

IEEE Standard for Framework and Process for Large-Scale Deep Learning Model Evaluation

IEEE Computer Society

Developed by the
Artificial Intelligence Standards Committee

IEEE Std 3378™-2025

STANDARDS

IEEE Standard for Framework and Process for Large-Scale Deep Learning Model Evaluation

Developed by the

Artificial Intelligence Standards Committee
of the
IEEE Computer Society

Approved 10 September 2025

IEEE SA Standards Board

Abstract: The tasks that large-scale deep learning models can support are specified in this standard. The evaluation metrics and evaluation requirements of large-scale deep learning models are addressed in this standard.

Keywords: deep learning, evaluation metrics, evaluation requirements, IEEE 3378, large-scale model

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2025 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved. Published 27 January 2026. Printed in the United States of America.

IEEE is a registered trademark in the U.S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated.

PDF: ISBN 979-8-8557-2733-3 STD28349
Print: ISBN 979-8-8557-2734-0 STDPD28349

IEEE prohibits discrimination, harassment, and bullying.

For more information, visit <https://www.ieee.org/about/corporate/governance/p9-26.html>.

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher.

Important Notices and Disclaimers Concerning IEEE Standards Documents

IEEE Standards documents are made available for use subject to important notices and legal disclaimers. These notices and disclaimers, or a reference to this page (<https://standards.ieee.org/ipr/disclaimers.html>), appear in all IEEE standards and may be found under the heading “Important Notices and Disclaimers Concerning IEEE Standards Documents.”

Notice and Disclaimer of Liability Concerning the Use of IEEE Standards Documents

IEEE Standards documents are developed within IEEE Societies and subcommittees of IEEE Standards Association (IEEE SA) Board of Governors. IEEE develops its standards through an accredited consensus development process, which brings together volunteers representing varied viewpoints and interests to achieve the final product. IEEE standards are documents developed by volunteers with scientific, academic, and industry-based expertise in technical working groups. Volunteers involved in technical working groups are not necessarily members of IEEE or IEEE SA and participate without compensation from IEEE. While IEEE administers the process and establishes rules to promote fairness in the consensus development process, IEEE does not independently evaluate, test, or verify the accuracy of any of the information or the soundness of any judgments contained in its standards.

IEEE makes no warranties or representations concerning its standards, and expressly disclaims all warranties, express or implied, concerning all standards, including but not limited to the warranties of merchantability, fitness for a particular purpose and non-infringement. IEEE Standards documents do not guarantee safety, security, health, or environmental protection, or compliance with law, or guarantee against interference with or from other devices or networks. In addition, IEEE does not warrant or represent that the use of the material contained in its standards is free from patent infringement. IEEE Standards documents are supplied “AS IS” and “WITH ALL FAULTS.”

Use of an IEEE standard is wholly voluntary. The existence of an IEEE standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard.

In publishing and making its standards available, IEEE is not suggesting or rendering professional or other services for, or on behalf of, any person or entity, nor is IEEE undertaking to perform any duty owed by any other person or entity to another. Any person utilizing any IEEE Standards document should rely upon their own independent judgment in the exercise of reasonable care in any given circumstances or, as appropriate, seek the advice of a competent professional in determining the appropriateness of a given IEEE standard.

IN NO EVENT SHALL IEEE BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Translations

The IEEE consensus balloting process involves the review of documents in English only. In the event that an IEEE standard is translated, only the English language version published by IEEE is the approved IEEE standard.

Use by artificial intelligence systems

In no event shall material in any IEEE Standards documents be used for the purpose of creating, training, enhancing, developing, maintaining, or contributing to any artificial intelligence systems without the express, written consent of IEEE SA in advance. “Artificial intelligence” refers to any software, application, or other system that uses artificial intelligence, machine learning, or similar technologies, to analyze, train, process, or generate content. Requests for consent can be submitted using the Contact Us form.

Official statements

A statement, written or oral, that is not processed in accordance with the IEEE SA Standards Board Operations Manual is not, and shall not be considered or inferred to be, the official position of IEEE or any of its committees and shall not be considered to be, or be relied upon as, a formal position of IEEE or IEEE SA. At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that the presenter’s views should be considered the personal views of that individual rather than the formal position of IEEE, IEEE SA, the Standards Committee, or the Working Group. Statements made by volunteers may not represent the formal position of their employer(s) or affiliation(s). News releases about IEEE standards issued by entities other than IEEE SA should be considered the view of the entity issuing the release rather than the formal position of IEEE or IEEE SA.

Comments on standards

Comments for revision of IEEE Standards documents are welcome from any interested party, regardless of membership affiliation with IEEE or IEEE SA. However, **IEEE does not provide interpretations, consulting information, or advice pertaining to IEEE Standards documents.**

Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments. Since IEEE standards represent a consensus of concerned interests, it is important that any responses to comments and questions also receive the concurrence of a balance of interests. For this reason, IEEE and the members of its Societies and subcommittees of the IEEE SA Board of Governors are not able to provide an instant response to comments or questions, except in those cases where the matter has previously been addressed. For the same reason, IEEE does not respond to interpretation requests. Any person who would like to participate in evaluating comments or revisions to an IEEE standard is welcome to join the relevant IEEE SA working group. You can indicate interest in a working group using the Interests tab in the Manage Profile & Interests area of the [IEEE SA myProject system](#).¹ An IEEE Account is needed to access the application.

Comments on standards should be submitted using the [Contact Us](#) form.²

Laws and regulations

Users of IEEE Standards documents should consult all applicable laws and regulations. Compliance with the provisions of any IEEE Standards document does not constitute compliance to any applicable regulatory requirements. Implementers of the standard are responsible for observing or referring to the applicable regulatory requirements. IEEE does not, by the publication of its standards, intend to urge action that is not in compliance with applicable laws, and these documents may not be construed as doing so.

¹ Available at: <https://development.standards.ieee.org/myproject-web/public/view.html#landing>.

² Available at: <https://standards.ieee.org/about/contact/>.

Data privacy

Users of IEEE Standards documents should evaluate the standards for considerations of data privacy and data ownership in the context of assessing and using the standards in compliance with applicable laws and regulations.

Copyrights

IEEE draft and approved standards are copyrighted by IEEE under U.S. and international copyright laws. They are made available by IEEE and are adopted for a wide variety of both public and private uses. These include both use by reference, in laws and regulations, and use in private self-regulation, standardization, and the promotion of engineering practices and methods. By making these documents available for use and adoption by public authorities and private users, neither IEEE nor its licensors waive any rights in copyright to the documents.

Photocopies

Subject to payment of the appropriate licensing fees, IEEE will grant users a limited, non-exclusive license to photocopy portions of any individual standard for company or organizational internal use or individual, non-commercial use only. To arrange for payment of licensing fees, please contact Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA 01923 USA; +1 978 750 8400; <https://www.copyright.com/>. Permission to photocopy portions of any individual standard for educational classroom use can also be obtained through the Copyright Clearance Center.

Updating of IEEE Standards documents

Users of IEEE Standards documents should be aware that these documents may be superseded at any time by the issuance of new editions or may be amended from time to time through the issuance of amendments, corrigenda, or errata. An official IEEE document at any point in time consists of the current edition of the document together with any amendments, corrigenda, or errata then in effect.

Every IEEE standard is subjected to review at least every 10 years. When a document is more than 10 years old and has not undergone a revision process, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE standard.

In order to determine whether a given document is the current edition and whether it has been amended through the issuance of amendments, corrigenda, or errata, visit [IEEE Xplore](#) or [contact IEEE](#).³ For more information about the IEEE SA or IEEE's standards development process, visit the IEEE SA Website.

Errata

Errata, if any, for all IEEE standards can be accessed on the [IEEE SA Website](#).⁴ Search for standard number and year of approval to access the web page of the published standard. Errata links are located under the Additional Resources Details section. Errata are also available in [IEEE Xplore](#). Users are encouraged to periodically check for errata.

³ Available at: <https://ieeexplore.ieee.org/browse/standards/collection/ieee>.

⁴ Available at: <https://standards.ieee.org/standard/index.html>.

Patents

IEEE standards are developed in compliance with the [IEEE SA Patent Policy](#).⁵

Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. If a patent holder or patent applicant has filed a statement of assurance via an Accepted Letter of Assurance, then the statement is listed on the IEEE SA Website at <https://standards.ieee.org/about/sasb/patcom/patents.html>. Letters of Assurance may indicate whether the Submitter is willing or unwilling to grant licenses under patent rights without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination to applicants desiring to obtain such licenses.

Essential Patent Claims may exist for which a Letter of Assurance has not been received. The IEEE is not responsible for identifying Essential Patent Claims for which a license may be required, for conducting inquiries into the legal validity or scope of Patents Claims, or determining whether any licensing terms or conditions provided in connection with submission of a Letter of Assurance, if any, or in any licensing agreements are reasonable or non-discriminatory. Users of this standard are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. Further information may be obtained from the IEEE Standards Association.

IMPORTANT NOTICE

Technologies, application of technologies, and recommended procedures in various industries evolve over time. The IEEE standards development process allows participants to review developments in industries, technologies, and practices, and to determine what, if any, updates should be made to the IEEE standard. During this evolution, the technologies and recommendations in IEEE standards may be implemented in ways not foreseen during the standard's development. IEEE standards development activities consider research and information presented to the standards development group in developing any safety recommendations. Other information about safety practices, changes in technology or technology implementation, or impact by peripheral systems also may be pertinent to safety considerations during implementation of the standard. Implementers and users of IEEE Standards documents are responsible for determining and complying with all appropriate safety, security, environmental, health, data privacy, and interference protection practices and all applicable laws and regulations.

⁵ Available at: <https://standards.ieee.org/about/sasb/patcom/materials.html>.

Participants

At the time this IEEE standard was completed, the Large-Scale Deep Learning Model Working Group had the following membership:

Dahua Lin, Chair
Peng Li, Vice Chair
Hui Jiang, Secretary

<i>Organization Represented</i>	<i>Name of Representative</i>
Ant Research.....	Peicheng Wang
China Mobile Limited.....	Xi Cao
China Automotive Engineering Research Institute Co., Ltd.	Qingqing Li
China Electronics Standardization Institute	Chenghao Ma
China Southern Power Grid Co., Ltd.	Peng Li
China Telecom Research Institute	Qing Liu
Guangdong OPPO Mobile Telecommunications Corp., Ltd.....	Xiaopu Wang
Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security	Xiangjuan Jia
Hangzhou Qulian Technology Co., Ltd.	Yijian Zhang
Hisense Group Holdings Co., Ltd.	Changyu Liu
Huawei Cloud Technologies Co., Ltd.....	Zimu Zheng
IEIT Systems Co., Ltd.	Baihong Li
iFLYTEK Co.,Ltd.....	Qiaoling Liang
Mashang Consumer Finance Co., Ltd.....	Yufeng Xie
Midea Technology Co., Ltd.	Feifei Feng
North China Electric Power University	Hua Wu
Peking University	Chao Yang
PKU-Changsha Institute for Computing and Digital Economy	Haipeng Gou
Qingdao Haier Technology Co., Ltd.....	Lei E
SenseTime Technology Co., Ltd.....	Geng Wu
Shanghai AI Lab	Dahua Lin
Siheria Technologies Co., Ltd.....	Baoxuan Hong
Software Engineering Center of China Academy of Science.....	Peng Chen
State Grid Hunan Electric Power Co., Ltd. Electric Power Research Institute	Jian Xiao
Tencent	Yongxia Wang
WeBank Co., Ltd.	Yuxiang Gao
Xi'an Jiaotong University.....	Xiaohua Wang
Zhejiang University	Xiaofeng Chen

The Standard for Framework and Process for Large-Scale Deep Learning Model Evaluation Group had the following membership during the development of this standard:

Kai Chen, Chair
Chenghao Ma, Vice Chair
Feifei Feng, Vice Chair
Geng Wu, Vice Chair
Jianong Li, Secretary

The Working Group gratefully acknowledges the contributions of the following participants. Without their assistance and dedication, this standard would not have been completed.

Yasen Cai
Yunfei Cai
Ge Chen
Kai Chen
Huipeng Deng
Weihong Deng
Haodong Duan

Penghui Hong
Hui Jiang
Jianong Li
Min Li
Mingxuan Li
Shaoqing Lu

Jinran Nie
Xian Shi
Wenxiu Sun
Zekun Wang
Dingheng Zeng
Songyang Zhang
Xingcheng Zhang

The following members of the entity Standards Association balloting group voted on this standard. Balloters may have voted for approval, disapproval, or abstention.

OxSenses Corporation	IBM
Alibaba China Co., Ltd.	Imaginary Universes, LLC
Beijing Tsingzhi Shuyuan Tech Inc.	Institute of Blockchain and Data Security
CETC Big Data Research Institute Co., Ltd	Intel Corporation
China Academy of Information and Communications Technology	Lenovo Group Limited
China Electronic Standardization Institute	Lockheed Martin
China Mobile Chengdu Institute of Research and Development	Mashang Consumer Finance Co., Ltd.
China Telecommunications Corporation	Northeastern University
China Three Gorges University	Peking University
Chongqing University of Posts and Telecommunications	Primeton Information Technologies, Inc.
CRRRC Zhuzhou Institute Co., Ltd.	Shanghai AI Lab
Dell Inc.	Shanghai Lingshu Zhonghe Information Technology Co., Ltd.
The First Affiliated Hospital of Zhejiang University, School of Medicine	Skysys Intelligent Technology Co., Ltd.
Haier Group Corporation	Tongyuan Software Co., Ltd.
Hangzhou Qulian Technology Co., Ltd.	WeBank Co., Ltd.
Hisense Group Holdings Co., Ltd.	Wuxi SensingNet Industrialization Research Institute
	Zhejiang University

When the IEEE SA Standards Board approved this standard on 10 September 2025, it had the following membership:

Lei Wang, *Chair*
Jon Walter Rasdahl, *Vice Chair*
David J. Law, *Past Chair*
Alpesh Shah, *Secretary*

Edward Au	Guido R. Hiertz	Daleep C. Mohla
Ted Burse	Ronald W. Hotchkiss	Annette D. Reilly
Xiaofeng (Alfred) Chen	Tyler L. Jaynes	Robby Robson
Doug Edwards	Thomas Koshy	Daniel Sabin
Nehad El-Sherif	Howard Li	F. Keith Waters
J. Travis Griffith	Xiaohui Liu	Sha Wei
Deborah R. Hagar	Kevin W. Lu	Luyang (Eric) Zhang
	Hiroshi Mano	

Introduction

This introduction is not part of IEEE Std 3378-2025, IEEE Standard for Framework and Process for Large-Scale Deep Learning Model Evaluation.

Large-scale deep learning models have become an important technical means for the development of artificial intelligence, playing an important role in leading industrial transformation. Artificial intelligence-related institutions have successively researched and developed more than 100 large-scale deep learning model products and evaluation lists, making it difficult for users to effectively evaluate the technical level and service capabilities of artificial intelligence products in the industry. This standard defines the tasks that can be supported by large-scale deep learning models, evaluation metrics, and evaluation methods.

Contents

1. Overview	11
1.1 Scope	11
1.2 Word usage.....	11
2. Normative references.....	11
3. Definitions, acronyms, and abbreviations	12
3.1 Definitions	12
3.2 Acronyms and abbreviations	12
4. Framework.....	12
5. General requirements.....	13
5.1 Resource pool	13
5.2 Tools.....	14
5.3 Data resources.....	17
5.4 Models	17
5.5 Industry applications.....	18
5.6 Service platform/cluster.....	18
6. Evaluation metrics	18
6.1 Evaluation dimensions.....	18
6.2 Evaluation metrics of understanding capability.....	18
6.3 Evaluation metrics of generation capability	30
7. Evaluation requirements.....	37
7.1 Evaluation method.....	37
7.2 Evaluation data sets	39
7.3 Evaluation environment.....	39
7.4 Evaluation execution	39
7.5 Evaluation tools	40
7.6 Evaluation implementation.....	40
Annex A (informative) Calculation method for evaluation metrics	41
A.1 Objective evaluation method	41
A.2 Objective evaluation method	43
A.3 Grading of large-scale deep learning model capabilities	45
Annex B (informative) Bibliography.....	46

IEEE Standard for Framework and Process for Large-Scale Deep Learning Model Evaluation

1. Overview

1.1 Scope

This standard applies to the design, development, application, and evaluation of large-scale deep learning models for artificial intelligence. The standard specifies a technical reference architecture of large-scale deep learning models. Furthermore, the standard provides general requirements for basic software and hardware functions of these learning models and their performance. The standard also includes a framework for the evaluation of such large-scale deep learning models.

1.2 Word usage

The word *shall* indicates mandatory requirements strictly to be followed in order to conform to the standard and from which no deviation is permitted (*shall* equals *is required to*).^{6, 7}

The word *should* indicates that among several possibilities one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required (*should* equals *is recommended that*).

The word *may* is used to indicate a course of action permissible within the limits of the standard (*may* equals *is permitted to*).

The word *can* is used for statements of possibility and capability, whether material, physical, or causal (*can* equals *is able to*).

2. Normative references

The following referenced documents are indispensable for the application of this document (i.e., they must be understood and used, so each referenced document is cited in text and its relationship to this document is explained). For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments or corrigenda) applies.

There are no normative references in this standard.

⁶ The use of the word *must* is deprecated and cannot be used when stating mandatory requirements; *must* is used only to describe unavoidable situations.

⁷ The use of *will* is deprecated and cannot be used when stating mandatory requirements; *will* is only used in statements of fact.

3. Definitions, acronyms, and abbreviations

3.1 Definitions

For the purposes of this document, the following terms and definitions apply. The *IEEE Standards Dictionary Online* should be consulted for terms not defined in this clause.⁸

deep learning: A method of creating rich hierarchical representations by training neural networks with many hidden layers.

inference: Draw a conclusion from the given premise.

test data: Data used to evaluate the performance of the final machine learning model.

3.2 Acronyms and abbreviations

API application programming interface

BLEU bilingual evaluation understudy

4. Framework

The large-scale deep learning model framework under the functional perspective is shown in Figure 1, including resource pool, tools, data resources, models, industry applications, and service platform/cluster. Among them:

- a) The resource pool includes hardware resources such as computing resources, storage resources, and network resources, and software resources such as resource virtualization/scheduling.
- b) Tools include data tools and modeling tools.
- c) Data resources include general data, domain data, and private data.
- d) Models include a basic large-scale deep learning model and a customized large-scale deep learning model. The basic large-scale deep learning model can support single-modal or multi-modal data, and a customized large-scale deep learning model can fine-tune the basic large-scale deep learning model according to user needs and customize the large-scale deep learning model to be suitable for the production environment.
- e) Industry applications provide large-scale deep learning model downstream task-matching services for users in various industry scenarios.
- f) Service platform/cluster runs through all levels, providing orchestration, deployment, model reasoning, operation, and management to support large-scale deep learning models and related services.

⁸ *IEEE Standards Dictionary Online* is available at: <http://dictionary.ieee.org>. An IEEE account is required for access to the dictionary, and one can be created at no charge on the dictionary sign-in page.

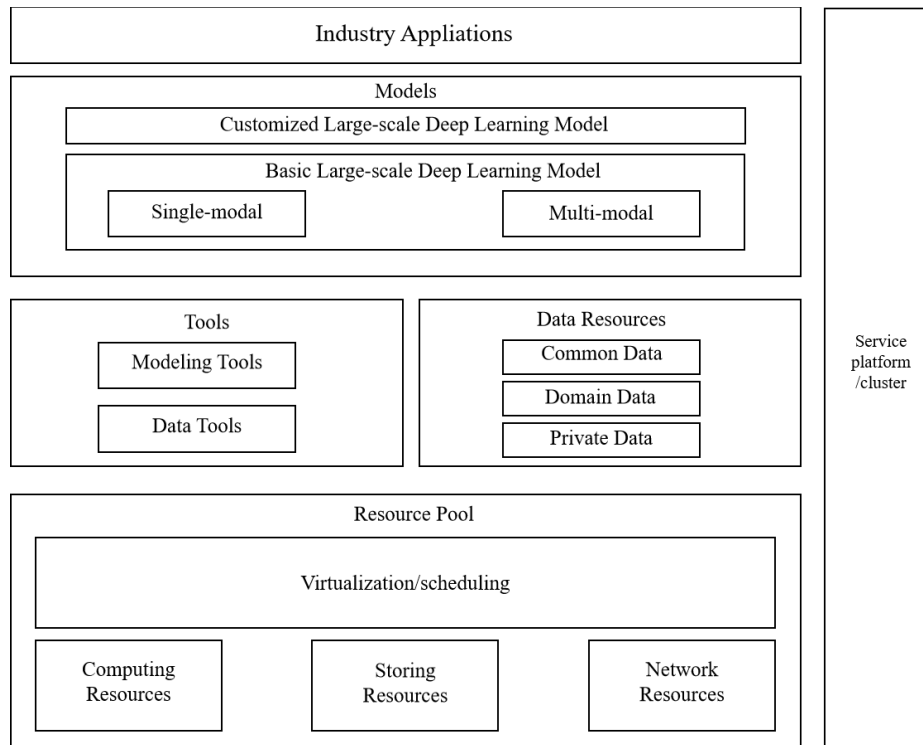


Figure 1—Framework for large-scale deep learning models

5. General requirements

5.1 Resource pool

5.1.1 Computing resources

A physical computing facility (e.g., central processing unit, graphics processing unit, field programmable gate array, neural network processor, tensor processor) or virtual computing device that provides computing and data processing capabilities for model training and inference.

- a) Model training or inference for at least one modality (e.g., text, image, speech) shall be performed.
- b) Shall have hardware acceleration functions for artificial intelligence computing and be equipped with distributed training and inference computing acceleration libraries.
 - 1) Training server: shall have no fewer than four 100GE network ports. Power module and fan module shall have hot plug and backup function (such as 2 + 2 redundancy, N + 1 redundancy, etc.).
 - 2) Inference server: the total memory bandwidth shall not be less than 800 GB/s. It shall have no fewer than two high-speed serial computer expansion bus standard (PCIe) expansion slots. Power module and fan module shall have a hot plug and backup function (such as 1 + 1 redundancy, N + 1 redundancy, etc.).
- c) Should have the preprocessing function based on hardware acceleration (such as image and video codecs).
- d) Key-value caching should be available.

5.1.2 Storing resources

Storage resources for large-scale deep learning model training and inference, including storage servers, etc. Storage resources are used to provide data storage and model storage:

- a) Shall have distributed storage and access function of data sets and a redundant backup mechanism.
- b) Shall have a standard file system interface, such as a portable operating system interface.
- c) The storage bandwidth should not be less than 200 GB/s, and the number of input/output operations should not be less than 2 million times/s.
- d) Should have a memory computing function.
- e) Should create a storage pool with storage servers or hard disks as units, and the storage pool should be able to identify and manage different types of storage media, such as solid state drives and hard disks.

5.1.3 Network resources

Network resources for large-scale deep learning model training and inference, including intra-cluster switches and routers:

- a) Shall be suitable for high-speed network communication protocols, such as 100GE, remote direct memory access based on converged Ethernet, etc.
- b) Forwarding packet rate should not be less than 4000 Mpps.
- c) Shall have load-balancing function.
- d) Reliable networking schemes should be available, such as link aggregation, cross-device link aggregation group double live (M-LAG), etc.
- e) Should have a fully connected network of 40GE, 100GE, 200GE, or 400GE in the server cluster.
- f) Should have the mapping function between the physical switch and the logical switch, and the failure of a single physical switch should not affect the execution of training and reasoning tasks.

5.2 Tools

5.2.1 Data tools

5.2.1.1 Data acquisition tools

Data tool functions for data acquisition include:

- a) The data collection task setting function shall be provided, including collection requirements, quantity, data source, category (such as text, voice, picture, and video, etc.), and scope (such as topic, content, etc.) of the collected data.
- b) Shall have the function of collecting multiple types of data, including text, video, image, audio, etc.
- c) Shall have the recording function of the source, time, and mode of data collection.
- d) Shall have structured, semi-structured, and unstructured data access functions.
- e) Data quality detection and data cleaning functions should be provided.