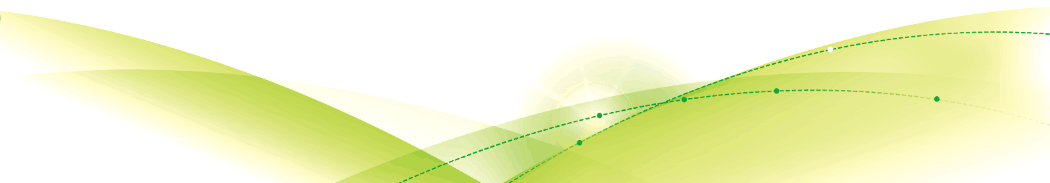


[编号ODCC-08-2017]

100G框式交换机硬件规格书

[文档副标题]



开放数据中心委员会 2017-08-18 发布

技术规格书制定单位：

百度在线网络技术（北京）有限公司

**前言**

本规格书详细描述了百度自研100G框式交换机的硬件规格及核心设计方案。

本规格书的主要内容包括百度自研100G框式交换机的机框物理形态描述、单板物理形态及硬件架构、整机单板构成及系统架构、BMC硬件设计及功能要求、BIOS功能要求、I2C和UART要求等等。

规格书供开发者参考之用，由于作者水平有限，其局限性和其他不足在所难免，敬请指正以便完善。

# 前言

目前数据中心里普遍采用两层的spine-leaf的扁平化组网结构，spine层一般采用成熟的商用框式核心交换机。但受商用框式核心交换机端口密度的限制，组网规模有限，通常单集群服务器规模在1W+，且集群与集群之间还存在较大的网络收敛比，这样的设计对于服务器需求量较大的业务而言，部署和运行都面临许多的难处。

要解决这一问题，基于现有商用的方案，可以采用调整现有集群的上下行收敛比，引入CLOS的架构，搭建出更大规模的网络。本规格书定义的100G框式交换机即是针对此场景的产品。

# 设计思路

本规格书定义的框式交换机，与主流商用框式产品的设计思路会有所差异。设计时将尽量以交换芯片为基本切分单元，弱化机框内部管理，让每个交换芯片各保持其独立性。

设计思路的一些特点：

* 定义两种角色Chassis，端口密度均为128x100G，分别应用于数据中心CLOS网络中的SPINE和LEAF层，共享相同的机框和线卡；
* 舍弃传统“主用主控+备用主控”的控制平面，将硬件集中式的控制平面分布式下沉到各单板，同时在机框外引入新的控制平面（纯软）；
* 机框内部交换芯片全部采用BCM56960，以两层CLOS互连，SPINE和LEAF之间传输完整以太报文；
* 1+1冗余的监控平面；
* 设立BMC管理平面，同样下沉到各单板；
* 转发平面的交换芯片和控制平面的控制CPU为一一对应关系；
* 机框内各单板包括线卡和交换网板均运行独立的OS，地位对等。

128x100G交换机逻辑上可以划分为：转发平面、控制平面、管理平面和监控平面。转发平面由线卡和交换网板上的交换芯片组成，控制平面由线卡和交换网板上的CPU组成，管理平面由线卡和交换网板上的BMC组成，监控平面是两块监控卡组成，两块监控卡为主备关系。其中，转发平面、控制平面和管理平面各自集成在单板上。

# 硬件规格

## 机框

### 尺寸

高度：机框高度6U;

宽度：满足19英寸机柜固定安装要求；

深度：满足19英寸机柜固定安装要求（数据中心机柜外框净深为1.2m，实际由于PDU和理线槽等存在，同时需要预留足够的空间安装，建议风扇模块拔出后（考虑更换风扇）把手到机框前面板眉头之间的距离不超过900mm）；

### 结构

**设计要求：**

* 内部基于直接垂直正交（零背板）架构设计；
* 线卡、监控卡槽位满足图示位置布局；

采用正交结构的考虑：

* 成熟的技术方案；
* 信号完整性好；
* 为更高密度的产品做技术储备；

**单板布局要求：**

交换机采用垂直正交结构，明确指定各个单板的位置关系，主要包括监控卡、线卡和交换网板三部分。PSU在前面板插拔运维，AC Input在后面板；风扇的设计在交换网板上。PSU和风扇的大小、厂家不做要求，厂家根据自己散热方面的专长自行发挥。



图 1 前面板位置关系

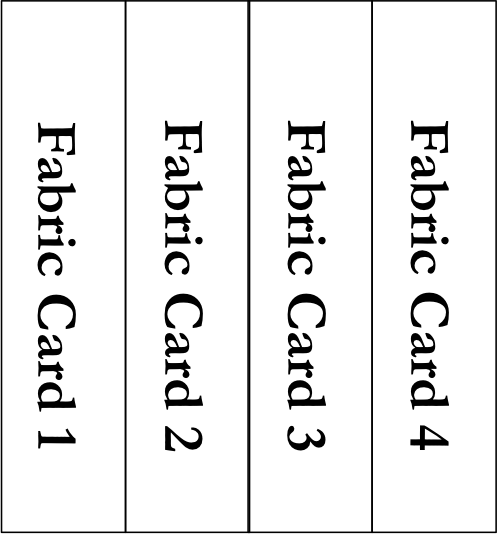


图 2 后面板位置关系

### 散热

机框散热采用前后风道。机框上配置的风扇模块和电源模块都必须是前进风后出风。

### 丝印

建议将LOGO和设备型号放在机框眉头。如果整机空间不够，也可以放在Monitor Card上。

机框正面左右两侧需要有线卡ID的丝印，编号1-8。

机框背面交换板槽位处需要有交换板ID的丝印，编号9-12。

示意图仅表示位置关系。



图 3 前面板丝印示意图

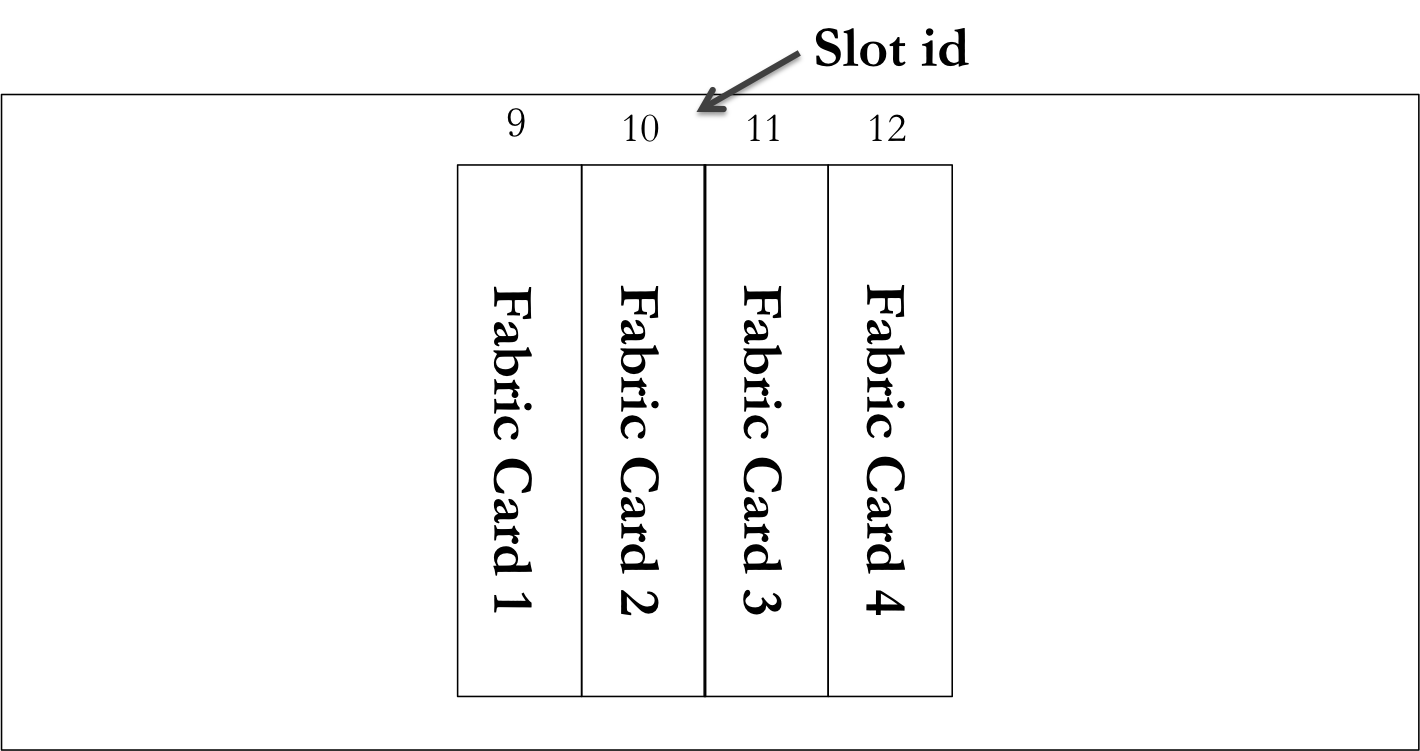


图 4 后面板丝印示意图

## 单板

本机框支持的单板一共3类：

* 线卡（Line Card，简称LC）
* 交换网板（Fabric Card，简称FC）
* 监控板（Monitor Card，简称MC）：提供整个机框的监控和管理。

**总体设计要求：**

* LC和FC硬件设计方案保持统一；
* 交换芯片使用Tomahawk 56960；
* 每颗交换芯片搭配一颗本地CPU和一颗BMC；
* 实现两类交换板，一类带交换芯片（Type A），一类不带交换芯片（Type B）；
* 每个LC和FC单板配置两个离线按键，作用有二：一、扳手解锁；二、产生硬件中断上报软件系统。
* 每个单板需要支持可热插拔；

### 线卡

**功能要求：**

1. 支持对单板远程power reset/off/on/SOL；
2. 提供32个100G业务端口，支持主流100G光模块(100G端口支持降速到40G使用)；
3. 提供2个串口和2个USB口；

**设计要求：**

1. 32个100G QSFP28端口，分成2组，每组16端口，2组端口之间最好能有明显的区分度；
2. 每个QSFP28端口支持1个绿色状态指示灯（绿色常亮表示Link，绿色闪烁表示Active）；
3. 2个RJ45 RS232串口，默认波特率115200，RS232 to RJ45 Pinout 遵循 YOST 标准(http://yost.com/computers/RJ45-serial/)
4. 2个USB接口；
5. 2个红绿双色LED状态指示灯（RUN/ALM），每组交换单元一个；
6. 支持紧固螺钉、离线按键以及扳手，支持无工具维护；
7. 能兼容Type A和Type B交换板；
8. CPU与交换芯片采用PCIe互连，但是D-1527 CPU片内集成的两对Serdes也连到BCM56960上留作扩展；
9. 每16个端口使用一定宽度的带颜色丝印标记；

**硬件参数（两组左右对称）：**

* CPU：Broadwell DE D-1527,4核，x2
* BMC: AST2520, x2
* Switch Chip：BCM56960，x2
* DDR4：至少4GB, x2，带ECC
* Storage：M.2 SSD 64GB, x2

**面板参考布局：**

线卡端口按照16个一组，分两组，每组从下往上从左往右编号，从1到16，线卡端口布局示意图如下。因为单板中间的散热比较好，所以推荐下图中示意图的布局方式。

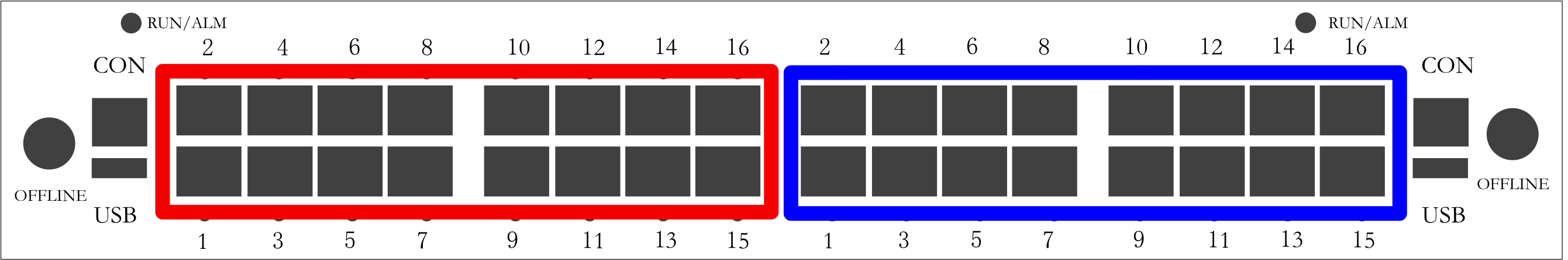


图 5 LC前面板布局示意图

### 交换板

根据CLOS组网中角色的不同，需要设计两种交换板：

* Type A：带交换芯片的交换板；
* Type B：不带交换芯片的交换板，板卡起纯互连作用；

#### Type A交换板

选用Type A交换板的Chassis内部逻辑框图为：

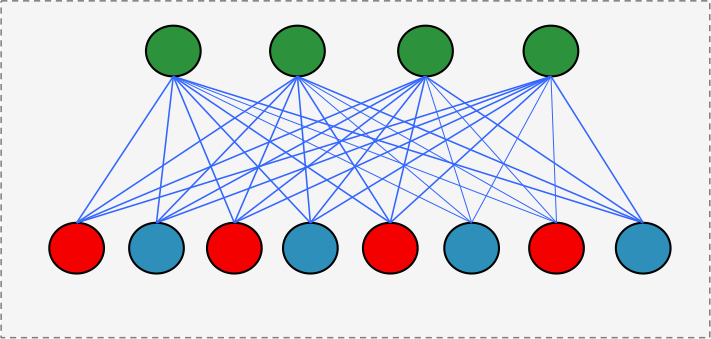


图 6 TypeA交换板逻辑框图

根据对称性可以设计硬件结构完全相同的交换板。对应交换机的正交结构示意图如下：

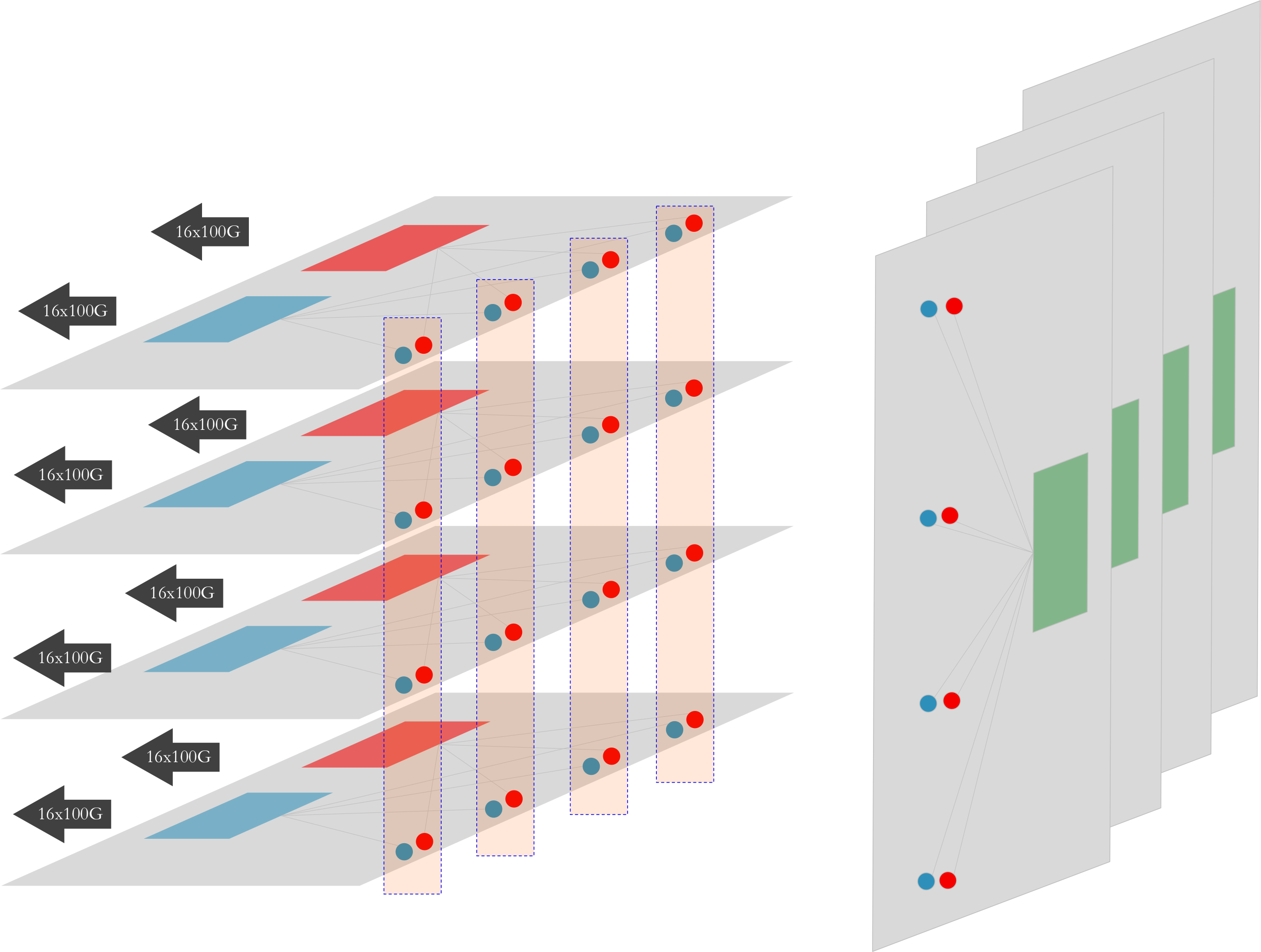


图 7 TypeA正交结构示意图

**功能要求：**

1. 支持对单板远程power reset/off/on/SOL；
2. 直接对外提供串口和USB口；

设计要求：

1. 1个RJ45 RS232串口，默认波特率115200，RS232 to RJ45 Pinout 遵循 YOST 标准(http://yost.com/computers/RJ45-serial/) ，可以承载在监控板上；
2. 1个USB接口，可以承载在监控板上；
3. 1个红绿双色LED指示灯(Run/ALM)，可以承载在监控板上；
4. 支持紧固螺钉、离线按键以及扳手，支持无工具维护；
5. CPU与交换芯片采用PCIe互连，但是D1527 CPU片内集成的两对Serdes也连到BCM56960上留作扩展；
6. 支持**机械**防呆设计，避免Type A和Type B混用；
7. 支持双BIOS；

**硬件参数：**

* CPU：Broadwell DE D-1527,4核
* Switch Chip：BCM56960
* DDR4：至少4GB，带ECC
* Storage：M.2 SSD 80GB
* BMC: AST2520

#### Type B交换板

选用Type B交换板的Chassis内部逻辑框图为：

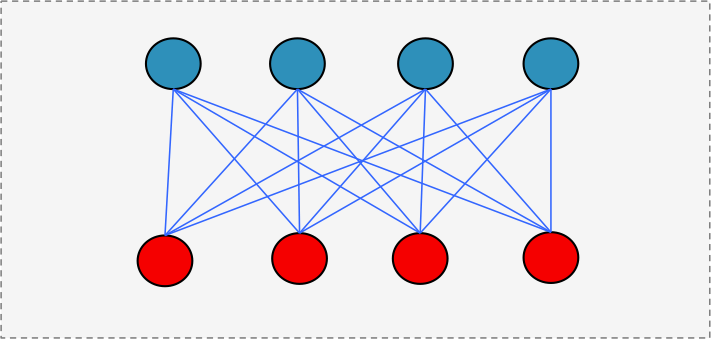


图 8 TypeB内部逻辑框图

该结构采用的是线卡之间直接互连，流量无阻塞方向为南北向。对应Chassis的正交结构示意图如下：

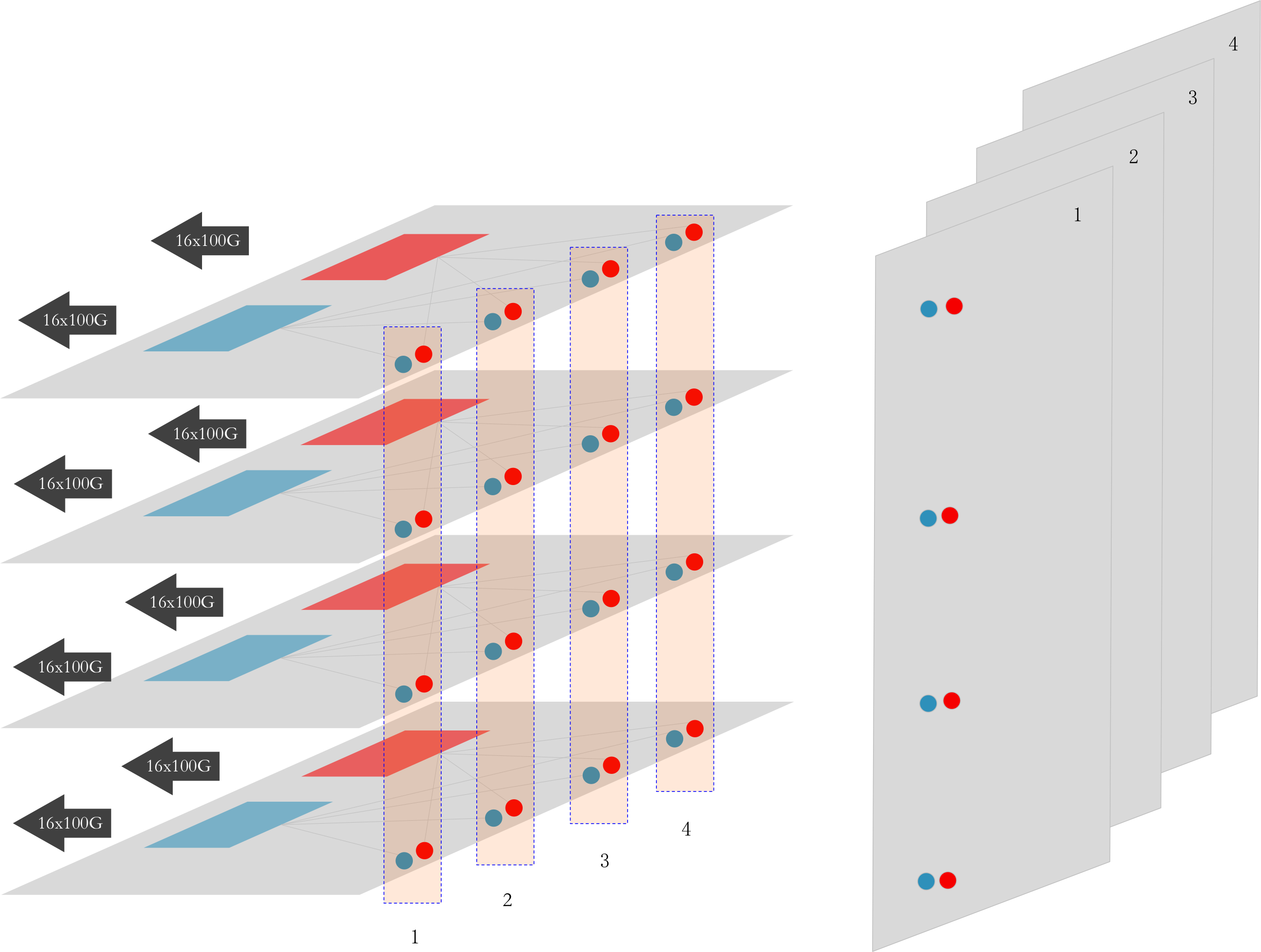


图 9 TypeB正交结构示意图

Type B交换网板设计示意图如下图所示：

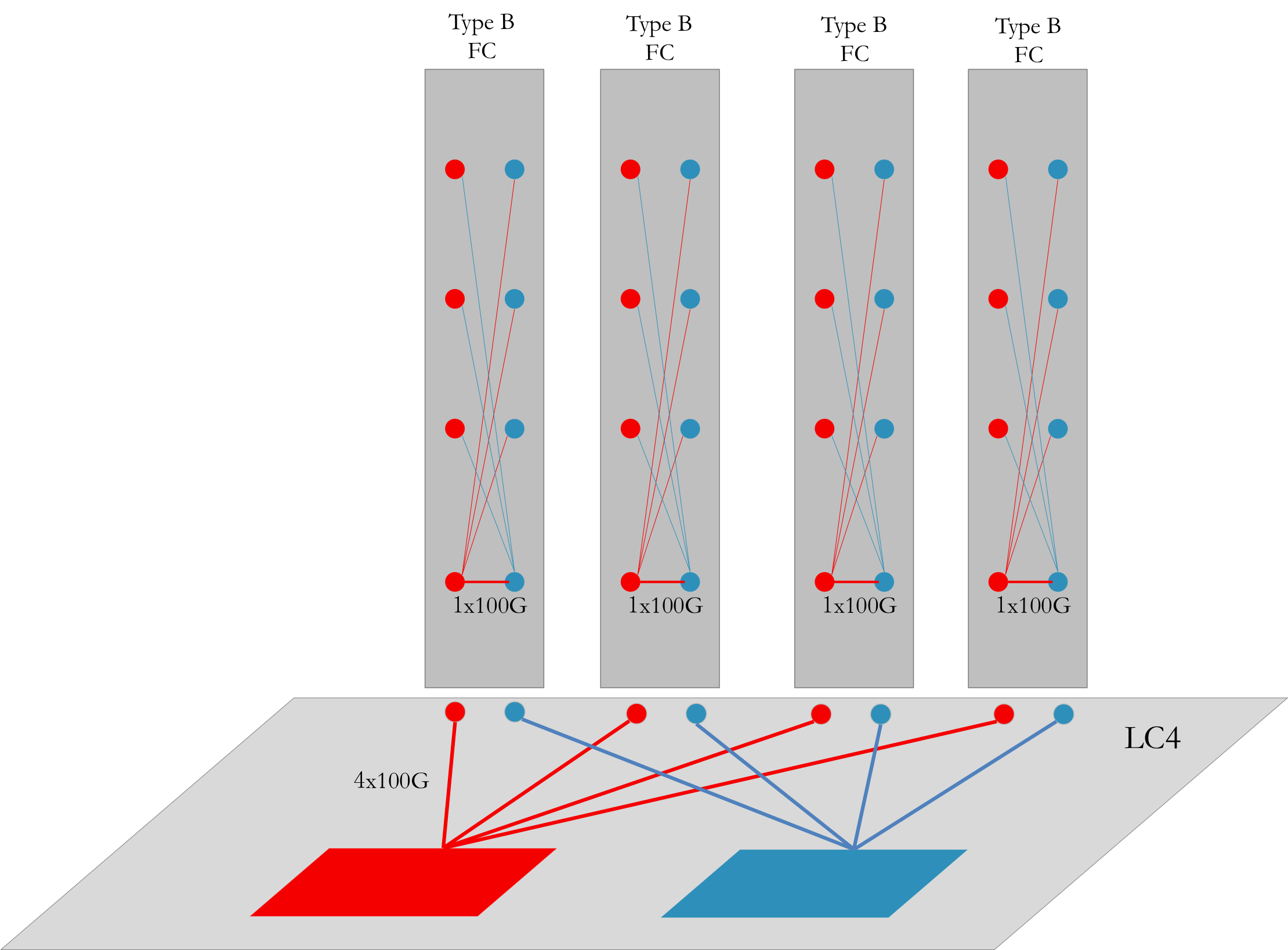


图 10 TypeB设计示意图

从图中可以看到，不带交换芯片的4个Type B交换网板的设计是完全相同的。

**设计要求：**

1. 支持紧固螺钉、离线按键以及扳手，支持无工具维护；
2. 支持机械防呆设计，避免Type A和Type B混用；

### 监控板

对监控板的内部硬件设计以及软件实现，未进行定义，仅定义功能需求。

监控板提供的功能如下：

* 温度监控与风扇管理；
* 电源管理；
* 单板管理口网络和BMC网络的汇聚;
* 机框级别运行状态监控与展示（部件在位管理、单板完备性指示，健康状态指示）；
* 响应线卡和交换网板对共享硬件资源的访问请求和应答；
* 扩展交换网板上的串口和USB口；
* 监控板以主备方式工作，默认左边MC卡为主用监控卡，支持故障时自动切换；
* 监控板可以单独Power on/Off、Reboot每个SLOT单板；可以Reboot整机；

**设计要求：**

* 10个红绿双色LED指示灯（主备/电源/风扇/4个交换网板/线卡/监控板/交换网板类型）：
* 1个RJ45 ETH口，10M/100M/1000M自适应，BMC和以太网管理口融合，共享一个端口； 为方便操作人员判断通断，Slave MC卡的ETH口插上网线后指示灯点亮，但流量不通。
* 每个线卡和交换网板均有独立的IP；
* 1个RJ45 RS232串口，默认波特率115200，RS232 to RJ45 Pinout 遵循 YOST 标准(http://yost.com/computers/RJ45-serial/)
* 1个USB Host接口；
* 1个全局reset按钮；
* 4个交换网板RJ45串口和4个交换网板USB口；

监控板面板布局参考：

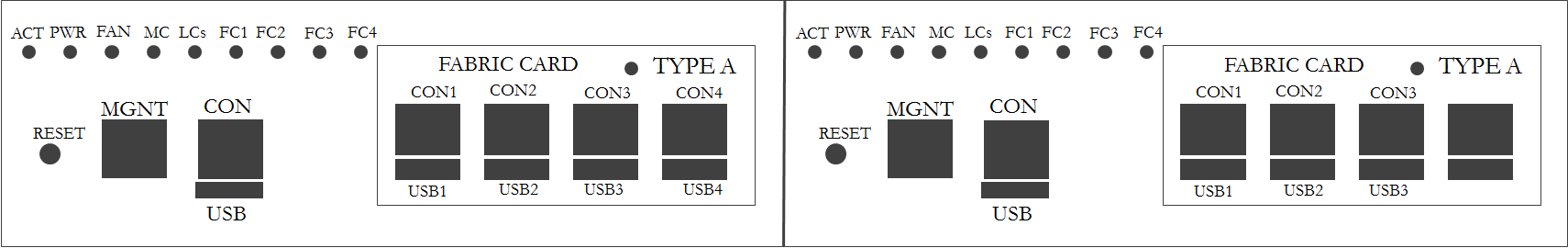


图 11 监控板面板布局示意图

## 电源

实际数据中心中采用的AB路双系统供电，为了避免单系统供电故障或电力切换导致系统下电或复位，要求电源模块支持N+N备份，N根据机框实际功耗对应配置电源模块的数量。推荐选用2+2备份。

**设计要求：**

1. 支持220VAC和240V HVDC；
2. N+N备份，任意模块可热插拔；
3. 每个电源模块自带2个状态指示灯（In和Out）；
4. 电源输入端子必须放在后面，电源本体从前面插拔运维；
5. 电源模块内部风扇支持前进风后出风；
6. 支持选配IEC和中国国家标准插头电源线；
7. 电源模块上必须有紧固设计，避免电源模块松动；
8. 电源输入端子必须有紧固设计，避免电源线松动；

## 风扇

**设计要求：**

1. N+1备份，任意模块可热插拔；
2. 前进风后出风风道；
3. 每个风扇自带状态指示灯；
4. 风扇必须由监控卡控制，不能由线卡或交换网板控制；
5. 推荐采用风扇墙设计；
6. 风扇能根据系统中各个热传感器的温度智能调速；
7. 在更换交换板时涉及到拔下风扇（5-10min），此时散热必须满足工作要求。

## 光模块

上电启动后，OS接管之前，光模块和AOC的Tx\_Disable是有效状态（不发光）。建议通过关闭PCB上Cage的电源来实现。

## LED指示灯

### 监控卡LED指示灯

|  |  |  |
| --- | --- | --- |
| 指示灯 | 颜色 | 含义 |
| ACT  主备状态指示灯 | 绿色 | 常亮：表示主用；  常灭：表示备用；  常灭：监控卡启动期间； |
| TYPE A  交换网板类型指示灯 | 绿色 | 常亮：表示Type A交换网板，适用于Spine lane；  常灭：表示Type B交换网板，适用于Leaf Plane；  常灭：监控卡启动期间； |
| MC  监控卡状态指示灯 | 绿色 | 闪烁：表示监控卡已经上电，处在启动阶段；闪烁频率4Hz  常亮：表示监控卡处于正常运行状态； |
| 红色 | 常亮：表示有故障，需要人工干预； |
|  | 绿色 | 常亮：表示所有的电源模块工作正常；  常灭：监控卡启动期间； |
| PWR  电源模块状态指示灯 | 红色 | 常亮：表示可能原因有：   * 一个或多个电源模块工作异常，在位不输出； * 一个或多个电源模块在位无输入； |
| FAN  风扇模块状态指示灯 | 绿色 | 常亮：表示所有的风扇模块工作正常；  常灭：监控卡启动期间； |
| 红色 | 常亮：表示可能原因有：   * 一个或多个风扇模块工作异常； * 一个或多个风扇模块不在位； |
| LCs  线卡状态指示灯 | 绿色 | 常亮：表示所有线卡在位且工作正常；  常灭：监控卡启动期间； |
| 红色 | 常亮：表示可能原因有：   * 有一块或多块线卡不在位； * 有一块或多块线卡异常； |
| 黄色 | 常亮：表示LCs和FCs正在复位中； |
| FC1  交换板1状态指示灯 | 绿色 | 常亮：表示交换板1在位且工作正常；  常灭：监控卡启动期间； |
| 红色 | 常亮：表示可能原因有：   * 交换板1未在位； * 交换板1有异常； * 交换板1插错（Type B）； |
| FC2  交换板2状态指示灯 | 绿色 | 常亮：表示交换板2在位且工作正常；  常灭：监控卡启动期间； |
| 红色 | 常亮：表示可能原因有：   * 交换板2未在位； * 交换板2有异常； * 交换板2插错（Type B）； |
| FC3  交换板3状态指示灯 | 绿色 | 常亮：表示交换板3在位且工作正常；  常灭：监控卡启动期间； |
| 红色 | 常亮：表示可能原因有：   * 交换板3未在位； * 交换板3有异常； * 交换板3插错（Type B）； |
| FC4  交换板4状态指示灯 | 绿色 | 常亮：表示交换板4在位且工作正常；  常灭：监控卡启动期间； |
| 红色 | 常亮：表示可能原因有：   * 交换板4未在位； * 交换板4有异常； * 交换板4插错（Type B）； |

### 线卡LED指示灯

|  |  |  |
| --- | --- | --- |
| 指示灯 | 颜色 | 含义 |
| RUN/ALM  运行状态指示灯 | 绿色 | 闪烁：表示单板已经上电，处于启动阶段；闪烁频率4Hz  常亮：表示单板处于正常运行状态； |
| 红色 | 常亮：表示有故障，需要人工干预；例如 CPU Hang机，系统过温等； |
| 黄色 | 常亮：表示release完成，可以拔下单板；或者表示一般故障：BMC Hang机、总线故障、备用BIOS启动等； |

### 交换卡LED指示灯（可选）

|  |  |  |
| --- | --- | --- |
| 指示灯 | 颜色 | 含义 |
| RUN/ALM  运行状态指示灯 | 绿色 | 闪烁：表示单板已经上电，处于启动阶段；闪烁频率4Hz  常亮：表示单板处于正常运行状态； |
| 红色 | 常亮：表示有严重故障，需要人工干预；例如 CPU Hang机，系统过温等； |
| 黄色 | 常亮：表示release完成，可以拔下单板；或者表示一般故障：BMC Hang机、总线故障、备用BIOS启动等； |

### 电源模块状态指示灯

|  |  |  |
| --- | --- | --- |
| 指示灯 | 颜色 | 含义 |
| IN STATUS  运行状态指示灯 | 绿色 | 常亮 ：表示电源模块输入正常； |
| 红色 | 闪烁： 表示电源模块无输入； |
| 常灭 | 表示电源模块异常，需要更换，可能原因：   * 供电异常； * 电源模块异常； * LED灯异常； |
| OUT STATUS  运行状态指示灯 | 绿色 | 常亮 ：表示电源模块输出正常； |
| 红色 | 闪烁： 表示电源模块输出异常； |
| 常灭 | 表示电源模块异常，需要更换，可能原因：   * 供电异常； * 电源模块异常； * LED灯异常； |

### 风扇状态指示灯

|  |  |  |
| --- | --- | --- |
| 指示灯 | 颜色 | 含义 |
| STATUS  运行状态指示灯 | 绿色 | 常亮 ：表示风扇模块工作正常； |
| 红色 | 闪烁： 表示风扇模块有异常； |
| 常灭 | 表示风扇模块异常，需要更换，可能原因：   * 供电异常； * 风扇软件异常； * LED灯异常； |

几点说明：

1、整机上电时，两个MC卡同时启动期间或者启动完成后，主用和备用MC卡的指示灯，除了ACT指示灯（主用亮，备用灭）和MC指示灯（本板故障的情况下）外，其他指示灯状态都同步。

2、对于MC卡之MC指示灯、LC/FC之Run/ALM指示灯如果不能做到上电开机时绿灯闪烁，也可以以上电开机时红灯常亮、OS booting时绿灯闪烁、OS booting完成后绿灯常亮来实现。以达到区分上电开机和启动完成之状态。

3、100G端口（Port）指示灯点灯逻辑为：交换机上电启动后，端口指示灯黄色（红绿混色）闪烁一次后全熄灭（持续时间2s左右），直到百度OS booting完成接管交换机。百度OS接管交换机后，端口指示灯按实际Link/Active状态显示。交换机整机reboot或者手动复位时，端口指示灯点灯逻辑同上电启动。

# 设计细节

## 软件概述

根据Chassis角色的不同，两种交换机运行时可以等价为8台或12台TOR。每个TOR运行相同而独立的OS。

线卡和交换网板硬件上如果有差异化设计，则在启动时可以首先获取自己的slot id（编号1-8或1-12），进而开始差异化运行。

单板上的OS一方面要完成交换芯片的控制和管理功能，另一方面也要具备交换芯片之外的硬件组件的管理，既包含单板上也包含机框级别。

这里的硬件外设管理主要有包含如下功能点：

* 系统SN查询（机框/监控板1/监控板2/单板自身/PSU/光模块）；
* 光模块信息查询（仅针对本地光模块，包括资产信息和DDM信息）；
* 风扇状态查询；
* 电源状态查询（状态、SN、温度、风扇、输入电压电流功率，输出电压电流功率）；
* 温度查询（本板温度，监控卡温度）；

除了单板上的OS以外，监控卡OS不对外开放，此部分的软件实现由硬件合作伙伴完成，本规格书仅从功能上进行定义。

## BIOS描述

* 硬件需求：
* BIOS FLASH: SPI FLASH 8 MB
* 双备份，Golden Backup

## BMC描述

* BMC管理系统硬件需求如下：

|  |  |
| --- | --- |
| 组件 | 型号/类型 |
| SOC | AST2520 （备选AST1250） |
| DRAM | DDR4 512MB—不强制 |
| SPI FLASH | 双备份 |
| EEPROM | 4KB |
| LAN | NC-SI (实现ShareNIC) |

* BMC系统要求：BMC先于CPU启动完成，达到BMC启动后可以通过SOL查看CPU Boot信息之目的。
* BMC reset和CPU reset解耦，互相不影响。同时，BMC可以reset CPU（硬复位）；CPU可以reset BMC（硬复位）。

## Mid-plane

机框除了监控卡，线卡和交换网板以外，还有一个Mid-plane，主要用于控制信号。包括电源、I2C、Ethernet、各类在位信号等。

Mid-plane上还会包含一个EEPROM，用于存储机框的SN。

## I2C互连

在当前的设计里，每个带CPU的线卡或交换网板除了需要访问本地的I2C slave设备外，还需要能够访问机框上的共享资源，比如风扇，电源，Mid-plane上的EEPROM，监控卡上的EEPROM和Sensor。

监控卡统一访问共享I2C slave资源，提供结果给单板。

线卡上QSFP28的各路I2C总线和整机上的其他I2C总线必须实现互相隔离。

QSFP28的I2C总线和管理信号通过CPLD/MUX/Expander直接连接到Local CPU（不通过BMC）。如果某路I2C短路或挂死，系统必须实现自动发现、隔离、恢复故障。

## 管理网互连

线卡、交换网板以及监控卡之间通过内部以太网的方式互连，每个单板上的CPU以及BMC的以太网络同时连接到监控卡1和监控卡2上的LAN Switch。

为了实现CPU或BMC同时到两个监控卡的稳定连接，每个CPU或BMC外部需要使用一颗小的LAN Switch芯片。

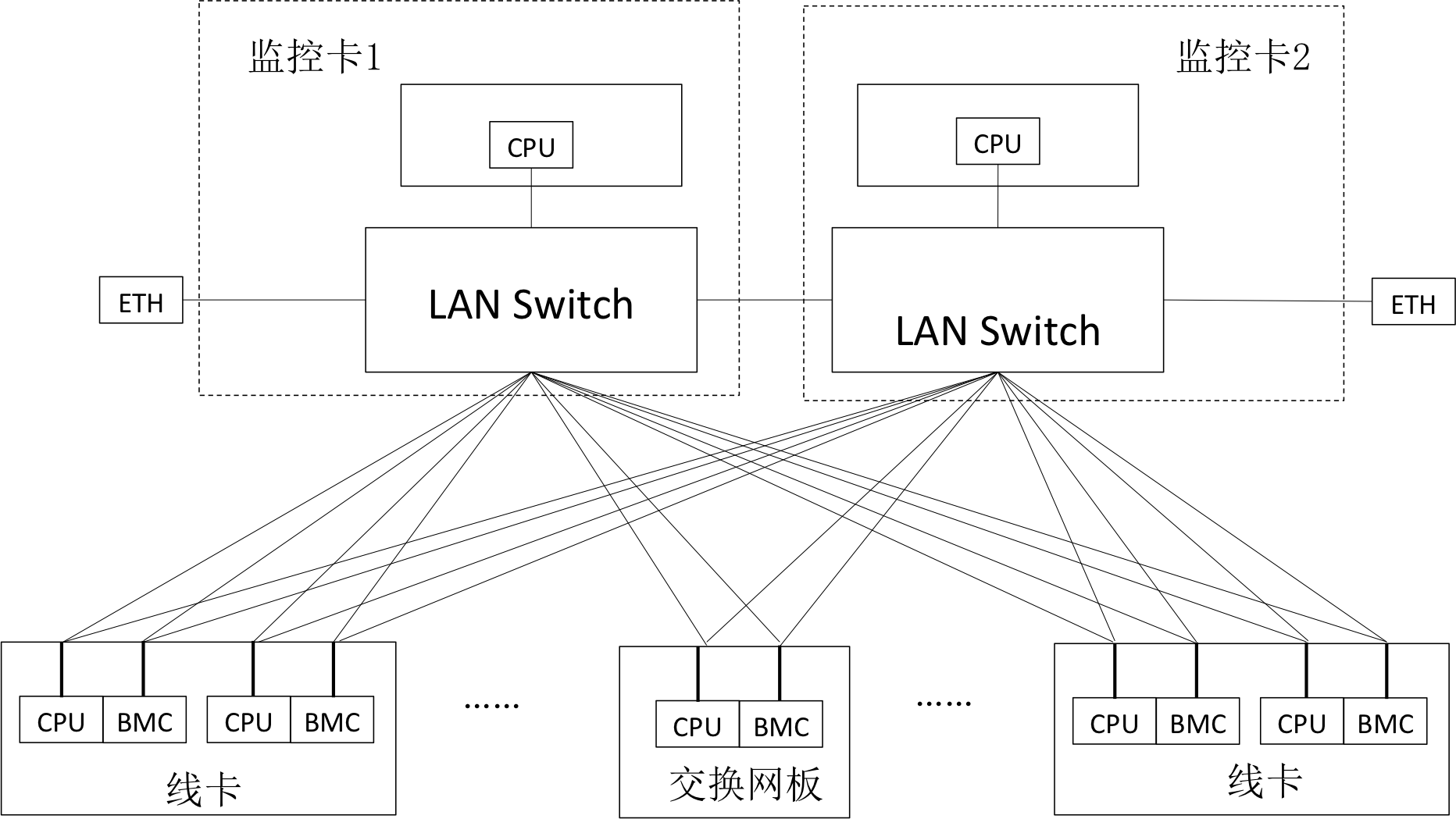


图 12 管理口互联示意图

说明：

* 每个CPU和BMC只有一个IP地址，不允许分别使用双端口分别和主备监控卡相连；
* 非Active的监控卡上与CPU和BMC相连的链路保持down，只有Active侧为up。
* CPU和BMC侧一个ethernet端口转换为两路分别连接两块监控卡的方案可选：NCSI NIC （同时连接CPU+BMC）+ LAN Switch或NIC (CPU侧)+ PHY（BMC侧） + LAN switch;

## UART连接

为了软件驱动的统一性，特做如下规定：

1、CPU的UART0直接接前面板的Console口(默认)；

2、CPU的UART0直接接BMC的UART（SOL功能实现或者其他信息交互）；

建议连接的示意图如下：



图 13 UART互联示意图

## 槽位ID

每个线卡和交换网板（Type-A）槽位需要能够获取槽位id。实现的思路：每个单板提供两组5个pin的GPIO线（对应两套交换系统），这些GPIO线连接到Mid plane，Mid plane对每个单板进行硬编码，00001~01100，0对应低电平，1对应高电平。

## EEPROM

出于资产管理的需要，机框、监控卡、线卡和交换网板都需有对应的EEPROM用来存储SN、MAC、硬件版本、固件版本等方面的信息。

线卡、监控卡和交换网板的EEPROM放置于各自板内，存储的信息包括：SN、MAC、硬件版本、固件版本、板号等。板号只对Type B交换网板有意义，用于监控板做交换网板完备性检查使用，其他单板板号为0。同一块线卡上有两块EEPROM，两块EEPROM中存储的SN相同，MAC不同。

机框部分由于没有电子设计，因此机框的资产信息所对应的EEPROM可以承载在Mid plane上，该EEPROM存储的信息包括：产品型号、SN、硬件版本。

## Reset按钮

每个监控卡上一个Reset按键，主要的作用是：持续按下5秒后，清除整机软件配置，然后Reset系统。为了防止运维人员正常操作时误动作，需将按键做成内缩样式。

## 离线(Release)按键和扳手

每个单板要求支持离线按键和扳手，并且扳手的解锁必须是在离线按键按下之后。按键和扳手的配合需要满足如下设计需求：

* 防止误操作，避免不小心触碰按键导致扳手解锁；
* 在按下按键到拔出单板之间有时间间隙（大于1s）可以通知CPU（Type B除外），让软件做配置清除的操作；

建议解决方案：

做成双按键设计，左右两个按键分别解锁，按下一个就解锁弹出一个扳手，但是只有左右两边都解锁后才出发硬件中断上送CPU。

## 风扇

交换网板做在风扇内部，这样做的好处是散热好做，但是需要注意的是运维交换网板时，风扇也要被移出断电，对散热有挑战，设计时需要考虑此场景。

## 交换网板防呆设计

由于设计了Type A和Type B两种交换网板，为了避免实际运维过程中可能出现的板卡滥用或混用问题，必须增加机械防呆设计。主要的要求是：当选定第一块板的类型后，后续所有的板只能选择同样类型的才能插入。

## Thermal要求

要求在满足所有要求的前提下，尽可能提升散热性能，满足强制目标：在45°C环境温度下，能支持工作温度上限为55°C的光模块。

风扇调速策略的原则是：使用BMC或者CPLD等纯硬件手段实现，调速过程不需要CPU和OS参与。风扇调速策略由ODM厂商确认，但是不读取MAC芯片内部sensor作为调速参考。因为读取MAC芯片内部温度需要CPU调用SDK读取，CPU参与此过程。

一般情况下，外部sensor和MAC内部温度有一定的差异，可以通过测量内部传感器和外部传感器的温度曲线，然后再算法里面修正两者差异进行规避。

风扇的转速差较大的时候（推荐设定值和实际值相差大于15%），需要报警。

当整机或者某个单板局部出现温度过高时，要有过温保护（OTP）。推荐的做法是关闭过温电路的电源，达到保护目的。关闭电源后，必须要做到可以远程手动开启被关闭电源。

## CPLD

系统中的可编程逻辑器件CPLD必须支持在线升级功能。

## 诊断和监控

**1、重启监控**

系统必须记录如下重启类型，建议通过CPLD记录或者BMC记录或者记录到NVRAM等其他介质，以便问题定位查询。

* + HW reset events
    - Power cycle event (PSU AC power cycle)
    - IPMI power cycle event (DC power cycle)
    - CPU power down/on event
  + Linux reboot events
    - Manual reboot (reboot cmd)
    - Manual shutdown (shutdown/poweroff cmd)
    - Linux WDT reboot
    - Kernel panic
  + BMC reboot events (CPU not reset)
    - BMC soft reset (via IPMI cmd)
    - BMC firmware upgrade reset
    - BMC hard reset (reset button，如果无reset button，此项忽略)
  + Thermal trip events (BMC not reset)
    - CPU temp too high trip
    - ASIC（Tomahawk）temp too high trip
    - Environment temp too high trip
    - All fans failed trip

**2、电压监控**

需具有电压监控功能，监控单板的所有VR电压，并提供电压超出阈值告警和log记录功能。同时支持命令行查询单板电压幅值的接口，供运维时查询单板VR电压。

# 工作环境要求

工作环境温度：0-45°C

工作海拔：0~2000米

工作环境湿度：相对湿度（非凝露）10%-90%

说明：环境温度和海拔并不需要同时满足。当海拔超过950米时，温度需要降额：每升高100米，降额0.3度。

# 工艺与器件要求

## 可靠性要求

**耐腐蚀要求：**要求整机支持耐腐蚀功能：通过三防漆涂覆及其他工艺处理，实现对板卡、电源、风扇及其他关键部件的保护。满足G3环境下，35°工作温度，24小时/天满载，5年工作寿命要求。

**MTBF要求：** 30°C环境温度下，MTBF≥87,600小时(10年)；

## 工艺要求

PCB和连接器表面禁止使用化银工艺。

## 器件要求

禁止使用带化银工艺的连接器、开关器件等。

12V及以上的供电电路上禁用钽电容。