

HCIA Big Data-Introduction

Eng. Hagar Muhammed

Lecture 1



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

content

- Big Data definition.
- Challenges Faced by Traditional Data Processing.
- The opportunities of big data.
- Big data applications.
- Kunpeng computing.
- Summary.

Big Data Era

Thermal
energy



Steam Age
1760s - 1840s
United Kingdom

Electricity



Electric Age
1860s - 1920s
Europe and United
States

Computers and
communications



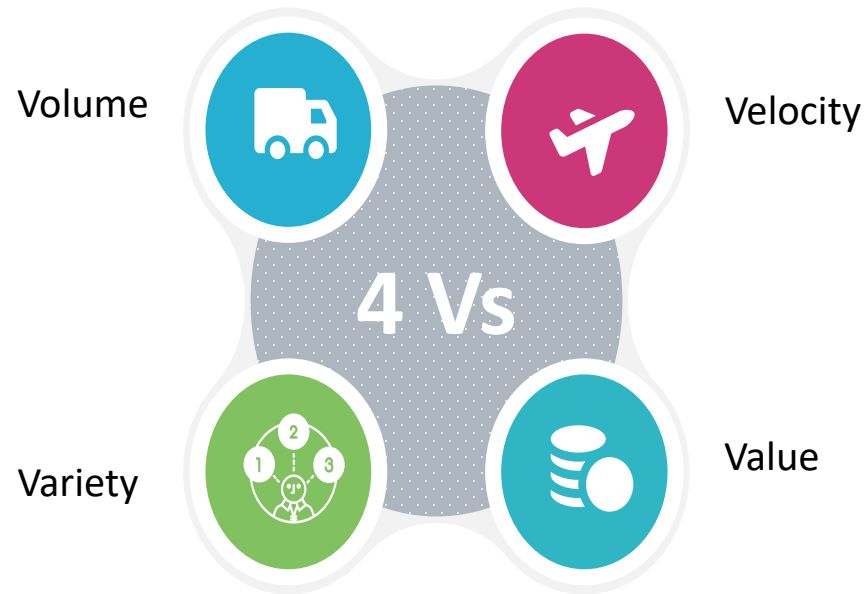
Information Age
1940s - 2010s
United States

Cloud, Big data, IoT,
Intelligence



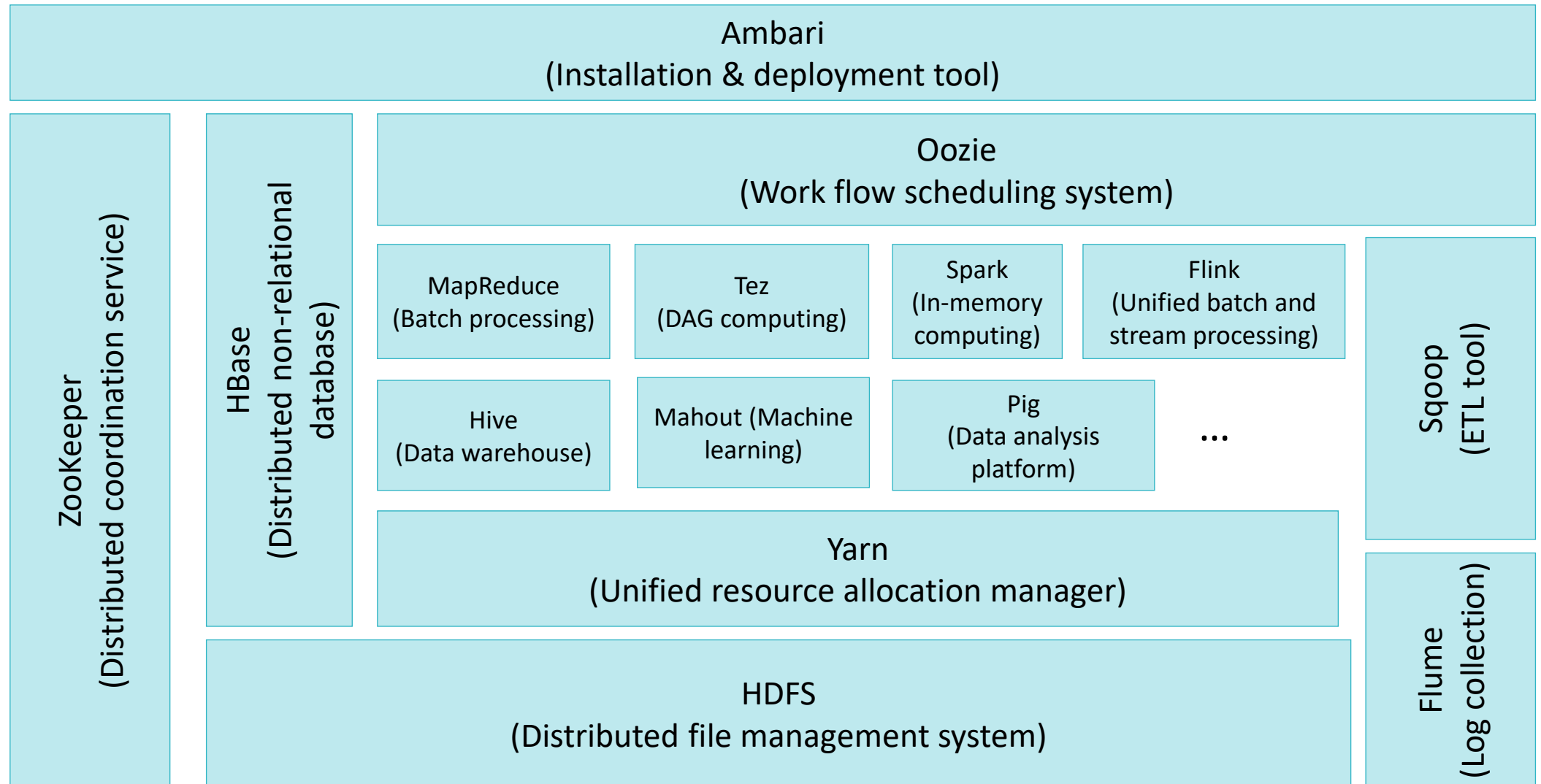
Intelligent Era
2010s - ?

- Big data: refers to datasets that are too large or complex to be captured, managed, and dealt with by traditional data-processing application software.



	Big Data Processing	Traditional Data Processing
Data scale	Large (in GB, TB, or PB)	Small (in MB)
Data type	Multiple (structured, semi-structured, and unstructured data)	Single (mainly structured data)
Mode-data relationship	Modes are set after data is generated. Modes evolve as data increases.	Modes are set before data is generated (ponds are built before fish are put in).
Scalability	Highly scalable architecture for handling large-scale data	Limited scalability, suitable for smaller datasets
Tool	No size fits all	One size fits all
Data Quality	Data quality may vary (incomplete, noisy, inconsistent)	Assumed to have high data quality and consistency
Storage	Distributed file systems (Hadoop, HDFS)	Relational databases (SQL-based systems)
Processing Model	Parallel and distributed processing	Sequential processing

Hadoop Big Data Ecosystem



Main Computing Modes of Big Data Applications

- **Batch computing**
 - Processes massive data in batches. Major technologies include MapReduce and Spark.
- **Stream computing**
 - Calculates and processes streaming data in real time. Major technologies include Spark, Storm, Flink, Flume, and Dstream.
- **Graph computing**
 - Processes large-scale graph structure data. Major technologies include GraphX, Gelly, Giraph, and PowerGraph.
- **Query and analysis computing**
 - Storage management and query analytics of massive data. Main technologies include Hive, Impala, Dremel, and Cassandra.

Challenges Faced by Traditional Data Processing

Costly storage of massive
data volumes

Insufficient batch processing
performance

Lack of streaming data
processing

Limited scalability

Single data source

External value-added data
assets

Challenge 1: Indefinite Big Data Requirements

- Many enterprise business departments do not understand the application scenarios and values of big data, making it difficult to raise accurate requirements for big data. Since the requirements are not clear and the big data department is a non-profit department, the enterprise decision-makers are worried about the low input-output ratio. They are hesitant to set up big data departments, or even delete a large amount of valuable historical data due to the lack of application scenarios.

Challenge 2: Data Silos

- Data fragmentation is the most important challenge for enterprises when implementing big data.
 - In a large enterprise, different types of data are scattered in different departments. As a result, data in the same enterprise cannot be shared, and the value of big data cannot be brought into full play.

Challenge 3: Low Data Availability and Poor Quality

- Many large- and medium-sized enterprises generate a large amount of data on a daily basis, but they do not pay much attention to the preprocessing phase of big data. As a result, data processing is not standard. In the big data preprocessing phase, valid data is obtained after extraction, conversion, cleaning and denoising. Sybase data shows that if the availability of high-quality data is improved by 10%, enterprise profits can increase by more than 20%.

Challenge 4: Data-related Management Technologies and Architectures

- Traditional databases are not suitable for processing PB-level data.
- Traditional databases do not support data diversity, especially compatibility with structured, semi-structured and unstructured data.
- Massive data O&M needs to ensure data stability, support high concurrency, and reduce server loads.

Challenge 5: Data Security Issues

- Networked life makes it easier for criminals to obtain information about people, and has given rise to new forms of crime that are difficult to track and prevent.
- Ensuring user information security has become an important issue in the big data era. In addition, the increasing amount of big data poses higher requirements on physical security of data storage, and therefore poses higher requirements on multiple data copies and disaster recovery mechanisms.

Challenge 6: Shortage of Big Data Talent

- Each phase of big data construction depends on professional personnel. Therefore, a professional big data construction team with management skills and big data application experience must be cultivated. Every year, hundreds of thousands of big data-related jobs are created around the world. In the future, there will be a talent gap of more than one million. Therefore, universities and enterprises must work together to find and cultivate talent.

Challenge 7: Trade-off Between Data Openness and Privacy

- As big data applications become increasingly important, data resource openness and sharing have become key to maintaining advantages in the data war. However, data openness will inevitably infringe on some users' privacy. Protecting citizens' and enterprises' privacy while promoting data openness, application, and sharing, and strengthening privacy legislation are key challenges in the big data era.

Opportunity 1: Big Data Mining as a New Core of Business Analysis

- The focus of big data has gradually shifted from storage and transmission to data mining and application, which will profoundly affect enterprise business models. Big data can bring direct profits to enterprises and the positive feedback provides competitive advantages that are difficult to replicate.
 - Big data technologies can help enterprises integrate, mine, and analyze the huge amount of data at their disposal, build a systematic data system, and improve their own structures and management mechanisms.
 - With the growth of consumers' personalized requirements, big data has been applied across a wide range of fields, altering the development path and business models of most enterprises.

Opportunity 2: Big Data Provides Support for Information Technology Applications

Big data processing and analysis have become the support for next-generation information technology applications.

Next-generation information technologies, such as the mobile Internet, IoT, social networking, digital home, and e-commerce, use big data to aggregate generated information. These technologies process, analyze, and optimize data from different sources in a unified and comprehensive manner, and results are returned to a wide range of applications to further improve the user experience and create new business, economic, and social value.

Therefore, big data has the potential to accelerate social transformations. However, to unleash this energy, we need strict data governance, insightful data analysis, and an environment that stimulates management innovation.

Opportunity 3: Big Data as a New Engine for Growth in the Information Industry

- The business value and market requirements of big data have become a new engine that drives growth in the information industry.
 - As the value recognition of big data increases, the market demand will surge, and new technologies, products, services, and business forms oriented to the big data market will emerge on an ongoing basis.
 - Big data will create a high-growth market for the information industry. In the hardware and integrated device field, big data will face challenges such as effective storage, fast read/write, and real-time analysis, which will have an important impact on the chip and storage industry. In addition, the demands for integrated data storage and processing servers and in-memory computing are emerging. In the software and service field, the huge value of big data brings urgent requirements for quick data processing and analysis, which will lead to unprecedented prosperity of the data mining and business intelligence markets.

Big Data Era Leads the Way

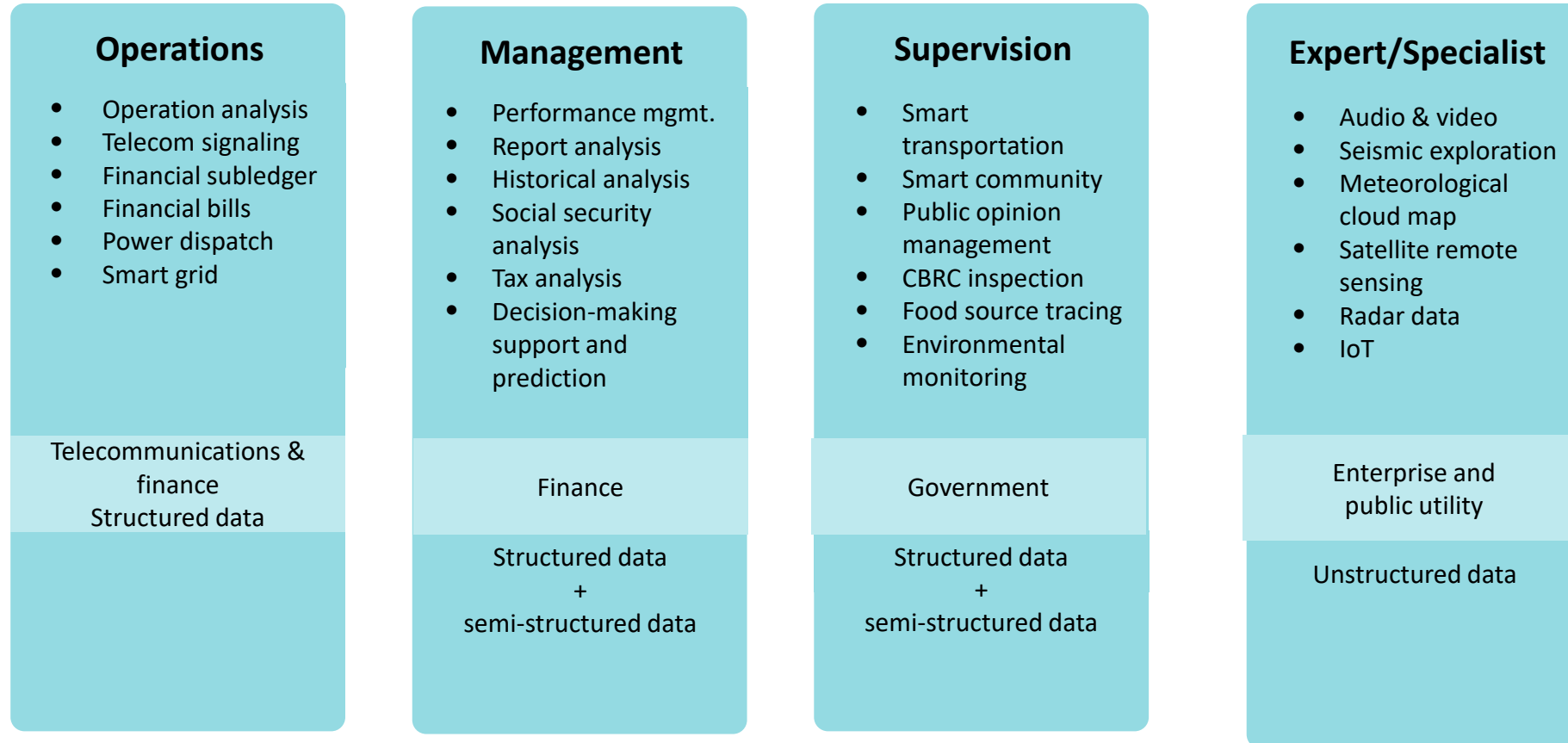
Data has penetrated every industry and business domain.

Insights into the essence (business), trend prediction, and future orientation are the key elements of the big data era.

The future guides the present, and the present ensures the future.



Application Scenarios of Enterprise-Level Big Data Platforms



Industries such as telecom, finance, and government have strong requirements for data analysis. New Internet technologies have been adopted to process big data with low value density.



Traditional Customers ➔ Receive services in a fixed location and fixed time.
Passively receive data.
Trust market information.
Passively receive propagation.

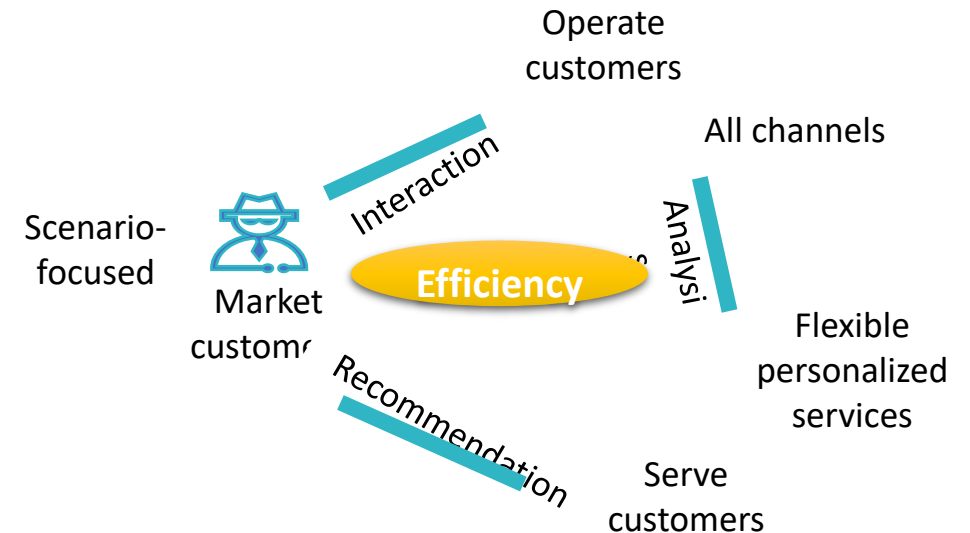
Offer standardized industrial services.
Focus on processes and procedures.
Passively receive information from a single source.
Contact customers by account managers.
Interact in fixed

Traditional
Finance

Obtain service analysis and create data anytime and anywhere.
Seek meaningful experiences.
Review the details.
Involved in creating content, products, and experiences.

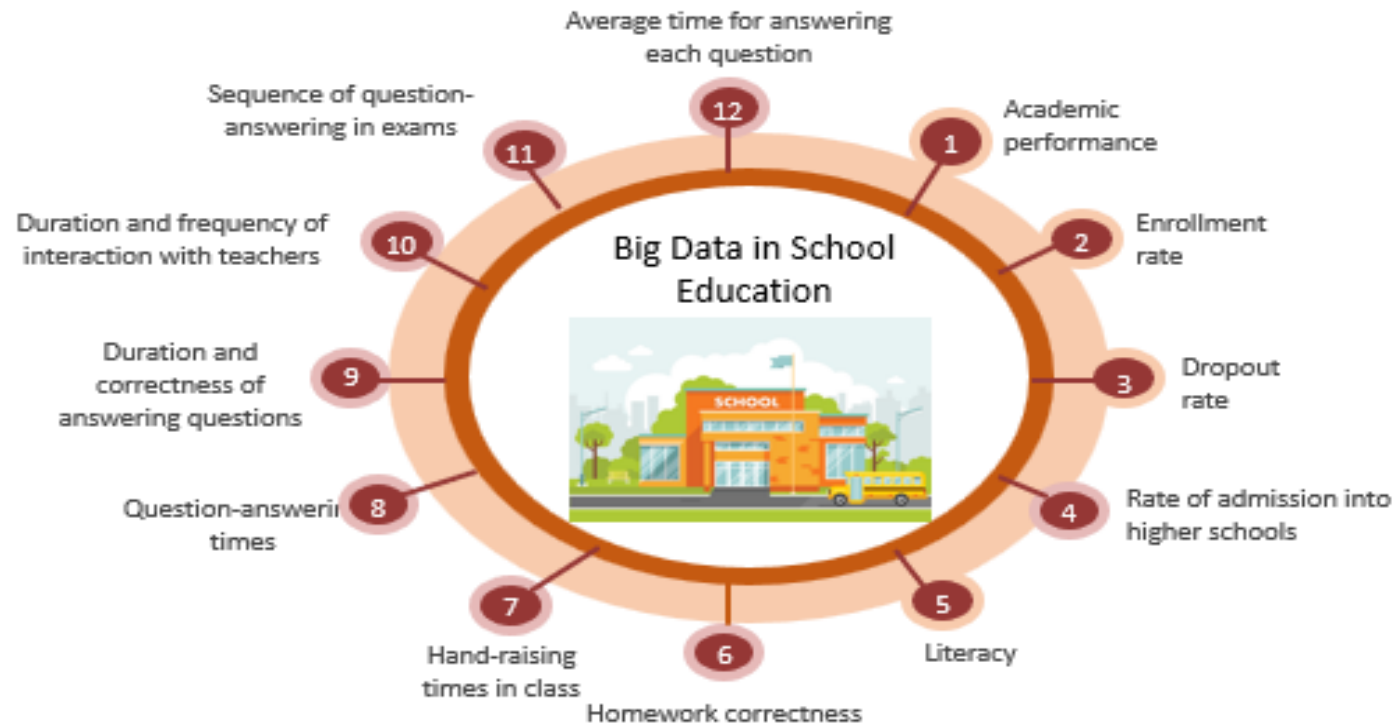
◀ New Customers

Importance of Data Mining



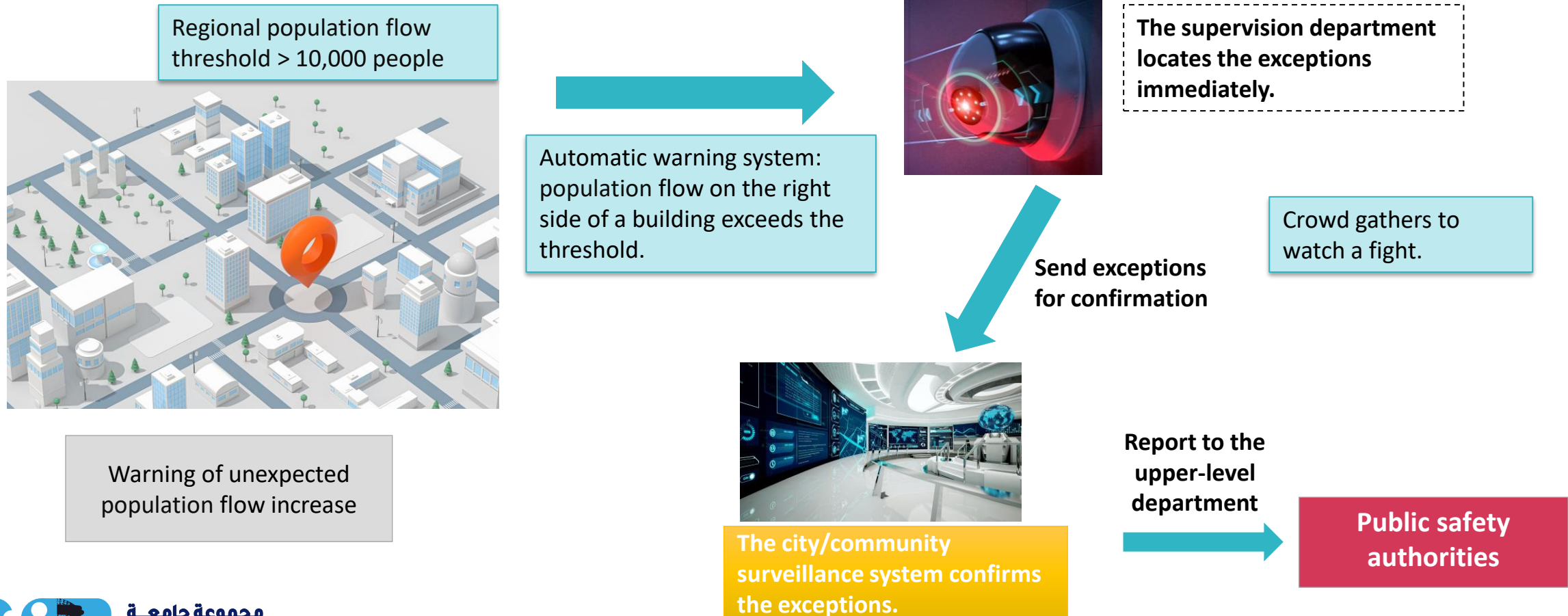
Big Data Application - Education

Big data analysis has been applied in all facets of education.



Public Safety Scenario - Automatic Warning and Linkage

Used to monitor and analyze urban population flows.



Traffic Planning Scenario - Multi-dimensional Traffic Crowd Analysis

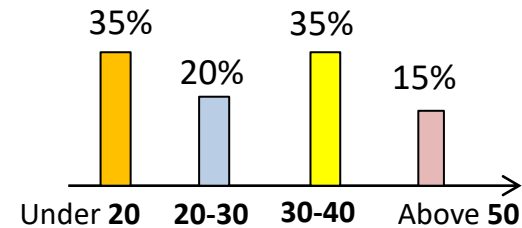


Traffic prediction suggestions
based on crowds

Area where historical
people flows exceed the
threshold

- North gate of the Beijing Workers' Sports Complex: The population flow exceeds 500/hour.
- Sanlitun: The population flow exceeds 800/hour.
- Beijing Workers' Sports Complex: 1,500 people/hour

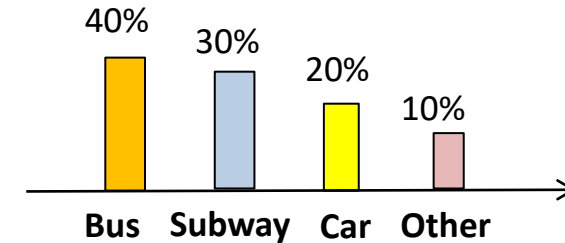
Analysis by crowd



Road network planning



Analysis by travel mode proportion



Bus route planning



Big Data Application - Clean Energy

Clean Energy Power Supply in Qinghai for Nine Consecutive Days

Coal Consumption Reduced by **800,000 Tons**

CO₂ Emissions Reduced by **1.44 Million Tons**



"Intelligence + Data": Clearing the Air and Water

Advantages of Kunpeng Computing

2

Advantages

1

Advantageous ecosystem
shared between Kunpeng and
Arm accelerates
development.

Industry applications
incubated and optimized in
the Chinese market
contribute to a healthy
global value chain.

Typical Applications in the Kunpeng Computing Industry

Driven by technologies like 5G, AI, cloud computing, and big data, different industries have raised differing requirements for computing platforms, such as device-cloud homogeneity, intelligent processing of massive diversified data, and real-time analysis. The powerful computing power of the Kunpeng processor will play an important role in the digitalization of whole industries.

Big Data

- ✓ 30% higher performance
- ✓ Encryption/Decryption engine
- ✓ Converged deployment with x86 servers

Distributed Storage

- ✓ Reduced compression time
- ✓ IOPS improvement
- ✓ Compression/Decompression engine

Database

- ✓ Higher performance
- ✓ RoCE low-latency network
- ✓ NUMA optimization algorithm

Native Apps

- ✓ ARM native homogeneous apps without performance loss
- ✓ 5 million native applications

Cloud Services

- ✓ BMS memory bandwidth improvement
- ✓ Higher multi-core integer performance of ECSs
- ✓ Hybrid deployment of Kubernetes containers

Building Server Computing Power Based on Huawei Kunpeng Processors

High Computing Efficiency

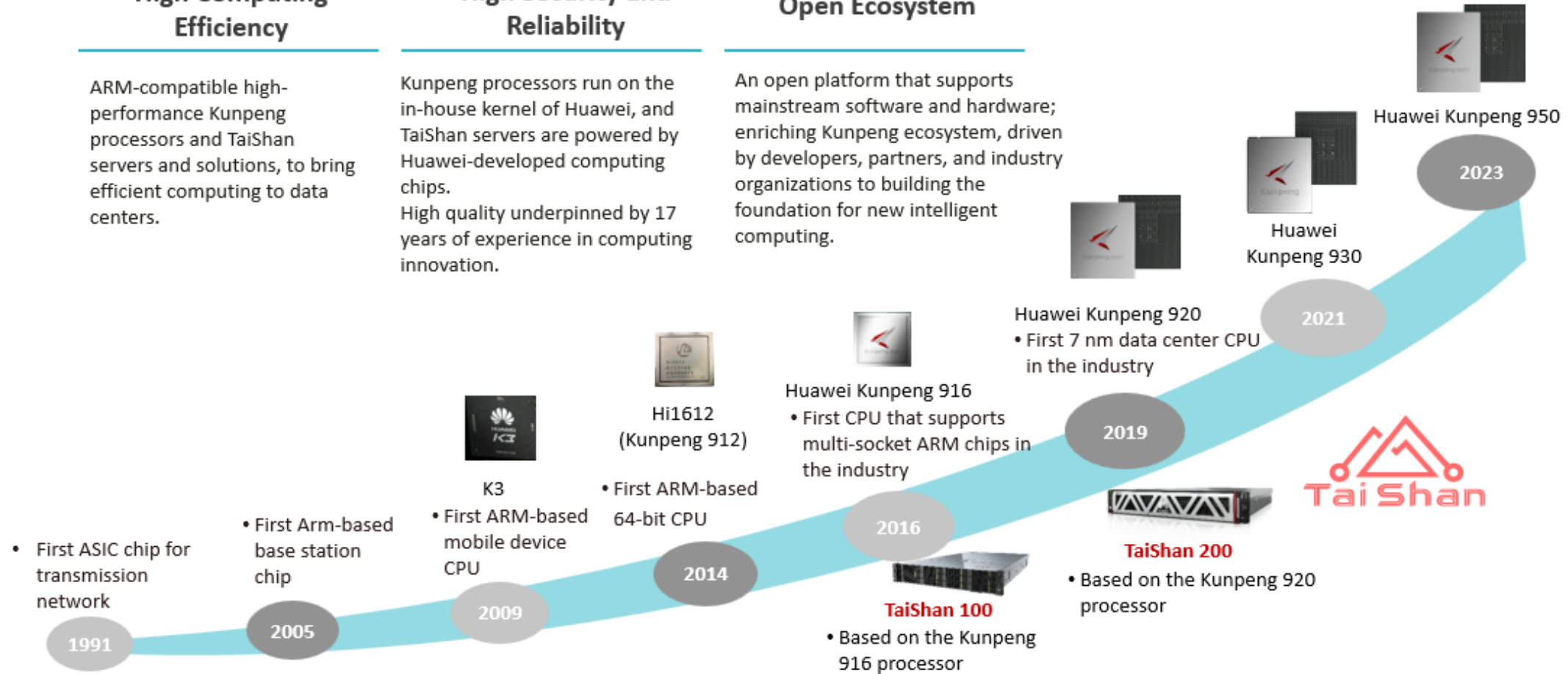
ARM-compatible high-performance Kunpeng processors and TaiShan servers and solutions, to bring efficient computing to data centers.

High Security and Reliability

Kunpeng processors run on the in-house kernel of Huawei, and TaiShan servers are powered by Huawei-developed computing chips. High quality underpinned by 17 years of experience in computing innovation.

Open Ecosystem

An open platform that supports mainstream software and hardware; enriching Kunpeng ecosystem, driven by developers, partners, and industry organizations to building the foundation for new intelligent computing.



Huawei's Big Data Solution

Kunpeng Big Data Solution

Huawei's secure and controllable Kunpeng big data solution provides one-stop high-performance big data computing and data security capabilities to solve basic problems such as data security, efficiency, and energy consumption in intelligent big data construction within the public safety sector.

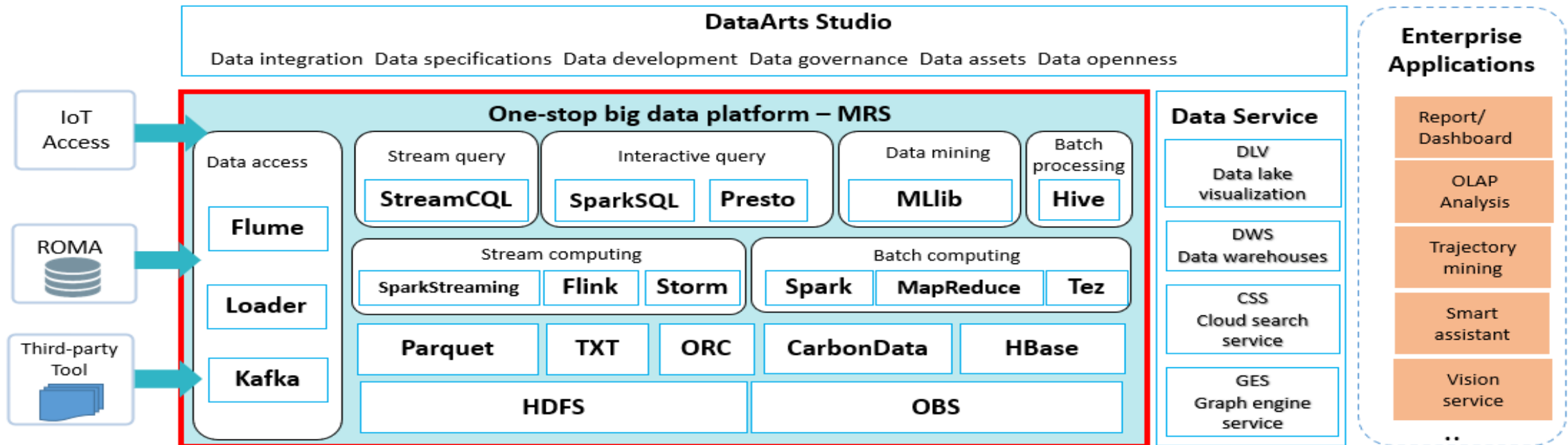
● Big Data Pro

This solution adopts the public cloud architecture with **decoupled storage and computing**, **ultimate scalability**, and the **highest possible efficiency**. The highly elastic Kunpeng computing power is used as the computing pool and an OBS that offers ultimate data durability is used as the storage pool. The resource utilization of big data clusters can be greatly improved, and big data costs can be halved.

Big data services and AI training and inference platform, which feature high performance and reliability, enable quick digitalization and intelligent transformation.

Huawei Cloud Big Data Services

- One-stop service for data development, testing, and application



100% compatibility with open-source ecosystems + plug-in management of third-party components; one-stop enterprise platform, separated storage and compute + Kunpeng optimization for better performance.

Summary

- This chapter provides an overview of big data, including the definition of big data, four Vs of big data, application fields of big data, and challenges and opportunities in the big data era.
- Huawei has proposed the Kunpeng solution to address these challenges. Thanks to Huawei Kunpeng processors and TaiShan servers, Huawei continuously has improved computing power and developed the Huawei Kunpeng computing industry. The Huawei Kunpeng ecosystem is designed for the big data sector, and offers numerous public cloud services to help partners implement data intelligence faster and more effectively.

Thank You
For Your Attention