deploy.zone

container deployment with SmartOS zones

a talk about:
practices and tools

David Grandinetti
@dbgrandi

Follow

OH: “Do programmers have any specific superstitions?”

“Yeah, but we call them best practices.”

9:02 PM - 6 Sep 2014

4,332 RETWEETS **2,395** FAVORITES

a talk about:
practices and tools

- not best practice
- but "current practice“
- still evolving as we learn

about SkyLime

- mostly consulting & managed internet services
- web, email, dns, db, … = core.io infrastructure
- hosted in multiple datacenters
- many instances of the same (similar) software

deploy.zone

- documentation of our practices
- tooling that supports us

Agenda

- SmartOS Zones
- ZFS delegated datasets
- reprovision
- pkgsrc
- SMF
- zoneinit
- mdata
- MIBE
- dz

SmartOS Zone

- OS Virtualization
- managed with „imgadm“ and „vmadm“

ZFS delegated dataset

- ZFS inside a zone
- `{… delegate_dataset: true, …}`
- snapshots, zfs send / receive
- backups (znapzend)
- `zfs set compression=lz4 $DDS/mail`

Stateless or delegated_dataset?

- try stateless by default (easy to scale, HA)
- push state to special zones with delegated_dataset
- similar to http://12factor.net/processes
- Example services that might use delegated datasets:
  mysql, redis master, samba fileserver, imap server, …
- Examples that can be made stateless:
  redis slaves, MX servers, web app workers, …

vmadm reprovision

- `vmadm reprovision <uuid> [-f <filename>]`
-
```
echo '{ "image_uuid": "d192eea8-ed68-11e3-bbc1-df80db1cbe67" }' |\
    vmadm reprovision 40bbc4e1-c6b3-454d-8a59-2cb3e3647850
```

- Change (or restore) the base image of a vm
- and execute provision step again
- keeps data on delegated datasets!

pkgsrc

- framework for building over 15,000 open source software packages
- native package manager on SmartOS, NetBSD, and Minix
- also available on OS X, Linux, …
- easy to build from source
- but binary packages available (`pkgin`)
- signed packages since 2014Q4

pkgsrc on the web

- http://pkgsrc.joyent.com - docs and pkgs from joyent
- http://pkgsrc.se - package browser
- http://pkgsrc.org - official
- http://pkgsrc.smartos.skylime.net- ipv6 mirror
- http://www.perkin.org.uk- blog of jperkin
- http://saveosx.org - docs for pkgsrc on osx

pkgsrc

- binary packages are great
- sometimes you need a specific version or a patch
- very easy to roll own packages with pkgsrc!
- own your complete software stack

pkgsrc

- Security Updates:
  `pkgin up && pkgin fug`
- Security alerts (put this in crontab / monitoring):
  `pkg_admin audit`
- Reprovision for quarterly releases
- 2014Q4 is a LTS release (3y security fixes)

SMF

- Service Management Facility (SMF)
- the init system
- features:
  - service dependencies
  - parallel starting of services
  - automatic service restart after failure
  - failure notifications by integrating FMA
- services are described by XML manifests

SMF

- `svcs`: examine state of your services
- `svcadm`: enable, disable, and restart a service
- `svccfg`: load manifest files (XML) that maintain configurations for each service
- `svcprop`: retrieves properties on a service (useful when writing custom scripts)
- http://wiki.smartos.org/display/DOC/Basic+SMF+Commands
- many manifests included in pkgsrc - if not submit pullrequest

zoneinit

- finalizes a new zone just provisioned
- https://github.com/joyent/zoneinit
- during (re-)provisioning zoneinit sources bash-snipptes in `/var/zoneinit/includes`
- good time to:
  - enable smf services for the first time
  - make final customizations
  - setup zfs filesystems if needed on delegated dataset

mdata

- `mdata-list`, `mdata-put`, `mdata-get`
- access `customer_metadata` of the vm
- with `sdc:` prefix access to the full manifest

MIBE

- Machine Image Build Environment
- https://github.com/joyent/mibe
- transforms a build plan („mi-repository“) into an image
- heavy changes to the global zone, not that easy to use

mi-repository

- https://github.com/joyent/mi-example
- essentials:
  - copy/
  - packages
  - customize
  - manifest

mibed

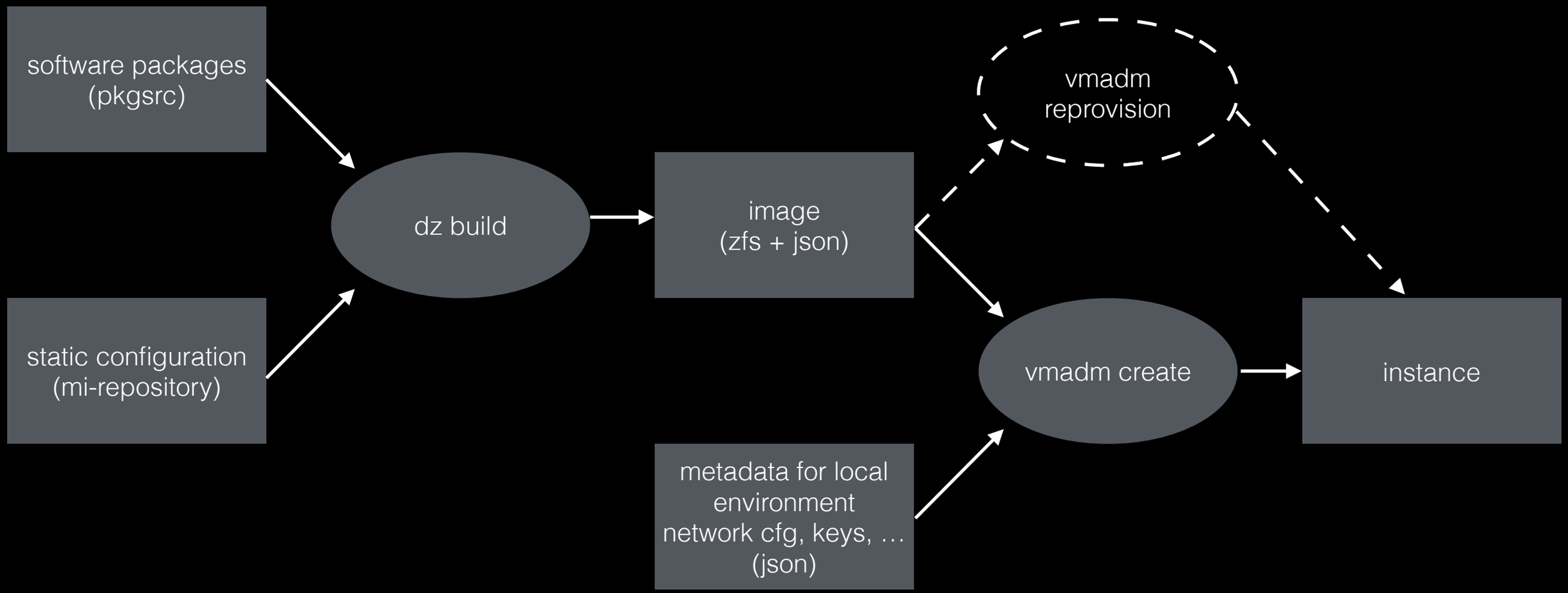
- not actually a daemon
- https://github.com/wiedi/mibed
- executes mibe on `git push`
- still heavy changes to the global zone
- easy to use

```
# git push mibed master
Counting objects: 5, done.
Delta compression using up to 4 threads.
Compressing objects: 100% (3/3), done.
Writing objects: 100% (3/3), 287 bytes | 0 bytes/s, done.
Total 3 (delta 2), reused 0 (delta 0)
remote: Image dc0688b2-c677-11e3-90ac-13373101c543 (base64 13.4.2) is already installed,
remote:
remote: build_smartos - version 1.0.0
remote: Image builder for SmartOS images
remote:
remote: * Sanity checking build files and environment..                    OK.
remote: * Halting build zone (9c0ff371-57f9-41b1)..                        OK.
remote: * Configuring build zone (9c0ff371-57f9-41b1) to be imaged..       OK.
remote: * Booting build zone (9c0ff371-57f9-41b1)..                        OK.
remote: * Copying in mi-graphite-1360a6ba40c2234ce558a5311216d3906f515c97/copy files..OK.
remote: * Creating image motd and product file..                          OK.
remote: * Installing packages list..                                      OK.
remote: * Executing the customize file..                                  OK.
remote: * Halting build zone (9c0ff371-57f9-41b1)..                        OK.
remote: * Un-configuring build zone (9c0ff371-57f9-41b1)..                OK.
remote: * Creating image file and manifest..                             OK.
remote:
remote: Image:    /opt/mibed/mibe/images/graphite-13.2.6.zfs.gz
remote: Manifest: /opt/mibed/mibe/images/graphite-13.2.6.dsmanifest
remote:
remote: Uploading to dsapid...
remote: ################################################################################ 100.0%
remote:
remote: UUID: 0cb4bc40-d79e-11e3-ae44-8ffa98d772ea
remote:
To git@mibed.example.com:mi-graphite
   764949b..1360a6b  master -> master
```

dz

- the deploy-zone utility
- `npm install -g deploy-zone`
- `dz build --host smartos.local --basevm <uuid> --publish http://imageserver.tld .`
- uses ssh to call `imgadm create`

zone configuration



dz

- dz also supports managing vms
- as it uses ssh, no centralized cloud needed
- local cache learns about vm uuids

dz

```
# dz host list
datacenter  host                    ram        storage  tags
----------  ----------------------  ---------  -------  ----
de-bln-f15  host-a.dev.example.com   3.87 GB    2.63 TB
de-bln-f15  host-b.dev.example.com   7.87 GB    2.63 TB
de-muc-ipx  host-c.prod.example.com  107.99 GB  5.25 TB

# dz list
host                  uuid                                  type  ram    quota  state    hostname            alias
--------------------  ------------------------------------  ----  -----  -----  -------  ------------------  -------------------
 host-a.dev.example.com  ab68ea68-d192-40a8-97c6-deca4f1710b1  OS     4096     10  running      fifo.f.fruky.net  fifo
 host-a.dev.example.com  8662a174-12c9-412d-8e23-b0110c68a107  OS     2048     10  running      blog.example.com  blog
 host-a.dev.example.com  0b1660f2-0894-48d0-9434-3320bab94c88  OS     1024   6000  running                 store-a  store-a
 host-a.dev.example.com  c11f26ba-0395-4723-9c2b-0d98931ce182  KVM    1024     10  stopped                          netbsd
 host-b.dev.example.com  06709ee8-ae32-48e4-a690-78276eb13825  LX     2048    100  running                          lx
 host-b.dev.example.com  594f44de-1bb4-417e-895b-c6bc129299e9  OS      256     10  stopped                          dhcpv6-a
host-c.prod.example.com  a4f7644b-4508-4721-ab73-57e3855fbfab  OS     2048     64  running  datasets.example.com  datasets.example.com
host-c.prod.example.com  261b3522-8ec6-40f3-ad48-e5882d8e0a98  OS     2048     10  running  graphite.example.com  graphite.example.com
host-c.prod.example.com  782c2631-85d1-44d4-9645-d5701fa42736  OS     2048     10  running   jenkins.example.com  jenkins.example.com

# dz start c11f26ba-0395-4723-9c2b-0d98931ce182

# dz shell d61b272f-2aea-4944-b8da-20c7b2c5331a
[Connected to zone 'd61b272f-2aea-4944-b8da-20c7b2c5331a' pts/6]
Last login: Thu Nov  6 07:55:53 on pts/2

 ✣ CORE.IO NS

[ core-ns 14.2.1 | https://github.com/skylime/mi-core-ns ]

[root@ns ~]#
```

better mi-repos

- document which mdata variables to use
- document which services will be provided
- if mdata not set: generate & mdata-put
- always create specific zfs filesystems for delegated datasets

demo

thx!
questions?

http://www.skylime.net
https://github.com/wiedi/
https://github.com/skylime/

mailto:sw@core.io
@wied0r on twitter
wiedi on irc