Case Study

intel.

Intel® SGX Empowers Alibaba Cloud to Build an End-to-End PPML Solution

Overview

Innovative digital technologies such as big data and artificial intelligence (AI) are reshaping the world rapidly. They are changing the social and economic development model by bringing more possibilities to people's lives. Knowing that data has become one of the most important assets, an increasing number of enterprises are seizing the trend of data-centric transformation. Meanwhile, however, their data elements are also faced with serious security risks in the process of storage, processing, and circulation both inside and outside organizations. Therefore, protecting data security is now one of the most important factors deciding success or failure of digital transformation.

Based on Intel® Software Guard Extensions (Intel® SGX), Intel has built BigDL PPML (Privacy Preserving Machine Learning) to secure the end-to-end big data and AI pipeline. In collaboration with Alibaba Cloud DataTrust, Intel has verified the PPML solution in its end-to-end workflow and relevant business scenarios, and demonstrated the best practice for rapidly building end-to-end privacy-preserving applications based on BigDL PPML.

Background: Data Fusion for Big Data and AI Challenged by Security Risks

Digital transformation, while highlighting the importance of data value and accelerating data flow, results in complex data storage, circulation, and processing among multiple parties. It is difficult for a single organization to prepare all data, especially those required by AI and big data applications. Therefore, multiple parties should cooperate to realize data convergence for utilization. Take the training of financial AI algorithms as an example, individual financial institutions cannot meet the needs with their own data. In this case, different parties can work together to create and maintain AI models, and share data eventually.

However, as a result of the increasing need for cross-institutional and cross-industry data fusion, analysis and modeling, data security risks have also accumulated dramatically. On the one hand, data can be copied and spread easily, making it difficult to track them once shared in the traditional security model. On the other hand, continuous data flow will lead to problems such as unclear division of responsibilities, difficult authority control, and hard tracking of accountability. Therefore, data security and reliability becomes top priority.

However, traditional security solutions to AI and big data applications often face the following challenges:

- Joint analysis and modeling require frequent data sharing and convergence, while traditional data security solutions are designed to protect data at rest and in transit, instead of data in use. As a result, some security threats may break through the lines of security defense, causing incidents such as data leakage.
- AI and big data applications involve multiple processes, such as data input, data analysis, machine learning, and deep learning. Vulnerabilities in any of these process may lead to serious consequences including data leakage. Therefore, it is crucial to ensure end-to-end security.
- Attacks against AI and big data encompass a wide range of known and unknown security threats, as well as a variety of attack techniques and tools. However, traditional solutions generally work on the software level, yet hardly protect the bottom layer of hardware, which hinders further improvement in its protection efficiency.
- Data security measures often resort to relatively complex calculations, which may lead to certain performance losses and a negative impact on the operational efficiency of data bank.

Solution: Alibaba Cloud E2E PPML Based on BigDL PPML

To help enterprises better protect end-to-end privacy of AI and big data applications, Alibaba Cloud and Intel worked together to validate Alibaba Cloud E2E PPML in end-to-end workflow and related business scenarios by synergizing BigDL PPML and Alibaba Cloud DataTrust.

BigDL PPML

BigDL, a unified open-source AI solution platform from Intel, makes it easier for data scientists and data engineers to build end-to-end, distributed AI applications. Using Intel® SGX, Intel's Trusted Execution Environment (TEE), and integrating with other hardware

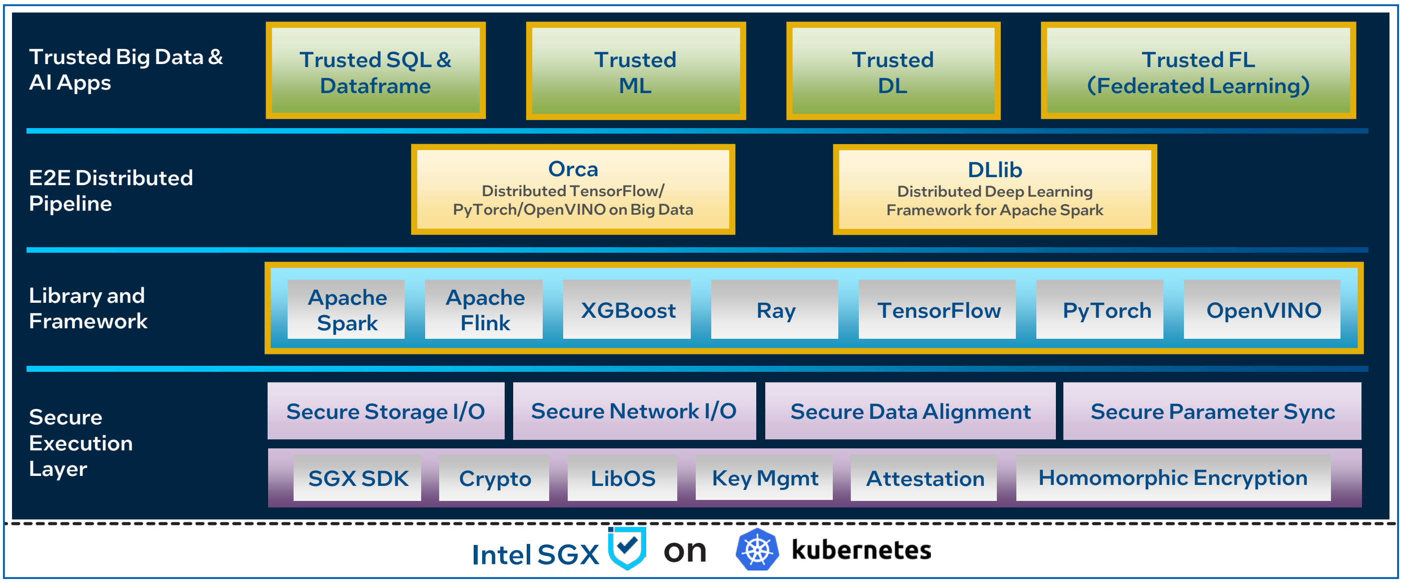


Figure 1. BigDL PPML software stack

and software security measures, BigDL has built a distributed PPML platform to protect end-to-end distributed AI pipeline (from data ingestion, data analysis, all the way to machine learning and deep learning).

As an important underlying technology of BigDL PPML, Intel® SGX bypasses a system's operating system (OS) and virtual machine (VM) software layers to provide significant additional protection against many of these attacks. It adds data security and addresses the need for more confidential computing. Intel® SGX offers hardware-based memory encryption that isolates specific application code and data in memory. Intel® SGX allows user-level code to allocate private regions of memory, called Enclaves, which are designed to be protected from processes running at higher privilege levels.

Intel® SGX is rigorously tested and widely deployed hardware-based data center trusted execution environment (TEE), with the smallest available attack surface within the system. In addition to helping defend against software-based attacks, Intel® SGX's attestation mechanisms also help users verify that their application and hardware have not been compromised and their processor has the latest security updates.

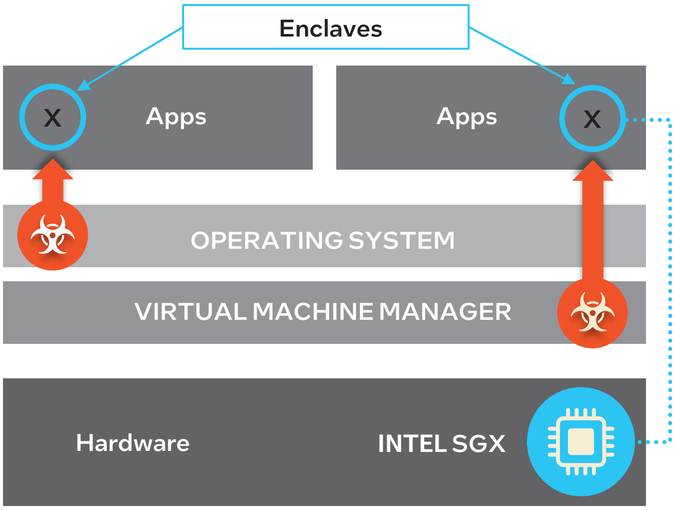


Figure 2. Intel® SGX protects at the bottom layer of hardware

Developers can use BigDL PPML platform to:

- Develop and run standard distributed AI applications (such as big data analysis, machine learning, and deep learning) with encrypted data;
- Protect computing processes and corresponding memory data using hardware-based security technologies such as Intel® SGX;
- Provide end-to-end security and privacy protection for AI applications, for example, creating and authenticating a trusted cluster in a K8s environment with Intel® SGX hardware; offering encryption and decryption capabilities for distributed data via Key Management System (KMS); and enabling secure distributed computing and data communication based on technologies such as Intel® SGX, encryption and decryption, TLS, and security authentication.

Alibaba Cloud DataTrust

Alibaba Cloud DataTrust is an industry-leading privacy enhanced computing platform based on privacy enhancing techniques including TEE, Secure Multi-Party Computation (MPC), Federated Learning (FL), and Differential Privacy (DP). It is committed to realizing secure flow of data value, and bringing the industry right, easy-to-use, and highly available products for secure data circulation.

Built on Intel® SGX, Alibaba Cloud DataTrust employs techniques like MPC and FL and utilizes Alibaba Cloud data center's rich application scenarios, thereby executing joint analysis, training, and prediction with multi-party data while ensuring data security, and providing enterprises with data-service-native solutions to secure data circulation and help boost business growth.

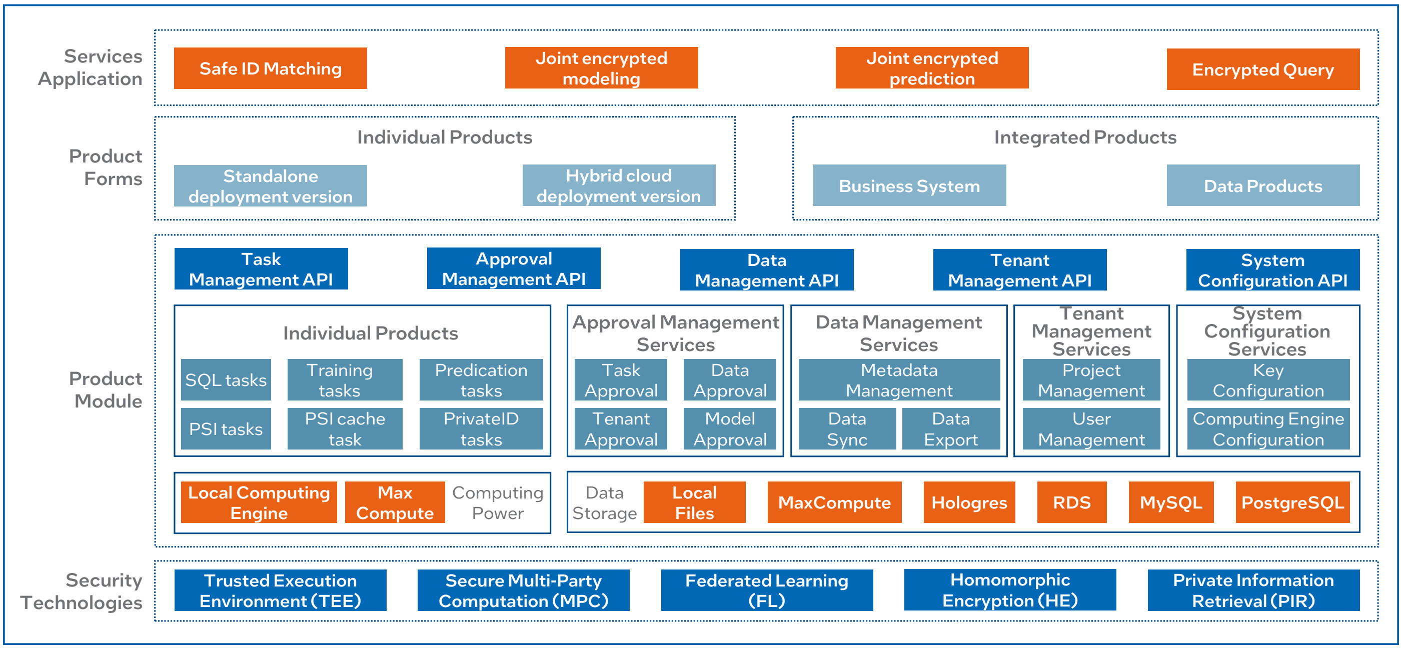


Figure 3. Architecture of Alibaba Cloud DataTrust

End-to-end Solution Workflow

Based on the core functions of privacy computing, BigDL PPML integrates more components of the end-to-end privacy-preserving computing workflow, such as Attestation Service, Key Management Service, and Kubernetes-based secure container deployment.

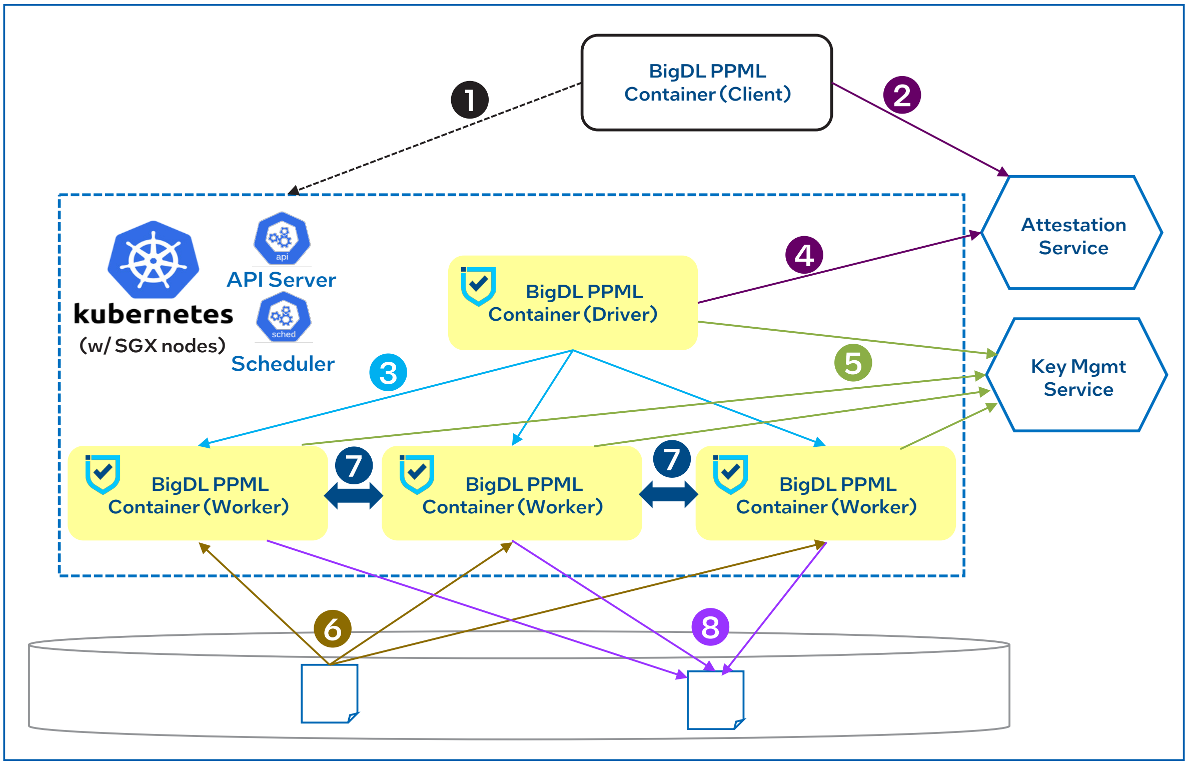
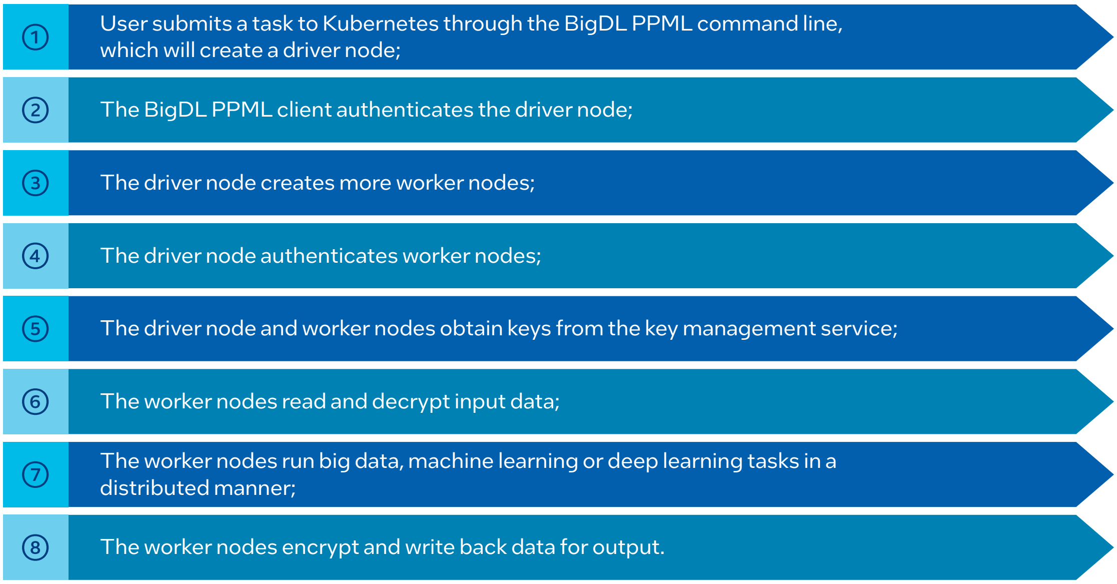

BigDL PPML Container (Client)
1
2
Attestation Service
kubernetes
(w/ SGX nodes)
api
API Server
sched
Scheduler
BigDL PPML Container (Driver)
4
3
5
Key Mgmt Service
BigDL PPML Container (Worker)
7
BigDL PPML Container (Worker)
7
BigDL PPML Container (Worker)
6
8

Figure 4. BigDL PPML based end-to-end secure computing workflow

In the above end-to-end secure computing workflow based on BigDL PPML, the functions of each process are described as follows:



① User submits a task to Kubernetes through the BigDL PPML command line, which will create a driver node;

② The BigDL PPML client authenticates the driver node;

③ The driver node creates more worker nodes;

④ The driver node authenticates worker nodes;

⑤ The driver node and worker nodes obtain keys from the key management service;

⑥ The worker nodes read and decrypt input data;

⑦ The worker nodes run big data, machine learning or deep learning tasks in a distributed manner;

⑧ The worker nodes encrypt and write back data for output.

Figure 5. Functions of end-to-end secure computing workflow based on BigDL PPML

BigDL PPML solution integrates the above workflow schemes that include: SGX-based trusted-computing core components supporting Apache Spark, Spark SQL, machine learning, and deep learning; abstracted client API of attestation service; abstracted client API of key management service; encrypted data transmission and storage; and customized K8s container images.

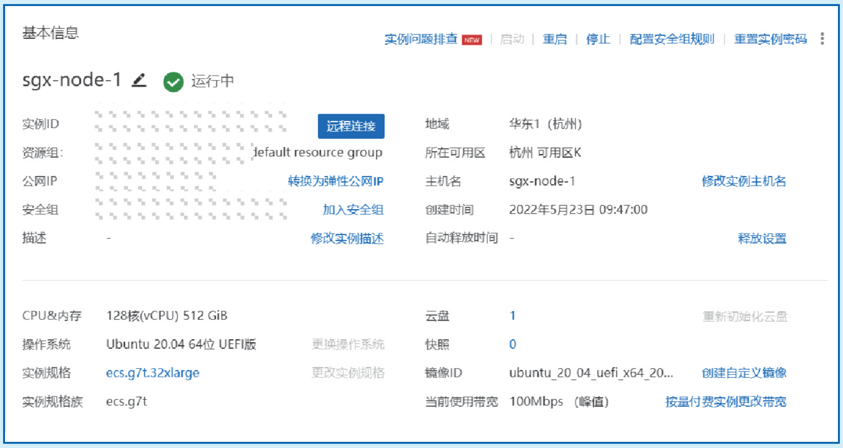
By using the above pre-configured workflow, developers can focus more on the development of business logic, and use BigDL PPML to ensure the end-to-end security and privacy of their applications. Users can significantly improve the development efficiency of private computing applications and greatly shorten the time to private computing solutions.

Application Practice

Alibaba Cloud DataTrust runs the Spark SQL instance to validate the BigDL PPML solution. The basic steps are as follows:

1. Create a secure ECS instance on Alibaba Cloud

Create a g7t instance with encrypted memory as shown in Figure 5. After creation, confirm specifications of the instance in the list below:



基本信息

sgx-node-1 运行中

实例ID	[redacted]	地域	华东1（杭州）
资源组:	[redacted] default resource group	所在可用区	杭州 可用区K
公网IP	[redacted]	主机名	sgx-node-1
安全组	[redacted]	创建时间	2022年5月23日 09:47:00
描述	-	自动释放时间	-
CPU&内存	128核(vCPU) 512 GiB	云盘	1
操作系统	Ubuntu 20.04 64位 UEFI版	快照	0
实例规格	ecs.g7t.32xlarge	镜像ID	ubuntu_20_04_uefi_x64_20...
实例规格族	ecs.g7t	当前使用带宽	100Mbps （峰值）

Figure 6. ECS Instance Specifications

2. Prepare operation environment for BigDL PPML

Firstly, deploy Kubernetes cluster, Intel® SGX plugin and NFS service, obtain the docker image of BigDL PPML, and generate security keys and passwords. Secondly, perform Kubernetes security configuration, including RABC configuration and Kubernetes secret generation. Finally, start the BigDL PPML client container.

3. Run a user sample with BigDL PPML on ECS to test end-to-end security protection

Firstly, enter the BigDL PPML client container, generate appID, appKey and KMS key, and use the KMS key to encrypt the input data. Secondly, configure spark-executor-template.yaml, and place the encrypted data and KMS key in the NFS path. Finally, submit the task to the Kubernetes cluster and run the sample program.

Through the above attestation process, run industry benchmark TPC-DS based queries on Alibaba Cloud ECS g7t.32xlarge instances. The test configurations are as follows:

Taking the geometric mean of the time spent by 99 query statements as the metric, the running time of BigDL PPML based on Intel® SGX is 1.89 times that of the case without Intel® SGX protection[1].

Test data shows although there will be certain performance loss after Intel® SGX is enabled, such performance loss is within an acceptable range. Moreover, the performance loss brought by Intel® SGX is usually significantly lower than that of traditional security solutions, allowing it to save computing resources while protecting data security.

Table 1. Test Configurations

Test Configurations	
System Configurations	3-node clusters (g7t.32xlarge Alibaba Cloud ECS instance) 2x Intel® Xeon® Platinum 8369B processor, 64 cores, hyper-threading enabled, 256 GB total memory, 256 GB EPC, Ubuntu 20.04.2 LTS, 5.17.0 kernel
Software Configurations	BigDL 2.1.2-SNAPSHOT, LibOS Graphene commit 1b8848b, Spark 3.1.2, Java 1.8.0_192
Workload Configurations	TPC-DS based queries implemented by databricks' spark-sql-perf

[1] Data quoted from the test of Alibaba Cloud in June 2022. Test configurations: 3-node clusters (g7t.32xlarge Alibaba Cloud ECS instance), 2x Intel® Xeon® Platinum 8369B processors, 64 cores, hyperthreading enabled, 256 GB total memory, 256 GB EPC, Ubuntu 20.04.2 LTS, and 5.17. 0 kernels. Test the runtime with and without Intel® SGX enabled. Intel does not control or audit third-party data. You should review this content and consult other sources to evaluate accuracy.

Benefit: Driving Secure Flow of Data Value

Alibaba Cloud end-to-end PPML solution based on BigDL PPML keeps the advantages of TEE. Compared with traditional data security solutions, it provides higher level of security and data utility, at lower performance loss.

Leveraging this solution, enterprises will build an end-to-end security workflow, with protection capabilities for multiple stages of AI and big data applications, including data ingestion, data analysis, machine learning, and deep learning, to avoid security threats. Meanwhile, the solution enables higher level of data protection from the bottom layer of hardware, defending against attacks difficult to prevent by traditional security solutions, thereby reducing the leakage risks of important data.

With this solution, enterprises will provide secure data fusion services. Instead of disclosing original data, the joint analysis, training, and prediction are authorized to use logic data for application only, thus meeting the security needs of scenario-based data fusion. Business needs like autonomy, controllability and security can also be satisfied, providing customers with a transparent and controllable environment for safe circulation. It also offers easy access to and exit from the management interface, with permanent data control right. In addition, using cutting-edge security technologies and packaged for various business scenarios, this solution works well on secure circulation of enterprise data.

Below are typical application scenarios of this solution:

- **Global refined operation:** Under the premise of protecting individual privacy and data security, the brand owner builds digital and intelligent operation capabilities on all-domain data from linked platforms and third parties, optimizes the complex of consumers, products and marketplaces, and promotes business growth.
- **Joint intelligent risk control:** Keeping original data within their own environment, businesses or institutions use privacy-enhancing computing technology to realize risk control over multi-party data, improve the efficiency of risk identification, and drive healthy business growth.
- **Ads recommendation:** On the premise of protecting consumer privacy and the security of first-party and second-party data, implement joint modeling based on secure data, improve algorithm accuracy and advertising effectiveness, and promote sustainable and efficient business growth.

Summary and Prospects

With continuous introduction of laws and regulations on data security and privacy protection, it is more crucial than ever for organizations to secure the privacy of customer data. Powered by PPML, organizations will continue to explore powerful AI while reducing the risks of massive sensitive data during processing and analysis.

The BigDL PPML solution, based on Intel® SGX, BigDL, and many other security components, has created a platform to ensure data security and the performance of big data and AI. The BigDL PPML workflow was validated jointly by Alibaba Cloud and Intel. The cooperation showcased the best practice for developing end-to-end privacy-preserving applications using BigDL PPML, and demonstrated the significant role of BigDL PPML in accelerating the development of those applications. Intel and Alibaba Cloud will carry forward the achievements, and further innovate and practice end-to-end privacy protection, helping users achieve safer data fusion and accelerate to tap data value.

intel® Alibaba Cloud

Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.