napp-it

ZFS Storage Server
Build examples
Content:

1. Checklist

2. Cases
   2.1 Silent Cases Silverstone CS 380 or Fractal Design Define R5
   2.2 other small Storage cases
   2.3 Expanderless 19" cases for 3,5"
   2.4 Expanderless 19" cases for 2,5"
   2.5 Ready to use system Zstor Cube SuperMicro barebone 24 x 2.5"
   2.6 19" cases with expander up to 90 x 3.5"
   2.7 19" cases with expander up to 72 x 2.5"
   2.8 Flexible 19" case, different backplanes

3. Mainboards
   3.1 Low Power, Low Cost Celeron/Xeon-D
   3.2 Low Power, Low Cost Socket 1151
   3.3 Full featured single Xeon 2011-R3
   3.4 Full featured dual Xeon 2011-R3
   3.5 Next gen socket 3647 mainboards
      SM X11 SPH-nCTPF, X11 SPH-nCTF
      1TB RAM, SAS+10G, max 10 NVMe

4. Add ons, HBAs

5. Example: Small workgroup/ lab storage
6. Example: Silent workgroup/ office storage
7. Example: High capacity Sata storage
8. Example: High capacity/ HA SAS storage
9. Example: Petabyte Storage
10. Z-RAID Appliance redundancy/ failover
11. 40G Ethernet connectivity examples
12. more docs
1. Checklist

Basically you can use any desktop or serversystem as a webmanaged ZFS NAS appliance.
The minimal hardware demand for a 64 bit Solarish like OmniOS, OpenIndiana or Oracle Solaris is

- 2 GB RAM (ECC or non-ECC) for any Poolsize – but you should not enable dedup with low RAM
- a 64 bit CPU (AMD or Intel). If you can decide, prefer frequency over number of cores.
- Sata (AHCI mode)
  - a nic (Realtek but prefer Intel)
  - a 32GB bootdisk
  - disks for a datapool, should be at least two disks for a mirror

These are the minimal demands. If you care about your data and want a better performance than
pure disk performance, then you should add

- ECC RAM. This is important as a RAM failure is the only type of failures ZFS cannot protect against.
  But even without ECC, ZFS is more secure than older filesystems without ECC as it protects against all problems
  in the chain controller <-> disk, powerloss problems and undo of unwanted modifications due snaps.
If you have the option, use always ECC for any storage system. From hardware this requires ECC support from
CPU and mainboard. Use an Intel mainboard with a serverchipset like C232 or C236 and a ECC capable CPU
up from an Intel G4400 (socket 1151) or a Xeon (any socket).

- Add more RAM. While Solarish works with 2 GB, more RAM is used as a readcache that can dramatically
  improve readperformance. With more RAM you can achieve that > 80% of all reads are delivered from RAM.
- Prefer solutions with 10G Intel nics as they have the best driver support and performance.
  You can enable bridging on your server. Your NAS acts then like a 10G switch over your 1G/10G nics.
- Add an HBA with an LSI chipset, preferred with a raidless IT firmware (ex LSI 9207) if you need either
  more ports or if you want to use it in an All-In-One setup where you can virtualize storage beside other VMs.

Such systems can be ordered quite cheap if you check offers from Dell example the PowerEdge T20 or T130
or HP systems like the HP G8 Microserver or the ML10 Gen9. Main problem of these systems are a limited
expandability regarding PCI-e slots, disk bays or available CPU options.

If you want to overcome these limitations, you can built your own setup or buy built to order systems that are
assembled by a distributor or vendor as a ready to use system.

I will concentrate on mainboards from SuperMicro as this is the number one option for
ZFS storage with a complete storage product line with the following use cases

1. Silent Solutions with up to 8 x 3,5" diskbays, 8 x 2,5" backplane and 2 x 2,5" bootdisks and optionally some
   PCI-e NVMe disks. Such a system can be placed near the desktop in an office or team environment

2. 19" cases for expanderless SAS/Sata storage with up to 24 disks
   With 10TB Sata disks you can go up to over 200 TB raw storage or with a 24 x 2,5" case you can
   build high performance SSD only storage with up to 90 TB when using the 3,8 TB Samsung SSD.

3. 19" cases for SAS solutions. With a 90 x 3,5" bay SuperMicro case and 8 TB SAS disks, you can achieve
   around 500 GB per case. For a Petabyte you need two cases. There are also options with 72 x 2,5" SAS disks.

With an expander you can use Sata disks as SAS is tunnelling the Sata protocol but this is not suggested
in a production environment. Professional storage vendors deny support for such a config.
2. Silent/ low budget cases

2.1 Silverstone CS 380

Case suitable for ITX, uATX and ATX mainboards
with 3 x 120mm fan and 8 x Sata/ SAS dualpath hotplug backplane

possible Addons:
2,5" Backplane (8 x Sata SSD in 2 x 5,25" Slots)
ex SuperMicro Sata Mobile Rack M14TQC (or other Sata/SAS mobile racks)
http://www.supermicro.com/products/accessories/mobilerack/CSE-M14TQC.cfm

Use case:
SoHo low power/ low noise system with up to 16 hotplug disks ex for a fast ZFS pool
from up to 8 SSDs (VM datastore) and a second 8 disks based pool for filer and backup
up to low budget HA systems with two virtualized storage heads under ESXi
each with an SAS HBA in pass-through mode to dualpath SAS disks)

Mainboard:
For a low budget NAS (sub 500 $/Euro) with socket 1151, G44xx or G45xx CPU with a small M.2
for OS + 8 onboard Sata https://www.supermicro.nl/products/motherboard/Xeon/C236_C232/X11SSH-F.cfm
also available with additional LSI SAS HBA for AiO systems with ESXi + local datastore + Slog with an Intel M.2
Optane 800P and optionally10G onboard.

A perfect mainboard would be one of the SuperMicro X10 SDV line
with 10G networking and LSI HBA with16 x SAS/Sata onboard (sub 1000 $/Euro system)
or with more powerful CPUs, https://www.supermicro.nl/products/motherboard/Xeon3000/#1667
2.1b A silent case: Fractal Design Define R5 (black, white, titanium)

From vendor specifications

“The Define R5 case reaches the highest level of silent computing through strategically placed dense sound-absorbing material, ModuVent™ fan vent covers and finely tuned Dynamic Series fans.

- ATX, Micro ATX, Mini ITX motherboard compatibility
- 7 expansion slots
- 2 - 5.25” bays (removable)
- 8 - 3.5” HDD positions (can also accommodate 2.5” units); 2 - 2.5” dedicated SSD unit positions
- 4 - ModuVent™ plates - three in the top and one in the side
- 9 - fan positions (2 Fractal Design Dynamic GP14 140mm fans included)
- Filtered fan slots in the front and bottom
- CPU coolers up to 180mm in height
- ATX PSUs up to 190/170 mm with a bottom 120/140mm fan installed; when not using any bottom
- fan location longer PSUs up to 300mm can be used
- Graphics cards up to 310 mm in length with the top HDD cage installed; with the top cage removed, graphics cards up to 440 mm in length may be installed
- 20 - 35 mm of space for cable routing behind the motherboard plate
- Velcro straps included for easy cable management
- Front door can switch opening direction via dual mounting system
- Left side panel features Quick Release System for easy access and provides a secure closure of side panel
- Right side panel features smart captive thumbscrews so no thumbscrews are lost
- Colours available: Black, Titanium (black case, titanium front panel), White
- Case dimensions (WxHxD): 232 x 451 x 521mm
- Case dimensions - with feet/screws/protrusions: 232 x 462 x 531mm
- Net weight: 10.7 kg

possible Addons: 2.5” Backplane
SuperMicro Mobile Rack M28SACB-OEM

1. 8x 2.5” Hot-swap SAS3/SATA3 HDDs
2. Overheat LED and Alarm
3. Drive Activity / Failure LED
4. 2x 5.25” Drive Bay Enclosure
5. 2x Mini SAS HDD Connectors
6. Fan-less Subsystem
2.2 other small cases

U-NAS ex NSC 400/600/800 (mini-ITX)
www.u-nas.com

Fractal Design Node 804 (up to uATX)

Micro ATX and Mini ITX motherboard compatibility
8 - 3.5″ HDD positions
2 - 2.5″ dedicated SSD unit positions
2 - Extra positions for either 3.5″ or 2.5″ drives
5 expansion slots
1 additional space in the front for a 12.7mm slim/slimline ODD
10 - Fan positions (3 x 120mm Silent Series R2 fans included)
Filtered fan slots in front, top and bottom
CPU coolers up to 160 mm in height
PSU compatibility: ATX PSUs up to 260 mm deep
Graphics card compatibility: Graphics cards up to 320mm in length.
Graphics cards up to 290 mm in length may be installed if a fan is installed in the lower position in the front.
Velcro strap for easy cable management
Clear Window side panel included
Colors available: Black
Case dimensions (WxHxD): 344 x 307 x 389 mm
Net weight: 6 kg
Package dimensions (WxHxD): 370 x 468 x 412 mm
Package weight: 7.7kg

Silverstone CS381 (expected Q1 2019)
uATX, FlexATX, 8 x 3,5" hotswap + 2 x 3,5"
https://www.servethehome.com/silverstone-cs381-8-bay-matx-case-shown/

Silverstone CS 380 (available)
ATX, uATX, FlexATX, 8 hot swapable Dualpath SAS
(perfect solution for a HA Cluster in a Box)
2.3 Expanderless 19” cases with 12 to 36 bays for 3.5” Sata and SAS disks
Build to Order system

Expanderless Options:

8 disks Supermicro, available but I would prefer 12 bay versions due same size.

12 disks SuperMicro
ex SuperChassis 826BA-R920LPB, 920 Watt PSU, 3 x miniSAS connector

16 disks SuperMicro
SuperChassis 836BA-R920B, 920 Watt PSU, 4 x miniSAS connector
SuperChassis 836A-R1200B, 1200 Watt PSU, 4 x miniSAS connector

24 disks SuperMicro
SuperChassis 846A-R900B, 900 Watt PSU, 6 x miniSAS connector
SuperChassis 846A-R920B, 920W Gold PSU, 6 x miniSAS connector
SuperChassis 846A-R1200B, 1200 Watt PSU, 6 x miniSAS connector

36 disks SuperMicro
Superchassis SC847BA-R1K28LPB, 1280W with a backplane without expander (iPASS connectors)

Attention: These 19" cases are intended to use in a serverrom.
For office/ near desktop use, these systems are way too loud
2.4 Expanderless 19" cases with 24 bays for 2.5" Sata and SAS disks
Build to Order system

24 x 2,5" expanderless:
SuperMicro SuperChassis 216BA-R920LPB, 1200W PSU, 6 x miniSAS, optional 2 x Sata bootdisk

The above Case can be combined with any mainboard and HBA. It is intended for a serverroom as it is quite loud

2.4 SuperChassis 216BA-R920LPB barebone
Prebuild Barebone system, you only need to add CPUs and RAM

Key Features

1. 920W Redundant High-efficiency Platinum Level Power Supply
2. 2U High Density 2.5" HDD Chassis 24x 2.5" Hot-swap SAS/SATA HD bays
3. Mini-I-pass (SFF 8087) Connectivity for clean cable routing
4. 7x Low-profile Expansion Slot
5. Optional rear hot-plug 2x 2.5" HDD drive bays

Integrated Board

Key Features

1. Dual socket R3 (LGA 2011) supports Intel® Xeon® processor E5-2600 v4/ v3 family; CPU up to 9.6GT/s
2. Up to 2TB ECC 3DS LRDIMM, up to 6x 2400 MHz; 16x DIMM slots
3. 1 PCIE-3.0 x16, 6 PCIE-3.0 x8 (slot 1-3 occupied by controllers)
4. Dual 10GBase-T LAN w/ Intel® X540
5. 24x 2.5" Hot-swap SAS3/SATA3 direct attached drive bays; 2x 2.5" optional hot-swap drive bays (rear)
6. SAS3 via 3x LSI 3008 controller, IT mode
7. Server remote management, IPMI 2.0 / KVM over LAN / Media over LAN
8. 3x 8cm hot-swap redundant PVWM fans
9. 920W Redundant Power Supplies Platinum Level (94%)
2.5 Zstor Cube
Ready to use NAS/SAN system with napp-it preinstalled

http://zstor.de/de/zstor-gs-cube8-mini-cube-storage-server.html

<table>
<thead>
<tr>
<th>Product</th>
<th>Mini Cube 8x 3.5''/2.5'' Bay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>One INTEL® Xeon® Processor E3-1230v6 3.5GHz, 4 cores and 8 threads</td>
</tr>
<tr>
<td>Memory</td>
<td>4x 16GB DDR4 UDIMM ECC 2400 MT/s total 64GB</td>
</tr>
<tr>
<td>Chipset</td>
<td>Intel® C236 chipset, Integrated Video Controller VGA, Integrated BMC with KVM</td>
</tr>
<tr>
<td>Host Bus Adapter</td>
<td>Onboard Broadcom / LSI 3008 SAS/SATA HBA</td>
</tr>
<tr>
<td>Connection</td>
<td>1x VGA, 2x USB 2.0 ports side, 1x USB 3.0 front, 2x USB 2.0 ports back 2x USB 3.0 ports back, 1x IPMI RJ-45 Ethernet Remote Management, 2x 10GE RJ-45 Ethernet</td>
</tr>
<tr>
<td>Network</td>
<td>2x onboard 1 GBE INTEL i219 RJ-45 1x onboard iPMI RJ-45 incl. Remote Management with KVM optional 1x PCI-E 10GEE INTEL 02596ES Dual SFP+ optional 1x PCI-E 10 GBE INTEL X540 Dual RJ-45</td>
</tr>
<tr>
<td>PCI-E Slots</td>
<td>1x HH-HL x16 mech., x8 elec. 1x HH-HL x8 mech., x8 elec.</td>
</tr>
<tr>
<td>Disk Bay</td>
<td>8x 3,5''/2.5'' HDD/SSD Hot-swap SATAIII 1-1.7TB per Disk, 1x 2.5'' SSD internal for OS Boot up to 96TB on a small foot print, mixture of HDs and SSDs possible</td>
</tr>
<tr>
<td>Operating System</td>
<td>Omnis 2FS File System with napp-it GUI, other operating systems on request</td>
</tr>
</tbody>
</table>
2.6 Enclosures with Expander for up to 90 3.5" disks  
Build to Order system

2.6.1 SuperChassis 847BE2C-R1K28LPB (12G)

**Key Features**

1. 36x (24 front + 12 rear) 3.5" hot-swap SAS/SATA drive bays supporting SAS3/2 or SATA3 HDDs with 12Gbps throughput
2. Optional rear 2.5" removable HDD
3. Redundant 1280W Platinum Level (1+1) power supplies with PMBus
4. 7x low-profile expansion slots; 7x 8cm (middle) hot-swap cooling fans and adjustable air shroud
5. E1C: Single SAS3 (12Gbps) expander backplane; E2C: Dual SAS3 (12Gbps) expanders backplane
6. Mini SAS HD (SFF 8643) connectivity on backplane

2.6.2 SuperChassis 946ED-R2KJBOD (90 disks Jbod, up to 500 TB per enclosure)
2.7 Enclosures with Expander for up to 72 2,5” disks  
Build to Order system

Supermicro SuperChassis 417E26–R1400LPB

72 x 2,5” disks with expander

2.8 Variable enclosure Chenbro RM417  
an ultra flexible case concept with different and selectable 3,5” or 2,5” backplanes with or without expander

Backplanes: 6Gb/s mini-SAS expanderless or mini-SAS or with expander on-board  
Disks: up to 36 x 3,5” or 72 x 3,5” or like in above example 24 x 3,5” + 24 x 2,5”
3. SuperMicro Mainboard options with current chipsets
Build to Order system

3.1 Low power, low cost System X10SDV-2C-7TP4F

Attention:

The board works with ESXi and OmniOS, I had problems with Solaris 11.3
Drivers for the X550 are on last tests (available from OmniOS 151019)

The board supports vt-d, so you can use it in an AiO setup

The boards is also available as
X10SDV-4C-7TP4F  (4C/8HT)
X10SDV-7TP4F   (8C/16HT) and
X10SDV-7TP8F   (16C/32HT)
The board is based on socket 1151 what limits RAM to 64GB but it offers more flexibility regarding CPU that can range from a cheap G4400 up to a Xeon, all with ECC and vt-d support. This board offers 10G and an onboard 12G SAS controller.

The board is also available with 1G and without the SAS controller.

Drivers for the X550 are on last tests (available from OmniOS 151019)

You can install OmniOS currently not from an USB installer stick.

Setup options:
- use the preconfigured napp-it systemimage
- use an Sata DVD drive to setup OmniOS manually

Needed bios settings
- Advanced > Boot: Install Windows 7 USB support: enabled
- Boot > Boot mode: Legacy
3.3 SuperMicro X10 SRL-F Single Xeon

This board offers more flexibility regarding RAM, nics, controllers and NVMe adapters as it offers up to 1TB RAM and 7 PCI-e slots

This board is also available in editions with 10G or 12G SAS onboard

<table>
<thead>
<tr>
<th>Key Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Single socket R3 (LGA 2011) supports Intel® Xeon® processor E5-2600 v4T/v3 and E5-1600 v3 family</td>
</tr>
<tr>
<td>2. Intel® C612 chipset</td>
</tr>
<tr>
<td>3. Up to 1TB ECC 3DS LRDIMM, up to DDR4- 2400 MHz; 8x DIMM slots</td>
</tr>
<tr>
<td>4. 7x PCI-E slots total: 2 PCI-E 3.0 x8, 2 PCI-E 3.0 x8 (in x16), 2 PCI-E 3.0 x4 (in x8) or 1 x8 + 1 x0 (auto-switch), &amp; 1 PCI-E 2.0 x4 (in x8)</td>
</tr>
<tr>
<td>5. Intel® i210 Dual port GbE LAN</td>
</tr>
<tr>
<td>6. 10x SATA3 (6Gbps) via C612</td>
</tr>
<tr>
<td>7. 1x VGA, 2x COM, 1x TPM</td>
</tr>
<tr>
<td>8. 4x USB 3.0 ports, 8x USB 2.0 ports</td>
</tr>
<tr>
<td>9. 2x SuperIO with built-in power</td>
</tr>
</tbody>
</table>
3.4 SuperMicro X10 DRH-iT Dual-Xeon

This board offers the most options as you can add up to 2 TB RAM with 7 PCI-e slots and 10 GB onboard. This series is also available as X10DRH-CLN4 with onboard 12G SAS 3008 controller but without 10G then.
3.5 SuperMicro X11 SPH-nCTPF (SFP+) and X11 SPH-nCTF (10G-Base T)

This single Xeon socket 3647 systems are my intended main future storage platform as they offer lanes for up to 10 NVMe (2 OcuLink and M.2 onboard, up to seven more via NVMe HBA), onboard 10G and LSI 3008 and max 1TB ECC RAM at a price of around 500 Euro without CPU and RAM.

I was able to install the newest OmniOS 151024. Nic (10Gbase-T and SFP+) worked as well as the LSI and the NVMe OcuLink ports with two Optane 900P. But USB give troubles.

With an initial bios setting I was not able to install from USB or with an USB keyboard.
I needed the following Bios setting to make it working

- Advanced Boot feature: enable Windows 7 USB Keyboard support=yes
- Chipset Southbridge:
  Legacy USB support=enable
  XHCI=disable

After booting of OmniOS, i got the console messages:
xhci0: failed to map device registers: -1

An additional problem was that caps-lock is inverted.
Caps Lock off=uppercase
3.6 SuperMicro X11SDV-4C-TLN2F mITX

This single Xeon mITX 60W system is my upcoming suggestion for an upperclass mITX storage system. It comes with a 4core/8 threads Intel Xeon, max 512GB ECC RAM and 2 x 10G Base-T onboard. You can connect 4 x Sata and one U.2 NVMe at the Oculink port (or alternatively 4 x Sata via Oculink). An additional 4x PCI-e port can hold an additional SAS adapter for more disks.

The same board is available in different editions with more nics or cores.
see https://www.servethehome.com/supermicro-x11sdv-4c-tln2f-review-with-intel-xeon-d-2123it/

### Key Features

1. Intel® Xeon® Processor D-2123IT, 4-Core, 8 Threads, 60W
2. System on Chip
3. Up to 256GB Registered ECC RDIMM or 512GB ECC LRDIMM, DDR4-2133MHz in 4 DIMM slots
4. 1 PCI-E 3.0 x8
   1 PCI-E 3.0 x4 NVMe Internal Port via OCuLink
5. Dual LAN with 10GBase-T with Intel® X557
6. Up to 8 SATA3 (6 Gbps) ports; RAID 0,1,5,10
   4 SATA ports via OCuLink (or PCIe3.0 x4 for NVMe)
7. 2 USB 3.0 ports (rear)
   2 USB 2.0 ports (via header)
8. 12V DC or ATX Power Input
4. Add-Ons

L2ARC readcache device. This is an SSD to extend the rambased readcache. As RAM is much faster and the L2Arc requires a few percent RAM for organisation, first try to add RAM if you need more Cache. If you add an L2Arc it should be much faster than your pool, even sequentially as you need to write to the pool and the cache.

If you need powerloss protection on writes with enabling sync, you should add an Slog. An Slog must offer very low latency, very high iops and internal powerloss protection, Good examples are Intel S3700 or Intel P3600/3700 or the new P3608.

Logilink mounting brackets for internal 2,5" SSDs

Mobile Racks/ backplanes for 5,25" bays
https://www.supermicro.nl/products/chassis/mobileRack/
http://www.icydock.com/goods.php?id=175

Adapter for 2,5" SSDs in a 3,5" backplane

Sata OS independent Raid-1 enclosure (2 x 2,5" in a 3,5" enclosure)

HBA controller, use LSI/Avaga HBA, best with raidless IT firmware, see
http://www.avagotech.com/products/server-storage/host-bus-adapters/

6G HBA adapter

LSI 9207-8 (IT mode per default)
LSI 9211-8 (IR mode, can be reflashed to IT mode)

LSI 9201-16 (16 Port)

OEM 6G HBA
IBM M 1015 (works after a reflash with the LSI 9211 IT firmware)

12G HBA

ATTO ExpressSAS H12xxx (new: ATTO, a media specialist supports Illumos up from 2018)
https://www.atto.com/products/adapters

LSI 9300, 9305 andf 9400 (9304 not on all supported OSes check your OS release)

10G network adapters
SFP+
Intel X520-D1
Intel X520-D2 (dual port)

Intel X550, Drivers are available in OmniOS 151014/18

TP45 10Gbase-T
Intel X540-T1
Intel X540-T2 (dual port)

40G network adapter
Intel X710 single or dual QSFP+ (driver i40e available in current Solaris and OmniOS 151014/18)
5. Examples

5.1 Small workgroup filer, backup or Lab use
like HP Microserver G8, ML10 G9 or Dell T20, T130

example: HP Proliant ML10-G9 with an Intel G4400 (vt-d capable) and 16 GB ECC

5.1.1 Barebone NAS setup
Your mainboard offers 4-6 Sata ports in AHCI mode and at least 2 PCI-e slots

- Use a small Sata SSD as system disks, prefer
  a small enterprise class SSD like Intel DC S351x

- Use a pool from a mirror of two 3,5" disk (use 24/7 NAS disks, up to 10TB, I prefer
  HGST Ultrastar) or a Raid-Zn from up to 5 disks. Do NOT use MSR/ archive disks.

or use or add a pool from a mirror of 2 SSDs up to 3,8TB or a Raid-Zn. Prefer enterprise class SSDs like a
Samsung SM/PM 863 or use upper class desktop SSDs with overprovisioning like the Sandisk Extreme Pro or
add at least a manual overprovisioning to a new or secure erased SSD of about 10% to keep write performance
high under load.

Options:
- Add a backplane with 4 x SSDs to one 5,25" bay or if you have two 5,25" bays,
  add a SSD backplane for 8 SSSs (see 2.1)

- Add a 3,5" hotplug capable drive bay where you can hot insert/remove a 3,5"
  disk for backups. Create a single disk pool and sync important data from your data-
  pools to this disk (autojob). Replace this backup pool regularly with other disks.

- 10G adapter (use an Intel X520 or X540)
- NVMe as an L2ARC (prefer more RAM)

5.1.2 Napp-in-one (ESXi + virtualized NAS + other operating systems)
Lab use

- Use an Sata SSD for ESXi and a local datastore where you place OmniOS on it
- add an LSI HBA in raidless IT mode in pass-through mode for OmniOS with disks
- use an SSD only mirror for VMs and a 3,5" disks for filer and backup use
- you can skip the Slog device and add an L2ARC if you cannot add/afford more RAM
6. Silent workgroup filer for video editing or lab/office use with high capacity

6.1 Fractal Design R5 case, silent, low power demands (or other small cases)
- use a mainboard from the X10 SDV series with a 2 core, 4 core, a core or 16 core CPI and 16 x SAS
- use an Sata bootdisk like the Intel enterprise class S3510-80

High capacity pool
- add up to 8 internal 3,5" disks like HGST Ultrastar 24/7 NAS (never use SMR archive disks)
Example: 6 disks in Raid-Z2. This gives up to 40 TB when using 10 TB disks

High performance/iops pool
- add a 4 or 8 bay 2,5" backplane into the 5,25" slots
- create an SSD pool from mirrors of 2 SSDs up to 3,8TB or Raid-Z2 of up to 8 SSD.
With high iops SSD you can use Raid-Z for a higher capacity, no need for Raid-10 alike setups

Prefer enterprise class SSDs like Intel S3500/3610/3700 or Samsung SM/PM 863 or use upper class desktop SSDs with over provisioning like the Sandisk Extreme Pro or add at least a manual over provisioning to a new or secure erased SSD of about 10% to keep write performance high under load.

6.2 Napp-in-one (ESXi + virtualized NAS + other operating systems)
- Use an Sata SSD for ESXi and a local datastore where you place the napp-it ZFS appliance template onto
- Pass-through the SAS controller to OmniOS to have real disk access for ZFS with native drivers
- Use an SSD only pool without Slog for VMs and regular disks for filer/backup use
- Use enough RAM (count 4 GB for ESXi and OmniOS, add the needs for your VMs and then add the amount of RAM that you want to use as ZFS readcache. RAM can go from 8 GB to 128 GB.

care about: Pass-through of NVMe does not work at the moment but you can use them as ESXi vdisks

6.3 Fractal Design R5 case, silent, ultra high iops demands
- use a mainboard from the X10 SR (single Xeon) or DR series (dual Xeon)
- with onboard SAS HBA or HBAs as PCI-e devices (ex LSI 9207, LSI 9003)
- with onboard 10G or with an additional Intel X520 or X540

- use an Sata bootdisk like the Intel enterprise class S3510-80

- create a pool from a mirror or Raid-Zn with up to 6 NVMe disks.
Use Intel P750. P3600, P3700 or P3608

and/or
- create an SSD pool with up to 16 SSDs in a Raid-10 or Raid-Zn config.
With high iops SSD you can use Raid-Z for a higher capacity, no need for Raid-10 alike setups
- add as much RAM as affordable/possible as readcache to improve read performance

An SSD only pool from SSDs does not require an additional L2ARC device unless you do not use a device that performs much better than your pool SSDs like an P3608 with regular SSDs in a pool.

An SSD only pool does not require an additional Slog device for sync writes. This may be only useful if your pool SSDs lack powerloss protection with an Slog that performs much better on sync writes than your pool SSDs with ultra low latency, high write iops and powerloss protection.
7. Very high capacity, price sensitive (Sata disks)
Use an expanderless 19" case with a miniSAS backplane (up to 36 x disks)

- add HBA controller according to the needed port numbers
  ex: a 24 bay case require one 16 port HBA and one 8 port HBA or 3 x 8 port HBA.

Example with a 24 bay 19" case:

- mainboard from the X10SDV line with 16 port HBA onboard
- 10G nics onboard (very new, drivers for OmniOS are not yet available but on the way)
- an additional LSI 9207 that comes with IT mode firmware or another HBA

7.1 Use case: high capacity with iops as a concern

- use a multi Raid-10 setup (Create a pool from a mirror vdev, add mirror vdevs) as
  iops performance scale with number of vdevs and sequential performance with number of disks

Create a pool with a raid-1 vdev of 4TB HGST Ultrastar, add more mirror vdevs up to 11 vdevs.
This means 22 disks with a capacity of 44 TB and the write iops of 11 disks and the read iops of 22 disks.
You can calculate around 150 iops per disk (limited by disk rpm and latency) what means 1650 write iops from
disks and 3300 read iops from disks. Storage values are higher with read and write caching and compress.

The sustained sequential write performance can go up to around 11 x 200 MB/s (2200 MB/s) and the sequential
sustained read performance (as ZFS can read from both disks of a mirror simultaneously) up to 4400 MB/s.

Readcaching is the key for performance, so add RAM. Only if you cannot add or afford more RAM and require
a larger readcache, add an L2ARC to extend the RAM. Be aware that an L2ARC SSD readcache is much slower
than RAM readcache. It also requires some percent of RAM to organize the L2ARC. As every new write must go
to the pool and to the L2ARC you should care of write performance of an L2ARC as well or it may affect write
throughput negative.

Good L2ARC: Intel S3610-400 or an NVMe like a Intel P750-400

7.2 Use case: highest capacity (filer or backup)

- use a two Raid-Z2 or Z3 setup (Create a pool from a Z2 vdev, add Z2 vdevs)
  create a pool with a 10 disk z2 or an eleven disk z3 vdev, add a second identical vdev now or later.
  I would avoid to use much more disks per vdev as this affects resilver or scrub time negatively
- add at least one hotspare disk.

Such a setup can go up to 200 TB with 10TB Sata disks

- optionally add an L2ARC or ZIL device ( see 6.1)
- optionally create one pool from SSDs (high iops) and one from 3,5" disks (high capacity)
- optionally use a SuperMicro X10 SR/DL board with up to 6 slots for NVMe disks for a second high iops pool

7.3 Napp-in-one (ESXi + virtualized NAS + other operating systems)
Lab use or as a backup/ failover system

- similar 6.2
8. Ultra high capacity or HA capable with SAS disks
Use 19” cases with expander
Use for HA configurations or capacity ranges from 200 TB to multiple Petabyte

8.1 High availability storage
based on two storageheads and a storagenode with dualpath SAS
- use two 19” systems (storageheads) with an external SAS connector
- use one Jbod storagenode with two external SAS connectors (each for one of the SAS ports)

Connect both storageheads via SAS to your storagenode. As SAS disks offer two ports, you can connect a single disk to both heads simultaneously. You need a HA software like RSF-1 from high-availability.com to manage the failover from one head to the other in case the primary head fails.

An option are SuperMicro Twin servers with two mainboards in one case.
This allows an immediate service failover of services even on a complete storagehead failure. If you also want to allow a complete storagenode failure, you need two heads and two nodes.

8.2 High capacity storage
Use a 19” case like a SuperChassis 847E26-R1400LPB with expander for 36 disks and a mainboard like a SM X11 SSH-CTF with a G4400 or Xeon, 64GB RAM that comes with 19G and onboard 12G SAS.

9 Petabyte storage
based on a storageheads and one or more 90 bay Jbod cases
- use a case like a SuperChassis 946ED-R2KJBOD (90 disks JBOD, up to 900 TB raw per enclosure)
- add up to 90 12G SAS disks like the HGST Ultrastar HE8 or He10, 512e, SAS12G SAS disks

From your 90 bays, use at least 2 hotspares what leaves a maximum of 88 disks
If you organize the pool in 8 vdevs with 11 disks, this gives a max usable capacity of 720 TB per case.
If you use new 10TB HE, you can get 720TB usable per enclosure (up to 900 TB raw).

If you want a more performance orientated setup you can use 14 vdevs from 6 disks each.
With 84 disks with room for hotspares, you have 560TB usable per enclosure.

- add a 19” Storagehead with a board like the X10 DRH-iT, 128–512 GB RAM, Dual Xeon and one ore more 12G SAS HBA SAS controller with external SAS connectors according to the number of enclosures.

You need a storagehead with a lot of RAM, at least one 12G external 2 SAS port (LSI/Avage 9300 8e).

For performance, plan two SAS ports per enclosure ex 2 x LSI 9300-8e If you need a large L2ARC cache, add an Intel P3608
- Use one or more 10G links for uplink to a switch or use an Intel X710 (QSFP+/40G) as a 40G uplink.
Drivers for the new X710/40G QSFP+ are on the way.

Such a setup has an enormous throughput. If you want to add an L2ARC for a very large readcache, you must use one that does not limit throughput on new writes as they must go to the pool and to the L2ARC. An 1.6 TB Intel P3608 with up to 5GB/s read throughput and 850k read iops may give you what you want. Enable sequential L2ARC caching in napp-it tunings. With 1.6 TB cache, use at least 256GB RAM.

Such an enclosure gives up to 900TB raw.
- If you need HA, use two storageheads with RSF-1 in an MPIO setup

Some SAS abling options:

Special settings with many disks:
If you use the napp-it tuning options with a short timeout like 8s, you may find connectivity problems like high iostat failure rates. Increase timeout settings then to a higher value. Solaris default timeout is 60s. This may be too high as this is the time ZFS is waiting for a disk to answer.

If you do not use MPIO, optionally disable MPIO in your /kernel/drev/"controller".conf
10. Redundant Storage with data and service failover (Z-RAID SSF v.2)

This solution is beta in the current dev edition, release expected napp-it 2019.01

High availability solutions are usually based on a Dualpath Mpio SAS storage with redundant SAS controller and two storage heads. They allow a complete head or controller failure without a service outage as they offer services over a switchable failover ip.

Advantage of a typical HA solution (Cluster in a Box)
- best performance (SAS datapath)
- allow a head failure without a service outage
On a Master head failure, the standby Slave imports the pool and continue to offer services over same ip.

Problem of a typical HA solution
- costs and complexity as you need shared SAS storage and at least two physical servers

Z-RAID SSF vCluster vs traditional Cluster in a box solutions

The main idea behind a napp-it ZFS vCluster is to reduce complexity and costs of a ZFS Cluster but to maintain most of its performance and availability. This is achieved by the following

- use ESXi and storage virtualisation: One All-in-One instead of two servers / mainboards
- use the features of ESXi to share virtual disk controllers and raw disks or vdisks between VMs
- use ZFS itself for NFS and SMB service management. This is possible due the kernelbased services on Solarish
- use napp-it for failover management and Stonith

Use cases for Z-RAID SSF
- affordable high available / high performance NFS/SMB filers (other services optional)

Details, see

11. High performance 40G Connection between Appliances

For Z-Raid SSF performance, network performance is essential. Ethernet 1G is similar to old ATA disks with around 100MB. This is not fast enough for appliance Z-RAID.

10G Ethernet can give up to 1000MB/s. This is enough for a Z-RAID and offers enough performance for a traditional disk based pool or many use cases.

The upcoming QSFP+ standard offers 40G ethernet or 4 x 10G SFP+ over a breakout cable. This is in the performance area of a local high performance pool or a very fast NVMe solution. As the price of 40G adapters are quite similar with 10G server adapters, this is the best solution beside FC/IB solutions. While FC/IB may have advantages regarding latency, iSCSI is mainstream and QSFP+ is available in a lot of ethernet switches.

For 40G connectivity you can use either a new MPO (multi-fibre push-on) connector with a copper MPO DAC cable up to 5m for local connectivity, a 12 fiber MPO optical tranceiver and MPO fiber patchcables for connectivity inside a serverroom or between different brands. For long distance connectivity you can use CWDM tranceivers with an LC fiber connector. They use 4 different colors simultaneously over a single fiber to achieve 40G over a traditional Multimode (up to 300m) or Singlemode (long distance) fiber.
40G QSFP+ for a same room Appliance <-> Appliance failover setup
- 2 x Intel XL 710 DA1 (around 450 Euro each)
- QSFP + DAC copper cable (1 to 5 m), either genuine Intel or compatible
  Intel XLDACBL1 (.BL5), Intel compatible cables up from 90 Euro

40G QSFP+ for a local Appliance <-> Appliance failover setup
- 2 x Intel XL 710 DA1 (around 450 Euro each)
- 2 x Intel Tranceiver Intel E40GQSFP5SR or compatible (Intel compatible up from 200 Euro)
- MPO Patchkabel or 2 x breakout cable MPO -> 12 x LC (multimode cabling) --> MPO

40G QSFP+ for a remote Appliance <-> Appliance failover setup
- 2 x Intel XL 710 DA1 (around 450 Euro each)
- 2 x CWDM Tranceiver (several 10G links over different colors) for 10G over
  Multimode od Singlemode LWL (ask for compatibility with Intel)

40G QSFP+ with a 40G Site to Site or building to building connectivity
- 2 Switches with QSFP+ ports ex H3C
- 2 x CWDM Tranceiver (several 10G links over different colors) for 10G over
  Multimode od Singlemode LWL (ask for compatibility with your switch type)
  ex (H3C) HPEX 140 for 2km (JL286A) or 10km (JG661A) or compatible up from 800 Euro

Connect your Appliance to the QSFP+ port of a switch
- 2 x QSFP+ MPO Tranceiver for your Switch ex HPE X140 MPO (JG709A)
  or compatible up from 200 Euro
- 2 x Intel XL 710 DA1 (around 450 Euro each) for the appliances
- 2 x MPO Tranceiver Intel E40GQSFP5SR or compatible (Intel compatible up from 200 Euro)
- 2 MPO Patchcable to connect the XL710 to the switch, price depends on length up from 100 Euro

If you simply want 40G between a QSFP+ switch and your storage appliance
- 1 x Intel X710 (around 450 Euro)
- 1 x MPO Tranceiver Intel E40GQSFP5SR or Intel compatible up from 200 Euro
- 1 x MPO Tranceiver for your Switch ex HPE X140 MPO (JG709A) or Intel compatible up from 200 Euro
- 1 MPIO patchcable, price depends on length up from 100 Euro

prices are estimated prices in Germany without tax, example from www.flexoptix.net

12. more docs

napp-it Homepage:
http://www.napp-it.org

How to setup the napp-it ZFS storage server

How to setup napp-in-one (virtualized storage server on ESXi)

Performancetuning with 10G and SMB2

Download napp-it ToGo (ready to use images for a barebone setup or ESXi template)
http://napp-it.org/downloads/index_en.html

How to setup OmniOS manually
http://napp-it.org/downloads/omnios_en.html

Intel Optane, a game-changing technology

read also
Top Hardware Components for napp-it and OmniOS NAS Servers (servethehome.com)
https://www.servethehome.com/buyers-guides/top-hardware-components-napp-omnios-nas-servers/