facebook
Server Board Design

Harry Li
August 17, 2011
Design Target

- Web Tier application (Type I)
  - CPU core and frequency
  - Memory speed

- Memcache Tier application (Type II)
  - Memory density

- Service Tier application (Type VI)
  - CPU core and frequency
  - Memory density
  - Flash card adds another cache layer
Standard server motherboard

- Standard form factor
  - Standard dimension, mounting holes, IO location and key component placement.

- Wide target audience
  - Many features to accommodate many requirements, increased cost

- Unused components
  - Feature rich system, increases component cost for Facebook

- Low power efficiency
  - Power conversion is optimized for cost, increases end user TCO
What We Did

▪ Dual source CPU and motherboards
  ▪ Limits supplier risk, encourages aggressive pricing

▪ Feature Reduced
  ▪ Pay for what you use, reduces cost

▪ Optimize I/O
  ▪ Front access I/O for better serviceability.
  ▪ External PCI-E link connection for future extensibility.

▪ High efficiency power conversion
  ▪ Regulators supply power to memory and CPUs, efficiency (>91%) reduces TCO
What We Did

- Multiple build options
  - Component configurations for memcache and web servers

- Upgradeable CPUs
  - Provisions to accept AMDs next generation processor, reduces future development cost – enables low cost upgrade path

- Reboot on LAN feature
  - Allows system to reboot via remote command, replaces traditional remote control features, such as relying on BMC.

- Hot Pluggable Debug Card
  - Provide system status code and serial connection.
Here they are
Intel 2 sockets motherboard
Here they are

Intel 2 sockets motherboard

- Custom spread design
- 2 Intel Xeon 5500 or 5600 series processors
- Intel 5000 I/O hub with 24 PCI-E lanes
- 288GB maximum memory
- Intel 82574 or 82580 NIC, reboot on LAN
- Custom debug card
- >91% efficient voltage regulators
- Onboard HDD power
Here they are

AMD 2 sockets motherboard
Here they are

- **AMD 2 sockets motherboard**
  - Custom shadow design
  - 2 AMD Opteron 6100 series 12- and 8- core processors
  - AMD SR5650 chipset, 22 PCI-E lanes
  - 384GB maximum memory
  - Intel 82574 or 82580 NIC, reboot on LAN
  - Custom debug card
  - 93% efficient voltage regulators
  - Onboard HDD power
Chassis

- 1.5U (2.65 in.) tall
- Integrate cable management
- 60mm fans
- Spring loaded plungers
- Tool-less motherboard standoffs
- Hard drive mounting cage (up to 6 drives)
- 4 rear facing fans
- Front cable access only
What we do for 2012

- **Next generation Intel platform for 2012**
  - Higher performance benefits from new microarchitecture and more core count
  - More memory channel, and higher memory speed
  - PCI-E directly from CPU
  - More PCI-E lanes

- **Upgrade AMD CPU for 2012**
  - Higher performance benefits from more core count.
New System Design

- New system architecture in Freedom chassis
  - Same 1.5U chassis form factor
  - Two servers in one chassis, with tray design for easy hot-swap either server
  - Boot HDD front access
  - More I/O expansion (2x PCI-E cards and 1x Mezzanine per each server)

- Improved Management and Debug
  - Improved debug card design for easy plug & play.
  - Serial-over-LAN
  - Remote power on/off control, in addition to reset control
## System Overview

### System Components

<table>
<thead>
<tr>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x PSU</td>
</tr>
<tr>
<td>4x System FAN (60x38 mm)</td>
</tr>
<tr>
<td>1x Mid-plane</td>
</tr>
<tr>
<td>AC/DC cable to front</td>
</tr>
<tr>
<td>2x MB Tray</td>
</tr>
<tr>
<td>2x HDD</td>
</tr>
<tr>
<td>All Front Access I/O</td>
</tr>
</tbody>
</table>
MB Tray in Chassis

- Guide
- Stopper
- Tray Ejector
## I/O Extensibility

<table>
<thead>
<tr>
<th>I/O Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>2x Standard Profile PCI-E card (x8 each)</td>
</tr>
<tr>
<td>1x Mezzanine card (PCI-E x8)</td>
</tr>
<tr>
<td>External PCI-E Link (x4 each)</td>
</tr>
</tbody>
</table>
Mid-plane

- Power delivery
  - Deliver DC power from PSU to installed MB
- Hot swap controller
  - Over current protection and short protection
  - System power consumption reporting
- Support system FAN
  - Bridge control and tachometer signals between system FAN and MB
Server Management

- Intel DCMI in BMC-less environment
  - Leverage management engine inside Intel south bridge (PCH)
  - Remote access and local management without BMC
- SOL (Serial Over LAN)
  - Share on-board NIC
  - Virtual COM port interface to local system
- Remote power control
  - Power on/off and system reset through standard IPMI protocol
- SEL and Sensor monitoring
  - Provide event log
  - Provide sensor monitor/control, such as temperature and FAN speed control
2012 Server Deployment

- **Web Tier rack has 50% more density**
  - Per column, increase from 30 servers to 45 servers

- **Drive TCO down 30%**
  - Higher performance server
  - Higher density to share infrastructure cost

- **Serviceability**
  - Tray hot swappable
  - Front access I/O and hot pluggable debug card
  - More remote management capabilities