



从IA到IE-华为大数据认证全揭秘



关注微信公众号
享终身免费培训



关于本节课

• 目标

- 华为大数据认证概述
- HCIA-面向大数据集群搭建与管理的入门级认证
- HCIP-面向大数据开发的中级认证
- HCIE-涵盖AI与机器学习的专家级认证
- 从认证看大数据的发展趋势



目录

1. 华为大数据认证概述
2. HCIA-面向大数据集群搭建与管理的入门级认证
3. HCIP-面向大数据开发的中级认证
4. HCIE-涵盖AI与机器学习的专家级认证
5. 从认证看大数据的发展趋势

华为大数据认证-HCIA-Big Data认证



定位:

大数据技术的普及，华为FusionInsight HD 解决方案的功能特性、常见且重要大数据组件基础技术原理及操作演练的考核和认证。

考试形式: 笔试 (单选、双选、判断题)

考试代码: H13-711

考试名称: HCIA-Big Data V2.0

考试时长: 90分钟

通过分数: 600/1000

考试费用: 200 USD

重认证:

随着ICT领域技术的发展，华为认证的内容将不定期进行更新和优化，通过重认证，可以保证认证的有效性；并表明认证与ICT技术的最新发展趋势保持一致。

证书有效期: 3年

重认证有效期: 3年

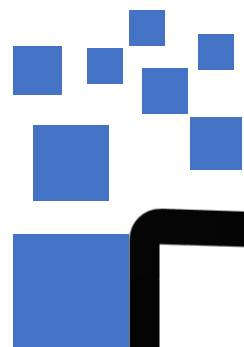
华为大数据认证-HCIA-Big Data认证

考试内容:

- 1.大数据行业的发展趋势，大数据特点等
- 2.FusionInsight HD解决方案架构、功能特性及华为在大数据行业的成功案例
- 3.常用且重要大数据组件基础技术原理（包括HBase, Hive, Loader, MR, YARN, HDFS, Spark, Flume, Kafka, ZooKeeper, Streaming）



华为大数据认证- HCIP-Big Data Developer



定位:

对大数据平台和核心组件开发应用的场景理解, 流程和开发方法等能力构建, 考核和认证

考试形式: 笔试 (单选、双选、判断题)

考试代码: H13-723

考试名称: HCIP-Big Data Developer V1.0

考试时长: 90分钟

通过分数: 600/1000

考试费用: 300 USD

重认证:

随着ICT领域技术的发展, 华为认证的内容将不定期进行更新和优化, 通过重认证, 可以保证认证的有效性; 并表明认证与ICT技术的最新发展趋势保持一致。

证书有效期: 3年

重认证有效期: 3年

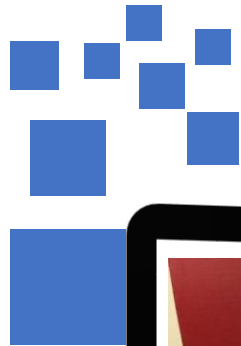
华为大数据认证- HCIP-Big Data Developer

考试内容:

HCIP-Big Data Developer V1.0考试覆盖FusionInsight应用开发流程, HDFS, HBase, Hive, Solr, Kafka, MR, YARN, Spark, Flume, Loader, Oozie, Redis, Storm等组件开发应用场景、流程、接口/方法和开发实践等所涉及的相关知识点



华为大数据认证- HCIE-Big Data



定位:

对大数据分析领域中大型分布式并行处理数据仓库平台的架构原理、优化设计、应用开发和大数据挖掘领域的基础理论、常用挖掘算法、应用设计以及开发，旨在推动业界大数据行业的专家型人才培养

考试形式: 笔试、实验、面试

资质要求: 建议5年以上相关工作经验

笔 试: (单选、双选、判断)

60分钟、 300 USD

实 验: 8小时上机实验 8000CNY

面 试: 1个小时 (HCIA和HCIE内容)

重认证:

随着ICT领域技术的发展，华为认证的内容将不定期进行更新和优化，通过重认证，可以保证认证的有效性；并表明认证与ICT技术的最新发展趋势保持一致。

证书有效期: 2年

重认证有效期: 3年

华为大数据认证- HCIE-Big Data

考试内容:

HCIE-Big Data V2.0 考试覆盖:

(1) 华为大型并行处理分布式数据仓库平台FusionInsight LibrA 产品架构、关键技术特性、开发和管理工具的使用、数据库系统设计及优化、基于FusionInsight LibrA的应用开发等;

(2) 数据挖掘的理论 (包括数据预处理、数据仓库、分类、聚类、关联规则等) 以及使用基于华为云的企业智能平台EI (Enterprise Intelligence) 的基础服务 - MLS (Machine Learning Service, 机器学习服务) 进行数据深度分析和挖掘;

(3) 面试还包括华为认证HCIA-Big Data 和HCIP-Big Data Developer 课程的相关内容。





目录

1. 华为大数据认证概述
- 2. HCIA-面向大数据集群搭建与管理的入门级认证**
3. HCIP-面向大数据开发的中级认证
4. HCIE-涵盖AI与机器学习的专家级认证
5. 从认证看大数据的发展趋势

HCIA面向大数据集群搭建与管理的入门级认证



培训对象：

准备参加HCIA-Big Data V2.0 认证考试的人员，希望掌握大数据的技术原理及部署实施与运维的人员

学习基础：

了解大数据技术基本原理和 Hadoop 的基础知识。了解SQL语言。

HCIA面向大数据集群搭建与管理的入门级认证



HCIA-Big Data认证包括但不限于：大数据行业与技术趋势介绍；

HDFS分布式文件系统、MapReduce分布式离线批处理计算引擎和Yarn资源协调

Spark2x基于内存的分布式计算引擎、HBase分布式NoSQL数据库、Hive分布式数据仓库

Streaming分布式流计算引擎、Loader数据转换、Flume海量日志聚合、Kafka分布式消息

订阅系统、ZooKeeper集群分布式协调服务等11个常用且重要的大数据组件技术原理与架构，

华为大数据解决方案产品FusionInsight HD及成功案例介绍；基于独立大数据组件的实战演练

以及大数据综合应用实战。

HCIA面向大数据集群搭建与管理的入门级认证



通过HCIA-Big Data认证,

将证明您已经系统掌握常用且重要的大数据组件技术原理与架构, 能够运用华为大数据解决方案FusionInsight HD进行海量数据的导入和导出、分布式文件系统HDFS的基础操作、分布式数据库HBase客户端及表操作、分布式数据仓库Hive的常用HQL语句查询, 以及了解典型场景的综合应用等。使您具备大数据售前、大数据售后技术支持、大数据项目管理、大数据运维管理、大数据开发、大数据分析等岗位所必备的知识技能。

HCIA面向大数据集群搭建与管理的入门级认证



认证价值:

个人: 通过HCIA-Big Data V2.0认证, 将证明您系统理解大数据架构、技术原理, 掌握华为大数据FusionInsight HD方案, 具备大数据平台的规划设计、部署和运维管理能力, 胜任大数据平台管理员工作岗位。

企业拥有通过HCIA-Big Data认证的工程师,

意味着企业掌握了大数据技术基础原理与架构等相关知识, 掌握了运用大数据技术、产品规划并建设大数据平台的系统性能力, 并具备大数据平台管理和日常运维能力/具备运用华为大数据解决方案FusionInsight HD和开源Hadoop生态系统进行数据初步处理的能力, 并且为大数据应用开发业务奠定了坚实的基础

HCIA课程大纲



第01章 大数据行业与技术趋势

- 1.大数据时代
- 2.大数据的应用领域
- 3.大数据时代的机遇和挑战
- 4.华为大数据解决方案

第02章 HDFS分布式文件系统技术

- 1.HDFS概述及应用场景
- 2.HDFS在FusionInsight HD产品的位置
- 3.HDFS系统架构
- 4.关键特性介绍

第03章 MapReduce分布式离线批处理和Yarn资源协调

- 1.MapReduce和Yarn基本介绍
- 2.MapReduce和Yarn功能与架构
- 3.Yarn的资源管理和任务调度
- 4.增强特性

HCIA课程大纲



第05章 HBase分布式NoSQL数据库

- 1.HBase 基本介绍
- 2.HBase 功能与架构
- 3.HBase 关键流程
- 4.HBase 华为增强特性

第06章 Hive分布式数据仓库

- 1.Hive概述
- 2.Hive功能与架构
- 3.Hive基本操作

第07章 Streaming分布式流计算引擎

- 1.Streaming简介
- 2.系统架构
- 3.关键特性介绍
- 4.StreamCQL介绍

HCIA课程大纲



第08章 Flink流计算处理和批处理平台

- 1.Flink概述
- 2.Flink原理与技术架构
- 3.Flink在FusionInsight HD中的集成情况

第09章 Loader数据转换

- 1.Loader简介
- 2.Loader作业管理

第10章 Flume海量日志聚合

- 1.Flume简介及架构
- 2.Flume关键特性介绍
- 3.Flume应用举例

第11章 Kafka分布式消息订阅系统

- 1.Kafka简介
- 2.Kafka架构与功能
- 3.Kafka关键流程

HCIA课程大纲



第12章 ZooKeeper集群分布式协调服务

- 1.ZooKeeper简介
- 2.ZooKeeper在FusionInsight HD产品的位置
- 3.系统架构
- 4.关键特性介绍
- 5.与组件的关系

第13章 FusionInsight HD 解决方案介绍

- 1.FusionInsight HD 企业版概述
- 2.FusionInsight HD 特性介绍
- 3.FusionInsight HD 成功案例



目录

1. 华为大数据认证概述
2. HCIA-面向大数据集群搭建与管理的入门级认证
- 3. HCIP-面向大数据开发的中级认证**
4. HCIE-涵盖AI与机器学习的专家级认证
5. 从认证看大数据的发展趋势

HCIP-面向大数据开发的中级认证



培训对象：

准备参加 HCIP-Big Data Developer 认证考试的人员。希望掌握大数据平台和核心组件开发应用的人员

学习基础：

了解大数据技术基本原理和 Hadoop 的基础知识。了解SQL语言、有开发语言基础，比如Java

HCIP-面向大数据开发的中级认证



HCIP-Big Data Developer认证包括但不限于：FusionInsight应用开发流程，HDFS, HBase, Hive, Solr, Kafka, MR, YARN, Spark, Flume, Loader, Oozie, Redis, Storm等组件开发应用场景、流程、接口/方法和开发样例讲解，开发实践

HCIP-面向大数据开发的中级认证



通过HCIP-Big Data Developer认证,

将证明您深入理解大数据平台与技术及应用, 掌握大数据平台和关键组件的开发流程、接口, 掌握典型场景的开发方法, 胜任大数据开发者工作岗位。

HCIP-面向大数据开发的中级认证



认证价值:

个人: 通过HCIP-Big Data Developer认证, 将证明您深入理解大数据平台与技术及应用, 掌握大数据平台和关键组件的开发流程、接口, 掌握典型场景的开发方法, 胜任大数据开发者工作岗位

企业拥有通过HCIP-Big Data Developer认证的工程师,

意味着企业掌握了运用大数据平台和组件实现大数据解决方案规划、实施、开发等综合能力

HCIP课程大纲



Module01 FusionInsight应用开发总指导

- 1.大数据整体
- 2.大数据开发流程
- 3.典型案例

Module02 HDFS 应用开发

- 1.HDFS应用场景
- 2.HDFS业务过程
- 3.环境搭建和程序运行
- 5.常用接口
- 6.HDFS组件应用实践

Module03 HBase应用开发

- 1.HBase应用场景
- 2.HBase业务过程
- 3.环境搭建和程序运行
- 5.HBase开发接口
- 6.表设计
- 7.HBase组件应用实践

HCIP课程大纲



Module04 MapReduce和YARN应用开发

- 1.MR和YARN应用场景
- 2.MR过程
- 3.环境搭建
- 4.程序运行
- 5.MR开发接口
- 6.MR组件应用实践

Module05 Hive应用开发

- 1.Hive应用场景
- 2.Hive业务流程
- 3.环境搭建
- 4.程序运行
- 5.Hive开发接口
- 6.Hive组件应用实践

HCIP课程大纲



Module06 Solr应用开发

- 1.Solr应用场景
- 2.Solr业务流程
- 3.环境搭建
- 4.程序运行
- 5.Solr开发接口
- 6.Solr组件应用实践

Module07 Kafka应用开发

- 1.Kafka应用场景
- 2.Kafka业务流程
- 3.环境搭建
- 4.程序运行
- 5.Kafka开发接口
- 6.Kafka组件应用实践

HCIP课程大纲



Module08 Spark应用开发

- 1.Spark应用场景
- 2.Spark业务过程
- 3.环境搭建
- 4.程序运行
- 5.Spark开发接口
- 6.Spark组件应用实践

Module09 Redis应用开发

- 1.Redis应用场景
- 2.Redis业务流程
- 3.环境搭建
- 4.程序运行
- 5.Redis组件应用实践

HCIP课程大纲



Module10 Streaming应用开发

- 1.Streaming应用场景
- 2.Streaming业务流程
- 3.环境搭建和程序运行
- 5.Streaming组件应用实践

Module11 Flume应用开发

- 1.Flume应用场景
- 2.Flume业务流程
- 3.环境搭建和程序运行
- 5.Flume组件应用实践

Module12 Oozie应用开发

- 1.Oozie应用场景
- 2.Oozie业务流程
- 3.环境搭建和程序运行
- 5.Oozie组件应用实践



目录

1. 华为大数据认证概述
2. HCIA-面向大数据集群搭建与管理的入门级认证
3. HCIP-面向大数据开发的中级认证
- 4. HCIE-涵盖AI与机器学习的专家级认证**
5. 从认证看大数据的发展趋势

HCIE-涵盖AI与机器学习的专家级认证



培训对象：

准备参加HCIE-Big Data V2.0 认证考试的人员，希望掌握大数据分析和大数据挖掘的应用和实践的人员

学习基础：

掌握大数据技术基本原理和 Hadoop 的基础知识和SQL语言、有开发语言基础，比如Python

HCIE-涵盖AI与机器学习的专家级认证



HCIE-Big Data认证包括但不限于：

- (1) 华为企业级的大型并行处理分布式数据仓库平台FusionInsight LibrA的架构、功能特性、数据库设计、数据库数据导入导出、数据库调优及数据库开发等；
- (2) 数据挖掘的理论（包括数据预处理、数据仓库、分类、聚类、关联规则等）以及使用华为企业级大数据挖掘平台FusionInsight Miner和基于华为云的企业智能平台EI（Enterprise Intelligence）的基础服务 - MLS（Machine Learning Service，机器学习服务）进行数据深度分析和挖掘。

HCIE-涵盖AI与机器学习的专家级认证



通过HCIE-Big Data认证,

将证明您系统理解并掌握大型并行处理数据仓库平台的架构原理、分布式关系型数据库的设计和优化及应用开发; 掌握大数据挖掘的基本原理, 常用算法、常用挖掘方法、对华为大数据挖掘解决方案FusionInsight Miner和华为云机器学习服务MLS的熟练使用, 胜任数据仓库开发、大数据分析、大数据数据挖掘和人工智能的相关岗位。

HCIE-涵盖AI与机器学习的专家级认证



认证价值:

个人：通过HCIE-Big Data V2.0认证，将证明您系统理解并掌握大型并行处理数据仓库平台的架构原理、分布式关系型数据库的设计和优化及应用开发；掌握大数据挖掘的基本原理，常用算法、常用挖掘方法、对华为大数据挖掘解决方案FusionInsight Miner和华为云机器学习服务MLS的熟练使用，胜任数据仓库开发、大数据分析、大数据数据挖掘和人工智能的相关岗位。

企业拥有通过HCIE-Big Data认证的工程师，

企业拥有HCIE-Big Data V2.0 认证的工程师，意味着企业掌握了运用大数据技术进行大数据分析应用、大数据挖掘应用的能力

HCIE-涵盖AI与机器学习的专家级认证



第一章 分布式数据库概述

- 1.数据库体系架构基本知识
- 2.MPP特性介绍
- 3.业界产品说明

第二章 FusionInsight LibrA基本组件简介

- 1.FusionInsight LibrA 逻辑架构
- 2.FusionInsight LibrA 物理架构
- 3.FusionInsight LibrA 数据库对象

第三章 FusionInsight LibrA 产品特性和关键技术

- 1.架构和支持的操作系统
- 2.面向应用开发的基本功能
- 3.数据库安全
- 4.图形化工具
- 5.高级特性

HCIE-涵盖AI与机器学习的专家级认证



第四章 FusionInsight LibrA配套工具集

- 1.Database Manager
- 2.Data Studio

第五章 FusionInsight LibrA安全管理

- 1.访问控制
- 2.用户管理
- 3.权限管理模型
- 4.对象权限
- 5.安全审计

第六章 FusionInsight LibrA数据库管理系统并发控制

- 1.配置负载均衡
- 2.事务与隔离级别
- 3.LibrA锁机制介绍

HCIE-涵盖AI与机器学习的专家级认证



第七章 FusionInsight LibrA数据库性能监控

- 1.系统资源监控方式
- 2.性能相关视图
- 3.使用命令行监控集群性能
- 4.使用DM工具监控集群性能

第八章 FusionInsight LibrA数据迁移

- 1.数据迁移场景划分
- 2.GDS工具适用场景和使用方法
- 3.copy from/to适用场景和使用方法
- 4.开源ELT工具适用场景和使用方法
- 5.gs_dump/gs_restore适用场景和使用方法
- 6.gsql工具适用场景和使用方法
- 7.Roach工具适用场景和使用方法
- 8.从其他数据库中迁移数据

HCIE-涵盖AI与机器学习的专家级认证



第九章 FusionInsight LibrA SQL 介绍

1. SQL相关机制和原理
2. 利用EXPLAIN进行SQL调优
3. 数据库设计
4. 数据库逻辑设计
5. 存储模式
6. 数据分布
7. 表分区
8. 索引

第十章 FusionInsight LibrA 数据库设计

1. 数据库逻辑设计
2. 存储模式设计
3. 数据分布设计
4. 表分区设计
5. 索引设计

HCIE-涵盖AI与机器学习的专家级认证



第十一章 FusionInsight LibrA应用程序开发指导

1. FusionInsight LibrA驱动介绍
2. ODBC应用程序开发
3. JDBC应用程序开发
4. 基础开发规范

第十二章 预备知识和数据介绍

1. 什么是数据挖掘
2. 数据和属性类型
3. 数据的统计描述

第十三章 数据预处理

1. 数据清理
2. 数据集成
3. 数据规约
4. 数据变换

HCIE-涵盖AI与机器学习的专家级认证



第十四章 数据仓库介绍

1. OLAP和OLTP概念
2. 数据仓库和数据集市概念
3. 多维数据模型
4. 概念分层
5. ROLAP/MOLAP/HOLAP
6. 方体物化

第十五章 分类 (Classification)

1. 分类和回归的概念
2. 决策树 (ID3、C4.5和CART)
3. 朴素贝叶斯分类
4. bootstrap
5. 组合分类 (装袋、提升、随机森林)
6. 后向传播
7. 支持向量机 (SVM)
8. 类不平衡数据

HCIE-涵盖AI与机器学习的专家级认证



第十六章 聚类 (Clustering)

1. 聚类概念
2. k-均
3. k-中心点 (PAM、CLARA、CLARANS)

第十七章 离群点检测

1. 离群点概念
2. 离群点检测方法

第十八章 关联规则 (Association Rule)

1. 基本概念、购物篮分析
2. 项集
3. 频繁项集挖掘和Apriori算法

第十九章 FusionInsight Miner 和 华为云机器学习服务MLS

1. FusionInsight Miner与MLS介绍
2. DEMO 演示



目录

1. 华为大数据认证概述
2. HCIA-面向大数据集群搭建与管理的入门级认证
3. HCIP-面向大数据开发的中级认证
4. HCIE-涵盖AI与机器学习的专家级认证
5. 从认证看大数据的发展趋势

BigData名词由来

EMC²

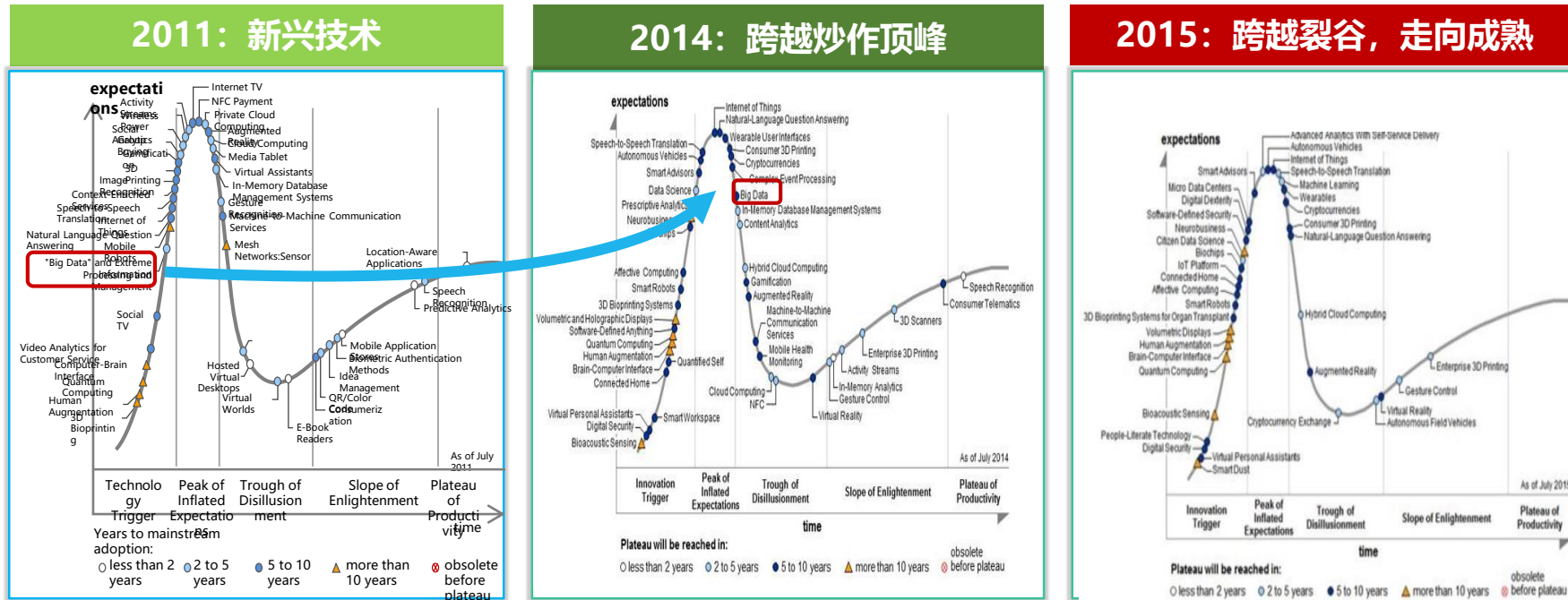
2011年5月，在“云计算相遇大数据”为主题的EMC World 2011会议中，EMC抛出了Big Data概念。



20世纪90年代，数据仓库之父的Bill Inmon就经常提及Big Data。



大数据趋势：跨过裂谷、走向成熟



50%的企业已经投资和使用大数据，33%的企业正在规划，企业的持续投资，逐步催熟大数据产业发展。云计算和大数据已不在2015年的图上，代表它们已不再是“新兴”技术，而是成为了“主流”技术了。

大数据：不同的声音



什么是大数据

维基百科: “大数据是指无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集合”

IDC: 一般会涉及2种以上数据形式, 数据量100T以上, 且是高速、实时数据流; 或者从小数据开始, 但数据每年增长60%

Gartner: 大数据的四个V: **Volume**、**Variety**、**Velocity**、**Value**

Volume: 数据量巨大

- 集中储存/集中计算已经无法处理巨大的数据量

每天9500万条



每天25TB日志数据



每天上传168TB视频



Variety: 种类和来源多样化

- 日志/图片/视频/文档/地理位置 ...

非结构化数据, 80%



结构化数据, 20%

80%

年增长率

60%

40%

20%

0%

32%

63%

结构化数据 非结构化数据

Velocity: 分析处理速度快

- 海量数据的及时有效分析



电子商务

- 实时反应, 以促成订单, 增强用户粘性。



社交网络

- 快速分析用户行为, 实现精准营销。



移动互联网广告

- 迅速调整, 实现对于广告需求的精确满足。



物联网

- 日本大地震后9分钟发出海啸预警

Value: 价值密度低, 商业价值高

- 大量的不相关信息的进行复杂深度分析, 深挖价值



精准营销深入洞察



统一监控分析

另外IBM有大数据5V特征定义, 增加了一个Veracity (真实性)

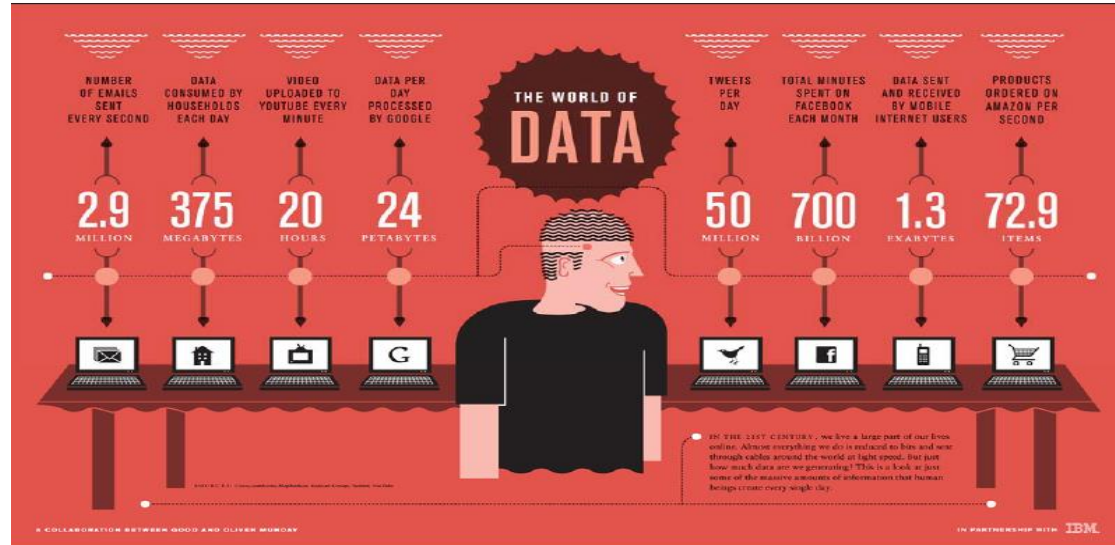
华为对大数据的理解



- 大数据是指通过对海量数据的智能存储、智能挖掘和智能分析，面向具体的行业应用，把巨量离散的、碎片化的数据加工形成具有商业价值的**数据处理技术**，以帮助企业/机构更准确地做出趋势判断、进行精准营销和实现精细化运营等。
- 大数据的数据量应足够大，一般是P级，数据的时间跨度应足够长，以真正实现从微观数据到宏观规律的跃迁。

海量的数据从哪里来-人

在web 2.0的时代，人们从信息的被动接受者变成了主动创造者



全球每秒钟发送**2.9百万封**电子邮件，一分钟读一篇的话，足够一个人昼夜不息的读5.5年...
每天会有**2.88万个**小时的视频上传到Youtube，足够一个人昼夜不息的观看3.3年...
推特上每天发布**5千万条**消息，假设10秒钟浏览一条信息，这些消息足够一个人昼夜不息的浏览16年...
每天亚马逊上将产生**6.3百万笔**订单...
每个月网民在Facebook 上要花费**7千亿分钟**，被移动互联网使用者发送和接收的数据高达**1.3EB**...
Google 上每天需要处理**24PB**的数据...

海量的数据从哪里来-机器



Boeing: 飞机每个引擎**3分钟**产生**1TB**数据, 波音787 **6小时**飞行产生**240TB**数据

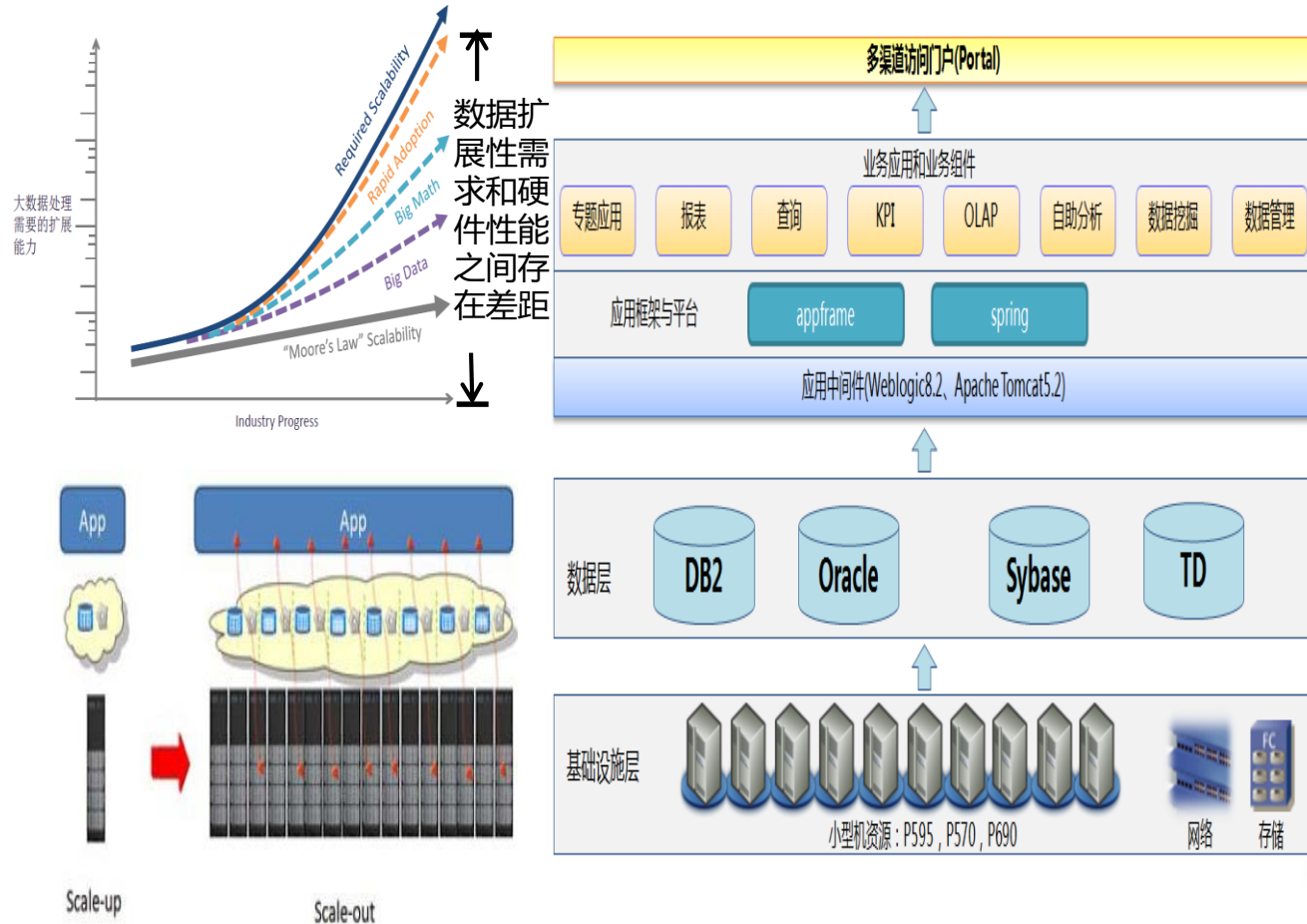
CERN: 大型强子对撞产生**1PB/s**的数据
SKA: 2015年存储需要**1EB**

云化IDC建设
催生了数据大集中

Facebook: 每天产生**50TB**的日志数据, 衍生分析数据超过**100TB**

“机器制造”和“人工制造”共同贡献了海量数据，集中式的数据中心建设加速数据集中。

传统的数据处理面临新的问题，呼唤新技术

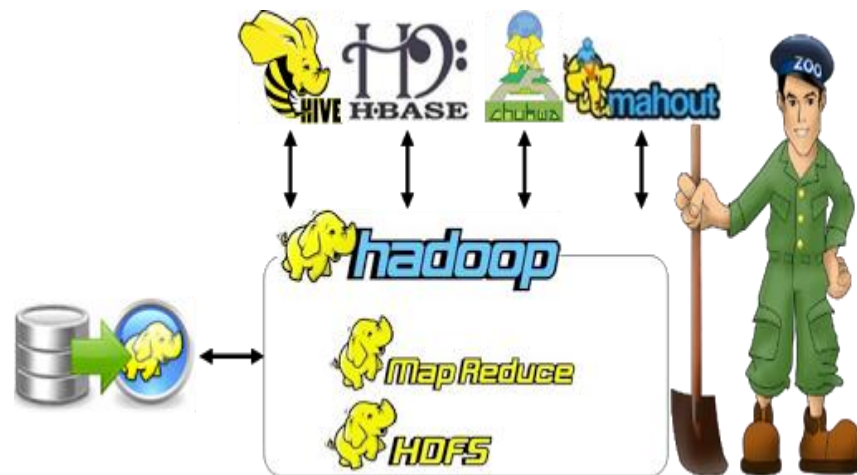


- 海量数据的高存储成本
- 大数据量下的数据处理性能不足
- 流式数据处理缺失
- 有限的扩展能力
- 单一数据源

新的业务需求，需要新的大数据处理平台

揭开Hadoop的面纱

- Hadoop是Apache基金会的一个项目总称，主要由HDFS、MapReduce和HBase等组成。HDFS是对Google GFS的开源实现，MapReduce是对Google MapReduce的开源实现，HBase是Google BigTable的开源实现。
- Hadoop 来源于其创始人Doug Cutting的儿子给一头黄色大象取的名字。
- Hadoop最初只与网页索引有关，迅速发展成为分析大数据的领先平台。



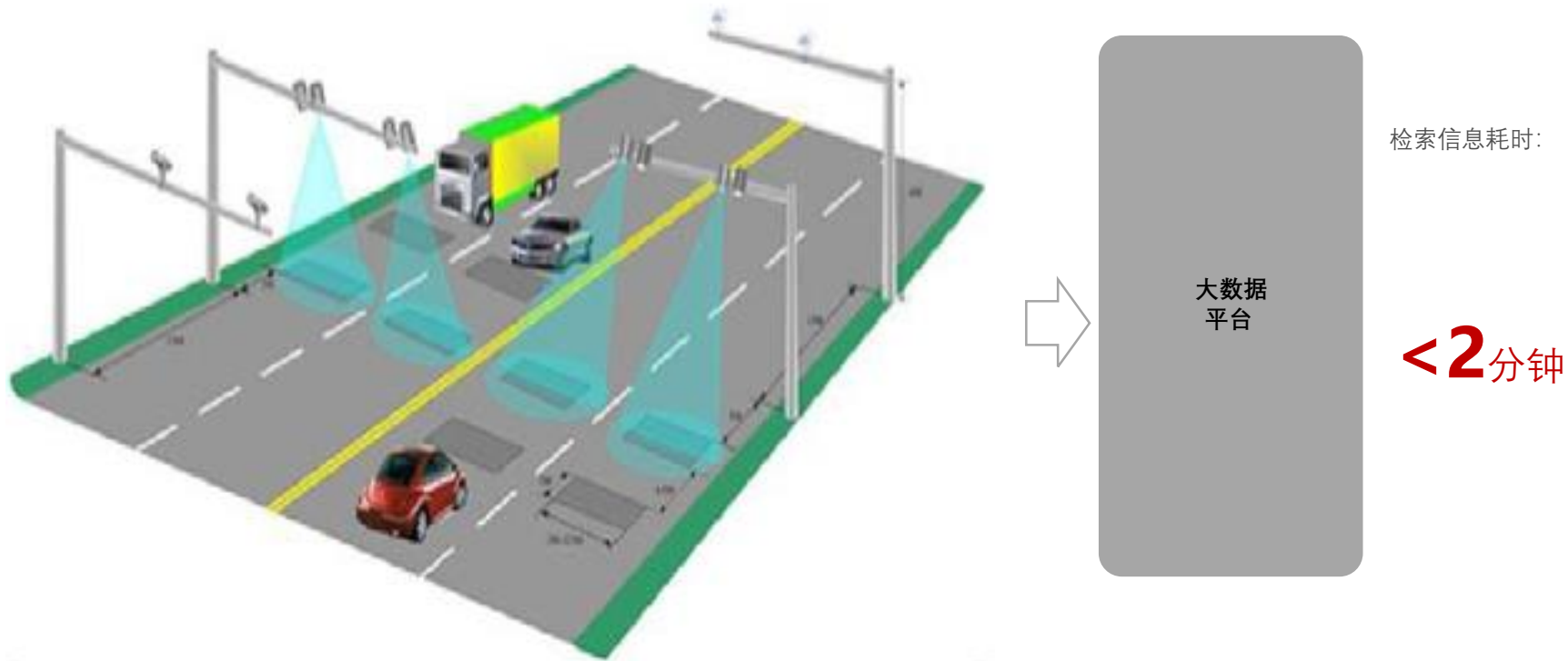
金融：大数据让银行更了解客户与识别潜在风险



公安：大数据实现在海量数据中快速检索出价值信息

国内某大城市1天的卡口数据：**3000+**万条记录，**1+**亿张照片，约**100TB**

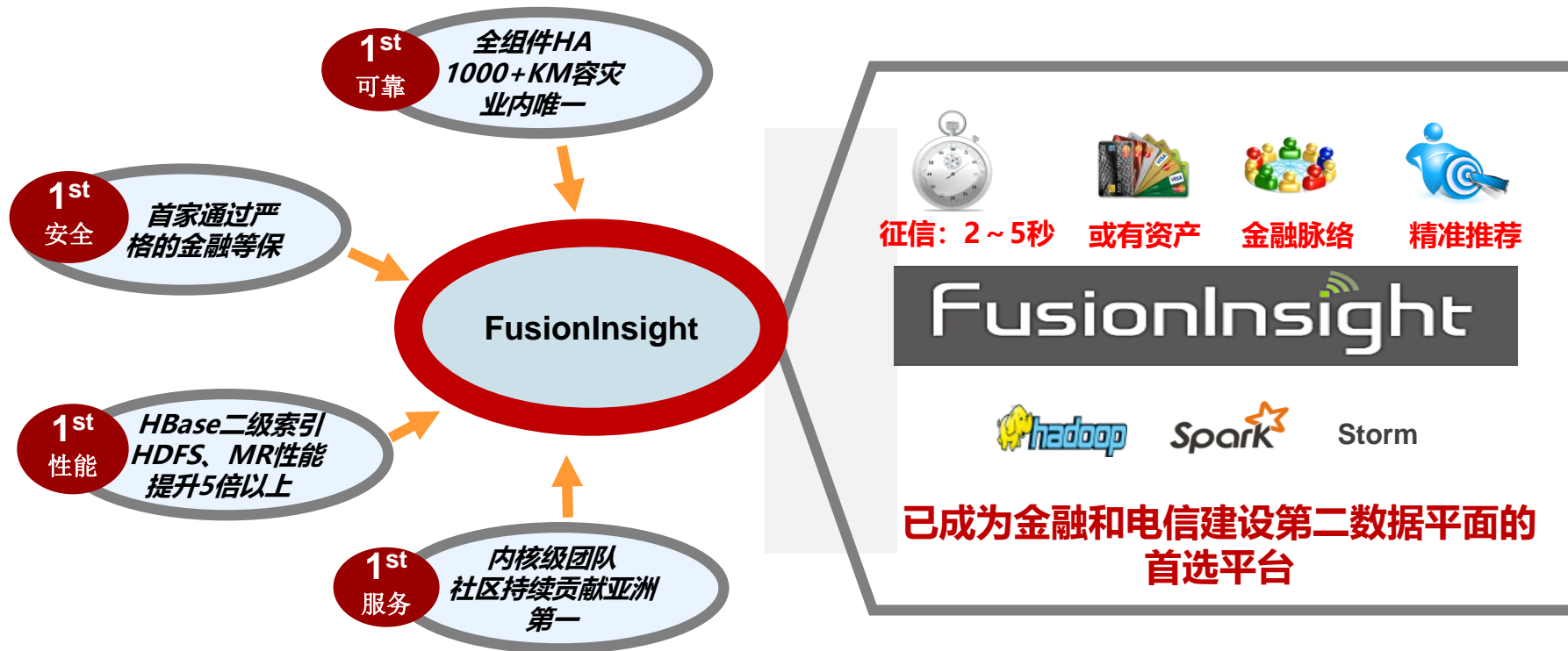
传统数据库平台模糊检索耗时：**1~2天**



华为大数据方案



FusionInsight:让数据“慧”说话



企业版的关键在与工程团队的能力

某大银行CIO：“我们把大数据应用视作是生命线，肯定是采用企业版，因为搞开源软件不是我们的主业。”

在选合作伙伴的时候，我们一定考虑门当户对，因为**只有**有力的合作伙伴才能保证**会使用**3-5年的供应、合作安全”



会定位内核级问题 (拔尖的个人)

定位内核级问题的团队 (依赖团队而不是精英个人)

能够独立完成支撑关键业务特性的内核级开发



HUAWEI

能够带领社区，引领社区完成面向未来的内核级特性开发

能够创建新的社区顶级项目，并且得到生态系统认可

强大的掌握代码的团队，才能造就成功的企业级数据平台软件

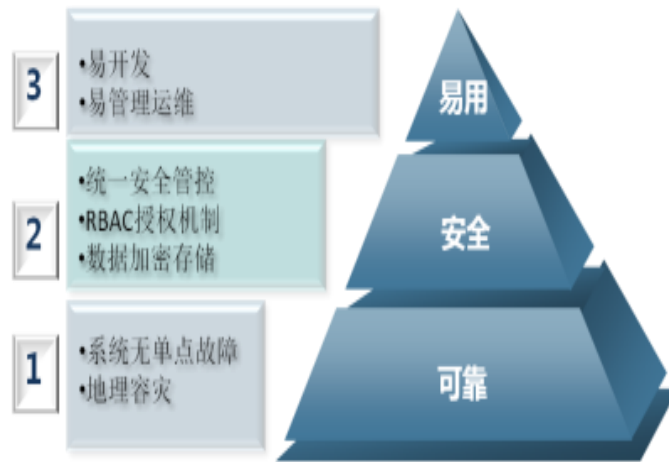
Hadoop社区全球贡献第4，开源到企业级蜕变

华为团队社区问题/补丁贡献

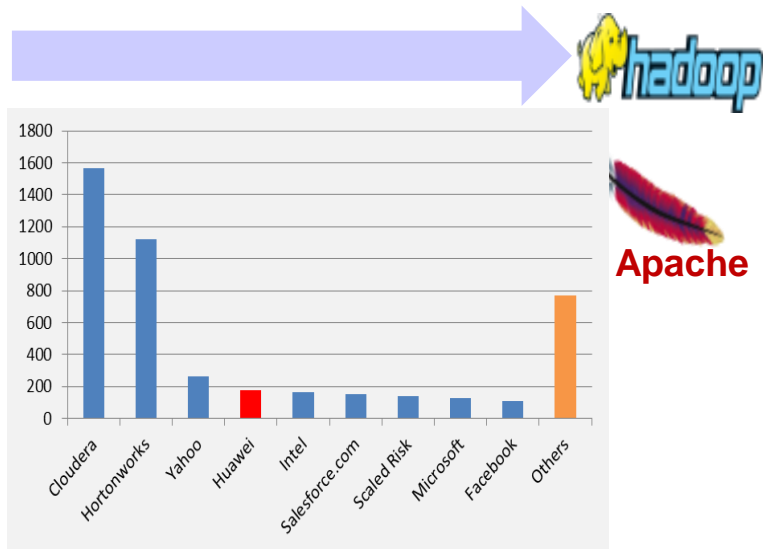
采纳社区精华，去除开源Bug:

- 谨慎选择稳定基线版本
- 认真评估高版本补丁影响范围评估和回合策略
- 采用数万个测试用例，确保企业版本稳定性

年份	提交	解决
2011	110	98
2012	189	155
2013	207	178



2013年社区贡献最新统计，全球第四，亚洲第一



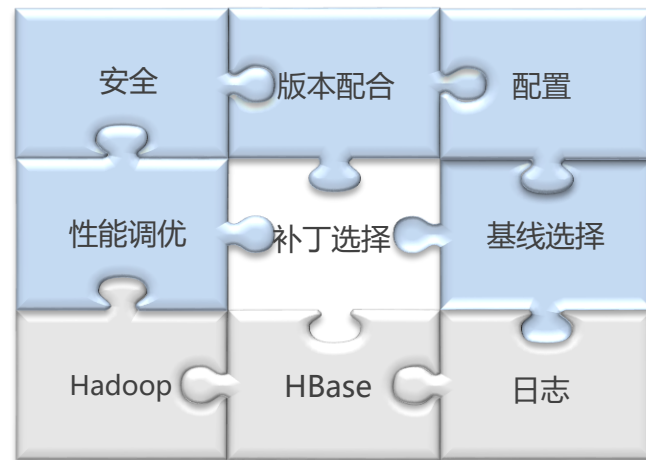
No.4 & No.1

8 Hbase Contributors

1 Hbase Committer

1 HDFS Committer

1 HDFS PMC



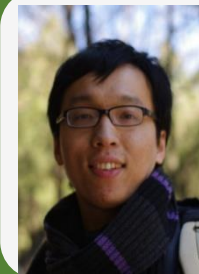
数据分析和挖掘领域顶尖人才，多项创新研究成果



Prof. Qiang Yang 杨强
中央研究院 Noah's Ark Lab 主任
世界级数据挖掘和人工智能专家
香港科技大学教授
IEEE Fellow, IAPR Fellow

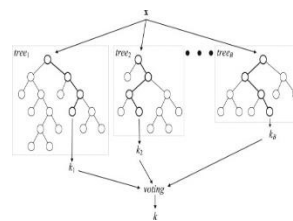


Dr. Hang Li 李航
中央研究院 Noah's Ark Lab 首席科学家
原微软亚洲研究院主任研究员
多个国际会议领域主席
已发表过上百篇学术论文
个人拥有20多项美国发明专利

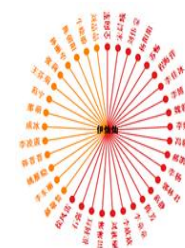


戴文渊 中央研究院诺亚方舟实验室主任，2012实验室首席科学家。
引领国际“数据”科学的研究，特别是在数据挖掘、机器学习和人工智能等领域享有盛誉

倾向预测



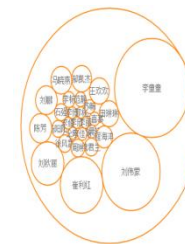
人物刻画



主题提取

$$\text{information-gain} \quad w(t_k) = - \underbrace{\sum_C P(C) \log P(C)}_{\text{"current" entropy}} + \underbrace{P(t_k) \sum_C P(C|t_k) \log P(C|t_k) + P(\bar{t}_k) \sum_C P(C|\bar{t}_k) \log P(C|\bar{t}_k)}_{\text{right hand side}}$$

关系估计



特征管理分析



自动特征构建

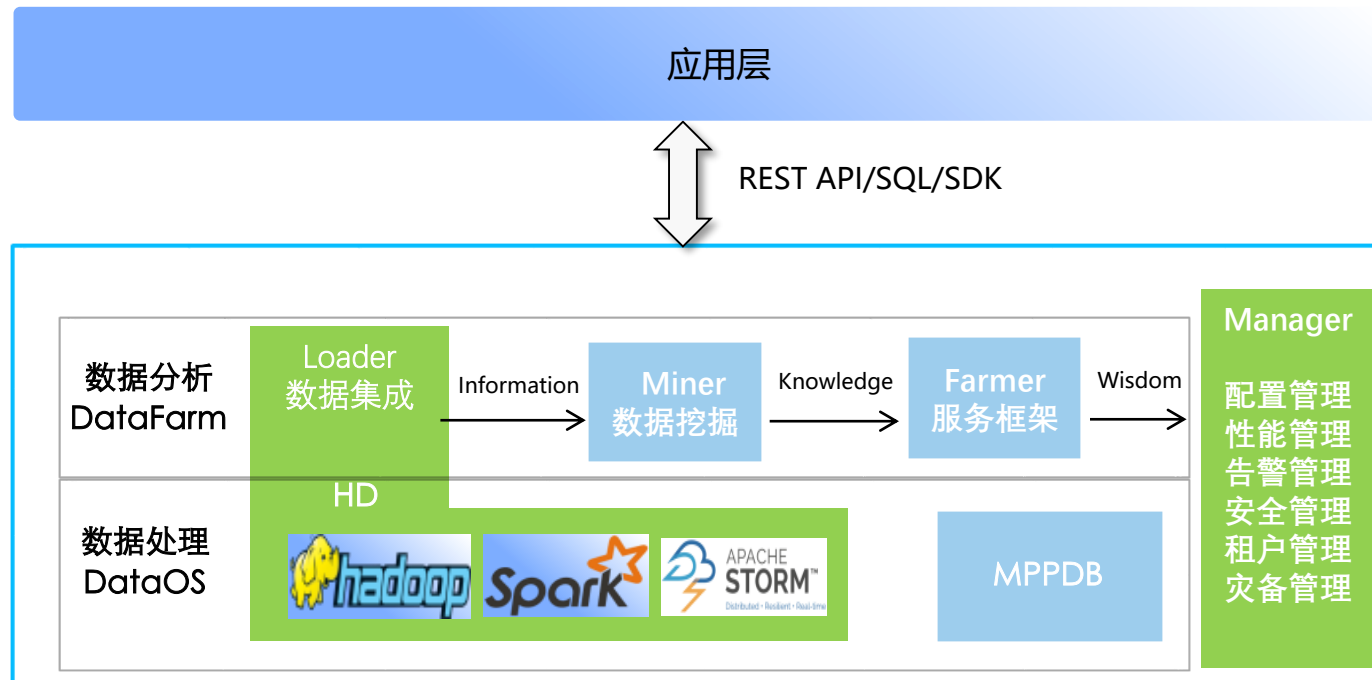


全面覆盖大数据领域的关键技术

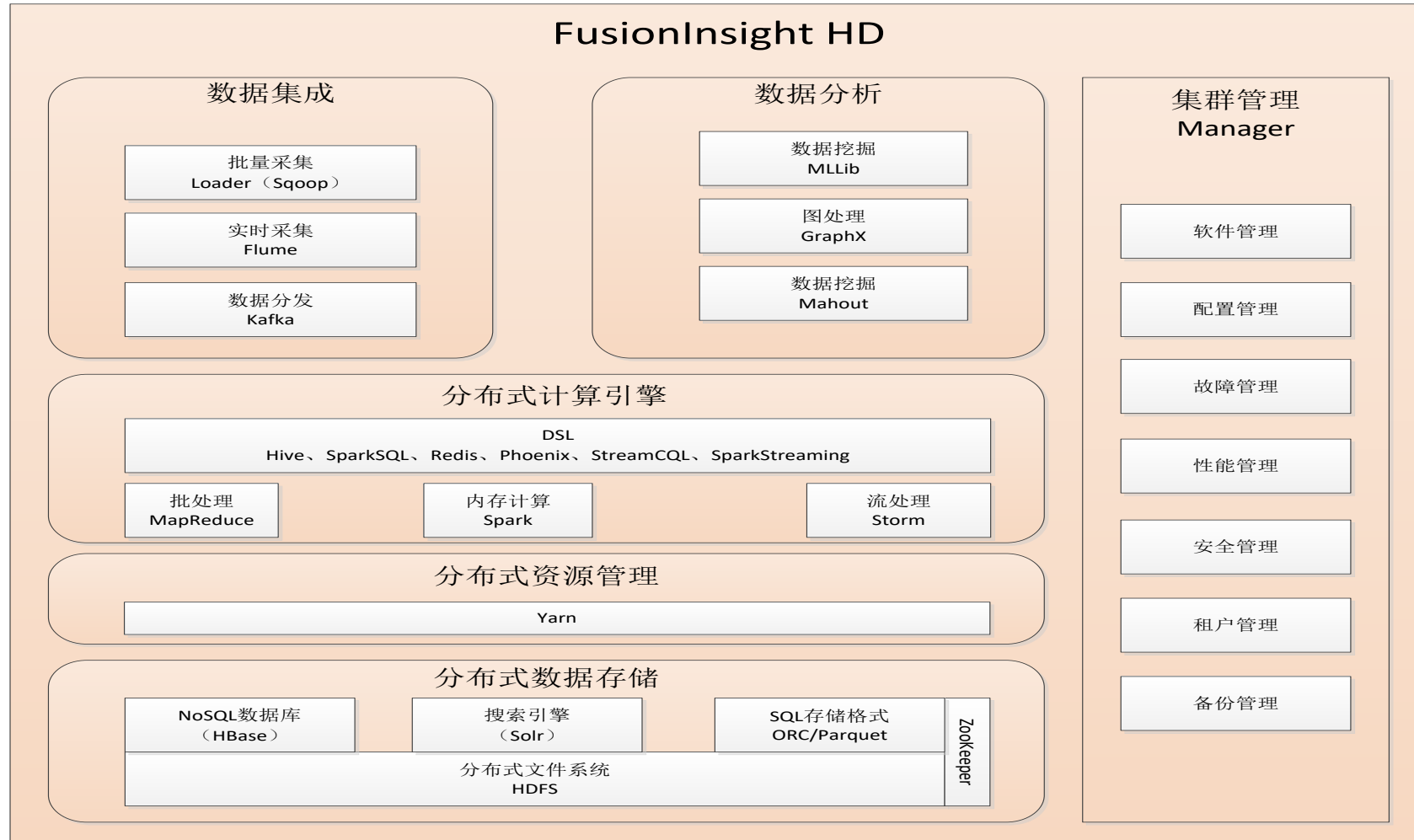


FusionInsight 产品族

- ✓ FusionInsight HD的Hadoop层提供大数据处理环境，基于社区开源软件增强，按照场景选择业界最佳实践
- ✓ FusionInsight Manager是一个分布式系统管理框架，管理员可以从单一接入点操控分布式集群，包括系统管理（OM/NTP/灾备）、数据安全管理和数据治理



FusionInight HD产品架构



易运维：向导式安装

The screenshot shows the FusionInsight Hadoop Manager installation wizard. The top navigation bar includes steps: 1. Select Version, 2. Set SSH Access, 3. Discover Hosts, 4. Set Rack, 5. Select Services, 6. Define Topology, 7. Configuration, and 8. Confirm. The 'Discover Hosts' step is circled in red. Below, the 'Configuration' step is also circled in red. The configuration interface shows a task description: 'Select services, roles, and hosts, and view and configure cluster parameters.' There are dropdown menus for 'type' (set to 'All'), 'Service' (set to 'KrbServer'), 'Role', and 'Host'. A search box is also present. A table lists parameters for the 'KrbServer' service:

Parameter	Value	Description
kdc_timeout	1000	[Desc] Maximum time (in milliseconds) allowed for a Java process to connect to the KDC. The a...
LDAP_SEARCH_TIMEOUT	3000	[Desc] Maximum waiting duration of kerberos server search principal on ldap server. [Default] 3...
max_retries	3	[Desc] Maximum number of retries allowed for a Java process to connect to the KDC. The acce...

Buttons for 'Back' and 'Next' are visible at the bottom right.

易安装：

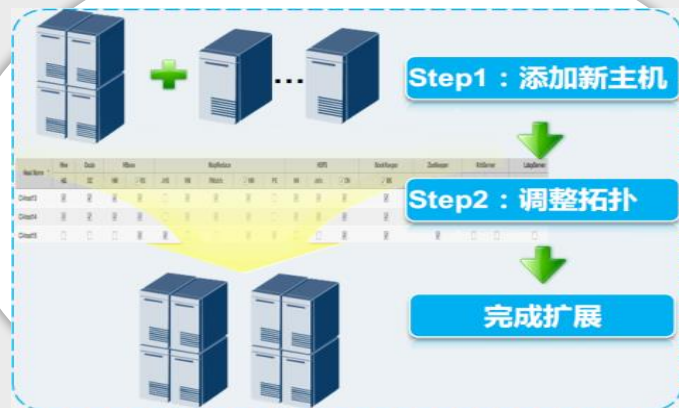
- 向导式操作，Step by step引导完成安装；
- 自动环境准备，内置脚本完成操作系统配置
- 分布式并行安装，10分钟完成集群安装。

易配置：

- 模板化配置：配置参数整合为模板，配置简单明了
- 关联配置自动同步，避免配置修改引发的多处人工修改

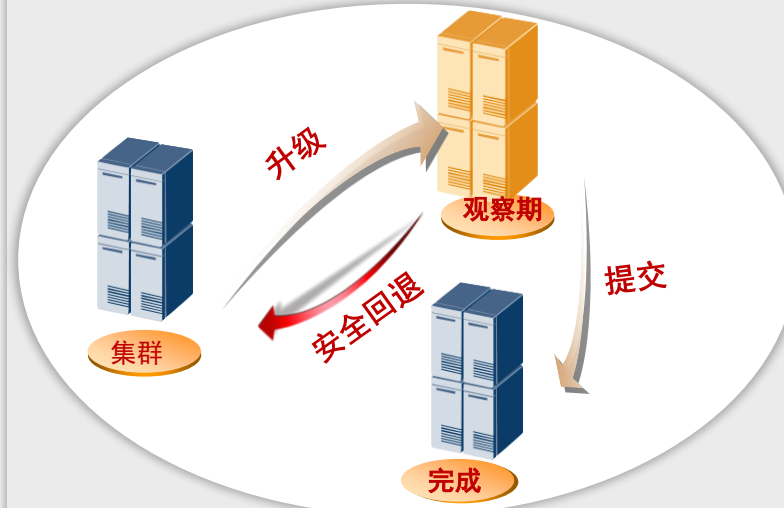
易运维：图形化快速升级和平滑扩容向导

平滑在线扩展



- 向导式界面操作
- 在线扩展，不影响业务
- 支持在线调整集群拓扑
- 各组件灵活扩展

快速升级



- 向导式界面操作
- 升级过程异常则自动安全回退
- 支持观察期手工触发安全回退
- 稳定运行后确认提交升级
- 6分钟完成集群升级

华为开发者社区 (大数据)

开发者社区FusionInsight HD开发资料地址:

http://developer.huawei.com/ict/cn/rescenter/CMDA_FIELD_FUSION_INSIGHT_HD?developlan=O

ther

开发资源 > 大数据 > FusionInsight HD

FusionInsight HD

支持的语言: Other

了解产品

论坛

V1.0

通用文档:

FusionInsight HD 产品介绍	下载	HBase架构原理	下载
Hbase开发指导文档	下载	HDFS架构原理	下载
HDFS开发指导文档	下载	Hive架构原理	下载
Hive开发指导文档	下载	Kerberos架构原理	下载
MapReduce和YARN架构原理	下载	MapReduce和YARN开发指导文档	下载
Spark架构原理	下载	Spark开发指导文档	下载
大数据基础知识	下载		

示例:

大数据 MapReduce Java Demo	下载
-------------------------	----

远程实验室

在华为开发者社区可以申请试用大数据环境，体验大数据平台开发。

远程实验室地址 (<http://esdkremotelab.huawei.com/RM/Diagram/Index/e20a331b-b738-427b-be44-ba8e36b605f4?diagramType=Topology>)

HUAWEI DEVELOPER

产品和解决方案 API 服务与支持 开发者大赛 动手实验室

开发装备 远程实验室 IDE开发工具

“零”出差，“零”成本远程实验室

对接真实设备远程调试

华为远程实验室为开发者提供了7*24小时、免费、云化实验室环境

[进入实验室](#)

远程实验室是什么

真实设备，全业务体验
华为远程实验室提供的所有实验室环境均为真实物理设备环境，为开发者提供和现场一致的调试环境，覆盖华为多种产品。

随时随地，远程接入
开发者可以在任何时间、任何地方，远程接入华为远程实验室进行开发调试。

按需使用，自助预约
华为远程实验室所有环境均支持提前预约时间，开发者可以在线上自助预约远程实验室，预约到期前系统会自动释放。

开发问题支持渠道

The image shows two screenshots from the Huawei Developer website. The top screenshot displays the navigation menu with '支持' (Support) highlighted. Below it, three support channels are listed: '支持' (Support), '问题与需求' (Issues and Requirements), and '支持热线' (Support Hotline). The '支持热线' section is annotated with a red arrow and the text '电话热线' (Telephone Hotline). The bottom-left screenshot shows the '提交问题' (Submit Issue) form, and the bottom-right screenshot shows a forum post for 'FusionInsight HD'.

<https://devcenter.huawei.com/usercenter/addDefectTicket>

<http://developer.huawei.com/ict/forum/forum-400089.html>

积极参与大数据方面的生态建设工作

HUAWEI

大赛日程 大赛奖项 技术领域 评选规则 学习视频 大赛论坛

在一起，梦飞扬
Together: Dream Big, Fly High

2017 华为开发者大赛
2017 HUAWEI DEVELOPER CHALLENGE

2017年华为开发者大赛

2016年上海联通和
华为公司共同举办
沃+数据应用大赛



Unite 'n Upgrade

上海联通沃 + 开放数据应用大赛

华为 FusionInsight 让数据“慧”说话

关于“沃+”

大赛数据命题

大赛奖金

大赛日程

大赛申明

大赛组织

参赛培训

报名通道

总结

01

华为大数据认证概述

02

HCIA-面向大数据集群搭建与管理的入门级认证

03

HCIP-面向大数据开发的中级认证

04

HCIE-涵盖AI与机器学习的专家级认证

05

从认证看大数据的发展趋势

Thank You

全国统一咨询热线：400-690-6115

北京|上海|广州|深圳|天津|成都|重庆|武汉|济南|杭州|西安

easthome.com