IBM System Blue Gene Solution: Blue Gene/Q Hardware Overview and Installation Planning

- Understand Blue Gene/Q hardware components
- Learn about physical installation requirements
- Prepare the Blue Gene/Q site

James Milano
Pamela Lembke

ibm.com/redbooks
Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

Second Edition (May 2013)

This edition applies to Version 1, Release 1, Modification 0 of IBM Blue Gene/Q (product number 5733-BGQ).

© Copyright International Business Machines Corporation 2013. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
## Contents

**Notices**  ......................................................... vii  
**Trademarks**  ................................................. viii  

**Preface**  ........................................................ ix  
The team who wrote this book  ................................................... ix  
Now you can become a published author, too!  ........................................ x  
Comments welcome  ......................................................... x  
Stay connected to IBM Redbooks  ................................................. xi  

**Summary of changes**  ........................................... xiii  
May 2013, Second Edition  ................................................. xiii  

**Chapter 1. Overview of Blue Gene/Q components**  ........................................... 1  
  1.1 Blue Gene/Q system overview  ................................................. 2  
  1.2 Blue Gene/Q hardware overview  ................................................. 3  
  1.3 Compute card  ......................................................... 4  
  1.4 Node board  ......................................................... 5  
    1.4.1 Node board torus  ......................................................... 5  
  1.5 Midplane  ........................................................ 5  
    1.5.1 Midplane torus  ......................................................... 5  
  1.6 Service card  ......................................................... 5  
  1.7 I/O drawer  ......................................................... 6  
    1.7.1 I/O links  ......................................................... 7  
    1.7.2 PCIe adapter  ......................................................... 7  
  1.8 Clock fanout card  ......................................................... 7  
  1.9 Compute rack  ......................................................... 8  
  1.10 I/O rack  ......................................................... 10  
  1.11 Cables  ........................................................ 10  

**Chapter 2. Blue Gene/Q interconnect buses**  ........................................... 11  
  2.1 Interconnection from compute card to compute card  ................................................. 12  
  2.2 Interconnection from compute card to I/O compute card  ................................................. 13  
  2.3 Interconnection from I/O drawer to federated switch  ................................................. 13  
  2.4 Interconnection from service node to compute card and I/O compute card  ................................................. 13  
  2.5 Interconnection from I/O node used as front end node to federated switch  ................................................. 14  
  2.6 Blue Gene/Q local memory bus  ................................................. 14  
    2.6.1 Memory size  ......................................................... 15  

**Chapter 3. Installation planning**  ........................................... 17  
  3.1 Blue Gene/Q basic installation requirements  ................................................. 18  
  3.2 Rack dimensions  ......................................................... 19  
    3.2.1 Compute rack layout configurations  ................................................. 25  
    3.2.2 I/O drawer and rack layout configurations  ................................................. 26  
    3.2.3 Power consumption  ......................................................... 26  
    3.2.4 Acoustics  ......................................................... 27  
  3.3 Power and electrical requirements  ................................................. 27  
    3.3.1 Power cords and circuit breaker  ................................................. 27  
    3.3.2 Power consumption  ......................................................... 28  
  3.4 Uninterruptible power supply  ................................................. 30
Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.
Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

- Redbooks (logo) ®
- Blue Gene/L®
- Blue Gene/P®
- Blue Gene/Q®
- Blue Gene®
- Global Technology Services®
- GPFS™
- IBM®
- Power Systems™
- Redbooks®

3D TRASAR and Nalco are trademarks of Nalco Company.

Other company, product, or service names may be trademarks or service marks of others.
Preface

This document is one of a series of IBM® Redbooks® written specifically for the IBM System Blue Gene® supercomputer, IBM Blue Gene/Q®. Blue Gene/Q is the third generation of massively parallel supercomputers from IBM in the Blue Gene series. This document provides an overview of components that comprise a Blue Gene/Q system. It helps System Planners and Customer Engineers plan for the installation of the Blue Gene/Q system. Information is provided about the physical requirements for the machine room where the Blue Gene/Q system is to be located. Examples of these requirements include floor (weight and cutouts), cooling, and electrical specifications.

Authors

This book was produced by a team of specialists from around the world working at the International Technical Support Organization (ITSO), Rochester Center.

James Milano is an IT Specialist at the ITSO, Rochester Center. In addition to writing Redbooks, he is a member of the IBM Blue Gene/Q installation team. Before working on this project, Jim was employed by IBM as an Electronic Engineer and then worked as an Infrastructure Specialist for Celestica. He has over 25 years of experience in working with both hardware and software. Jim has a degree in Electronic Engineering and is a member of the Phi Theta Kappa international honor society.

Pamela Lembke is a Hardware Compliance Engineer at IBM Rochester. She has five years of installation planning experience with IBM products and also works daily with the environmental compliance teams for IBM Server and Storage products. She holds a bachelors degree in Chemical Engineering from Michigan Technological University.

Special thanks to the following people for their contributions to this project:

Thomas Budnik
Matt Mattingly

Thanks to the following people for their contributions to this project:

Mark R Gallant
Fumiyasu Ishibashi
Will Jose
Takashi A Ono
Achim Rehor
Shaun Ryan
Jim A Wood

IBM Global Technology Services®

Franck Jorda

IBM Integrated Supply Chain

Paul Coteus
Todd Takken

IBM Research
Now you can become a published author, too!

Here’s an opportunity to spotlight your skills, grow your career, and become a published author - all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at: ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online Contact us review Redbooks form found at: ibm.com/redbooks
- Send your comments in an email to: redbooks@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. JLU Building 107-2
  3605 Highway 52N
  Rochester, MN 55901-7829
Stay connected to IBM Redbooks

- Find us on Facebook:
  http://www.facebook.com/IBMRedbooks
- Follow us on Twitter:
  http://twitter.com/ibmredbooks
- Look for us on LinkedIn:
  http://www.linkedin.com/groups?home=&gid=2130806
- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:
- Stay current on recent Redbooks publications with RSS Feeds:
  http://www.redbooks.ibm.com/rss.html
Summary of changes

This section describes the technical changes made in this edition of the book and in previous editions. This edition might also include minor corrections and editorial changes that are not identified.

Summary of Changes
for SG24-7872-01
for IBM System Blue Gene Solution: Blue Gene/Q Hardware Overview and Installation Planning
as created or updated on May 10, 2013.

May 2013, Second Edition

This revision reflects the addition, deletion, or modification of new and changed information described below.

New information
  ▶ Added 3.7, “Coolant monitor functions and overview” on page 35.
Overview of Blue Gene/Q components

This chapter provides an overview of the Blue Gene/Q system and describes the following system components:

- Service nodes
- Front end nodes
- File servers
- Networks

This chapter also introduces the following hardware components:

- Compute card
- Node board
- Midplane
- Service card
- I/O drawer
- Clock fanout card
- Compute rack
- I/O rack
- Cables
1.1 Blue Gene/Q system overview

The IBM Blue Gene/Q system is the third generation of massively parallel supercomputers in the IBM System Blue Gene Solution series. The Blue Gene/Q system, shown in Figure 1-1, is architecturally evolved from the IBM Blue Gene/P®.

![Blue Gene/Q system architecture](image)

The components of the Blue Gene/Q system architecture are:

**Blue Gene/Q system** The Blue Gene/Q system consists of one or more racks of specialized compute and I/O hardware explained in detail in this document. The Blue Gene/Q system contains densely packaged compute nodes, I/O drawers, and service cards. The I/O drawers connect to the functional local area network (LAN) to communicate with file servers, front end nodes (FENs), and the service node (SN). The service cards connect to the control subnets that communicate with the SN.

**File servers** Blue Gene/Q is a diskless system, so file servers must be present. A high-performance parallel file system is expected. Blue Gene/Q is flexible, accepting different file systems supported by Linux. Typical parallel file systems are IBM General Parallel File System (GPFS™) and Lustre.

**Functional LAN** The functional LAN interconnects the Blue Gene/Q system and other support systems with the parallel file system. The LAN is a high-speed network with either quad data rate (QDR) InfiniBand or 10 Gbps Ethernet. Blue Gene/Q only supports a single type of network in an installation. However, bridges and gateways can be added to enable external systems to access the file servers from different network types.
Service node

The service node provides a single point of control and administration for the Blue Gene/Q system. It is possible to operate a Blue Gene/Q system with a single service node, as shown. However, it is also possible to include distributed subnet service nodes (SSN) for high scalability.

Control subnets

The service node controls the Blue Gene/Q hardware using private Ethernet subnets. A large Blue Gene/Q system is divided into multiple subnets for scalability. The uplink from the SN (or SSN) is 10 Gbps Ethernet, which fans into numerous 1 Gbps Ethernet links into the Blue Gene/Q system. These networks must be private, and cannot be shared with other systems.

Front end nodes (FEN)

Front end nodes, also known as login nodes, are servers used for application development and job launch. These systems are installed with compilers, debuggers, job launch and scheduler commands, and other software development tools.

Site LAN

The site LAN provides connectivity for users and administrators to access the FENs and the SN. It is possible for the site LAN and functional LAN to be the same network.

Key hardware features of the Blue Gene/Q include:

- Each compute rack of Blue Gene/Q contains 1,024 compute nodes.
- Scalable to 512 compute racks.
- Engineered with fewer moving parts and with greater redundancy.
- Densely packaged with highly efficient air and liquid cooling.

In addition to the 1024 compute nodes per compute rack, there are the following components:

- Associated input/output (I/O) nodes
- A central service node (SN) computer
- Several front end node (FEN) computers for compiling and job submission
- A storage system

The largest supported Blue Gene/Q system is 256 racks or 262,144 compute nodes. The racks are connected into a 5-dimensional toroidal network carried over optical cables for processor-to-processor communication. For larger systems, a simple change can be made to support 512 racks, or 100 PetaFLOPs of peak performance.

In Blue Gene/Q, connections to file systems and other such equipment are through I/O nodes, which are housed eight to a drawer. The I/O nodes connect to the compute nodes (see 1.7.1, “I/O links” on page 7). The I/O drawers are either in separate racks or in I/O enclosures on top of the compute racks, sometimes referred to as “top hats.” Physically, I/O nodes are connected to compute nodes through optical cables.

The I/O drawers are connected to the storage system through Peripheral Component Interconnect Express (PCIe) 2.0 adapter cards available on every drawer. The SN and FENs are accessed through an Ethernet control network or an InfiniBand network.

1.2 Blue Gene/Q hardware overview

As shown in Figure 1-2 on page 4, a Blue Gene/Q system is made up of multiple components including a Blue Gene/Q compute rack and an optional Blue Gene/Q I/O rack. Additional hardware is associated with the storage subsystem, the primary service node (SN), the front end node (FEN), and the communications subsystem.
The Blue Gene/Q compute racks are cooled by either water or air. Water is used for the processing nodes. Air is used for the power supplies and the I/O drawers that are mounted in the Blue Gene/Q rack. The airflow of the power supplies is front and back to top. The Blue Gene/Q compute rack is available in either a one midplane or a two midplane solution.

Eight air-cooled I/O nodes are housed in each I/O drawer. The I/O drawers can be installed either in compute racks or in industry-standard 19-inch I/O racks. In the compute rack, up to four I/O drawers, two per midplane, can be configured using the I/O enclosure (top hat). The placement of I/O drawers in the 19-inch rack configuration is advisable in a large system installation where the number of I/O nodes cannot be accommodated in the compute racks.

1.3 Compute card

A compute card consists of one Blue Gene/Q compute application-specific integrated circuit (ASIC) and 72 SDRAM DDR3 memory chips. Thirty-two such cards plug into a node board.

Each compute card has an interintegrated circuit (I²C) addressable EEPROM¹ for storing vital product data. The data can be both written and read. In practice, however, it is written once at the factory and repair information is added if it is repaired. Each compute card also contains an I²C addressable thermal sensor device and a populated I²C addressable analog-to-digital voltage conversion device. The latter device reads both the instantaneous voltage and the maximum and minimum voltages recorded since reset.

¹ Electrically erasable programmable read-only memory
1.4 Node board

The node board, which connects the 32 compute cards into a 5-dimensional torus, is water cooled and contains the dc-dc supplies for the compute cards.

The node board also contains the field programmable gate array (FPGA) complex that provides the following access:

- Joint Test Action Group (JTAG) access to the Blue Gene/Q compute (BQC) and Blue Gene/Q link (BQL) application-specific integrated circuits (ASICs)
- I²C access to the optical transceivers and other programmable devices

Each BQC ASIC contains an on-chip thermal diode. The conversion to a digital signal occurs on the node board using an external analog-to-digital conversion chip. These chips communicate directly with the FPGA function over an I²C bus. Each BQC ASIC also contains a digitizing process and a temperature monitor (PTMON) circuit. This circuit sends temperature data directly to the node board FPGA. The node board FPGA in turn sends the averaged result back to the compute cards, along with output from the dc-dc converters indicating a power draw. This approach enables appropriate measures to be taken on the compute card in the event of thermal overload or current overdraw.

1.4.1 Node board torus

Each Blue Gene/Q node board has 32 compute cards organized as a 5-dimensional torus of a length 2. The signaling rate is 4 Gbps. All links leaving the node board to other node boards in the same midplane pass through the midplane connector. All links leaving the midplane pass through 8 link chips, also mounted on the node board. These link chips in turn drive optical transceivers, which signal at 10 Gbps, through a set of fiber optics cables connecting to other midplanes and I/O drawers.

1.5 Midplane

Node boards are inserted from both the front and rear of the vertical midplane. The service card is in a central position in the front of the midplane. A Blue Gene/Q compute rack can contain either one or two fully populated midplanes.

1.5.1 Midplane torus

The Blue Gene/Q midplane interconnection is a 512 node, 5-dimensional torus of size $4 \times 4 \times 4 \times 4 \times 2$. All connections leaving the midplane are intercepted by 128 link chips, eight per node board, and then sent to local optical transceivers.

1.6 Service card

The service card distributes the clock signals and the 1 Gb Ethernet connection to the compute rack node boards. The service card is roughly 3 cm in height and plugs horizontally into the middle of each midplane in the front of the rack.
The service card also provides all other rack functions, such as local sensing, initialization of the rack 1 Gb concentrator, and local monitoring of rack temperature and voltage. All data is captured through the local FPGA and accessed through the 1 Gb Ethernet.

1.7 I/O drawer

The Blue Gene/Q I/O board is housed in its own I/O drawer enclosure. The I/O drawer contains eight I/O compute cards. The I/O compute cards have the same compute chips as the compute cards on a node board. However, the compute cards on a node board have a water-cooled cold plate. The I/O compute cards have an air-cooled heat sink. The I/O drawer itself is completely air cooled.

Figure 1-3 shows a fully populated I/O drawer. The eight I/O compute cards, near the fans, are air-cooled. They use attached finned heat sinks instead of the cold plates. The dc-dc power converters are spread across the board. They deliver redundant power to all components on the board including the required service card function (FPGAs and clock redrivers) that is integrated into the I/O board. The link chips and their optical transceivers are upstream of the dc-dc converters. They are shown with their tall, pillar-like heat sinks. Below the link chip and the optical transceiver heat sinks are a cold plate, thermal interface, and connector system similar to the one used on the node boards.

![Diagram of I/O drawer](image)

The eight PCIe cards, shown with their individual tail stocks protruding, are in green. Also visible, mounted on the I/O board tail stock, are the optical bulkhead connectors, and in the lower-right corner, the differential clock input connector. Not shown is an I²C connector for monitoring the I/O board bulk-power enclosure. Each I/O rack contains one bulk-power enclosure, with six bulk-power modules.
Rails are mounted on the I/O enclosure to permit sliding a drawer in and out of the rack. Three fan modules, with two fans each for a total of six fans, cool the I/O drawer. If one fan is disabled, the five working fans provide sufficient airflow to cool the I/O drawer until the module with the failing fan can be replaced.

The I/O board contains the dc-dc supplies for the compute and PCIe cards. Because the power converters are direct soldered, the board is itself a large, multi-voltage power supply. The bulk power supply is monitored through one or more of the I/O drawers using a power management bus (PMBus) cable connection on the I/O drawer tail stock.

The I/O board also contains the FPGA complex. This complex provides JTAG access to the BQC and BQL ASICs and I²C access to the optical transceivers and other programmable devices. Unlike the node boards, which are accessed through the service card, I/O boards are controlled directly from the service host through the 1 Gb Ethernet.

1.7.1 I/O links

Each of the eight I/O compute cards has two I/O connections back to the compute node: one bidirectional 2 GB + 2 GB link to each of two compute ASICs. These links are carried on 12-fiber optical cables. In total, they consume two additional link chips.

1.7.2 PCIe adapter

An I/O board supports eight PCIe 2.0 adapter cards. Currently half-height or full-height and half-length or full-length adapter cards of up to 25 W are supported. All Blue Gene/Q I/O boards come equipped with eight PCIe 2.0 adapter cards. The following three adapter card choices are available:

- Dual 10 Gb Ethernet (GbE)
- Quad-data-rate (QDR) InfiniBand
- Mixed InfiniBand and Ethernet (only available in the I/O drawers used as front end nodes)

The PCIe 2.0 specifications are available at this link: http://www.pcisig.com/specifications/pciexpress/base2

1.8 Clock fanout card

The clock fanout card provides the clock fanout function.

Note: The clock fanout card diagram shown in Figure 1-4 on page 8 is only an example. It is not to be used in an actual installation.

The clock card circuit is designed to optionally contain a clock source and to serve as a system clock-master fanout. Every Blue Gene/Q system contains one clock master. The clock card receives direct 5 V power from either the lower or upper bulk power enclosure (BPE). It is mounted near the bottom of a Blue Gene/Q rack to allow for the shortest possible clock cables. Alternatively, for systems that use overhead cabling, the clock card can be mounted near the top of the rack. The system-master clock chip, including a spread-spectrum function, is controlled over the I²C from the service card. The base frequency is 100 MHz, or 1/16 of the processor frequency, and is distributed to every BQC and BQL chip.
1.9 Compute rack

The Blue Gene/Q compute rack, shown in Figure 1-5 on page 9, is custom designed to accommodate two midplane assemblies.
Figure 1-5  Blue Gene/Q compute rack

The rack is 52 inches deep, including doors, and is accessed from the front and back. The modest airflow is drawn from the front and rear of the rack and ejected out the top. Cold air, which is drawn from perforated floor tiles in the aisles in the front or back of the rack, is required to cool the bulk power modules.

Blue Gene/Q compute rack specifications:

- 82.5 in. maximum height.
- 52 in. deep from outside of door to outside of door to allow approximately 36-inch deep aisles and rack rows on 8-foot centers.
- 48 in. wide.
- Internal cable trays.
- Separate individual 51 V cables to 64 direct current assembly (DCA) cards. The service card gets 5 V from a separate cable from the BPEs on each of the two midplanes.
- Midplanes installed with a midplane stiffener.
Two card cages per midplane to guide horizontal cards into the vertical midplane from both front and back.

- Caster lockdowns to lock the rack to the floor.

### 1.10 I/O rack

The 19-inch Blue Gene/Q I/O rack is air cooled with standard front-to-back airflow. Up to 12 I/O drawers are configured in the rack and positioned close to the compute racks. Ensure that the ambient air entering the compute rack ac/dc power supplies is within the range indicated in Table 3-12 on page 34. Rear-door heat exchangers can be used to keep all aisles as cold aisles within the Blue Gene/Q configuration.

### 1.11 Cables

There are six basic cable types in Blue Gene/Q:

- Optical, 48-channel, multimode cables for extending torus connections between midplanes.
- Electrical, single-conductor, differential coaxial cables. These cables are used to distribute the differential clock signals from clock card to clock card between racks and from clock card to service card and I/O drawers.
- Electrical cables for the 1 Gbps Ethernet connections from the control switch to the service cards and I/O drawers.
- Active optical cables (AOCs) that accept electrical InfiniBand signals, convert electrical signals to optical on the cable ends, and use optical fiber within the cables. The AOCs provide an interface between the PCIe 2.0 adapter cards and the federated switch or other InfiniBand equipment.
- 10 Gb fiber Ethernet.
- Optical, 12-fiber, multimode cables that provide logical connections between I/O drawers and node boards.
Chapter 2. Blue Gene/Q interconnect buses

In this chapter, the various types of interconnections for the Blue Gene/Q system are described.
2.1 Interconnection from compute card to compute card

Compute-card-to-compute-card communication is performed through a 5-dimensional torus interconnect. This interconnection carries 2 gigabytes per second (GBps) of raw data, simultaneously, in both directions on all 10 links. At a 10% packet overhead, the bandwidth for the payload, in both directions on a single link, is 3.6 GBps. The latency of each hop is approximately 35 ns. The maximum number of hops depends on the system size.

For example, a 96-rack system can have a compute card topology of \(16 \times 16 \times 16 \times 12 \times 2\). The maximum number of hops is therefore \(8 + 8 + 8 + 6 + 1 = 31\). This results in a one-way hardware latency for the longest packet route of 2.5 µs in the absence of contention. The direct neighbor one-way latency is 0.3 µs.

Table 2-1 describes the compute card torus dimensionality of various Blue Gene/Q components.

<table>
<thead>
<tr>
<th>Block</th>
<th>Number of nodes</th>
<th>A-B-C-D-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute card</td>
<td>1</td>
<td>1-1-1-1-1</td>
</tr>
<tr>
<td>Node board</td>
<td>32</td>
<td>2-2-2-2-2</td>
</tr>
<tr>
<td>Compute midplane</td>
<td>512</td>
<td>4-4-4-4-2</td>
</tr>
<tr>
<td>Compute rack</td>
<td>1,024</td>
<td>4-4-4-8-2</td>
</tr>
<tr>
<td>Dual compute racks</td>
<td>2,048</td>
<td>4-4-8-8-2</td>
</tr>
<tr>
<td>128 compute rack system</td>
<td>131,072</td>
<td>16-16-16-16-2</td>
</tr>
</tbody>
</table>

The links internal to a midplane (\(4 \times 4 \times 4 \times 4 \times 2\)) are through circuit cards. The links leaving a midplane are on the surface of this 5-dimensional cube. They connect through a link chip to an optical transceiver. In contrast to IBM Blue Gene/L® and Blue Gene/P, the Blue Gene/Q system does not use split redirection cables. This simplification of the system is enabled by the larger number of compute blocks that can be created with the 5-dimensional torus.

All optical links have spare fibers that can be used to replace failed links without an application failure. This is accomplished by hardware identification of excessive error counts. Next, a reliability, availability, and serviceability (RAS) message is sent to the control system to spare out the faulty fiber. The routing logic continues to retry sends until the optics are repaired, which allows for a lossless recovery mechanism.

In addition to powerful Reed-Solomon error detection codes with retry on the links, the links are covered by an additional independent cyclic redundancy check (CRC). This CRC is used to give further confidence that no corrupted packets that escaped the Reed-Solomon check were transferred. The link chip also contains single-bit correction hardware for optical links to significantly reduce the rate of retries.

The compute node torus network contains an embedded collective network that handles floating-point additions and subtractions and common integer logical and arithmetic operations in hardware. This collective network allows for the building of binary-like trees that dramatically speed up collective operations.

The compute card network also carries the global barriers, which were carried on a separate bus for Blue Gene/L and Blue Gene/P. These barriers are formed and controlled by additional information carried by acknowledgment and token packets.
2.2 Interconnection from compute card to I/O compute card

The interconnection from a compute card to an I/O compute card uses the same physical layer and protocol used in the interconnection from a compute card to a compute card. The protocol in the Blue Gene/Q compute node fabric allows for directly targeting an I/O node. As a result, no software forwarding is necessary at the compute-to-I/O-node boundary.

The bandwidth of these I/O links is the same as the compute-to-compute case (2 GBps raw data rate). Each I/O node has two input links that come from two different compute nodes in the compute cluster. As a result, the input bandwidth is balanced with respect to the PCIe bandwidth ($2 \times 2$ GBps = 4 GBps).

2.3 Interconnection from I/O drawer to federated switch

This interconnection is based on PCIe 2.0 cards, which can be plugged into an I/O drawer. The links are 4 GBps InfiniBand (QDR-IB) links and dual 10 Gb Ethernet links. Every I/O compute card has one of these PCIe cards.

2.4 Interconnection from service node to compute card and I/O compute card

The service nodes are connected to the compute and I/O racks through a 1 or 10 Gb Ethernet infrastructure. The number of aggregated Ethernet links is one per row of racks. This interconnection carries the boot and diagnostic network. The network for each row is sourced by a single Ethernet link. It is fanned out in an Ethernet switch into a dedicated 1 Gb Ethernet link that serves each midplane.

Every Blue Gene/Q ASIC: compute card (compute node), I/O compute card (I/O node), and link chip, has a unique address that defines its location in the Blue Gene/Q machine. For example, a compute ASIC is defined by rack row and column, midplane (0 - 1), node board (0 - 15), and its compute card position on a node board. The rack's row, column, and midplane number are set by three hexadecimal switches that are accessible through the service card tailstock. The rack, row, and midplane address is displayed in the liquid crystal display (LCD) panel, just to the left of the hexadecimal switches on the service card tailstock.

The rack and midplane number is accessible by the service card (Pal/S) field-programmable gate array (FPGA). The ASIC location on a midplane is handled through Joint Test Action Group (JTAG) port numbers. At system initialization, the host broadcasts to all service cards asking for their license plate (the unique ID on each service card). Then a unique request is sent to each service card asking for the address of the hexadecimal switches and the master bus (midplane 0 or 1). This information is stored in a table.

This rack and midplane address is used for access by the service bus to provide the connection between the logical Ethernet address and the physical midplane location. Starting from the host computer, a unique electrical Gb cable goes from the host computer to a multiport Gb Ethernet switch on the service card.

From this switch, 16 ports go to the 16 node boards, while the 17th goes to the iCon chip. Lastly, a connection on the tailstock also goes to the iCon configuration port. This port is first configured by the host. Then iCon acts as a simple 2-way switch to direct subsequent traffic to the Gb Ethernet switch.
The 16 midplane ports go through physical layer (PHY) chips to FPGA chips on each node board. The node board FPGAs contain JTAG buses. To control up to 18 pairs of transmit and receive optical transceivers, the JTAG buses connect to the following:

- Thirty-two Blue Gene/Q compute chips (BQC)
- Eight or nine Blue Gene/Q link chips (BQL)
- I²C buses

The entire network, from host to ASICs, sensors, power supplies, and fan control is termed the service bus.

2.5 Interconnection from I/O node used as front end node to federated switch

Blue Gene/Q supports using I/O nodes as front end nodes. These nodes are Blue Gene/Q I/O drawers that are used for code compilation and other functions. If used, they are connected to the rest of the system through an InfiniBand link or Ethernet switch.

2.6 Blue Gene/Q local memory bus

The BQC ASIC contains two memory controllers. Each memory controller interfaces to a 16-byte (128-bit) data bus contained within a 144-bit memory bus. The two memory controllers run in parallel, each accessing a different cache line of sequential addresses.

The system memory is direct-attached JEDEC standard synchronous, dynamic random access memory (SDRAM) DDR3 running with a data rate of 1.33 GBps. Each memory controller activates a rank of 18 devices. Each rank is 8 data bits in width. Two ranks of memory are supported.

The data bus is terminated at the dynamic random-access memory (DRAM) during writes using the on-die termination (ODT) feature of the DRAM. The data bus is terminated at the bulk power controller (BPC)ASIC during reads using MCTT0. ODT must be enabled during a write and disabled during a read. MCTT0 must be enabled during a read and disabled during a write. Because the data nets are short, only the highest (lowest current) terminator values are used.

The address bus is terminated on the compute and I/O circuit card at the end of a daisy chain net to nine DRAM sites. There are two address buses per memory controller. A small voltage regulator is contained on each compute card and I/O compute card to provide the termination voltage.

The error-correcting code (ECC) is 72 symbols long. Sixty-four symbols are data, one is available for other purposes, and seven symbols are check symbols. Each symbol is 8 bits. There are four symbols per DRAM, one each from the four beats of data covered by the code. The code is robust. Before a hard-chip-failure, the code can correct every single-symbol failure, every double-symbol failure, detect every triple-symbol failure, and correct the vast majority of chip failures. It still detects all of the rest of the failures, even the ones it cannot correct. After a hard-chip-failure, the code corrects every single-symbol failure and detects every double-symbol failure.
2.6.1 Memory size

The minimum memory capacity is 8 GB, consisting of a single rank of 1 Gb monolithic DDR3 SDRAMs. The maximum memory capacity is 64 GB, consisting of a 4-rank memory system using dual-stacked 4 Gb DDR3 SDRAMs. Currently systems are available with 16 GB of memory from two ranks of 2 Gb DRAM.
Installation planning

This chapter addresses the physical layout of a Blue Gene/Q configuration including ac power requirements, cooling needs, and basic system information.

Because Blue Gene/Q is so unique, there are several atypical requirements for the data center that will house the unit.

- **Power:** Each Blue Gene/Q rack has a maximum input power consumption of 106 kVA. Therefore, the magnitude of the electrical infrastructure is much larger than typical equipment.

- **Cooling:** Because of the large load, each rack is cooled by air and water. The building facility water must support the Blue Gene/Q cooling system. A raised floor is not a requirement for air cooling the Blue Gene/Q.

- **Size:** The Blue Gene/Q racks are large and heavy. Appropriate actions must be taken before, during, and after installation to ensure the safety of the personnel and the Blue Gene/Q equipment.

The customer is responsible for the facility planning and data center layout before delivery of the Blue Gene/Q system. Specific details about customer responsibilities are provided throughout this chapter. IBM offers facility planning services and can be contracted for a layout. IBM can also serve as a general contractor for facility remodeling or new facility construction.
3.1 Blue Gene/Q basic installation requirements

Table 3-1 through Table 3-4 on page 19 provide the Blue Gene/Q compute rack and I/O rack dimensions and weights. The access route to the final installation site must be defined before delivery. Hallways, doors, elevators, and passageways must be large enough for a full-size Blue Gene/Q. Ensure an adequate turning radius around corners, provide overhead clearance, and adhere to weight restrictions. Because of its width, the Blue Gene/Q compute rack requires double doors at all room entrances.

The Blue Gene/Q site must have at least an 8-foot (2.44 m) height clearance to allow space for installation tools. A material handler, such as the battery-powered Alum-a-lift, is used to install the Blue Gene/Q. See “Alum-a-lift tool” on page 80 for more information.

All floors that will be in contact with the Blue Gene/Q must support the full weight of the Blue Gene/Q. This includes any hallways, elevators, or storage rooms where the Blue Gene/Q might sit or that it might be rolled over to get to the installation location.

An area of at least 10 feet by 23 feet is required to remove the Blue Gene/Q from the shipping pallet.

Table 3-1  Blue Gene/Q compute rack and I/O rack dimensions

<table>
<thead>
<tr>
<th>Item description</th>
<th>Height</th>
<th>Width</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute rack plus I/O enclosure (top hat)</td>
<td>2082 mm</td>
<td>1219.2 mm</td>
<td>1219.2 mm</td>
</tr>
<tr>
<td></td>
<td>82 inches</td>
<td>48 inches</td>
<td>48 inches</td>
</tr>
<tr>
<td>Compute rack with I/O enclosure and side covers(^a)</td>
<td>2082 mm</td>
<td>1219.2 mm</td>
<td>1321 mm</td>
</tr>
<tr>
<td></td>
<td>82 inches</td>
<td>48 inches</td>
<td>52 inches</td>
</tr>
<tr>
<td>Compute rack shipping dimensions, with packaging</td>
<td>2273 mm</td>
<td>1194 mm</td>
<td>1321 mm</td>
</tr>
<tr>
<td></td>
<td>89.5 inches</td>
<td>47 inches</td>
<td>52 inches</td>
</tr>
<tr>
<td>Compute rack without I/O enclosures as shipped</td>
<td>1806.4 mm</td>
<td>1034.6 mm</td>
<td>1219.2 mm</td>
</tr>
<tr>
<td></td>
<td>71.1 inches</td>
<td>40.73 inches</td>
<td>48 inches</td>
</tr>
<tr>
<td>I/O rack</td>
<td>2015 mm</td>
<td>644 mm</td>
<td>1098 mm</td>
</tr>
<tr>
<td></td>
<td>79.3 inches</td>
<td>25.4 inches</td>
<td>43.4 inches</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with side panels</td>
<td></td>
</tr>
<tr>
<td>I/O rack with shipping packaging</td>
<td>2095.5 mm</td>
<td>1219.2 mm</td>
<td>1219.2 mm</td>
</tr>
<tr>
<td></td>
<td>82.5 inches</td>
<td>48 inches</td>
<td>48 inches</td>
</tr>
</tbody>
</table>

\(^a\) Side covers add 0.6 inches (1.5 cm) to each side of the rack. Side covers are only used on the end of the rack rows. A single rack with side covers has a width of 49.2 inches.

Table 3-2  Blue Gene/Q compute rack weights

<table>
<thead>
<tr>
<th>Blue Gene/Q components</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double midplane rack, with covers, no I/O drawers, dry.</td>
<td>1973 kg</td>
</tr>
<tr>
<td></td>
<td>4350 lbs</td>
</tr>
<tr>
<td>Double midplane rack with I/O drawer capability, I/O enclosure (top hat), and covers, dry.</td>
<td>2016 kg</td>
</tr>
<tr>
<td></td>
<td>4445 lbs</td>
</tr>
<tr>
<td>Single midplane rack with covers, no I/O drawers, no I/O enclosure (top hat), dry.</td>
<td>1236 kg</td>
</tr>
<tr>
<td></td>
<td>2725 lbs</td>
</tr>
</tbody>
</table>
Table 3-3  Blue Gene/Q compute rack shipping weights

<table>
<thead>
<tr>
<th>Blue Gene/Q components</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single midplane rack, with I/O enclosure (top hat) and covers, dry.</td>
<td>1279 kg 2820 lbs</td>
</tr>
<tr>
<td>Individual I/O drawers.</td>
<td>31 kg 68 lbs</td>
</tr>
<tr>
<td>Water per rack. <strong>Note</strong>: Racks are shipped dry, so this weight only applies after the installation is complete.</td>
<td>18 kg 40 lbs</td>
</tr>
<tr>
<td>Maximum point load at castor.</td>
<td>719 kg 1586 lbs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I/O rack</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/O rack, including covers and doors, line cord, and bulk power</td>
<td>343 kg 755 lbs</td>
</tr>
<tr>
<td>Each individual I/O drawer</td>
<td>31 kg 68 lbs</td>
</tr>
<tr>
<td>Cables for each I/O drawer</td>
<td>6.4 kg 14 lbs</td>
</tr>
<tr>
<td>Rear door heat exchanger</td>
<td>29.9 kg 66 lbs</td>
</tr>
</tbody>
</table>

3.2 Rack dimensions

Figure 3-1 on page 20 provides a front view of a Blue Gene/Q compute rack. The dimensions are for the rack as shipped without an I/O enclosure or right-side bulk power enclosures (BPEs).
Figure 3-1 Compute rack as shipped with no I/O enclosure (top hat) or right side BPEs, front view (internal components omitted), mm (in)

Figure 3-2 on page 21 provides a bottom view of the same Blue Gene/Q compute rack. The dimensions for casters are to the center of the pivoting point, not the load.
Figure 3-3 on page 22 provides a top-down view of the Blue Gene/Q compute rack. The blue diagonal lines represent the bulk power exhaust. The pink dotted lines represent the I/O drawer exhaust, when I/O drawers are present.
**Figure 3-3** Compute rack, top-down view

**This dimension (see the upper-right corner in the figure) is the distance from the rack to the back of the covers. Air can occupy this area.**

Figure 3-4 on page 23 provides a front view of a Blue Gene/Q compute rack with an I/O enclosure and field-installed baffles. Figure 3-5 on page 24 provides a bottom view of the same configuration.
Figure 3-4  Compute rack with top I/O enclosure and field-installed baffles, front view (internal components omitted)
Table 3-5 lists the I/O cables that can be used to connect a Blue Gene/Q rack to the I/O rack. The two 12-fiber bundles, blue and brown, connect I/O nodes in compute racks to the compute node cards in the same rack. The individual cables connect I/O drawers in an I/O rack to node boards in a compute rack or to connect an I/O torus.

Table 3-5 12-fiber I/O data cables

<table>
<thead>
<tr>
<th>Part number</th>
<th>Length</th>
<th>Band color</th>
</tr>
</thead>
<tbody>
<tr>
<td>45D9545</td>
<td>26.6</td>
<td>Gray</td>
</tr>
<tr>
<td>45D9544</td>
<td>18.7</td>
<td>Red</td>
</tr>
<tr>
<td>45D9543</td>
<td>10.3</td>
<td>Orange</td>
</tr>
<tr>
<td>45D9542</td>
<td>7.8</td>
<td>Yellow</td>
</tr>
<tr>
<td>45D9541</td>
<td>6.2</td>
<td>Green</td>
</tr>
<tr>
<td>45D9552</td>
<td>Bundle of 4</td>
<td>Blue</td>
</tr>
<tr>
<td>45D9553</td>
<td>Bundle of 4</td>
<td>Brown</td>
</tr>
<tr>
<td>45D9533</td>
<td>3.3</td>
<td>Violet</td>
</tr>
</tbody>
</table>

Figure 3-6 on page 25 shows an example floor plan for a Blue Gene/Q room with separate I/O racks.
3.2.1 Compute rack layout configurations

While the Blue Gene/Q compute rack is approximately square, it does have a front and a back. The service card is only in the front of the rack and any I/O drawers in the compute rack slide out to the front. All Blue Gene/Q central electronics complex rows, and racks within a row, need to face the same direction (see Figure 3-6) in a front-to-back layout. Because there are power supplies taking in air on both sides of the rack, the aisle is set up as a cold/cold aisle.

Pitch is the distance from the front of one row of equipment racks to the front of the adjacent row of equipment racks. The minimum pitch of the compute rack rows is 7 feet, or 3½ 24-inch tiles. The maximum pitch is 8 feet, or four 24-inch tiles. The pitch for installations with 600 mm tiles is 2.4 meters, or four tiles. Compute racks within a compute row are on a 4-foot pitch (122 cm).

Table 3-6 shows possible rack layout configurations for Blue Gene/Q. After determining the number of racks for your facility, a recommended layout can be picked from the table. If you need a different configuration from what is listed in Table 3-6, contact your IBM representative.

<table>
<thead>
<tr>
<th>Number of racks</th>
<th>Rows</th>
<th>Columns</th>
<th>Number of racks</th>
<th>Rows</th>
<th>Columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2 racks (1 midplane)</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>24</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>
3.2.2 I/O drawer and rack layout configurations

The I/O drawer can be placed in the I/O enclosure (top hat) of a compute rack or in the IBM 19-inch rack. Up to two I/O drawers are allowed per rack midplane. Airflow of the I/O drawers in the compute rack is front to top (see Figure 3-3 on page 22). That is, the airflow is front to back within the drawer, but the drawer does not extend to the end of the rack and into the back aisle. When placed in a 19-inch rack, the airflow direction is front to back. The rear door heat exchanger can be used to keep the exhaust air cool. The Blue Gene/Q system does not require a hot aisle/cold aisle setup.

3.2.3 Service clearance

Between the compute rows, the minimum clearance is 36 inches. There is no side clearance requirement for the compute rack. However, the Blue Gene/Q compute racks doors must be able to open 180°. Racks or other equipment deeper than 4 feet must not be positioned within 30 in. of the Blue Gene/Q compute rack. Service clearance for the I/O rack is also 36 inches in the front and back of the rack. For additional information about I/O rack clearances, see http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/iphad/f7014t42rack.htm

Note: The link provides Model 7014-T42 rack specifications. The Blue Gene/Q Model 207-200 FC 0215 rack uses the same covers and requires the same clearances.

<table>
<thead>
<tr>
<th>Number of racks</th>
<th>Rows</th>
<th>Columns</th>
<th>Number of racks</th>
<th>Rows</th>
<th>Columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>32</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>32</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>48</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>48</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>48</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>8</td>
<td>56</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
<td>64</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2</td>
<td>64</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>2</td>
<td>72</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>4</td>
<td>80</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>6</td>
<td>96</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>8</td>
<td>96</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>4</td>
<td>96</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>
3.2.4 Acoustics

The acoustic values for the Blue Gene/Q system follow:

- Compute rack with no I/O drawers: 7.4 bels
- Compute rack with 1 I/O drawer: 7.8 bels
- Compute rack with 4 I/O drawers: 8.0 bels
- I/O rack with 12 I/O drawers: 8.25 bels

3.3 Power and electrical requirements

This section describes the power and electrical requirements that must be met at a Blue Gene/Q installation site.

3.3.1 Power cords and circuit breaker

The Blue Gene/Q has one line cord per power enclosure. A system with two midplanes has four line cords, two per side. A system with one midplane has two line cords, one per side. The lower midplane power cords connect near the bottom of the front and back of the rack. The upper midplane power cords connect near the middle of the front and back of the rack. Therefore, the length of cable available for connecting power in the vertical direction varies. Place the required receptacles within reach of the line cord:

- For line cords overhead, the available lower and upper cord lengths for connections from the top of the rack are 2.35 m (7.7 ft.) and 1.5 m (4.8 ft.) to reach the facility connector.

  **Note:** The lower power cable takes approximately 203 cm (80 in.) of cable inside the rack. The upper cable requires 117 cm (46 in.) of cable inside the rack.

- For line cords laid under the floor, the available cord lengths from the bottom of the rack are 2.8 m (9.25 ft.) and 2 m (6.7 ft.) to reach the facility connector.

  **Note:** The lower cable takes approximately 68.6 cm (27 in.) of cable inside the rack. The upper cable requires 146 cm (57.5 in.) of cable inside the rack.

Also, note the following:

- The bend radius of the power cord is 190 mm (7.5 in.).
- Any extra lengths of power cord must be managed by the customer.
- Each compute rack must have its own circuit breaker.
- All Blue Gene/Q configurations use a four pole, five wire (4P 5W) connector that requires a neutral.
- Blue Gene/Q racks are not ideally balanced 3-phase loads. Under normal operating conditions, there is small phase-loading imbalance. Running at reduced capacity in a degraded state might cause a significant phase-loading imbalance. A neutral conductor with the same current rating as the three power phases must be present.
- Each line cord powers one side of a midplane. There is no line cord redundancy with Blue Gene/Q.
3.3.2 Power consumption

Table 3-7 contains information about the electrical specifications for the Blue Gene/Q rack.

<table>
<thead>
<tr>
<th>Compute rack</th>
<th>I/O rack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage range</td>
<td>380 - 415 VAC</td>
</tr>
<tr>
<td>System rating</td>
<td>40A</td>
</tr>
<tr>
<td>Voltage tolerance</td>
<td>+10% to -10%</td>
</tr>
<tr>
<td>Frequency</td>
<td>47 - 63 Hz</td>
</tr>
<tr>
<td>Per-line ac input</td>
<td>41A</td>
</tr>
<tr>
<td>Inrush current</td>
<td>165 Amps peak per phase. See note 1.</td>
</tr>
<tr>
<td>Plug rating</td>
<td>63A</td>
</tr>
<tr>
<td>Number of line cords</td>
<td>2 per midplane 4 per rack</td>
</tr>
<tr>
<td>Cord size</td>
<td>6AWG</td>
</tr>
<tr>
<td>Cord length</td>
<td>3.5 m 11.5 ft.</td>
</tr>
<tr>
<td>Recommended breaker size</td>
<td>50A</td>
</tr>
<tr>
<td>Recommended receptacle</td>
<td>HBL563R6W</td>
</tr>
</tbody>
</table>

Table 3-8 outlines the Blue Gene/Q power consumption.

<table>
<thead>
<tr>
<th>Maximum design input power. See note 2.</th>
<th>LINPACK measured power. See note 3.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Gene/Q rack Two midplanes</td>
<td>106 kVA</td>
</tr>
<tr>
<td>Blue Gene/Q rack Single midplane</td>
<td>53 kVA</td>
</tr>
<tr>
<td>I/O rack</td>
<td>16.5 kVA</td>
</tr>
</tbody>
</table>

Typical Blue Gene/Q rack power consumption is about 65 kW. See note 4.
Notes:
1. See Figure 3-7 on page 29 and Figure 3-8 on page 30 for inrush waveform decay.
2. The same power supplies feed the compute cards and I/O drawers. The maximum input power is constant.
3. No I/O drawers installed on top of the compute rack. Add 0.6 kW for each I/O drawer.
4. Typical compute rack power consumption is an estimate. It assumes that the system is operating in the recommended environmental envelope, has no hardware failures, and is an average workload application. Actual results can vary.

Figure 3-7 illustrates the input current drawn by the central electronics complex rack.

Figure 3-7  Blue Gene/Q central electronics complex rack inrush waveform

Figure 3-8 on page 30 illustrates the input current drawn by the I/O rack.
3.4 Uninterruptible power supply

An uninterruptible power supply often serves two purposes:

- Provides backup power for critical customer operations or those operations that might lose data if the system is suddenly without power (such as storage/disk writing operations)
- Provides power conditioning for the equipment

For Blue Gene/Q compute racks attached to external disks, an uninterruptible power supply is not required. However, IBM does recommend using an uninterruptible power supply system or motor generator for the host equipment, including service node, front end nodes, switches, and associated storage devices to provide regulated power. If compute racks have I/O enclosures (top hats) with I/O, the customer can decide if an uninterruptible power supply is needed.

In general, a conscious decision must be made by each customer of a Blue Gene/Q solution relative to backup power implementation. The primary consideration for the backup power decision is system availability. When high system availability is important, backup power is recommended. A secondary consideration is the quality of power feeds. If low quality power is anticipated, an uninterruptible power supply provides power conditioning benefits.

3.5 Thermal requirements

This section describes Blue Gene/Q heating, cooling, and air environment requirements.
3.5.1 Rack heat load and cooling requirements

For the Blue Gene/Q compute rack, approximately 91% of the cooling is provided by water and 9% of the cooling is accomplished by air. For the air cooling portion, the air is drawn into the rack from both the front and back. Hot air is exhausted out the top of the rack. For compute racks, a hot/cold aisle is not required.

Figure 3-9 shows the rack flow rates by inlet water temperature. The shaded region is the required operating zone. The graph shows both liters per minute (left y axis) and gallons per minute (right y axis). Moving along the red arrow increases the thermal performance of the system. The maximum pressure entering the rack cannot exceed 95 pounds per square inch (psi). The rack entrance begins at the quick connect that is part of the rack. It does not include any hose from this quick connect point to the facility.

![Figure 3-9 Rack coolant flow rate by inlet temperature](image)

The Blue Gene/Q has two mechanisms to prevent overheating. Internal sensors measure component temperatures and stop any workload if there is an over temperature situation. Also, a coolant monitor sends an alarm and stops workloads if the water temperature is too high.

Figure 3-10 on page 32 shows the temperature change of the water based on the flow rate and power draw of the Blue Gene/Q.
Rack pressure drop, which varies by flow rate, is as shown in Figure 3-11. Rack pressure drop includes the quick connect and the hose that is attached to the rack. An additional pressure drop for any added hose that connects the Blue Gene/Q hose to the facility water must be included in the pressure drop calculations. Table 3-9 shows the additional pressure drop.

![Figure 3-10 Coolant temperature increase by flow rate and rack power](image)

**Figure 3-10** Coolant temperature increase by flow rate and rack power

<table>
<thead>
<tr>
<th>Rate Flow Rate V [GPM]</th>
<th>Water-Cooled Rack Power P [kW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>91</td>
<td>80</td>
</tr>
<tr>
<td>80</td>
<td>60</td>
</tr>
<tr>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

![Figure 3-11 Rack water flow pressure drop versus rack flow rate](image)

**Figure 3-11** Rack water flow pressure drop versus rack flow rate

<table>
<thead>
<tr>
<th>Flow rate (gpm)</th>
<th>Additional pressure drop</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.36 psi/meter of hose</td>
</tr>
</tbody>
</table>

**Table 3-9 Pressure drop per hose length**
For all installations, the operating point of water must be 2°C above the dew point of the room so that condensation does not occur in the data center. The listed 16°C lower boundary is intended to prevent condensation on the pipes in the system. If the humidity is controlled so that the dew point is decreased to a lower temperature, the water temperature can also be decreased as low as 7.5°C. Table 3-10 and Table 3-11 show the rack heat load and airflow requirements.

**Table 3-10  Rack heat load**

<table>
<thead>
<tr>
<th></th>
<th>Blue Gene/Q rack at maximum power draw</th>
<th>I/O rack at maximum power draw</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>100 kW 342 kBTU/hr</td>
<td>16 kW</td>
</tr>
<tr>
<td>Heat to water</td>
<td>91 kW 311 kBTU/hr</td>
<td>N/A</td>
</tr>
<tr>
<td>Heat to air</td>
<td>9 kW 31 kBTU/hr</td>
<td>16 kW</td>
</tr>
</tbody>
</table>

**Table 3-11  Rack airflow requirements**

<table>
<thead>
<tr>
<th></th>
<th>Blue Gene/Q rack at maximum power draw</th>
<th>I/O rack at maximum power draw</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total per rack. Maximum includes four I/O drawers</td>
<td>2200 cubic feet per minute (CFM)</td>
<td>3350 CFM</td>
</tr>
<tr>
<td>Total for bulk power supplies. Two midplane solution</td>
<td>1200 CFM&lt;sup&gt;a&lt;/sup&gt;</td>
<td>350 CFM</td>
</tr>
<tr>
<td>Total for each I/O drawer</td>
<td>250 CFM</td>
<td>250 CFM</td>
</tr>
</tbody>
</table>

<sup>a</sup> Airflow is split evenly between the front and back of the rack.

For the I/O rack, the air is drawn from the front to the back of the rack. Air inlet temperatures must be within the required limits. If this can be achieved by alternative (nontraditional) cooling methods, a hot/cold aisle is not needed. Each I/O rack has one set of power supplies (350 CFM). Airflow calculations must include the power supplies and airflow for the total number of I/O drawers.

While not required, the IBM Rear Door Heat eXchanger can be used on the back of the Blue Gene/Q MTM 0207-200 (formerly 7014-T42) I/O rack. The planning information for the Rear Door Heat eXchanger is available on the IBM website or on the Power Systems™ information center:

3.5.2 Air environmental requirements

Table 3-12 outlines the Blue Gene/Q environmental requirements.

<table>
<thead>
<tr>
<th></th>
<th>Allowable operating range</th>
<th>Recommended operating range</th>
<th>Storage</th>
<th>Shipping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>See ASHRAE white paper.</td>
<td>See ASHRAE white paper.</td>
<td>1°C - 60°C</td>
<td>-40°C - 60°C</td>
</tr>
<tr>
<td>Relative humidity</td>
<td>See ASHRAE white paper.</td>
<td>See ASHRAE white paper.</td>
<td>5% - 80%</td>
<td>5% - 100%</td>
</tr>
<tr>
<td>Maximum altitude</td>
<td>0 - 3000 ft., derated 1°C for every additional 1000 ft. up to 10000 ft.</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>


Blue Gene/Q follows the American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) 2008 Class 1 specifications for recommended and allowable operating regions. The white paper addressing this topic is available at:


IBM requires system operation within the recommended range. Operation outside of the recommended range can result in voiding of IBM Warranty or IBM Maintenance Service Agreements.

For all installations, the operating point of water must be above the dew point of the room so that condensation does not occur in the data center. The listed lower boundary is intended to prevent against condensation on the pipes in the system. If the humidity is controlled so that the dew point is decreased to a lower temperature, the water temperature can also be decreased.

For the I/O rack, the air is drawn from the front to the back of the rack. Air inlet temperatures must be within the required limits. If this can be achieved by alternative (nontraditional) cooling methods, a hot/cold aisle is not needed. While not required, the IBM Rear Door Heat eXchanger (RDHX) can be used on the back of the Blue Gene/Q MTM 0207-200 I/O rack. The planning information for the Rear Door Heat eXchanger is available on the IBM website or on the Power Systems information center:


### 3.6 Rack-level and facility-level water leak detection

A coolant monitor is installed inside each Blue Gene/Q rack to eliminate potential problems from a rack-level leak. A leak inside the rack that is higher than a specific flow rate is detected
and isolated by the coolant monitor; however, leaks less than that flow rate and leaks outside of the rack, such as a leak from a facility pipe, are not detected. A facility-level leak detection system, such as floor sensors, must be installed. It can detect leaks too small for the coolant monitor to identify or leaks outside of the Blue Gene/Q rack.

### 3.7 Coolant monitor functions and overview

Figure 3-12 shows the coolant monitor and its components.

![Figure 3-12 Coolant monitor](image)

All of the sensors on the manifold have specific calibration data. This calibration is performed by the vendor. This calibration data is stored inside the coolant monitor. For this reason, the sensors and the coolant monitor must remain as a matched set. There is a MAC address label on the components to show that they are matched. It is important to save the calibration data during the installation phase and after every calibration, which allows the data to be restored if there is a FRU replacement of the coolant monitor.

### 3.7.1 Error detection

If any of the following conditions are true, both the coolant flow and 51 V power to the node and I/O boards are shut off:

- A slow leak is detected
- A catastrophic leak is detected
- The inlet coolant temperature is too low
- The inlet coolant pressure is too high

If the inlet water flow is too low, the 51 V power to the node and I/O boards is shut off, but the coolant flow remains on.

---

1 Image source: Courtesy of Proteus Industries
If the inlet coolant temperature is too low, the inlet coolant pressure is too high, or the inlet water flow is too low, the coolant monitor must observe the operating conditions to be outside the threshold for 30 seconds before it takes action.

All other errors are indicated with a RAS message only.

If the 51 V power to the node and I/O boards is shut off or coolant flow is shut off, the reason for the shutoff is written to a register. This message is latched so that debugging can occur. When the rack is returned to normal functionality, this shutoff cause register must be cleared.

### 3.7.2 Calibration

There are two methods of calibration for the coolant monitors: the flow sensor calibration and the hydrodynamic impedance calibration.

**Flow sensor calibration**

As water passes over the flow sensor, a diaphragm vibrates. The frequency of this vibration is converted to a flow reading. Anomalies, such as air bubbles or debris, can cause incorrect readings over a short period of time. These effects cannot be monitored. It is important to have correct filtering, as described in 3.9.3, “Filtration” on page 42, and the rack must be bled, as described in the *IBM System Blue Gene Solution: Blue Gene/Q Hardware Installation and Maintenance Guide*, SG24-7974 Redbooks publication:

- When the flow sensors are calibrated, the purpose is to match the two flow sensors to give the same flow reading on the supply and return sensors. In this calibration, the return sensor is assumed to be correct. This calibration adjusts the supply sensor calibration to better match the return sensor. At lower flow rates, more noise enters the readings, making the sensors less accurate.
- The sole purpose of matching the supply and return flow sensors is to capture slow leaks and to avoid false slow leak shutdowns.
- The vendor calibrates the flow sensors, but each rack has unique characteristics that require this calibration. This calibration is done at IBM manufacturing, but it might be necessary to perform this calibration in the field.

**Hydrodynamic impedance calibration**

The hydrodynamic impedance calibration sets a unique flow versus pressure drop profile for every rack. The hydrodynamic impedance function has two purposes:

- It identifies when node card hoses are not correctly reconnected after a service action.
- It identifies when something internal to the rack causes a change in flow, such as a kinked hose or a clog.

The flow versus pressure drop curve should be upward trending, as shown in Figure 3-13 on page 37.
The operating point of the rack (pressure drop and flow rate) at any particular point is compared to the operating curve. If the operating point varies too far from the curve, an impedance error is posted.

A high impedance alarm (status "1" is displayed) means that the flow is higher than expected for that given pressure. The operating point is below and to the right of the calibration curve. This position indicates that the impedance is low, which allows a higher flow rate.

A low impedance alarm (status "2" is displayed) means that the flow is lower than expected at that given pressure. The operating point is above and to the left of the calibration curve. This position indicates that the impedance is high, which allows a lower flow rate. A low impedance alarm message is usually sent when a node board is disconnected for service.

Impedance errors do not affect rack functionality if all node hoses are plugged in and are not kinked, and if there are no clogged quick connects. A clogged quick connect leads to lower flow rates. The lower flow rates reduce heat transfer from the compute cards to the water and raise the temperature of the compute card. A hydrodynamic impedance error and higher temperature for a node card might indicate that the quick connect is clogged.

If an impedance error exists after a service action is completed, a notification is sent through the control system, but the service action is not allowed to complete.

It is acceptable to have one false hydrodynamic impedance alarm per week per rack. This alarm does not cause an action to be taken by the coolant monitor.

When the flow sensors are calibrated, the hydrodynamic impedance calibration is set at the same time. There is a separate calibration script to set only the hydrodynamic impedance.

The slow leak algorithm compares the difference of the supply and return flow sensors over a specified duration, sums these differences over the duration, and compares the result to the slow leak threshold. If the sum is higher than the threshold, the coolant monitor determines that there is a leak and takes action. The slow leak algorithm is designed to capture leaks greater than 0.45 gallons per minute, and takes action after three consecutive time cycles where there is a slow leak. This prevents false alarms. If a slow leak action is taken but there...
are no leaks present, the flow sensors might have drifted out of calibration. Flow sensor calibration typically resolves this situation.

The catastrophic leak algorithm does not compare the differences between the flow sensors. The catastrophic leak algorithm is a complicated function that analyzes many different variables that indicate a large leak. The catastrophic leak algorithm is expected to capture leaks greater than 3.75 gallons per minute.

**Coolant monitor connections**

The coolant monitor connects to the solenoid valve and the pressure and flow sensors through cables. It is important to connect these cables correctly. Swapping the pressure sensor cable with the flow sensor cable damages the flow sensor. Swapping the supply cables with the return cables causes incorrect data to be passed to the control system. The cables are keyed to prevent swapping the flow and the pressure sensor cables, but there are no keys to prevent the supply and return cables from being swapped.

The coolant monitor connects to the system with a cable harness. One 15-pin D shell connects to the coolant monitor. On the other end of the harness, there are three connectors. One connector connects to the service card for communication to the control system. The other two connectors connect to BPEs. In a two-midplane rack, the coolant monitor connects to the front two BPEs. In a single-midplane rack, the coolant monitor connects to the bottom two BPEs.

**Attention:** This cable cannot be hot plugged or unplugged. This applies to either the coolant monitor side of the cable harness or either of the BPE sides of the cable harness. Doing so can damage the coolant monitor.

**Coolant monitor threshold values**

Table 3-13 lists the level 1 system alarm thresholds for a two-midplane system.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>X1L</th>
<th>X1L Action</th>
<th>X1H</th>
<th>X1H action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply flow rate (VS)</td>
<td>GPM</td>
<td>17</td>
<td>Message</td>
<td>30</td>
<td>Message</td>
</tr>
<tr>
<td>Return flow rate (VR)</td>
<td>GPM</td>
<td>17</td>
<td>Nothing</td>
<td>30</td>
<td>Nothing</td>
</tr>
<tr>
<td>Supply water temperature (TS)</td>
<td>Degrees C</td>
<td>DP + 2</td>
<td>Message</td>
<td>25</td>
<td>Message</td>
</tr>
<tr>
<td>Return water temperature (TR)</td>
<td>Degrees C</td>
<td>18</td>
<td>Nothing</td>
<td>37</td>
<td>Nothing</td>
</tr>
<tr>
<td>Supply water pressure (pS)</td>
<td>PSI</td>
<td>20</td>
<td>Message</td>
<td>95</td>
<td>Message</td>
</tr>
<tr>
<td>Differential water pressure (Δp)</td>
<td>PSI</td>
<td>7</td>
<td>Message</td>
<td>45</td>
<td>Message</td>
</tr>
<tr>
<td>Ambient air temperature (TAIR)</td>
<td>Degrees C</td>
<td>16</td>
<td>Nothing</td>
<td>40</td>
<td>Nothing</td>
</tr>
</tbody>
</table>
Table 3-14 lists the level 2 system alarm thresholds for a two-midplane system.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>X1L</th>
<th>X1L Action</th>
<th>X1H</th>
<th>X1H action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambient relative humidity (RH)</td>
<td>%</td>
<td>25</td>
<td>Nothing</td>
<td>60</td>
<td>Nothing</td>
</tr>
<tr>
<td>Ambient dew point (DP)</td>
<td>Degrees C</td>
<td>5.5</td>
<td>Message</td>
<td>15</td>
<td>Message</td>
</tr>
<tr>
<td>System power utilization (Q)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>kW</td>
<td>0</td>
<td>Nothing</td>
<td>100</td>
<td>Nothing</td>
</tr>
<tr>
<td>Flow impedance tolerance (e)</td>
<td></td>
<td>1.7</td>
<td>Message</td>
<td>1.7</td>
<td>Message</td>
</tr>
</tbody>
</table>

<sup>a</sup> Based on water flow and temperature rise

---

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>X2L</th>
<th>X2L action</th>
<th>X2H</th>
<th>X2H action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply flow rate (VS)</td>
<td>GPM</td>
<td>15</td>
<td>Shut off</td>
<td>35</td>
<td>Message</td>
</tr>
<tr>
<td>Return flow rate (VR)</td>
<td>GPM</td>
<td>0</td>
<td>Nothing</td>
<td>100</td>
<td>Nothing</td>
</tr>
<tr>
<td>Supply water temperature (TS)</td>
<td>Degrees C</td>
<td>DP + 1</td>
<td>Shut off power and water</td>
<td>30</td>
<td>Message</td>
</tr>
<tr>
<td>Return water temperature (TR)</td>
<td>Degrees C</td>
<td>0</td>
<td>Nothing</td>
<td>46</td>
<td>Nothing</td>
</tr>
<tr>
<td>Supply water pressure (pS)</td>
<td>PSI</td>
<td>15</td>
<td>Message</td>
<td>110</td>
<td>Shut off Power and Water</td>
</tr>
<tr>
<td>Differential water pressure (Δp)</td>
<td>PSI</td>
<td>7</td>
<td>Message</td>
<td>50</td>
<td>Message</td>
</tr>
<tr>
<td>Ambient air temperature (TAIR)</td>
<td>Degrees C</td>
<td>15</td>
<td>Message</td>
<td>46</td>
<td>Message</td>
</tr>
<tr>
<td>Ambient relative humidity (RH)</td>
<td>%</td>
<td>20</td>
<td>Nothing</td>
<td>80</td>
<td>Nothing</td>
</tr>
<tr>
<td>Ambient dew point (DP)</td>
<td>Degrees C</td>
<td>0</td>
<td>Message</td>
<td>17</td>
<td>Message</td>
</tr>
<tr>
<td>System power utilization (Q)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>kW</td>
<td>0</td>
<td>Nothing</td>
<td>100</td>
<td>Nothing</td>
</tr>
<tr>
<td>Flow impedance tolerance (e)</td>
<td></td>
<td>100</td>
<td>Nothing</td>
<td>100</td>
<td>Nothing</td>
</tr>
</tbody>
</table>

<sup>a</sup> Based on water flow and temperature rise
Table 3-15 lists the level 1 system alarm thresholds for a single-midplane system.

Table 3-15  System alarm thresholds for a single-midplane system

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>X1L</th>
<th>X1L Action</th>
<th>X1H</th>
<th>X1H action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply flow rate (VS)</td>
<td>GPM</td>
<td>8.5</td>
<td>Message</td>
<td>15</td>
<td>Message</td>
</tr>
<tr>
<td>Return flow rate (VR)</td>
<td>GPM</td>
<td>8.5</td>
<td>Nothing</td>
<td>15</td>
<td>Nothing</td>
</tr>
<tr>
<td>Supply water temperature (TS)</td>
<td>Degrees C</td>
<td>DP + 2</td>
<td>Message</td>
<td>25</td>
<td>Message</td>
</tr>
<tr>
<td>Return water temperature (TR)</td>
<td>Degrees C</td>
<td>16</td>
<td>Nothing</td>
<td>37</td>
<td>Nothing</td>
</tr>
<tr>
<td>Supply water pressure (pS)</td>
<td>PSI</td>
<td>20</td>
<td>Message</td>
<td>95</td>
<td>Message</td>
</tr>
<tr>
<td>Differential water pressure (Δp)</td>
<td>PSI</td>
<td>7</td>
<td>Message</td>
<td>45</td>
<td>Message</td>
</tr>
<tr>
<td>Ambient air temperature (TAIR)</td>
<td>Degrees C</td>
<td>16</td>
<td>Nothing</td>
<td>40</td>
<td>Nothing</td>
</tr>
<tr>
<td>Ambient relative humidity (RH)</td>
<td>%</td>
<td>25</td>
<td>Nothing</td>
<td>60</td>
<td>Nothing</td>
</tr>
<tr>
<td>Ambient dew point (DP)</td>
<td>Degrees C</td>
<td>5.5</td>
<td>Message</td>
<td>15</td>
<td>Message</td>
</tr>
<tr>
<td>System power utilization (Q)^a</td>
<td>kW</td>
<td>0</td>
<td>Nothing</td>
<td>50</td>
<td>Nothing</td>
</tr>
<tr>
<td>Flow impedance tolerance (e)</td>
<td></td>
<td>1.7</td>
<td>Message</td>
<td>1.7</td>
<td>Message</td>
</tr>
</tbody>
</table>

a. Based on water flow and temperature rise

Table 3-16 lists the level 2 system alarm thresholds for a single-midplane system.

Table 3-16  System alarm thresholds for a single-midplane system

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>X2L</th>
<th>X2L Action</th>
<th>X2H</th>
<th>X2H action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply flow rate (VS)</td>
<td>GPM</td>
<td>7.5</td>
<td>Shutoff power</td>
<td>17.5</td>
<td>Message</td>
</tr>
<tr>
<td>Return flow rate (VR)</td>
<td>GPM</td>
<td>0</td>
<td>Nothing</td>
<td>50</td>
<td>Nothing</td>
</tr>
<tr>
<td>Supply water temperature (TS)</td>
<td>Degrees C</td>
<td>DP + 1</td>
<td>Shut off power and water</td>
<td>30</td>
<td>Message</td>
</tr>
</tbody>
</table>
3.8 Water and water infrastructure requirements for systems using CDUs

For small Blue Gene/Q systems, IBM suggests using coolant distribution units (CDUs) to provide the appropriate conditioned water. Suppliers of CDUs include Eaton-Williams and Coolcentric by Vette Corp:

- [http://www.coolcentric.com/](http://www.coolcentric.com/)

CDUs can be customized to work on either a raised floor or non-raised floor environment by moving the connection locations and direction on the CDU. Work directly with the CDU supplier to determine the most appropriate connector locations for Blue Gene/Q. Although the hoses can come out the top or the bottom, Blue Gene/Q does not route the hoses from above the rack down to the floor. See Figure 3-14 on page 45 and Figure 3-15 on page 46 for the hose location on Blue Gene/Q. If using the IBM supplied barb, the CDU requires a 1-inch national pipe thread (NPT) female fitting. See 3.11, “Hose placement and installation” on page 44.

The distance from the CDU to the Blue Gene/Q rack can vary by installation. There are set lengths of hose that exit the Blue Gene/Q rack and a length of hose that can be cut to size from the CDU to the Blue Gene/Q rack. Placement of the CDU must account for the length of hose required to run both horizontally and vertically, including the bend radii, to reach the Blue Gene/Q. See 3.11, “Hose placement and installation” on page 44 for hose information.

### Parameter | Unit | X2L | X2L Action | X2H | X2H action
--- | --- | --- | --- | --- | ---
Return water temperature (TR) | Degrees C | 0 | Nothing | 46 | Nothing
Supply water pressure (pS) | PSI | 15 | Message | 110 | Shut off power and water
Differential water pressure (Δp) | PSI | 7 | Message | 50 | Message
Ambient air temperature (TAIR) | Degrees C | 15 | Message | 46 | Message
Ambient relative humidity (RH) | % | 20 | Nothing | 80 | Nothing
Ambient dew point (DP) | Degrees C | 0 | Message | 17 | Message
System power utilization (Q)a | kW | 0 | Nothing | 50 | Nothing
Flow impedance tolerance (e) | 100 | Nothing | 100 | Nothing

a. Based on water flow and temperature rise
For a Blue Gene/Q solution using CDUs that also uses at least one full rack (two midplanes) and a one half rack (single midplane), butterfly or globe flow valves must be installed. The valves adjust the water flow to the different racks appropriately. The valves are installed on the CDU manifold.

Discuss the water chemistry requirements and options for the Blue Gene/Q system with a vendor that specializes in water systems and treatment. See Chapter 5, “Water treatment specification” on page 67 for details about water requirements.

3.9 Water and water infrastructure requirements for systems using facility water loops

For large installations, water conditioning at the facility level is recommended. Discuss the water chemistry requirements and options for the Blue Gene/Q system with a vendor that specializes in water systems and treatment. See Chapter 5, “Water treatment specification” on page 67 for details about water requirements.

IBM has fully tested and qualified the specific concentrations of Nalco products defined in this document. These products protect the Blue Gene/Q against copper pitting and corrosion. To plan for such a facility-level system, consult Nalco to acquire and implement the correct chemicals to treat your water system (www.nalco.com). For contact information, see Chapter 5, “Water treatment specification” on page 67.

The following items must be in place with the facility loop to install Blue Gene/Q:

- Piping
- Valves
- Filtration
- Pressure regulation and balancing
- Water chemical treatment

3.9.1 Piping

All metal piping must be copper, stainless steel, or polyvinyl chloride (PVC) to prevent scaling. It must be clean to eliminate contaminants, such as pipe dope from the threaded fittings. To avoid scaling and complicated water chemistry, do not use iron or steel piping. Piping must be large enough, as dictated by industry best practices, to avoid excessive water velocity and undue pressure drop.

3.9.2 Valves

For a Blue Gene/Q solution that implements any half rack (single midplane) configuration, butterfly or globe valves must be installed. The valves are used to adjust the water flow rate to the different racks appropriately. Place the valves just after water branches off to the racks from the main line.

3.9.3 Filtration

Forty-mesh, duplex basket strainers, such as those available from Eaton-Hayward, are required for Blue Gene/Q operation:

http://www.eaton.com/Filtration/
3.9.4 Pressure regulation and balancing

Differential pressure is supply minus return. The differential pressure in the main pipes that feed the Blue Gene/Q racks must be maintained by a closed-loop control system to deliver the water flow requirements specified in Figure 3-9 on page 31. Typically, to achieve this goal, the facility pumps are driven by variable frequency drives (VFDs).

If necessary, install balancing valves to eliminate differences in pressure across a large room.

3.9.5 Water chemical treatment

IBM recommends installing a centralized system for water monitoring and control. Contact a water treatment specialist, such as Nalco, to ensure that the correct water treatment system is installed based on site-specific water condition. See “Water treatment specification” on page 67 for information about system-side water requirements. IBM has fully tested and qualified the specific concentrations of Nalco products defined in that section. These products protect the Blue Gene/Q system against copper pitting and corrosion.

3.10 Customer responsibilities for water

The loop that routes water to the Blue Gene/Q racks (either through a CDU or a facility loop) must be a closed loop that is dedicated to the Blue Gene/Q racks. The introduction of other equipment might violate the requirements stated in “Water treatment specification” on page 67. The only exception is for IBM rear-door heat exchangers (including the 19” rack and the iDataPlex rack rear-door heat exchange), which can be on the same loop as the Blue Gene/Q racks.

The Blue Gene/Q water cooling architecture differs significantly from other IBM systems and drives unique water chemistry and water delivery requirements.

3.10.1 Overview

Other IBM water-cooled racks have a water-to-water heat exchanger at the point where the water enters the rack. In these systems, the secondary loop is all contained within the rack. IBM has monitored and done extensive testing on those closed loop solutions supplied to the customer.

The Blue Gene/Q system does not have a water-to-water heat exchanger, such as a modular water-cooling unit (MWU) or a CDU, built inside the rack. Therefore, the Blue Gene/Q system relies on the heat transfer occurring outside the rack through an off-the-shelf CDU or a facility water-to-water heat exchanger. The CDU has several purposes:

► To provide an interface to transfer heat from one water source to another
► To create a separate closed water loop, referred to as the secondary loop, that flows water through the Blue Gene/Q system

3.10.2 Unique Blue Gene/Q Requirements

Because the secondary loop begins outside of the rack, the water that enters the Blue Gene rack is subjected to much more stringent requirements than the water that enters other IBM water-cooled racks. The Blue Gene/Q system design relies on the customer to implement the proper closed loop control using either an off-the-shelf CDU or a custom facility-level closed
loop heat exchanger. IBM cannot test every possible water infrastructure combination. Therefore, a customer-controlled closed secondary loop requires the customer to tightly control the water chemistry, contaminants, particle size, temperature, and flow rate.

It is extremely important that the Blue Gene/Q system secondary water loop is controlled from a chemistry perspective. Failure to follow the directions provided in the “Water treatment specification” can lead to leaks or clogged pathways inside the system. Therefore, the customer is responsible for all of the following items:

- Procuring, correct handling, administering, and disposing of all water and treated water for the Blue Gene/Q system.
- Ensuring that the packaging or form factor of the water and any added chemicals is suited for the customer's facility. Also ensure that the facility can effectively and safely administer the water and any added chemicals for the Blue Gene/Q system.
- Ensuring that the chemical properties of the treated water meet facility requirements and the requirements documented in “Water treatment specification” on page 67.
- Ensuring that material safety data sheets (MSDSs) for the treated water and any added chemicals are obtained from the chemical supplier.
- Ensuring that all MSDS safety requirements are communicated and followed by all personnel who handle or are otherwise exposed to the treated water.
- Communicating the safety requirements for the treated water to IBM service personnel.
- Providing appropriate personal protective equipment to IBM service personnel for use during system maintenance.
- Providing and managing the materials necessary for the containment and cleanup of any water or chemical spills or escapes that might occur during system maintenance.
- Researching and implementing a method for the proper disposal of any water, chemicals, or other materials used in any clean-up activities according to the requirements in their local jurisdiction.

Nalco MSDSs are attainable by contacting Nalco at:

http://www.nalco.com/msds.htm

### 3.11 Hose placement and installation

Two hoses are attached to the Blue Gene/Q to connect it to the supply and return water outside of the rack. They can be connected either under the floor or above the rack, and have a quick connect on the end. The length of hose outside of the rack is shown in Figure 3-14 on page 45 and Figure 3-15 on page 46.
Figure 3-14  Hose return and supply dimensions (mm/in.) for below-floor water drop locations, side view
The locations of the hose exits from the rack are shown in Figure 4-1 on page 51, Figure 4-2 on page 52, and Figure 4-3 on page 53. For installation purposes on raised floors less than 4 feet high, the facility hose connection must be in the aisle closest to the Blue Gene/Q rack water connection. The hose connection cannot be in the aisle opposite to the hose connection on floors less than 4 feet high because of possible limited access issues when feeding the power cord by hand under the raised floor.

A 14-foot hose kit, shown in Figure 3-16 on page 47, connects the Blue Gene/Q to the facility water system. The kit is shipped before the Blue Gene/Q racks are shipped.
These hoses must be in place before the installation of the Blue Gene/Q racks. The facility
end of the provided hoses is a cut hose without a fitting. The customer is responsible for the
assembly of the hose to the water system. Each hose ships with a 1-inch National Pipe
Thread Male (NPTM) barb (44V6056) and clamp (45D5039) that can be used for assembly.
The customer can also provide their own solution. The 14-foot hose kit is available with
feature code 0214 (part number 66Y4590). The kit is shipped per rack and contains two
hoses, two barbs, four clamps, three supply labels, and three return labels. See Figure 3-16.
A request price quotation (RPQ) is available that contains 6-foot hoses. Hoses longer than
14 feet are not recommended because the added length might increase the pressure drop
across the system. Additionally, only use hoses provided by IBM with the Blue Gene/Q
system.

The clamp is a product of Oetiker, part number 16703242. For recommended clamping tools
and specifications, see:

[www.oetiker.com](http://www.oetiker.com)

Use a thread sealant, such as Loctite 554, on the 1-inch NPT threads from the host fitting to
the facility fitting. Threads must be cleaned and dry before applying the thread sealant.
Stainless-steel-to-stainless-steel joints must also use a primer (such as Loctite Activator
7471). Thread sealant must be applied to at least four full threads, leaving the first thread free.
The fitting must be tightened at least ¾ of a turn past hand tight. Teflon tape or other similar
materials must not be used because they can break off and clog components.

**Hose:**

- This hose can be cut to the appropriate length for the facility; however, it must be
  cleaned so no particles are inside before installation.
- A hose clamp must be implemented.
- The hose inside diameter is 25.4 mm ±0.5 mm.
- The hose outside diameter is 34.54 mm ±0.76 mm.
The bend radius of the hose is 6 inches (152.4 mm). Tight bends or kinks in the hose cause cooling problems for Blue Gene/Q systems. Keep some slack in the hoses for easy installation.
Site layout requirements

This chapter covers the site requirements for a Blue Gene/Q installation.
4.1 Raised floor preparations and requirements

A raised floor is not required for Blue Gene/Q or the I/O node. However, without a raised floor, all water, power, and cable lines must be routed overhead. They must be at least 8 feet (2.44 m) from the floor surface to allow clearance for the lift that is used during installation and service actions. See “Alum-a-lift tool” on page 80.

If a raised floor is installed, minimize electrostatic discharge (ESD) by using conductive floor panels. Pedestals must be firmly attached to the structural (concrete) floor using an adhesive and must be attached into the grounding system, along with the stringers and floor tiles. Details about ESD are in the IBM System Information Center article, “Static Electricity and Floor Resistance”:


Figure 4-1 on page 51, Figure 4-2 on page 52, and Figure 4-3 on page 53 show a top view of floor tile cutouts. Several of the cutouts span two tiles. Because of this setup, additional pedestals are probably required to support the Blue Gene/Q load on the floor. It is the customer’s responsibility to verify that the raised floor meets the requirements to hold the Blue Gene/Q system. If the raised floor uses a stringer system, the stringer must be left intact. Cables, and so forth, are routed around the stringer. Additionally, the cutouts can be enlarged outward to the edge of the tile, removing the thin section of tile between the cutout and tile edge. Hose cutouts denote the back of the rack. Racks are placed within rows front to back (not front to front).

**Note:** Cutouts in the raised floor tiles often have sharp edges. Treat the edges according to local building and electrical codes. Eliminating sharp edges ensures a safe environment for personnel and minimizes abrasion to cables that are run through an opening.

Figure 4-1 on page 51, Figure 4-2 on page 52, and Figure 4-3 on page 53 show the approximate location of the rack levelers or eyebolts (plus sign with circle). The location is subject to change based on the position of the rack on the floor.
Figure 4-1  Floor tile cutouts, top view, 24-inch floor tile, 7-foot pitch row-to-row (mm [top] and in. [bottom]), top view
Figure 4-2  Floor tile cutouts, top view, 24-inch tile, 8-foot pitch row-to-row (mm [top] and in. [bottom])
Notes about Figure 4-3:

1. One leveler from each rack extends onto a tile in the aisle. Access to the floor area under this tile is restricted.

2. The cutout pitch along a Blue Gene/Q row is constant at 48 inches (1219.2 mm). For 600 mm tiles only, the cutouts shift down the row.

3. Rack tie down bolts can be installed by the customer and used to secure the rack to the concrete subfloor.

Rack tie down bolts can be installed by the customer and used to secure the rack to the concrete subfloor. The rack is tied down through the leveler foot using a through hole. See Figure 4-4 on page 54 for the hardware dimensions and location. A floor clearance hole at least 6.35 mm (0.25 inches) larger than the bolt diameter is recommended for positional tolerance. Also for positional tolerance, rows with four or more racks must slot the tolerance hole along the row 38 mm (1.5 inches).
4.2 Grounding

IBM equipment, unless double insulated, has power cords containing an insulated grounding conductor. The conductor, color-coded green or green with yellow stripe, connects the frame of the equipment to the ground terminal at the power receptacle.

The power receptacles for IBM equipment are identified in the equipment documentation and must match the equipment power plug. In some cases, there might be options for different manufacturer equivalent receptacles. IBM equipment plugs must not be modified or replaced to allow connection of the power cord to existing connectors or receptacles. To do so might create a safety hazard and void the product warranty.

The connectors or receptacles for IBM equipment must be installed to a branch circuit with an equipment grounding conductor. The grounding conductor is connected to the grounding bus bar in the branch-circuit distribution panel. The grounding bus bar in the panel is then connected back to the service entrance or a suitable building ground by an equipment grounding conductor.

Information technology equipment must be properly grounded. It is recommended that an insulated green wire ground the same size as the phase wire be installed between the branch circuit panel and the receptacle.

For personnel safety, the ground must have sufficiently low impedance to limit the voltage to ground and to facilitate the operation of protective devices in the circuit. For example, the ground path must not exceed 1 Ω for 120-volt, 20-ampere branch circuit devices.

To ensure a good ground and reasonable efficiency of operation, per National Electrical Code (NEC) 210.19(A)(1), FPN No 4, the maximum voltage drop from the branch panel to the outlet being used by the Blue Gene/Q rack must not exceed 5%.

All grounds entering the room must be interconnected somewhere within the building to provide a common ground potential. This includes any separate power sources, lighting and convenience outlets, and other grounded objects, such as building steel, plumbing, and ducts.
work. If using a transient voltage surge suppressor (TVSS), follow the requirements of NEC article 285.

The equipment grounding conductor must be electrically attached to both the enclosure of the computer power center and the connector grounding terminal. Conduit must not be used as the only grounding means, and it must be connected in parallel with any grounding conductors it contains.

### 4.2.1 Verifying the grounding impedance and the power source

The following procedure confirms that the correct facilities power requirements are met for the new system being installed. This procedure verifies that the power supplied to the receptacles and system line cords is correct, in terms of power needs and safety. This is a recommendation from IBM only and must be performed by a qualified electrician. The customer can conduct any additional testing they see fit to verify their power source.

Within this procedure, you will:

- Confirm that the receptacle is wired safely and correctly.
- Verify that the expected voltage is present at all receptacles to be used for this system.

This procedure is performed at the receptacle that supplies power to the Blue Gene/Q.

**Caution:** Do not touch the receptacle or the receptacle faceplate with anything other than test probes before you have determined that the receptacles meet the tested requirements.

**Caution:** All measurements are to be made with the receptacle faceplate properly installed.

Use the following steps to verify that the grounding impedance is correct:

**Caution:** The following steps are to be performed with the facilities branch circuit breakers turned off and with lock-out tags attached to the circuit breakers.

1. Locate and turn off the branch circuit breakers for each affected ac outlet. Attach lockout tag S229-0237, or equivalent, which reads “Do Not Operate” to each circuit breaker.

2. Safety check for receptacles with a *metal* housing or case: Some receptacles have metal housings. For receptacles of this type, perform the following steps using a multimeter (DVM). Ensure that the probes are in contact with metal if the receptacle housing is painted:
   a. Check for less than 1 V ac from the receptacle housing to any grounded metal structure in the building. The grounded structure can be a raised floor metal structure, a water pipe (best), building steel, or a similar structure.
   b. Check for less than 1 V ac from each of the four pins to any grounded metal structure in the building. The grounded structure can be a raised floor metal structure, a water pipe (best), building steel, or a similar structure.
3. Safety check for receptacles with a plastic housing/case. If the receptacle has a non-metallic (plastic) housing but is connected to a rigid or flexible metallic conduit, perform the following steps using a multimeter (DVM):
   a. Check for less than 1 V ac from the metallic conduit to any grounded metal structure in the building. The grounded structure can be a raised floor metal structure, a water pipe (best), building steel, or a similar structure.
   b. Check for less than 1 V ac from each of the four pins to any grounded metal structure in the building. The grounded structure can be a raised floor metal structure, a water pipe (best), building steel, or a similar structure.

In the following steps, if the measured impedance is greater than 1 Ω and the test instrument used was a digital multi-meter, verify that the grounding impedance is correct. Use the ECOS Electronics Corporation 1020, 1023, B7106, or another approved ground impedance tester. 

*If the following resistance reading is greater than 1 Ω, do not proceed!* Make necessary wiring corrections before continuing.

4. Check the resistance from the ground pin of the outlet receptacle to the receptacle case. Check resistance from the ground pin to building ground. The reading must be less than 1.0 Ω. This reading indicates the presence of a continuous grounding conductor.

5. Check for infinite resistance between the three power pins. Each pin is associated to one of the three phases. Make the measurements between pins 1 to 2, 2 to 3, and 1 to 3. This is a check for a wiring short.

*If the resistance reading is other than infinity, do not proceed!* Make necessary wiring corrections before continuing.

Do not turn on the branch circuit breaker until all the previous steps are satisfactorily completed for each receptacle.

6. Repeat checks 1 through 6 for the remaining ac power receptacles.

7. If all of the receptacles passed the previous tests, *the qualified person who put on the lockout tags* can remove the lock-out tags (S229-0237) or equivalent that state Do Not Operate. Then turn on the branch circuit breakers.

In the following steps, you measure for appropriate voltages between the power (phase) pins:

1. With a digital multimeter and probes, verify that the voltage at the receptacle is correct. Make your measurements between pins 1 to 2, 2 to 3, and 1 to 3:
   a. For 380 - 415 volts alternating current (VAC), the acceptable phase-to-phase voltage measurement range is 342 - 440 VAC.
   b. For 480 VAC, the acceptable phase-to-phase voltage measurement range is 432 - 509 VAC.

2. Repeat the previous voltage checks for the remaining ac power receptacles.

### 4.3 Cable trays and cabling

Figure 4-5 on page 57 shows how cables are routed within the Blue Gene/Q system.
Each compute row needs two cable trays running from one end of the row to the other end of the row. Each column needs one tray per column, plus an additional tray at the end of the row.

For ease of installation, a solid cable tray (versus a wire mesh cable tray) is requested, but not required. Do not use interconnected systems. Do not install any covers or tops on the trays. Also, no spillways are required between trays. Whether cabling is above or below the floor has no impact on inter-rack cable trays as shown in Figure 4-5.

Fiber optic cables must not be pressed directly around a 90 degree corner. They cannot be pulled tight against any edge. As shown in Figure 4-6, the fiber optic bend radius must be 2 in. (4 in. diameter) to maintain signal integrity.
The maximum cable tray distances are:

- **Below-floor cable trays**, see Figure 4-9 on page 59:
  - Along the row, the bottom of the cable tray is no more than 2 feet below the top of the raised floor.
  - Along the column, the bottom of the cable tray is no more than 1.25 feet below the top of the raised floor.

- **Above-floor cable trays**: All trays must be between 8 - 11.5 feet above the floor.

### 4.3.1 Below-floor cabling

Figure 4-7 and Figure 4-8 on page 59 show below-floor cable trays. The cable trays along the Blue Gene/Q columns are positioned above the trays for the row, at the dimensions listed in Figure 4-9 on page 59. This cable tray system is designed around cable length. The cable tray design for a facility might need to be adjusted slightly if water pipes are already installed. However, first review with IBM any major changes to the cable tray layout that increases the amount of cable required. Lowering cable trays further below the raised floor is an example of such a change.

![Figure 4-7: Below-floor cable trays, side view](image-url)

**Figure 4-7** Below-floor cable trays, side view
### 4.3.2 Above-floor cabling

Place above-floor cable trays in locations equivalent to below floor cable trays. If cables are routed from the cable trays through ducts to the system, the ducts must stop 82.5” (210 cm) above the floor, which is the height of the Blue Gene/Q rack.

In Figure 4-10 on page 60, the red hash areas show the top locations where cables exit the rack. In Figure 4-11 on page 61, the red hash areas show the side locations where cables exit the rack.
The red hash areas show the cable top exit locations.

Figure 4-10  Above-floor cable trays, top-down view
4.3.3 Cable tray size calculation

The required cable tray dimensions vary with the number of Blue Gene/Q racks in the system. Switches must be placed within two tiles of the 19-inch I/O racks to avoid using compute rack cable trays. Otherwise, additional or larger cable trays might be needed.

Determining the size of the cable trays is a multiple step process. Calculations are completed as a cross sectional area of the tray. The tray dimensions that meet the cross sectional area can be determined by the customer:

1. Determine the initial amount of cross sectional area needed for the Blue Gene/Q central electronics complex racks (48 fiber optic cables) from Table 4-1 on page 62. Determine a separate value for both row and column cable trays.

2. For I/O drawers on the central electronics complex rack, add additional area to the trays to account for the PCI cables coming from I/O drawers mounted on top of the Blue Gene/Q central electronics complex racks. Each I/O drawer requires 1 in.\(^2\) of cable tray space from the I/O drawer to its switch. Depending on the switch location and the path the cables take to the drawer, the adder for I/O drawers might need to be added to both row and column cable trays. Add additional area for any other non Blue Gene/Q cables routed through these trays.
3. For I/O drawers in the I/O rack, add additional area to the trays to account for the optical cables coming from the I/O drawers at the end of the racks. Each I/O drawer requires 0.5 in.\(^2\) of cable tray for each I/O drawer at the end of a row.

4. Multiply the entire value by 2.5 to account for installation space and future service needs. Round up to the nearest appropriate cable tray size available.

A general formula for calculating the cable tray size for a given area is:

- **Using A:** Total cable-tray cross-sectional area = (value from Table 4-1 + (number of I/O drawers \(\times\) 1 in.\(^2\) per I/O drawer)) \(\times\) 2.5

- **Using B:** Total cable-tray cross-sectional area = (value from Table 4-1 + (number of I/O drawers \(\times\) 0.5 in\(^2\) per I/O drawer)) \(\times\) 2.5

<table>
<thead>
<tr>
<th>Rows</th>
<th>Columns</th>
<th>Row cable-tray cross-sectional area (in(^2))</th>
<th>Column cable-tray cross-sectional area (in(^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>
Example cable tray size calculation

Figure 4-12 is the example used in the calculation.

To calculate the cable tray size:

1. Go to Table 4-1 on page 62 to find the size required for a 2-row by 4-column layout of central electronics complex cables:
   - For the rows: 5 in²
   - For the columns: 5 in²

2. A) Additional tray space is needed for the I/O drawer in a compute rack, 1 in² for each I/O drawer, or 1 in² in this example.
   B) N/A for this example

3. Multiply the area required by central electronics complex cables plus the area required by I/O drawer cables by 2.5:
   - ROW:
     - = \text{(value from Table 4-1 on page 62 + (number of I/O drawers } \times 1 \text{ in/sq. per I/O drawer))} \times 2.5
Again, if the system has the I/O drawers at the end of the rows in racks, whatever the total cable tray area required, the cable tray width must be at least 5 inches.

\[
\text{COLUMN:} \quad = (\text{value from Table 4-1 on page 62} + (\text{number of I/O drawers} \times 1 \text{ in/sq. per I/O drawer})) \times 2.5
\]

\[
= (5 + 1 \times 1) \times 2.5 \\
= 15 \text{ in}^2
\]

These values are rounded up to the nearest available and appropriate cable tray size.

Similar layouts are needed for above-floor cable trays.

## 4.4 Particulate contamination and air quality

Systems are being installed in increasingly diversified industries. Some industries, as a by-product of their processes, cause the atmosphere to contain measurable quantities of gases and solid particles that are potentially harmful to electronic equipment. Urban areas that are highly industrialized might have levels of gases and solid particles that cause an unacceptable environment exposure to exist throughout an entire area.

IBM is concerned with two classes of atmospheric contaminants: solid particles and gases. Solid particles in the air are referred to as particulates. Water vapor can combine with these tiny, solid particles and form compounds. Such matter is said to be hygroscopic. It can be harmful, depending on the particulate composition. Gases can form harmful acids or bases when combined with treated water. Because of the ability to absorb moisture, the relative humidity and temperature are significant factors in an unacceptable environment.

High concentrations of gases, such as the ones in the following list, are associated with industrial processes and are known to cause corrosion and failure of electronic components:

- Sulfur dioxide
- Nitrogen dioxide
- Ozone
- Acidic gaseous chlorine

In addition to gases, some industrial processes produce particulate contamination. These particles can settle (in the form of dust) in surrounding areas even though the process producing the particles might be some distance away.

Industries engaged in processing petroleum, chemicals, primary metals, food, mining, and paper have a higher probability of encountering an unacceptable environment. However, contamination can be a result of construction, cleaning, or other activities that can occur anywhere.

Airborne particulates, including metal flakes or particles, might pose a risk to the server. In addition, reactive gases acting alone or in combination with other environmental factors, such as humidity or temperature, might pose a risk to the server. Risks that are posed by the presence of excessive particulate levels or concentrations of harmful gases include damage that might cause the server to malfunction or cease functioning altogether. The environmental specifications define limits for particulates and gases that are intended to avoid such damage. The limits must not be viewed or used as definitive limits. Numerous other factors, such as
temperature or moisture content of the air, can influence the impact of particulates or environmental corrosives and gaseous contaminant transfer. In the absence of specific limits that are defined in the environmental specifications, you must implement practices that maintain particulate or gas levels that are consistent with the protection of human health and safety. If IBM determines that the levels of particulates or gases in your environment damaged the server, IBM might condition provision of repair or replacement of servers or parts on the implementation of appropriate remedial measures to mitigate such environmental contamination. Implementation of such remedial measures is a customer responsibility.

The operating environment requirements are:

- **Gaseous contamination**: Severity level G1 as per ANSI/ISA 71.04-1985; see note 1. The reactivity rate of copper coupons must be less than 300 Angstroms per month (Å/month, ≈ 0.0039 µg/cm²-hour weight gain). See note 2. In addition, the reactivity rate of silver coupons must be less than 300Å/month (≈ 0.0035 µg/cm²-hour weight gain). See note 3. The reactive monitoring of gaseous corrosiveness must be conducted approximately 5 cm (2 in.) in front of the rack on the air inlet side at one-quarter and three-quarter frame height off the floor or where the air velocity is much higher.

- **Particulate contamination**: Data centers must meet the cleanliness level of International Organization for Standardization (ISO) 14644-1 class 8. For data centers without airside economizer, the ISO 14644-1 class 8 cleanliness might be met by the choice of the following filtration:
  - The room air can be continuously filtered with minimum efficiency reporting value (MERV) 8 filters.
  - Air entering a data center can be filtered with MERV 11 or preferably MERV 13 filters.

For data centers with airside economizers, the choice of filters to achieve ISO class 8 cleanliness depends on the specific conditions present at that data center.

The deliquescent relative humidity of the particulate contamination must be more than 60% RH3.

Data centers must be free of zinc whiskers. See note 4.

**Notes:**


2. The derivation of the equivalence between the rate of copper corrosion product thickness growth in Å/month and the rate of weight gain assumes that Cu2S and Cu2O grow in equal proportions.

3. The derivation of the equivalence between the rate of silver corrosion product thickness growth in Å/month and the rate of weight gain assumes that Ag2S is the only corrosion product.

4. Surface debris is randomly collected from 10 areas of the data center on a 1.5 cm diameter disk of sticky electrically conductive tape on a metal stub. If examination of the sticky tape in a scanning electron microscope reveals no zinc whiskers, the data center is considered free of zinc whiskers.
4.5 Homologation statement

This product is not intended to be connected directly or indirectly by any means whatsoever to interfaces of public telecommunications networks.
This chapter describes the specific water treatment requirements for the system-side cooling loop that carries water to cool the electronics in the Blue Gene/Q supercomputer. The ongoing monitoring and maintenance are also specified.

The system-side cooling loop hardware consists mainly of corrosion-resistant alloys of copper and stainless steels. Ethylene propylene diene monomer (EPDM) rubber forms the inner lining of all the hoses in the system. The chemistry of the cooling water must be properly maintained to avoid system disruption or shutdown due to any of four common cooling water-related problems: corrosion, microbiological growth, scale formation, and fouling.

The details of the system-side water treatment depend on whether the local municipality allows the disposal of water that contains some cleaning chemicals down a sanitary drain. If the local municipality does not allow the disposal of contaminated water down a sanitary drain, a de-ionizing bypass can be included in the system-side loop. The bypass enables cleaning the water to purity levels corresponding to resistivity greater than 1 MΩ.cm before pouring the water down the drain. The customer is responsible for verifying the local regulations before disposing of any water.
5.1 Water-related problems

Correct treatment of the system-side water is necessary to avoid the following four major water-related problems: corrosion, microbiological growth, scale formation, and fouling. Any of these problems can significantly reduce cooling efficiencies and increase risk of system downtime.

5.1.1 Corrosion

Corrosion can take many forms. The common forms of corrosion relevant to the system-side cooling loop are pitting corrosion and galvanic corrosion.

Pitting corrosion is the localized attack of a metal surface. In copper tubes, shown in Figure 5-1, pitting can lead to water leaks with a typical mean time to failure of about two years.

Galvanic corrosion arises when two metals that are wide apart in the galvanic series are in electrical contact and immersed in the same water environment. The potential difference that arises between the two metals in contact forces electrons to flow from the less noble to the more noble metal. On the less noble metal surface, the metal corrodes. The corrosion gives off electrons that are consumed on the more noble metal surface by a reduction reaction that can take many chemical forms. Examples are the reduction of metal ions or the consumption of oxygen and water to form hydroxyl ions. Even when not in electrical contact, aluminum can be galvanically attacked by copper. The reaction occurs because dissolved copper ions in low concentrations are deposited on the aluminum surface, thus forming the galvanic corrosion couple.

5.1.2 Scaling

Scaling is the deposition of dense, adherent material on the cooling loop surfaces. Scaling occurs when the solubilities of salts in water are exceeded because of high concentrations or because of increased temperature.
5.1.3 Fouling

Fouling of cooling loops is the deposition of non-scale-forming substances, such as corrosion products and organics.

5.1.4 Microbiological growth

Microbiological growth in water systems can lead to deposition, fouling, and corrosion within the cooling loop. Prevention of microbiological growth involves:

- Making sure that the cooling loop hardware is assembled from components that are free of biological organisms.
- Treatment with biocides to control the bacteria population.

To avoid biological growth, the water cooling loops must be shipped and stored dry. Every effort must be made to blow out the water and dry the water cooling loop as much as possible before shipping and storage.

As shown in Figure 5-2, microbiological issues can develop a biofilm, which can foul low-flow areas and create accelerated corrosion where biofilm comes in contact with surface metal.

![Figure 5-2 Accelerated corrosion](image)

5.2 Avoiding water-related problems

The following best practices are recommended to avoid water-related problems:

- Design clean: Restrict the water-wetted metallurgies to copper alloys and stainless steels. Avoid using steel hardware that can rust and foul the water cooling loop.
- Build clean: Ensure that the cooling loop components are clean and free of bacteria. The cooling loop assembly must be free of soldering or brazing fluxes. Clean water must be used in the assembly operations. Any residual water must be blown out of the assembly. The finished assembly must be clean and dry.
- Ship clean: Any residual water from the assembly or testing operations must be blown out from the cooling loop before shipping to avoid corrosion and microbiological growth. As a final step, use nitrogen gas to dry the system. Plug ends and ship the system with the cooling loop pressurized with nitrogen gas.
Install clean: The cooling loop must be kept clean during the installation step. Brazing is preferred over soldering. All flux residues must be cleaned off. Fill the system with clean water. If possible, include a secondary step to de-ionize the water in the cooling loop before the addition of biocide and corrosion inhibitors.

Maintain clean: Monitor and maintain the pH, the bacteria count, and the corrosion inhibitor concentration.

5.3 Water quality chemical requirements

For reference, Table 5-1 lists the parts per million (concentrations) permitted for certain chemicals and minerals. Make the following measurements before adding any chemicals to the water loop. It is the starting point that defines a clean water base.

<table>
<thead>
<tr>
<th>Chemical/mineral</th>
<th>Maximum value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All metals</td>
<td>( \leq 0.10 ) ppm</td>
</tr>
<tr>
<td>Calcium (CaCO3)</td>
<td>( \leq 1.0 ) ppm</td>
</tr>
<tr>
<td>Magnesium (CaCO3)</td>
<td>( \leq 1.0 ) ppm</td>
</tr>
<tr>
<td>Manganese (Mn)</td>
<td>( \leq 0.10 ) ppm</td>
</tr>
<tr>
<td>Phosphorus (P)</td>
<td>( \leq 0.50 ) ppm</td>
</tr>
<tr>
<td>Silica (SiO2)</td>
<td>( \leq 1.0 ) ppm</td>
</tr>
<tr>
<td>Sodium (Na)</td>
<td>( \leq 0.10 ) ppm</td>
</tr>
<tr>
<td>Bromide (Br)</td>
<td>( \leq 0.10 ) ppm</td>
</tr>
<tr>
<td>Nitrate (NO2)</td>
<td>( \leq 0.50 ) ppm</td>
</tr>
<tr>
<td>Chloride (CaCO3)</td>
<td>( \leq 0.50 ) ppm</td>
</tr>
<tr>
<td>Nitrate (CaCO3)</td>
<td>( \leq 0.50 ) ppm</td>
</tr>
<tr>
<td>Sulfate (CaCO3)</td>
<td>( \leq 0.50 ) ppm</td>
</tr>
<tr>
<td>Conductivity (( \mu )S)</td>
<td>( \leq 1.0 ) ( \mu )S/cm (^a)</td>
</tr>
<tr>
<td>pH</td>
<td>6.5 - 8.0</td>
</tr>
<tr>
<td>Turbidity (nephelometric turbidity units [NTU])</td>
<td>( \leq 1.0 ) ppm</td>
</tr>
</tbody>
</table>

\(^a\) Measure conductivity at 20 - 25°C. Conductivity increases ~5% for every degree C rise in temperature

5.4 Plumbing material requirements

All piping must be composed of specific materials to prevent scaling and improper reactions with the chemistry of the water within the system. Do not seal any threaded joints with teflon tape because particles from the tape can enter the water stream and create clogs. Instead, use a thread sealant, such as Loctite, to seal threaded fittings. Piping must be large enough, as dictated by industry best practices, to avoid excessive water velocity and undue pressure drop.
Material selection and installation are complex issues governed by building codes and other local requirements. Authorities, such as building inspectors, fire departments, insurance providers, and code compliance officers, might have jurisdiction. The customer is encouraged to investigate with appropriate authorities before planning and installing cooling distribution systems. The following information is provided for chemical-compatibility purposes only.

The following alloys must be avoided in the plumbing system:

- Aluminum and aluminum alloys.
- Brass with more than 15% zinc.
- Free-cutting brasses, especially the ones that contain lead. One example of such a brass is copper alloy C36000, called free-cutting yellow brass.
- Leaded brasses are especially a concern when subjected to high tensile stress.
- Steels that are not stainless steel.
- Stainless steels that are not properly solution treated.

The following materials are preferred:

- Copper alloys:
  - Lead-free copper alloys with less than 15% zinc.
- Stainless steels:
  - Low carbon stainless steels are preferred.
  - Must be solution treated.
  - Must be passivated.
  - Avoid sensitization during welding.
  - Avoid brazing; welding is preferred.
- Polyvinyl chloride (PVC). Not allowed inside IBM products due to flammability concerns, but can be used at a facility level. Consult the appropriate authorities that have jurisdiction.
- EPDM rubber is the preferred material for hoses:
  - Flammability rating must be Canadian Standards Association (CSA) or Underwriters Laboratory (UL) VW-1 or better.
  - Peroxide-cured hoses are preferred.

Metal joining operations:

- Avoid solder joints that come in contact with water. Solder joints are porous; they leach flux residue into the cooling loop. Solder joints might pass inspection and pressure test as manufactured, but still be unreliable.
- Brazed joints are preferred for joining copper pieces.
- Do not use braze joints to join stainless steels. Tungsten inert gas (TIG) and metal inert gas (MIG) welding are preferred for joining stainless steels. Avoid sensitization. The welded assembly must be passivated.

### 5.5 Water quality requirements

The water required to initially fill the system-side cooling loop must be reasonably clean, bacteria-free water that contains less than 100 colony-forming units per milliliter (CFU/mL). For example, the following types of water meet the criteria: de-mineralized water, reverse osmosis water, de-ionized water, or distilled water.
The water must be filtered with in-line 40-mesh or finer duplex basket strainers. Redundant strainers are preferred so that they can be concurrently maintained during the operation of the system.

Figure 5-3 shows an optional method of obtaining de-ionized water in the cooling loop. It is mainly used in large cooling loops. If this method is used, pass the water through the de-ionizer before any racks are connected to the water loop.

**Important:** Ensure that the system begins with clean water by de-ionizing the water using de-ionizing cartridges installed in the cooling loop, as shown in the Figure 5-3. Even if de-ionized water was used to fill the system, a de-ionizing step is prudent for two reasons. First, it ensures that the starting water is de-ionized. Second, it removes any ions that might have leached off the walls of the cooling loop.

When the water needs to be de-ionized, valves V2 and V3 can be opened and valve V1 partially closed to bypass some of the water through the de-ionizing canister. During this de-ionizing step, the cooling loop and the computers can keep operating normally. (This process is only applicable for refreshing the water loop as described in 5.12, “Refreshing the system-side water” on page 77.) When the de-ionization is complete, the V2 and V3 valves are closed and V1 fully opened.

The de-ionization step raises the resistivity of the water greater than 1 MΩ.cm. Under normal operation, the V2 and V3 valves are closed and V1 valve is fully open.

![Figure 5-3   De-ionizing water in the cooling loop](image)

### 5.6 Water dosing and testing equipment

The following equipment is recommended:

- **De-ionizing equipment:** Optional, mainly used in large cooling loops:
  - Canister, with de-ionizing cartridge, through which some of the water flow can bypass when the water must be de-ionized. See 5.3, “Water quality chemical requirements” on page 70.

- **Dosing equipment.** The following equipment is used to feed the appropriate doses of chemicals into a Blue Gene/Q system:
  - Recommend using a stainless steel or fiberglass chemical shot feeder:
    - System volumes less than 100 gallons, use a 0.1-gallon size feeder.
    - System volumes less than 1000 gallons, use a 1-gallon size feeder.
    - System volumes greater than 1000 gallons, use a 2.5-gallon feeder.
  - Chemical pump as per Nalco or another water-treatment contractor specification.
– 3D TRASAR® controller (#060-TR5500.88) for systems larger than 250 gallons to enable precise and continuous monitoring of system water chemistries: conductivity, pH, corrosion rates, turbidity.

- Testing equipment. The equipment listed in Table 5-2 is recommended.

### Table 5-2  Recommended testing equipment

<table>
<thead>
<tr>
<th>Component</th>
<th>Part number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azole test kit</td>
<td>Nalco 460-P3119.88</td>
<td>Triazole Reagent Set, 25 mL, 100 tests</td>
</tr>
<tr>
<td></td>
<td>Nalco 500-P2553.88</td>
<td>UV lamp with power supply, 115 VAC</td>
</tr>
<tr>
<td></td>
<td>Nalco 400-P0890.88</td>
<td>Nalco DR/890 Colorimeter</td>
</tr>
<tr>
<td></td>
<td>Nalco 500-P1204.88</td>
<td>25 mL graduated cylinder</td>
</tr>
<tr>
<td>Nalco bacteria test kit</td>
<td>Nalco 500-P3054.88</td>
<td>Bacteria dip slides</td>
</tr>
<tr>
<td>Water resistivity monitor</td>
<td>Nalco 400-C006P.88</td>
<td>With 0 - 10 MΩ.cm range</td>
</tr>
</tbody>
</table>

### 5.7 Predelivery water treatment protocol

If any plumbing unit had water passed through it before shipping, whether the water was treated or not, follow the steps in this section to eliminate the potential for microbiological fouling, general fouling, or corrosion during the shipping and storage phase.

Use nitrogen to purge any residual water out of the system and provide a nitrogen blanket for shipping. Avoid use of compressed air for shipment because compressed air can lead to oxygen-based corrosion issues. If the supply of nitrogen is limited, the residual water can be blown out first by using compressed air. Follow this step with a thorough purging with nitrogen gas.

### 5.8 Materials required for site system startup

The following items must be available to correctly and safely complete the initial system startup:

- De-ionizing cartridges of the appropriate capacity (optional).
- Nalco treatment chemicals in appropriate quantities:
  - Systems with 20 gallons or less coolant: Suggest using a prepackaged cleaner and inhibitor solution: Nalco 460-CCL2567 or Nalco CCL2567 and Nalco 460-CCL100 or Nalco CCL100.
- A method to add chemicals. Use installed system chemical shot feeder or an appropriately sized chemical feed pump.
- Source of demineralized water, reverse osmosis water, de-ionized water, or distilled water.
- Correct personal protective equipment.
5.9 Initial site water system startup

Requirements follow for the treatment procedures for small (20 gallons or less) and large (20 gallons and more) Blue Gene/Q systems.

5.9.1 Initial treatment procedure for systems of 20 gallons or less

This section contains cleaning and dosing procedures.

**Cleaning procedure**

Perform this cleaning procedure on the cooling loop before any Blue Gene/Q racks are hooked up to the system:

1. The system must be empty. If not, drain the system completely.
2. Remove all mesh filters from filter housings.
3. Ensure that bypass hoses are connected between the supply and return portions of the cooling loop to ensure the cleaning of all areas of flow.
4. One of the two following cleansing procedures can be used:
   - Chemical cleaning: This process is the most effective way to clean the plumbing loop:
     i. Add the required volume of cleaning solution (recommended Nalco 460-CCL2567 or Nalco CCL2567 at 100% concentration).
     ii. Circulate the cleaning solution for a minimum of 30 minutes (longer if time permits) to ensure that the cleaner reaches all areas of flow.
     iii. Drain the system completely, disposing of the cleaner per local regulations.
     iv. Refill with demineralized water, reverse osmosis water, de-ionized water, or distilled water.
     v. Circulate the water for 15 minutes.
     vi. Drain the system completely, disposing of the cleaner per local regulations.
     vii. Immediately proceed to the inhibitor dosing procedure.
   - De-ionized water cleaning: This process can be used if the cleaning chemical cannot be obtained or if local laws prevent disposal of the chemicals:
     i. Completely fill the system with demineralized water, reverse osmosis water, de-ionized water, or distilled water.
     ii. De-ionize the water by bypassing some of the water flow through the de-ionizing cartridges. Circulate the water normally through the complete system until the resistivity of the water increases above 1 MΩ.cm.
     iii. Proceed to the inhibitor dosing procedure.
Inhibitor dosing procedure
To use the inhibitor dosing procedure:
1. Install new or cleaned mesh filter cartridges in the filter housings.
2. There are two dosing procedures:
   a. If the system was cleaned using cleaning solution Nalco 460-CCL2567 or Nalco CCL2567, proceed as follows:
      i. Fill the coolant reservoir with Nalco 460-CCL100 or Nalco CCL100.
      ii. The system is ready for use.
   b. If the system was cleaned using de-ionized water only, dose as follows:
      i. Inject 100 ppm of Nalco H-550 biocide.
      ii. Inject 120 ppm of Nalco 3DT-199 to achieve 40 ppm of azole concentration.
      iii. Confirm the azole residual using the Nalco azole test kit. See Table 5-2 on page 73.
      iv. The system is ready for use.

5.9.2 Initial treatment procedure for systems of greater than 20 gallons

This section contains cleaning and dosing procedures.

Cleaning procedure
Perform this cleaning procedure on the cooling loop before any Blue Gene/Q racks are hooked up to the system:
1. The system must be empty. If not, drain the system completely.
2. Remove all mesh filters from the filter housings.
3. Ensure that bypass hoses are connected between the supply and return portions of the cooling loop to ensure the cleaning of all areas of flow.
4. One of the two following cleansing procedures can be used:
   – Chemical cleaning: This process is the most effective way to clean the plumbing loop:
     i. Fill the system with demineralized water, reverse osmosis water, de-ionized water, or distilled water.
     ii. Add the required volume of cleaning solution (recommended Nalco 2567 at 1% concentration).
     iii. Circulate the cleaning solution for a minimum of four hours.
     iv. Drain the system completely using all drain ports available, disposing of the cleaner per local regulations.
     v. Refill with demineralized water, reverse osmosis water, de-ionized water, or distilled water.
     vi. Circulate the water for one hour.
     vii. Drain the system completely using all drain ports available, disposing of the cleaner per local regulations.
     viii. Refill with demineralized water, reverse osmosis water, de-ionized water, or distilled water.
     ix. Circulate for 15 minutes.
     x. Immediately proceed to the inhibitor dosing procedure.
– De-ionized water cleaning: This process can be used if the cleaning chemical cannot be obtained or if local laws prevent disposal of the chemicals:
  i. Completely fill the system with demineralized water, reverse osmosis water, de-ionized water, or distilled water.
  ii. De-ionize the water by bypassing some of the water flow through the de-ionizing cartridges. Circulate the water normally through the complete system until the resistivity of the water increases above 1MΩ.cm.
  iii. Proceed to the inhibitor dosing procedure.

Inhibitor dosing procedure
To use the inhibitor dosing procedure:
1. Install new or cleaned mesh filter cartridges in the filter housings.
2. The dosing procedure for systems of greater than 20 gallons is the same regardless of the cleaning technique:
   b. Circulate for 30 minutes.
   c. Add 120 ppm of Nalco 3DT-199 to achieve 40 ppm of azole concentration.
   d. Circulate for 30 minutes.
   e. Confirm the azole residual using the Nalco azole test kit. See Table 5-2 on page 73.
   f. The system is ready for use.

5.10 Water system monitoring and maintenance

It is important to conduct a bacteria test on a quarterly basis. Add 100 ppm of Nalco H-550 biocide if the bacteria count is greater than 1000 CFU/mL.

On large systems that have more than 250 gallons of water, install a Nalco 3D TRASAR controller on the system cooling loop.

It is important to conduct an azole test on an annual basis. Add Nalco 3DT-199 to bring the azole concentration to the required 40 ppm level.

5.11 Adding additional racks to an existing system

Racks arrive from IBM ready for installation. To install additional racks:
1. Install the racks, and open the flow to the existing system.
2. Be sure that the automated water make-up on the chiller coolant reservoir is on. If there is not an automated water make-up feature, top off the system-side reservoir.
3. Within two hours of installing the new racks, add 100 ppm of Nalco H-550 biocide.
4. Add 120 ppm of Nalco 3DT-199 to achieve 40 ppm of azole concentration.
5. Confirm the azole residual using the Nalco azole test kit. See Table 5-2 on page 73.
5.12 Refreshing the system-side water

A situation might arise where the water must be refreshed. That is, all the contamination in the water must be removed and biocide and corrosion inhibitor re-added. To refresh the water, follow one of these two procedures:

1. If no water can go down the sanitary sewer:
   a. Insert new deionization cartridges into the canisters. Bypass some of the water through the de-ionizing cartridges until the resistivity of the water increases above 1 MΩ.cm. See “Water quality chemical requirements” on page 70. During this period, the computers and the cooling system can be left on and are fully functional.
   b. Stop bypassing the water flow through the de-ionizing filter, and add new filter cartridges in the filter housings, if applicable.
   c. Inject 100 ppm of Nalco H-550 biocide.
   d. Circulate for 30 minutes.
   e. Inject 120 ppm of Nalco 3DT-199 to achieve 40 ppm of azole concentration.
   f. Circulate for 30 minutes.
   g. Confirm the azole residual using the Nalco azole test kit. See Table 5-2 on page 73.
   h. The water has been refreshed.

2. If water can be poured down the sanitary sewer:
   a. Pour the water down the drain with the permission of the local municipality.
   b. Fill the system by using one of the following procedures:
      • Systems less than 20 gallons, see 5.9.1, “Initial treatment procedure for systems of 20 gallons or less” on page 74 for the cleaning and dosing procedure
      • Systems greater than 20 gallons, see 5.9.2, “Initial treatment procedure for systems of greater than 20 gallons” on page 75 for the cleaning and dosing procedure

5.13 Readying the system for movement or storage

If the computers must be moved to another site or put in storage, drain the system-side water in one of two ways:

- De-ionize the system-side water and pour it down any municipal drain. To ensure that the water is now pure de-ionized water, its resistivity must be greater than 1 MΩ.cm.
- Pour the system-side water down a sanitary drain with the permission of the local municipality.

5.14 Disposing of de-ionizing cartridges

The de-ionizing cartridges must be disposed of in accordance with the local municipality ordinances.
5.15 Disposing of water

IBM is not responsible for the disposal of water. The customer is responsible for determining the local regulations that govern disposal of water. The customer is responsible for disposing of water per the governing local regulations.

5.16 Troubleshooting

This section explains how to handle some typical issues that might be encountered during an installation.

- Poor cooling performance. Contact your IBM client representative.
- Reduced flow issues. Contact your IBM client representative.
- Chemical pump issues (where installed and used). Follow the procedures provided by the installer, contact your site water treatment contractor, or contact Nalco, as described in 5.17, “Nalco contact information” on page 78.
- 3D TRASAR alarms or operational issues. Contact Nalco, as described in 5.17, “Nalco contact information” on page 78.
- Discolored water might be an indication of corrosion and or microbiological issues. Refresh the water supply, as described in 5.12, “Refreshing the system-side water” on page 77.
- Slime in the flow meter areas might be an indication of corrosion and or microbiological issues. Refresh the water supply, as described in 5.12, “Refreshing the system-side water” on page 77.
- Elevated microbial counts:
  - For systems of 20 gallons or less, refresh the water supply, as described in 5.12, “Refreshing the system-side water” on page 77.
  - For systems of greater than 20 gallons, add 100 ppm of Nalco H-550 biocide. Retest the bacteria content 24 - 48 hours after biocide dosing. If the bacteria level is not less than 100 CFU/mL, contact your site water treatment contractor or contact Nalco, as described in 5.17, “Nalco contact information” on page 78.

5.17 Nalco contact information

Do not place chemical orders through the local Nalco representative and do not direct technical questions to the local Nalco representative. The local Nalco representatives do not have in-depth knowledge of the requirements described in this document.

Place chemical orders through:

Sandi Jenkins - District Administrator
sjenkins@nalco.com
860-623-6855

Send all technical questions about the treatment, including deviations from the specification, to:

Michael Cannizzaro - District Manager
mcannizzaro@nalco.com
860-462-8664
Appendix A. Installation tools

This appendix describes tools used to install the Blue Gene/Q hardware.
A.1 Alum-a-lift tool

An Alum-A-Lift lifting device, model A250-84, assists with the safe installation and servicing of nodes and I/O drawers in the Blue Gene/Q racks. The lift operates on a 12 V dc battery, which must be charged. The Alum-a-Lift weighs 240 lbs (~109 kg). For more information, see:

http://www.alumalift.com/

**International customers:** Include an email address in the web form. Alum-a-Lift uses the email address to contact customers for the appropriate shipping address.

The lift comes standard with a National Electrical Manufacturers Association (NEMA) 5-15 plug to charge the product. It also comes with a plug adapter. For a list of supported countries for this adapter, see:


A grease to lubricate the lead screw on the lift can be obtained from the supplier. To order, go to:

http://www.alum-a-lube.com/

This service grease does not ship with the lift; instead, the supplier must be contacted to obtain the grease. This grease is only needed for lift maintenance and is not required when the Blue Gene/Q is initially installed.
Glossary

**3D TRASAR controller**  NaAlco equipment unit that enables precise and continuous monitoring of system water chemistries: conductivity, pH, corrosion rates, and turbidity.

**ac**  Alternating current.

**ANSI**  American National Standards Institute.

**AOC**  Active optical cables.

**application-specific integrated circuit**  In computer chip design, an integrated circuit created by first mounting an array of unconnected logic gates on a substrate and later connecting these gates in a particular configuration for a specific application. This design approach allows chips for various applications to be made from the same generic gate array, thus reducing production costs.

**ASHRAE**  American Society of Heating, Refrigerating and Air Conditioning Engineers.

**ASIC**  See application-specific integrated circuit.

**Azole**  Copper (yellow metal) corrosion inhibitor.

**bacteria test kit**  Agar dip slides that enable the user to test for and monitor bacteria.

**bit (b)**  A single, indivisible binary unit of electronic information.

**Blue Gene/L**  The Blue Gene/L system is the original massively parallel supercomputer designed and built in collaboration with the Department of Energy’s NNSA/Lawrence Livermore National Laboratory.

**Blue Gene/P**  The Blue Gene/P system is the second generation of a massively parallel supercomputer in the IBM System Blue Gene Solution series.

**Blue Gene/Q**  The Blue Gene/Q system is the third generation of a massively parallel supercomputer in the IBM System Blue Gene Solution series.

**BPC**  Bulk power controller.

**BPE**  Bulk power enclosure.

**BQC**  Blue Gene/Q compute chip.

**BQL**  Blue Gene/Q link chip.

**byte**  A collection of 8 bits.

**CDU**  Coolant distribution unit.

**CFM**  Cubic feet per minute.

**CFU**  Colony-forming unit.

**cluster**  A set of nodes connected through a scalable network technology.

**CN**  See compute node.

**compute card**  Consists of one Blue Gene/Q compute ASIC and 72 SDRAM-DDR3 memory chips. Thirty-two water-cooled compute cards plug into a node board.

**compute node**  The element of Blue Gene/Q that supplies the primary computational resource for execution of a user application.

**core**  A single chip that houses a central processing unit (CPU) and is a component in the larger circuit design of a computer.

**corrosion coupon rack**  Allows for internal monitoring of corrosion by using preweighed coupon specimens, which are later removed and reweighed to document any corrosion.

**CRC**  Cyclic redundancy check.

**CSA**  Canadian Standards Association.

**DCA**  Direct current assembly.

**dc**  Direct current.

**DRAM**  Dynamic random-access memory.

**ECC**  Error-correcting code.

**EEPROM**  See electrically erasable programmable read-only memory.

**electrically erasable programmable read-only memory**  A type of nonvolatile memory.

**EPDM**  Ethylene propylene diene monomer.

**ESD**  Electrostatic discharge.

**federated gigabit Ethernet switch**  Connects the I/O nodes of Blue Gene/Q to external resources, such as the front end node.

**FEN**  See front end node.

**FPGA**  Field programmable gate array.
FPN  Fine-print note.

front end node (FEN)  A separate system that provides the user with an operational interface to the Blue Gene/Q system. Also see login node.

GBps  Gigabyte per second.

GB  See gigabyte.

gigabyte  In decimal notation, 1,073,741,824 bytes, when referring to memory capacity; in all other cases, it is defined as 1,000,000,000.

GPFS  IBM General Parallel File System.

I/O node  An air-cooled component that provides connections to file systems, compute nodes, and other equipment.

I²C  Inter-integrated circuit.

input/output (I/O)  Describes any operation, program, or device that transfers data to or from a computer.

ION  See I/O node.

ISA  International Society of Automation.

ISO  International Organization for Standardization.

ITSO  International Technical Support Organization.

JTAG  Joint Test Action Group.

LAN  Local area network.

LCD  Liquid crystal display.

LED  Light-emitting diode.

LINPACK  A collection of Fortran subroutines that analyze and solve linear equations and linear least-squares problems.

LN  See login node.

login node  A separate system, functionally identical to a front end node, that provides the user with an operational interface to the Blue Gene/Q system through an I/O node. Also see front end node.

management node  See service node.

MERV  Minimum efficiency reporting value.

midplane  An intermediate packaging component of Blue Gene/Q. Multiple node boards plug into a midplane to form the basic scalable unit of Blue Gene/Q.

MIG  Metal inert gas.

mL  Milliliter.

MSDS  Material safety data sheets.

MTM  Machine type/model.

Nalco 2567  Nalco pre-cleaner to prepare the system for the correct water quality, in concentrated form. Available in 5-, 15-, or 55-gallon containers:

> Nalco 460-CCL2567  Part number available in the United States. Pre-mixed pre-cleaner. Available in 1- or 5-gallon containers.

> Nalco CCL2567  Part number available outside the United States. Pre-mixed pre-cleaner. Available in 1- or 5-gallon containers.


Nalco 460-CCL100  Part number available in the United States. Pre-mixed inhibitor and preservative. Azole concentration is 40 ppm. Available in 1- or 5-gallon containers.

Nalco CCL100  Part number available outside the United States. Pre-mixed inhibitor and preservative. Azole concentration is 40 ppm. Available in 1- or 5-gallon containers.

Nalco H-550  Nalco biocide treatment for the control of microbiological organisms in concentrated form. This biocide degradation rate depends on the pH of the water, but generally fully degrades within 48 hours of introduction to the system. Available in 5-, 15-, or 55-gallon containers.

NEC  National electrical code.

NEMA  National Electrical Manufacturers Association.

node  An independent operating-system partition that operates under a single instance of an operating-system image.

node board  An intermediate packaging component of Blue Gene/Q. Compute cards are plugged into a node board. Multiple node boards plug into a midplane to form the basic scalable unit of Blue Gene/Q.

NPT  National pipe thread.

NPTM  National Pipe Thread Male.

NTU  Nephelometric turbidity units.

ODT  On-die termination.

PCIe  Peripheral Component Interconnect Express.

PHY  Physical layer.
PMBus  Power management bus.

PPE  Personal protection equipment that consists of gloves, chemical resistant apron, and face shield.

Primary water loop  This is the building chilled water loop.

PTMON  Process and temperature monitor.

PVC  Polyvinyl chloride.

QDR  Quad data rate.

RAS  See reliability, availability, and serviceability.

RDHX  Rear-door heat exchanger.

reliability, availability, and serviceability  Includes those aspects of hardware and software design and development, solution design and delivery, manufacturing quality, technical support service and other services that contribute to assuring that the IBM offering is available when the client wants to use it. Also ensures that it reliably performs the job, that if failures do occur, they are nondisruptive and repaired rapidly, and that after repair the user can resume operations with a minimum of inconvenience.

RPQ  Request price quotation.

scalable  A system attribute that increases in performance or size as some function of the peak rating of the system.

SDRAM  See synchronous, dynamic random access memory.

secondary water loop  For systems that use facility water to cool a secondary loop, the system-side closed water loop that flows through the Blue Gene/Q systems.

service node  A separate system used for initial program load and to manage the Blue Gene/Q system. Sometimes also referred to as “management node.”

SN  See service node.

SSN  Subnet service node.

synchronous, dynamic random access memory  A type of dynamic random access memory (DRAM) with features that make it faster than standard DRAM.

TIG  Tungsten inert gas.

tori  The plural form of the word torus.

torus  In Blue Gene/Q, a 5-dimensional network for processor-to-processor communication.
Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

For information about ordering these publications, see “How to get Redbooks”. Note that some of the documents referenced here may be available in softcopy only:

- *IBM System Blue Gene Solution: Blue Gene/Q Application Development Manual*, SG24-7948-00
- *IBM System Blue Gene Solution: Blue Gene/Q Code Development and Tools Interface*, REDP-4659
- *IBM System Blue Gene Solution: Blue Gene/Q Hardware Installation and Maintenance Guide*, SG24-7974-00
- *IBM System Blue Gene Solution: Blue Gene/Q Safety Considerations*, REDP-4656
- *IBM System Blue Gene Solution: Blue Gene/Q System Administration Manual*, SG24-7869-00

Other publications

This publication is also relevant as a further information source:

- *General Parallel File System HOWTO for the IBM System Blue Gene/Q Solution*, SC23-6939-00

Online resources

These websites are also relevant as further information sources:

- PCI Express: “PCle Base 2.1 Specification PCIe Card Electromechanical 2.0 Specification”
  
  http://www.pcisig.com/specifications/pciexpress/base2

- IBM System Information Center: “Planning for the installation of IBM Rear Door Heat eXchanger for 7014-T42”
  

- IBM System Information Center: “Static Electricity and Floor Resistance”
  
How to get Redbooks

You can search for, view, or download Redbooks, Redpapers, Technotes, draft publications and Additional materials, as well as order hardcopy Redbooks publications, at this website:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services
Index

Numerics
3D TRASAR controller 73, 76

A
active optical cables 10
Alum-a-Lift 80
ASIC 4–5, 13–14
Azole 75–77

B
Blue Gene/L 12
Blue Gene/P 2, 12
Blue Gene/Q compute 3–5, 7–8, 13–14, 18–19, 25–26, 30–31
Blue Gene/Q link 5, 7
bulk power enclosure 7

cables 3, 5, 7, 9–10, 12–13, 19, 27, 50, 56–59, 61–64
clock fanout card 7–8
cluster 13
colonies-forming unit 71, 76
compute card 4–7, 12–14, 29
compute node 7, 12–13
compute rack 3–5, 8, 10, 12, 18–23, 25–29, 31, 61, 63
coolant distribution unit 41–42
corrosion 64–65, 67–70, 73, 77
cyclic redundancy check 12

D
DRAM 14–15

E
ECC 14
EEPROM 4
ESD 50

F
FPGA 5–7, 13–14
front end node 3

G
General Parallel File System 2

I
I/O compute card 14
I/O drawer 4, 6–7, 13, 18–19, 21, 26, 33, 61–64
I/O node 13, 50
I/O rack 3, 6, 10, 18–19, 24, 26–30, 33–34
I2C 4–7, 14
ITSO ix–x

J
JTAG 5, 7, 13–14

L
LAN 2–3

M
material safety data sheets 44
MERV 65
midplane 4–5, 8–10, 12–14, 18–19, 25–28, 33, 42

N
Nalco 460-CCL100 75
node 2–7, 10, 12–14, 30, 50	node board 4–6, 12–14	node boards 5–7, 10, 13
NTU 70

O
on-die termination 14

P
PCIe 3, 6–7, 10, 13
PMBus 7
power management bus 7
process and temperature monitor 5
PVC 42, 71

Q
quad data rate 2, 7, 13

R
RAS 12
rear-door heat exchanger 34
Redbooks ix–x, 92
Redbooks Web site
Contact us x
Redbooks web site 86

S
scalable 3
SDRAM 4, 14
servers 2–3, 64–65
service card 5–7, 9–10, 13, 25
service node 2–3, 13, 30
Site LAN 3

© Copyright IBM Corp. 2013. All rights reserved.
site LAN  3
subnet service node  3

T
  torus  3, 5, 10, 12
  transient voltage surge suppressor  55
  Tungsten inert gas  71

U
  Underwriters Laboratory  71

V
  VFD  43

W
  water loop  72
IBM System Blue Gene Solution: Blue Gene/Q Hardware Overview and Installation Planning

This document is one of a series of IBM Redbooks written specifically for the IBM System Blue Gene supercomputer, IBM Blue Gene/Q. Blue Gene/Q is the third generation of massively parallel supercomputers from IBM in the Blue Gene series. This document provides an overview of components that comprise a Blue Gene/Q system. It helps System Planners and Customer Engineers plan for the installation of the Blue Gene/Q system. Information is provided about the physical requirements for the machine room where the Blue Gene/Q system is to be located. Examples of these requirements include floor (weight and cutouts), cooling, and electrical specifications.