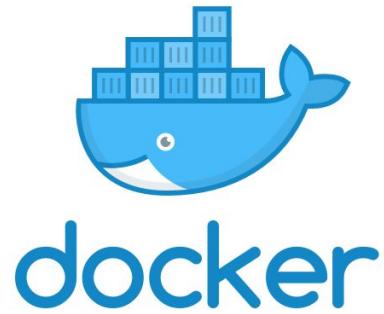


# Handling duplicates in Dockerfiles families: Learning from experts

**Mohamed A. Oumaziz**, Jean-Rémy Falleri, Xavier Blanc,  
Tegawendé F. Bissyandé and Jacques Klein

University of Bordeaux, France  
University of Luxembourg, Luxembourg

# Context



# Context

# Use an official Python runtime as a parent image

FROM python:2.7-slim

# Set the working directory to /app

WORKDIR /app

# Copy the current directory contents into the container at /app

COPY . /app

# Install any needed packages specified in requirements.txt

RUN pip install -r requirements.txt

# Make port 80 available to the world outside this container

EXPOSE 80

# Run app.py when the container launches

CMD ["python", "app.py"]

# Context

# Use an official Python runtime as a parent image

**FROM** python:2.7-slim

# Set the working directory to /app

**WORKDIR** /app

# Copy the current directory contents into the container at /app

**COPY** . /app

# Install any needed packages specified in requirements.txt

**RUN** pip install -r requirements.txt


# Make port 80 available to the world outside this container


**EXPOSE** 80


# Run app.py when the container launches


**CMD** ["python", "app.py"]


# Motivation


 [docker-library](#) / [python](#)


 Watch ▾


 84


 Star 959


 Fork 429

 Code















 Issues 12

 Pull requests 2

 Projects 0

 Insights

[python](#) / [Dockerfile](#)

>  <a href="#">2.7/jessie/Dockerfile</a>
 <a href="#">3.5/jessie/Dockerfile</a>
 <a href="#">3.6/jessie/Dockerfile</a>
 <a href="#">2.7/stretch/Dockerfile</a>
 <a href="#">3.5/stretch/Dockerfile</a>
 <a href="#">3.6/stretch/Dockerfile</a>
 <a href="#">3.7/stretch/Dockerfile</a>
 <a href="#">2.7/alpine3.8/Dockerfile</a>
 <a href="#">2.7/alpine3.9/Dockerfile</a>
 <a href="#">3.5/alpine3.8/Dockerfile</a>
 <a href="#">3.5/alpine3.9/Dockerfile</a>
 <a href="#">3.6/alpine3.8/Dockerfile</a>
 <a href="#">3.6/alpine3.9/Dockerfile</a>
 <a href="#">3.7/alpine3.8/Dockerfile</a>

# Motivation

## Update to 3.5.6, pip 19.0.3

master

docker-library-bot committed on 20 Feb

1 parent 9bb5d46    commit af2cf72d9c6c304d041c88db3e79242fd8c9ce60

Showing 6 changed files with 6 additions and 6 deletions.

Unified   Split

3.5/alpine3.8/Dockerfile

View file

@@ -102,7 +102,7 @@ RUN cd /usr/local/bin \	
102     && ln -s python3-config python-config	102     && ln -s python3-config python-config
103 +	103
104     # if this is called "PIP_VERSION", pip explodes with "ValueError: invalid truth value '<VERSION>'"	104     # if this is called "PIP_VERSION", pip explodes with "ValueError: invalid truth value '<VERSION>'"
105 -ENV PYTHON_PIP_VERSION 19.0.2	105 +ENV PYTHON_PIP_VERSION 19.0.3
106	106
107     RUN set -ex; \	107     RUN set -ex; \
108     \	108     \

3.5/alpine3.9/Dockerfile

View file

3.5/jessie/Dockerfile

View file

3.5/jessie/slim/Dockerfile

View file

3.5/stretch/Dockerfile

View file

3.5/stretch/slim/Dockerfile

View file

@@ -100,7 +100,7 @@ RUN cd /usr/local/bin \	
100     && ln -s python3-config python-config	100     && ln -s python3-config python-config
101	101
102     # if this is called "PIP_VERSION", pip explodes with "ValueError: invalid truth value '<VERSION>'"	102     # if this is called "PIP_VERSION", pip explodes with "ValueError: invalid truth value '<VERSION>'"
103 -ENV PYTHON_PIP_VERSION 19.0.2	103 +ENV PYTHON_PIP_VERSION 19.0.3

# Motivation

## Alpine: Ensure thread stack size is large enough (#211)

master (#248)

JayH5 authored and tianon committed on 18 Dec 2017

1 parent ee9df91 commit 8717dc2523c8093990cb958bb8c47ade6f851c05

Showing 8 changed files with 24 additions and 0 deletions.

Unified Split

View file

```
3 2.7/alpine3.4/Dockerfile
@@ -60,6 +60,9 @@ RUN set -ex \
    --enable-shared \
    --enable-unicode=ucs4 \
    && make -j "$(nproc)" \

63 && make install \
64 \
65 && runDeps="$( \

60 --enable-shared \
61 --enable-unicode=ucs4 \
62 && make -j "$(nproc)" \
63 + # set thread stack size to 1MB so we don't segfault before we hit sys.getrecursionlimit()
64 + # https://github.com/alpinelinux/aports/commit/2026e1259422d4e0cf92391ca2d3844356c649d0
65 + EXTRA_CFLAGS="--DTHREAD_STACK_SIZE=0x100000" \
66 && make install \
67 \
68 && runDeps="$( \
```

View file

```
3 2.7/alpine3.6/Dockerfile
@@ -60,6 +60,9 @@ RUN set -ex \
    --enable-shared \
    --enable-unicode=ucs4 \
    && make -j "$(nproc)" \

63 && make install \
64 \
65 && runDeps="$( \

60 --enable-shared \
61 --enable-unicode=ucs4 \
62 && make -j "$(nproc)" \
63 + # set thread stack size to 1MB so we don't segfault before we hit sys.getrecursionlimit()
64 + # https://github.com/alpinelinux/aports/commit/2026e1259422d4e0cf92391ca2d3844356c649d0
65 + EXTRA_CFLAGS="--DTHREAD_STACK_SIZE=0x100000" \
66 && make install \
67 \
68 && runDeps="$( \
```

View file

```
3 3.4/alpine3.4/Dockerfile
```

View file

# Motivation



Commits on Dec 22, 2015

More places that Python 3.4 needs the backport...

tianon committed on 22 Dec 2015



00c226b



Backport pip fix from 3.5 to 3.4

tianon committed on 22 Dec 2015



f9739c6



More places that Python 3.4 needs the backport...

[Browse files](#)

master (#1)

tianon committed on 22 Dec 2015

1 parent f9739c6 commit 00c226b82eee61c6c68adf813d9f7177d2efa52a

Showing 2 changed files with 2 additions and 2 deletions.

Unified Split

2 3.4/slim/Dockerfile

[View file](#)

@@ -47,7 +47,7 @@ RUN set -x \

```
47  && make -j$(nproc) \
48 +  && make install \
49  && ldconfig \
50 -  && pip3 install --no-cache-dir --upgrade pip==$PYTHON_PIP_VERSION \
51  && find /usr/local \
52      \(\ -type d -a -name test -o -name tests \) \
53      -o \(\ -type f -a -name '*.pyc' -o -name '*.pyo' \) \
```

```
47  && make -j$(nproc) \
48  && make install \
49  && ldconfig \
50 +  && pip3 install --no-cache-dir --upgrade --ignore-installed pip==$PYTHON_PIP_VERSION \
51  && find /usr/local \
52      \(\ -type d -a -name test -o -name tests \) \
53      -o \(\ -type f -a -name '*.pyc' -o -name '*.pyo' \) \
```

2 3.4/wheezy/Dockerfile


[View file](#)



# Research question

“ Do official projects maintain families of Dockerfiles,  
and why? ”


# RQ1 - Corpus




[Explore](#) [Pricing](#) [Sign In](#) [Sign Up](#)


[Docker EE](#) [Docker CE](#) [Containers](#) [Plugins](#)


Filters (1) [Clear All](#)


Docker Certified 

☐  Docker Certified

Images

☐ **Verified Publisher**   
Docker Certified And Verified Publisher Content

☒ **Official Images**   
Official Images Published By Docker

Categories 

☐ Analytics

☐ Application Frameworks

☐ Application Infrastructure

☐ Application Services

☐ Base Images

☐ Databases

☐ DevOps Tools

☐ Featured Images


☐ Messaging Services

☐ Monitoring

☐ Operating Systems

1 - 25 of 150 available images.

Most Popular



couchbase

Updated 15 minutes ago

Couchbase Server is a NoSQL document database with a distributed architecture.

Container


Linux

x86-64

Storage

Application Frameworks

OFFICIAL IMAGE




10M+

395

Downloads

Stars



redis

Updated 16 minutes ago

Redis is an open source key-value store that functions as a data structure server.

Container

Windows

Linux

IBM Z

PowerPC 64 LE

386


ARM 64

ARM

x86-64

Databases

OFFICIAL IMAGE




10M+

6.7K

Downloads

Stars



ubuntu

Updated 16 minutes ago

Ubuntu is a Debian-based Linux operating system based on free software.

Container

Linux

ARM 64

ARM

x86-64

IBM Z


PowerPC 64 LE

386

Base Images

Operating Systems

OFFICIAL IMAGE



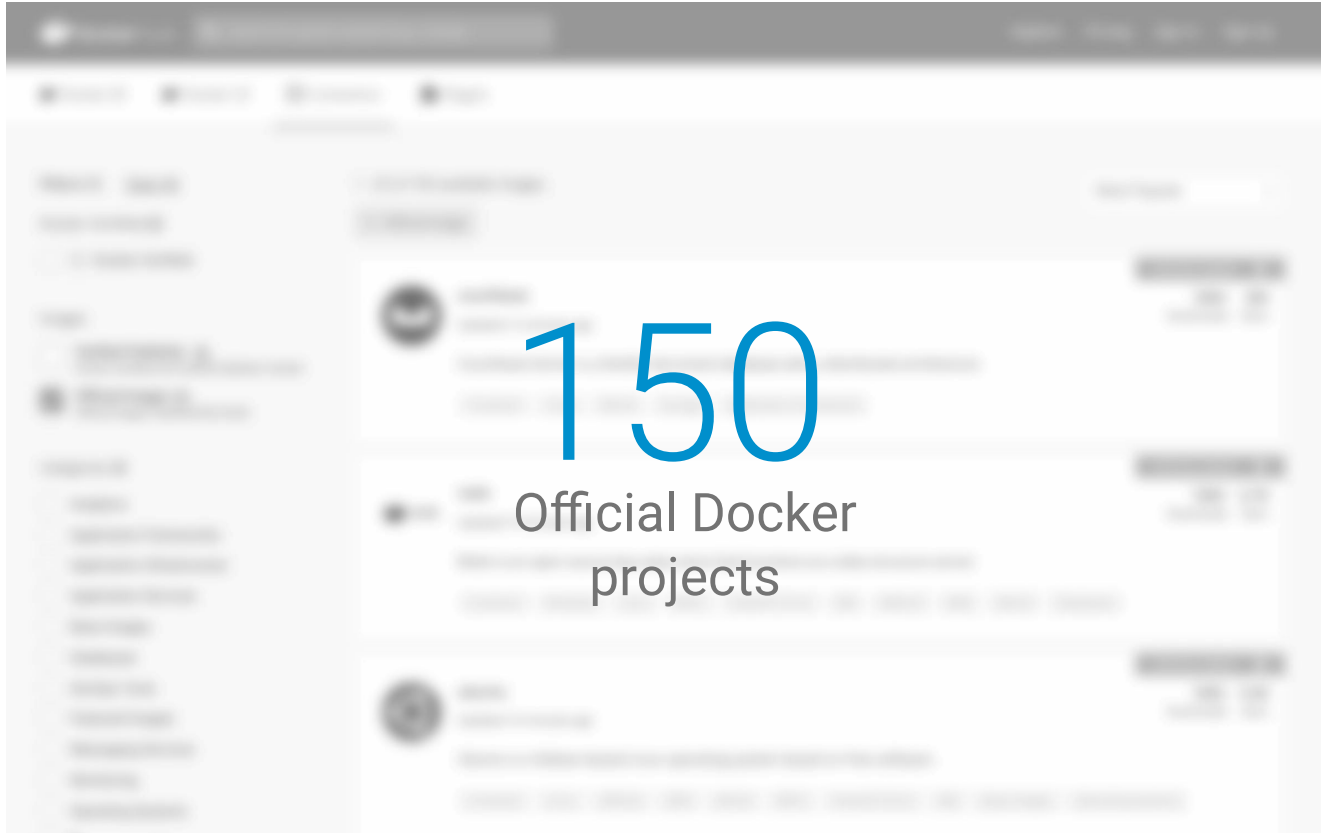
10M+

9.4K

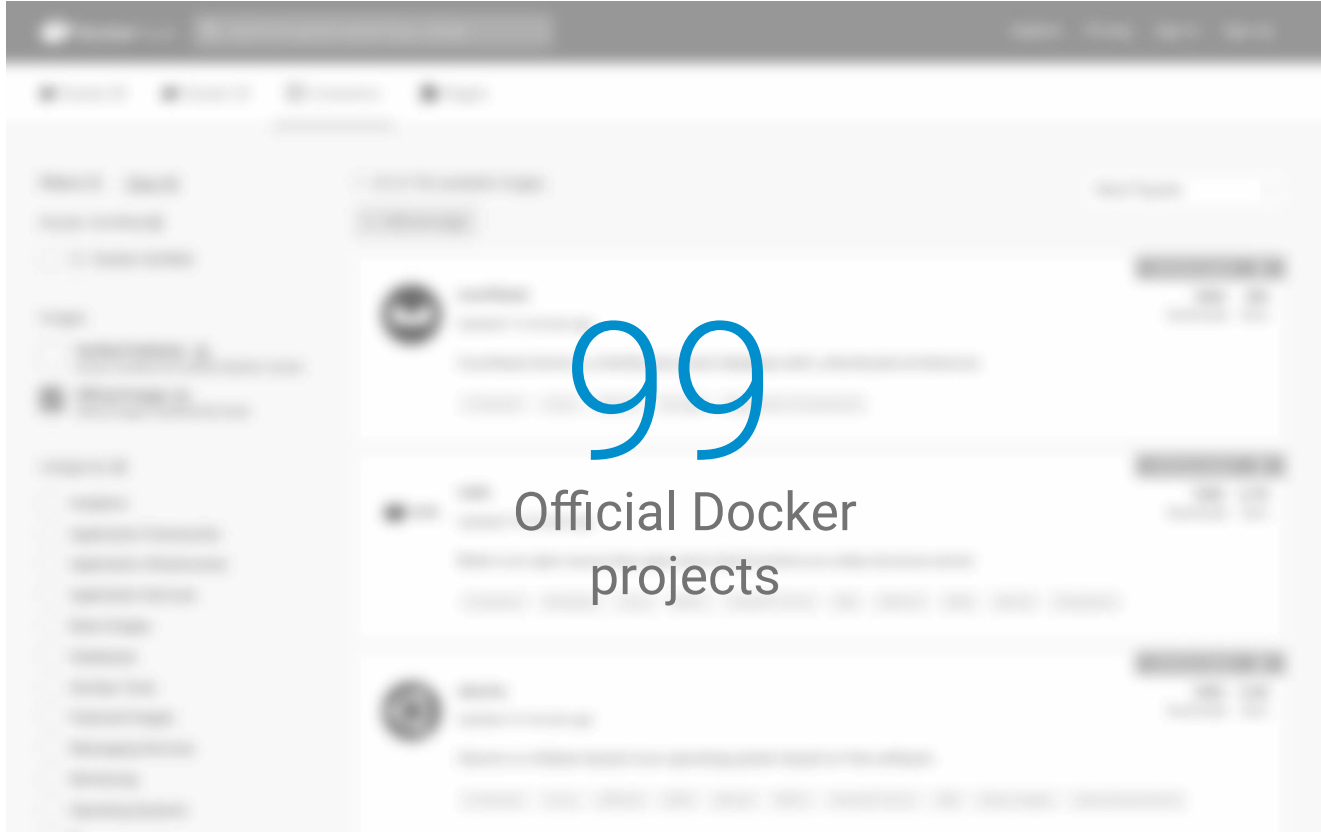
Downloads

Stars

# RQ1 - Corpus



# RQ1 - Corpus

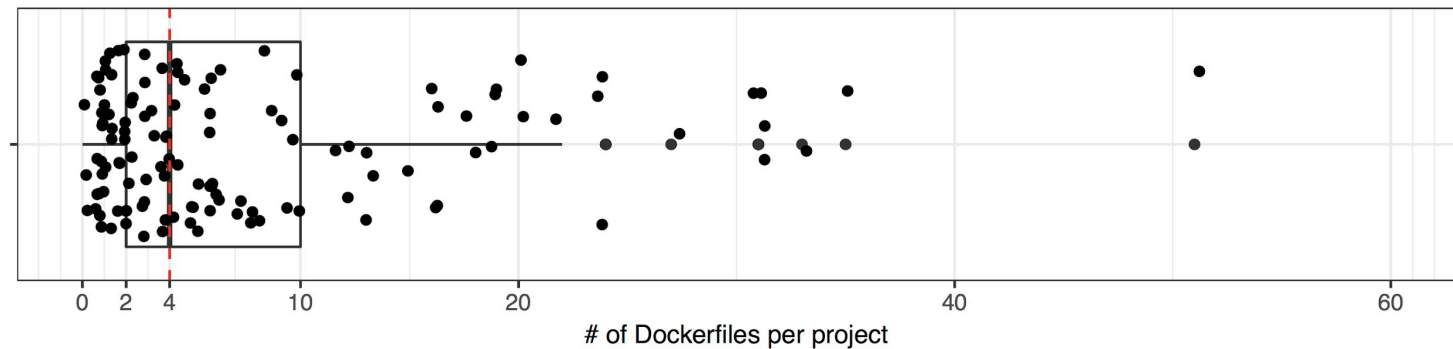


## RQ1 - Results

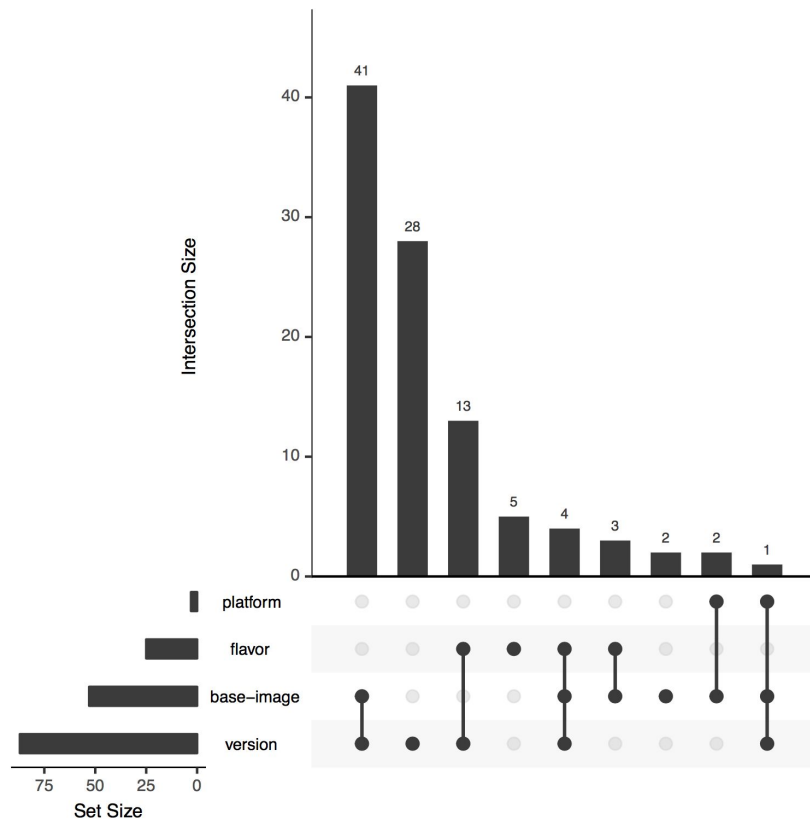
1300  
Dockerfiles

## RQ1 - Results

1300  
Dockerfiles



# RQ1 - Results



# Research question

“ Do duplicates arise in Dockerfiles families and why? ”



## RQ2 - Results

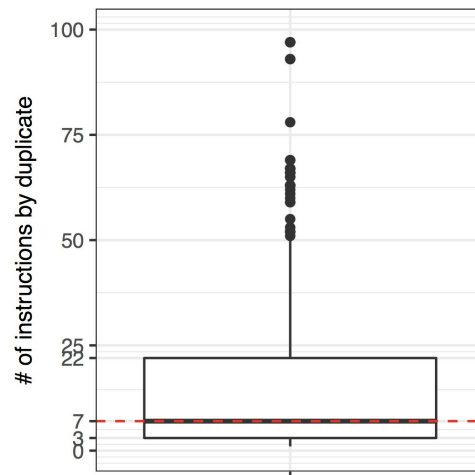
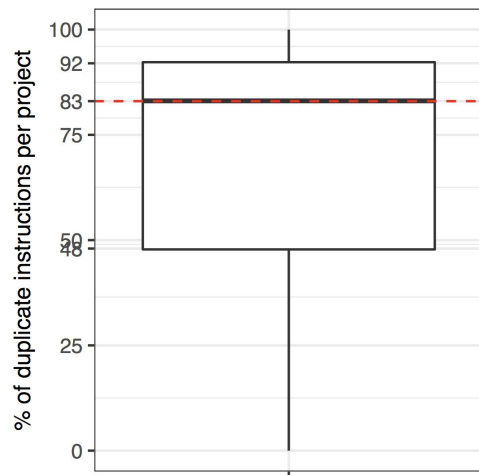
877

Duplicates

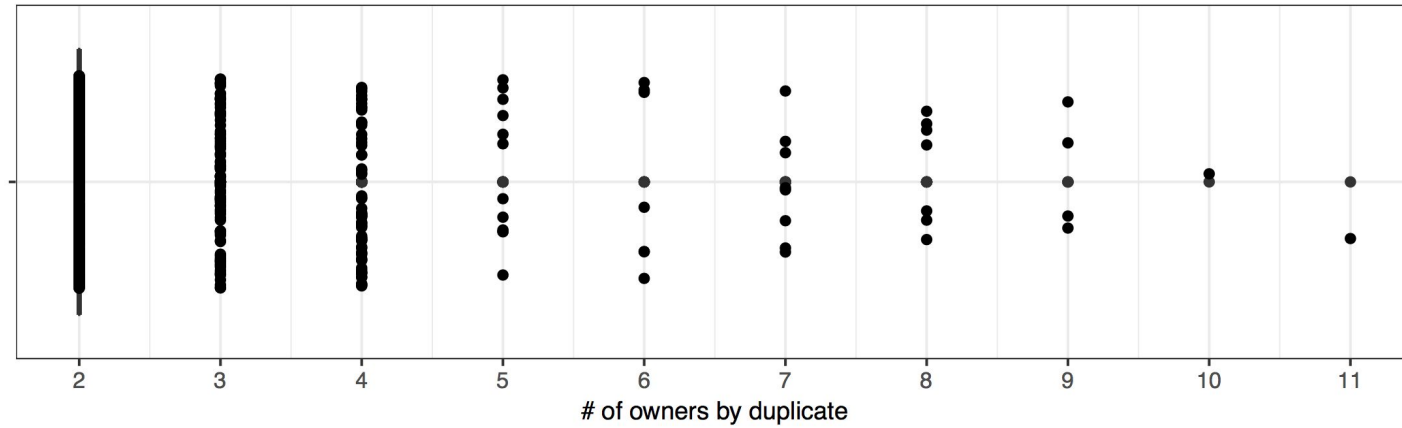
## RQ2 - Results

877

Duplicates



## RQ2 - Results



## RQ2 - Reasons

```
1  ENV XWIKI_VERSION=8.4.5
2  ENV XWIKI_URL_PREFIX "http://maven.xwiki.org/.../
   ${XWIKI_VERSION}"
3  ENV XWIKI_DOWNLOAD_SHA256 52ed122c44984748a729a78
   \
4  4c94cb70ccf0d2fa34c2340d0fd45c75deb3b0bc9
5  RUN rm -rf /usr/local/tomcat/webapps/* && \
6  mkdir -p /usr/local/tomcat/temp && \
7  mkdir -p /usr/local/xwiki/data && \
8  curl -fSL "${XWIKI_URL_PREFIX}/xwiki-ent-web-${
   XWIKI_VERSION}.war" -o xwiki.war \&& \
9  echo "$XWIKI_DOWNLOAD_SHA256 xwiki.war" |
   sha256sum -c - && \
10 unzip -d /usr/local/tomcat/webapps/ROOT xwiki.war
   && \
11 rm -f xwiki.war
```

Software installation  
and configuration

## RQ2 - Reasons

```
1  RUN apt-key adv --keyserver pgp.mit.edu \  
2  --recv-keys 1614552  
    E5765227AEC39EFCFA7E00EF33A8F2399  
3  RUN echo "deb http://download.rethinkdb.com/apt  
    xenial main" \  
4  > /etc/apt/sources.list.d/rethinkdb.list
```

## Dependency management

## RQ2 - Reasons

```
1  VOLUME /spiped
2  WORKDIR /spiped
3  COPY *.sh /usr/local/bin/
4  ENTRYPOINT ["docker-entrypoint.sh"]
5  CMD ["spiped"]
```

Runtime configuration

## RQ2 - Experts survey

“ 68% of experts have faced duplicates in the past. ”

“ 64% of experts who encountered duplicates  
don't need a tool to detect duplicates. ”

“ 94,1% of experts who encountered duplicates performed  
identical changes in a single commit in the past. ”

“ 75% of experts qualify duplicates as annoying or error-prone. ”

## Research question

“ What are the tools used by official Docker projects for managing duplicates? ”



## RQ3

```
1  sed -ri '  
2    s/^(ENV KIBANA_MAJOR) .*/\1 '$version' '/;  
3    s/^(ENV KIBANA_VERSION) .*/\1 '$fullVersion'  
      '/;  
4    s/^(ENV KIBANA_SHA1) .*/\1 '$sha1' '/;  
5  , '$version/Dockerfile'
```

Find and replace

## RQ3

```
1  RUN apt-get update && \  
2  apt-get --no-install-recommends -y install \  
3  curl \  
4  libreoffice \  
5  unzip \  
6  procps \  
7  <% if (db == 'mysql') print 'libmysql-java' \  
8  if (db == 'postgres') print 'libpostgresql-jdbc  
    -java' %> && \  
9  rm -rf /var/lib/apt/lists/*
```

## Template processors

## RQ3

```
1  for version in "${versions[@]}"; do
2      ...
3      if [ "$addSuite" ]; then
4          cat >> "$version/Dockerfile" <<-EOD
5              RUN echo 'deb http://deb.debian.org/debian
                      $addSuite main' > /etc/apt/sources.list.d
                      /$addSuite.list
6          EOD
7      fi
8  done
```

## Generators

## RQ3 - Experts survey

“ 56% of experts don't use tools because it's too much time consuming or difficult to implement. ”

“ 57% of experts not using tools would love to have a tool built for avoiding duplicates. ”

“ 63% of experts who are using tools are satisfied. ”

## Wrapping up

- Docker official projects maintain multiple Dockerfiles at the same time.
- The reasons for maintaining several Dockerfiles are to support multiple: versions, base-images, flavours and platforms
- Copy-paste of Dockerfile instructions is frequent.
- Three main reasons behind duplicates (Software installation and configuration, dependency management, runtime configuration).
- Three main tools are used to avoid duplicates (Find and replace, template processors and generators).
- Survey performed on 25 experts maintaining projects in our corpus.