ecsub のデータフロー

2018.4.16

Ai Okada

このドキュメントについて

- ecsub は AWS ECS を使用してバッチ処理を行うツールです。
- このドキュメントでは `ecsub submit` コマンド実行時のデータ の流れを解説します。

リポジトリ: <u>https://github.com/aokad/ecsub</u>

ecsub data flow 1 - 初期状態

local directory

--tasks ~/examples/taskswordcount.tsv

tasks.tsv

tasksファイルです。 入出力ファイルのパスや環境変数 が記載されています。 local directory or s3 bucket

--script ~/examples/wordcount.sh

word count. sh

s3 bucket

--aws-s3-bucket
s3://bucket

temporary 用 S3バケットです。

--wdir /tmp/edsub

temporary 用ローカルディレクトリです。中間ファイルやログが出力されます。

tasksファイルに記載された入力 ファイルです。**S3**バケットに配置 します。

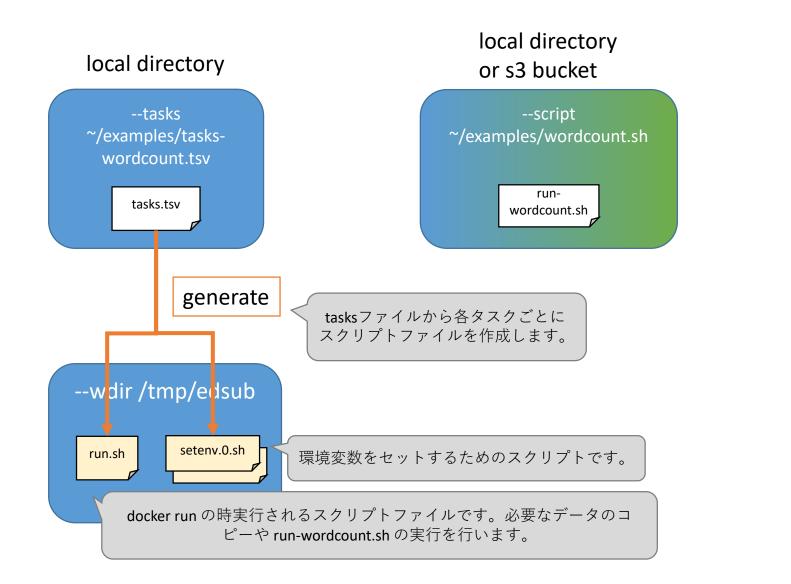
実際の解析を行うスクリプトです。

tasksファイルに記載された出力 ファイルです。 INPUT_FILE

Hamlet.txt

OUTPUT FILE

ecsub data flow 2 - スクリプト作成



s3 bucket

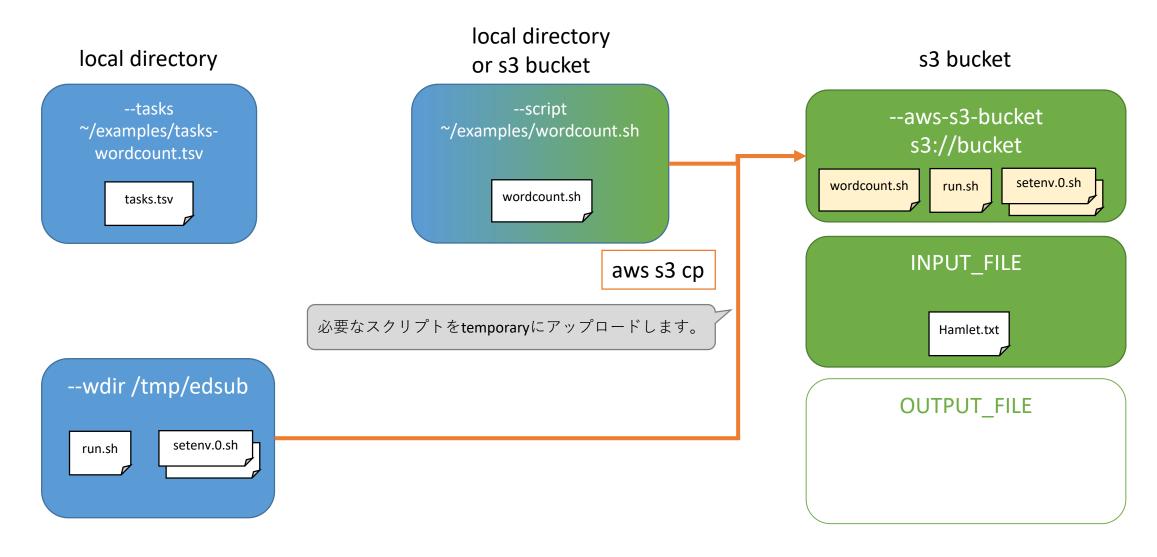
--aws-s3-bucket s3://bucket

INPUT_FILE

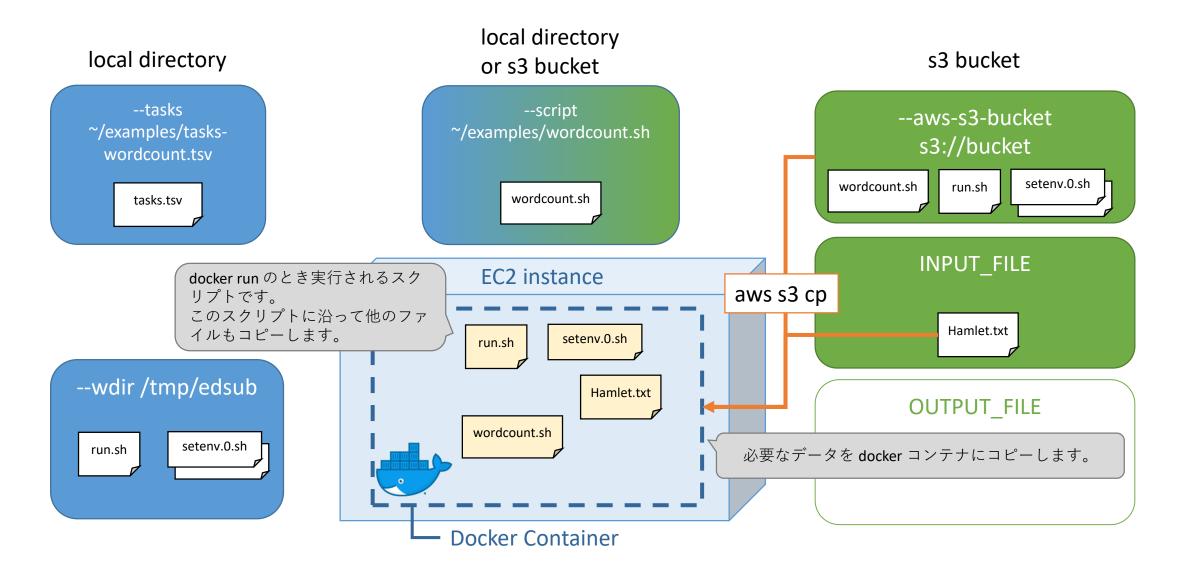
Hamlet.txt

OUTPUT FILE

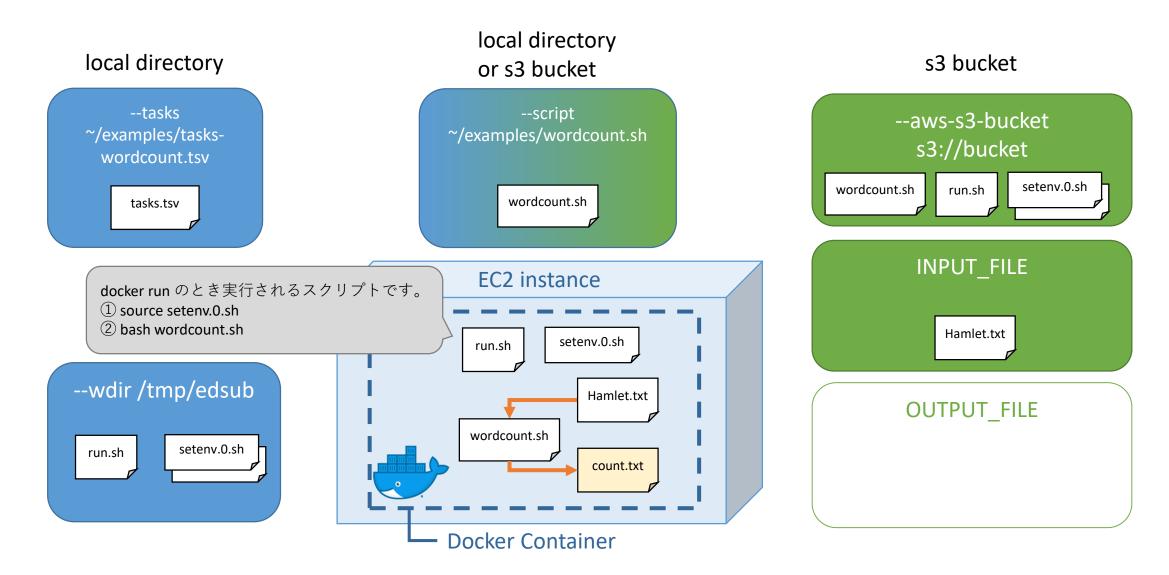
ecsub data flow 3 - scriptをS3にupload



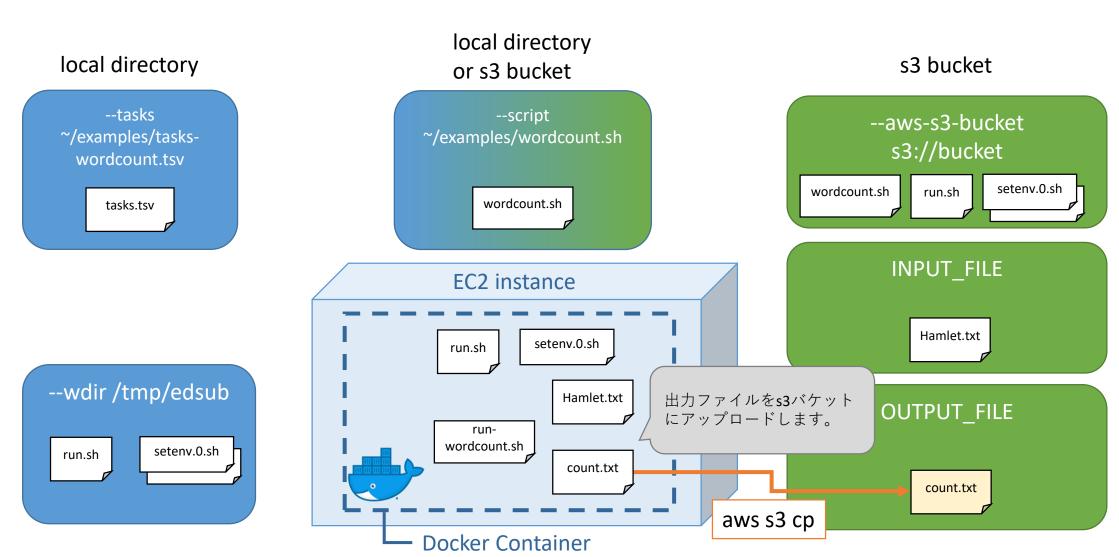
ecsub data flow 4 - docker コンテナ起動



ecsub data flow 5 - スクリプト実行



ecsub data flow 6 - outputをS3にupload



ecsub data flow 7 - EC2 インスタンス削除

local directory

--tasks ~/examples/taskswordcount.tsv

tasks.tsv

--wdir /tmp/edsub

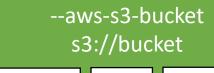
local directory or s3 bucket

--script ~/examples/wordcount.sh

word count. sh

EC2 インスタンスは削除

s3 bucket



wordcount.sh run.sh setenv.0.sh

INPUT_FILE

Hamlet.txt

OUTPUT_FILE

最終成果物です。 count.txt

Final product

ecsub submit 実行時フローまとめ

Submit

ec2 instance 作成

Run docker container:

docker run -e ENV:"s3://xxx/setenv.0.sh" -e EXEC:"s3://xxx/wordcount.sh" ¥ \${ユーザ指定のイメージ}¥

"/bin/bash -c apt-install awscli; aws s3 cp s3://xxx/run.sh ./; bash run.sh"

=== run.sh (ecsub により自動作成) ===
aws s3 cp \$ENV ./setenv.0.sh
aws s3 cp \$EXEC ./wordcount.sh
source ./setenv.0.sh
aws s3 cp \$S3_INPUT1 \$INPUT1
aws s3 cp \$S3_SCRIPT \$SCRIPT --recursive
bash ./wordcount.sh
aws s3 cp \$OUTPUT \$S3_OUTPUT

=== setenv.O.sh (ecsub により自動作成) ===
export S3_INPUT_FILE="s3://xxx/wordcount/input/hamlet.txt"
export INPUT_FILE="/AWS_DATA/xxx/wordcount/input/hamlet.txt"
export S3_SCRIPT="s3://xxx/wordcount/python"
export SCRIPT="/AWS_DATA/xxx/wordcount/python"
export S3_OUTPUT_FILE="s3://xxx/output/hamlet-count.txt"
export OUTPUT_FILE="/AWS_DATA/xxx/output/hamlet-count.txt"

=== wordcount.sh (ユーザが用意する) ===
python \${SCRIPT}/wordcount.py \${INPUT_FILE} \${OUTPUT_FILE}

