HUAWEI

Huawei Cloud Academy

Huawei Kunpeng Computing Big Data Solution

Grow With Intelligence

Foreword

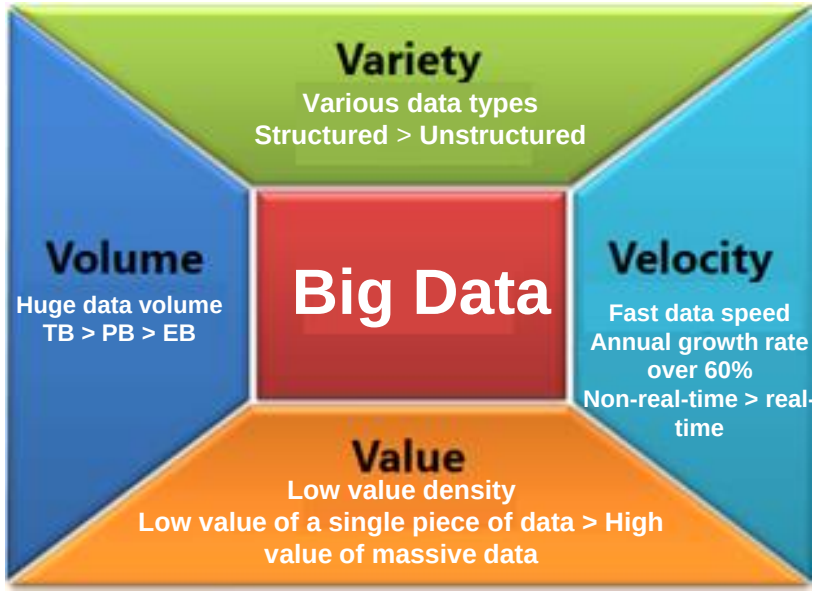
This course introduces some basic concepts and applications of big data, analyzes the advantages of Kunpeng computing big data based on the current big data trends and Kunpeng features, and briefly introduces the procedure for porting big data components to Kunpeng.

Objectives

- Upon completion of this course, you will be able to:
  - Understand the Kunpeng big data ecosystem.
  - Understand the advantages of Kunpeng in big data.
  - Understand how to port big data components to Kunpeng.

Contents

Big Data Concepts and Application Scenarios

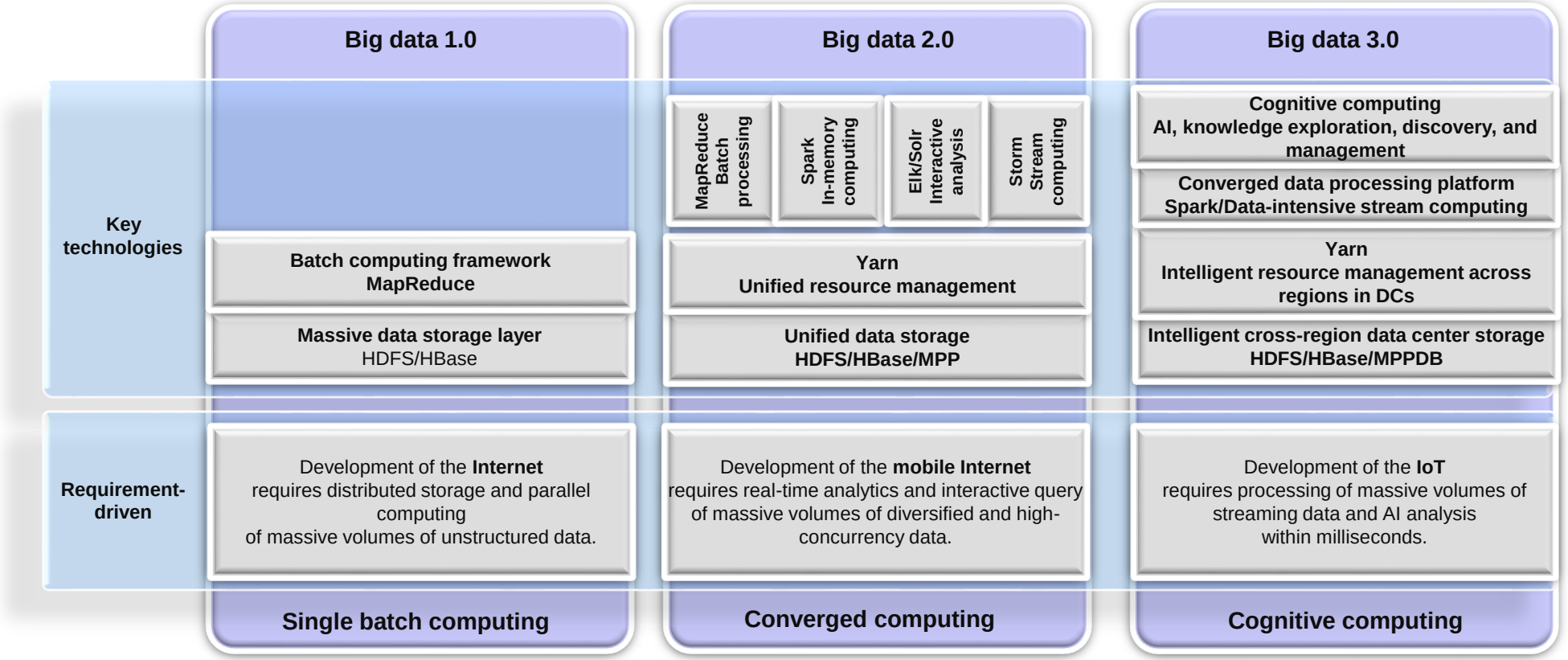


"Big data consists of datasets that are too large to collect, use, manage, and process in an acceptable time."

"Big data refers to large amounts of **unstructured or structured** data from a variety of sources."

Operations	Management	Supervision	Specialized Fields	Internet
• Operations analysis • Telecom signaling • Financial subledger • Financial bills • Power distribution • Smart grid	• Performance • Reports • Files • Social security analysis • Tax analysis • Decision-making support and prediction	• CBRC inspection • Food source tracing • Environmental monitoring	• Audio & video • Seismic exploration • Meteorological cloud map • Satellite remote sensing • Radar data • IoT	• Fine-grained operations • User experience optimization • Strategy analysis
Telecom/Finance/ Electric power	**Carriers/ Finance**	**Government/ Public Security**	**Government**	**Internet**

Big Data Trend: Parallel Computing Framework Has Become Mainstream



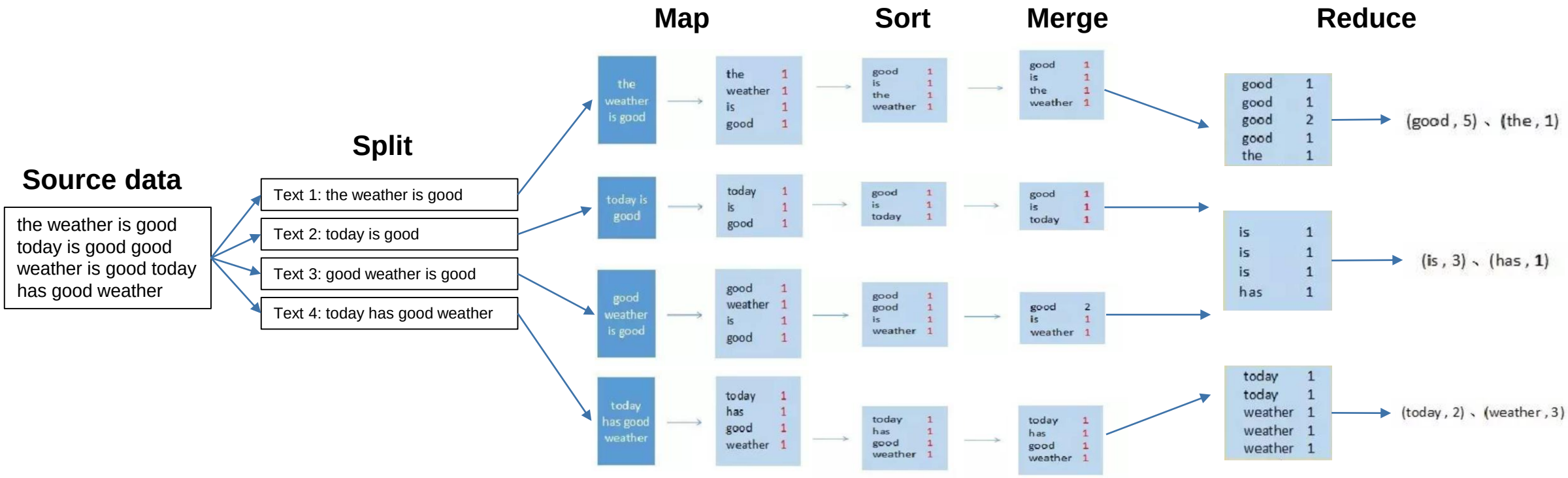
Computing Requirements of Big Data

A single server fails to process massive volumes of data.

Distributed parallel computing framework becomes a standard solution.

High concurrency is the key to improve performance.

Big Data MapReduce Parallel Computing Model Fits the Multi-Core Kunpeng Architecture



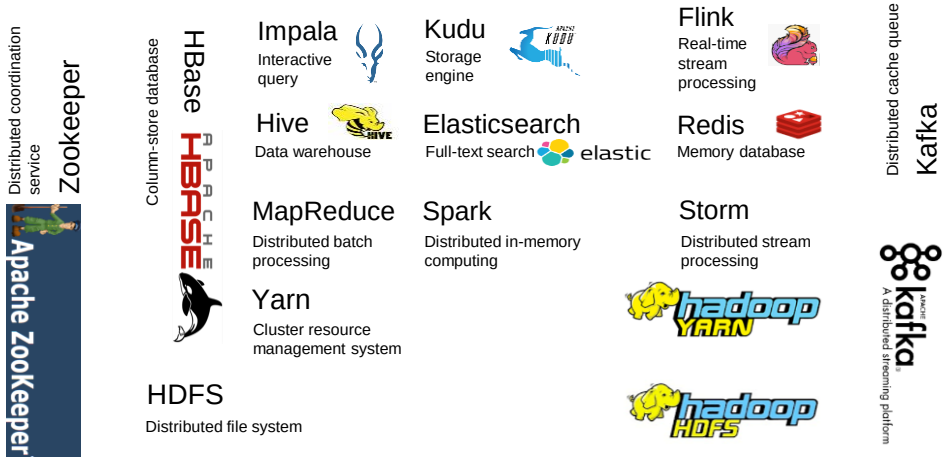
- Map: A big data set is divided into several small data sets for analytics. Each small data set has an independent thread for parallel analytics and computing.
- Reduce: The analytics results of small data sets are consolidated and returned to users.
- **Kunpeng multi-core computing improves MapReduce's I/O concurrency and accelerates big data computing.**

Contents

Kunpeng Computing Big Data Ecosystem - Open Source and Business

Full support for open-source big data

- Support for big data components of open-source Apache
- Support for big data components of open-source Hortonworks Data Platform (HDP) and Ambari
- Support for big data components of Cloudera's Distribution Including Apache Hadoop (CDH) *



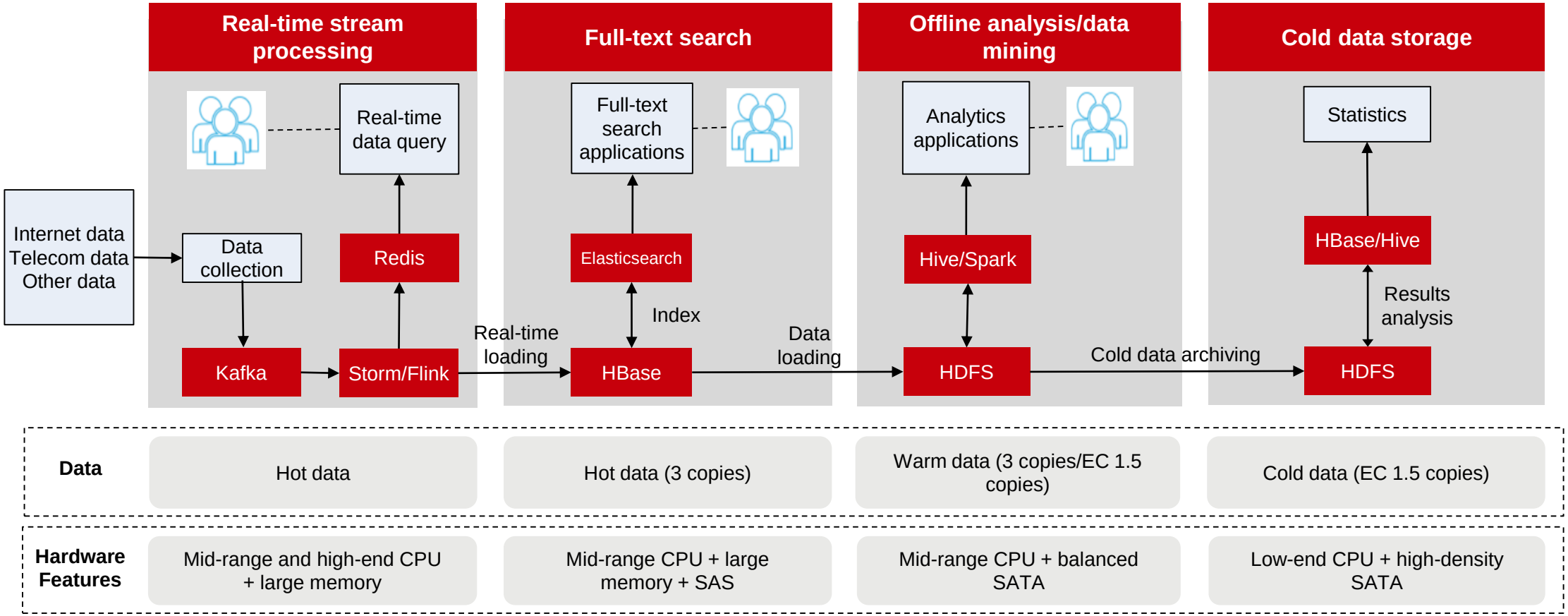
The ARM ecosystem fits to the open-source community.

- Core components such as Hadoop, Hive, HBase, Spark, Flink, Elasticsearch, and Kudu in the open-source community support the ARM ecosystem.

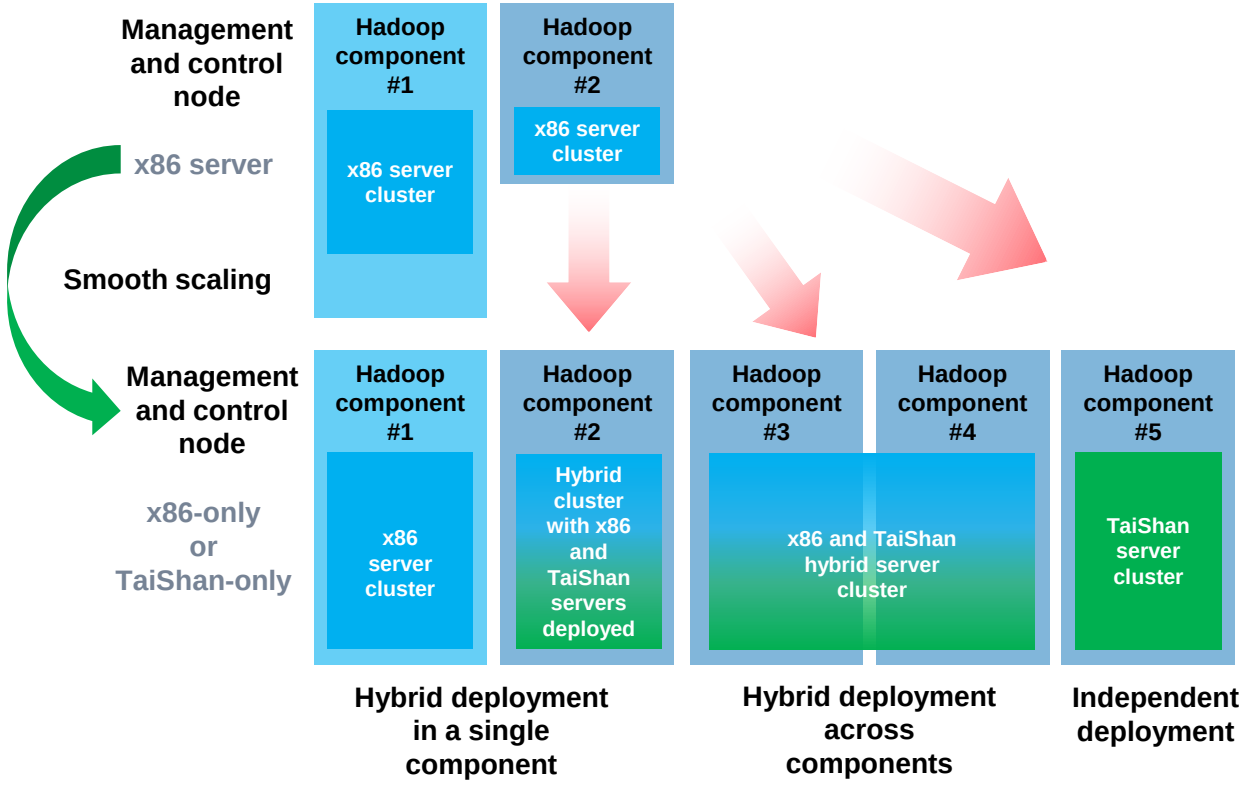
- **Continuous integration**: Code is compiled and packed for automatic tests to ensure the compatibility quality of code in ARM.
- **Software package release**: The community allows you to download the ARM software packages, and therefore you do not need to download the source code or compile and package the code.
- **Feature acceptance**: The community code trunk accepts feature patches. You do not need to download the source code to incorporate it to patches, and recompile and package the patches.

Kunpeng big data has been certified in the public safety, carrier, finance, and Internet industries, as well as mainstream China home-made software.

Typical Big Data Configuration Solution



Hybrid Deployment Based on FusionInsight or Other Commercial Software



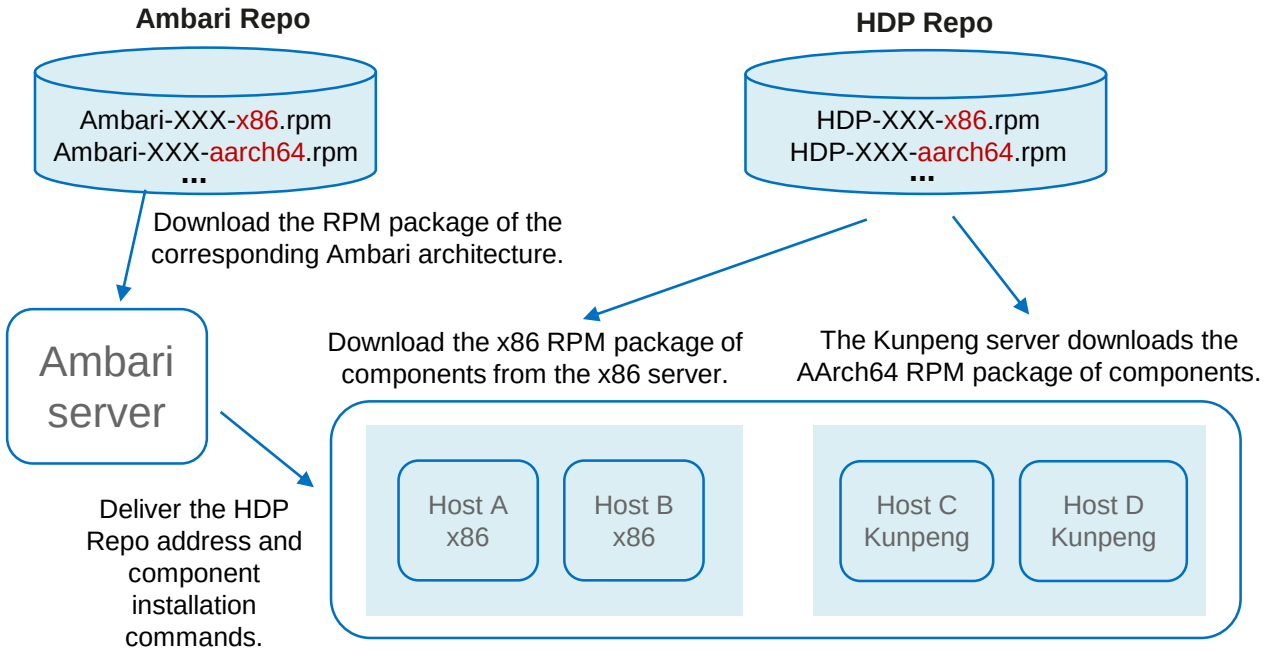
Components

- Supported components: HDFS, Yarn (MapReduce), Hive, Spark, Flink, HBase, Elasticsearch, Storm/Kafka/Flume, and GraphBase
- Unsupported components: Redis, Solr, Elk, Hue, Loader, Oozie and SmallFS. You are advised to use Kunpeng or x86 servers for independent deployment.

Deployment Procedure

- Check whether the operating system (OS) is supported in FusionInsight 6.5.1.
- Upgrade the FusionInsight cluster to FusionInsight 6.5.1.
- Install and connect the TaiShan servers to scale out the FusionInsight cluster.

Hybrid Deployment Based on HDP Open-Source Software



Components

- Supported components: HDFS, Yarn (MapReduce), Hive, Spark, Flink, HBase, Elasticsearch, Storm/Kafka/Flume, and GraphBase
- Unsupported components: Redis, Hue, Sqoop, and Oozie. You are advised to use Kunpeng or x86 servers for independent deployment.

Deployment Procedure

- Ensure that the OS and JDK versions meet the hybrid deployment requirements.
- Port the Ambari and required big data components to Kunpeng.
- Prepare the software packages of the x86 and Kunpeng versions and create the Yum repository based on the *Ambari Deployment Guide*.
- On the Ambari web page, configure the Yum repository address and add nodes.

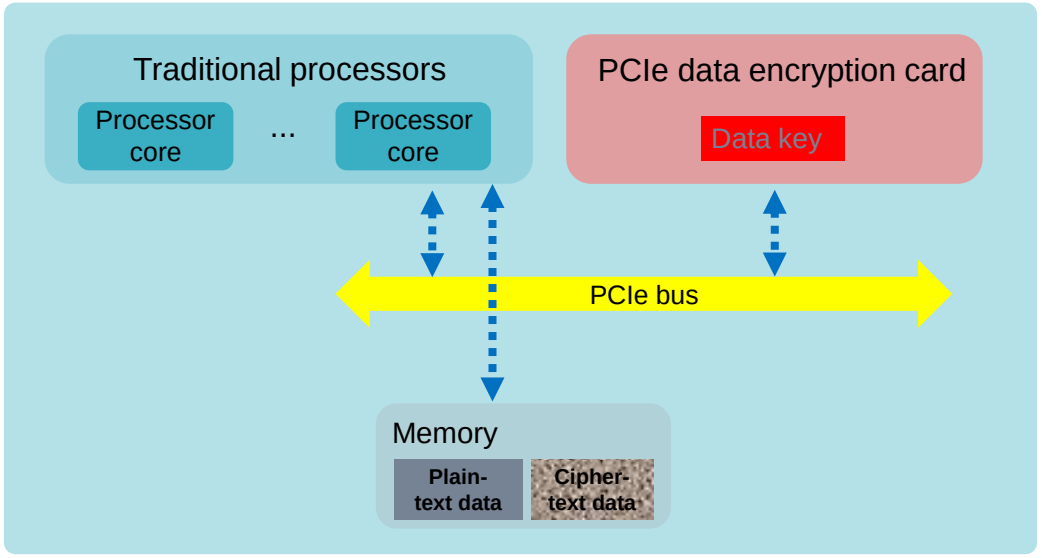
Kunpeng Encryption/Decryption Ensures Secure Plaintext Data Transmission

Traditional PCIe encryption card solution

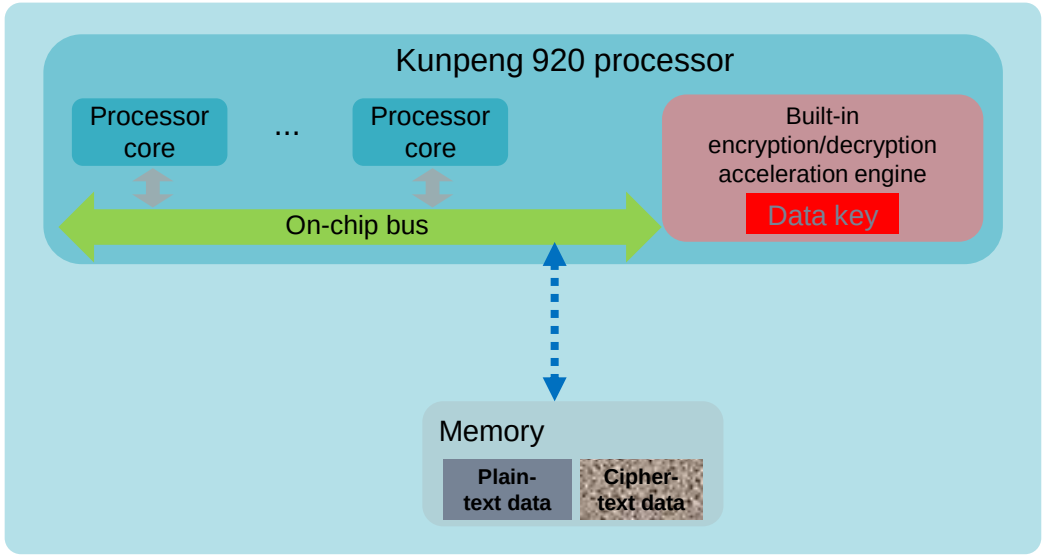
- Plaintext data is transmitted through the PCIe bus, which may cause data leakage.

Kunpeng encryption/decryption solution

- Kunpeng has a built-in encryption/decryption acceleration engine, which does not occupy computing resources.
- Plaintext data is transmitted only through the on-chip bus, ensuring security.
- SM3/SM4 encryption algorithm acceleration is supported.



Built-in encryption engine for higher security



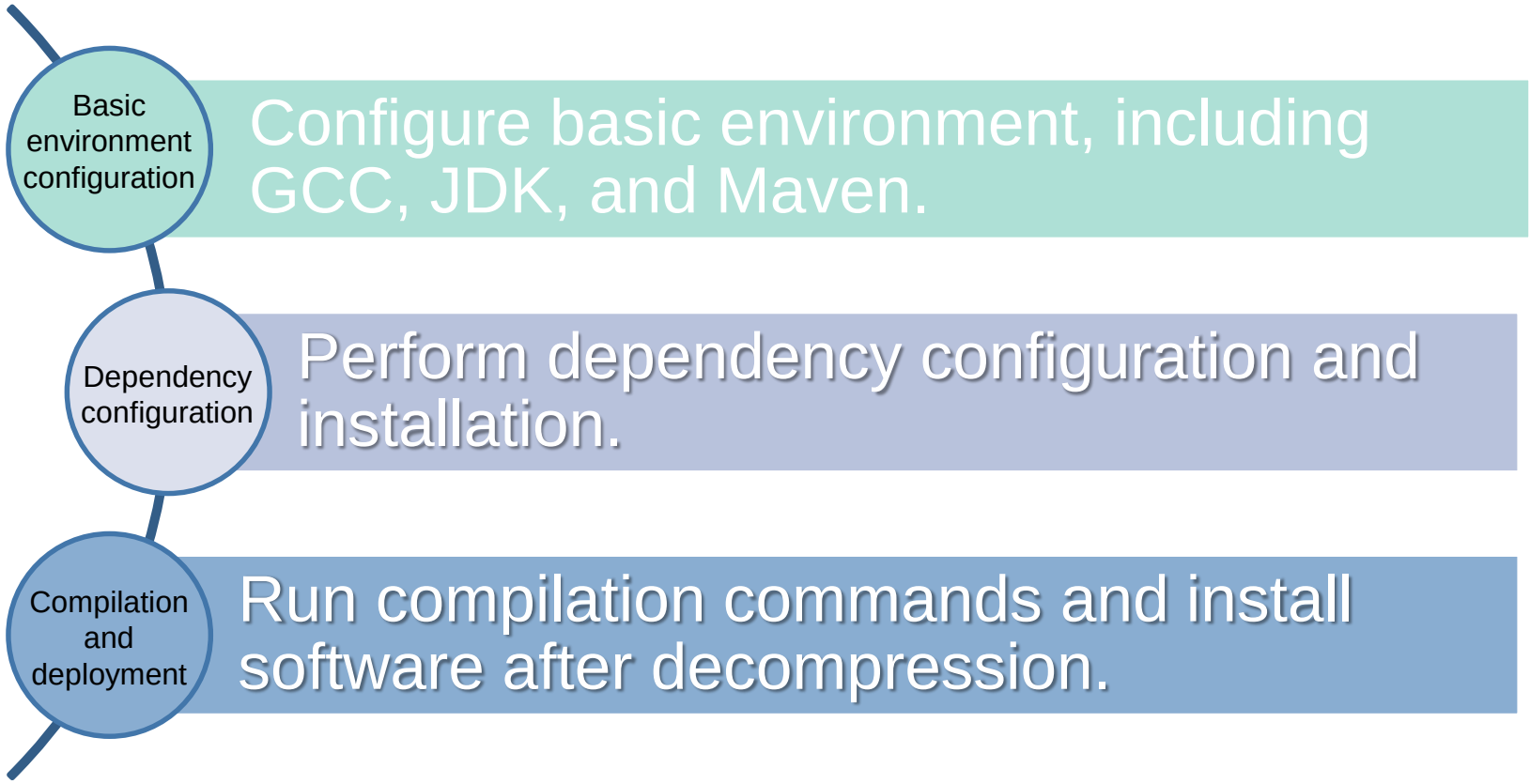
Processor resources are released without compromising service performance.

Why Kunpeng Big Data?

- **High performance**: Kunpeng multi-core processors enable high-concurrency I/Os.
- **Smooth scale-out**: Hybrid deployment of Kunpeng and x86 servers implements smooth scale-out of clusters on the live network.
- **Secure encryption/decryption**: Encryption/Decryption are implemented based on the built-in hardware of chips, and Chinese cryptographic algorithm is supported.
- **Prosperous ecosystem**: Mainstream open-source software, China home-made commercial software, and software and hardware decoupling are supported.

Contents

Kunpeng Big Data Software Porting



The simple and common configuration allows you to download basic dependency provided by Huawei Kunpeng repository and to select your desired versions.

Summary

This course analyzes the core advantages of Kunpeng big data based on the basic features of big data and the features of the Kunpeng processor, and briefly introduces the procedure for porting big data components to Kunpeng.

Recommendations

HUAWEI CLOUD Kunpeng ecosystem highlights:

https://bbs.huaweicloud.com/forum/thread-27853-1-1.html

Huawei Kunpeng image repository:

https://mirrors.huaweicloud.com/kunpeng/maven/

HUAWEI

Huawei Cloud Academy

Thank You.

Grow With Intelligence