China Unicom Drives Data Center Energy Saving and Emissions Reduction with Intel Intelligent Energy Management Solution

"Under the guidance of our new strategy, China Unicom has formulated the Action Plan for Building New Digital Information Infrastructure and the Action Plan for the Integration of Computing and Networking to coordinate and drive the deep integration of new data centers and cloud networks. Through cooperation with industry leaders such as Intel, we will explore technological innovation, improve management efficiency, and implement the national dual carbon policy."

– Kang Kai
Manager, China Unicom Cloud Network Operation Center

"The ongoing environmental crisis highlights the importance of reducing our carbon footprint through innovative technologies. The partnership between Intel and China Unicom on data center energy conservation and emissions reduction reflects the commitment of both companies to sustainability and provides a useful industry reference for improving the energy efficiency of servers. We hope to use the fruits of our partnership to drive high-quality industry development with an eye toward digital transformation and sustainability."

– Carl Li
Managing Director, China Enterprise and Government, Global OEM Intel Cloud and Enterprise Solution Group

Overview

China has clearly outlined the "dual carbon" goals of peak carbon by 2030 and carbon neutrality by 2060. In addition to lowering environmental pressures, the achievement of these dual carbon goals also serves as a prerequisite for sustainable development. In order to better respond to national policies, promote the sustainable development of the society and the economy, and contribute to the global response to climate change, China Unicom has successfully implemented a large number of successful technical and management measures with a focus on the construction of sustainable digital information infrastructure.

At present, China Unicom's sustainable development strategy has been widely implemented across the data center sector, effectively driving the realization of low-carbon development. China Unicom has partnered with Intel to make full use of Intel Intelligent Energy Management solution to further drive the energy conservation and emissions reduction of data centers. The solution can forecast and optimize energy consumption through software and artificial intelligence (AI) models. This improves the operational energy efficiency of data centers while meeting service-level agreements (SLAs) for business workloads without the need for any SW application changes. At present, the solution has been trialed for China Unicom big data and other scenarios. Tests have shown that the solution saves 28% more energy when compared to the baseline configuration that did not use Intel Intelligent Energy Management solution[1]. China

[1] Data from internal test results conducted by China Unicom in October 2022. Test Configuration: 2x Intel® Xeon® Gold 5318N processors (24 cores, 2.1 GHz), 512 GB total memory, 2x 960 GB SSDs, 2x 1.6 TB PCIe NVMe SSDs, 10x 12 TB SATA HDDs, 1x dual-port 1G NIC, 2x dual-port 25 GE NICs. The test was performed on OLAP – Decision Support using 240TB of data. Both Intel Intelligent Energy Management and the baseline configuration were run for an hour, during which energy consumption was recorded. As every 2.5 minutes of the test represents an actual hour, the duration of the test simulated a full day. Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Unicom and Intel also plan to further expand the application of the solution to more use case/workload scenarios while improving the sustainable development of China Unicom.

Background: Lowering Data Center Energy Consumption is the Key to Realizing the Dual Carbon Goals

With the increasing environmental and energy supply crises, more and more people are turning their attention to environmental protection. Reducing energy consumption and establishing a sustainable economy in which man and nature can develop in harmony helps reduce environmental pressure and is a prerequisite for sustainable development. China's dual carbon action plan outlines that peak carbon and carbon neutrality will be achieved during the "14th Five-Year Plan" and "15th Five-Year Plan" periods, respectively. These goals shall be realized through sustainable energy and low-carbon transformation, energy conservation, carbon reduction, and carbon peaking in the industrial sector and urban and rural construction.

Nevertheless, overall energy consumption keeps growing. A research report shows that global energy consumption increased by 5.8% in 2021[2], surpassing pre-pandemic levels as a result of accelerated economic activities. This worsens the challenges facing low-carbon development. The sustainable development of cloud and data centers is the key to realizing energy conservation and emissions reduction across society and the economy as a whole. Data shows that the energy consumption of data centers in China exceeded 200 billion kWh by the end of 2020, accounting for 2.7% of the country's total energy consumption. It is projected that this figure will reach 270 billion kWh by 2022[3].

As a leading global telecommunications operator, China Unicom outlined that it would "continuously improve the sustainability of communications network infrastructure, and help the industry reach a new level of sustainable, low-carbon, and high-quality development" in the "14th Five-Year Action Plan on Carbon Peak and Carbon Neutrality"[4] published in 2021. China Unicom seeks to minimize the energy consumption of communication cloud construction, management, and operations while ensuring the efficient development of communication services such as 5G.

To reduce the energy consumption of data centers, the industry generally aims to optimize power usage effectiveness (PUE) of data centers and reduce the energy consumption of auxiliary equipment such as air conditioners. The problem is that PUE becomes bottlenecked when reduced to a certain level, while certain solutions aimed at reducing PUE are complex and require large amounts of deployment and construction labor. In addition to extending the construction period, this also results in a significant increase to the total cost of ownership (TCO).

In consideration of these factors, China Unicom tried to further promote energy conservation and emissions reduction through data center IT equipment. To realize this transformation, it would be necessary to enhance the precision of CPU energy consumption adjustments and improve the energy efficiency of the core piece of data center IT equipment, the servers themselves. But this strategy also faces a variety of challenges.

- Although CPU processes and technologies are constantly improving, the higher demands of data centers workloads for computing power and power density have caused the overall thermal design power (TDP) of CPU chips to continue to rise. This poses a significant challenge to the overall energy consumption of servers.
- Traditional CPU energy consumption control solutions typically adopt flexible orchestration to realize partial redundant host hibernation, deep energy conservation of idle cores, and regulation of light load cores. These solutions typically adopt a static adjustment strategy, which cannot be flexibly adjusted and configured in accordance with operational demands. The host and kernel are also typically shut down as a whole, making it difficult to realize finer-grained control and resulting in unsatisfactory energy consumption reduction.
- Traditional CPU energy consumption control solutions cannot precisely gain insights into or project CPU workloads and cannot effectively meet SLAs while accurately controlling energy consumption.

Solution: Using Intel Intelligent Energy Management Solution to Realize the Energy Saving and Emissions Reduction of Data Centers

China Unicom and Intel began their partnership in 2021 to further drive the sustainable development of data centers. The partners formulated a comprehensive data center energy conservation and emissions reduction plan. Starting from 5G core networks, they have gradually expanded into sectors such as big data, and continuously driven the implementation of smart energy conservation solutions. The partners are focusing the current stage of their cooperation on the energy conservation and emissions reduction of servers. They have adopted Intel Intelligent Energy Management solution to forecast and control energy consumption in a more accurate and intelligent manner, improving the operational energy efficiency of data centers while meeting service-level agreements (SLAs) for business workloads.

Intel Intelligent Energy Management

Intel Intelligent Energy Management is an energy conservation and emissions reduction solution empowered by Intel AI solutions and server platform technologies. The solution is capable of forecasting and optimizing energy consumption through AI models and software solutions to improve the operational energy efficiency of data centers. The solution uses components such as intelligent telemetry, Chronos time series data analysis, Intel® Xeon® platform-level energy efficiency controls, and Container Runtime Interface Resource Manager (CRI-RM) dynamic resource management policies to effectively meet SLAs for business workloads without the need for any application changes.

[2] https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html, last accessed in November 2022.

[3] Data from the National Energy Administration.

[4] https://baijiahao.baidu.com/s?id=1702806555402999078&wfr=spider&for=pc, last accessed in November 2022.

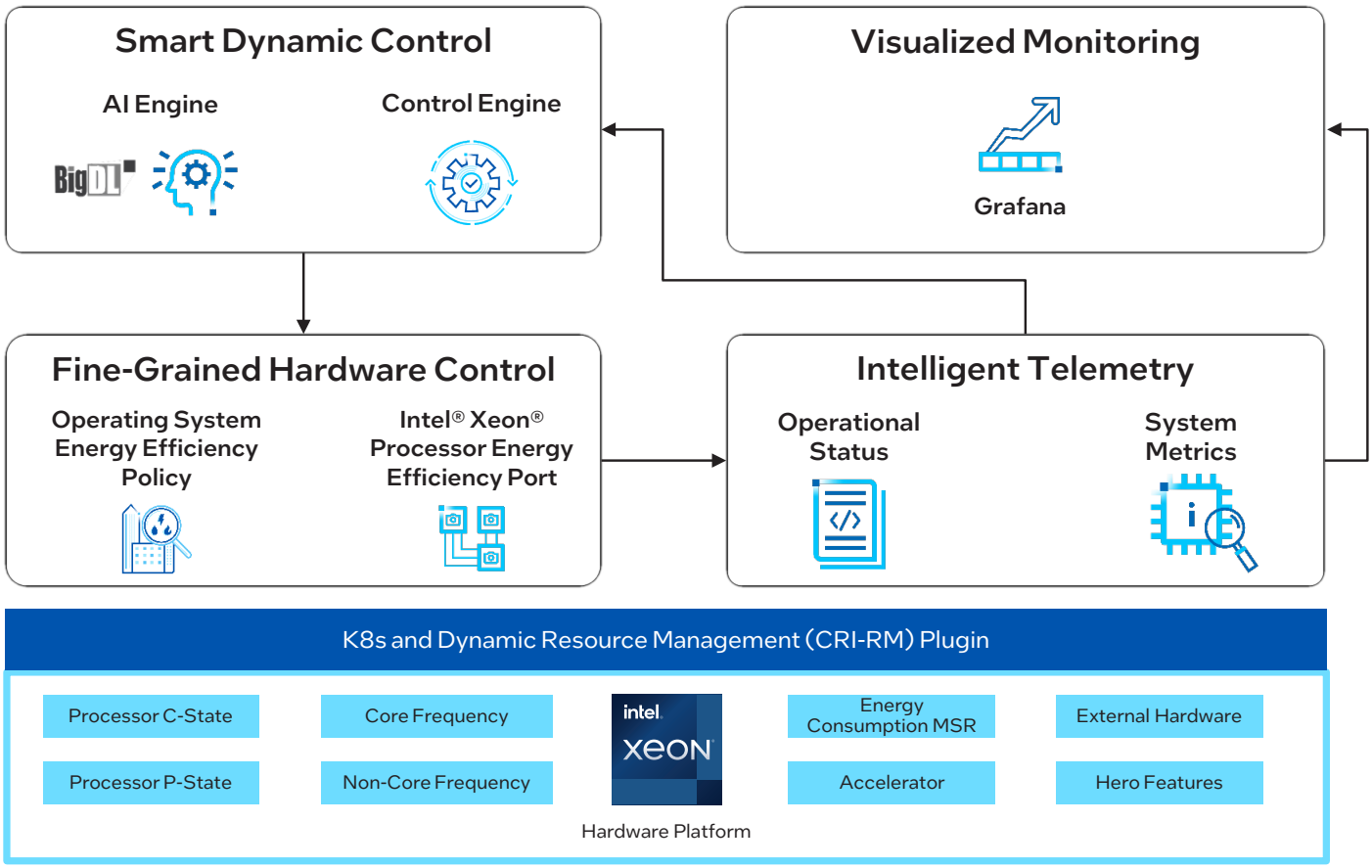

Figure 1. Intel Intelligent Energy Management solution architecture

Intelligent Telemetry

Intel Intelligent Energy Management adopts the Prometheus open-source monitoring system. The Prometheus cloud-native standard telemetry tool is used to collect and aggregate metrics as time series data. Prometheus provides the collectd system performance collection daemon, which can publish data in a number of ways. Intel has also developed high-performance Python plugins to collect additional metrics to further improve performance and reduce overhead.

Chronos-based Time Series Data Analysis

The Chronos framework is derived from BigDL, Intel's open-source unified big data analysis and AI platform. Chronos empowers model training with the ability to train regression models with full parameters, automatically analyze and extract key parameters, and limit predictive model training to key parameters only. In model inference, the Chronos framework can forecast workloads, identify better control parameters in query networks, and apply better control parameters through CRI-RM webhooks.

Intel Xeon Platform-Level Energy Efficiency Controls

Intel Intelligent Energy Management provides Intel Xeon platform-level power control options, integrating energy regulation at the operating system level, CPU Turbo, drivers, EEP control, SAPM control, dynamic switching of processor energy efficiency, and P\C\S state adjustment to provide better configurations for different scenarios. The solution provides fine-grained hardware control knobs to dynamically switch and control a variety of processor algorithms.

Container Runtime Interface-based Resource Manager (CRI-RM) Dynamic Resource Management Policies

This feature can dynamically divide system resources across nodes and work with the Kubernetes scheduler to realize logical job scheduling at the node level, and efficiently adapt the features of the Intel platform to the Kubernetes cluster environment. The CRI-RM Balloon Policy supports the control of CPU core frequency and non-core frequency. Users can create CPU resource pools based on the CRI-RM Balloon Policy, define balloon types for different types of workloads, and dynamically change CPU classes and balloon size configurations.

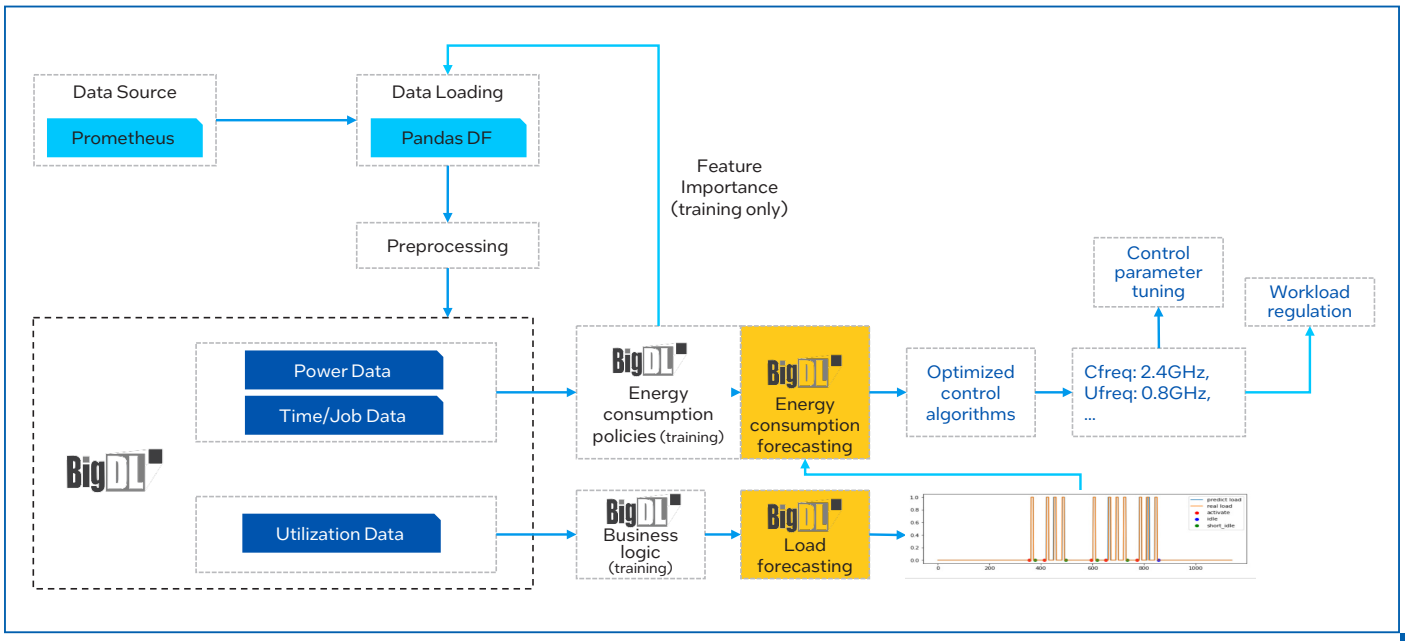

Figure 2. Intel Intelligent Energy Management AI control implementation architecture

Intel Intelligent Energy Management solution is capable of performing intelligent telemetry on a variety of Intel Xeon Scalable processor operating indicators. In addition to providing telemetry support, the solution inputs telemetry data into closed-loop control logic and realizes fine-grained control at the operating system, processor energy efficiency, and processor energy state levels to provide enhanced energy efficiency control for a variety of scenarios. Intel Intelligent Energy Management can be easily expanded from single-node closed-loop control to cluster closed-loop control (as shown in Fig. 1) to improve the overall energy efficiency of data centers.

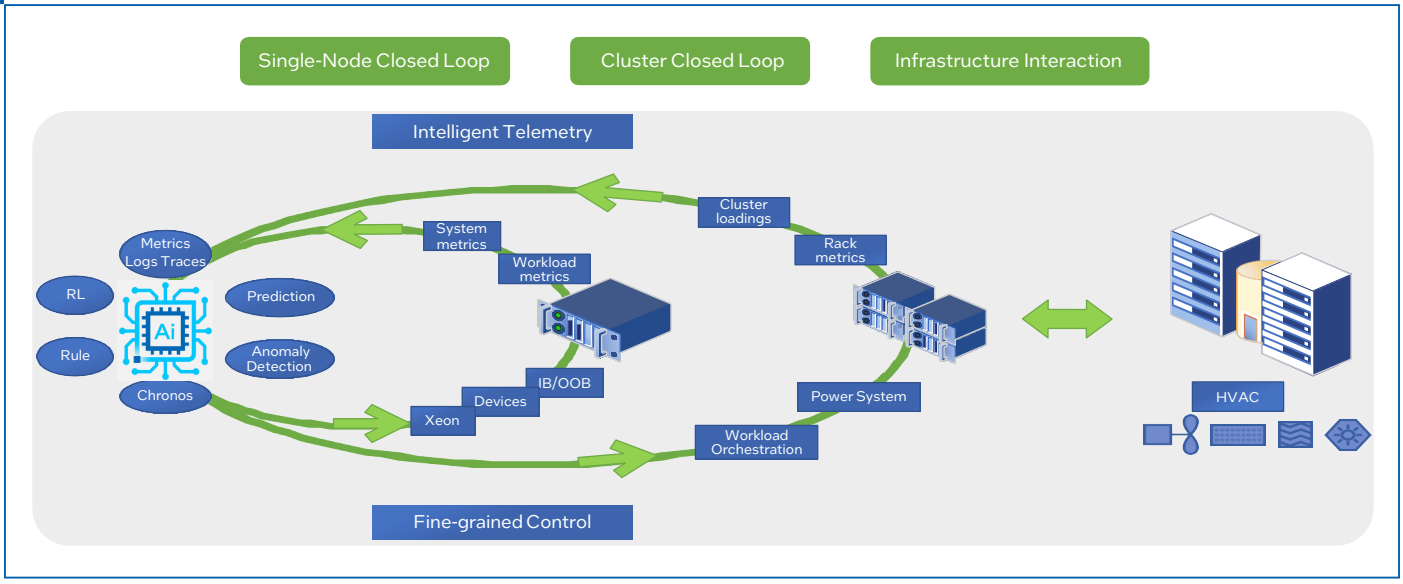

Figure 3. Intel Intelligent Energy Management solution

Compared with traditional server energy conservation solutions, Intel Intelligent Energy Management provides more intelligent peak and valley forecasting, puts more redundant hosts to sleep during downtime, and significantly improves energy conservation. Intel Intelligent Energy Management also provides smart regulation in accordance with SLAs for business workloads, enabling more fine-grained control and power control vectors to achieve a better balance between SLA requirements and energy conservation.

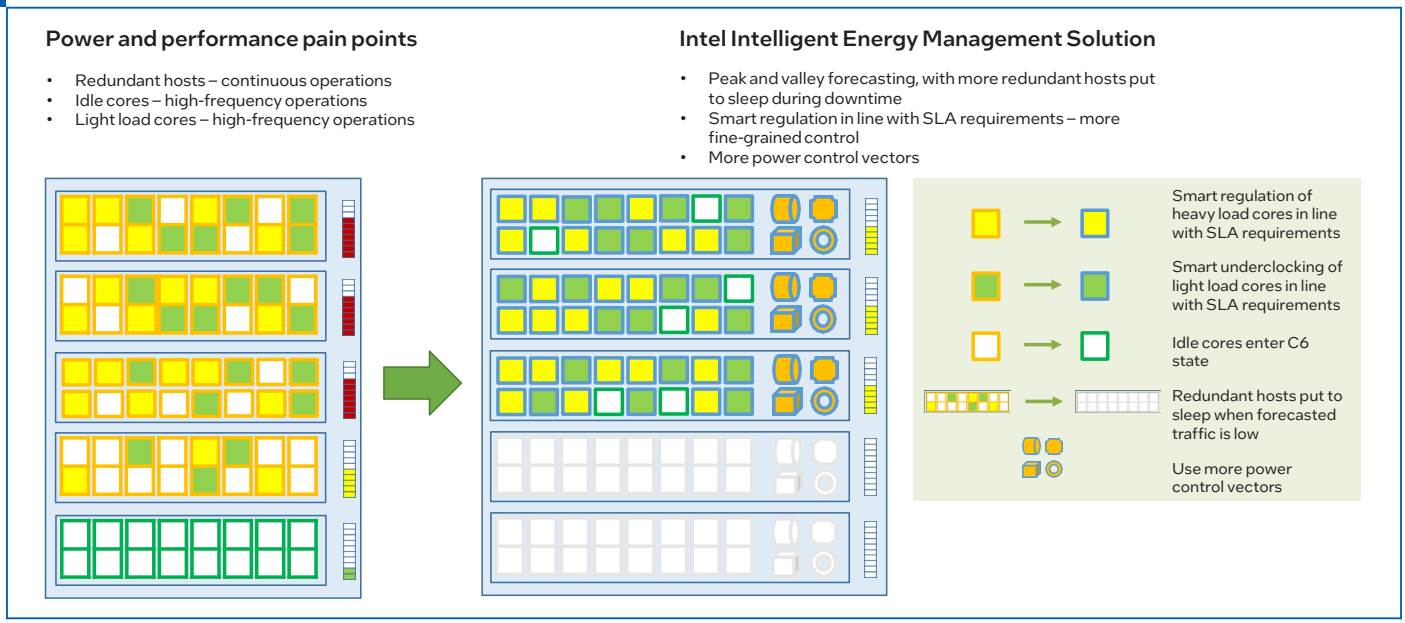

Figure 4. Intel Intelligent Energy Management provides smarter and more dynamic power control

China Unicom Adopts Intel Intelligent Energy Management Solution to Optimize Data Center Energy Consumption

China Unicom analyzed the operational data and processor usage of 5G network elements, big data, and other services of its communications cloud resource pool. The analysis shows that volume differs with time of day and is characterized by peaks and valleys. During service, the operational status of the server can be adjusted to realize dynamic energy conservation and emissions reduction by detecting changes in service workloads and quality indicators in real time.

Supported by the rich components and integrated optimization strategies of Intel Intelligent Energy Management, China Unicom worked to improve energy conservation and emissions reduction from the following aspects:

- During server operation, the components supported by Intel Intelligent Energy Management are used to perform intelligent telemetry on a variety of operating states and provide relevant information for closed-loop control logic.
- Modeling based on operational data. Intel Intelligent Energy Management APIs can be used to rapidly perform automated feature generation, along with operations such as scaling and filling the gaps in time series data.
- Realize hyperparameter search, retrieve higher-quality sets of hyperparameters according to the forecast target, optimize the model and data processing, and develop a time series projection model.
- Use the model to perform inference (or performance evaluation and optimization) on real-time operational data to obtain final processor usage forecasting data.
- Realize AI control based on projection data to enable C/UFreq that closely follows workload changes, meeting the SLA requirements of business workloads while improving energy conservation.

During the early stage of their partners, China Unicom and Intel tested the 5GC network element application scenarios. The partners are now working on the testing and deployment of solutions for big data services. The workloads of China Unicom's big data operations show clear fluctuations over time, with large differences between peaks and valleys. Adopting energy conservation measures during valleys is projected to have significant effects.

Based on the identified characteristics, China Unicom designed an energy conservation prototype based on Intel Intelligent Energy Management. The prototype starts with batch scenarios. Energy conservation is realized through peak and valley load characteristics, defaulting to high-performance configuration during peaks, automatically switching to low-power configuration during valleys, and configuring the CPU cfreq and ufreq as required.

Test data shows that Intel Intelligent Energy Management energy conservation solution saves 28.6% more energy than the baseline configuration that did not use Intel Intelligent Energy Management solution[5]. This is largely due to the solution's support for adaptive cfreq scaling. As the baseline configuration remains at high frequencies, Intel Intelligent Energy Management consumes far less power.

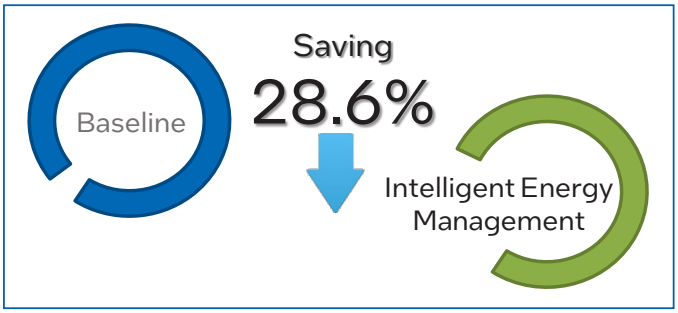

Figure 5. Test data comparison[6]

If extrapolated to the entire cloud resource pool, the solution has the potential to save millions of kWh of electricity every year. When combined with the energy conservation and consumption reduction provided by existing data center PUE solutions, it has the potential to reduce CO_2 emissions by thousands of tons per year.

Outlook

The China Unicom's energy conservation and emissions reduction strategy empowered by Intel Intelligent Energy Management has achieved the expected results. With the goal of realizing energy conservation without modifying hardware, the parties have significantly reduced the energy consumption of 5GC network elements and servers in big data services, all the while meeting the SLA requirements of business workloads. Through the combination of server energy saving and emissions reduction with other energy conservation measures, China Unicom enhances data center sustainability at a larger scale and facilitates the construction of sustainable data centers. Intel has also released the "Intel Technology Framework for Sustainable Data Centers" to jointly drive the green and sustainable development of data centers with customers and the industry chain as a whole.

On the basis of their current achievements, China Unicom and Intel plan to further enhance energy conservation by expanding their cooperation in the following aspects:

- Expand Intel Intelligent Energy Management solution to more scenarios to further China Unicom's realization of the dual carbon goals.
- Expand the scope of energy conservation to cover networks, external hardware, data center environments, and other areas while minimizing overall data center energy consumption.
- Take advantage of the performance core and efficient core policies of next-gen Intel processors to efficiently schedule cores based on workload to realize the ideal balance between performance and energy consumption.
- Precisely control CPU energy consumption while driving the application of innovative energy conservation technologies in more accelerators (such as GPUs).
- Realize the application of innovative technologies such as real-time reinforcement learning, automated machine learning (AutoML), and reinforcement learning to further improve energy efficiency.

Through cooperation and exploration in these areas, Intel will help China Unicom formulate a more efficient and intelligent data center energy consumption management strategy, significantly enhance energy conservation and emissions reductions, and implement an action plan for constructing faster and better new digital information infrastructure.

intel®

[5,6] Data from internal test results conducted by China Unicom in October 2022. Test Configuration: 2x Intel® Xeon® Gold 5318N processors (24 cores, 2.1 GHz), 512 GB total memory, 2x 960 GB SSDs, 2x 1.6 TB PCIe NVMe SSDs, 10x 12 TB SATA HDDs, 1x dual-port 1G NIC, 2x dual-port 25 GE NICs. The test was performed on OLAP – Decision Support using 240TB of data. Both Intel Intelligent Energy Management and the baseline configuration were run for an hour, during which energy consumption was recorded. As every 2.5 minutes of the test represents an actual hour, the duration of the test simulated a full day. Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.