

Using z/OS to Access a Public Cloud - Beyond the Hand-Waving

Stephen Goetze

Kirk Wolf

Dovetailed Technologies, LLC



Thursday, March 5, 2015: 8:30AM – 9:30 AM

Session 16785



#SHAREorg



SHARE is an independent volunteer-run information technology association that provides **education, professional networking and industry influence.**



Trademarks

- Co:Z® is a registered trademark of Dovetailed Technologies, LLC
- AWS®, Amazon Elastic Compute Cloud®, and Amazon S3® are registered trademarks of Amazon Web Services, Inc
- z/OS®, DB2®, zEnterprise® and zBX® are registered trademarks of IBM Corporation
- Java® are registered trademarks of Oracle and/or its affiliates
- Hadoop® and HDFS™ are either registered trademarks or trademarks of the Apache Software Foundation
- Hortonworks® is a registered trademark of Hortonworks Inc.
- Linux® is a registered trademark of Linus Torvalds

Dovetailed Technologies

We provide z/OS customers world wide with innovative solutions that enhance and transform traditional mainframe workloads:

- Co:Z Co-Processing Toolkit for z/OS
 - z/OS Enabled SFTP, z/OS Hybrid Batch, z/OS Unix Batch integration
 - uses IBM Ported Tools for z/OS - OpenSSH
- JZOS
acquired by IBM in and now part of the z/OS Java SDK

Agenda

- Introduce z/OS Hybrid Batch Processing
- Hello World Example
- Security Considerations
- z/OS Hybrid Batch and public cloud services
 - Example: generate and publish PDF documents using input from z/OS data sets.
 - Run and managed as a z/OS hybrid batch job
 - 1 million PDFs (60 GB) generated and published on a public cloud Hadoop cluster in an hour for < \$25
 - All example code, JCL, and tools available free
- Summary
- References

Acknowledgement

Mike Cox of IBM's Advanced Technical Support first conceived and prototyped this example using z/OS and Co:Z to drive a zBX (power) Hadoop cluster.

Hybrid Computing Models

Well Known:

- Linux or Windows as user-facing edge, web and application servers (zBX, zLinux, or on a non-z platform)
 - z/OS provides back-end databases and transaction processing
- zBX as special purpose appliances or optimizers
 - DB2 Analytics Accelerator
 - DataPower

Another Model: **z/OS Hybrid Batch**

- zBX/zLinux/Linux/Unix/Windows integrated with z/OS batch

z/OS Hybrid Batch Processing

1. The ability to execute a program or script on a virtual server from a z/OS batch job step
2. The target program may already exist and should require little or no modification
3. The target program's input and output are redirected from/to z/OS spool files or datasets
4. The target program may easily access other z/OS resources: DDs, data sets, POSIX files and programs
5. The target program's exit code is adopted as the z/OS job step condition code

Network Security provided by OpenSSH.

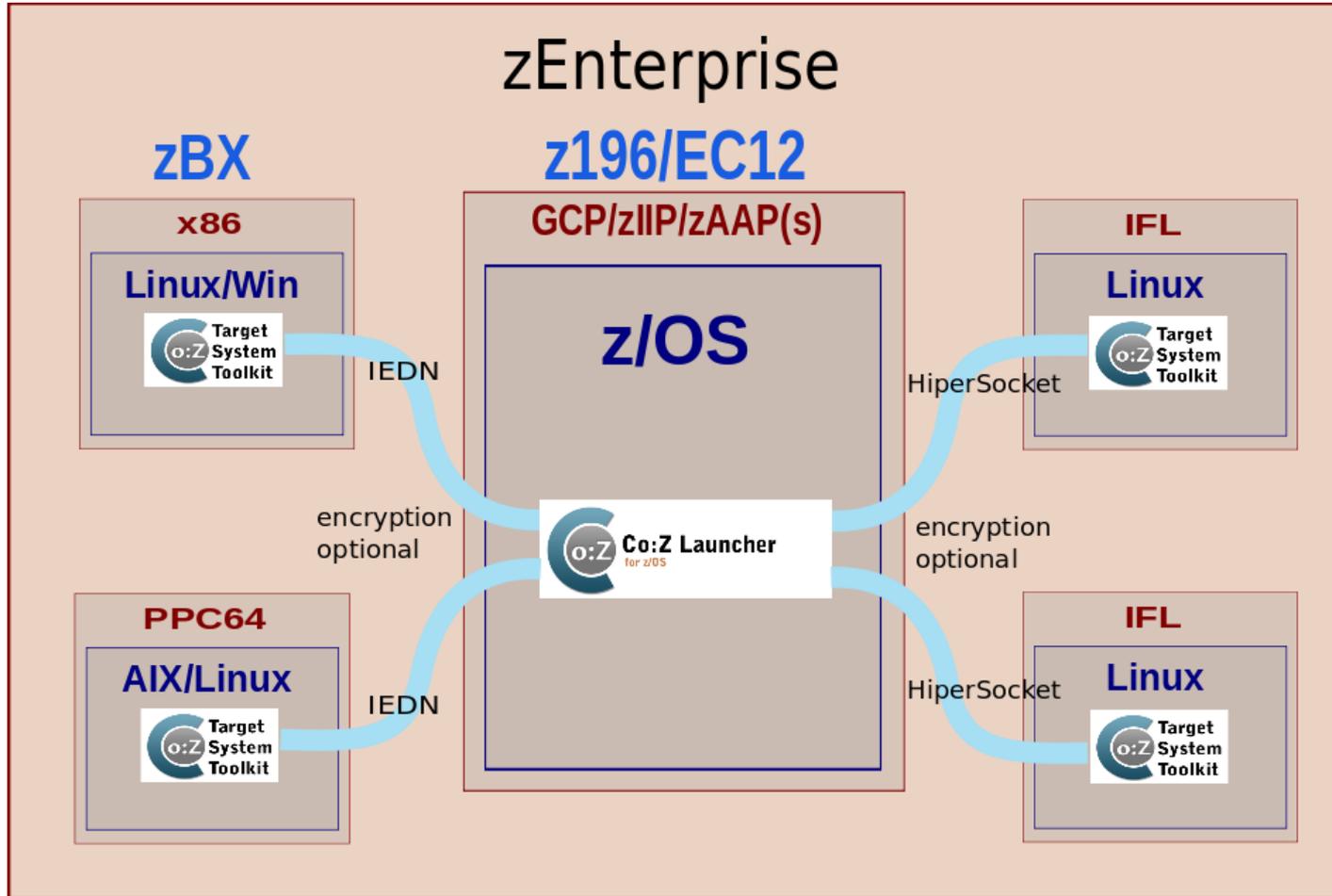
Data security governed by SAF (RACF/ACF2/TSS)

Requires new enablement software...

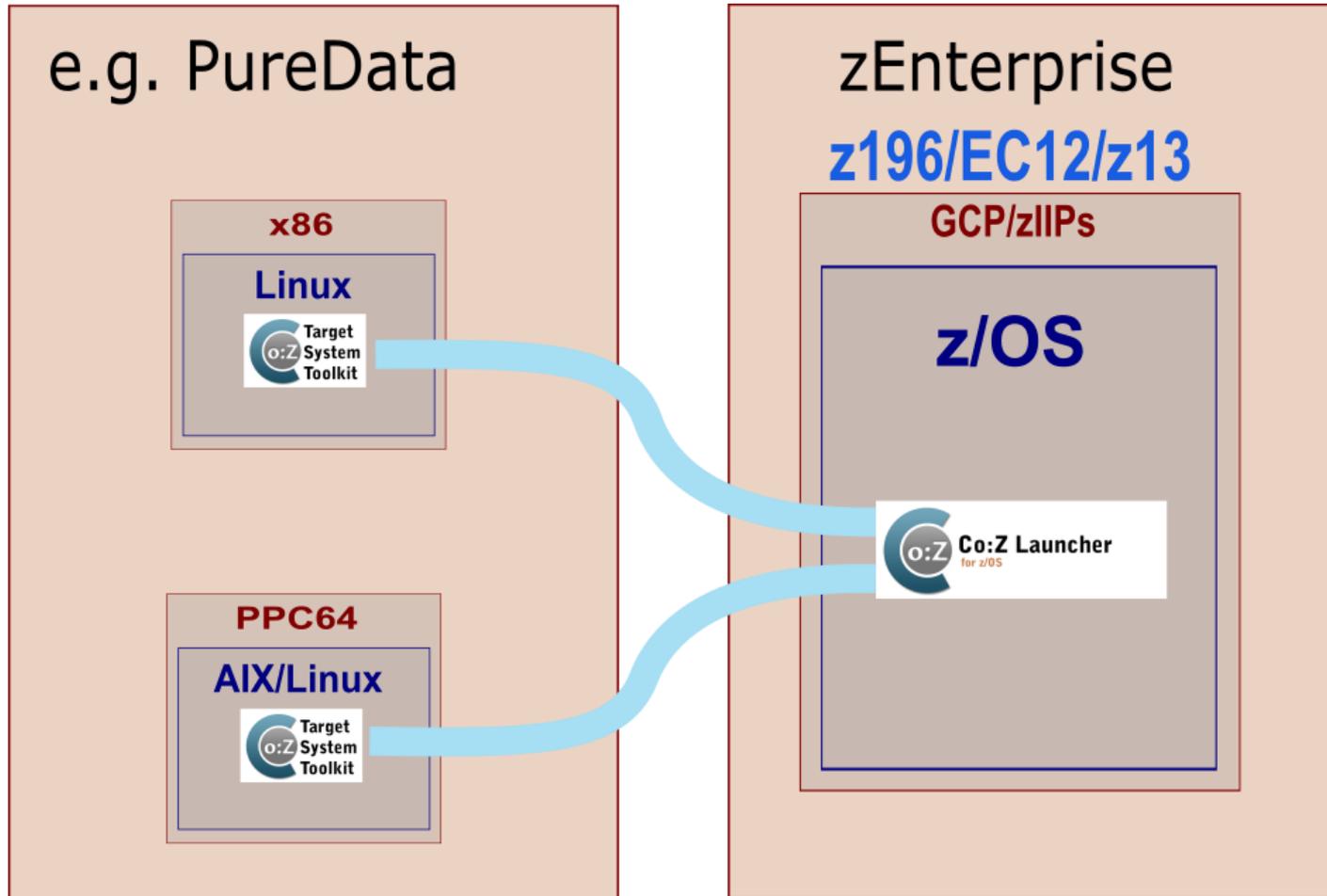
Co:Z Co-Processing Toolkit

- Implements z/OS Hybrid Batch model
- Co:Z Launcher starts a program on a target server and automatically redirects the standard streams back to jobstep DDs
- The target program can use Co:Z DatasetPipes commands to reach back into the active jobstep and access z/OS resources:
 - **fromdsn/todsn** – read/write a z/OS DD or data set
 - **fromfile/tofile** – read/write a z/OS Unix file
 - **cozclient** – run z/OS Unix command
- Free (commercial support licenses are available)
- Visit <http://dovetail.com> for details

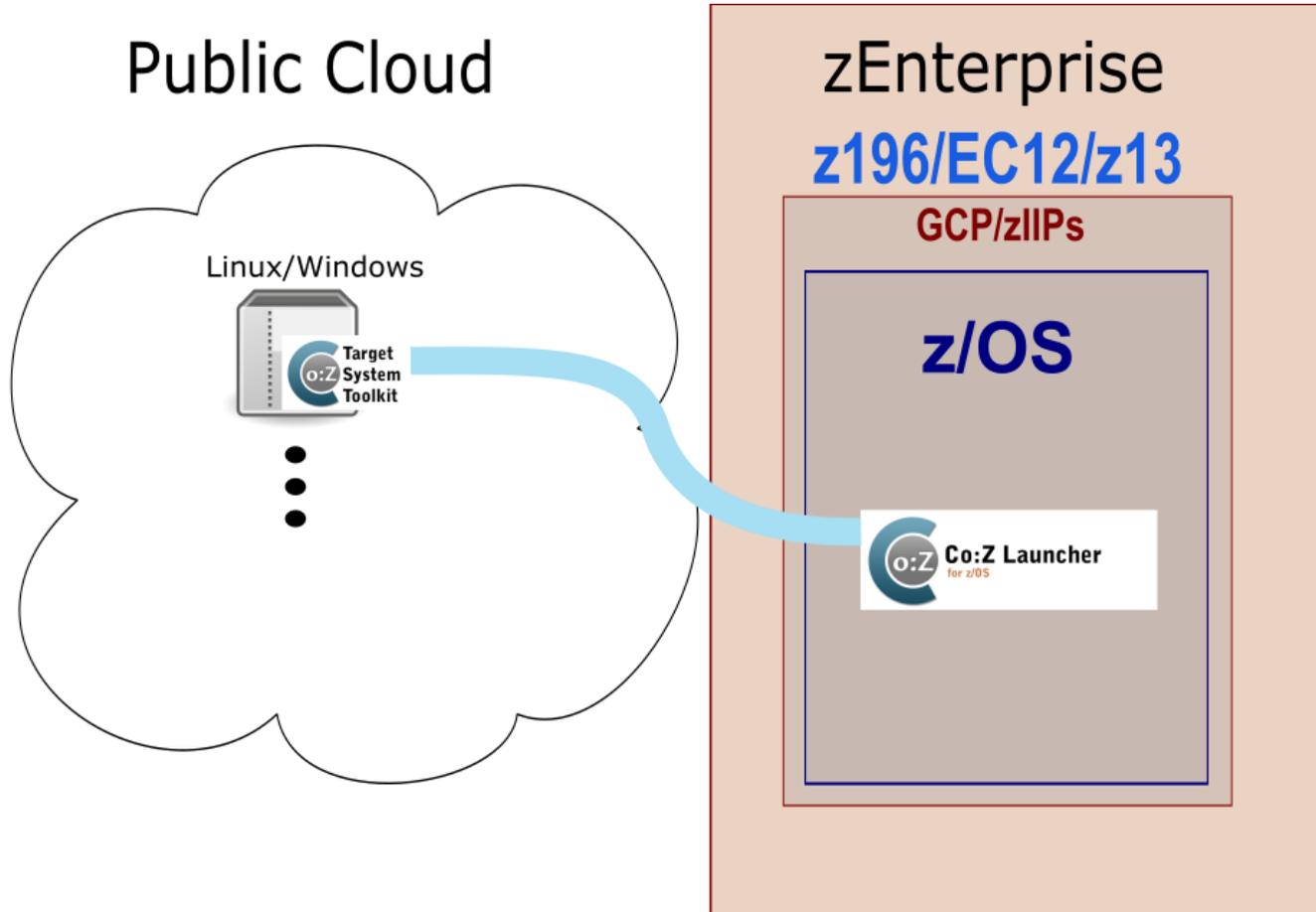
zEnterprise Hybrid Batch Processing



Private Cloud Hybrid Batch Processing



Public Cloud Hybrid Batch Processing



Hybrid Batch – Hello World

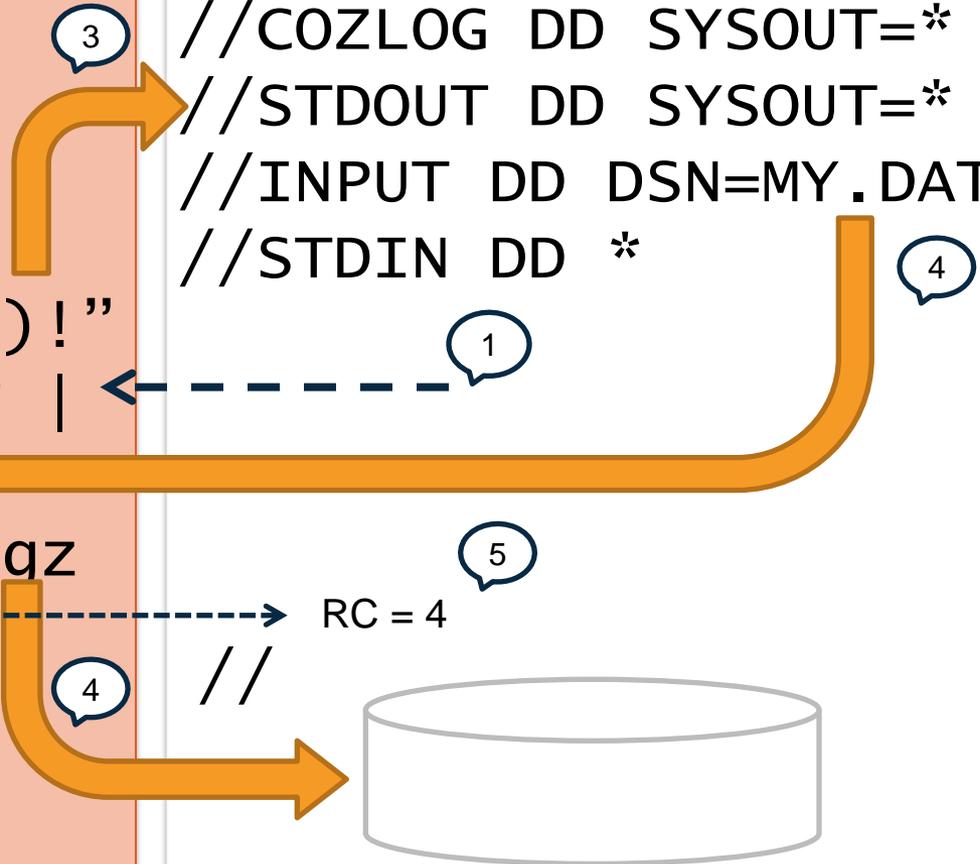
- Simple example illustrating the principles of Hybrid Batch Processing
- Launch a process on a remote Linux server
 - Write a message to stdout
 - In a pipeline:
 - Read the contents of a dataset from a jobstep DD
 - Compress the contents using the Linux gzip command
 - Write the compressed data to the z/OS Unix file system
 - Exit with a return code that sets the jobstep CC

Linux

```
echo "Hello $(uname)!"  
fromdsn -b DD:INPUT |  
gzip -c |  
tofile -b /tmp/out.gz  
exit 4
```

z/OS

```
//HYBRIDZ JOB ()  
//RUN EXEC PROC=COZPROC,  
// ARGS='u@linux'  
//COZLOG DD SYSOUT=*  
//STDOUT DD SYSOUT=*  
//INPUT DD DSN=MY.DATA  
//STDIN DD *
```



Hello World: Hybrid Batch

1. A script is executed on a virtual server from a z/OS batch job step
2. The script uses a program that already exists -- `gzip`
3. Script output is redirected to z/OS spool
4. z/OS resources are easily accessed using `fromdsn`, `tofile`, etc...
5. The script exit code is adopted as the z/OS job step CC

Hello World – DD:STDOUT

```
Hello Linux!
```

Hello World – DD:COZLOG

```
CoZLauncher[N]: version: 3.0.0 2014-11-12
cozagent[N]: version: 1.1.2 2013-03-19
fromdsn(DD:STDIN) [N]: 5 records/400 bytes read..
fromdsn(DD:INPUT) [N]: 78 records/6240 bytes read..
tofile(/tmp/out.gz) [N]: ... 1419 bytes written
todsn(DD:STDOUT) [N]: ... 13 bytes written
todsn(DD:STDERR) [N]: ... 0 bytes written
CoZLauncher[E]: u@linux target ... ended with RC=4
```

Hello World – DD:JESMSG LG

```
JOB01515 ----- FRIDAY, 7 SEPT 2014 -----  
JOB01515  IRR010I  USERID GOETZE  IS ASSIG..  
JOB01515  ICH70001I GOETZE  LAST ACCESS AT..  
JOB01515  $HASP373 HYBRIDZ  STARTED - INIT..  
JOB01515  -  
JOB01515  -STEPNAME PROCSTEP      RC      EXCP..  
JOB01515  -RUN      COZLNCH      04      1345..  
JOB01515  -HYBRIDZ  ENDED.  NAME-  
JOB01515  $HASP395 HYBRIDZ  ENDED
```

Co:Z Hybrid Batch Network Security

- OpenSSH is used to establish secure, authenticated network connections.
 - IBM Ported Tools OpenSSH client on z/OS
 - OpenSSH sshd server on target system
- By default, data transfer is tunneled (encrypted) over the ssh connection
 - Optionally, data can be transferred over raw sockets (option: ssh-tunnel=false)
 - This offers very high performance without encryption costs
 - Ideal for a secure network, such as zEnterprise HiperSockets or IEDN

z/OS Hybrid Batch and a Public Cloud

- Why use a public cloud?
 - Infrastructure, platforms, and services on demand
(see other sessions for lots of hand-waving on this)
- z/OS is generally not suited as a cloud services *provider*.
 - Operating system, language, software requirements
 - Services like Hadoop are designed to horizontally scale across commodity hardware platforms
- z/OS Hybrid Batch offers a solution
 - z/OS centric operational control
 - Data and process connectivity
 - Security

Using Co:Z Toolkit with Cloud Services

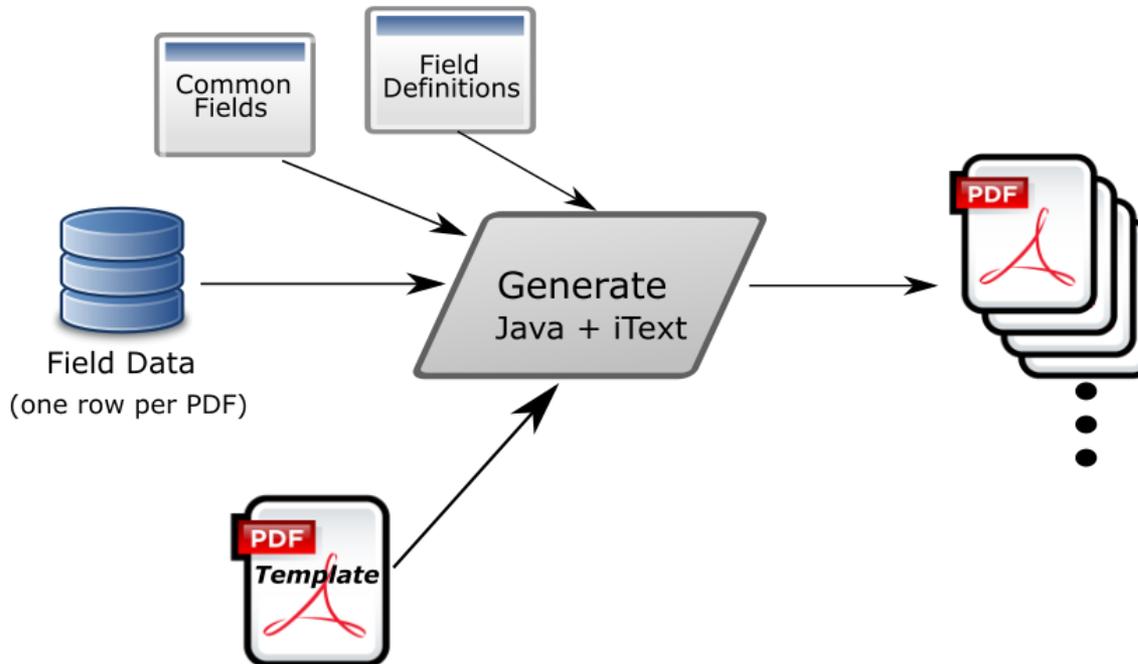
- The Co:Z Launcher and Dataset Pipes utilities facilitate:
 - Redirect input data from z/OS to the cloud
 - Data sets, POSIX files
 - Other sources, DB2, etc
 - Executing compute intensive jobs on cloud servers
 - Source maintained on z/OS
 - Control launch/termination of Hadoop clusters
 - Monitor progress in the job log
 - Redirect results to z/OS
 - Job spool
 - Data sets, POSIX files
 - Other destinations, DB2, etc

Generate, Publish Documents from z/OS data

- Example: generate and publish a million PDF documents using input from z/OS data sets
- Java driven PDF generation can be time consuming and expensive on the zSeries architecture
- z/OS hybrid batch can be used to:
 - Target Java execution to a Hadoop cloud cluster
 - Enable z/OS operations to retain control of scheduling
 - Publish the resulting PDFs in cloud storage
- For example code and JCL see:
 - <http://dovetail.com/products/casestudyitext.html>

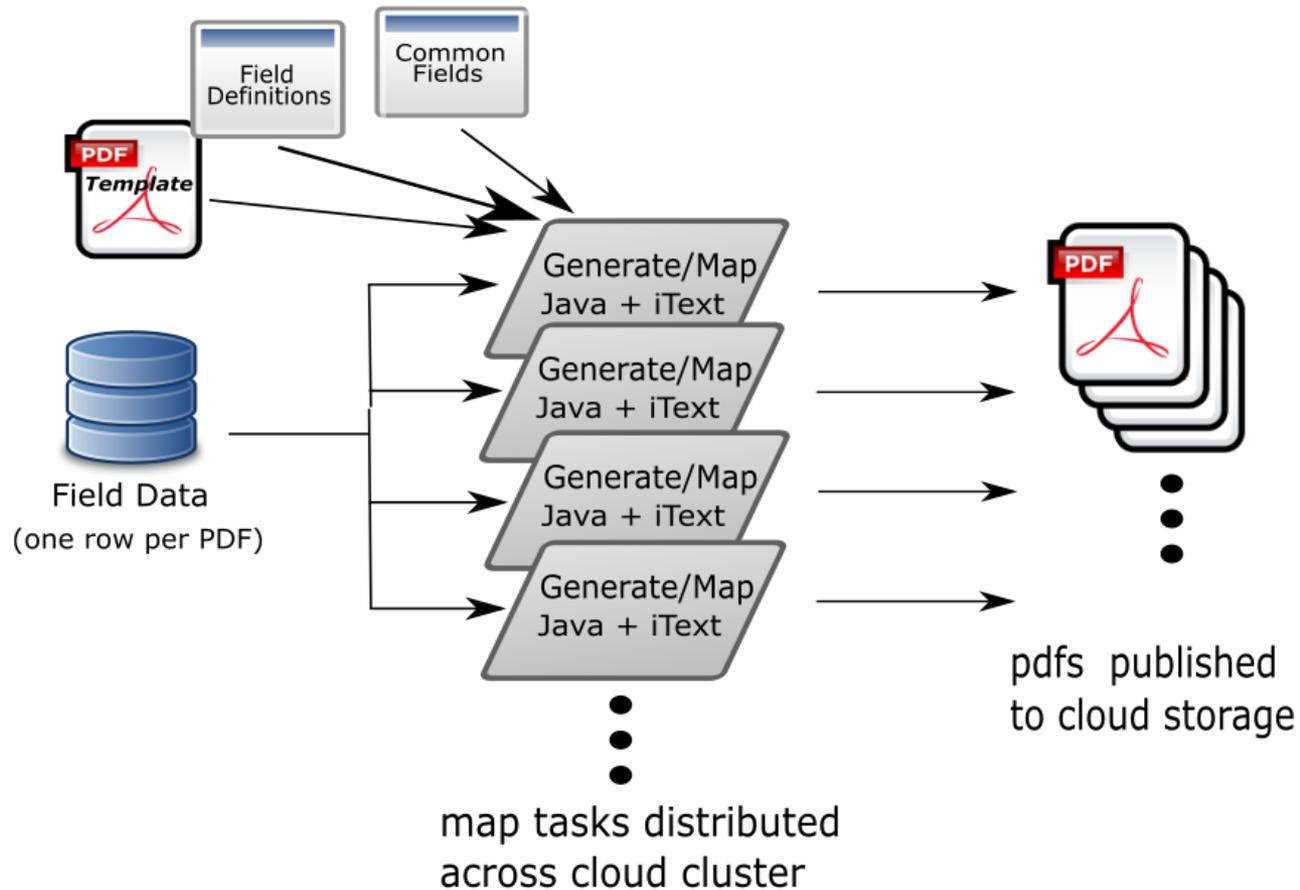
PDF Generation (a.k.a “stamping”)

(a single process implementation)



See <http://dovetail.com/products/casestudyitext.html>

PDF Generation using Hadoop cloud service



PDF Generation Example Overview

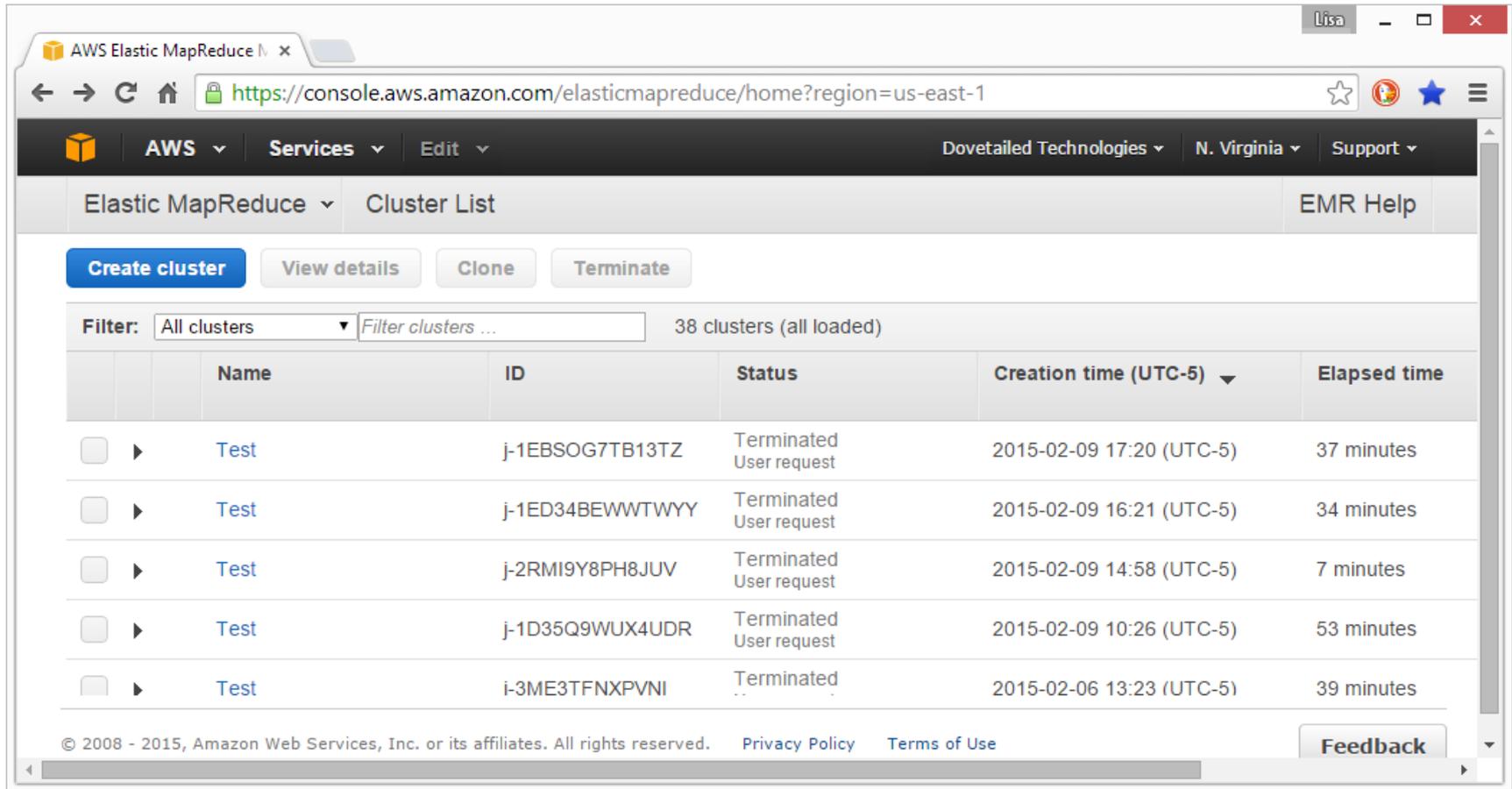
- Write a Java MapReduce application to generate PDFs.
 - See our example code to get started. You can use the free Hortonworks Hadoop sandbox for testing.
- Configure Amazon Web Services
- Create and Launch an Amazon Elastic MapReduce (EMR) Cluster
- Run Co:Z Launcher, executing the Java MapReduce on EMR
- Terminate the Amazon EMR Cluster
- View generated PDFs using Amazon S3 URLs

Configure Amazon Web Services

- Create an AWS Account at <http://aws.amazon.com>
- Create an Amazon S3 bucket
- Create an EC2 access key pair
- Create an Identity and Access Management (IAM) User
- Configure static website hosting
- Enable access to the published PDFs

Note: *The README in our downloadable example walks through these steps for proof of concept purposes; however, refer to aws.amazon.com for complete AWS documentation and recommendations for configuration and security.*

Create and Launch an Amazon EMR Cluster



The screenshot shows the AWS Elastic MapReduce console interface. At the top, there are navigation tabs for 'Elastic MapReduce' and 'Cluster List'. Below this, there are buttons for 'Create cluster', 'View details', 'Clone', and 'Terminate'. A filter dropdown is set to 'All clusters', and it indicates that 38 clusters are loaded. The main content is a table with the following columns: Name, ID, Status, Creation time (UTC-5), and Elapsed time. The table lists five clusters, all of which are in a 'Terminated' state due to a 'User request'.

Name	ID	Status	Creation time (UTC-5)	Elapsed time
Test	j-1EBSOG7TB13TZ	Terminated User request	2015-02-09 17:20 (UTC-5)	37 minutes
Test	j-1ED34BEWWTWYY	Terminated User request	2015-02-09 16:21 (UTC-5)	34 minutes
Test	j-2RMI9Y8PH8JUV	Terminated User request	2015-02-09 14:58 (UTC-5)	7 minutes
Test	j-1D35Q9WUX4UDR	Terminated User request	2015-02-09 10:26 (UTC-5)	53 minutes
Test	i-3ME3TFNXPVNI	Terminated	2015-02-06 13:23 (UTC-5)	39 minutes

<https://console.aws.amazon.com/elasticmapreduce/home?region=us-east-1#create-cluster:j-GBCN>

Elastic MapReduce > **Create Cluster**

Cluster Configuration

Cluster name

Termination protection Yes No
Prevents accidental termination of the cluster: to shut down the cluster, you must turn off termination protection. [Learn more](#)

Logging Enabled
Copy the cluster's log files automatically to S3. [Learn more](#)

Log folder S3 location

s3://<bucket-name>/<folder>/

Debugging Enabled
Index logs to enable console debugging functionality (requires logging). [Learn more](#)

Tags

Optional: Add up to 10 tags to your EMR cluster. A tag consists of a case-sensitive key-value pair. Tags on EMR clusters are propagated to the underlying EC2 instances. [Learn more](#) about tagging your Amazon EMR clusters.

Key	Value (optional)
<input type="text" value="Add a key to create a tag"/>	<input type="text"/>

Software Configuration

Hadoop distribution Amazon MapR
Use Amazon's Hadoop distribution. [Learn more](#)

AMI version

Determines the base configuration of the instances in your cluster, including the Hadoop version. [Learn more](#)

Use MapR's Hadoop distribution. [Learn more](#)

Applications to be installed	Version



AWS Elastic MapReduce | x

← → ↻ 🏠 <https://console.aws.amazon.com/elasticmapreduce/home?region=us-east-1#create-cluster:j-GBCN> ☆ 🔔 ☰

Hardware Configuration

i Specify the [networking](#) and [hardware](#) configuration for your cluster. If you need more than 20 EC2 instances, [complete this form](#).
[Request Spot instances](#) (unused EC2 capacity) to save money.

Network Use a Virtual Private Cloud (VPC) to process sensitive data or connect to a private network. [Create a VPC](#)

EC2 Subnet [Create a Subnet](#)

Type	Name	EC2 instance type	Count	Request spot	Bid price
Master	<input type="text" value="Master instance group -"/>	<input type="text" value="m3.xlarge"/>	<input type="text" value="1"/>	<input type="checkbox"/>	<input type="text"/>
Core	<input type="text" value="Core instance group - 2"/>	<input type="text" value="m1.large"/>	<input type="text" value="10"/>	<input type="checkbox"/>	<input type="text"/>

Security and Access

EC2 key pair Use an existing EC2 key pair to SSH into the master node of the Amazon EMR cluster. [Learn more](#)

IAM user access All other IAM users No other IAM users
 Control the visibility of this cluster to other IAM users. [Learn more](#)

AWS Elastic MapReduce

https://console.aws.amazon.com/elasticmapreduce/home?region=us-east-1#create-cluster:j-GBCNXQ

Bootstrap Actions

i Bootstrap actions are scripts that are executed during setup before Hadoop starts on every cluster node. You can use them to install additional software and customize your applications. [Learn more](#)

Bootstrap action type	Name	S3 location	Optional arguments
Custom action	Custom action	s3://dovetailed-technologies/install/download.sh	 

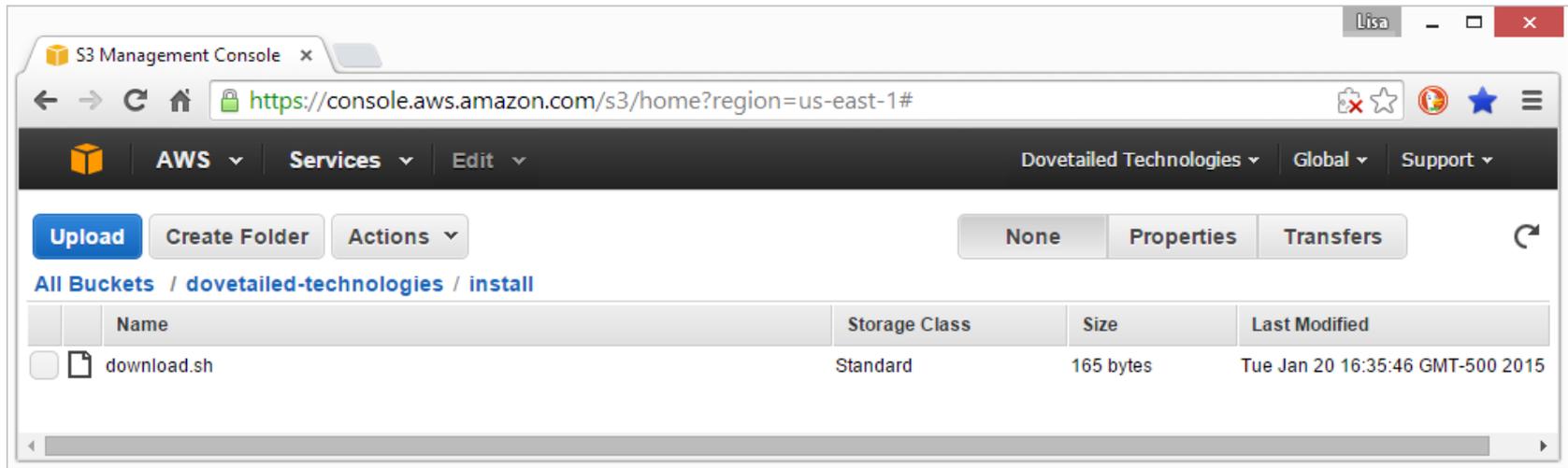
Add bootstrap action

Steps

i A step is a unit of work you submit to the cluster. A step might contain one or more Hadoop jobs, or contain instructions to install or configure an application. You can submit up to 256 steps to a cluster. [Learn more](#)

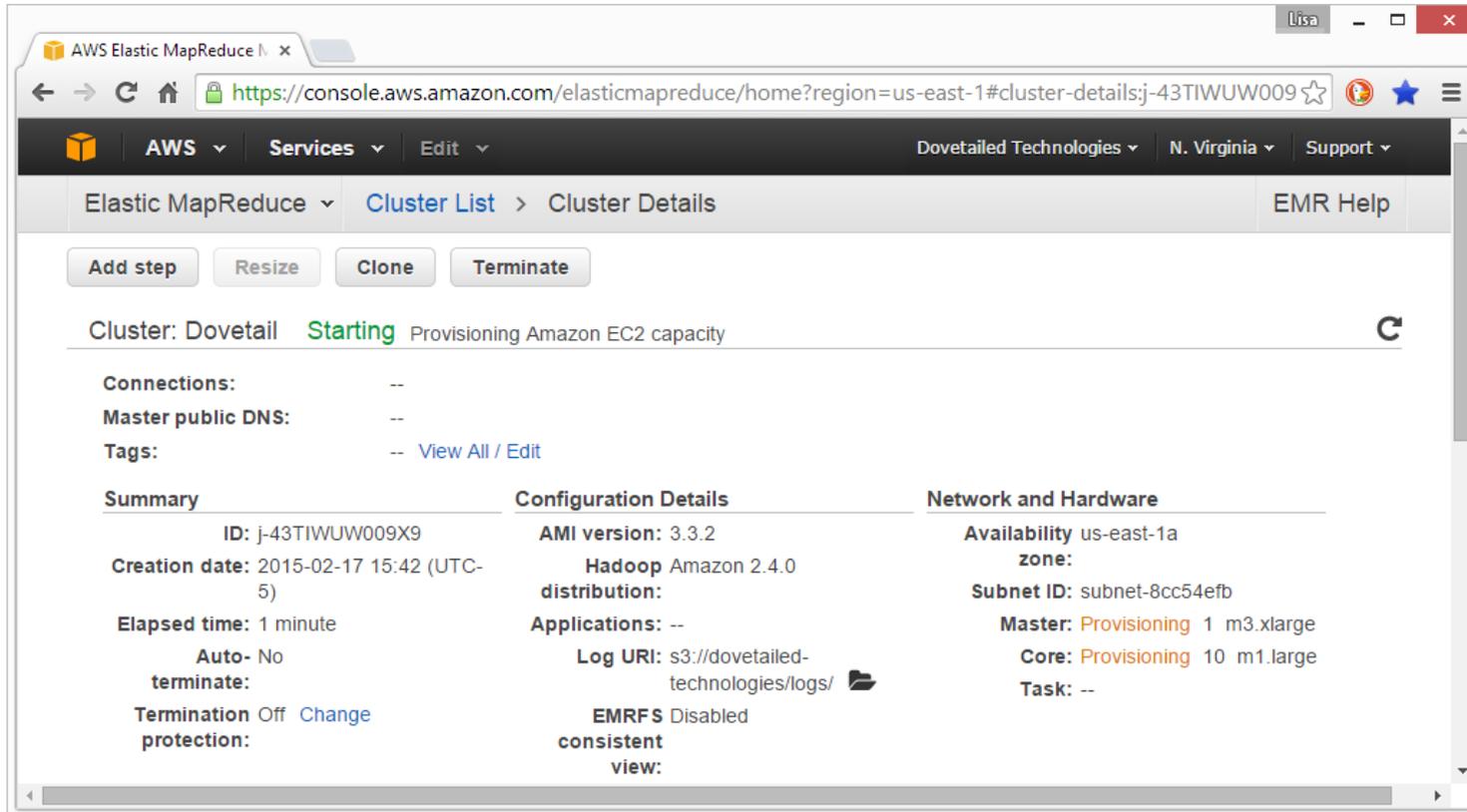
Name	Action on failure	JAR location	Arguments
Add step <input type="text" value="Select a step"/> <input type="button" value="Configure and add"/>			
Auto-terminate	<input type="radio"/> Yes	Automatically terminate cluster after the last step is completed.	
	<input checked="" type="radio"/> No	Keep cluster running until you terminate it.	

Bootstrap Action



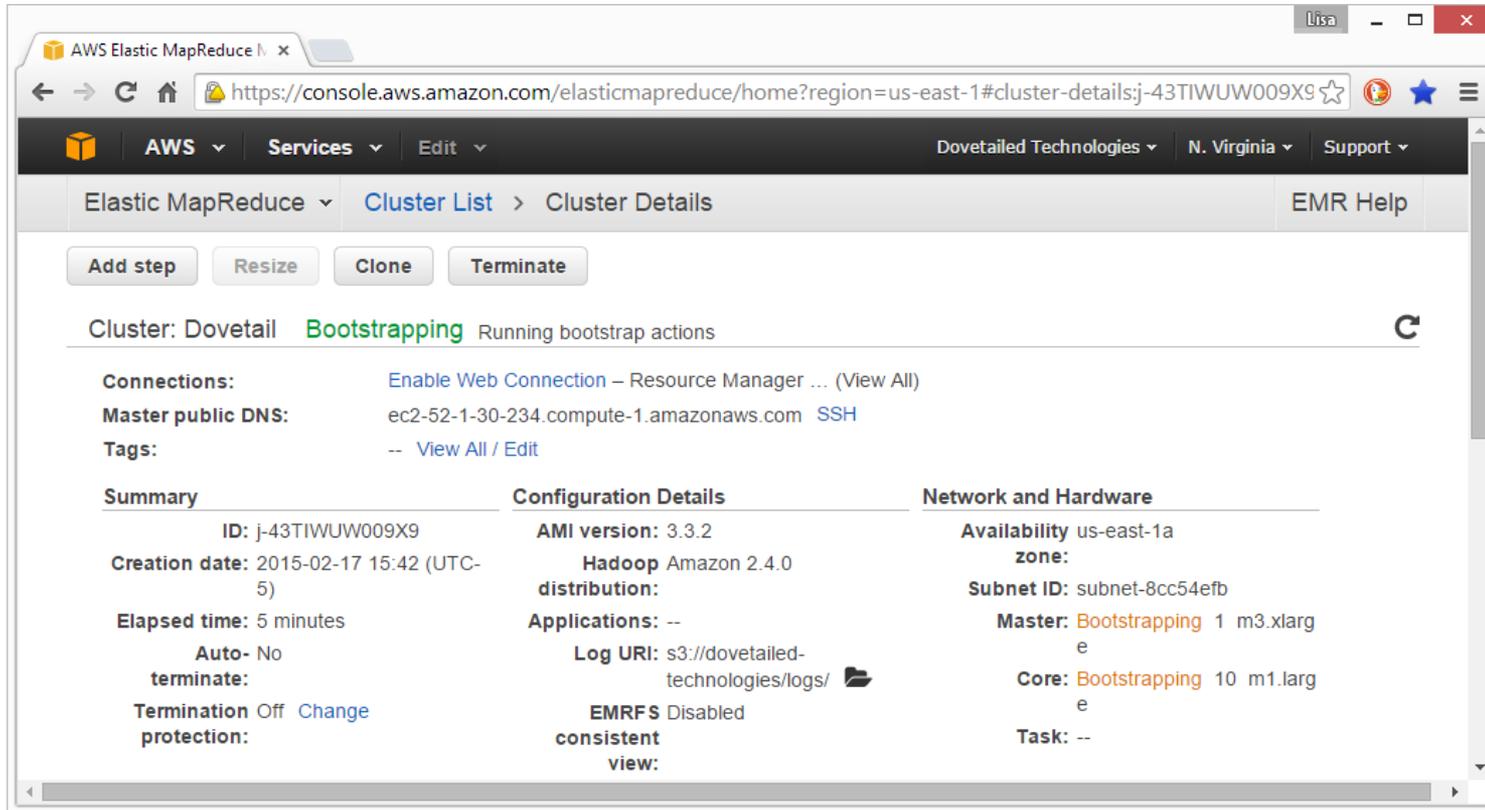
```
#!/bin/bash
wget http://www.dovetail.com/downloads/coz/coz-toolkit-1.1.2-1.x86_64.rpm
sudo rpm -i coz-toolkit-1.1.2-1.x86_64.rpm
mkdir -p /home/hadoop/tmp
```

Starting the Cluster



The screenshot shows the AWS Elastic MapReduce console interface. At the top, there's a navigation bar with 'AWS', 'Services', and 'Edit' menus. Below that, the breadcrumb trail reads 'Elastic MapReduce > Cluster List > Cluster Details'. A secondary navigation bar contains 'Add step', 'Resize', 'Clone', and 'Terminate' buttons. The main content area displays the cluster name 'Dovetail' in a green 'Starting' state, with a sub-status 'Provisioning Amazon EC2 capacity'. Below this, there are sections for 'Connections', 'Master public DNS', and 'Tags'. The core of the page is divided into three columns: 'Summary', 'Configuration Details', and 'Network and Hardware'. The 'Summary' column shows the cluster ID 'j-43TIWUW009X9', creation date '2015-02-17 15:42 (UTC-5)', elapsed time '1 minute', and 'Auto-terminate' set to 'No'. The 'Configuration Details' column lists 'AMI version: 3.3.2', 'Hadoop distribution: Amazon 2.4.0', 'Applications: --', 'Log URI: s3://dovetailed-technologies/logs/', and 'EMRFS consistent view: Disabled'. The 'Network and Hardware' column shows 'Availability zone: us-east-1a', 'Subnet ID: subnet-8cc54efb', and instance counts: 'Master: Provisioning 1 m3.xlarge', 'Core: Provisioning 10 m1.large', and 'Task: --'.

Starting the Cluster ...



Cluster: Dovetail **Bootstrapping** Running bootstrap actions

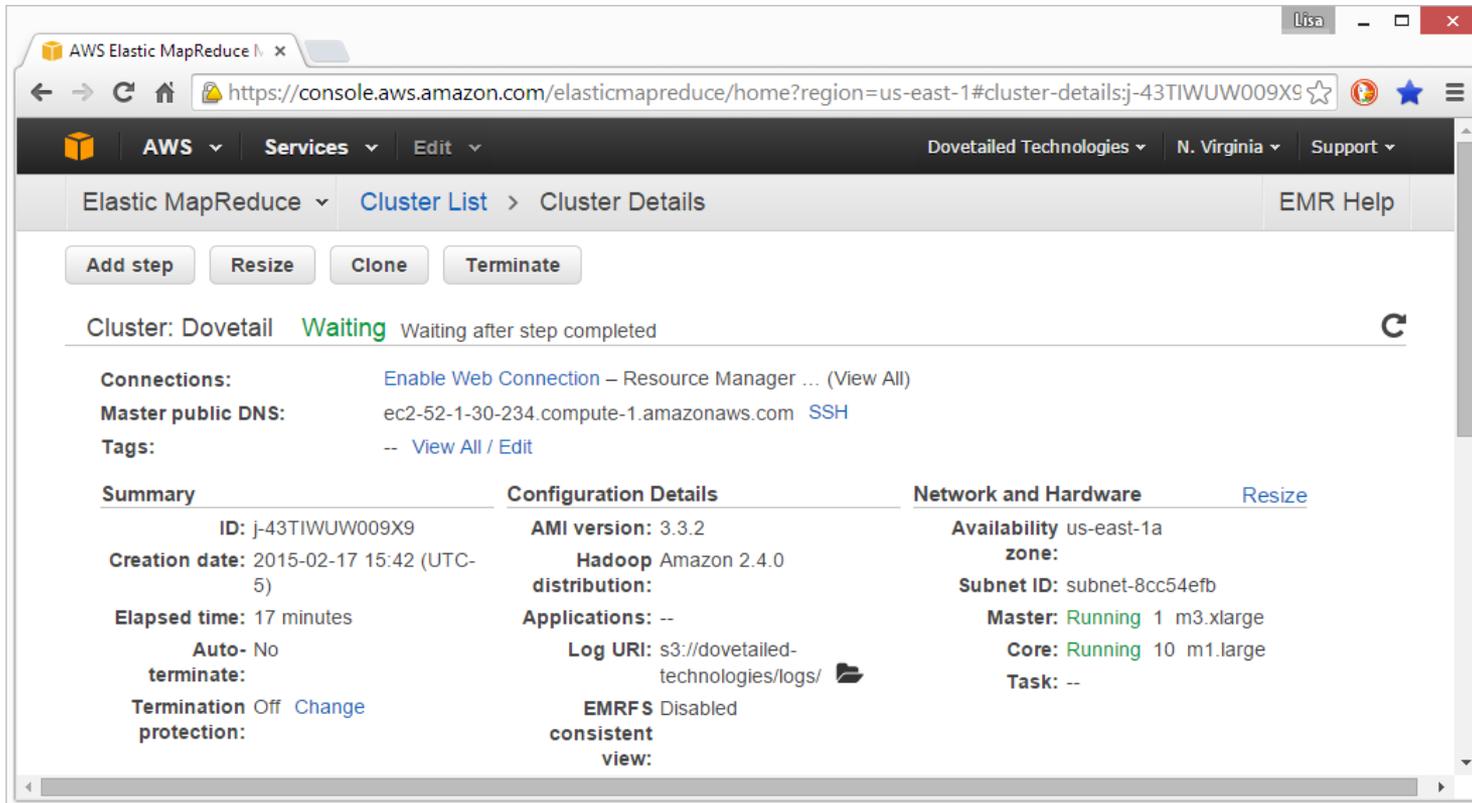
Connections: [Enable Web Connection](#) – Resource Manager ... (View All)

Master public DNS: ec2-52-1-30-234.compute-1.amazonaws.com [SSH](#)

Tags: -- [View All / Edit](#)

Summary	Configuration Details	Network and Hardware
ID: j-43TIWUW009X9	AMI version: 3.3.2	Availability zone: us-east-1a
Creation date: 2015-02-17 15:42 (UTC-5)	Hadoop distribution: Amazon 2.4.0	Subnet ID: subnet-8cc54efb
Elapsed time: 5 minutes	Applications: --	Master: Bootstrapping 1 m3.xlarge
Auto-terminate: No	Log URI: s3://dovetailed-technologies/logs/ 	Core: Bootstrapping 10 m1.large
Termination protection: Off Change	EMRFS consistent view: Disabled	Task: --

Cluster is Ready



The screenshot shows the AWS Elastic MapReduce console interface. At the top, there are navigation tabs for 'Elastic MapReduce', 'Cluster List', and 'Cluster Details'. Below this, there are buttons for 'Add step', 'Resize', 'Clone', and 'Terminate'. The main content area displays the cluster name 'Dovetail' with a status of 'Waiting' and the message 'Waiting after step completed'. Below this, there are sections for 'Connections', 'Master public DNS', and 'Tags'. The bottom section is divided into three columns: 'Summary', 'Configuration Details', and 'Network and Hardware'. The 'Summary' column shows the cluster ID, creation date, elapsed time, and termination protection. The 'Configuration Details' column shows the AMI version, Hadoop distribution, applications, log URI, and EMRFS consistent view. The 'Network and Hardware' column shows the availability zone, subnet ID, and the number of master, core, and task nodes.

Summary	Configuration Details	Network and Hardware
ID: j-43TIWUW009X9	AMI version: 3.3.2	Availability zone: us-east-1a
Creation date: 2015-02-17 15:42 (UTC-5)	Hadoop distribution: Amazon 2.4.0	Subnet ID: subnet-8cc54efb
Elapsed time: 17 minutes	Applications: --	Master: Running 1 m3.xlarge
Auto-terminate: No	Log URI: s3://dovetailed-technologies/logs/	Core: Running 10 m1.large
Termination protection: Off Change	EMRFS consistent view: Disabled	Task: --

- Transfer your mykeypair.pem to your z/OS .ssh directory
- Click on the SSH link and follow instructions to connect to the Master public host name from z/OS

z/OS Hybrid Batch JCL

- Defines inputs from z/OS data sets
 - Customer field data (one row per output PDF)
 - Common fields and field descriptions
 - a template PDF with form fields
- Submits the MapReduce job to the cluster
 - Transfers Java jar and submits the job to the hadoop master node
 - Streams input data to the cluster for processing
 - Streams output PDFs to Amazon S3, scalable cloud storage
- Transfers to z/OS the list of Amazon S3 URLs for each generated PDF

Co:Z Launcher: Define z/OS Data Set Inputs

```
//CZUSERR JOB ( ) , 'COZ' ,MSGCLASS=H ,NOTIFY=&SYSUID ,CLASS=A
//      SET PRFX=CZUSER.COZ.HADOOP.ITEXT
//DELOLD EXEC PGM=IEFBR14
//URLOUT  DD DSN=&PRFX..URLOUT ,
//          DISP=(MOD,DELETE,DELETE)
//STEP01  EXEC PROC=COZPROC , ARG='-LI '
//FLDDATA DD DISP=SHR,DSN=&PRFX..FLDDATA
//COMDATA DD DISP=SHR,DSN=&PRFX..CONFIG(COMDATA)
//COMMAP  DD DISP=SHR,DSN=&PRFX..CONFIG(COMMAP)
//FLDMAP  DD DISP=SHR,DSN=&PRFX..CONFIG(FLDMAP)
//TEMPLATE DD DISP=SHR,DSN=&PRFX..TEMPLATE
//URLOUT  DD DSN=&PRFX..URLOUT ,
//          DISP=(NEW,CATLG) ,
//          SPACE=(CYL,(200,100),RLSE) ,
//          DCB=(RECFM=VB,LRECL=1028,BLKSIZE=6144)
```

(Sample input field data)

```
//FLDDATA DD DISP=SHR,DSN=&PRFX..FLDDATA
10010|Shana V. Crane|(642) 468-1794|non.arcu@ignore.com|915-2786 Cursus Ave|Green Bay|...
10011|Orla A. Peck|(138) 218-6798|convallis.est.vitae@ignore.com|P.O. Box 392, 2726 Ma...
10012|Isadora N. Tyson|(603) 373-1585|pede.malesuada.vel@ignore.com|989-4958 Tempor Ave...
10013|Lesley X. Castro|(892) 940-5463|egestas@ignore.com|Ap #420-7231 Magna. Rd.|Huntin...
10014|Lisandra L. Russo|(625) 271-5861|sed.dolor.Fusce@ignore.com|Ap #802-4828 Eu Street...
10015|Kibo Z. Sykes|(751) 248-2339|Nunc.mauris.sapien@ignore.com|394-3122 Etiam Ave|Hart...
10016|Erin A. Shannon|(773) 202-0622|ac.turpis.egestas@ignore.com|Ap #941-8398 Mauris. R...
10017|Ian V. Mcclure|(561) 633-9148|vel.convallis@ignore.com|P.O. Box 686, 5766 Cras Ave...
10018|Omar F. Grant|(171) 220-6026|nec.diam@ignore.com|207-8265 Cursus. St.|Coatesville|...
10019|Calvin S. Walters|(417) 929-6722|vel.turpis@ignore.com|Ap #353-7720 Sed Rd.|Evansv...
10020|Hedley S. Guy|(623) 456-8704|magna@ignore.com|8223 Aenean St.|Garland|C6G 5W9|NB|A...
10021|Tamekah K. Mccray|(310) 180-5164|nascetur@ignore.com|9983 Libero. Rd.|Culver City|...
10022|Jessica H. Walter|(673) 189-5945|vulputate@ignore.com|Ap #119-768 In St.|Ocean Cit...
10023|Maile P. Holland|(735) 320-8493|dignissim@ignore.com|P.O. Box 732, 7222 Pede Road|...
10024|Alfreda Y. Snider|(398) 791-0593|libero.et.tristique@ignore.com|6816 Nam Avenue|Mo...
...
```

Co:Z Launcher JCL: Setup

```
//COZCFG DD *
target-user=hadoop
target-host=ec2-52-1-30-234.compute-1.amazonaws.com
# Note: StrictHostKeyChecking=no could be added to ssh-options
ssh-options=-oIdentityFile=/home/czuser/.ssh/mykeypair.pem
//STDIN DD *
#
# Define the output directory as a bucket URL
#
awsid=CKIBICAXYBHETULAZKEL
awskey=54TzLHOqqOpZ9XNOTnM8IZwW8PUQaewp5D2qaJfc
bucket=s3://${awsid}:${awskey}@dovetailed-technologies
outdir=${bucket}/output
```

Co:Z Launcher JCL: transfer Java jar

```
# Transfer application jar from z/OS to hadoop master node
source=/home/czuser/coz-itext-hadoop
fromfile -b ${source}/ItextPdfGenerator.jar > ItextPdfGenerator.jar
```

Co:Z Launcher JCL: Submit Job to EMR

```
# ${outdir} is a bucket in Amazon S3
# /user/hadoop is HDFS in the hadoop cluster
# /home/hadoop/tmp is a local file system temp directory
#
hadoop jar ItextPdfGenerator.jar          \
${outdir}                                 \
/user/hadoop                               \
/home/hadoop/tmp                           \
<(fromdsn DD:COMDATA)                      \
<(fromdsn DD:COMMAP)                       \
<(fromdsn DD:FLDDATA)                      \
<(fromdsn DD:FLDMAP)                       \
<(fromdsn -b DD:TEMPLATE)                  \
2500000 # max split size
test $? -ne 0 && exit 8
```

Co:Z Launcher: Transfer URLs to z/OS

```
#  
# Copy the URL list from S3 to a dataset,  
# one file per mapper task  
#     ../output/urls-m-xxxxx  
#  
hadoop fs -cat ${outdir}/urls* | todsn DD:URLOUT
```

Bash shell: process substitution

- Make a command (or pipeline) appear as a file name argument:
 - `<(cmd)` – is substituted with name of *readable* pipe file: `/dev/fd/nn`
 - `>(cmd)` – is substituted with name of *writable* pipe file: `/dev/fd/nn`
- for this example:

```
diff <(ls dir1) <(ls dir2)
```

the bash shell actually does something like:

```
mkfifo /dev/fd/63 /dev/fd/64 # temp named pipe files
ls dir1 > /dev/fd/63 & # forks a child process
ls dir2 > /dev/fd/64 & # forks another child process
diff /dev/fd/63 /dev/fd/64
rm /dev/fd/63 /dev/fd/64
```

- Very handy for enabling “data in flight” in hybrid batch processing...

Co:Z Launcher: DD:COZLOG

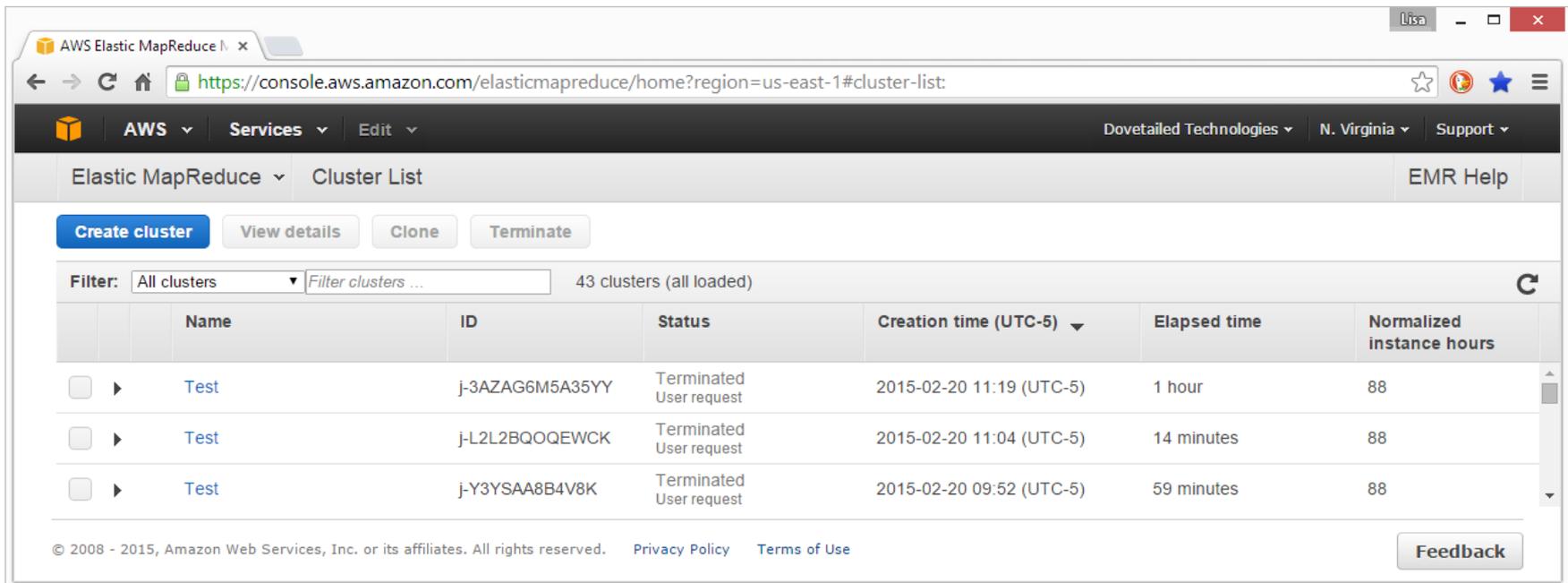
```
CoZLauncher[N]: version: 3.0.0 2014-11-12
cozagent[N]: version: 1.1.2 2013-03-19
fromdsn(DD:STDIN) [N]: 20 records/1600 bytes read..
fromfile(/home/czuser/coz-itext-hadoop/ItexPdfGenerator.jar) [N]: 7095063 bytes read..
fromdsn(DD:COMMAP) [N]: 1 records/26 bytes read..
fromdsn(DD:COMDATA) [N]: 1 records/40 bytes read..
fromdsn(DD:FLDMAP) [N]: 1 records/99 bytes read..
fromdsn(DD:TEMPLATE) [N]: 11 records/62716 bytes read..
fromdsn(DD:FLDDATA) [N]: 1000000 records/122727376 bytes read..
cozagent [22:33] Waiting for completion...
:
cozagent [22:53] Waiting for completion...
todsn(DD:URLOUT) [N]: ... 71888896 bytes written ...
todsn(DD:STDERR) [N]: ...353 records/32938 bytes written ...
todsn(DD:STDOUT) [N]: ...0 bytes written ...
CoZLauncher[I]: CoZAgent process (33620106) ended with RC=0
CoZLauncher[N]: hadoop@ec2-54-152-154-141.compute-1.amazonaws.com target command
'<default shell>' ended with RC=0
CoZLauncher[I]: CoZLauncher ended with RC=0
```

Co:Z Launcher: DD:STDERR

```
2015-02-20 16:25:55,790 INFO ItextPdfConfiguration -
  pdf-output-hdfs-directory: s3://<awsid>:<awskey>@dovetailed-technologies/output
  local-hdfs-root: /user/hadoop
  local-work-directory: /home/hadoop/tmp
  common-data: /dev/fd/63 (csv delimited by '|')
  common-map: /dev/fd/62 (csv delimited by '|')
  field-data: /dev/fd/61 (csv delimited by '|')
  field-map: /dev/fd/60 (csv delimited by '|')
  pdf-template: /dev/fd/59
  max-input-split-size: 2500000
2015-02-20 16:30:58,074 INFO JobSubmitter - number of splits:50
2015-02-20 16:30:58,617 INFO Job - The url to track the job: http://172.31.2.91:9046
2015-02-20 16:30:58,618 INFO Job - Running job: job_....
2015-02-20 16:31:09,819 INFO Job - map 0% reduce 0%
...
2015-02-20 16:56:32,575 INFO Job - map 100% reduce 0%
2015-02-20 17:16:09,155 INFO Job - Job job_... completed successfully
2015-02-20 17:16:09,292 INFO Job - Counters: 36
  File System Counters ...
  Job Counters
    Launched map tasks=50 ...
```

Terminate the EMR Cluster

- From the Amazon EMR interface, terminate the cluster
- Once terminated, the normalized instance hours provide the approximate number of compute hours.



The screenshot shows the AWS Elastic MapReduce console interface. The browser address bar displays the URL: <https://console.aws.amazon.com/elasticmapreduce/home?region=us-east-1#cluster-list>. The page title is "Elastic MapReduce Cluster List". There are navigation buttons for "Create cluster", "View details", "Clone", and "Terminate". A filter dropdown is set to "All clusters" and shows "43 clusters (all loaded)". The main content is a table with the following columns: Name, ID, Status, Creation time (UTC-5), Elapsed time, and Normalized instance hours. Three clusters are listed, all with a status of "Terminated User request".

	Name	ID	Status	Creation time (UTC-5)	Elapsed time	Normalized instance hours
<input type="checkbox"/>	Test	j-3AZAG6M5A35YY	Terminated User request	2015-02-20 11:19 (UTC-5)	1 hour	88
<input type="checkbox"/>	Test	j-L2L2BQOQEWCK	Terminated User request	2015-02-20 11:04 (UTC-5)	14 minutes	88
<input type="checkbox"/>	Test	j-Y3YSAA8B4V8K	Terminated User request	2015-02-20 09:52 (UTC-5)	59 minutes	88

© 2008 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy Policy](#) [Terms of Use](#) [Feedback](#)

Automate Cluster Launch and Termination

- A cluster can be created, launched and terminated using the AWS CLI
- Use the Co:Z Launcher properties exit to create and launch a cluster, setting the `target-host` property to the Master public host name

```
//COZCFG DD *
target-user=hadoop
target-host=
ssh-options=
properties-exit=/home/czuser/set-properties.sh
```
- Requires `ssh-options` to be set to the following in `set-properties.sh`:

```
-oStrictHostKeyChecking=no -oIdentityFile=/home/czuser/.ssh/mykeypair.pem
```
- Add a second step to the JCL that terminates active clusters

```
//STEP02 EXEC PROC=SFTPPROC
//SFTPIN DD *
/home/czuser/terminate.sh
/*
```
- The `set-properties.sh` and `terminate.sh` z/OS shell scripts use `ssh` to connect to a linux server and run AWS CLI commands.
- Download the example code to review the z/OS and linux sample scripts.

View the generated PDFs

- From ISPF, view the URLOUT data set

:

.

s3.amazonaws.com/dovetailed-technologies/output/**3661**/111957-m-00009.pdf

s3.amazonaws.com/dovetailed-technologies/output/**4239**/111958-m-00009.pdf

s3.amazonaws.com/dovetailed-technologies/output/**9127**/111959-m-00009.pdf

s3.amazonaws.com/dovetailed-technologies/output/**2479**/111960-m-00009.pdf

s3.amazonaws.com/dovetailed-technologies/output/**7964**/111961-m-00009.pdf

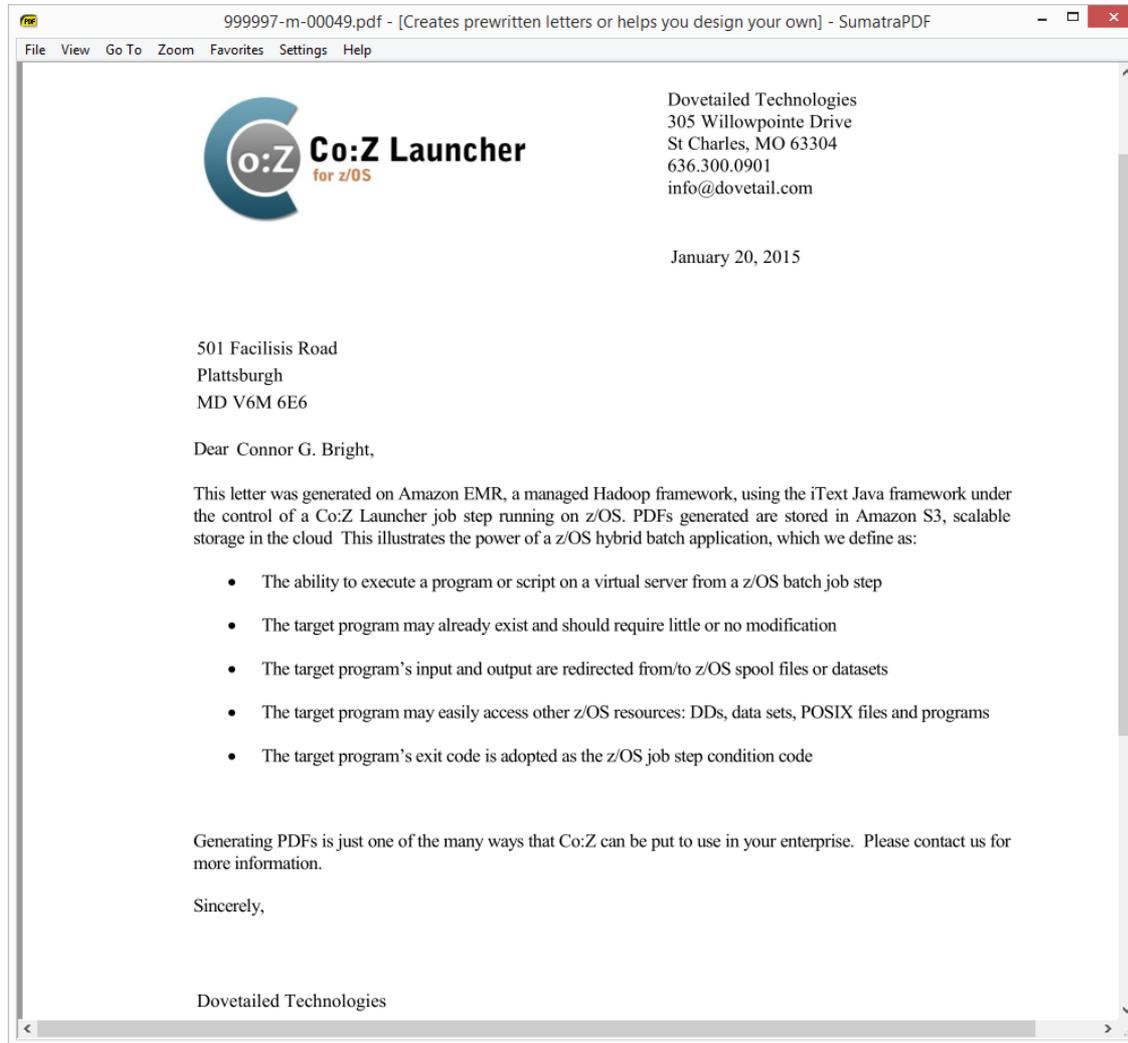
s3.amazonaws.com/dovetailed-technologies/output/**1198**/111962-m-00009.pdf

s3.amazonaws.com/dovetailed-technologies/output/**2276**/111963-m-00009.pdf

.

- Note the randomized part added to the S3 key
- Use the URL to access the PDF from a browser

A generated PDF



Complete your session evaluations online at www.SHARE.org/Seattle-Eval

Copyright © 2015 Dovetailed Technologies, LLC

The results

- Using a EMR 10-node (+1 master) “m3.xlarge” cluster
 - 4 vCPUs, 15 GB memory, Intel Xeon E5-2670 v2*
- Generated 1 Million PDFs (~ 60 GB) and transferred them to S3 storage.
 - Cluster startup - 5 minutes
 - Job execution - 58 Minutes (including startup)
 - EMR cost: \$3.85
 - S3 storage/transfer cost: \$12.00

z/OS: TOTAL TCB CPU TIME= .03 TOTAL ELAPSED TIME= 58.5

References

- Contacts: <http://dovetail.com>
info@dovetail.com
- For example code and JCL see:
<http://dovetail.com/products/casestudyitext.html>
- Hybrid Batch Information including additional case studies
<http://dovetail.com/solutions.html>
- Additional presentations, articles, webinars
<http://dovetail.com/webinars.html>
 - Introduction to z/OS Hybrid Batch Processing
 - z/OS Hybrid Batch Processing on the z/Enterprise
 - z/OS Hybrid Batch Processing and Big Data
- “Hadoop and System z” by Vic Leith and John Thomas, IBM
<http://www.redbooks.ibm.com/redpapers/pdfs/redp5142.pdf>

