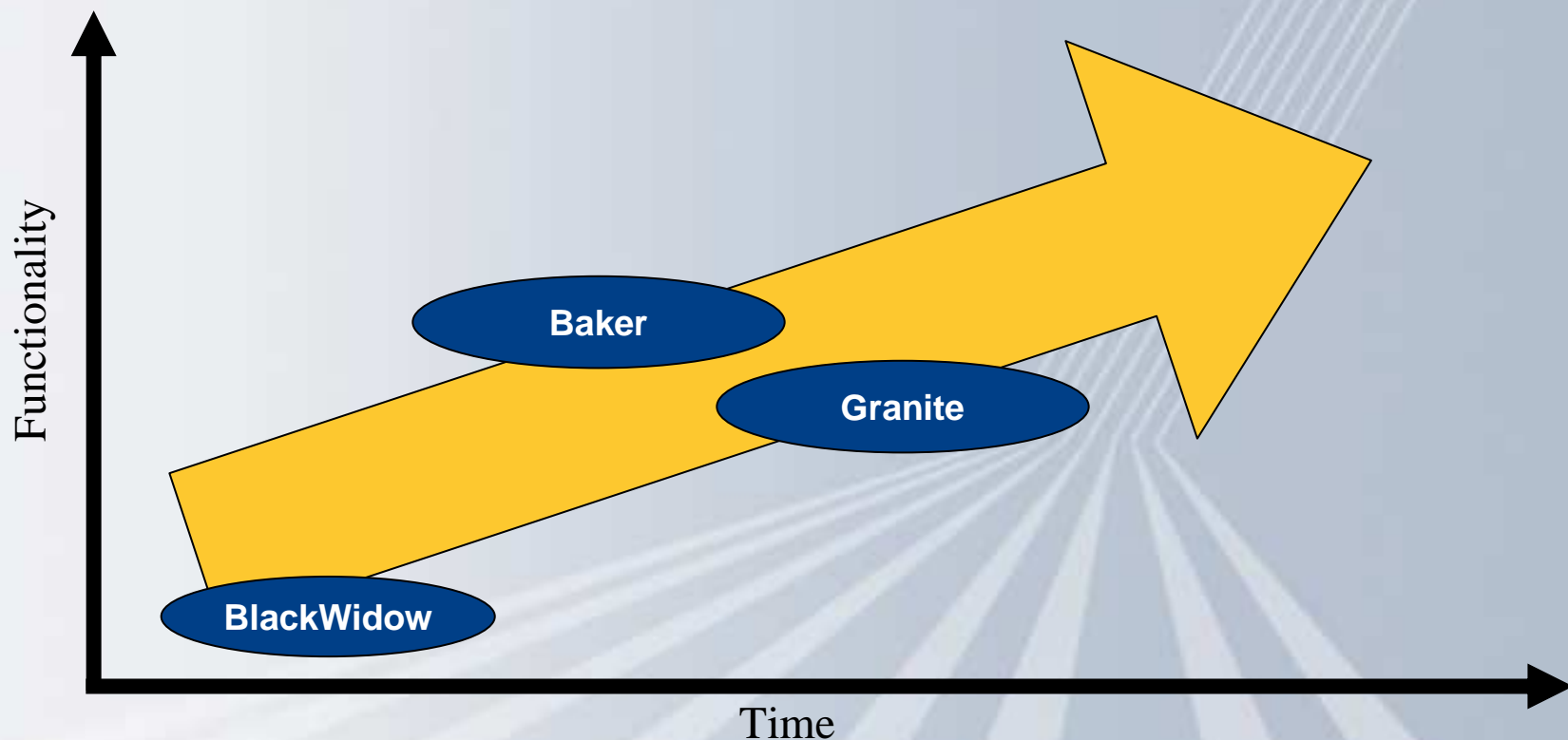# Mazama System Administration Infrastructure
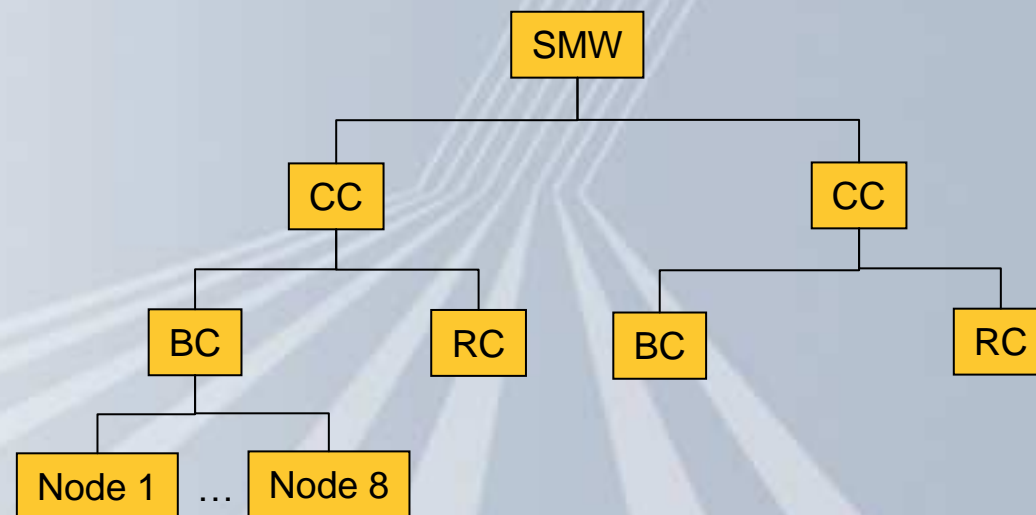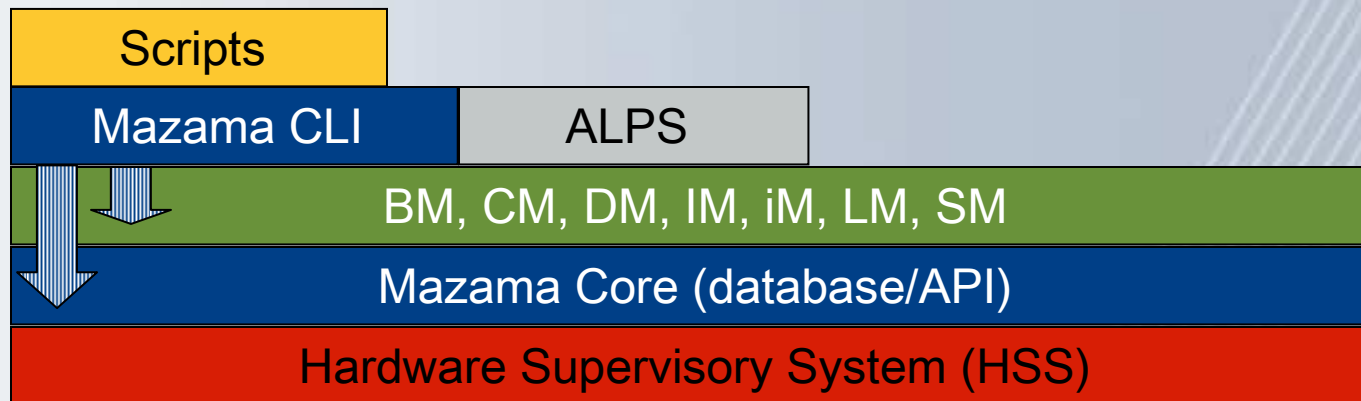
Jonathan Sparks

Introduction

- Next generation of system administration software.
- Interface for BlackWidow administration, integrated into Baker and beyond.

Mazama Architecture

Scripts

Mazama CLI | ALPS

BM, CM, DM, IM, iM, LM, SM

Mazama Core (database/API)

Hardware Supervisory System (HSS)

```
              SMW
         /          \
       CC            CC
      /   \         /    \
    BC    RC      BC     RC
   /  \
Node 1 … Node 8
```

Features

System Administration	System Governance
Command & Control Monitoring Image Management Node Attributes RAS	Configuration Provisioning Failure prediction & analysis Partition Management

Command and Control

- Day-2-day operations:
  - Boot (mzboot)
  - Dump (mzdump)
  - Power (mzpower)
  - Status (mzstate)

- Low-level state query operations.

Monitoring: logmanager

- Provides tools for analysis.
  - Searching
  - Payload analysis
  - Display

- Provides notification capabilities for messages.
  - Monitor and report on hardware and software components.
  - Monitor all logs and errors generated within the system.
  - Filter all messages generated within the cabinet and, based on site definable policies, are sent back to the SMW for storage and analysis.

Examples

- Continuous listing of all messages.

```
crayadm@nellie-smw:~> mzlslog -C1 --color -FtHpPm -t now-300
 Time        Hostname      Program       Priority    Message
…
```

- Summarize the top producers of messages/events.

```
crayadm@nellie-smw:~> mzlslog --summary -t0
```

component	program	level	count
---------	---------	---------	-----
	mzsmd	error	50755
	bwconsole	error	47738
c0-0c0s0n0	ec_console_log	info	42426
c0-0c0s0n3	ec_console_log	info	21006
r11-0c1s6p0	ec_hw_err_log	error	9869
	mzpower	error	9249
	bwconsole	info	8884
r11-0c0s4p3	ec_hw_err_log	error	8011

- Selectivley list all console messages from a known source.

```
crayadm@nellie-smw:~> mzlslog -tnow-3600 -p ec_console_log -c r16-1c1s5n0
…
```

Monitoring: Alarms

- Administrator can add and list custom alarm definitions.

- Definitions/actions can be:
  - Ignore: ignore messages/events in the logs.
  - Delete: delete messages/events from the logs.
  - Notify: Notify the administrator an event has happened.
  - Archive: Archive the log information.

Examples

- **Notify the admin of a fault condition**

```
crayadm@nellie-smw:~> mzlogdef -a -snotify -
  Nadmin@localhost -pec_l0_health -q"*VERTY health check
  fault*"
```

- **Ignore all ec_mesh_link_failed messages**

```
crayadm@nellie-smw:~> mzlogdef --create -s ignore -
  pec_console
```

- **Archive all messages to a file**

```
crayadm@nellie-smw:~> mzlogdef -c -sarchive -
  pec_hw_err_log -A ec_hw_err_log.txt
```

Image Management

- Definition: Image = Kernel + OS + kernel options.

- Install and upgrade compute node OS images.

- Allow site adaptations to compute node images.

Example:

```
crayadm@nellie-smw:~> mzimage
Images found (2):

ID:1  Name:dailyRM-3D042407 RefCount:0 Description:BW Release Master
      State:New 2007-04-24 12:51:10 Type:BW Release Master
      Ref Image ID:0 Creation Time:2007-04-24 12:51:10
      Image Path:file:///opt/mazama/RM/dailyRM-3D042407/
      Huge Mem Page Size:65536


ID:2  Name:dailyPM-3D042407 RefCount:0 Description:
      State:New 2007-04-25 11:57:13 Type:BW Partition Master
      Ref Image ID:1 Creation Time:2007-04-24 12:51:52
      Image Path:file:///opt/mazama/PM/dailyPM-3D042407/boot-
  dailyPM-3D042407-2.cpio
      Huge Mem Page Size:65536
```

Cray Inc. Confidential

Partition Management

- Definition: Partition = Number Nodes + Image.

- Aggregator for nodes and operations, such as boot, shutdown and dump.

- Allows the administrator to setup regions within the system running different images.

Example

```
crayadm@nellie-smw:~> mzpartition
Partitions found: 1

System attributes IP address: 192.168.0.0 Netmask: 255.255.0.0 Gateway:
   192.168.0.254

ID:1 (Default) Name:daily SysID:0  Image ID:2 Ideal Count:0 Min Count:4
        Base address: defaults Netmask: defaults Gateway: defaults
        Scripts
                Pre:  not set
                Post: not set


Partitions found: 1
```
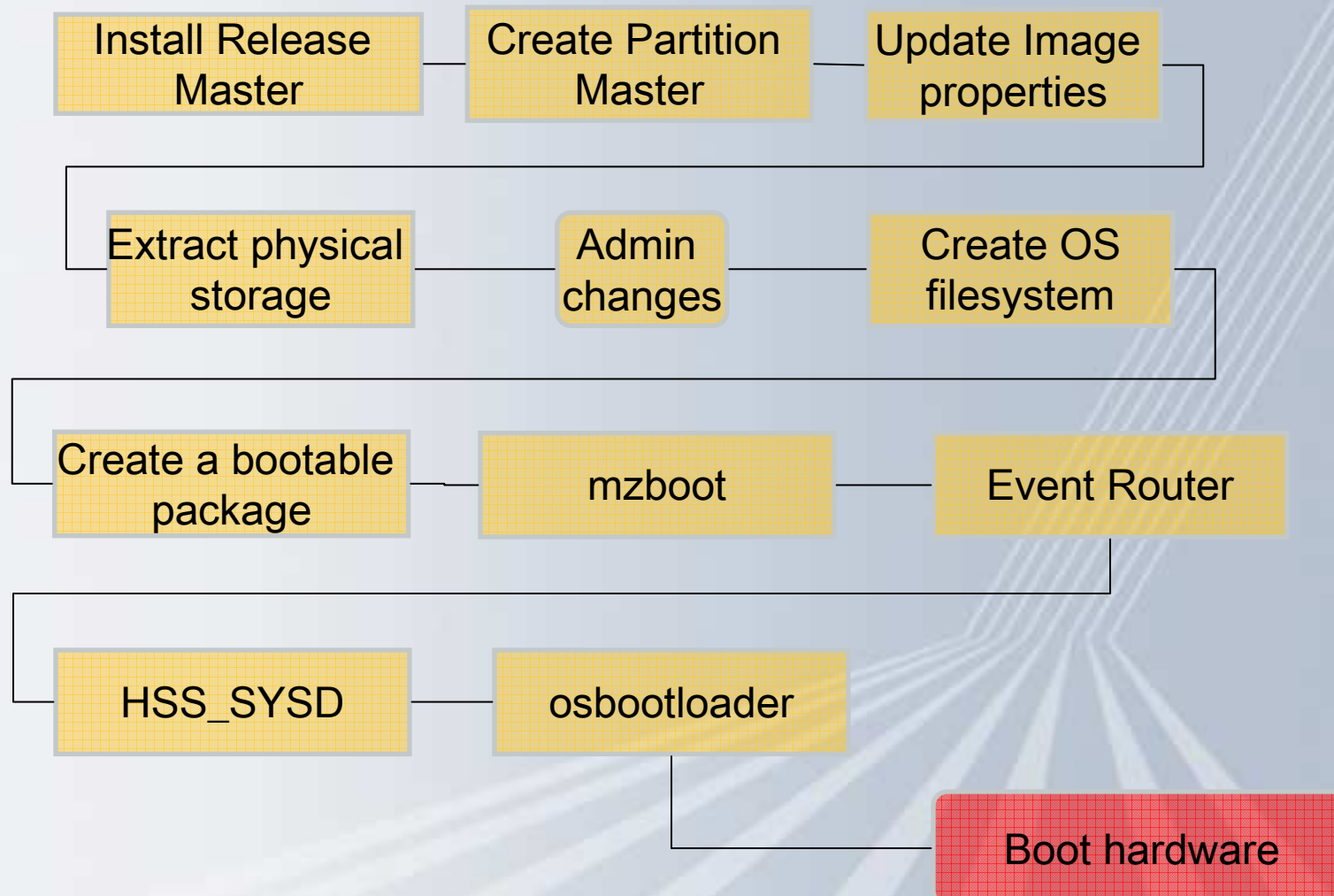
Node Attributes

- Describes the smallest schedulable entity within the BlackWidow system.

- Used by ALPS for scheduling.

- Aggregation of many system attributes in one UI.

Example

```
crayadm@nellie-smw:~> mzattr --summary
Type      Attr             Count   Value
------    ---------------  ------  -------------------------------
NODE      STATE            4       4 -> "BOOTING"
NODE      ADMIN_STATE      4       0 -> "UNAVAILABLE"
NODE      PARTITION_ID     4       1
NODE      MEM_SZ           4       32768 MB
NODE      HUGE_MEM_PSZ     4       65536 KB
NODE      HUGE_MEM_NPAGES  4       480
NODE      NSOCKETS         4       4
NODE      NODE_TYPE        4       1 -> "BlackWidow compute"
NODE      MACHINE_TYPE     4       0 -> "BlackWidow"
NODE      OS_TYPE          4       0 -> "BlackWidow"
NODE      OS_VERSION       4       FIX THIS OS VERSION!
NODE      KERNEL_VERSION   4       SLES10-042707
NODE      MPI_LABEL_0      4
NODE      MPI_LABEL_1      4
NODE      MPI_LABEL_2      4
NODE      MPI_LABEL_3      4
SOCKET    TYPE             16      0 -> "BlackWidow CPU"
SOCKET    SPEED            16      0 MHz
SOCKET    NCORES           16      1
SOCKET    MEM_SZ           16      32768 MB
SOCKET    MEM_SPEED        16      667 MHz
```

Use Case: Booting

```
Install Release          Create Partition          Update Image
    Master                    Master                  properties
```

```
Extract physical            Admin                   Create OS
    storage                 changes                 filesystem
```

```
Create a bootable
    package                  mzboot               Event Router
```

```
   HSS_SYSD               osbootloader
```

```
                                           Boot hardware
```

Questions?