IPU-POD$_{64}$ REFERENCE DESIGN

Build and test guide

Table of contents

1 Overview

The IPU-POD$_{64}$ reference design is a rack solution containing 16 IPU-M2000s, 1 to 4 host servers (default 1 host server in reference configuration), network switches and IPU-POD software. There are 64 Mk2 GC200 IPUs in total with four IPUs in each IPU-M2000.

For more information on IPU-POD systems available from Graphcore see https://www.graphcore.ai/products.

This guide is for properly trained service personnel and technicians who are required to install the IPU-POD$_{64}$.

Warning: Only qualified personnel should install, service, or replace IPU-POD$_{64}$ equipment.

If you have any questions then please contact your Graphcore representative or use the resources on the Graphcore support portal: https://www.graphcore.ai/support.

1.1 Acronyms and abbreviations

This is a short list that describes some of the most commonly used terms in this document.

BMC	Baseboard Management Controller – standby power domain service processor doing system hardware management
BOM	Bill of Materials
GW	Short for IPU-Gateway, a co-processor to the four IPUs in the IPU-M2000. It enables scaling with multiple IPU-M2000 units
IPU-Link	High speed communication links that interconnects IPUs within and between IPU-M2000 units. Special cables are required for IPU-Links between IPU-M2000 units
GW-Link	High speed communication link(s) that interconnect IPUs and IPU-GWs horizontally between IPU-M2000 units. Special cables are required for GW-Links between IPU-M2000 units
PDU	Power Distribution Unit
RDMA	Remote DMA
RNIC	RDMA Network Interface Controller
RoCE	RDMA over converged Ethernet
ToR	Top of Rack. Often used in combination with the ToR RDMA switch that is placed on top of the IPU-M2000 stacked units.

2 IPU-POD$_{64}$ reference design components

This section describes the components in the IPU-POD$_{64}$. Each IPU-POD$_{64}$ contains:

- 16 IPU-M2000s
- 1 server (default configuration is 1 host server, up to 4 can be supported)
- 1 1GbE management switch
- 1 100GbE ToR switch
- 2 power distribution units
- 1 rack
- Supplementary mounting components
- Cables

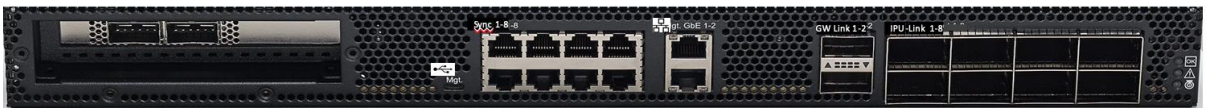
2.1 IPU-M2000

2.1.1 **Overview**

There are 16 IPU-M2000s in each IPU-POD$_{64}$ making a total of 64 IPUs: 4 IPUs per IPU-M2000.

The IPU-M2000 front panel contains:

- 2 RNIC ports
- 8 IPU-Link ports
- 2 management GbE ports (BMC/GW SoC management ports)
- 2 GW-Link ports
- 8 IPU-Link ports



Front panel

The IPU-M2000 back panel contains:

- 2 power connectors per IPU-M2000
- Fan units
- Unit QR code

Back panel

2.1.2 QR code label

There is a QR code label on the back panel of each IPU-M2000. The QR code contains the following information for each IPU-M2000:

- Company name (Graphcore)
- Serial number
- Part number
- BMC Ethernet MAC address
- GW Ethernet MAC address
- Graphcore support web URL (https://www.graphcore.ai/support)

2.2 Server

The default configuration of the IPU-POD$_{64}$ uses a single PowerEdge R6525 server but up to four servers can be connected. Contact Graphcore sales for details of other supported server types. This document describes the default server (PowerEdge R6525) installation only – other servers may have different installation requirements.

The default server configuration is described in section 4.1.

2.3 Switches

Each IPU-POD$_{64}$ contains two switches serving different purposes.

2.3.1 100GbE RoCE/RDMA switch (ToR switch)

The 100GbE RoCE/RDMA switch (also referred to as the ToR switch) is used by the end user's machine learning (ML) jobs as a data-plane, connecting the host servers running the Poplar® SDK with the IPUs running the ML model in the IPU-M2000s. The default ToR switch is an Arista DCS-7060CX-32S-F. Contact Graphcore sales for details of other supported switch types. This document describes the default switch (7060CX) installation only – other switches may have different installation requirements.

2.3.2 1GbE management switch

The 1GbE management switch is used for connecting the management ports together inside the rack. The default management switch is an Arista DCS-7010T-48-F. Contact Graphcore sales for details of other supported switch types. This document describes the default switch (7010T) installation only – other switches may have different installation requirements.

2.4 Power distribution units

Two power distribution units (PDUs) are installed in each IPU-POD$_{64}$. The default unit is an APC AP8886.

2.5 Rack

The IPU-M2000s, servers, switches, and PDUs are installed in an APC AR3300SP rack. This rack has a packing system designed to safely transport and unload the rack.

It is important to follow the instructions carefully when packing or unpacking the rack.

2.6 Supplementary mounting components

The supplementary components listed below also need to be installed.

- Cable organizer
- Blanking panel

2.7 Cables

Each IPU-POD$_{64}$ has three types of cabling:

- RJ45 cables
- OSFP cables
- QSFP cables

2.7.1 **RJ45 cables**

- Red: IPU-M2000 to IPU-M2000 within-rack IPU-Link connectivity
- Blue: Connecting IPU-M2000s to the management switch (BMC + GW management)
- Blue: Connecting servers to the management switch
- Yellow for connecting IPU-M2000s to the management switch (BMC only management)

2.7.2 **OSFP cables**

- IPU-M2000 to IPU-M2000 (IPU-Link) connectivity

2.7.3 **QSFP cables**

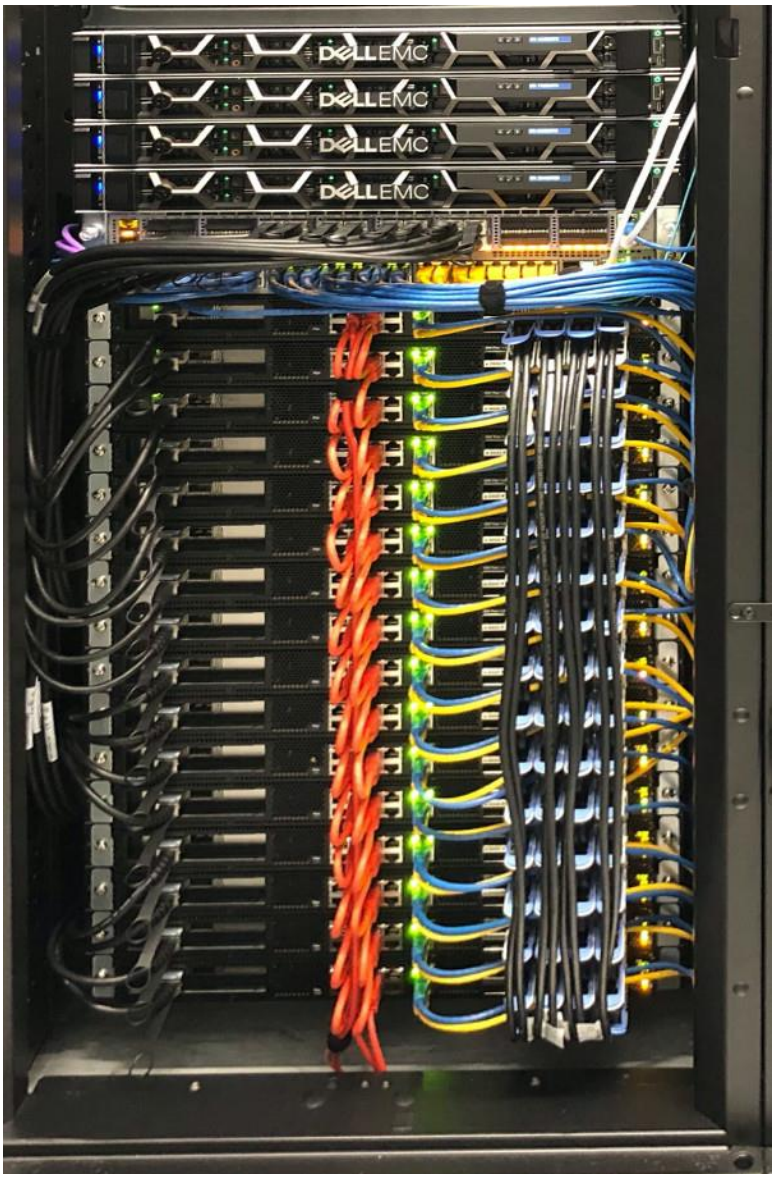
- IPU-M2000 to ToR switch connectivity
- Server to ToR switch connectivity

All cable connections are described in Section 0.

3 Rack assembly

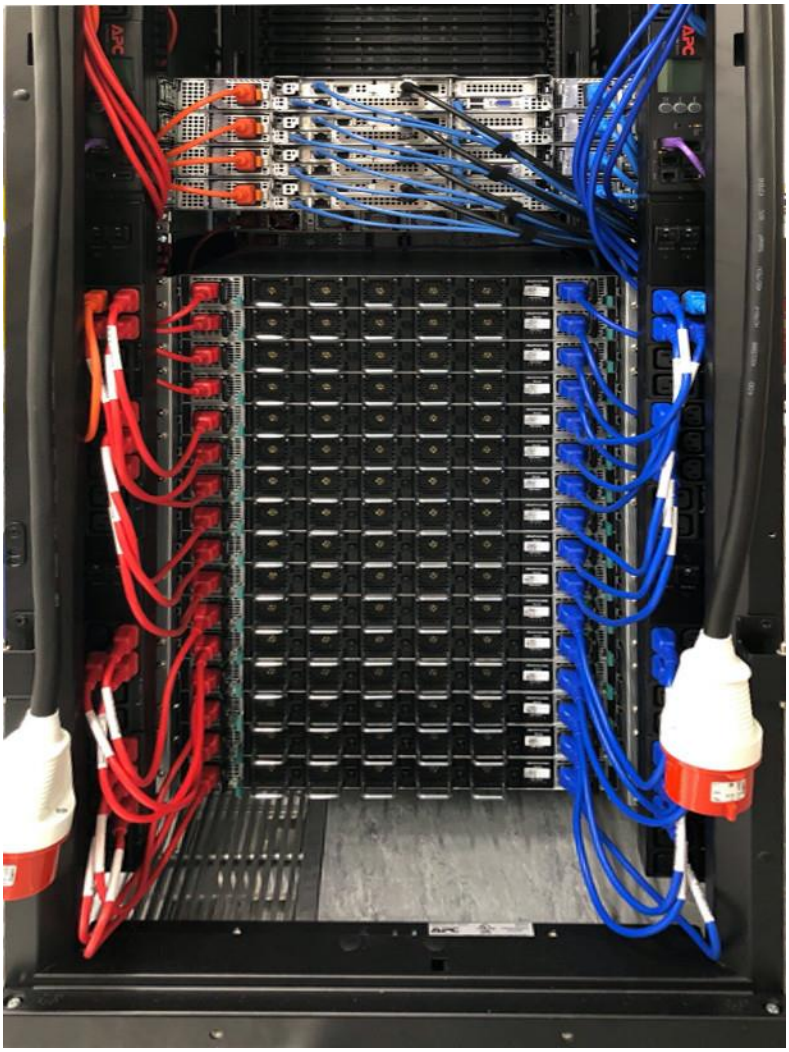
Please note the correct orientation of the IPU-M2000, server and switch units in the rack to ensure correct airflow.

The front interface of the IPU-M2000 units (connectivity ports) should be matched with the front door of the rack (cold aisle). The rear interface of the server and switches (power and fans) should be matched with the rear door of the rack (hot aisle).



Completed rack - cold aisle (four-server version)

Note that this photo shows a four-server version of the IPU-POD$_{64}$. The default reference design has one server which would be the server in the lowest position, closest to the switches.



Completed rack – hot aisle (four server version)

Note that this photo shows 3x blue RJ45 cables in each R6525 server. In the default build, servers 2 - 4 only have 2x blue RJ45 cables. See later sections for more information about server cabling.

Note also that this photo shows a four-server version of the IPU-POD$_{64}$. The default reference design has one server which would be the server in the lowest position, closest to the switches.

3.1 Equipment checklist

Description	Quantity (1 server)	Quantity (4 server)
Rack (AR3300SP)	1	1
Blanking panels (APC AR8136BLK)	23 pieces for 42U reference rack (delivered in packs of 10)	20 pieces for 42U reference rack (delivered in packs of 10)
AP8886 PDU	2	2
Hardware mounting kit (APC AR8100)	1	1
PDU bracket kit APC (AR7711)	2	2
Graphcore IPU-M2000	16	16
IPU-M2000 slider kits	16	16
Dell R6525 server	1	4
Arista DCS-7010T-48-F switch	1	1
Arista DCS-7060CX-32S-F switch	1	1
2m purple Ethernet	2	2
1.5m blue Ethernet	11	17
1m blue Ethernet	9	9
1m yellow Ethernet	12	12
1.5m yellow Ethernet	4	4
1m red Ethernet	2	2
0.15m red Ethernet	30	30
1m QSFP28	8	8
1.5m QSFP28	9	12
0.3m OSFP	60	60
1m OSFP	4	4
0.5m red 10A C14 to C15	12	12
1m red 10A C14 to C15	4	4
0.5m blue 10A C14 to C15	12	12
1m blue 10A C14 to C15	4	4
1m red C13 to C14	2	5
1.5m red C13 to C14	1	1

1m blue C13 to C14	2	5
1.5m blue C13 to C14	1	1
Velcro	1	1

3.2 Preparing the rack

3.2.1 **Rail distance**

The IPU-M2000 mounting system requires a rail-to-rail distance of 720mm. This document describes the adjustments required for an AR3300SP rack. If using a different rack this rail distance must be observed.

3.2.2 **Unpackaging the rack**

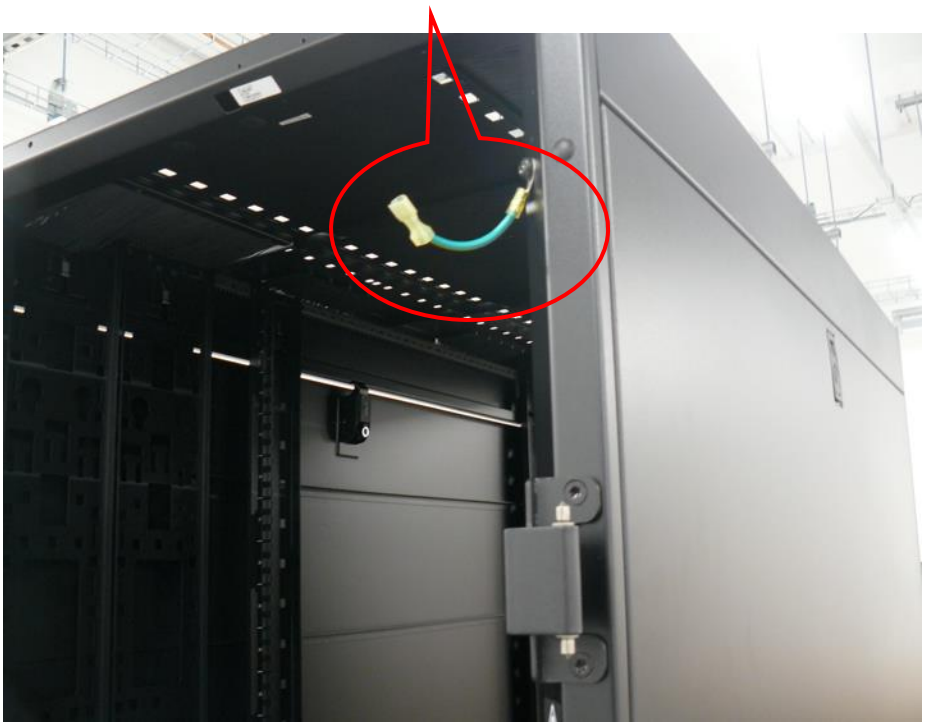
Follow the instructions to remove the outer packaging of the APC AR3300SP rack, ensuring that you safely store these materials for later repackaging. Do not remove the rack from the shock pallet.

Remove the white bag from the rack. This contains screws and cage nuts to be used in the assembly of the components into the rack.

3.2.3 Removing the side panels and doors

Remove the front and rear doors from the rack.

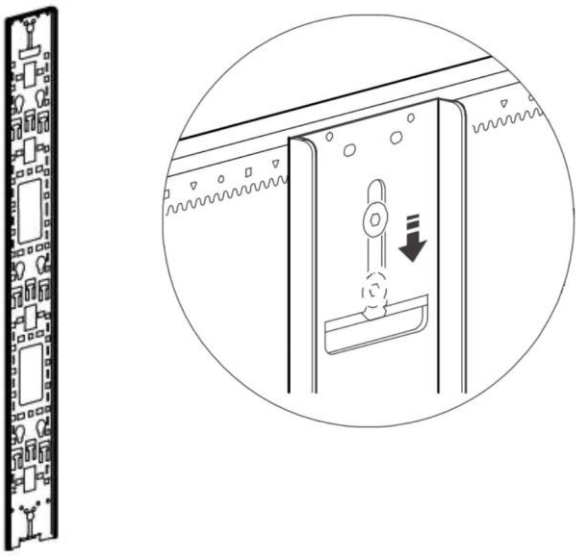
Ensure the earth straps are disconnected before the doors are removed.



Remove the top and bottom side panels:

3.2.4 Removing the vertical accessory channels

Using a Torx TX30 screwdriver, remove two accessory channels from the rack.

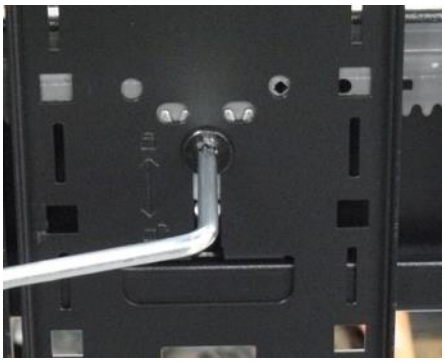


Accessory channel removal

3.2.5 Adjusting the rear accessory channels

Set the rear accessory channel to the furthest position in the rack.

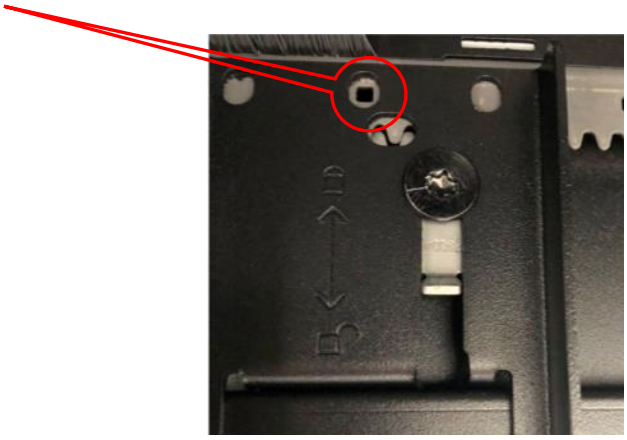
Tighten up the screws ensuring the teeth engage into the slots in the rail:



3.2.6 Adjusting the rear vertical rails

Using a Torx TX30 screwdriver, make both rear vertical rack rails loose and freely movable.

Position the rear vertical rack rails such that there is 20mm of distance between the rear face of the vertical rack rail and the racks rear frame. This should result in a square symbol being visible through the alignment window at the top and bottom of the rail.



Secure the rail into position by moving the TX30 screws back upwards such that the teeth engage with both the supporting rail. This must be done at the top and bottom of the bracket.

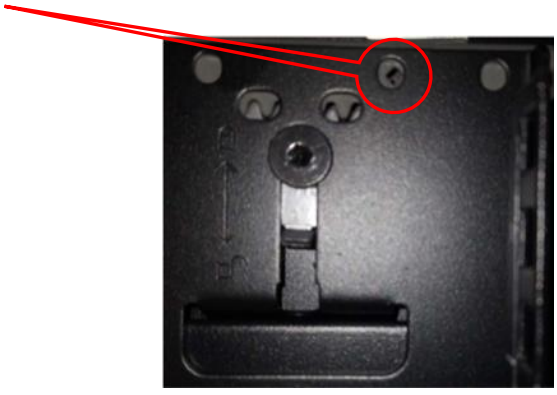
3.2.7 Adjusting the front vertical rails

Using a Torx TX30 screwdriver, make both front vertical rack rails loose and freely movable.

Install the accessory channels in the front of the rack (one on the left hand side, one on the right hand side) at the frontmost position possible, moving the TX30 screws back upwards such that the teeth engage with both the supporting rails - this must be done at the top and bottom of the bracket.

Note	To ensure the clips on the accessory channels align with the channel in the rack, lift the accessory channels through the cut-out in the top of the rack and then drop them down onto the channels.

Move the vertical rack rails tight against the vertical cable organisers such that only a single diamond symbol is visible through the alignment window at the top and bottom of the rail:



Secure the rail into position by moving the TX30 screws back upwards such that the teeth engage with both the supporting rails. This must be done at the top and bottom of the bracket.

3.2.8 Installing the rack rails

M2000 rack rail kit – unboxed

The IPU-M2000 rail kit comprises two mated inner and outer rack rails and an accessory bag containing screws. The inner rail affixes to the body of the IPU-M2000 and the outer rail affixes to the vertical rack rails in the server cabinet.

Firstly, separate the mated inner and outer rails:

1) Fully extend the rails by pulling on the end which has the captive thumb screw attached:

2) Whilst pulling on the thumb screw end of the rails, push the white plastic release tab towards the thumb screw end:

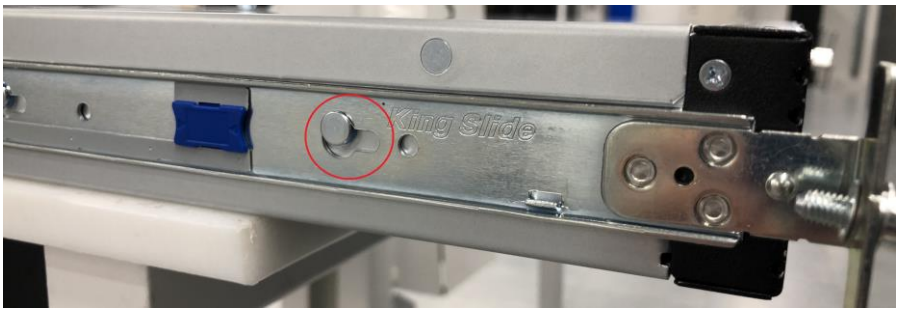
3) The inner and outer rails will now separate:

Mate the inner rails (the thinner of the two separated rails which has a captive thumb screw at one end) to the body of the IPU-M2000. Please note that the inner rails are mirrored and are not handed. As such, the procedure for inner rail fixing is the same for the left and right hand inner rails.

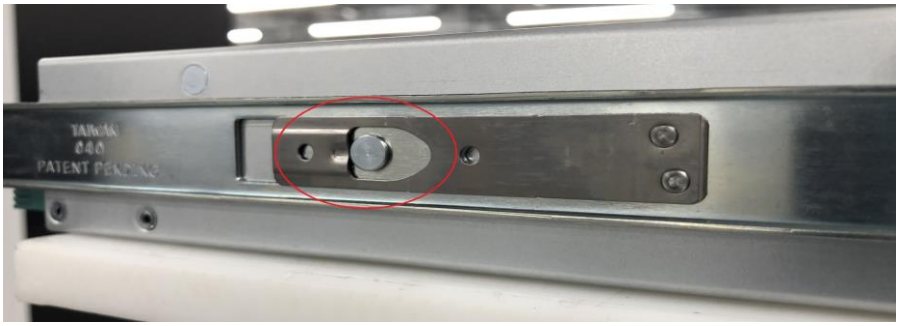
The inner rail should be oriented such that the captive thumb screw end is at the end of the IPU-M2000 containing the network receptacles.

To affix the inner rail to the body of the IPU-M2000:

1) Offer up the inner rail to the side of the IPU-M2000 and ensure that all fixing pins are sitting within the enlarged opening of the retention channel:



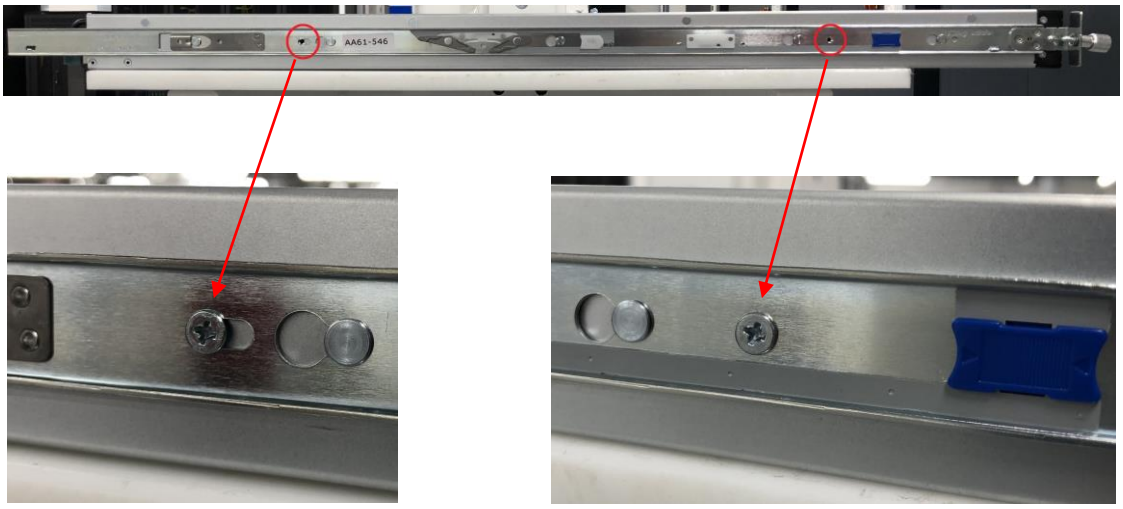
2) Push the inner rail towards the end of the IPU-M2000 containing the network receptacles, you should hear a click as the latching mechanism locks behind the head of a fixing pin:



3) Ensure all fixing pins are correctly engaged with their respective retention channel.

4) Locate the four flat head fixing screws from the rack rail accessory bag:

5) Using the above screws, affix the inner rail to the body of the M2000:



The inner rails are now securely affixed to the IPU-M2000 body.

Place the outer rails to one side for later use:

IPU-M2000 outer rails

3.2.9 Installing PDU brackets

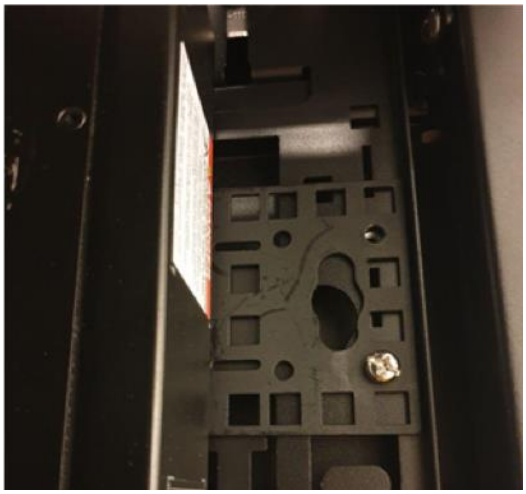
Install four cage nuts on the outside at the top and bottom to both of the accessory channels as shown below.

Top PDU bracket cage nuts

Bottom PDU bracket cage nuts

Screw the PDU support brackets to the inside of the cabinet.

The PDU brackets should be installed at the rear of the rack - one bracket on top with 9cm distance from the top of rack and one bracket on the bottom with 12cm distance from the bottom of rack. The figure below illustrates this. Follow the PDU bracket installation instructions.



PDU support bracket

3.3 Installing the equipment

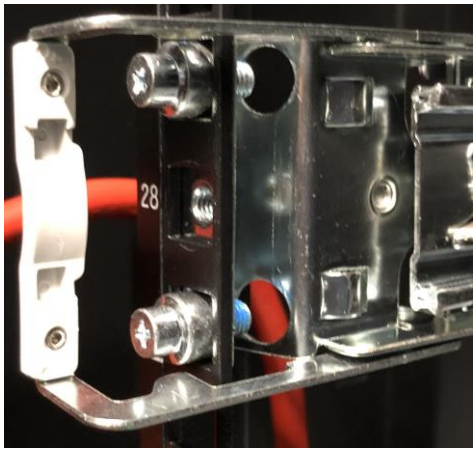
The following sections describe the installation of the IPU-M2000s, PDUs, servers, ToR and management switches into the rack.

3.3.1 **Installing the IPU-M2000s**

Earlier in the guide we affixed the inner rack rails to the IPU-M2000 body. We now need to install the outer rack rails into the rack.

It is possible to identify the front and rear of the outer rail by finding the large metal latching mechanism – this is to be located at the rear of the rack. The outer rail is also embossed with the text “FRONT” at the front end of the rail.

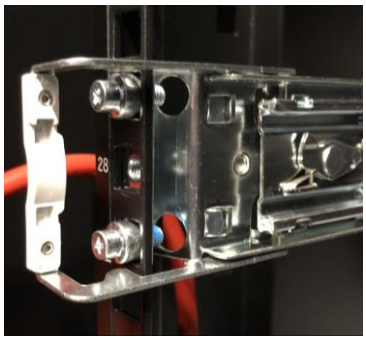
The Outer rail large metal latch end is to be installed at the rear of the rack:



For each rack U in U1 through U16 (inclusive), perform the steps below with both the left hand and right hand outer rack rails:

1) Pull on each end of the outer rail to adjust the rail length to suite your rack
2) Locate the front end of the outer rail and hold it behind the square holes in the vertical rack rail for your installation U. Pull the outer rail towards the vertical rack rail and the latching mechanism will click and hold the outer rail in place

3) Locate the rear end of the outer rail and slightly open the large metal latch, then press the upper and lower locating pins into the square holes. Release the large metal latch and the outer rail will now be secured to the vertical rack rail:



4) Included in the rack rail accessory bag are two screws and two washers. One screw with one washer should be screwed through the vertical rack rail and into the outer rack rail threaded hole. The washer should be used in such a way that the washer sits flush with the head of the screw –like a cup.

   This should be repeated for both outer rack rails.

To install an IPU-M2000 unit into the rack rails, perform the following steps:

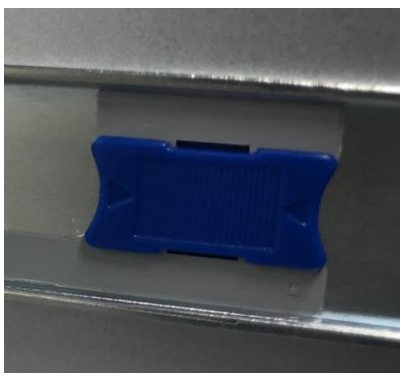
1) Pull the sliding rail located within the outer rack rail completely forward such that it locks into the fully extended position:

2) Place the IPU-M2000 onto an appropriately suited server lift and adjust the height such that it is suitable for the sliders. If a lift is not available, this is a two person operation.

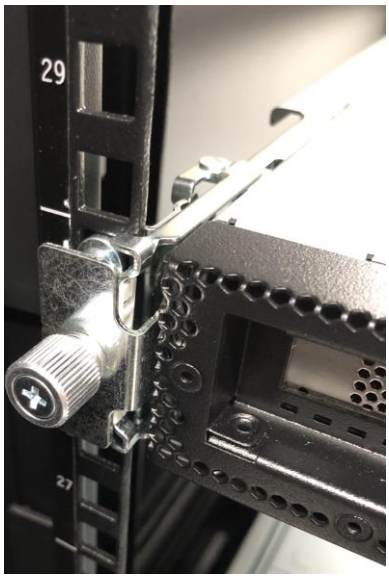
3) Slide the protruding inner rails into the receiving channel of the extended outer rails

4) Whilst the server lift is supporting the full weight of the IPU-M2000, slide the IPU-M2000 into the extended outer rails until you feel both sides engage a stopping mechanism

5) Simultaneously pull on the blue tabs for the release mechanism at each side of the IPU-M2000 and then push the IPU-M2000 unit fully into the rack:



6) Screw the captive thumb screw into the inner rack rail:



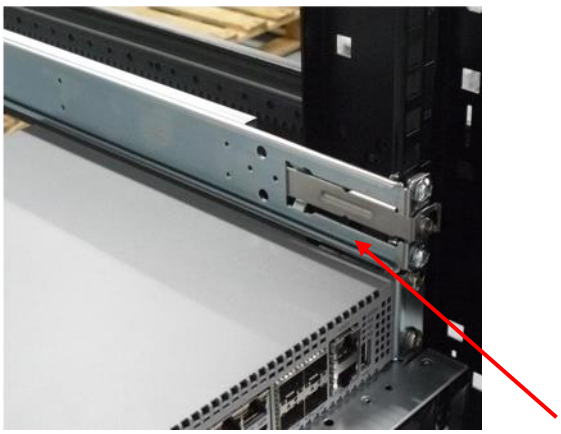
3.3.2 Installing the management switch

Insert 2 cage nuts, inside the rack, on either side of the rack into the top and bottom positions of location 17.

Place the management switch on top of the last IPU-M2000 and screw it into position using 4x M6 screws.

3.3.3 Installing the ToR switch

Fit the sliders for the ToR switch into position 18 on the rack ensuring both ends of the slider are pushed firmly into the mounting slots on the rack rail.



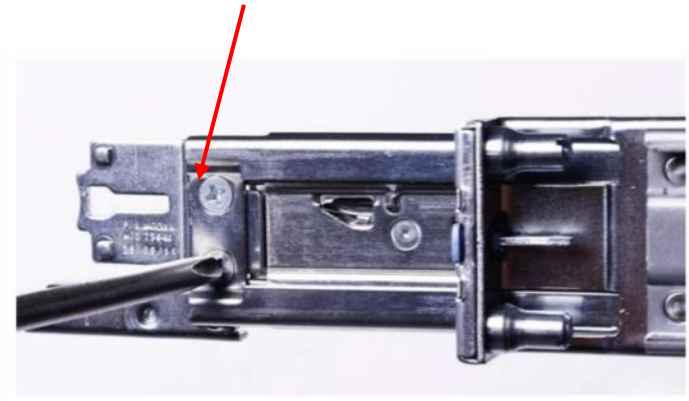
Insert the ToR switch ensuring the wheel on the switch is located in the groove on the slider.

3.3.4 Installing the PDUs

Install the two PDUs vertically – at the rear of the rack, one on the left side and one on the right side. Push the mains cable through the roof of the rack and then clip the PDUs onto the PDU bracket as shown below:

3.3.5 Installing the Dell R6525 server(s)

Remove and discard the cable management arm brackets from the rear of each tool-less sliding rail.



Install the tool-less sliding rail kit(s). The reference design has a single server installed in rack slot #19; in four server configurations they are installed in rack slots #19-#22.

Pull out the rail and fit the server to the rail ensuring the T pins on the side of the server locate in the slots on the rail. Ensure that the power supplies on the server face the rear of the rack.

Note	**Use an appropriate server lift or have two people installing the servers to ensure correct fitting**

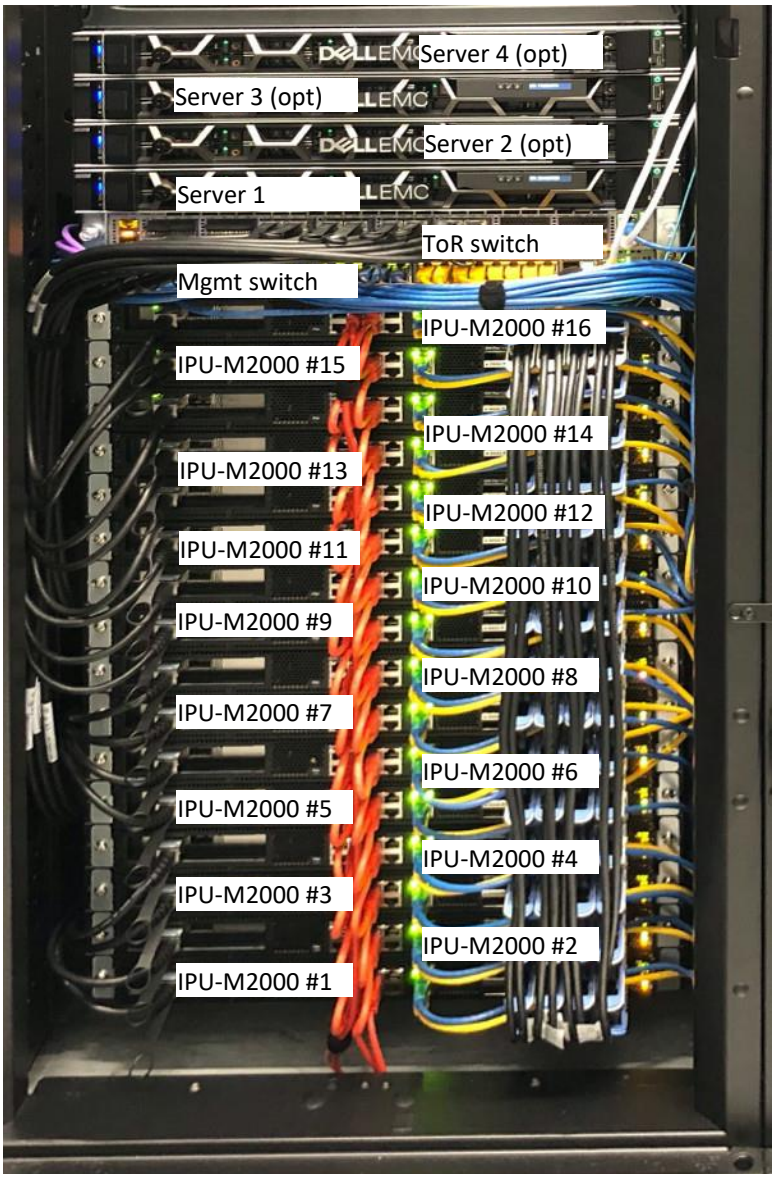
Push the server gently from the front to lock it into the slides then press the tab on the side of the slides and push the server fully home in the rack. Repeat the above process for each server if installing multiple servers.

Remove the Velcro tape from the light pipes on the rear of the servers.

Remove the small plastic tab on the left front side of the server bezel and clip the bezel in place on the front of the server ensuring the connection pins on the right hand side of the bezel line up with the connector on the server, as shown below:

3.4 Wiring the rack

The following sections detail the wiring of the rack and the dressing of the cables within the rack. For reference, the IPU-M2000s and server(s) are numbered, as shown below:

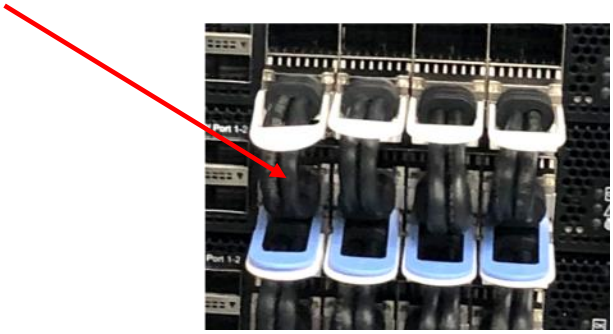


3.4.1 IPU-M2000 to IPU-M2000 IPU-Link connectivity (OSFP)

There are 8 OSFP IPU-Link ports on the right side of each IPU-M2000. Using the supplied 60x 0.3M OSFP cables, and starting at IPU-M2000 #1 (bottom-most IPU-M2000 in the rack), link the top row of four ports (5-8) to the bottom row of four ports (1-4) in the IPU-M2000 that is installed directly above (see figure and table below). This applies to all IPU-M2000s except for the top row (5-8) of the top-most IPU-M2000 (#16) and the bottom row (1-4) of the bottom-most IPU-M2000 (#1), which are connected together using the 1m OSFP cables.

Before attempting to install the OSFP cables, it is beneficial to manipulate the cable to form a tight loop. During manufacture and shipping, the cables can form quite a stiff shape, so manipulating the cables before installing them reduces stresses on the socket during install.

After installing a cable, pull gently on the black cable to ensure the plugs are firm in the sockets on the IPU-M2000.



Note	**The white tab is on the top of the cable when inserted into the IPU-M2000.**

IPU-M2000 to IPU-M2000 IPU-Links		Cables
IPU-M2000 # 15 IPU-Link ports 5,6,7,8	IPU-M2000 # 16 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 14 IPU-Link ports 5,6,7,8	IPU-M2000 # 15 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 13 IPU-Link ports 5,6,7,8	IPU-M2000 # 14 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 12 IPU-Link ports 5,6,7,8	IPU-M2000 # 13 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 11 IPU-Link ports 5,6,7,8	IPU-M2000 # 12 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 10 IPU-Link ports 5,6,7,8	IPU-M2000 # 11 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 9 IPU-Link ports 5,6,7,8	IPU-M2000 # 10 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 8 IPU-Link ports 5,6,7,8	IPU-M2000 # 9 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 7 IPU-Link ports 5,6,7,8	IPU-M2000 # 8 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 6 IPU-Link ports 5,6,7,8	IPU-M2000 # 7 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 5 IPU-Link ports 5,6,7,8	IPU-M2000 # 6 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 4 IPU-Link ports 5,6,7,8	IPU-M2000 # 5 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 3 IPU-Link ports 5,6,7,8	IPU-M2000 # 4 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 2 IPU-Link ports 5,6,7,8	IPU-M2000 # 3 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 1 IPU-Link ports 5,6,7,8	IPU-M2000 # 2 IPU-Link ports 1,2,3,4	OSFP 0.3m
IPU-M2000 # 1 IPU-Link ports 1,2,3,4	IPU-M2000 # 16 IPU-Link ports 5,6,7,8	OSFP 1m

IPU-M2000 OSFP port mapping

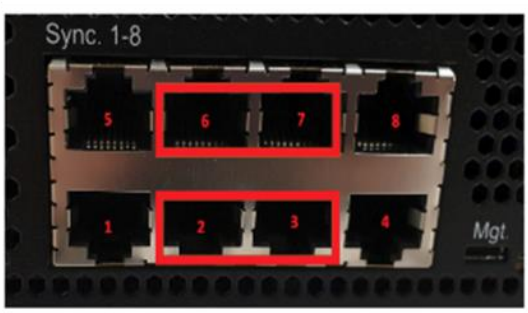
The figure below shows the final IPU-M2000 to IPU-M2000 IPU-Link cabling.

3.4.2 IPU-M2000 to IPU-M2000 Sync-Link cabling

Using the 0.15 M Red Ethernet RJ45 cable, wire the 16 IPU-M2000s as follows:

Starting from IPU-M2000 #1 at the bottom of the rack, insert one side of a cable into port 6 and one side of another cable into port 7.

Insert the other side of the cable from port 6 of IPU-M2000 #1 into Port 2 of IPU-M2000 #2 and the other side of the cable from Port 7 of IPU-M2000 #1 into Port 3 of IPU-M2000 #2.



Continue the cabling for all IPU-M2000s. When completed the top row (6,7) of the top-most IPU-M2000 (#16) connectors should be connected to the bottom row (2,3) of the bottom-most IPU-M2000 (#1) connectors with 1M Red Ethernet RJ45 cables.

The figure below shows the final IPU-M2000 in-rack Sync-Link cabling.

Sync-Link cabling

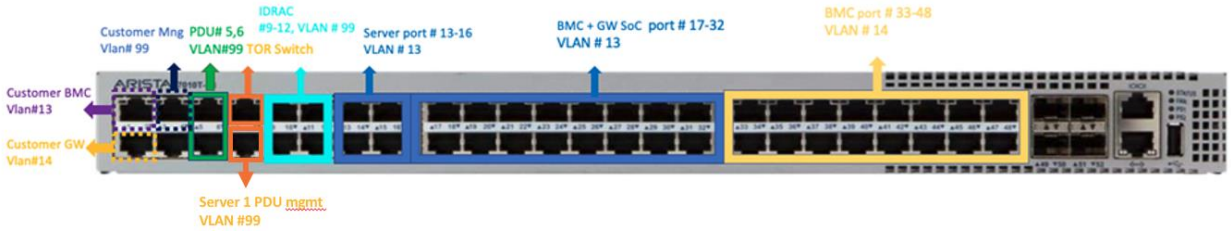
IPU-M2000 to IPU-M2000 Sync-Link connections		**Cables**
IPU-M2000 # 15 IPU-Sync ports 6-7	IPU-M2000 # 16 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 14 IPU-Sync ports 6-7	IPU-M2000 # 15 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 13 IPU-Sync ports 6-7	IPU-M2000 # 14 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 12 IPU-Sync ports 6-7	IPU-M2000 # 13 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 11 IPU-Sync ports 6-7	IPU-M2000 # 12 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 10 IPU-Sync ports 6-7	IPU-M2000 # 11 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 9 IPU-Sync ports 6-7	IPU-M2000 # 10 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 8 IPU-Sync ports 6-7	IPU-M2000 # 9 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 7 IPU-Sync ports 6-7	IPU-M2000 # 8 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 6 IPU-Sync ports 6-7	IPU-M2000 # 7 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 5 IPU-Sync ports 6-7	IPU-M2000 # 6 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 4 IPU-Sync ports 6-7	IPU-M2000 # 5 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 3 IPU-Sync ports 6-7	IPU-M2000 # 4 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 2 IPU-Sync ports 6-7	IPU-M2000 # 3 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 1 IPU-Sync ports 6-7	IPU-M2000 # 2 IPU-Sync ports 2-3	RJ45 0.15 red
IPU-M2000 # 1 IPU-Sync ports 2-3	IPU-M2000 # 16 IPU-Sync ports 6-7	RJ45 1m red

IPU-M2000 Sync-Link port mapping

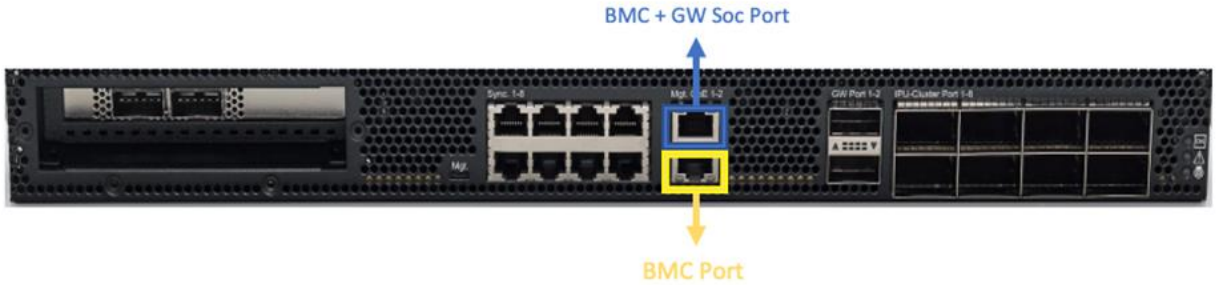
3.4.3 IPU-M2000 to management switch cabling (RJ45)

There are 2 Ethernet ports in the middle of each IPU-M2000 (see figure below). One of them is a BMC + GW SoC port (upper port) and the other is a BMC port (lower port). These are connected from each IPU-M2000 to the management switch with RJ45 cables. The cables required are:

- 12x RJ45/Yellow 1.0M and 4x RJ45/Yellow 1.5m (BMC)
- 8x RJ45/Blue 1.0m and 8x RJ45/Blue 1.5m (BMC + GW SoC)



Management switch



IPU-M2000

The port allocation is as follows:

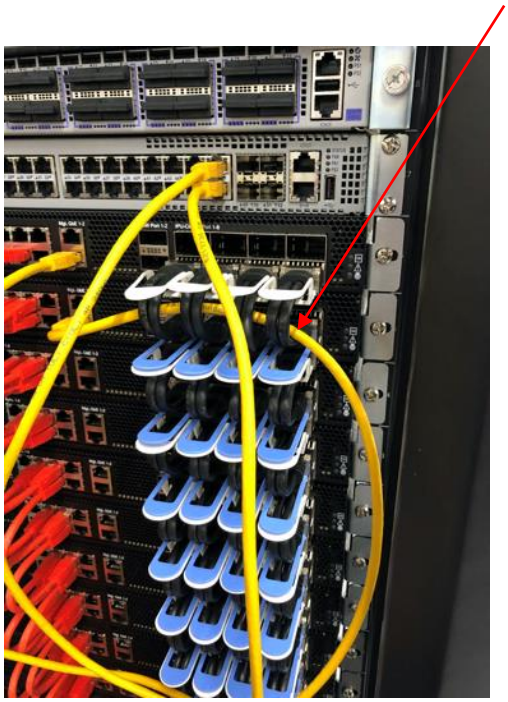
Port	Allocation
1	Customer uplink for BMC + GW
2	Customer uplink for BMC-only (future update)
3	Customer management interface
5,6	PDU management
7	ToR to management switch
9-12	1GbE server management (e.g. iDRAC)
13-16	Server data ports
17-32	IPU-M2000 BMC + GW combined management
33-48	IPU-M2000 BMC-only management (future update)

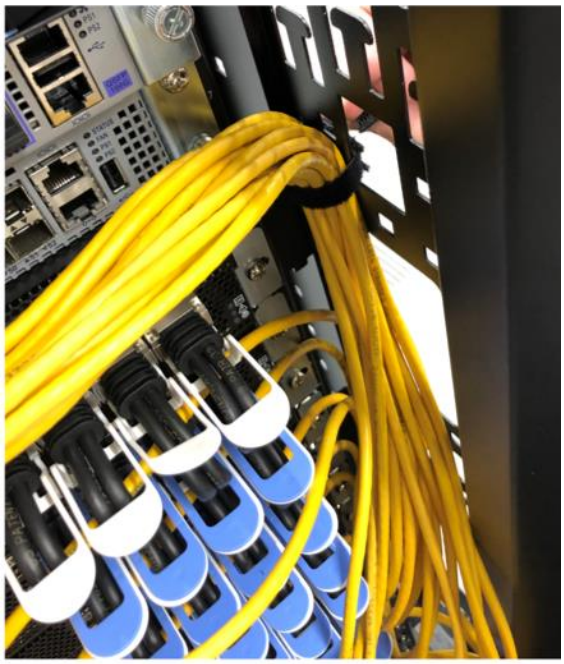
Management switch port mapping

3.4.4 Management switch – BMC wiring

Start wiring using a 1.0M yellow cable and insert one end into port 48 of the management switch.

Run the cable through the loop in the OSFP connector as show below and connect it to the BMC port on IPU-M2000 #16 (top IPU-M2000 in the rack).





Repeat the process for ports 47 to 37 using 1.0m yellow cables and ports 33 to 36 using 1.5m cables.

When all cables have been connected, dress the loom down the side of the cabinet and secure the bundle with a Velcro strip as shown in the picture above.

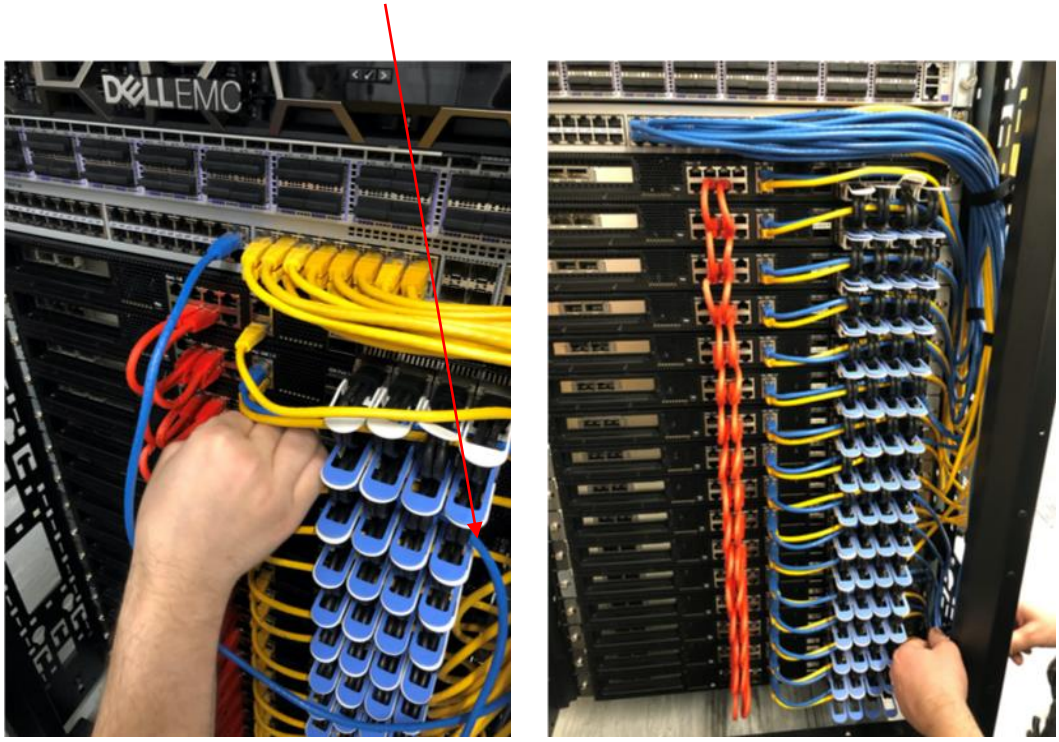
The port mapping between the management switch and the IPU-M2000 BMC sockets is shown in the following table.

IPU-M2000 management GW-only mapping		Cables
IPU-M2000 # 16 BMC port	Management switch port # 48	RJ45 1.0m yellow
IPU-M2000 # 15 BMC port	Management switch port # 47	RJ45 1.0m yellow
IPU-M2000 # 14 BMC port	Management switch port # 46	RJ45 1.0m yellow
IPU-M2000 # 13 BMC port	Management switch port # 45	RJ45 1.0m yellow
IPU-M2000 # 12 BMC port	Management switch port # 44	RJ45 1.0m yellow
IPU-M2000 # 11 BMC port	Management switch port # 43	RJ45 1.0m yellow
IPU-M2000 # 10 BMC port	Management switch port # 42	RJ45 1.0m yellow
IPU-M2000 # 9 BMC port	Management switch port # 41	RJ45 1.0m yellow
IPU-M2000 # 8 BMC port	Management switch port # 40	RJ45 1.0m yellow
IPU-M2000 # 7 BMC port	Management switch port # 39	RJ45 1.0m yellow
IPU-M2000 # 6 BMC port	Management switch port # 38	RJ45 1.0m yellow
IPU-M2000 # 5 BMC port	Management switch port # 37	RJ45 1.0m yellow
IPU-M2000 # 4 BMC port	Management switch port # 36	RJ45 1.5m yellow
IPU-M2000 # 3 BMC port	Management switch port # 35	RJ45 1.5m yellow
IPU-M2000 # 2 BMC port	Management switch port # 34	RJ45 1.5m yellow
IPU-M2000 # 1 BMC port	Management switch port # 33	RJ45 1.5m yellow

3.4.4.1. Management switch – BMC + GW SoC wiring

Start wiring using a 1.0m blue RJ45 cable and insert one end into port 32 of the management switch.

Run the cable through the loop in the OSFP connector as show below and connect it to the BMC+GW port on IPU-M2000 #16.



Repeat the process for ports 31 to 25 using 1.0m blue cables and ports 24 to 17 using 1.5m blue cables. When all cables have been connected, dress the loom down the side of the cabinet and secure the bundle with a Velcro strip.

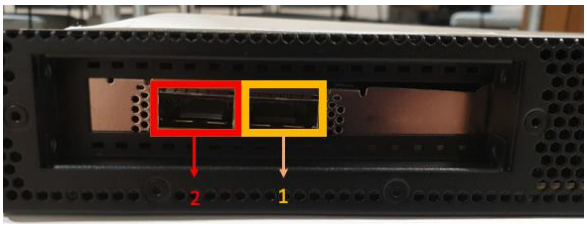
Using a 1.0m blue cable, connect Port 7 of the management switch to the top RJ45 connector on the ToR Switch.

With the IPU-M2000s numbered from bottom to top, the BMC + GW SoC ports should be connected to the management switch as shown below:

IPU-M2000 management BMC + GW port mapping		**Cables**
IPU-M2000 # 16 BMC + GW Soc port	Management switch port # 32	RJ45 1.0m blue
IPU-M2000 # 15 BMC + GW Soc port	Management switch port # 31	RJ45 1.0m blue
IPU-M2000 # 14 BMC + GW Soc port	Management switch port # 30	RJ45 1.0m blue
IPU-M2000 # 13 BMC + GW Soc port	Management switch port # 29	RJ45 1.0m blue
IPU-M2000 # 12 BMC + GW Soc port	Management switch port # 28	RJ45 1.0m blue
IPU-M2000 # 11 BMC + GW Soc port	Management switch port # 27	RJ45 1.0m blue
IPU-M2000 # 10 BMC + GW Soc port	Management switch port # 26	RJ45 1.0m blue
IPU-M2000 # 9 BMC + GW Soc port	Management switch port # 25	RJ45 1.0m blue
IPU-M2000 # 8 BMC + GW Soc port	Management switch port # 24	RJ45 1.5m blue
IPU-M2000 # 7 BMC + GW Soc port	Management switch port # 23	RJ45 1.5m blue
IPU-M2000 # 6 BMC + GW Soc port	Management switch port # 22	RJ45 1.5m blue
IPU-M2000 # 5 BMC + GW Soc port	Management switch port # 21	RJ45 1.5m blue
IPU-M2000 # 4 BMC + GW Soc port	Management switch port # 20	RJ45 1.5m blue
IPU-M2000 # 3 BMC + GW Soc port	Management switch port # 19	RJ45 1.5m blue
IPU-M2000 # 2 BMC + GW Soc port	Management switch port # 18	RJ45 1.5m blue
IPU-M2000 # 1 BMC + GW Soc port	Management switch port # 17	RJ45 1.5m blue

3.4.5 IPU-M2000 to ToR switch cabling (QSFP)

The next step is to connect the IPU-M2000s to the ToR switch. There are 2 RNIC ports on the left side of each IPU-M2000 (see figure below). Only **one** of them should be connected from each IPU-M2000 to the ToR switch with either the 1m or 1.5m QSFP cables supplied.



IPU-M2000 RNIC ports (QSFP)

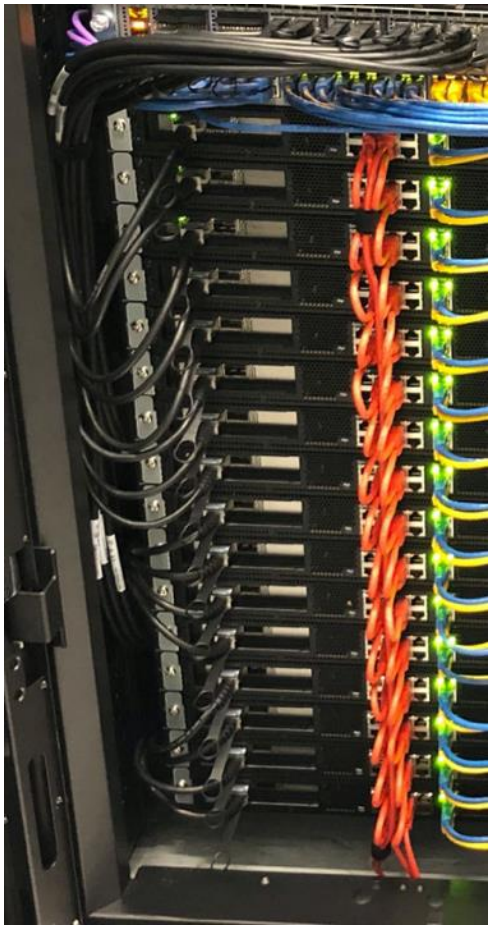
In order to manage the cables, the QSFP cables are divided into two different lengths:

- 8x QSFP 1.0m from IPU-M2000 #9-16 to ToR switch port # 16-9
- 8x QSFP1.5m from IPU-M2000 #1-8 to ToR switch port # 17-24

Start from Port 9 on the ToR switch and connect the cable to Port 2 of IPU-M2000 #16.

Continue cabling from ports 10 to 24 on the ToR switch to Port 2 on IPU-M2000 #15 to IPU-M2000 #1 respectively.

Dress the cables down the side of the cabinet. A set of four cables can be looped into cut-outs on the side of the cabinet. Tie the bundle together at the top with a Velcro strip.



The IPU-M2000s should be connected to the ToR switch as shown below, note that the IPU-M2000s are numbered from bottom to top #1 - #16:

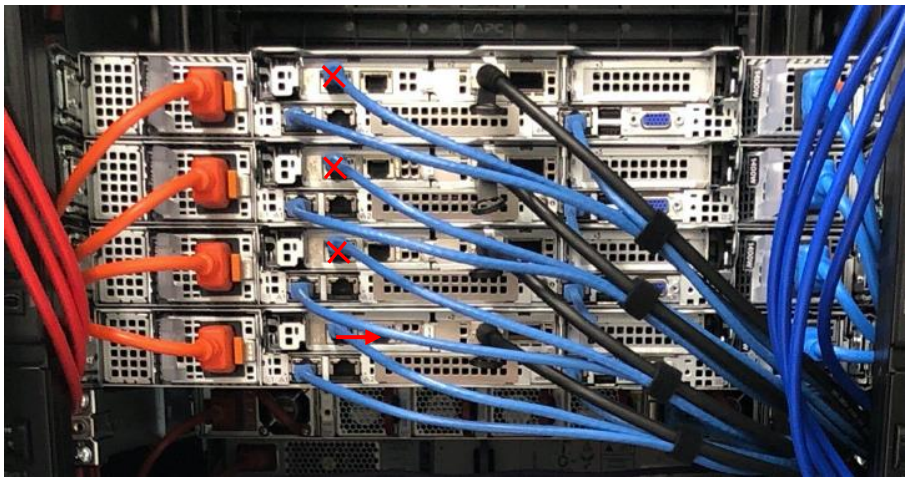
IPU-M2000 RNIC port mapping		Cables
IPU-M2000 # 16 port 2	ToR switch port # 9	QSFP 1.0m
IPU-M2000 # 15 port 2	ToR switch port # 10	QSFP 1.0m
IPU-M2000 # 14 port 2	ToR switch port # 11	QSFP 1.0m
IPU-M2000 # 13 port 2	ToR switch port # 12	QSFP 1.0m
IPU-M2000 # 12 port 2	ToR switch port # 13	QSFP 1.0m
IPU-M2000 # 11 port 2	ToR switch port # 14	QSFP 1.0m
IPU-M2000 # 10 port 2	ToR switch port # 15	QSFP 1.0m
IPU-M2000 # 9 port 2	ToR switch port # 16	QSFP 1.0m
IPU-M2000 # 8 port 2	ToR switch port # 17	QSFP 1.5m
IPU-M2000 # 7 port 2	ToR switch port # 18	QSFP 1.5m
IPU-M2000 # 6 port 2	ToR switch port # 19	QSFP 1.5m
IPU-M2000 # 5 port 2	ToR switch port # 20	QSFP 1.5m
IPU-M2000 # 4 port 2	ToR switch port # 21	QSFP 1.5m
IPU-M2000 # 3 port 2	ToR switch port # 22	QSFP 1.5m
IPU-M2000 # 2 port 2	ToR switch port # 23	QSFP 1.5m
IPU-M2000 # 1 port 2	ToR switch port # 24	QSFP 1.5m

IPU-M2000 RNIC port mapping

The figure below shows the final IPU-M2000 to ToR cabling.

3.4.6 **Dell R6525 server(s) wiring**

All cables should be routed from the rear of the server to the right-hand side (when viewed from the rear), then along the side of the rack using the cable management holes in the vertical rack rails.



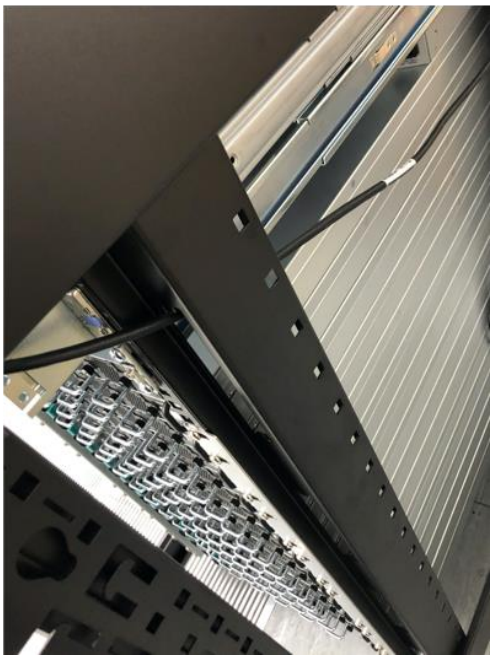
Note that the photo above shows the four-server version. The default build has one server (the one in the lowest position).

Note	Picture to be updated: this picture shows 3x blue RJ45 cables in each R6525 server. In the default build, servers 2 - 4 only have 2x blue RJ45 cables. The cables removed in the default build are marked with a red cross. The additional cable in server 1 moves to the port on the right, as indicated by the red arrow.

3.4.7 **ToR switch to Dell server(s)**

Using 4 of the 1.5m QSFP cables, connect the ToR switch to the server(s) as follows.

1) Feed the cables through the cut-outs in the side of the rack first.



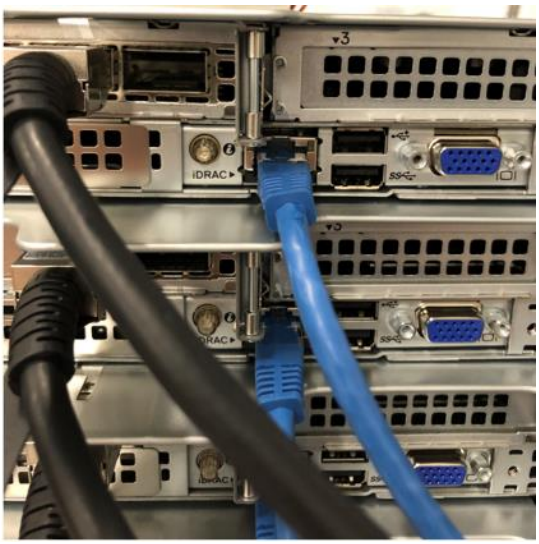
2) Connect the cables to the ToR switch and server: ToR switch ports 8,6,4,2 to each Dell server. ToR switch port 8 connects to the bottom server (server 1). Therefore if you are only using 1 server then only port 8 of the ToR switch is used.

3.4.8 Management switch to Dell server(s) - iDRAC

Using 4 of the 1.5m blue RJ45 cables connect the management switch to the Dell server(s) as follows:

Ports 9 to 12 on the management switch connect to the iDRAC connector on the server(s).

Port 9 is connected to the bottom server (server 1). If you only have one server then only port 9 on the management switch is used.



3.4.9 Management switch to Dell server(s) – network connector

Using 4 of the 1.5m blue RJ45 cables connect the management switch to the Dell server(s) as follows:

Ports 13 to 16 on the management switch connect to the network connector on the server(s). Port 13 is connected to the bottom server (server 1). If you only have one server then only port 13 on the management switch is used.

3.4.9.1. Management switch to Dell server(s) – switch management

Using 1 of the 1.5m blue RJ45 cables connect the management switch to the Dell server(s) as follows:

Port 8 from the management server is connected to the lowest server (server 1). This is used for control of the PDUs in the case where server 1 is used as the management server.

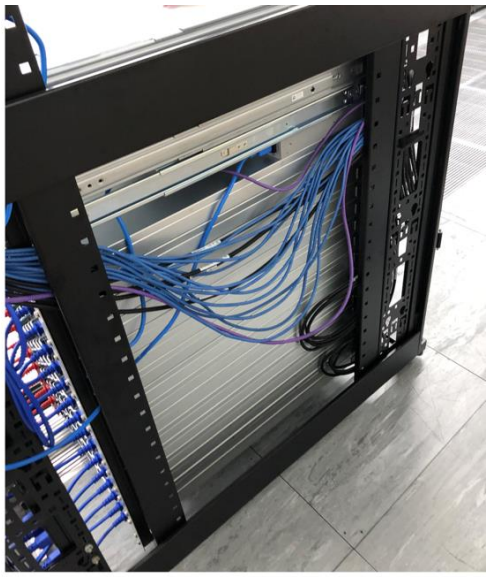
Connect from here to management server port 8

3.4.10 Management switch to PDUs

Using the two 2m purple RJ45 cables connect the management switch to the PDUs as follows:

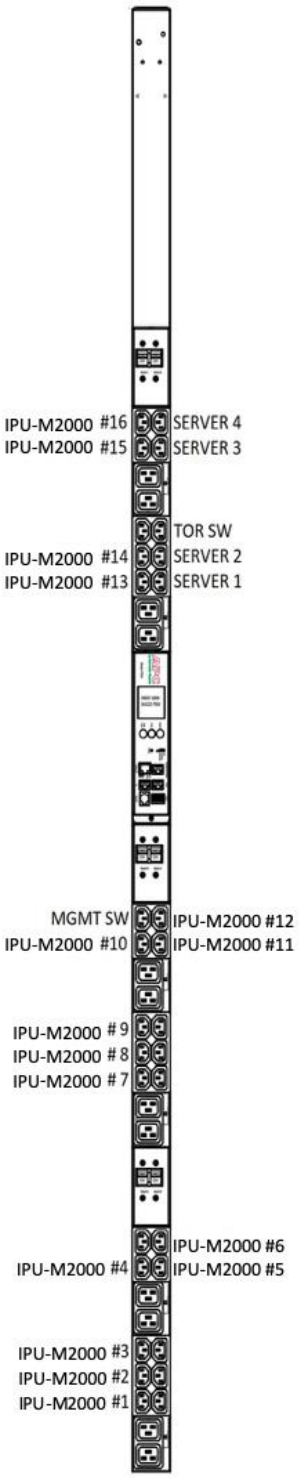
Management switch to PDUs connection		Cables
Ethernet port PDU left side	Management switch port # 5	RJ45 2m
Ethernet port PDU right side	Management switch port # 6	RJ45 2m

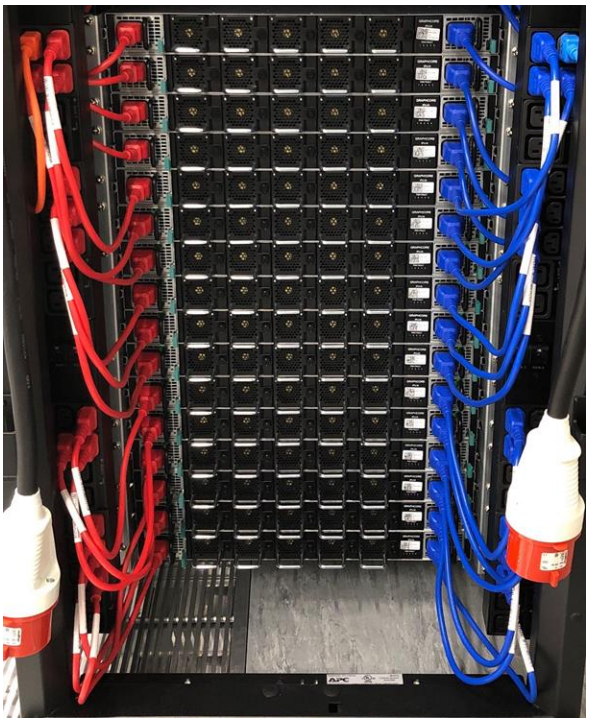
Allow all the cables going to the server to hang down in the rack as shown below. This allows the cables to be pulled slightly if it is necessary to remove them from a server.



3.5 Power cabling

The designated sockets on the PDUs for servers and IPU-M2000s are shown below.





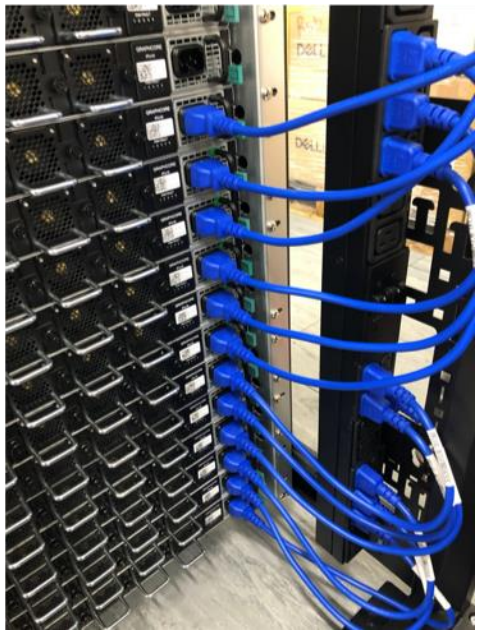
Power cabling

3.5.1 IPU-M2000 power cabling

Start by wiring IPU-M2000 #1 using 0.5m power cables. Ensure that only three IPU-M2000s are connected to the same bank on the PDU. Red cables to the left PDU and blue cables to the right PDU (as seen looking at the rear of the rack).

The following table defines the length of the power cable for each IPU-M2000:

Cable colour	Cable length	IPU-M2000
Blue	1m	IPU-M2000 #13 to IPU-M2000 #16
Red	1m	IPU-M2000 #13 to IPU-M2000 #16
Blue	0.5m	IPU-M2000 #1 to IPU-M2000 #12
Red	0.5m	IPU-M2000 #1 to IPU-M2000 #12



3.5.2 Server power cabling – Dell R6525

Using 1m C13 to C14 power cables (selecting the correct coloured cable to match the PDU colour), follow the wiring table below to connect the server(s) to the PDUs.

Server power wiring

Note: The photo above shows the four-server version. The default reference design has one server (the one in the lowest position).

3.5.3 Switch power cabling

Connect the mains cables to the management switch and ToR switch using C13 to C14 power cables as follows:

Switch	Cable length (red/blue)
Management	1.0m
ToR	1.5m

3.6 Completing the rack

The following steps describe completing the rack: fitting blanking panels and re-installing the doors and side panels.

3.6.1 **Blanking panels**

Install the supplied APC 1U blanking panels in every unoccupied rack slot at the front of the rack.

1U blanking panels

3.6.2 **Front and rear doors**

Re-install the front and rear doors. Ensure the earth cables are reconnected to the cable on the rack.

3.6.3 **Side panels**

Re-install the top and bottom side panels on each side of the rack.

4 IPU-POD$_{64}$ server and switch configuration

This chapter describes how to configure the server(s), and the network switches in an IPU-POD$_{64}$.

Note	Scripts for the server installation and switch configuration can be provided: please contact Graphcore support. These would be appropriate for configuration of a small number of IPU-POD$_{64}$ systems, or for adaption into a central administration service.

Note	An automated method using a USB stick and QR scanner can be provided: please contact Graphcore support. This method might be appropriate where many IPU-POD$_{64}$ need to be configured and tested as stand-alone racks

4.1 Server configuration

4.1.1 Hardware recommendations

The IPU-POD$_{64}$ reference design uses a single PowerEdge R6525 server but up to four servers can be connected. Contact Graphcore sales for details of other supported server types. This document describes the default server (PowerEdge R6525) installation only – other servers may have different installation requirements.

The recommended configuration of the Dell R6525 is as follows:

- Dell R6525 containing dual AMD EPYC 7742 processors
- 24x32GbE RDIMM PC4-25600 ECC registered dual-rank X4 1.2v
- 2x 480GbE SSD-SATA 6Gbps 2.5 inch hot-swap
- 7x 1TB NVME SSD PCIe 4x 3.1
- Dual port Gigabit BASE-T PCIe
- Single/dual port Mellanox ConnectX-5 EN 100Gb/s Ethernet

4.1.2 Storage configuration recommendations

The recommendation is to have two types of server storage: SSD-SATA for the operating system and NVME SSD for data storage.

Operating system:

- 2x 480GbE SSD-SATA units as a RAID 1 via hardware controller
- Partitioned to use ext4 file system

Data storage:

- 7x 1TB NVME SSD units as a logical RAID 6 managed with MDADM
- Partitioned to use xfs file system

4.1.3 Operating system recommendations

Please contact your Graphcore representative or use the support portal support.graphcore.ai for information about operating system support. This document describes the following operating systems:

- Ubuntu 18.04.4 LTS (bionic)
- CentOS 7.2

4.1.4 User accounts and groups

The following accounts are required as part of the default server configuration:

Accounts	
root	A root user account secured with a password is recommended
itadmin	An admin account secured with a password is recommended. Home folder located at /home/itadmin using bash shell.
ipuuser	An account dedicated to IPU software and IPU-POD management software is mandatory Home folder located at /home/ipuuser using bash shell.
poplaruser	An account dedicated to Poplar software is mandatory. Home folder located at /home/poplaruser using bash shell.

The following table gives the default usernames provided on the IPU-POD$_{64}$:

Login to:	Username	Password
IPU-M2000 BMC OS	`root`	The default passwords are available from Graphcore support support.graphcore.ai.
IPU-M2000 GW OS	`itadmin`	
Server as Poplar SDK user	`poplaruser`	
Server as IPU-POD admin user	`ipuuser`	
Server as IT admin user	`itadmin`	
Server iDRAC port	`root`	
100GbE RDMA switch	`admin`	
1GbE Management switch	`admin`	
PDU	`apc`	

The following table gives the required groups provided on the IPU-POD$_{64}$.

Groups	
root:	A root group to locate the root account is mandatory.
dhcpd:	A group to allocate the DHCP service is mandatory (usually is configured automatically while installing the DHCP service).
ipugroup:	A group to allocate *ipuuse*r is mandatory
poplargroup:	A group to allocate *poplaruser* is mandatory.
ipupodgroup:	A group to allocate both *ipuuser* and *poplaruser* is mandatory

Note that users need to have unique user IDs and group IDs.

4.1.5 Ubuntu OS Installation and packages

The Ubuntu OS should be installed from the following default public Ubuntu 18.04.4 repositories:

```
deb http://archive.ubuntu.com/ubuntu/ bionic main restricted
deb http://archive.ubuntu.com/ubuntu/ bionic-updates main restricted
deb http://archive.ubuntu.com/ubuntu/ bionic universe
deb http://archive.ubuntu.com/ubuntu/ bionic-updates universe
deb http://archive.ubuntu.com/ubuntu/ bionic multiverse
deb http://archive.ubuntu.com/ubuntu/ bionic-updates multiverse
deb http://archive.ubuntu.com/ubuntu/ bionic-backports main restricted universe multiverse
deb http://archive.canonical.com/ubuntu bionic partner
deb http://security.ubuntu.com/ubuntu bionic-security main restricted
deb http://security.ubuntu.com/ubuntu bionic-security universe
deb http://security.ubuntu.com/ubuntu bionic-security multiverse
```

In order to have a stable system where IPU related software can run, several packages need to be installed on the system via Aptitude package manager tool:

apt-transport-https	ibverbs-utils	openjdk-8-jdk	python3-virtualenv
autoconf	ipmitool	php-cli	python3-wheel
automake	jq	php-curl	qtcreator
bc	kcachegrind	policykit-1	rdma-core
build-essential	libaio-dev	protobuf-compiler	screen
ccache	libboost-all-dev	python-boto3	software-properties-common
clang	libeigen3-dev	python-dev	sshpass
cmake	libjson-c-dev	python-lxml	subversion
curl	libjson-c-doc	python-numpy	swig
direnv	libpci-dev	python-pip	sysfsutils
dkms	libpixman-1-dev	python-pytest	tar
emacs	libprotobuf-dev	python-recommonmark	tmux
ethtool	libtool	python-requests	u-boot-tools
exuberant-ctags	lldpad	python-setuptools	unzip
flex	m4	python-wheel	valgrind
g++	minicom	python-yaml	vim
gawk	moreutils	python2	virtualenv
gcc	net-tools	python3	wdiff
gdb	netcat	python3-dev	wget
git	parallel	python3-numpy	zip
golang-go	pciutils	python3-pip	
htop	perl	python3-setuptools	

4.1.6 CentOS OS installation and packages

In order to have a stable system where IPU related software can run, several packages need to be installed on the system via yum configuration manager:

bc	libaio-devel	python2-numpy	vim
centos-release-scl	libboost-devel	python2-pip	wdiff
clang	libibverbs-utils	python2-pytest	wget
cmake	libuser	python27-python-devel	
containerd.io	lldpad	qt5-qbase	
devtoolset-7	minicom	rdma-core	
dhcp	moreutils	rh-python36	
dkms	nano	rh-python38-python-lxml	
docker-ce	nc	rh-python38-python-numpy	
docker-ce-cli	net-tools	rh-python38-python-setuptools	
eigen3	ntp	rh-python38-python-wheel	
emacs	parallel	rh-python38-scldevel	
golang-go	pciutils-devel	rh-pyton38	
htop	php-cli	screen	
ipmitool	protobuf-devel	snapd	
java-latest-jdk	python-anymarkup	sshpass	
jq	python-boto3	sysfsutils	
json-c-devel	python-requests	tmux	
json-c-doc	python-wheel	uboot-tools	
kcachegrind	python2	valgrind	

4.1.1 Python packages

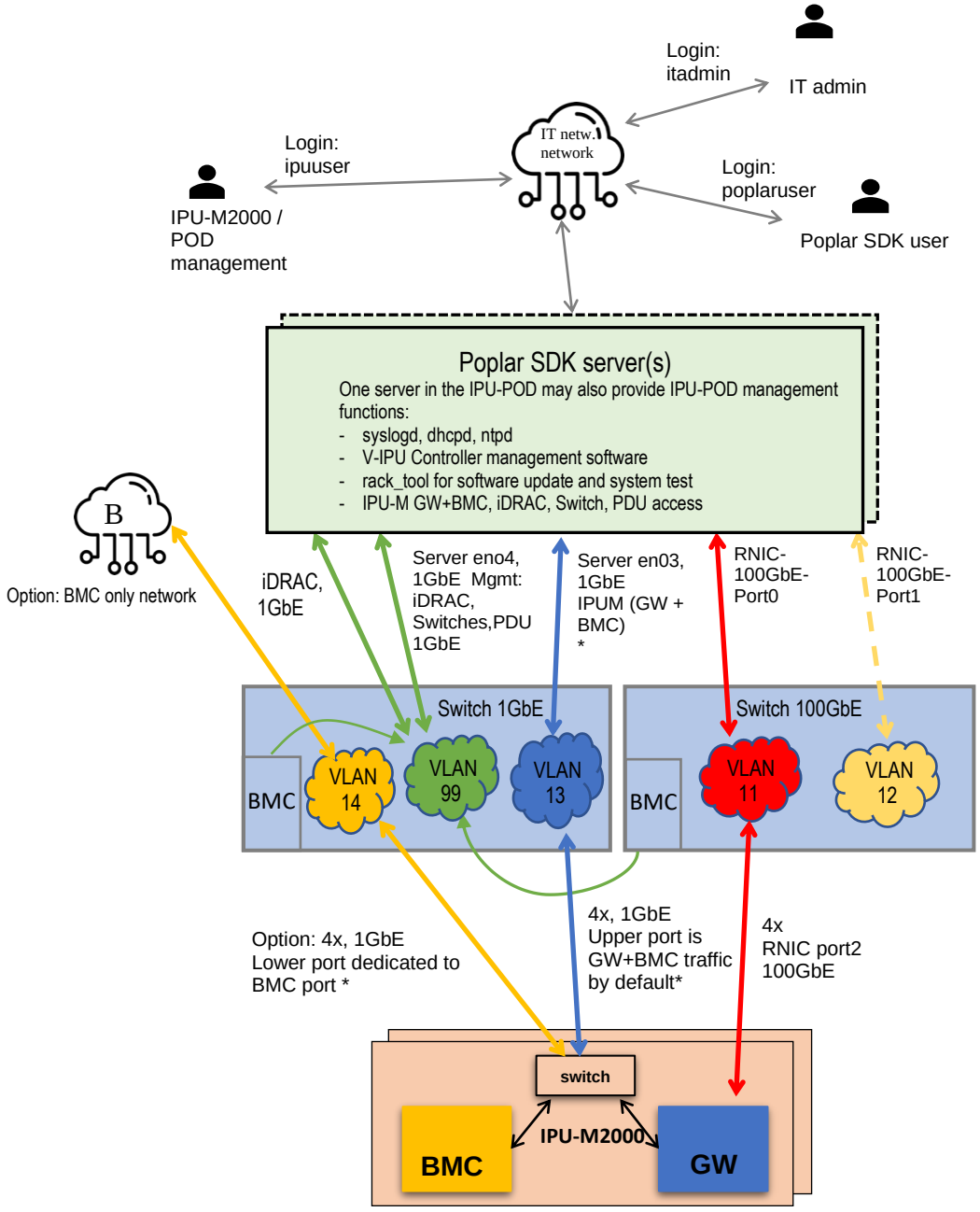
Several python packages are also required. They can be installed using the pip installation tool.

autograd	paramiko	pylint	scp
jstyleson	pep8	pyyaml	yapf
mock	pexpect	requests	

4.2 Network configuration

4.2.1 Overview

The following figure gives a logical overview of the network setup within the IPU-POD$_{64}$.



* Option: VLAN 14 and the cabling to the dedicated BMC port can be provided as an upgrade for customers that want an isolated BMC network. This separates the BMC and GW traffic inside the IPU-M2000 onto the two ports, BMC on lower port, GW on upper port. With this option enabled, VLAN 13 becomes a GW only VLAN. Please contact Graphcore Sales for more information.

IPU-POD$_{64}$ network overview

4.2.2 IPU-POD$_{64}$ network interfaces

Port	Role	Speed	IP address	Config from	VLAN*
IPU-M2000 - BMC	BMC only management (future)	1GbE	10.1.1.1-16/22	Static DHCP	14
IPU-M2000 - GW	BMC+GW management	1GbE	10.1.2.1-16/22	Static DHCP	13
Server - Port1	Mgmt IPU-M2000	1GbE	10.1.3.101/22	Local netplan file	13
Server - Port2	Div management iDRAC, switches & PDUs	1GbE	10.1.6.1/22	Local netplan file	99
Server - iDRAC	Server BMC	1GbE	10.1.6.4-7/22	Manual setup	99
Server - RNIC/Port0	RDMA IPU-M2000	100GbE	10.1.5.101/23	Local netplan file	11
Server - RNIC/Port1	RDMA NAS	100GbE	Site specific	Site specific	11
48x 1GbE + (4x 10G) Switch management port	CLI + Switch BMC management	1GbE	10.1.6.2/22	Manual setup	99
32x 100GbE + (4x 10G) switch management port	CLI + Switch BMC management	1GbE	10.1.6.3/22	Manual setup	99
PDU	Power dist. unit	1GbE	10.1.6.8-11	Dynamic DHCP	99

* port based VLAN in switches (VLAN 13,14 and 99 in 1GbE switch, VLAN 11 in 100GbE switch)

4.2.3 Management switch configuration

Several VLANs need to be configured on the switches to separate traffic for the different hardware and integrate DHCP properly on the system.

The default management switch is Arista DCS-7010T-48-F.

1) Up to 4 interfaces per rack for PDUs and switches management
2) Up to 4 interfaces per rack for server management interfaces
3) Up to 4 interfaces per rack for server facing BMCs and gateways
4) Up to 16 interfaces per rack for combined IPU-M2000 Gateway and BMC
5) Up to 16 interfaces per rack for BMC only connection (option)

Management switch VLAN assignments

These port groups are members of the switch's internal port based VLANs. The VLAN assignments are given below.

- VLAN 13 – IPU-M2000 BMC and GW traffic
- VLAN 14 – IPU-M2000 BMC only traffic (option)
- VLAN 99 – Device management

Device management is intended for switch management, PDU remote management and any other device in the system with remote management interfaces.

A switch configuration file can be provided – please contact Graphcore support for details.

4.2.4 ToR switch configuration

The default ToR switch is an Arista DCS-7060CX-32S-F.

1) Up to 4 interfaces per rack for server RNIC ports to site-specific network
2) Up to 4 interfaces per rack for server RNIC ports carrying IPU-M2000 traffic
3) Up to 16 interfaces per rack for IPU-M2000s RNIC ports

ToR switch VLAN assignments

These port groups are members of the switch's internal port based VLANs. The VLAN assignments are given in the figures below.

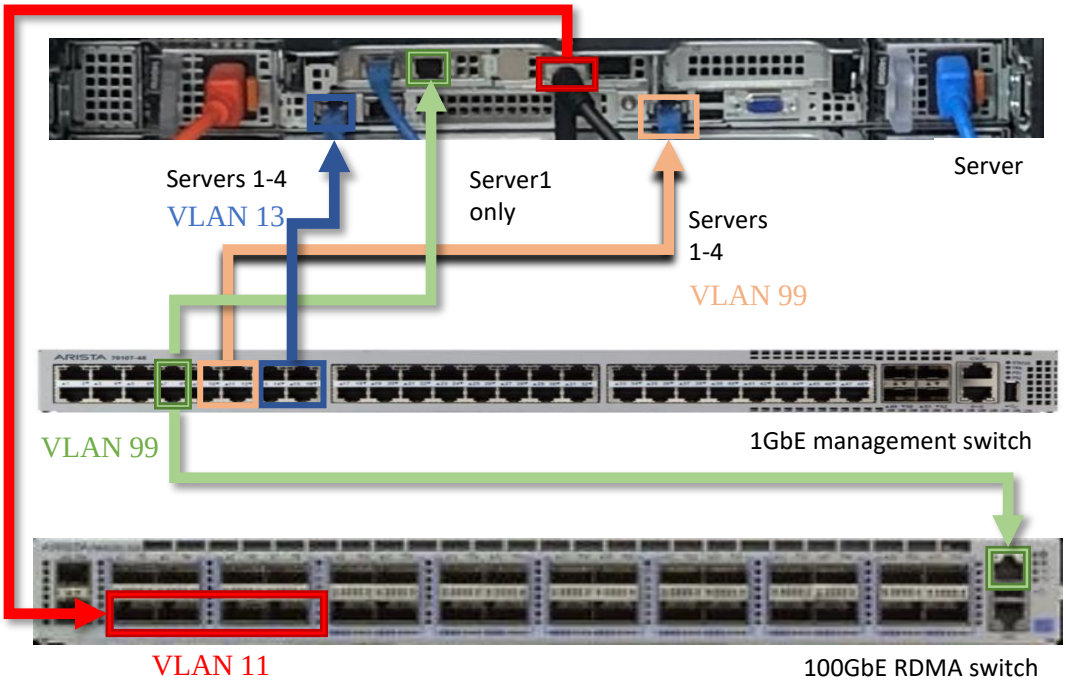
- VLAN 11 – 100Gb RNIC facing IPU-M2000s
- VLAN 12 – 100GbE RNIC facing customer network (optional: in case of uplink to NAS)

A switch configuration file can be provided – please contact Graphcore support for details.

4.2.5 IPU-POD$_{64}$ VLAN assignments

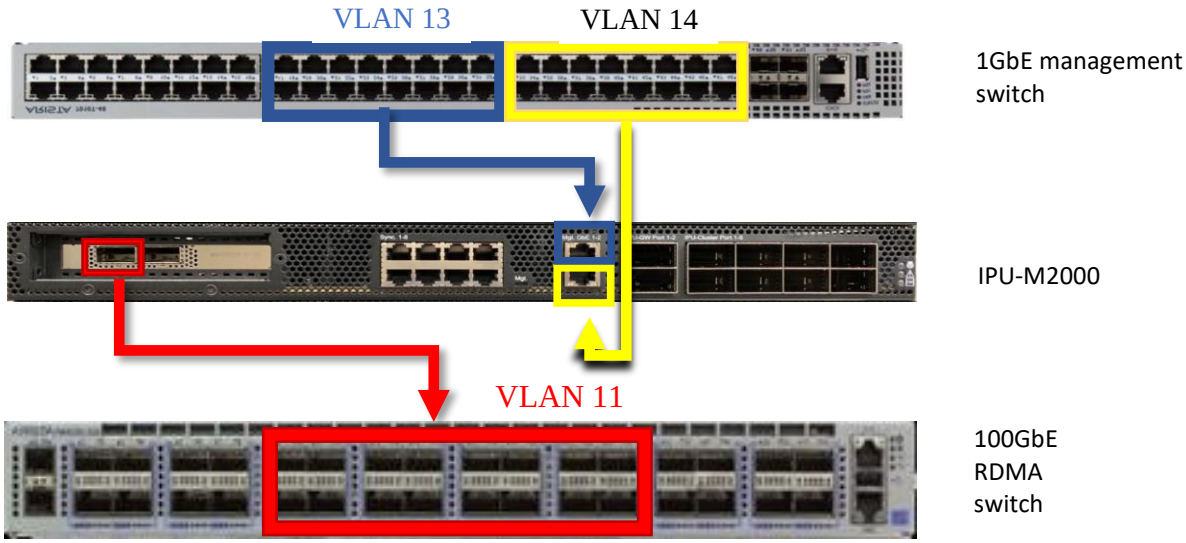
Each switch is configured independently based on the number of interfaces needed for the IPU-POD size (in this case IPU-POD$_{64}$). This section describes the interfaces used in an IPU-POD$_{64}$ with figures showing interface allocation.

The following figure shows the VLAN assignments for the server connecting between the server(s) and the switches. Four server-facing ports are shown for each group on the switches to allow for up to 4 servers in the POD.



Dell R6525 server VLAN connectivity

The following figure shows the VLAN assignments for the IPU-M2000s connecting to the switches.



IPU-M2000 VLAN connectivity

4.2.6 Server network configuration

- It is recommended to use the Netplan network manager to configure the server using netplan configuration files
- Two 1GbE baseT connections are required to the 1GbE management switch for Server 1. Only one 1GbE baseT connection is required to the 1GbE management switch for additional servers. Fixed IP addresses are required.
- One 100Gb connection is required to the 100GbE switch. A fixed IP address is required.
- Subnets for each interface should be capable to contain the number of devices of the installation

Example netplan configuration file:

- The default location for this file is /etc/netplan/01-netcfg.yaml.
- In this example:
  - Interface eno3 is facing 1GbE network with IPU-M2000 BMCs and Gateways.
  - Interface eno4 is facing user's network.
  - Interface enp59s0f1 is facing IPU-M2000 100GbE RNICs.

```
network:
  version: 2
  renderer: networkd
  ethernets:
    eno1np0:
      dhcp4: yes
    eno2np1:
      dhcp4: yes
    eno3:
      addresses:
        - 10.1.3.101/22
    eno4:
      dhcp4: yes
    enp59s0f0:
      dhcp4: yes
    enp59s0f1:
      addresses:
        - 10.1.5.101/23
```

4.2.7 Services: DHCP (Dynamic Host Configuration Protocol)

An ISC-DHCP-Server service is recommended to provide DHCP network configuration to IPU-M2000s. It can be installed from the Ubuntu or CentOS public repositories.

- File structure:
  - /etc/default/isc-dhcp-server (file)
    - This file contains the network interfaces which DHCP is going to use.
    - root:root 0644

```
INTERFACES="eno3 enp59s0f1"
```

  - /etc/dhcp/ (folder)
    - Folder containing DHCP related files.
    - root:dhcpd 0575
  - /etc/dhcp/dhcpd.conf /(file)
    - Main DHCP server configuration file
    - root:root 0444

```
default-lease-time 600;
max-lease-time 1200;
ddns-update-style none;
authoritative;
log-facility local7;
subnet 10.1.4.0 netmask 255.255.254.0 {
    option subnet-mask         255.255.254.0;
    range                      10.1.5.200 10.1.5.254;
}
include "/etc/dhcp/dhcpd.d/vlan-11.conf";
subnet 10.1.0.0  netmask 255.255.252.0 {
    option subnet-mask         255.255.252.0;
    range                      10.1.3.200 10.1.3.254;
}
include "/etc/dhcp/dhcpd.d/vlan-13.conf";
subnet 10.1.6.0 netmask 255.255.254.0 {
    option subnet-mask         255.255.254.0;
    range                      10.1.6.200 10.1.6.254;
}
include "/etc/dhcp/dhcpd.d/vlan-99.conf";
```

  - /etc/dhcp/dhcpd.d/ (folder)
    - Folder to contain IPU-M2000 network configuration files.
    - root:dhcpd 0770
  - /etc/dhcp/dhcpd.d/vlan-11.conf (file)

- Specific file related to IPU-M2000 100Gb RNIC interfaces. Will contain a relation between desired IPs and IPU-M2000 100GbE RNIC (Mellanox) interfaces MAC addresses.
- root:dhcpd 0660

```
host lr1_ipum1rnic { hardware ethernet 1c:34:da:67:17:91; fixed-address 10.1.5.1; }
host lr1_ipum2rnic { hardware ethernet 1c:34:da:67:1d:19; fixed-address 10.1.5.2; }
host lr1_ipum3rnic { hardware ethernet 1c:34:da:67:14:e9; fixed-address 10.1.5.3; }
host lr1_ipum4rnic { hardware ethernet 1c:34:da:67:23:91; fixed-address 10.1.5.4; }
```

- /etc/dhcp/dhcpd.d/vlan-13.conf (file)
  - Specific file related to IPU-M2000 1Gb BASE-T interfaces. Will contain a relation between desired IPs and IPU-M2000 1GbE BASE-T interfaces MAC addresses.
  - root:dhcpd 0660

```
host lr1_ipum1bmc { hardware ethernet 70:69:79:20:03:A8; fixed-address 10.1.1.1; }
host lr1_ipum1gw  { hardware ethernet 70:69:79:20:03:A9; fixed-address 10.1.2.1; }
host lr1_ipum2bmc { hardware ethernet 70:69:79:20:01:48; fixed-address 10.1.1.2; }
host lr1_ipum2gw  { hardware ethernet 70:69:79:20:01:49; fixed-address 10.1.2.2; }
host lr1_ipum3bmc { hardware ethernet 70:69:79:20:03:80; fixed-address 10.1.1.3; }
host lr1_ipum3gw  { hardware ethernet 70:69:79:20:03:81; fixed-address 10.1.2.3; }
host lr1_ipum4bmc { hardware ethernet 70:69:79:20:03:E0; fixed-address 10.1.1.4; }
host lr1_ipum4gw  { hardware ethernet 70:69:79:20:03:E1; fixed-address 10.1.2.4; }
```

The dhcp service is started using:

```
sudo systemctl enable dhcp
sudo systecmtl start dhcp
```

4.2.8 Services: NTP (Network Time Protocol)

NTP service is recommended to provide network time configuration to IPU-M2000s. It can be installed from the Ubuntu or CentOS public repositories.

File structure:

- /etc/ntp.conf (file)
  - NTP tool configuration file.
  - root:root 0444

```
disable monitor
driftfile /var/lib/ntp/drift
fudge 127.127.1.0 stratum 10
includefile /etc/ntp/crypto/pw
keys /etc/ntp/keys
restrict ::1
restrict 127.0.0.1
restrict default nomodify notrap nopeer noquery
server 127.127.1.0
server 0.pool.ntp.org iburst
server 1.pool.ntp.org iburst
server 2.pool.ntp.org iburst
```

The ntp service is started using:

```
sudo systemctl enable ntpd
sudo systecmtl start ntpd
```

4.2.9 Services: syslog

This is a software utility for forwarding log messages in an IP network.

File structure:

- /etc/rsyslog.d (folder)
  - Rsyslog tool configuration folder.
  - root:root 0755
- /etc/rsyslog.conf (file)
  - Rsyslog tool configuration file.
  - root:root 0444

```
module(load="imuxsock")
module(load="imudp")
input(type="imudp" port="514")
module(load="imtcp")
input(type="imtcp" port="514")
module(load="imklog" permitnonkernelfacility="on")
$ActionFileDefaultTemplate RSYSLOG_TraditionalFileFormat
$RepeatedMsgReduction on
$FileOwner syslog
$FileGroup adm
$FileCreateMode 0640
$DirCreateMode 0755
$Umask 0022
$PrivDropToUser syslog
$PrivDropToGroup syslog
$WorkDirectory /var/spool/rsyslog
$IncludeConfig /etc/rsyslog.d/*.conf
```

- /etc/rsyslog.d/99_ipum.conf (file)
  - Rsyslog rules configuration file.
  - root:root 0444

```
$template precise,"%fromhost-
ip%,%HOSTNAME%,%syslogpriority%,%syslogfacility%,%timegenerated::fu
lltime%,%syslogtag%,%msg%\n"
      :HOSTNAME, contains, "ipum" /var/log/ipulogs;precise
      & ~
```

- /etc/rsyslog.d/99_dhcpd.conf (file)
  - Rsyslog rules configuration file.
  - root:root 0444

```
local7.*         /var/log/dhcpd.log
```

5 IPU-POD$_{64}$ software installation and configuration

5.1 Management server

One server in the IPU-POD$_{64}$ is nominated as the management server – by default this is server 1. The following Graphcore software packages need to be installed on the management server:

a) **V-IPU software** – contains management and control software for IPU resource control, Built In Self Test (BIST) and monitoring of the IPU-M2000s and IPUs. There is a V-IPU Admin Guide and a V-IPU User Guide available.

b) **IPU-M2000 software** - contains the latest IPU-M2000 resident software. It also includes the server resident `rack_tool` which is required for doing operations from the management server related to all IPU-M2000s.

Note	For large deployments, the management functions can be provided by a separate high-availability server cluster outside the IPU-POD$_{64}$. Please contact Graphcore for more details.

5.2 V-IPU software installation and configurations

Please read carefully the release notes for the V-IPU software release before any software upgrade takes place. The release notes are available from the Graphcore download portal https://downloads.graphcore.ai as a separate downloadable entity from the same Web page where the V-IPU software release itself is found.

The release notes give the following details of the release:

a) Software version numbers.

b) Compatibility changes that may need to be understood before upgrading the V-IPU software.

c) Details to any special upgrade handling for this specific release.

d) An overview of fixed problems

e) An overview of remaining known issues with proposed workarounds, if any.

The software release bundle contains a set of server resident software components.

V-IPU should be installed from a user with root privileges (i.e. with sudo ./install.sh). Note that this install script installs the V-IPU controller to run as a service in the context of the root user. You may need to change to the itadmin user to do this since ipuuser does not have permission for root access.

5.3 IPU-M2000 software installation and configuration

Please read carefully the release notes for the IPU-M2000 software release before any software upgrade takes place. The release notes are available from the Graphcore download portal https://downloads.graphcore.ai as a separate downloadable entity from the same web page where the software release itself is found.

The release notes give the following details of the release:

a) Software sub-component version numbers

b) Compatibility statements to Poplar SDK versions

c) Compatibility changes from earlier releases that may need to be understood before upgrading the IPU-M2000s

d) Details to any special upgrade handling for this specific release

e) An overview of fixed problems

f) An overview of remaining known issues with proposed workarounds, if any.

The IPU-M2000 software release bundle contains a set of upgradable software and FPGA sub-components that is targeted to be executed on the IPU-M2000. The release also contains the tool `rack_tool` which is used for the software upgrade and other rack related tasks targeting the IPU-M2000s.

The `rack_tool` upgrade command performs the software upgrade. See details later for how to run this. The IPU-M2000 GW and ICU supports booting from one of two persistent software images, the active image or the standby image. When updating the software, the system will always update the standby image that is not running.

If an upgrade operation fails for one of the components, the user should not try to force booting from the now inconsistently upgraded standby image(s) for the various CPU systems inside the IPU-M2000.

The software install process can currently NOT be run at the same time as running ML jobs since the install process reboots the IPU-M2000 once complete.

When the update of the standby image is completed successfully, the IPU-POD$_{64}$ is immediately instructed to switch to the updated standby image, making it the active one. The previously running image now becomes the standby image. This is done via a service affecting reboot of each of the upgraded CPU/FPGA systems.

If for some reason the operator wants to revert to the previous software version, the standby image can be upgraded to the previous version in the same way as described above.

Note	Graphcore has only qualified the IPU-M2000 software release with exactly the documented set of software sub-component versions and any other version combinations of software components are not guaranteed.

Note that there is a BMC user guide available at https://docs.graphcore.ai/projects/bmc-user-guide/.

5.3.1 Download IPU-M2000 software update bundle

The management server needs to be loaded with the correct IPU-M2000 software update bundle before the software update of the IPU-M2000s can be performed. To perform the download, please follow these steps:

1. Login as “ipuuser”
2. Go to the Graphcore download portal https://downloads.graphcore.ai and download the latest release into the `$HOME /IPU-M_releases` directory
3. Follow the install instructions in the release notes, and perform:
   `tar xvfz <tar-ball.tgz>`

The unrolling of the IPU-M2000 software onto the management servers file system will automatically create a file tree with a leading directory containing the release version number. This allows several releases to be kept on the management server, in case there is a need to revert to a previous release onto the IPU-M2000s. If this is considered not needed, the older releases (both the unrolled files and the downloaded tar file) can be removed from the management server.

5.3.2 Software update of all IPU-M2000s

Having unrolled the software release onto the server file system, please follow these steps:

1. Change directory to:
   `cd $HOME/IPU-M_releases/IPU-M_SW_<release version>`
2. Run the commands:
   ```
   virtualenv -p python3 venv
   source venv/bin/activate
   pip3 install -r requirements.txt
   ./rack_tool.py upgrade
   ```

`rack_tool` will read a default config file to learn how to access the IPU-M2000s.
The default location of this file is: **`$HOME/ipuuser/.rack_tool/rack_config.json`**

This file can be edited by a site administrator that integrates the IPU-POD$_{64}$ into the site-specific network in cases where the default IPU-POD$_{64}$ IP address plan collides with the site-specific network.

The upgrade process will take several minutes and all the IPU-M2000s will be upgraded in parallel to make this time as short as possible.

The upgrade process at the end will perform a few reboots in order to activate the new software.

`rack_tool` finally verifies that the upgrade completes with all sub-components being upgraded to the same version. Please verify that this version corresponds to what is defined in the release notes to check that the upgrade procedures have been followed correctly.

5.3.3 IPU-M2000 GW's root file system config files

The IPU-POD$_{64}$ and IPU-M2000 support a simple concept where a set of IPU-M2000 Gateway OS config files residing on the management server will be copied to the IPU-M2000 Gateway after an update. This feature is optional.

With multiple IPU-POD$_{64}$ racks installed, this overwriting (overlay) concept will support a site-specific config control on top of the standard Gateway distribution that always brings default config files. The current implementation requires all IPU-M2000s to be identically configured.

Any config files on the Debian OS in the Gateway can be overwritten by these server-side maintained overlay files.

These files reside under `ipuuser` at **`$HOME/.rack_tool/root-overlay/.`**

Note	**Make sure the IP addresses in these files are correct and point to the management server's IP address**

The IPU-M2000 Gateway upgrade itself is destructive in the sense that it will overwrite all executables and corresponding file systems for the standby image. The content of the overlay config files is maintained and stored persistently on the management server. A single central management server can function as a central repository for a site wide setup of all IPU-M2000s.

Please refer to the `rack_tool` documentation that is include in the IPU-M2000 software bundle that needs to be unrolled on the management server for the `rack_tool` and corresponding "overlay" files to be supported.

5.4 `rack_tool`

`rack_tool` is a utility that is supplied with the IPU-M2000 release bundle when installed onto a management server. It is always found under the account that is performing IPU resource management (the default is the `ipuuser` account), as **`~/IPU-M_releases/IPU-M_<release-version>/rack_tool.py`.**

A suggested naming scheme for IPU-M2000s is: **lr1_ipum<n> (logical rack 1, IPU-M2000 #**n).

`rack_tool` is used for the following:

a) Installing on the IPU-POD$_{64}$ system for single or all IPU-M2000s

b) Querying the version of all IPU-M2000s

c) Connectivity test performed on all IPU-M2000 RDMA data-plane ports, GW and BMC management-plane ports across all IPU-M2000s that are listed in the `rack_tool`'s default config file

d) Restarting IPU-M2000's GW and BMC in different ways (power cycling or OS reboot)

e) Control power on/off for the GW part of the IPU-M2000

f) Running commands on several IPU-M2000s for troubleshooting

g) Updating the "root overlay" files onto all IPU-M2000s if NTP or syslog server has changed

h) Running hardware and connectivity tests (see section 6 for the built-in test capabilities)

i) Future: rewriting OS config files when an IPU-POD$_{64}$ is initially given a "logical rack number" in a row of IPU-POD$_{64}$ racks for a larger system. These rewrites of config files are mostly related to addressing and DNS naming of IPU-POD$_{64}$ interfaces of various kinds.

The supported options will evolve over time so please refer to the official help menu (`rack_tool --help`) or accompanying man pages for the installed `rack_tool` version on your system. The current supported options are listed below:

```
rack_tool.py upgrade [--help] [--golden] [--gw-root-overlay path to overlay]
rack_tool.py bist [--help]
rack_tool.py vipu-test [--help] [--vipu-path path to vipu binaries]
rack_tool.py status [--help] [--no-color]
rack_tool.py hostname [--help]
rack_tool.py install-key [--help]
rack_tool.py update-root-overlay [--help] [--overlay overlay directory]
rack_tool.py run-command [--help] -c command -d device
rack_tool.py bmc-factory-reset [--help]
rack_tool.py power-off [--help] [--hard]
rack_tool.py power-on [--help]
rack_tool.py power-cycle [--help]
rack_tool.py logging-server [--help] -a address -p port -d device
```

6 IPU-POD$_{64}$ manual installation tests

There is a division of responsibility between BMC management and V-IPU management when it comes to which parts of the system that they test.

BMC has support for chassis management, which means that it can verify correct hardware behaviour on most functional blocks within the chassis.

V-IPU management has support for connectivity tests and focuses on verifying correct cables, cabling for IPU-Links and GW-Links, as well as for the cabled IPU-Link network.

In combination, these two areas of built-in self-tests (BISTs) will cover most of the needs for system installation verification.

6.1 Running system BISTs

The `rack_tool` utility is included as part of the IPU-M2000 software release bundle. See section 0 which covers usage of `rack_tool` before running these tests. This is especially important if the tests are to be run on a system that has active users.

`$ ./rack_tool.py bist` - performs BMC chassis hardware testing

`$ ./rack_tool.py vipu-test` - performs V-IPU connectivity related tests

6.2 Troubleshooting

This section contains useful information about what to do if you encounter problems while installing and testing the rack. If you can't find the answer to your query here and are still experiencing problems, then please contact your Graphcore representative or use the resources on the Graphcore support portal: https://www.graphcore.ai/support.

6.2.1 BMC BISTs

`$ ./rack_tool.py bist`

This test will generate a very low level hardware verification report/log that will need to be analyzed by Graphcore support in case any errors are reported. The logs are located at "`./logs`" relative to current directory from which the command is executed.

The command "Done BIST on …" if the test is successful.

The command "Failed BIST on …" if the test fails.

The command will point to the log name generated in both cases.

6.2.2 V-IPU built in self tests

`$ ./rack_tool.py vipu-test`

The following section is based on excerpts from the V-IPU Admin Guide which should be consulted for a detailed and updated overview of BISTs. This guide is available here. The V-IPU User Guide is also useful and can be found here. The collection of V-IPU connectivity tests can be invoked by the **`./rack_tool.py vipu-test`** command or by directly using V-IPU CLI commands as described below.

The V-IPU Controller implements a cluster testing suite that runs a series of tests to verify installation correctness. A V-IPU cluster test can be executed against a cluster entity before

any partitions are created. It is strongly recommended to run all the test types provided by the cluster testing suite before deploying any applications in a cluster.

Assume we have created a cluster named “cluster1” formed by four IPU-M2000s (IPU-POD$_{64}$) using the command:

`vipu-admin create cluster cl0 --agents ipum1, ipum2, ipum3, ipum4 --mesh.`

`./rack_tool.py vipu-test` will create V-IPU machine VIRM agents for each IPU-M2000 and automatically create this cluster by applying this command.

The simplest way to run a complete cluster test for this cluster is to run `./rack_tool.py vipu-test`. The test performs the V-IPU self-tests shown below.

```
vipu-admin test cluster cluster1
```

```
Showing test results for cluster cl0
  Test Type     | Duration | Passed | Summary
-------------------------------------------------------------------------
  Version       | 0.00s    | 4/4    | All component versions are consistent
  Cabling       | 8.76s    | 4/4    | All cables connected as expected
  Sync-Link     | 0.35s    | 8/8    | Sync Link test passed
  Link-Training | 20.16s   | 76/76  | All Links Passed
  Traffic       | 42.00s   | 1/1    | Traffic test passed
  GW-Link       | 0.00s    | 0/0    | GW Link test skipped
```

The output above shows a successful test with no errors reported.

As the test results show, five test types were executed on “cluster1”. The results for each test type are printed one per line in the output. Each test type tested zero or more elements of the cluster as can be seen from the “Passed” column. Each test type is explained in detail in the rest of this section.

Note that the vipu-test command blocks the CLI until the cluster test is completed, and may take several minutes to complete. To avoid blocking the CLI for prolonged periods of time, cluster tests can be executed asynchronously with the --start, --status and --stop options.

Depending on the how the cluster is created, some of the link tests will be omitted.

In the above example the V-IPU GW link test is skipped since the GW link is not used in single IPU-POD$_{64}$ installation testing. Only when interconnecting several IPU-PODs together it makes sense to also tests the GW links.

Errors discovered during testing can be like the ones shown below. The error text being shown is if possible, indicating which ports are relevant for the problem detected. The port numbers used are aligned with the various connector´s numbering scheme described earlier in this document.

When a cluster test is running, some restrictions are imposed on the actions an administrator can perform to the system:

- Partition creation in a cluster where a test is in progress is forbidden.
- Removal of a cluster where a test is in progress is forbidden.

- Only one cluster test can be running at any given time on a V-IPU server, even if the V-IPU server controls more than one cluster.
- There is no persistence to the cluster test results. Only the results of the last test can be retrieved with the --status command, as long as the V-IPU server has not been restarted.

IPU-Link cabling test:

In order to verify that external IPU-Link cables are connected and properly inserted as expected in a cluster, the cabling test can be utilized. The cabling test will read the serial ID of the OSFP cables from each end of the links and verify that the cable connects the expected ports together.

Cabling tests are invoked by passing the `--cabling` flag to the test cluster command.

If the test fails, details about which connections that failed are displayed. This will give the user a hint to which cables to physically inspect and correct. Very often, a loose cable is the root cause of problems. Below is an example of a test run when the 4 OSFP cables between ipum1 and ipum2 in the cluster are not connected.

```
$vipu-admin test cluster cluster1 --cabling
```

```
Showing test results for cluster cluster1
 Test Type   | Duration | Passed | Summary
-------------------------------------------------------------------------------
 Cabling     | 21.77s   | 8/12   | ipum1 (IPU-Cluster Port 5) x--> ipum2 (IPU-
Cluster port 11) (cable not connected)
             |          |        | ipum1 (IPU-Cluster Port 6) x--> ipum2 (IPU-
Cluster port 12) (cable not connected)
             |          |        | ipum1 (IPU-Cluster Port 7) x--> ipum2 (IPU-
Cluster port 13) (cable not connected)
             |          |        | ipum1 (IPU-Cluster Port 8) x--> ipum2 (IPU-
Cluster port 14) (cable not connected)
-------------------------------------------------------------------------------
```

This is an indication of either faulty cabling or an incorrect cluster definition that doesn´t reflect the intended cabling.

Sync-Link test:

The Sync-Link test verifies the external Sync-Link cabling between IPU-M2000s. You can run a Sync-Link test by passing the `--sync` option to the test cluster command.

A failing Sync-Link test reports the cables which failed to satisfy the cluster topology that is being tested by pointing to the IPU-M2000s and Sync-Link port numbers of the failing Sync-Link. In the example command below, two Sync-Link cables between “ipum1” and “ipum2” fail:

- the link between “ipum1” Sync-Link port 6 and “ipum2” Sync-Link port 2
- the link between “ipum1” Sync-Link port 7 and “ipum2” Sync-Link port 3

This is an indication of either faulty cabling or an incorrect cluster definition that doesn´t reflect the intended cabling.

```
$vipu-admin test cluster cluster1 –sync
```

```
Showing test results for cluster cluster1
  Test Type   | Duration | Passed | Summary
-------------------------------------------------------
  Sync-Link   | 0.90s    | x/y    | Failed Sync Links:
              |          |        | ipum1:6 <--> 2:ipum2
              |          |        | ipum1:7 <--> 3:ipum2
-------------------------------------------------------
test (cluster): failed: Some tests failed.
```

IPU-Link training test:

IPU-Link training test verifies IPU-Link readiness for IPU-Links between and within IPU-M2000s (OSFP cables). An IPU-Link test can be invoked with the `--ipulink` option in the test cluster command. A failing test will indicate which IPU-Links are failing by pointing to the agent and cluster port enumeration of the failing IPU-Link. In the following example, we test a cluster where the IPU-Links have been disconnected between the first and second IPU-M2000 units.

```
$vipu-admin test cluster cluster1 –ipulink
```

```
Showing test results for cluster cluster1
  Test Type   | Duration | Passed  | Summary
-------------------------------------------------------------------------------------
  IPU-Link    | 34.57s   | x/y | Failed Links
              |          |     | ipum1:4 [pending g1x1] <--> ipum2:8 [pending g1x1]
              |          |     | ipum1:3 [pending g1x1] <--> ipum2:7 [pending g1x1]
              |          |     | ipum1:2 [pending g1x1] <--> ipum2:6 [pending g1x1]
              |          |     | ipum1:1 [pending g1x1] <--> ipum2:5 [pending g1x1]
-------------------------------------------------------------------------------------
test (cluster): failed: Some tests failed.
```

IPU-Link traffic test:

The traffic test acts as a smoke test for all IPU-Links of a cluster before deploying applications.

The traffic test can be invoked with the --traffic option. Note that for a traffic test to pass, a prerequisite is that the IPU-Link and IPU-Link training tests have passed.

```
$vipu-admin test cluster cluster1 --traffic
```

```
  Test      | Duration | Passed | Summary
-------------------------------------------------------------------------------
  Traffic   | 92.23s   | 3/4    | Traffic test failed
            |          |        | Errors encountered in traffic test 1
            |          |        | corrected link errors: 460
            |          |        | - error counter IPU-Link 1 in ipum1, IPU '1' is 250
            |          |        | - error counter IPU-Link 1 in ipum4, IPU '1' is 210
-------------------------------------------------------------------------------
test cluster (cluster1): failed: Some tests failed.
```

This example shows a situation where the IPU-link traffic test has failed due to too many correctable errors being detected. Should this occur please try reseating the IPU-Link cables associated with the referenced IPU-M2000 units. If that does not resolve the issue, please contact Graphcore support.

7 Automatic IPU-POD$_{64}$ configuration

This section describes using a golden laptop, USB stick and a scanner to set the IPU-POD$_{64}$ up automatically instead of running through the steps manually as described in previous sections. This method might be appropriate where many IPU-POD$_{64}$ need to be configured and tested as stand-alone racks, rather than installed and connected to a scripted central management service.

7.1 Devices and preparation

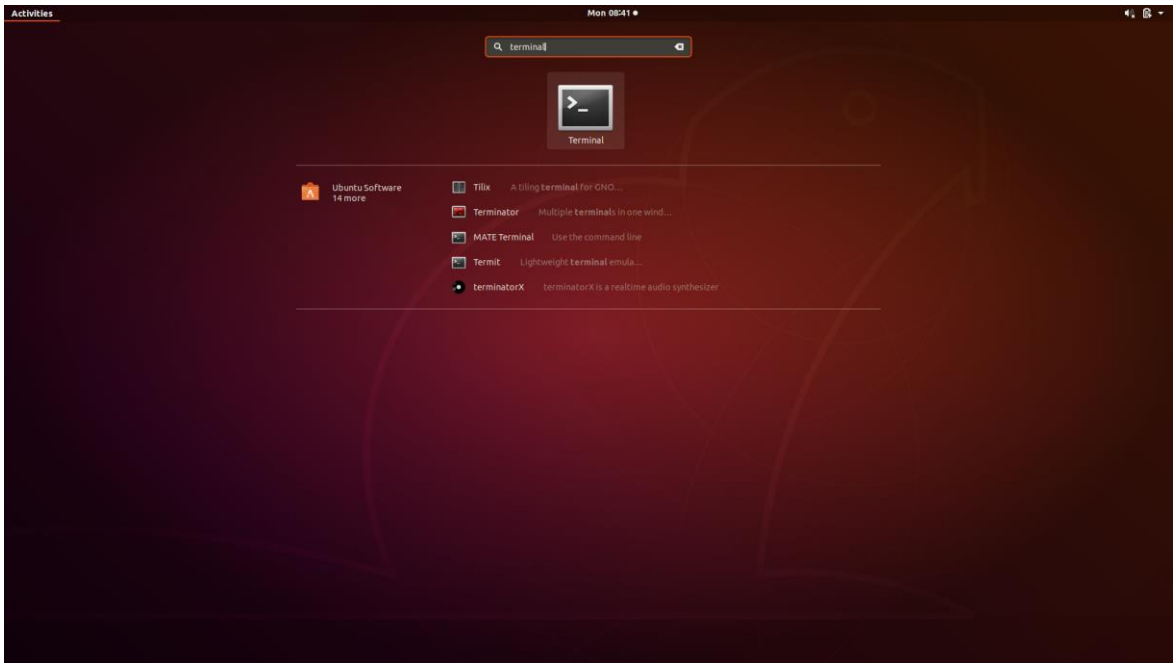
You will need:

- 1 Lenovo ThinkPad laptop (pre-configured by Graphcore)
- 1 Lenovo power supply
- 1 Sandisk USB memory stick (pre-configured by Graphcore)
- 1 barcode scanner
- 1 document containing configuration QR codes (available from Graphcore support)

7.2 Scanning and test

Follow the instructions below:

1. Make sure the laptop used for scanning has the USB stick connected
2. Power on the laptop and login with the user "`gctest`", the password is available from Graphcore support support.graphcore.ai
3. Press the Windows button, type "`terminal`" and press enter to open a terminal:



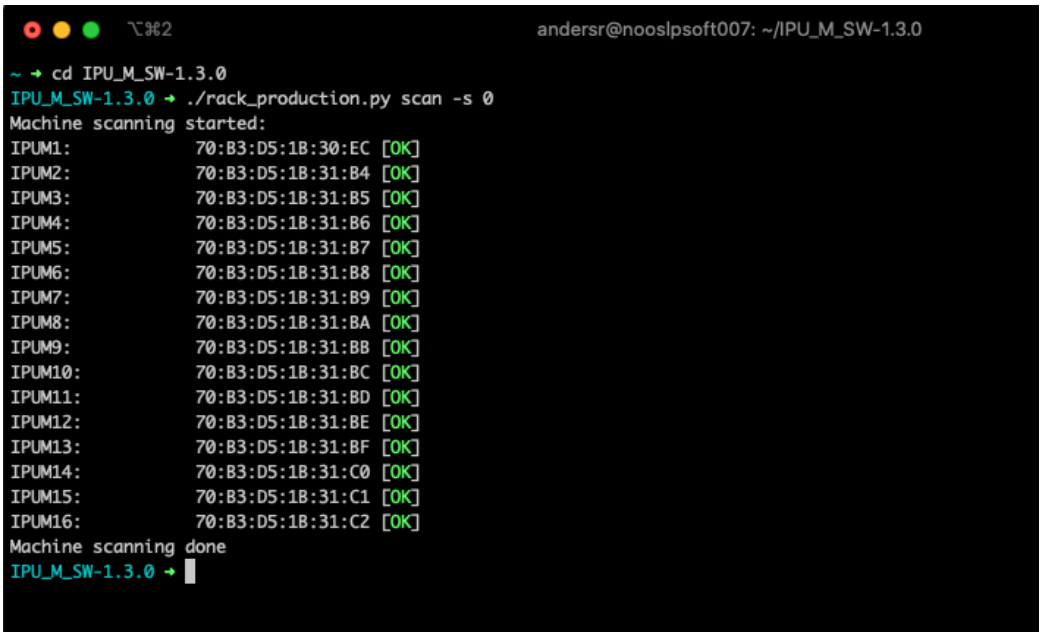
4. Change directory to the IPU-M2000 release to use (`$HOME/IPU-M_releases/IPU-M_SW_<release version>`).

5. Run the command "`./rack_production.py scan -s 0`"

6. Scan each of the 16 IPU-M2000s starting from the bottom-most IPU-M2000 (#1) up to IPU-M2000 #16. If you encounter problems while scanning the IPU-M2000s then check the troubleshooting section 7.2.2.

Note	**It's important to verify that each scan reports [OK] before moving on to scan the next IPU-M2000**

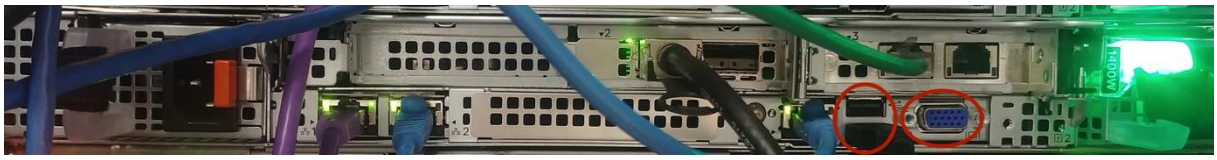
7. After all the IPU-M2000s have been scanned successfully, `rack_production.py` will exit and all the files needed are stored on the USB stick.



8. Unplug the USB Stick from the laptop and plug it into the management server (server 1, the lowest server).

9. Connect the scanner and a VGA monitor to the back side of the management server..

Note	**If debug is required, a keyboard can be connected to one of the USB connectors to manually run commands on the server.**



10. Reboot the machine by scanning the reboot QR code given in the **IPU-POD$_{64}$ QR codes and passwords for build and test** document. This document is available from Graphcore support.
11. The monitor should look like the picture below after the server has finished the reboot:

```
## Graphcore.ai - IPUM NODE ########################
Updated:          Thu 19 Dec 16:24:33 GMT 2019
OS:               Ubuntu 18.04.3 LTS
HOST:             ipu-pod-mgmt
IP:               192.168.1.203 10.1.3.101 10.1.6.1 10.1.5.101 10.240.0.1
Puppet run:       Thu 19 Dec 16:24:33 GMT 2019
CERT:             ipu-pod-mgmt
RAID:             [Pass]
CM:               [Pass]
```

12. After reboot, scan the QR codes given in the **IPU-POD$_{64}$ QR codes and passwords for build and test** document (available from Graphcore support) in order from 1 to 15. QR codes 1 to 14 are steps preparing for upgrade and self-test. QR code 15 is a full rack test and will take about one and a half hours.

A description of what is happening in each of the steps is given at the end of this section – see section 7.2.1.

Note	Make sure each step has completed before scanning the next QR label.

If nothing fails, then something similar to the output below will be seen:

```
#################################################################
#                 Authorized access only!                       #
# Disconnect IMMEDIATELY if you are not an authorized user! #
#         All actions are monitored and recorded                #
#################################################################

ipuuser@ipu-pod-mgmt:~$ cp -r /IPU_M_SW-v1.0.0-rc.48+1683409 /home/ipuuser/
ipuuser@ipu-pod-mgmt:~$ cp -r /prod_python3_venv /home/ipuuser/
ipuuser@ipu-pod-mgmt:~$ cp /rack_production.py /home/ipuuser/IPU_M_SW-v1.0.0-rc.48+1683409/
ipuuser@ipu-pod-mgmt:~$ cp /rack_config.json /home/ipuuser/IPU_M_SW-v1.0.0-rc.48+1683409/
ipuuser@ipu-pod-mgmt:~$ cd /home/ipuuser; mkdir -p /home/ipuuser/.local/bin; source .profile
ipuuser@ipu-pod-mgmt:~$ source prod_python3_venv/bin/activate
(prod_python3_venv) ipuuser@ipu-pod-mgmt:~$ cd /home/ipuuser/IPU_M_SW-v1.0.0-rc.48+1683409; ./rack_production.py
- vipu-server will be configured to be run as a service in this host
'vipu-server' -> '/home/ipuuser/.local/bin/vipu-server'
'vipu-cli' -> '/home/ipuuser/.local/bin/vipu-cli'
- Configuring systemd for user ipuuser
- Initialising vipu-server storage
Initialised storage: vipu-server.json
- Configuring vipu-server systemd service
- Enabling vipu-server systemd service to start at system boot
Created symlink /home/ipuuser/.config/systemd/user/default.target.wants/vipu-server.service -> /home/ipuuser/.config/systemd/use
r/vipu-server.service.
- Starting vipu-server service
- Use the command 'systemctl --user status vipu-server.service' to check the status of the service
Checking if all machines are up
Getting Mellanox MACs
Time spent: 1s
Restarting dhcp server
Starting upgrade of ipum1
Update complete for ipum1. Logfile at: /home/ipuuser/IPU_M_SW-v1.0.0-rc.48+1683409/logs/ipum1_upgrade.log
All machines were successfully upgraded
Time spent: 11m24s
Started BIST on ipum1
Done BIST on ipum1. Logfile at: /home/ipuuser/IPU_M_SW-v1.0.0-rc.48+1683409/logs/ipum1_bist.log
Time spent: 10m18s
create agent (ipum1): success.
create cluster (ipums): success.
 Test          | Duration | Passed | Summary
-----------------------------------------------------------
 Sync-Link     | 0.00s    | 0/0    | Sync Link test skipped
 Link-Training | 8.97s    | 16/16  | All Links Passed
 Traffic       | 11.16s   | 1/1    | Traffic test passed
-----------------------------------------------------------
Time spent: 20s
(prod_python3_venv) ipuuser@ipu-pod-mgmt:~/IPU_M_SW-v1.0.0-rc.48+1683409$
```

Rack installation and test is now complete.

7.2.1 Description of each QR code step

QR #1: Login to the server as itadmin

QR #2: Enter password. Passwords are available from Graphcore support

QR #3: Reboot:

```
sudo reboot
```

QR #4: Login to the server as itadmin again

QR #5: Enter password. Passwords are available from Graphcore support

QR #6: Copy the "IPU_M_release" to the "itadmin" home directory:

```
cp -r /IPU_M_SW* /home/itadmin/IPU_M_SW_prod
```

QR #7: Install requirements for itadmin:

```
pip3 install -r
/home/itadmin/IPU_M_SW_prod/maintenance_tools/requirements.txt
```

QR #8: Go to the home directory and prepare for vipu install:

```
cd /home/itadmin; mkdir -p /home/itadmin/.local/bin; source .profile
```

QR #9: Install vipu:

```
cd /home/itadmin/IPU_M_SW_prod/maintenance_tools/;
./rack_production.py vipu-install
```

QR #10: Login to the server as ipuuser:

```
ssh -o "StrictHostKeyChecking no" ipuuser@localhost
```

QR #11: Enter password for ipuuser. Passwords are available from Graphcore support

QR #12: Copy IPU_M release to ipuuser home directory:

```
cp -r /IPU_M_SW* /home/ipuuser/IPU_M_SW_prod
```

QR #13: Install requirements for ipuuser:

```
pip3 install -r
/home/ipuuser/IPU_M_SW_prod/maintenance_tools/requirements.txt
```

QR #14: Copy rack config to rack_tool config directory:

```
mkdir -p /home/ipuuser/.rack_tool; cp /rack_config.json
/home/ipuuser/.rack_tool/
```

QR #15: Go to install directory and run `rack_production.py`

```
cd /home/ipuuser/IPU_M_SW_prod/maintenance_tools/;
./rack_production.py setup -s 0
```

7.2.2 Scan troubleshooting

If you encounter problems while scanning the IPU-M2000s then check the following:

1) If the scan is reporting "[Fail, read error]":

   Something went wrong during analysis of the scan input. Try to scan the same QR code again.

2) If the scan is reporting "[Fail, duplicate]":

   The same QR code has been scanned twice.

3) If you start scanning the wrong QR code:

   Abort the scanning with "Ctrl+c" and restart the scanning process from the beginning.

If you are still experiencing problems please contact Graphcore support or your Graphcore representative.

7.2.3 Installation troubleshooting

- If the upgrade fails, try to restart the `rack_production.py` tool by scanning QR code 12 again. If it continues to fail, a manual inspection of the rack will be required.
- If BIST fails, the failing IPU-M2000(s) have to be replaced and the whole procedure from step 1 has to be restarted.
- If Sync-Link, Link-Training or Traffic test fail, some cables might be loose. Check for loose cables and restart the `rack_production.py` tool by scanning QR code 15 again.
- If VIPU-test fails with the following message:

```
test (cluster): failed: rpc error: code = DeadlineExceeded desc =
context deadline exceeded.
```

  Then do the following:

  1. To fix potential sync errors, run: `vipu-admin test cluster ipums --sync`

     It is also possible to run this command by scanning a QR code. This QR code is given in the **IPU-POD$_{64}$ QR codes and passwords for build and test** document (available from Graphcore support)

  2. To fix potential link errors, run: `vipu-admin test cluster ipums --training`

     It is also possible to run this command by scanning a QR code. This QR code is given in the **IPU-POD$_{64}$ QR codes and passwords for build and test** document (available from Graphcore support)

  3. Restart the `rack_production.py` tool by scanning QR code 15 again

- For all other failures, manual inspection is required.

7.2.4 Example output for Sync-Link or Traffic test failure

```
  Test           | Duration | Passed | Summary
-------------------------------------------------------------
  Sync-Link      | 0.05s    | 4/8    | Failed Sync Links:
                 |          |        | ag5:S0 <-> ag4:N0
                 |          |        | ag5:S1 <-> ag4:N1
                 |          |        | ag4:S0 <-> ag3:N0
                 |          |        | ag4:S1 <-> ag3:N1
-------------------------------------------------------------
  test cluster (cl1): failed: Some tests failed.

  Test           | Duration | Passed | Summary
-------------------------------------------------------------
  Sync-Link      | 0.19s    | 6/6    | Sync Link test passed
  Link-Training | 9.93s    | 76/76  | All Links Passed
  Traffic        | 12.08s   | 0/1    | Traffic test failed
                 |          |        | Errors encountered in traffic test 1
                 |          |        | corrected link errors: 55066
                 |          |        | - error counter on IPU Cluster
Link '1C' in Agent ag1, IPU '0' is 55066
-------------------------------------------------------------
```

7.2.5 Other useful commands

The QR codes for these commands are given in the **IPU-POD64 QR codes and passwords for build and test** document (available from Graphcore support).

Run only upgrade:

```
./rack_tool.py
upgrade
```

Run only BIST:

```
./rack_tool.py
bist
```

Run only vipu-test:

```
./rack_tool.py
vipu-test
```

8 Document revisions

8.1 Revision history

This document's revision history is as follows:

Version	Date	Notes
1.0	4th of December 2020	First release

9 Legal notices

Warranties & licences

This document is confidential and is provided subject to one or more confidentiality obligations between you and Graphcore, including a Non-Disclosure Agreement. The information disclosed to you hereunder (the “Materials”) is provided solely for the selection and use of Graphcore products. To this end, you shall only use the Materials internally and in connection with Graphcore products, and you shall at all times comply with any and all documentation and software licensing terms provided by Graphcore. To the maximum extent permitted by applicable law: (1) Materials are made available "AS IS" and with all faults, Graphcore hereby DISCLAIMS ALL WARRANTIES AND CONDITIONS, EXPRESS, IMPLIED, OR STATUTORY, INCLUDING BUT NOT LIMITED TO WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, OR FITNESS FOR ANY PARTICULAR PURPOSE; and (2) Graphcore shall not be liable (whether in contract or tort, including negligence, or under any other theory of liability) for any loss or damage of any kind or nature related to, arising under, or in connection with, the Materials (including your use of the Materials), including for any direct, indirect, special, incidental, or consequential loss or damage (including loss of data, profits, goodwill, or any type of loss or damage suffered as a result of any action brought by a third party) even if such damage or loss was reasonably foreseeable or Graphcore had been advised of the possibility of the same. Graphcore assumes no obligation to correct any errors contained in the Materials or to notify you of updates to the Materials or to product specifications. You may not reproduce, modify, distribute, or publicly display the Materials without prior written consent. Certain products are subject to the terms and conditions of Graphcore’s limited warranty. Graphcore products are not designed or intended to be fail-safe or for use in any application requiring fail-safe performance; you assume sole risk and liability for use of Graphcore products in such critical applications.

The European Directive 2012/19/EU on Waste Electrical and Electronic Equipment (WEEE) states that these appliances should not be disposed of as part of the routine solid urban waste cycle, but collected separately in order to optimise the recovery and recycling flow of the materials they contain, while also preventing potential damage to human health and the environment arising from the presence of potentially hazardous substances.

The crossed-out bin symbol is printed on all products as a reminder.

Waste may be taken to special collection site or can be delivered free of charge to the dealer when purchasing a new equivalent or without obligation to make a new purchase for equipment smaller than 25cm.

For more information on proper disposal of these devices, kindly refer to the public utility service.

Trademarks & copyright

Graphcore® and Poplar® are Registered Trademarks of Graphcore Ltd.

Colossus™, IPU-Core™, In-Processor-Memory™, Exchange Memory™, Streaming Memory™, IPU-Tile™, IPU-Exchange™, IPU-Machine™, IPU-M2000™, IPU-POD™, IPU-Link™, Virtual-IPU™, AI-Float™, IPU-Fabric™, PopART™, PopLibs™, PopTorch™ and PopVision™ are Trademarks of Graphcore Ltd.

All other trademarks are the property of their respective owners.

Design and specifications subject to change without prior notice.