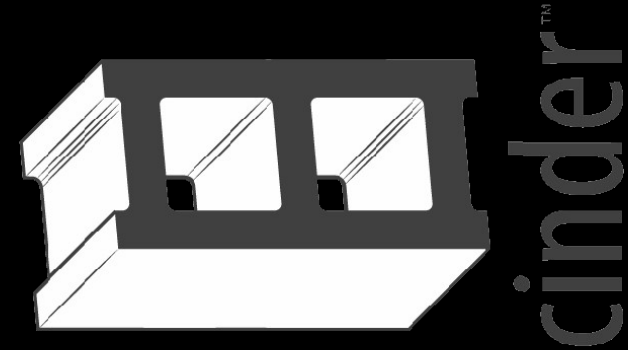


# OpenStack Cinder

## Deep Dive

**Havana Release**

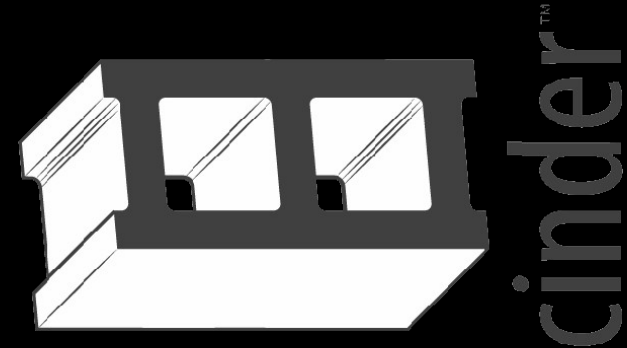
**Presented by: Avishay Traeger**



# Avishay Traeger

IBM Research - Haifa

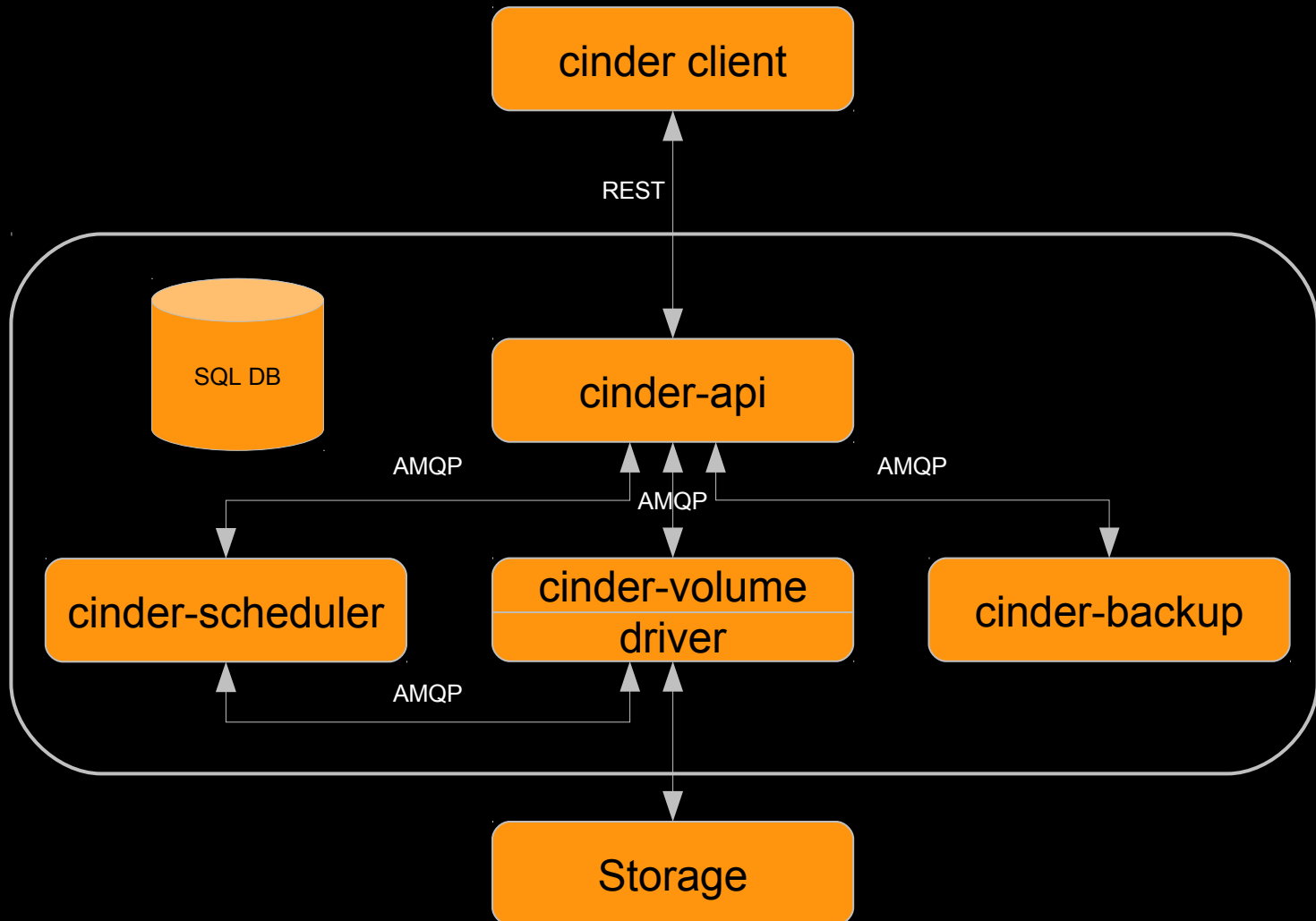
Cinder Core Team



# Cinder Overview

- Project exists since Folsom release, spun off Nova-volume
- Cinder manages block storage
  - Not object storage (Swift)
  - Not file-level storage (Manila)
  - Volumes attach to VM instances
  - Boot from volume
- Volumes have a life-cycle independent of VM instances

# Architectural Overview



# Cinder API

- Volume create/delete/list/show
  - Create from image, snapshot, volume
- Snapshot create/delete/list/show
- Backup create/restore/delete/list/show
- Volume attach/detach (called by Nova)
- Volume types
- Quotas

# Volume Types

- Admins can create tiers of storage
  - Specify requirements from storage
- Users can specify a tier when creating a volume

The screenshot shows a 'Create Volume' dialog box with the following fields and information:

- Volume Name:** A text input field containing 'web-volume'.
- Description:** A text input field containing 'Additional information here...'.
- Type:** A dropdown menu with a list of storage tiers: 'tier1', 'tier1b', 'tier2' (highlighted with a blue border), and 'tier2c'.
- Description:** A text block stating 'Volumes are block devices that can be attached to instances.'
- Volume Quotas:**
  - Total Gigabytes (350 GB):** A progress bar showing 650 GB Available.
  - Number of Volumes (3):** A progress bar showing 7 Available.
- Buttons:** 'Cancel' and 'Create Volume' buttons at the bottom right.

# Cinder Scheduler

- Chooses back-end to place a new volume on
- Configurable plugins for scheduler
  - Simple
  - Chance
  - Filter
- Most common is the filter scheduler
  - Has plug-able filters & weights

# Filter Scheduler: Example Flow

- Drivers continuously report capabilities and state
- Scheduler starts with list of all back-ends
- Filters out unsuitable back-ends
  - Insufficient free space
  - Insufficient capabilities
- Sorts according to weights (e.g., available space)
- Returns best candidate



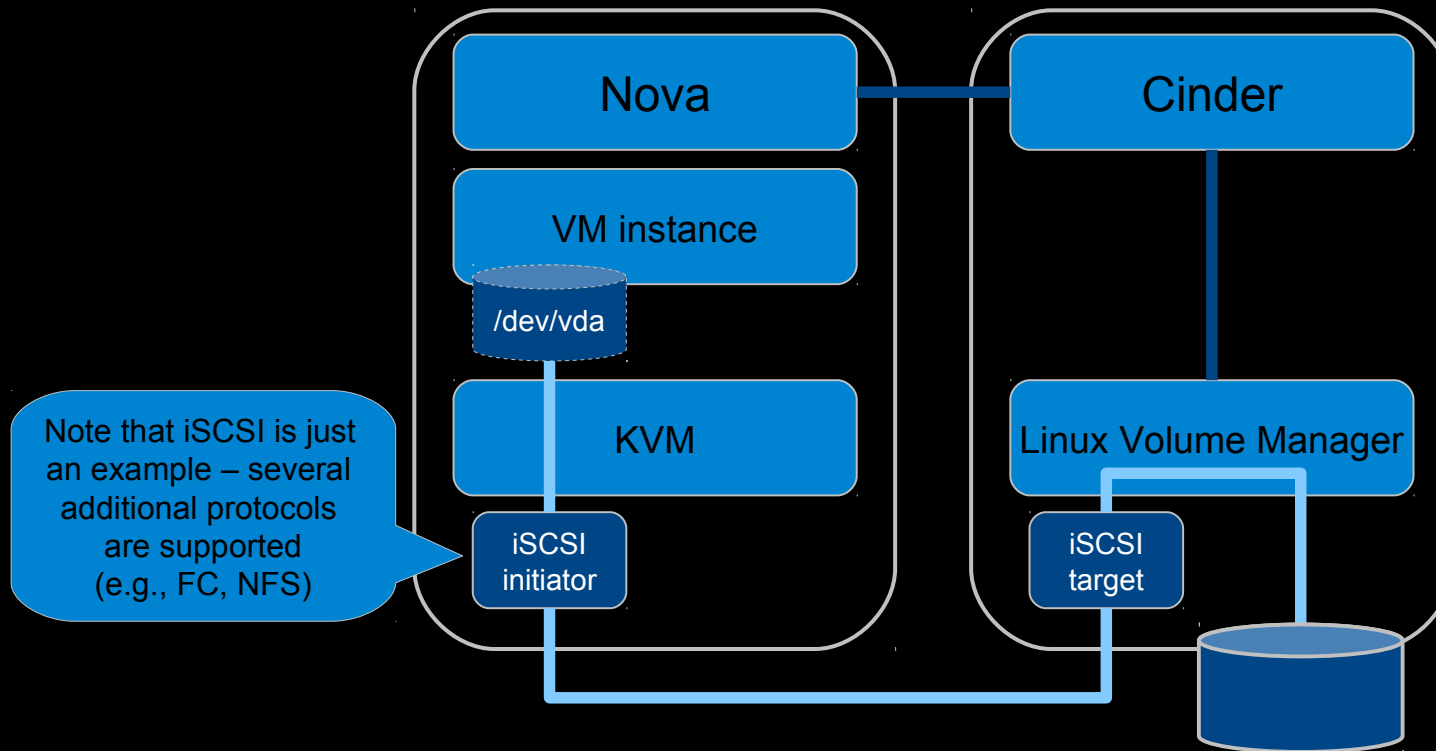
# Cinder Volume

- Manager contains generic code
  - e.g., High-level flow, DB & quota updates
- Drivers contain back-end specific code
  - Linux LVM
  - Storage controllers from various vendors
  - Distributed file systems
- Admin can run multiple cinder-volume instances
  - Each able to manage multiple back-ends
- Each back-end is generally configured to interact with one storage pool
- Multi-threading

# Cinder Backup

- Cinder can back up volumes
- Backups should allow recovery from
  - Volume data corruption
  - Storage failure
  - Site failure (provided that backups are safe)
- Plug-able driver architecture
  - OpenStack Swift
  - IBM Tivoli Storage Manager
  - Ceph

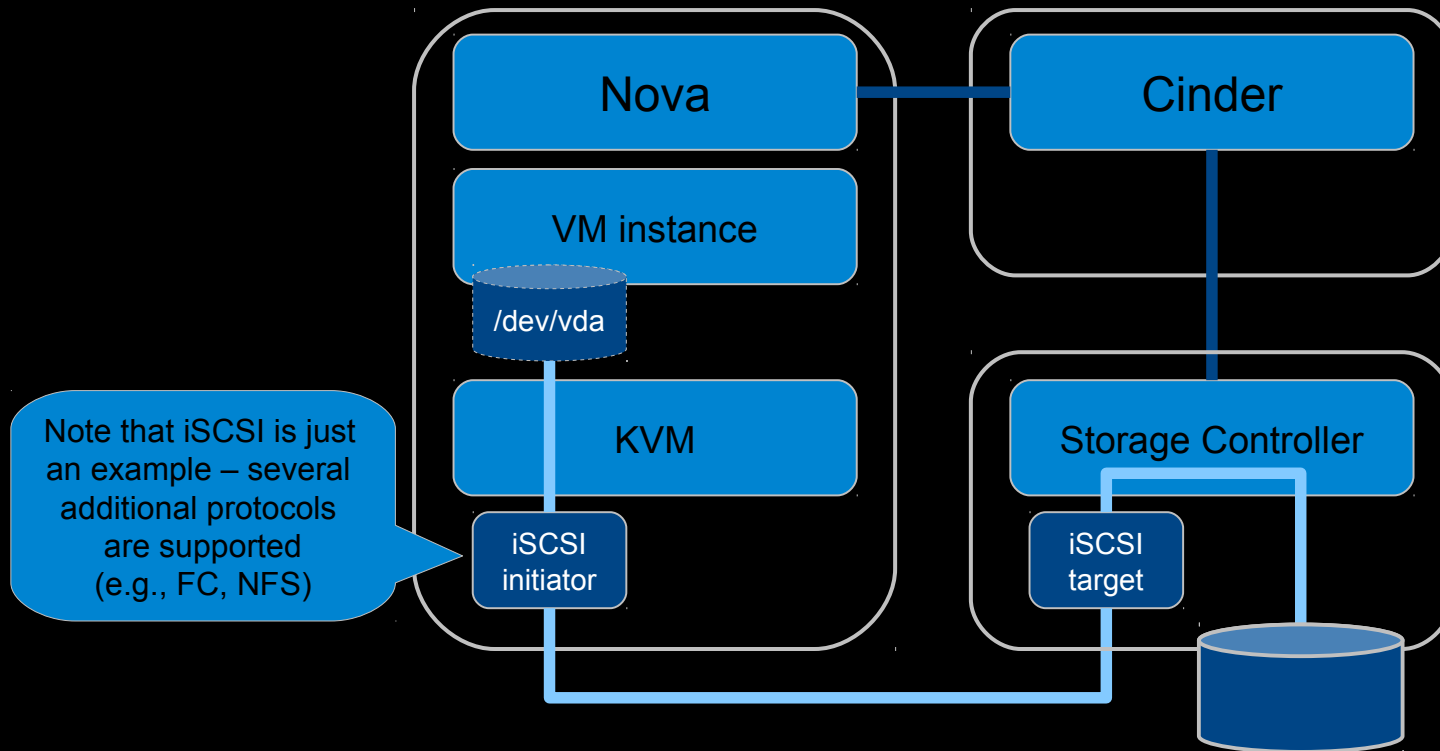
# Example: High-Level Attach 1



## Legend

- Persistent volume control
- Persistent volume data

# Example: High-Level Attach 2



## Legend

- Persistent volume control
- Persistent volume data

# Example Flow: Attach Volume

- Nova calls Cinder via its API, passing connection information
  - e.g., host name, iSCSI initiator name, FC WWPNS
- cinder-api passes message to cinder-volume
- Manager does initial error checking and calls volume driver
- Volume driver does any necessary preparation to allow the connection
  - e.g., give the nova host permissions to access the volume
- Volume driver returns connection information, which is passed to Nova
  - e.g., iSCSI iqN and portal, FC WWPNS
- Nova creates the connection to the storage using the returned information
- Nova passes the volume device/file to the hypervisor

# New Major Features in Havana

- Disk encryption
  - Encryption is done by Nova using dm-crypt, Cinder is made aware of encryption keys
- Volume Migration
  - Admin interface: `cinder migrate <volume-id> <target>`
  - Check if storage can migrate the volume
  - If not, create a new volume
    - If original volume is detached, Cinder server attaches both and runs 'dd'
    - If original volume is attached, Nova performs the copy (KVM-only in Havana)
- Volume rate limiting
  - Allows rate limiting per volume
  - Can be enforced by Nova (KVM-only in Havana) or by storage
- Extend Volume
- Transfer volume ownership
- Scheduler hints

# Drivers in Havana

- Coraid (AoE)
- Dell Equallogic (iSCSI)
- EMC VMAX/VNX (iSCSI)
- GlusterFS (GlusterFS)
- HP 3PAR (iSCSI/FC)
- HP LeftHand (iSCSI)
- Hitachi HUS (iSCSI)
- Huawei HVS/T-series/Dorado (iSCSI/FC)
- IBM DS8000 (FC)
- IBM GPFS (GPFS)
- IBM Storwize family/SVC (iSCSI/FC)
- IBM XIV (iSCSI/FC)
- Local disk partitions
- LVM (iSCSI)
- NetApp (iSCSI/NFS)
- Nexenta (iSCSI)
- NFS (NFS)
- RBD (Ceph)
- Scality SOFS (scality)
- Sheepdog (sheepdog)
- Solaris (iSCSI)
- SolidFire (iSCSI)
- Windows Server 2012 (iSCSI)
- Zadara (iSCSI)

# Features in the Works for Icehouse

- Read-only volumes
- Volume retype
- Volume replication
- ACLs for volumes
- Multi-attach
- API rate limits
- Volume import



# Thank you!

- Getting started with Cinder:  
<https://wiki.openstack.org/wiki/Cinder>
- Source code:  
<https://github.com/openstack/cinder>

Contact: [avishay@il.ibm.com](mailto:avishay@il.ibm.com)

Thanks to Mike Perez for slide flow inspiration!