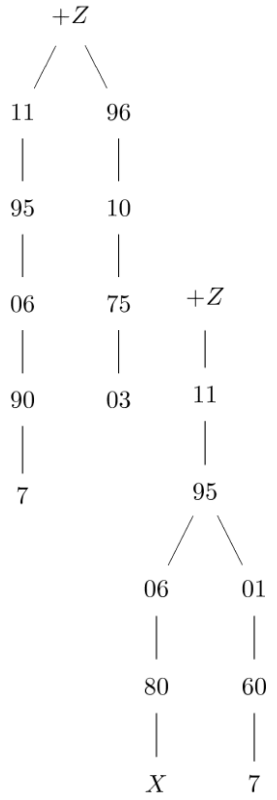
Österreichische Nationalbibliothek

# ADOCO (Austrian Books Online Download & Control)

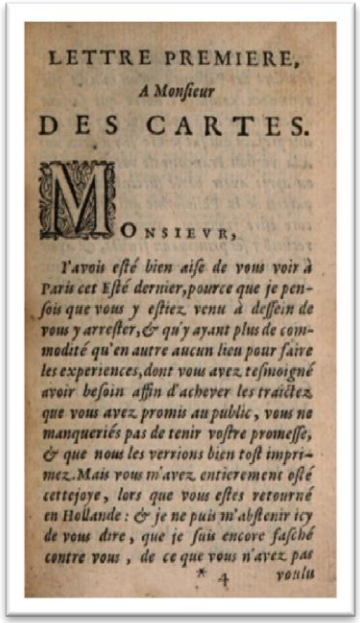
Digitalisierung → Download & Speicherung → Qualitäts-Kontrolle → Zugang

Google
Public Private Partnership






ADOCO

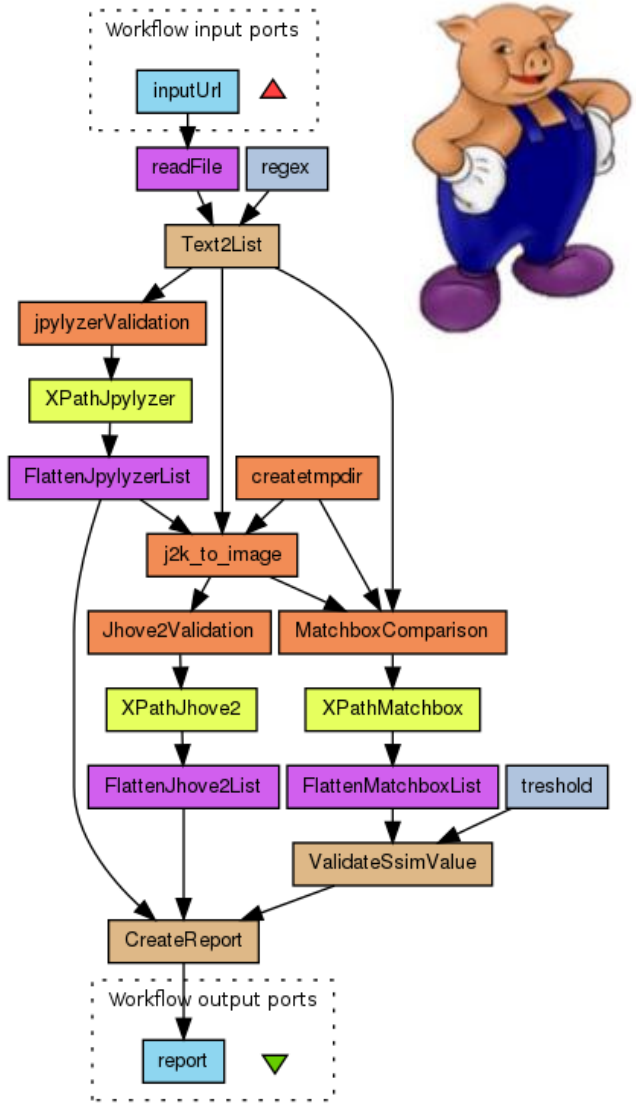

LETTRE PREMIERE,
A Monsieur
DES CARTES.
MONSIEVR,
J'avois esté bien aise de vous voir à Paris cet Esté dernier, pource que je pensois que vous y estiez venu à dessein de vous y arrester, & qu'y ayant plus de commodité qu'en autre aucun lieu pour faire les experiences, dont vous avez tesmoigné avoir besoin affin d'achever les traictez que vous avez promis au public, vous ne manqueriés pas de tenir vostre promesse, & que nous les verrions bien tost imprimez. Mais vous m'avez entierement osté cette joye, lors que vous estes retourné en Hollande: & je ne puis m'abstenir icy de vous dire, que je suis encore faschè contre vous, de ce que vous n'avez pas voulu

https://confluence.ucop.edu/display/Curation/PairTree


*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Dateiformat-Migration

- Bildformat-Migration
  - TIFF nach JPEG2000
    - Speicherbedarf reduzieren
  - JPEG2000 nach TIFF
    - Risiko: Dateiformat nicht mehr ausreichend verbreitet, unterstützt und verwendet
- Herausforderungen:
  - Integration verschiedener Werkzeuge der Datei-Format-Validierung, Migration und Qualitätssicherung
  - Qualitätssicherung rechenintensiv (abhängig von Methode)




SCAPE

Osterreichische Nationalbibliothek

# Vergleich verschiedener Digitaler Buchversionen

- Vergleich verschiedener Digitaler Buchversionen
  - Bilder wurden verändert
  - Bilder kommen von verschiedenen Scan-Quellen
- Herausforderungen:
  - Rechenintensiv
  - 170.000 Bände à ca. 400 Seiten
- SCAPE Software: Matchbox

# MapReduce in der Qualitätssicherung

- Verarbeitung von 60.000 Bänden, ~ 24 Millionen Seiten
- Taverna Tool service (Job-Ausführung per SSH)
- Verschiedene Hadoop-Jobs
  - Hadoop-Streaming-API
  - Hadoop Map/Reduce
  - Hive
- Workflow auf myExperiment verfügbar:
  http://www.myexperiment.org/workflows/3105
- Blogpost:
  http://www.openplanetsfoundation.org/blogs/2012-08-07-big-data-processing-chaining-hadoop-jobs-using-taverna

# Indikator für möglichen Textverlust

Korrekte Bildbeschneidung


N.° 100.

IL MESSAGGIERE TIROLESE

TRENTO Venerdì 13 Dicembre 1816.

VIENNA 28 Novembre.

Domenica 24, S. M. l'Imperatore e Re nella I. R. chiesa di corte si compiacque di coprire del berretto cardinalizio il principe arcivescovo d'Olmütz, conte di Traunmandorff, ed il principe vescovo di Gurk conte Salm. Gli II. RR. consiglieri intimi, i signori ciambellani e siniscalchi formarono l'accompagnamento in occasione di quest'atto solenne.

Attesa la morte del fu Re Federico di Wirtemberg, in preso per ordine supremo jeri mercoledì 27, il lutto a corte, che deve portarsi per quattro settimane.

Le notizie di Trieste recano che dall' 8 al 13 novembre giunsero in quel porto 41 grossi bastimenti mercantili, 25 austriaci, 7 inglesi, 3 turchi, 3 pontificj, 2 russo, 1 napolitano ed 1 francese. Il legno russo proveniente da Odessa e Costantinopoli aveva 1700 sacchi di grano; anche gli altri erano in parte carichi di commestibili.

( G. di V. )

PARIGI 28 Novembre.

Il 23 dopo la messa il sig. Canning venne ammesso all' udienza di S. M. accompagnato dall' ambasciator d' Inghilterra.

S. A. R. il duca di Cambridge è giunto il 28 a Stuttgard presso la Regina vedova sua sorella.

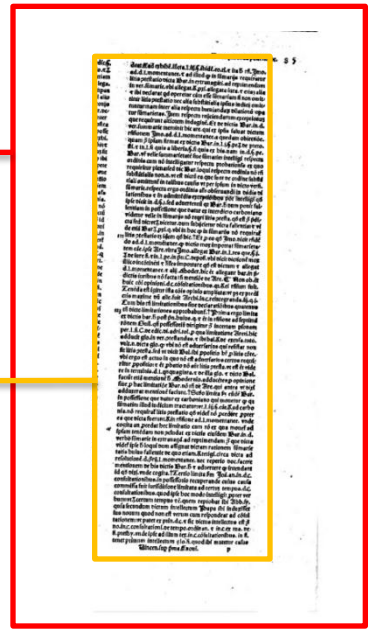
Il 26 alle ore una il sig. conte di Rostopschin, governatore di Mosca nel 1812, ebbe l'onore d'un' udienza da S. M., con cui si trattenne per mezz' ora.

A Tolosa la tranquillità è stata per alcuni giorni turbata. Il popolo ( istigato come assicurano i giornali francesi dai mal intenzionati ) tentò al mercato del grano del giorno 8 di costringere i venditori a calare il prezzo del grano dai 29 e 31 franchi per ogni ettolitro a quello di 24. Il commissario e gli agenti della polizia s'opposero a questa ingiusta pretensione; intanto il maire, i generali, i capi della guardia nazionale ed un distaccamento di questa stessa guardia, e delle truppe della guarnigione si radunarono sulla piazza del mercato per impedire ogni tassa arbitraria, che la plebe sconsigliata aveva in pensiero d' imporre, e le vendite continuarono al prezzo del mattino; ma un venditore che al lasciò lutinorire dalle grida, che risuonavano da ogni parte, avendo accettato il prezzo di 24 franchi, tutti gli altri furono obbligati a seguirne l' esempio.

Pel mercato dell' 11, si guarnì la piazza di numerose truppe, e tutto pareva tranquillo, quando una folla di popolo entrò gridando di bel nuovo: a 24 franchi. Si arrestarono subito i più turbolenti, ma il popolo faceva cenno di volerli liberare: allora i magistrati fecero avanzare un distaccamento di dragoni, i quali senza tirar un colpo sbarazzarono in un baleno la piazza e le strade che vi mettono capo. I sediziosi corsero al sobborgo di S. Stefano spargendo di trovarlo sprovveduto, ma alcuni distaccamenti di truppe dispersero anche quivi la folla. Nel giorno seguente 12, si presero tutte le precauzioni per ricondurre il buon ordine sul mercato del 13; all' alba del 13 tutta la guarnigione, tutta la guardia nazionale, ed i gendarmi vennero disposti in varj luoghi onde poter in un momento riunirsi, e ancorcersi a vicenda. Il mercato si fece quietamente. Il grano si vendeva 26 franchi, ed alcuni ectolitri di qualità inferiore si valutarono a 25, nessuno a 24. Ma in questo tempo gli abitanti del borgo di S. Cipriano avevano occupato il borgo aperto un mercato, ove vendevano a 24 franchi tre carra di grano prese per forza al padrone.

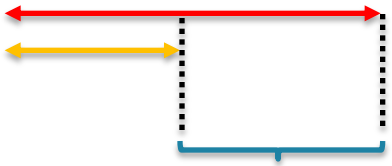
Un maggiore della guardia nazionale si pose alla testa di questa valorosa truppa ed entrò nel borgo, ove represse il tumulto. Un capo bisbiglio annunziava intanto, che sulla piazza stessa si sarebbero di bel nuovo attruppati i sediziosi. Di fatto verso le due sbucarono essi da tutte le strade sulla piazza armati di sassi. Il capo-squadrone dei

Fehlerhafte Bildbeschneidung



Bildbreite

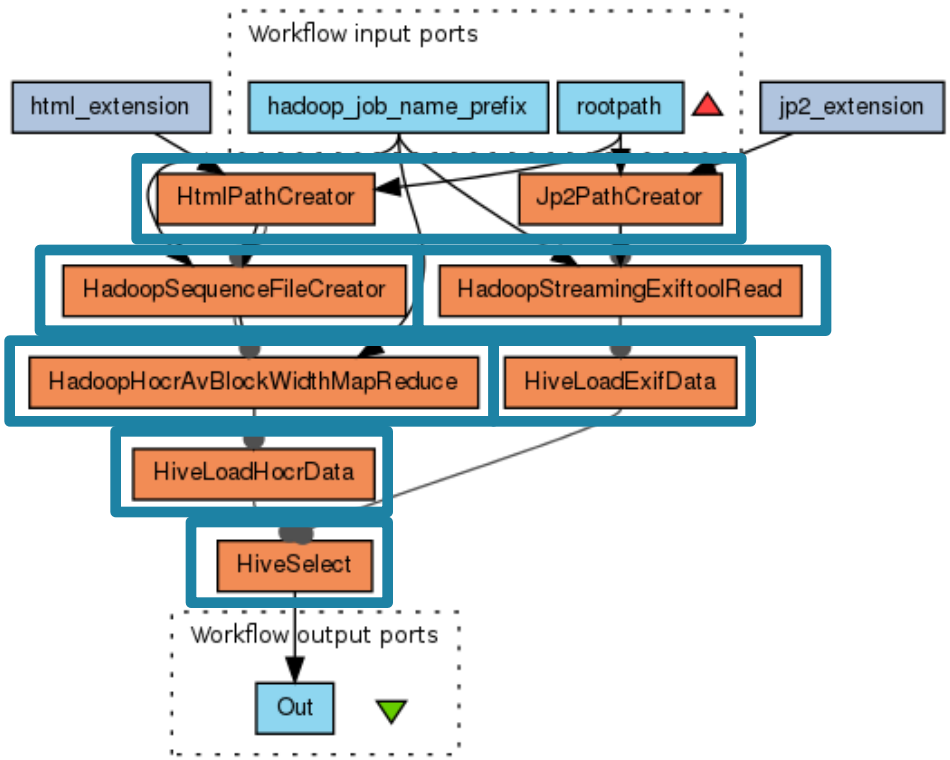
Blockbreite



Annahme: „Signifikante“ Unterschiede zwischen durchschnittlicher Blockbreite und Bildbreite sind Hinweise auf möglichen Textverlust wegen Bild-Beschneidungsfehler.

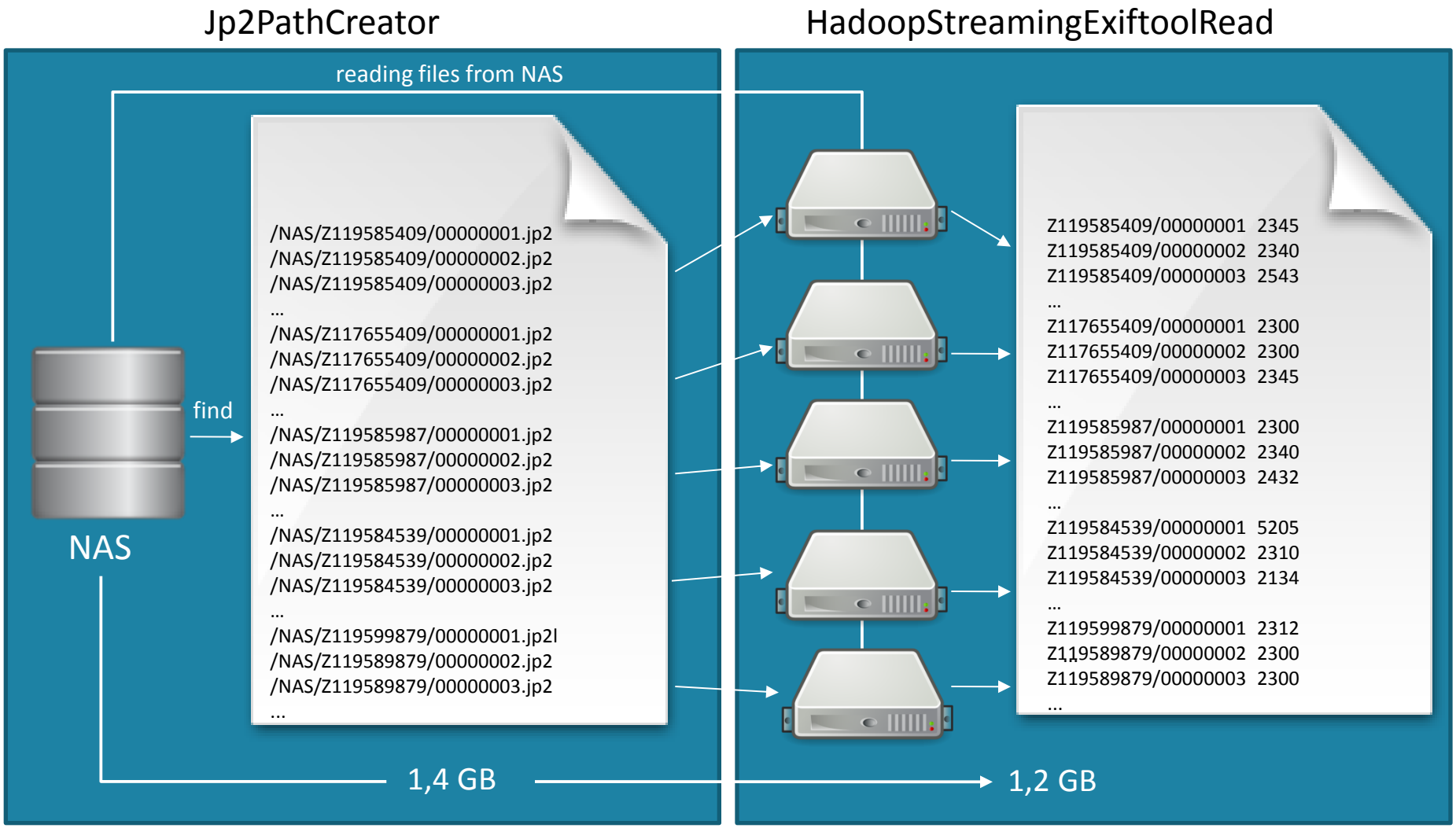
*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# MapReduce in der Qualitätssicherung



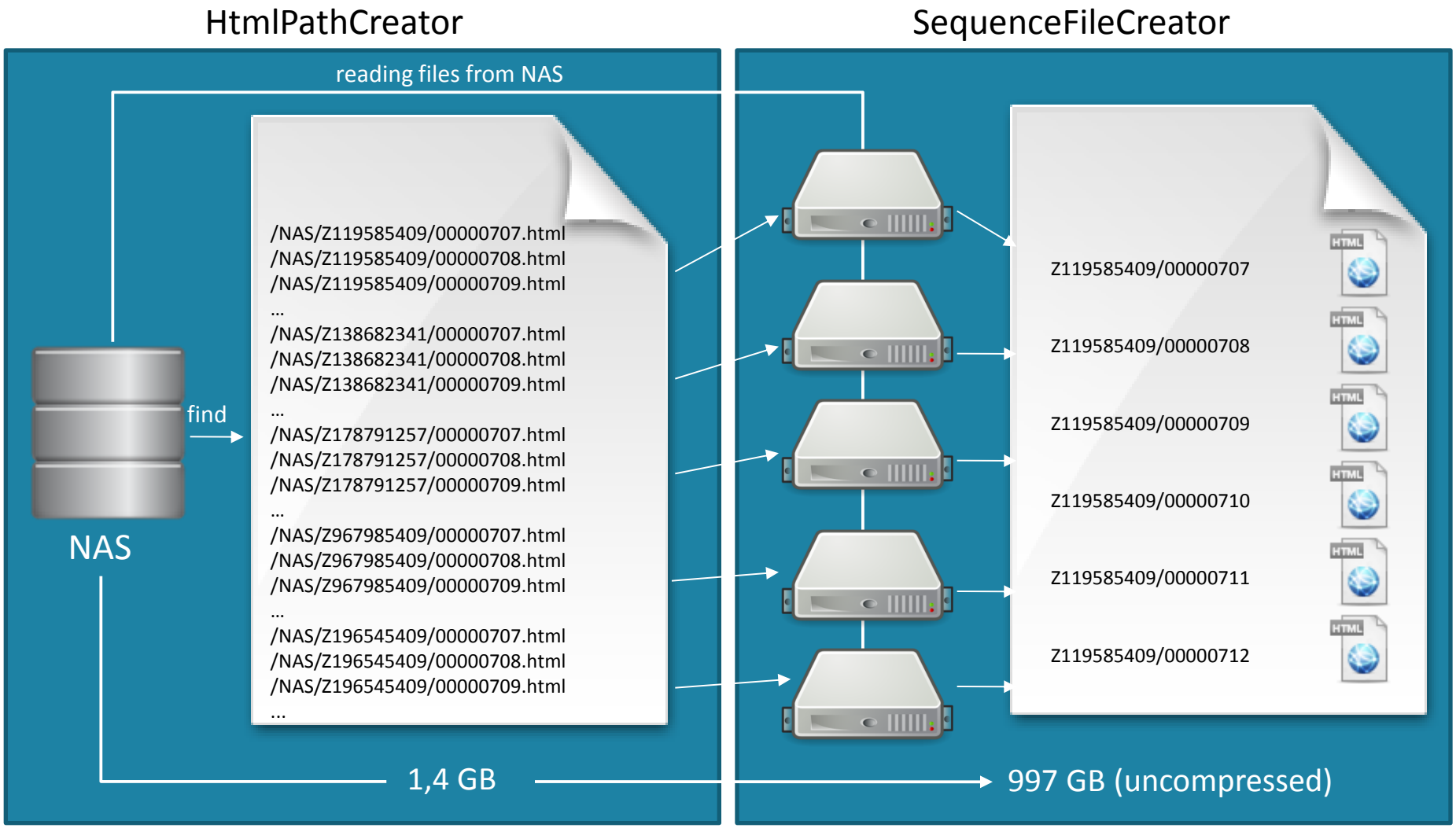

- Textdateien mit Pfaden zu Dateien erzeugen (JP2 & HTML)
- Bildmetadaten mit Exiftool lesen (Hadoop Streaming API)
- SequenceFile mit allen HTML Dateien erzeugen
- Durchschnittliche Blockbreite mit MapReduce berechnen
- Daten in Hive-Tabelle laden
- Test-Abfrage

# Lesen der Bild-Metadaten




60.000 Bücher(24 Million Seiten): ~ 5 h + ~ 38 h = **~ 43 h**

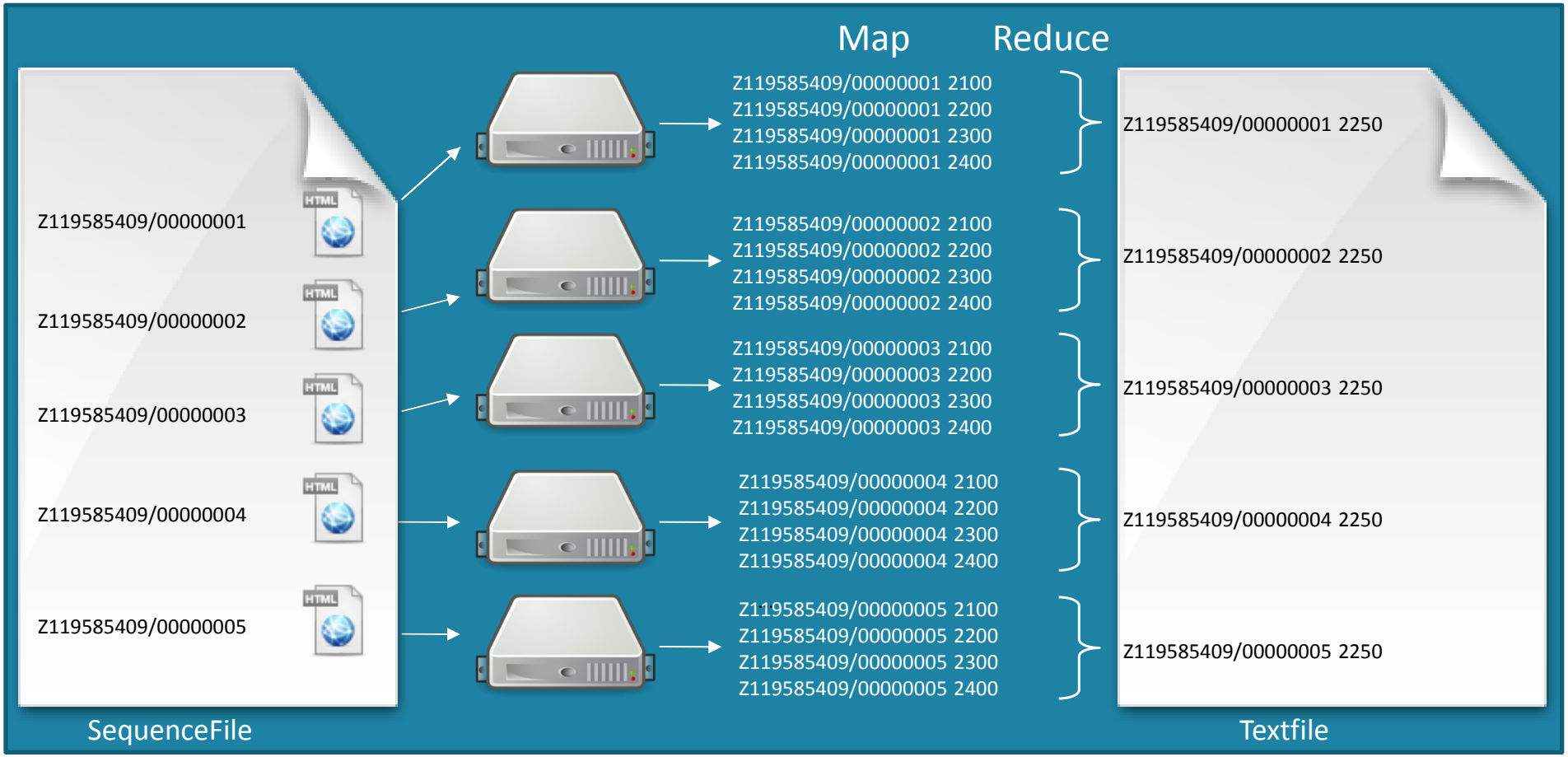
# Erzeugung eines SequenceFile




60.000 books (24 Million pages): ~ 5 h + ~ 24 h = **~ 29 h**

# Berechnung der durchschnittlichen Blockbreite

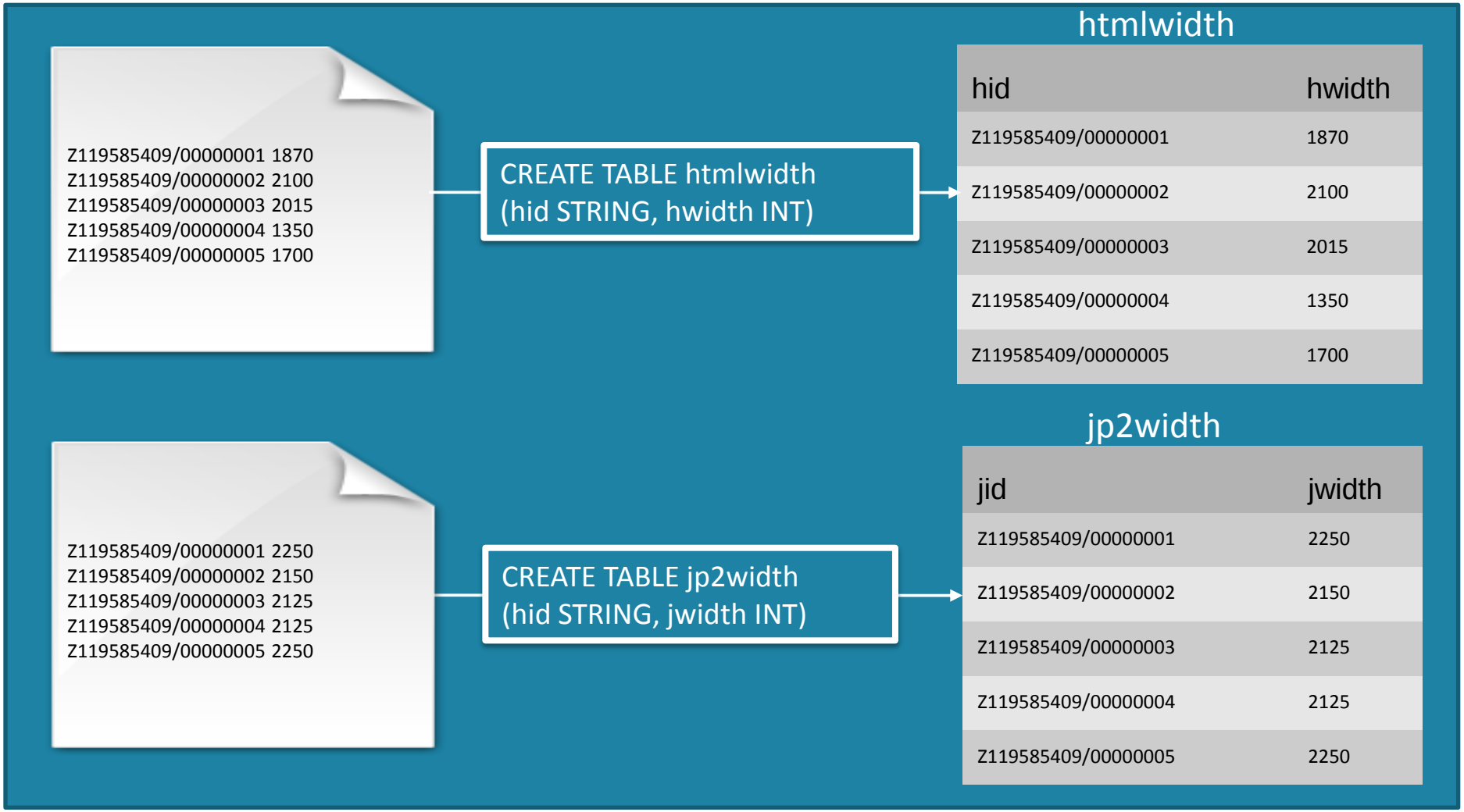
HadoopAvBlockWidthMapReduce




60.000 books (24 Million pages): **~ 6 h**

# Laden der Daten (Hive)

HiveLoadExifData & HiveLoadHocrData



# Abfragen (Hive)

HiveSelect

jp2width

| jid | jwidth |
|---|---|
| Z119585409/00000001 | 2250 |
| Z119585409/00000002 | 2150 |
| Z119585409/00000003 | 2125 |
| Z119585409/00000004 | 2125 |
| Z119585409/00000005 | 2250 |

htmlwidth

| hid | hwidth |
|---|---|
| Z119585409/00000001 | 1870 |
| Z119585409/00000002 | 2100 |
| Z119585409/00000003 | 2015 |
| Z119585409/00000004 | 1350 |
| Z119585409/00000005 | 1700 |

```
select jid, jwidth, hwidth from jp2width inner join htmlwidth on jid = hid
```

| jid | jwidth | hwidth |
|---|---|---|
| Z119585409/00000001 | 2250 | 1870 |
| Z119585409/00000002 | 2150 | 2100 |
| Z119585409/00000003 | 2125 | 2015 |
| Z119585409/00000004 | 2125 | 1350 |
| Z119585409/00000005 | 2250 | 1700 |

# Einige Erfahrungen auf den Punkt gebracht

- Für die digitalen Bibliothek bietet Apache Hadoop eine verlässliche Basis zum Aufbau einer kostengünstigen und skalierbaren IT-Infrastruktur
- Sorgfältige Auswahl der Werkzeuge aus dem Apache Ökosystem (weniger ist mehr)!
  - HBase, Hive, Pig, Oozie, etc.
- Mit neue Möglichkeiten kommen auch neue Fragestellungen
  - HDFS als Master- oder Bereitstellungsbereich?
  - Eigenen Cluster oder Infrastruktur mieten?
  - (Urheber-)Rechtliche Bedenken

*This work was partially supported by the SCAPE Project.*
*The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).*

# Danke für ihre Aufmerksamkeit!

## Weiterführende Informationen

- Projekt-Website: www.scape-project.eu
- Github-Repository: www.github.com/openplanets
- Projekt-Wiki: www.wiki.opf-labs.org/display/SP/Home

## Erwähnte SCAPE Werkzeuge

- SCAPE Plattform
  - http://www.scape-project.eu/publication/an-architectural-overview-of-the-scape-preservation-platform
- Jpylyzer – Jpeg2000 Validierung
  - http://www.openplanetsfoundation.org/software/jpylyzer
- Matchbox – Bildvergleich
  - https://github.com/openplanets/scape/tree/master/pc-qa-matchbox

SCAPE

Österreichische Nationalbibliothek