Implementation of seamless assistance with Google Assistant leveraging cloud computing

Jiaxin Huang[1,*], Yifan Zhang[2], Jingyu Xu[3], Binbin Wu[4], Bo Liu [5], Yulu Gong[6]

[1]Information Studies, Trine University, Phoenix USA

[2]Executive Master of Business Administration, Amazon Connect Technology Services (Beijing) Co., Ltd. Xi'an, Shaanxi, China

[3]Computer Information Technology, Northern Arizona University, Flagstaff, AZ, USA

[4]Computer Network Engineering, Cisco Systems, Beijing, China

[5]Software Engineering, Zhejiang University, HangZhou China

[6]Computer & Information Technology, Northern Arizona University, Flagstaff, AZ, USA

*Corresponding author: jiaxinhuang1013@gmail.com

Abstract. AI and cloud native are mutually reinforcing and inseparable. Due to the huge storage and computing power requirements, most AI applications need cloud support, especially large model applications If cloud native has influenced the software industry to a considerable extent in the past few years, the big model boom means that cloud native has become a standard option for developers.This paper describes the rise of AI model applications and their integration with traditional development workflows, pointing out the challenges that enterprises and developers face when integrating large models. With the rise of cloud-native technologies, the combination of artificial intelligence and cloud computing is becoming increasingly important. Cloud-native technologies provide the infrastructure needed to build and run resilient and scalable applications, while distributed infrastructure supports multi-cloud integration, enabling a unified foundation of "one cloud, multiple computing." As an intelligent voice Assistant, Google Assistant achieves a more intelligent, convenient and efficient user experience through applications in smart home control, enterprise customer service and healthcare. Finally, this paper points out the advantages of combining Google Assistant with cloud computing, providing a more intelligent, convenient, and efficient user experience.

Keywords: Artificial intelligence model application, Cloud-native technology, Google Assistant, User experience.

1. Introduction

Developers who are attempting to build software solutions with large models fall into two main categories. There are professional developers who are familiar with machine learning techniques and can train, fine-tune and deploy large models on their own. However, these developers face hardware infrastructure challenges, and training a large model like GPT-3.5 requires a huge GPU cluster to run for a long time. Another type of developer needs to eliminate hardware computing power constraints

and focus on the actual development and integration of model applications. In this context, the combination of AI and cloud-native technologies is increasingly important. Due to the huge demand for storage and computing power in large models, the vast majority of AI applications need to rely on cloud support especially for large model applications.. As a result, cloud-native technology has become a standard option for developers, and developing, deploying, and operating AI applications will become a standard skill and workflow for every developer. With the historical turning point of personal computer and mobile Internet innovation, the software industry is once again standing at a new historical starting point.

2. Related work

2.1. Cloud-native

The concept of cloud computing was first proposed by Dell in 1996. In 2006, Amazon took the lead in launching Elastic Compute Cloud (EC2) [1]services, and then more and more enterprises began to gradually accept the concept of cloud computing, and gradually migrate applications to the cloud, and enjoy the technical dividends brought by this new computing method. Cloud-native technologies provide the necessary infrastructure to build and run elastically scalable applications for a variety of cloud environments, including public, private, and hybrid clouds. Representative cloud-native technologies include containers, service grids, microservices, immutable infrastructure, and declarative apis that help build loosely coupled systems that are fault-tolerant, easy to manage, and easy to observe. Combined with automation, cloud-native technology enables engineers to easily cope with frequent and predictable major changes to their systems. Cloud computing provides sufficient computing and storage resources to enable intelligent assistants such as Siri, Alexa, and Google Assistant to perform complex tasks such as natural language processing, speech recognition, and machine learning, thereby greatly facilitating everyday life. Therefore, combining cloud-native technology with intelligent assistants can not only enhance the user experience, but also bring more efficient development and operation processes for enterprises.

2.2. Cloud-native distributed infrastructure

The cloud computing infrastructure architecture takes distributed multi-cloud as the core, builds the "one cloud and multiple computing" converged base, relies on the unified management of heterogeneous resources and the distributed task collaboration framework, builds a new service system with AI running through it, supports the integrated carrying capacity of general computing, intelligent computing, supercomputing, and network convergence services, and ensures the availability of full-link services. In terms of overall architecture, the hierarchical system of traditional cloud architecture is retained. In the cloud network resource construction, it emphasizes the distributed optimal layout of multiple types of resource pools. Diversity is emphasized in the software and hardware resource layer, which is further divided into CPU-based general computing infrastructure and intelligent computing infrastructure dominated by AI-accelerated chips such as [2]GPU. The distributed cloud platform manages multi-dimensional heterogeneous resources in a unified manner and implements efficient collaborative task scheduling. On the basis of infrastructure architecture, cloud service forms show a trend of generalization and intelligent development, carrying multiple business types, and providing rich industrial digital capabilities.

Because search engines need to process huge amounts of data, Google's two founders, Larry Page and Sergey Brin, designed a file system called "BigFiles" in the early days of the company. GFS(full name "Google File System") this distributed file system is a continuation of [3]"BigFiles". First, to introduce its architecture, GFS is divided into two main types of nodes:

1.Master node: mainly stores metadata related to data files, rather than chunks. Metadata includes a table that maps 64-bit labels to the location of the data block and its constituent file, the location of the copy of the data block, and which process is reading and writing the particular data block. The Master

node also periodically receives updates (" Heart-beat ") from each Chunk node to keep the metadata up to date.

2.Chunk node: As the name implies, it must be used to store Chunk. Data files are divided into chunks with a default size of 64MB. Each Chunk has a unique 64-bit label.

Among them, the distributed infrastructure also has intelligentization oriented functions. Due to the low degree of interaction among heterogeneous resources, high matching complexity, difficult to ensure the balance after adjustment, insufficient consideration of service characteristics and other problems, distributed infrastructure has the following functions: Therefore, the resource supply mode changes from providing fixed computing resources to flexibly adjusting the resource usage for specific service scenarios.

2.3. Cloud-native Core feature

The core characteristics of cloud infrastructure are wide distribution, high performance, and hyperscale:

(1) Widely distributed cloud network resources. Relying on the distributed cloud architecture, it can achieve extensive coverage of near-global infrastructure from the cloud resource pool of service providers and local cloud resource pools to the production site. To provide comprehensive connectivity and highly reliable network protection, integrating connectivity across air, sky, and sea.; Provide consistent services in different geographic resource pools, and provide one-click cloud network resource supply anytime, anywhere.

(2) High-efficiency hardware resource supply. Based on the green advanced multi-component computing power, more than ten times the computing performance improvement. Build a new type of intensive and efficient storage to provide mass storage requirements under the digital wave. Promote the innovation of system-level network disconnection cooperation system, and build a low-consumption high-speed interconnection network between 100,000 nodes[4].

(3) Ultra-large scale management and scheduling. The scale of data management and control continues to increase, providing multi-modal data management and scheduling of PB-level large data volumes; Support complex business requirements logic cumbersome, frequent interaction of modular management, to achieve complex business logic-oriented management scheduling; Massive data and highly complex algorithms drive the cloud platform to achieve unified management and control of more large-scale computing power.

A variety of learning methods based on AI models and even large models can design and generate adaptive intelligent planning and scheduling policies based on training models for business needs to improve the scheduling superiority of large-scale resources. Multi-level scheduling of Services Level Indicator (SLO), service level indicator (SLI), topology-aware scheduling, and off-line business mixing to maximize resource utilization.

2.4. Google Assistant

Google Assistant is an intelligent voice assistant developed by Google, which can parse and process voice commands uploaded by smart terminals. Google Home speaker is a smart speaker product launched by Google equipped with Google Assistant voice assistant. Speaker built-in Wi-Fi, Bluetooth and NFC communication, through Wi-Fi connected to the network to achieve Google Assistant services, Bluetooth and NFC[5] can realize the connection with other devices, expand the application of the speaker. Built-in 2 microphone array, the use of beamforming technology, while using noise reduction algorithm, to ensure that the speaker can activate the speaker in a noisy environment to connect Google Assistant for semantic recognition. GoogleAssistant to achieve smart Home control mainly relies on the Home Graph. HomeGraph is essentially a logical map of the home. Home Graph can pass the above information to Google Assistant so that it can execute the user request based on the corresponding before and after state.

The solution architecture of intelligent air conditioning products is generally composed of "intelligent air conditioning + cloud service + mobile APP", [6-7]in which the cloud service stores various attribute parameters of intelligent air conditioning, and the cloud service communicates with intelligent air

conditioning through the Internet to ensure the real-time property parameters of intelligent air conditioning on the cloud service. The appliance APP can control the appliance by changing the attribute parameters on the cloud service. The solution of intelligent air conditioning access to Google Assistant voice control is to connect the intelligent air conditioning cloud service with Google Assistant to transmit the relevant control and status parameters of intelligent air conditioning to achieve the voice control of intelligent home appliances. The system scheme is shown in Figure 1.
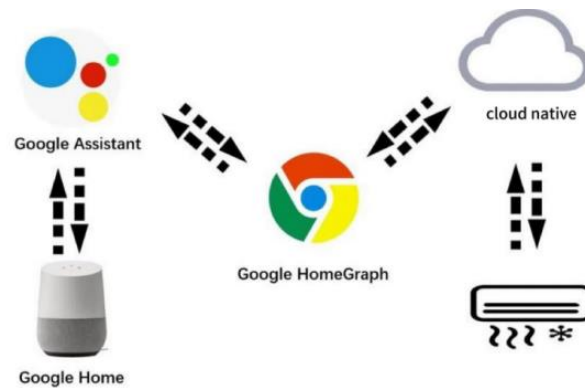

Figure 1. GoogleAssistant service implementation architecture diagram

Integrating this part of the content, we can see that cloud native technology provides the infrastructure needed to build and run elastic and scalable applications, while distributed infrastructure is multi-cloud as the core, establishing the integration basis of "one cloud and multiple computing" and supporting AI-powered service systems. [8-10] The implementation of this architecture diagram will provide the basis for us to further explore the actual application scenarios of Google Assistant.

3. Google Assistant application

Intelligent voice robots can provide intelligent customer service, intelligent outbound call, intelligent voice quality inspection and other services[11]. Through the intelligent voice system, it integrates with the existing customer service system and outbound call system of the enterprise, and realizes various functions including employee work order assistant, work order intelligent distribution, work order automatic troubleshooting, and judging whether the user can currently troubleshoot according to the established process.
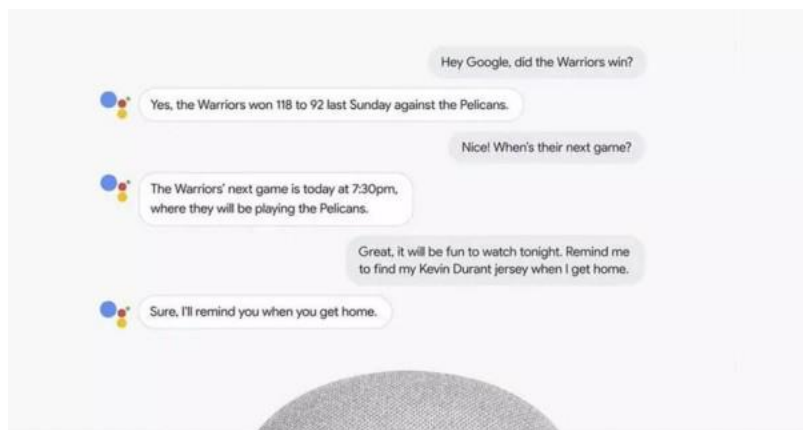

Figure 2. Logical thinking of Google Assistant voice assistant

3.1. Google Assistant search aids ecological construction

Google Assistant from the simple search assistance of Google Now to the ecological construction of today, reflecting Google's significant progress in the field of artificial intelligence. In the eyes of consumers, the functions of Google Assistant have become more comprehensive, and the understanding of voice commands and semantic recognition are more accurate. However, the deeper meaning behind this change is that Google is building an entire AI ecosystem around Google Assistant. In the development of artificial intelligence, algorithms, calculations and data are considered to be the three major elements. Google has solved the problem of algorithms and computing by open-source TensorFlow and numerous algorithms, as well as the development of Tpus[12]. Data is the only dividing line. By connecting different functions and applications, Google Assistant builds an AI ecosystem centered on user habits. This ecosystem enables different services to be related to each other, providing users with smarter services. In addition, Google has partnered with major products including Product Hunt, Food Network, LinkedIn, Uber, [13]IFTTT, and Spotify to integrate them into Google Assistant, further expanding its capabilities, To provide users with more comprehensive and convenient intelligent assistant services.

3.2. Google Assistant in different areas of application

First of all, the application of Google Assistant in the field of smart homes can be described as the most prominent. Through the intelligent voice assistant, users can control various smart devices in the home through simple voice commands, such as smart lamps, smart sockets, smart door locks, etc. This intelligent home control greatly improves the convenience and comfort of users' lives. Secondly, in the business field, Google Assistant is widely used in customer service and enterprise office scenarios. Enterprises can use Google Assistant's voice recognition and natural language processing capabilities to develop intelligent customer service robots for automatically answering customer inquiries, processing orders and other tasks, thereby improving customer service efficiency and user experience. Finally, in the field of health care, Google Assistant also plays an important role. Medical institutions can use Google Assistant to develop intelligent health assistants to provide medical and health information, appointment registration, drug consultation and other services. [14]At the same time, individual users can also obtain personalized health services such as health management advice and sports fitness guidance through Google Assistant, so as to better pay attention to and manage their health status. The application of this intelligent health assistant helps to improve the efficiency of medical services and the level of personal health management.

Overall, Google's new smart Assistant announcement could have a big impact on how we use tech products and services. Making voice commands easier and faster could change the way we interact with our devices in the same way that smartphones, led by Apple's iPhone, went mainstream more than a decade ago and ushered in the era of touch screens. Perhaps we can see this as the first step towards a world where humans are constantly in dialogue with inanimate objects.

3.3. Google Assistant Benefits

Google Assistant, as an intelligent voice assistant, has many advantages. First, Google Assistant enables natural language understanding and interaction, where users can ask questions or issue commands in natural language without paying too much attention to the syntax or writing of instructions. Second, Google Assistant integrates a wealth of functions and services, can provide users with a variety of practical functions such as search, schedule management, music playback, smart home control and so on, greatly facilitating the daily life of users. [15-16] Combined with cloud computing, Google Assistant achieves even more advantages. First, the cloud provides powerful computing and storage resources that enable Google Assistant to handle complex natural language processing, speech recognition, and machine learning tasks. In this way, Google Assistant can more accurately understand the user's instructions and intentions, and provide more accurate services and recommendations. Second, the elasticity and flexibility of cloud computing enables Google Assistant to serve users anytime, anywhere, whether on mobile devices or smart home devices. In addition, cloud computing also provides Google

Assistant with high availability and scalability, ensuring its stability and performance continue to improve. In summary, the combination of Google Assistant and cloud computing achieves a more intelligent, convenient and efficient user experience, providing users with a new way of voice interaction.

4. Conclusion

The combination of cloud computing and artificial intelligence will continue to show great potential in the future. Cloud computing provides powerful computing and storage resources for AI, enabling intelligent assistants like Google Assistant to perform complex natural language processing, speech recognition, and machine learning tasks to provide users with more accurate services and recommendations. By combining with cloud computing, Google Assistant can achieve higher availability and scalability, ensuring its stability and performance continue to improve. Therefore, the application prospect of cloud computing in the field of artificial intelligence is very broad, and it will bring users a smarter, more convenient and more efficient experience. With the continuous development and popularization of artificial intelligence technology, people's demand for intelligent assistants will further increase. In the future, intelligent assistants will not only be voice-controlled home devices or tools to provide daily services, but more likely to become intelligent partners in People's Daily life and work. Therefore, artificial intelligence assistants will become an important way of human-computer interaction in the future, bringing more convenience and fun to people's lives and work.

References

[1] Antonopoulos, Nick, and Lee Gillam. Cloud computing. Vol. 51. No. 7. London: Springer, 2010.

[2] Gong, Chunye, et al. "The characteristics of cloud computing." 2010 39th International Conference on Parallel Processing Workshops. IEEE, 2010.

[3] Furht, Borivoje, and Armando Escalante. Handbook of cloud computing. Vol. 3. New York: springer, 2010.

[4] Zhang, Qi, Lu Cheng, and Raouf Boutaba. "Cloud computing: state-of-the-art and research challenges." Journal of internet services and applications 1 (2010): 7-18.

[5] Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing— The business perspective. Decision support systems, 51(1), 176-189.

[6] Tulshan, Amrita S., and Sudhir Namdeorao Dhage. "Survey on virtual assistant: Google assistant, siri, cortana, alexa." Advances in Signal Processing and Intelligent Recognition Systems: 4th International Symposium SIRS 2018, Bangalore, India, September 19–22, 2018, Revised Selected Papers 4. Springer Singapore, 2019.

[7] Huang, Zengyi, et al. "Research on Generative Artificial Intelligence for Virtual Financial Robo-Advisor." Academic Journal of Science and Technology 10.1 (2024): 74-80.(12 个)

[8] Huang, Zengyi, et al. "Application of Machine Learning-Based K-Means Clustering for Financial Fraud Detection." Academic Journal of Science and Technology 10.1 (2024): 33-39.

[9] Xu, Z., Gong, Y., Zhou, Y., Bao, Q., & Qian, W. (2024). Enhancing Kubernetes Automated Scheduling with Deep Learning and Reinforcement Techniques for Large-Scale Cloud Computing Optimization. arXiv preprint arXiv:2403.07905.

[10] Xu, X., Xu, Z., Ling, Z., Jin, Z., & Du, S. (2024). Comprehensive Implementation of TextCNN for Enhanced Collaboration between Natural Language Processing and System Recommendation. arXiv preprint arXiv:2403.09718.

[11] López, Gustavo, Luis Quesada, and Luis A. Guerrero. "Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces." Advances in Human Factors and Systems Interaction: Proceedings of the AHFE 2017 International Conference on Human Factors and Systems Interaction, July 17– 21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8. Springer International Publishing, 2018.

[12] Berdasco, A., López, G., Diaz, I., Quesada, L., & Guerrero, L. A. (2019, November). User experience comparison of intelligent personal assistants: Alexa, Google Assistant, Siri and Cortana. In Proceedings (Vol. 31, No. 1, p. 51). MDPI.

[13] Gannon, Dennis, Roger Barga, and Neel Sundaresan. "Cloud-native applications." IEEE Cloud Computing 4.5 (2017): 16-21.

[14] Song, B., Xu, Y., & Wu, Y. (2024). ViTCN: Vision Transformer Contrastive Network For Reasoning. arXiv preprint arXiv:2403.09962.

[15] Zhang, Chenwei, et al. "Enhanced User Interaction in Operating Systems through Machine Learning Language Models." arXiv preprint arXiv:2403.00806 (2024)

[16] Ni, Chunhe, et al. "Enhancing Cloud-Based Large Language Model Processing with Elasticsearch and Transformer Models." arXiv preprint arXiv:2403.00807 (2024).