

华为云 技术 私享会



华为云
技术
私享会

华为云开放式技术与实践

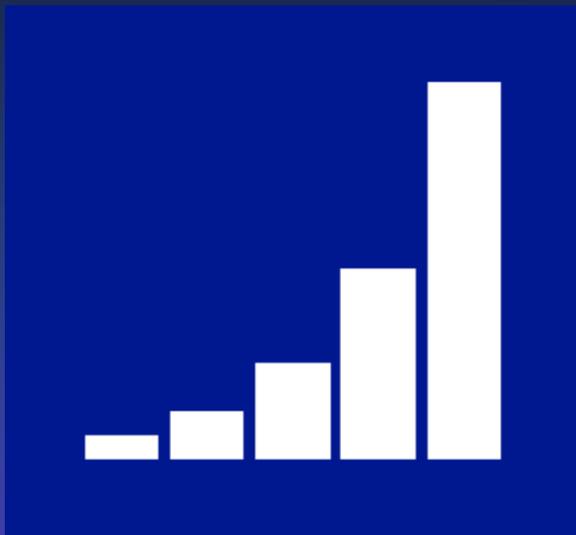
赵军



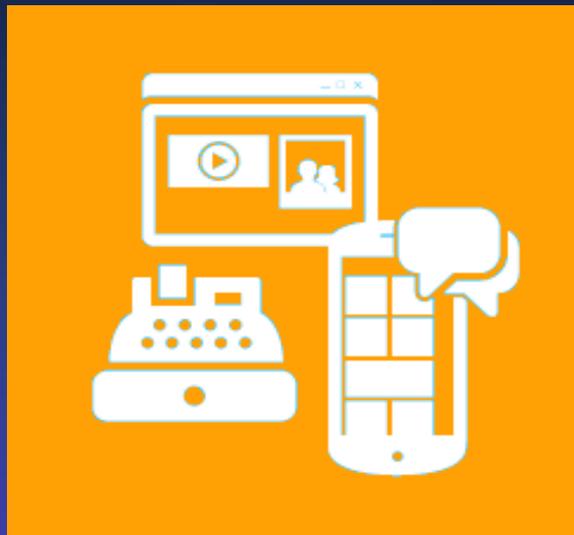
目录

- 华为云EI 总体架构
- MRS技术介绍 & 使用技巧
- 应用案例 & 未来规划

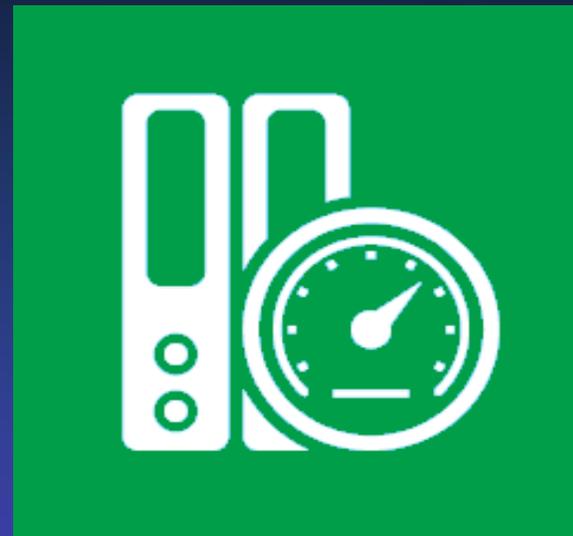
大数据带来的挑战



数据量越来越大



数据种类越来越多



数据产生的速度
越来越快

Hadoop:大数据的开源解决方案

- Apache开源项目，起始于2005年
- 针对解决数据量大，种类多，产生速度快的问题
- 强大的开源社区支持
- 日益丰富的生态系统

批处理

Map reduce

内存计算

Spark

SQL

Hive

Nosql

Hbase

流式

Kafka/Storm

搜索

Solr

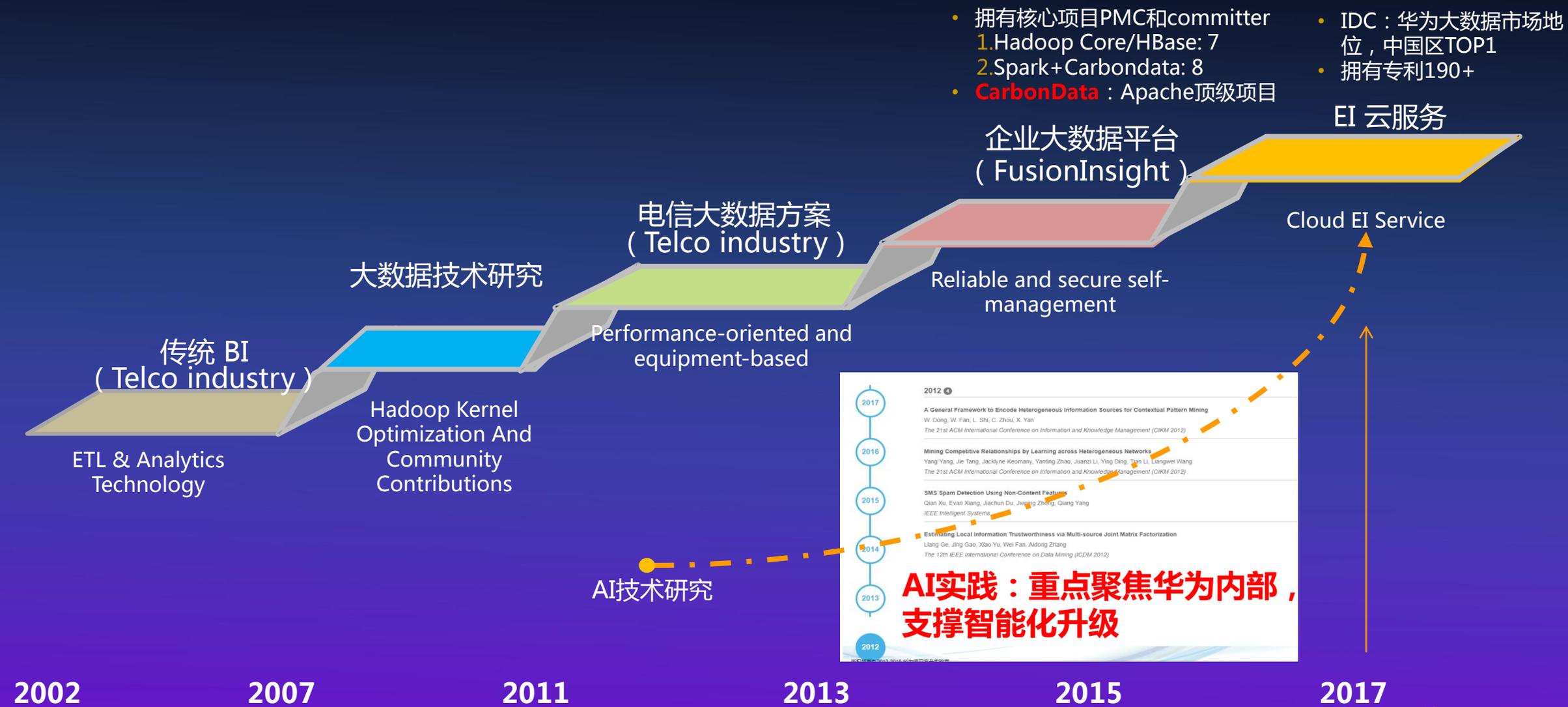
脚本

Pig

YARN: 资源管理系统

HDFS 文件系统 (Hadoop Distributed File System)

华为云大数据的发展历程



- 拥有核心项目PMC和committer
 - 1.Hadoop Core/HBase: 7
 - 2.Spark+Carbondata: 8
- **CarbonData** : Apache顶级项目
- IDC : 华为大数据市场地位, 中国区TOP1
- 拥有专利190+

AI实践：重点聚焦华为内部，支撑智能化升级

2017	A General Framework to Encode Heterogeneous Information Sources for Contextual Pattern Mining W. Dong, W. Fan, L. Shi, C. Zhou, X. Yan The 21st ACM International Conference on Information and Knowledge Management (CIKM 2012)
2016	Mining Competitive Relationships by Learning across Heterogeneous Networks Yang Yang, Jie Tang, Jacklynne Keomany, Yanling Zhao, Juandi Li, Ying Ding, Tan Li, Liangwei Wang The 21st ACM International Conference on Information and Knowledge Management (CIKM 2012)
2015	SMS Spam Detection Using Non-Content Features Qian Xu, Evan Xiang, Jiachun Du, Jierong Zhang, Qiang Yang IEEE Intelligent Systems
2014	Estimating Local Information Trustworthiness via Multi-source Joint Matrix Factorization Liang Ge, Jing Gao, Xiao Yu, Wei Fan, Aidong Zhang The 12th IEEE International Conference on Data Mining (ICDM 2012)
2013	
2012	

华为云EI：让企业更智能

物流



电商



IOT



制造



金融



智慧城市



运营商



Enterprise Intelligence

行业场景

车联网

智能物流

智能制造

智能水务

.....

数据分析

Data Lake Factory(workflow编排/调度/DataIDE)

MRS

CloudStream

UQuery
数据查询

AI基础平台
MLS/DLS/GES

AI通用服务
视觉/语音

数据计算

一站式大数据集群

实时流计算

CloudTable
表格存储

ElasticSearch
搜索引擎

DWS
数据仓库

数据存储

DLC、OBS

数据集成

DIS(数据接入)

CDM(数据迁移)

使能

- 基于实践提供丰富的行业算法和预置模型
- 规模的AI实践经验，使能企业快速步入智能时代

敏捷

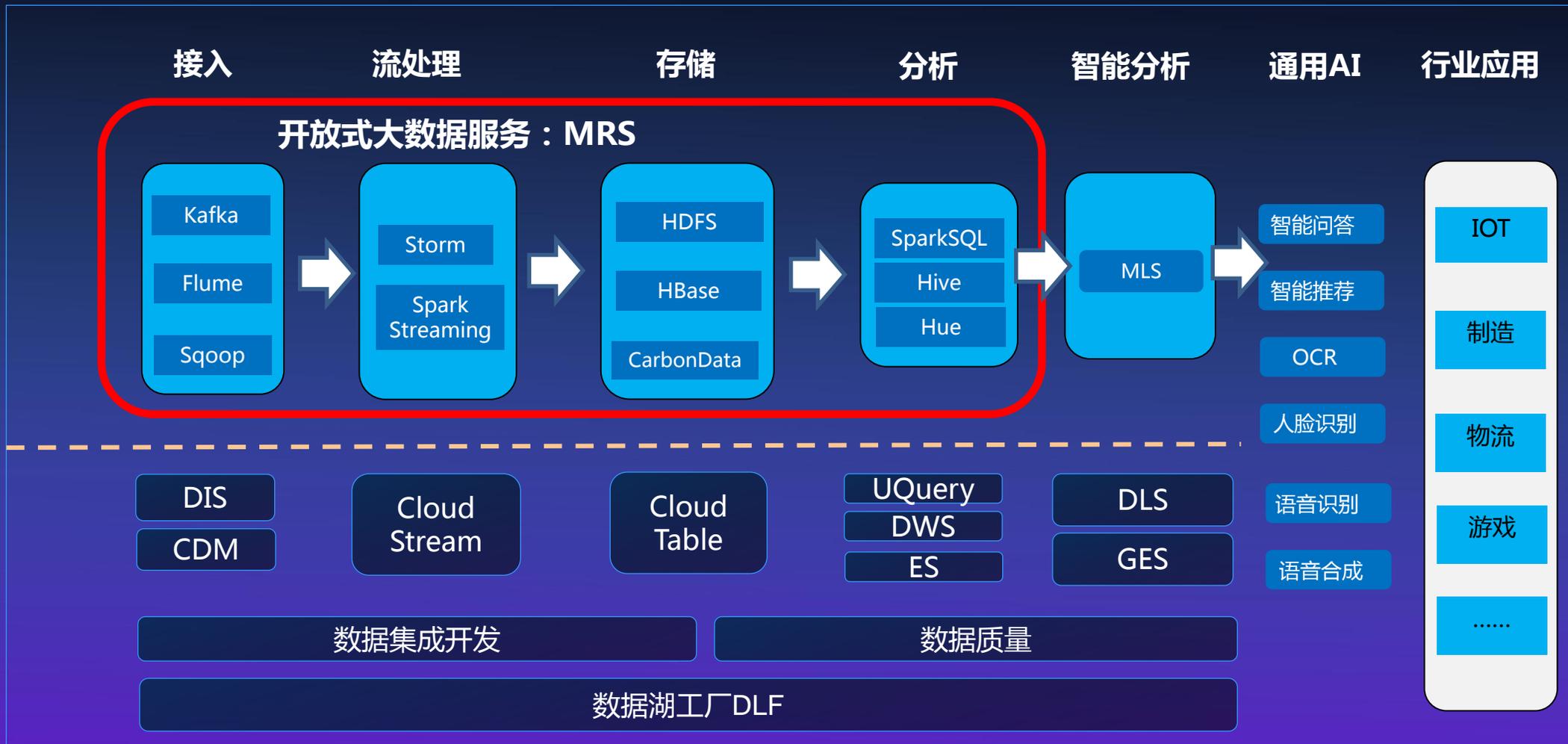
- 强大的SQL能力，业务移植便捷
- 丰富的工具支持，开发运维高效
- 完善的云服务，平滑迁移

可信

- 全组件HA、异地容灾、金融等保
- 开放共赢，可信赖的合作伙伴
- 企业级的大数据云服务，开放接口

华为EI云服务：围绕数据开发流程，构筑全生命周期全栈服务

华为云
技术
私享会



开放式大数据服务MRS的特点

开放性：通用接口，通用组件，使用和开源一样

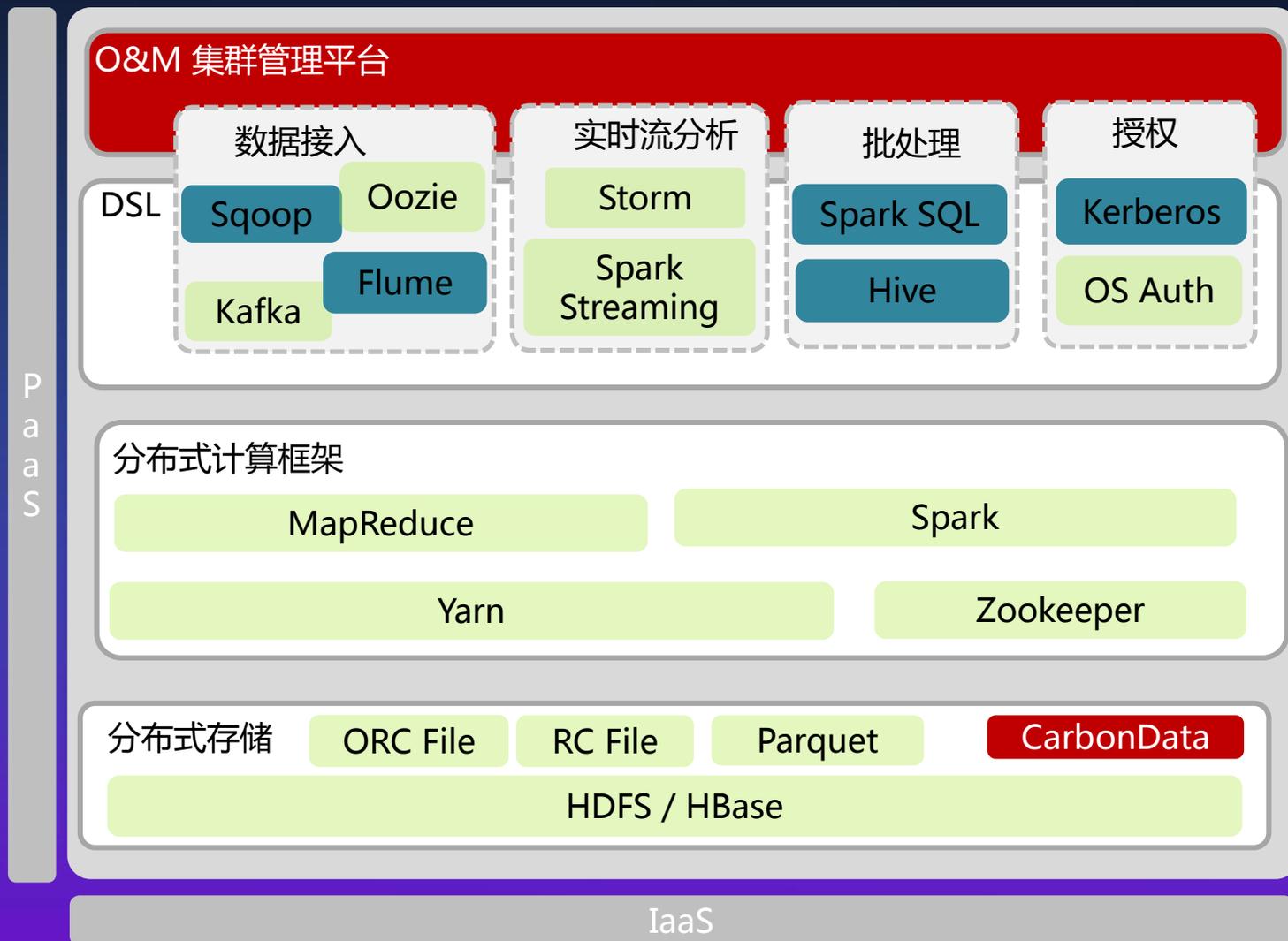
自主性：可控性强，用户**可视化**看到集群的所有信息（资源/状态/数据）并**控制**，可最短路径解决问题。

隔离性：所有的资源完全属于用户控制，具备物理隔离的能力，无需担心其他用户的争抢。
组件之间的访问，都在独立的私网（VPC）内访问，干扰少

全球化：标准的大数据服务，可以方便的在全球各主流公有云上部署

华为云MRS — 云时代全栈企业级大数据服务

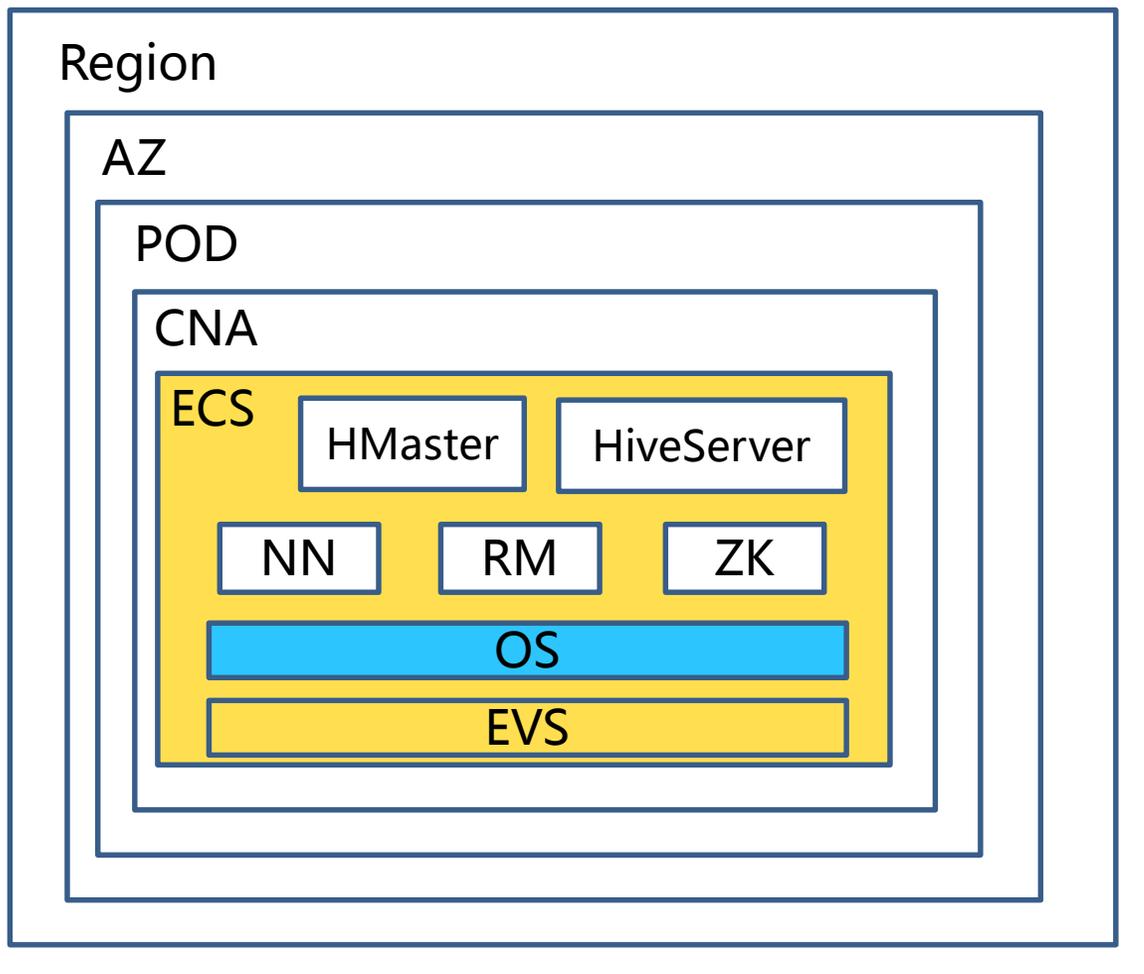
■ 轻量增强 ■ 深度增强 ■ 完全自研



- 企业级
- 易运维
- 高安全
- 低成本

MRS-可靠性

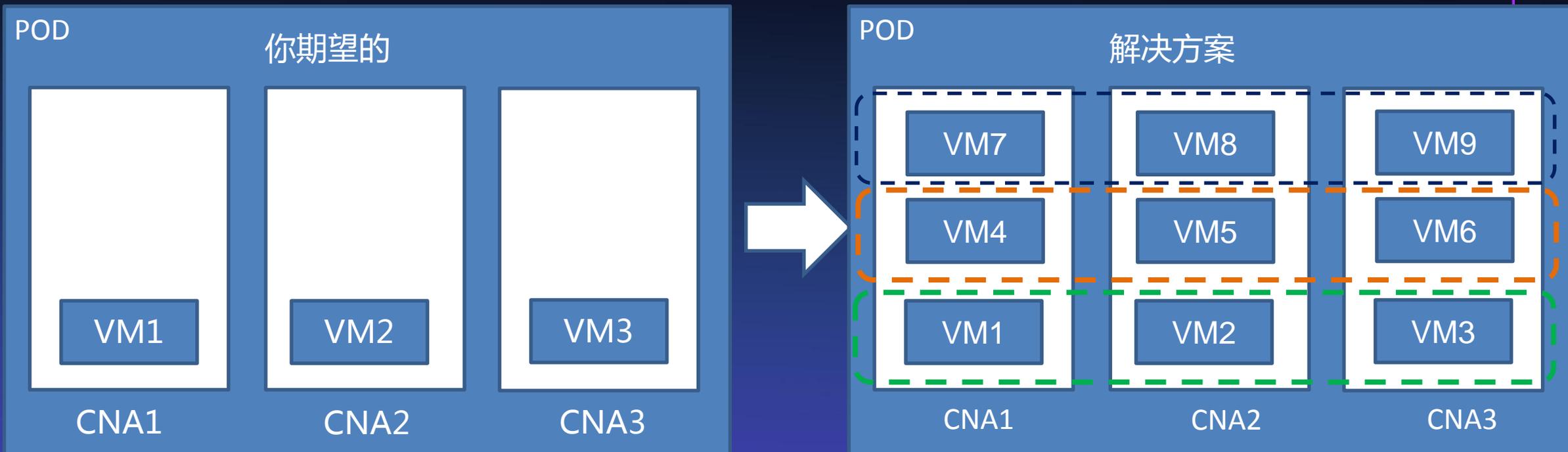
公有云



特点：

- 公有云上的技术栈很深
- 可靠性链条很长，需要考虑：
 - 进程可靠性
 - 操作系统可靠性
 - **虚拟机可靠性**
 - 存储可靠性
 - 网络可靠性
 - AZ故障
 - Region故障
 - 客户端故障

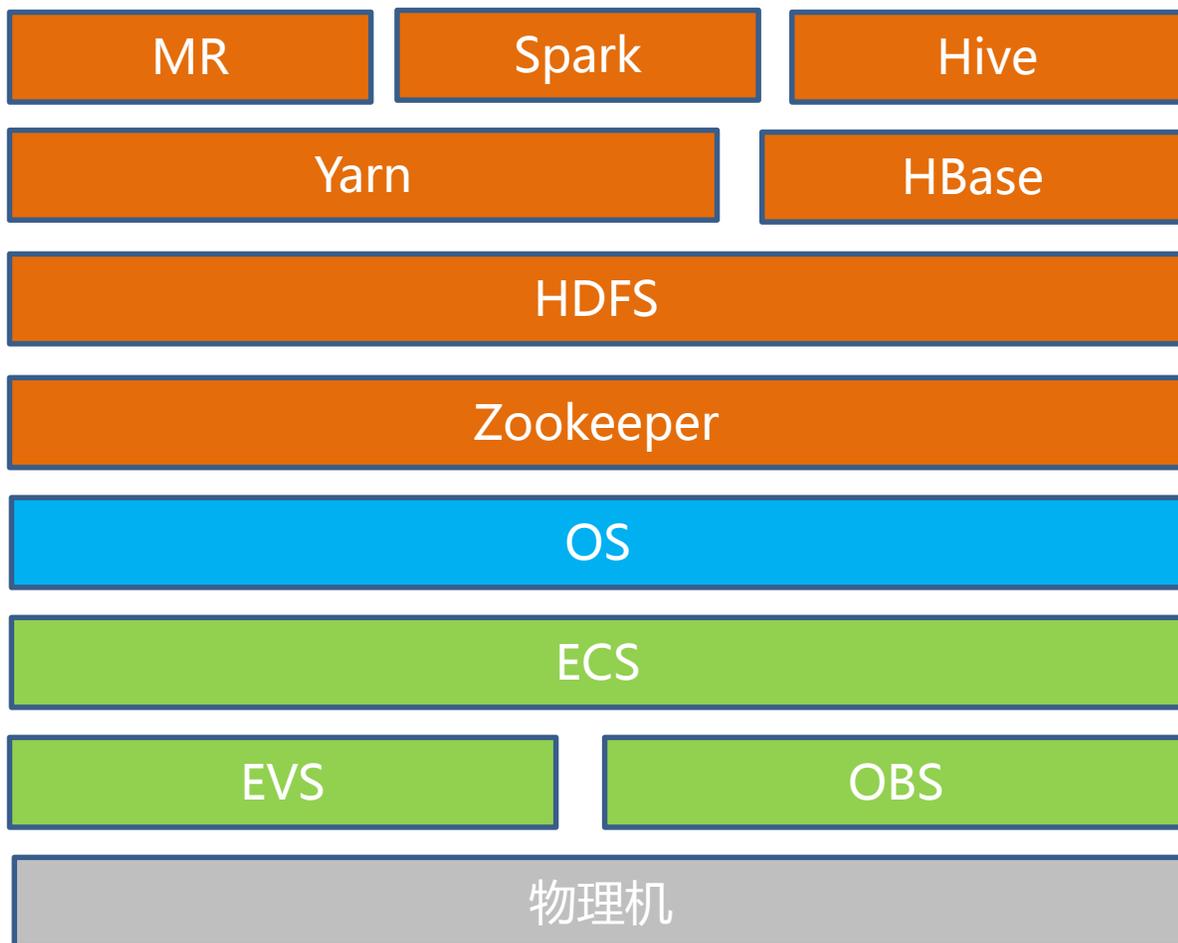
MRS可靠性-虚拟机可靠性



解决方案—反亲和

- 1.虚拟机创建使用反亲和组
- 2.反亲和组数据分配算法
 - 2.1 数据副本仅存在一个反亲和组内
 - 2.2 数据块复制时，将多个副本同时迁移到其他反亲和组内

MRS-性能



Hadoop的调优：

- 1、块大小
- 2、容器重用
- 3、短路径读
- 4、其他

OS的调优：

- 1、网卡多队列
- 2、sriov驱动
- 3、中断均衡
- 4、其他

IaaS的调优：

- 1、SAS/SATA/SSD，磁盘个数
- 2、OBS的HTTPS加密算法，OBS的快速上传特性，buffer使用内存
- 3、CPU和内存比，网卡多队列
- 4、其他

MRS-安全

用户/用户组管理

- 访问MRS Manager用户权限管理和审计
- Hadoop组件WEBUI服务支持单点登录认证

Kerberos认证

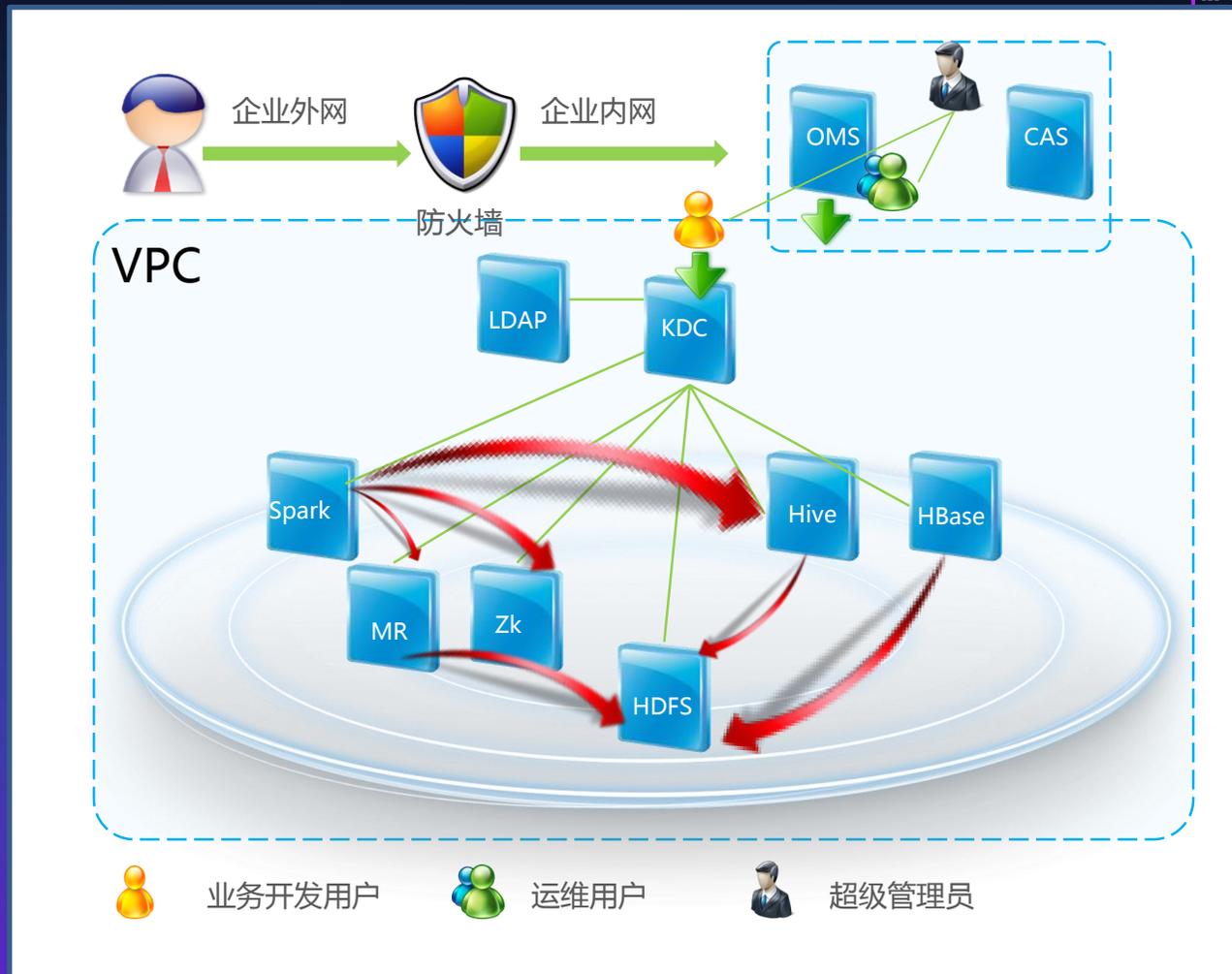
- Hadoop服务用户鉴权管理
- 用户鉴权信息存储HA

数据分角色存储

- Hive和HBase中数据分用户，角色受控存储，用户间数据隔离
- 细粒度的权限管理，可按照表/按列控制访问权限

数据加密

- Hive, MR和HBase在固化到HDFS中的数据均加密存储
- 细粒度的安全管理，可按照表/按列加密



MRS-可视化集群管理，运维便捷

The screenshot displays the MRS Manager interface with several key components:

- Service Overview (服务概况):** A table showing the health status of various services.

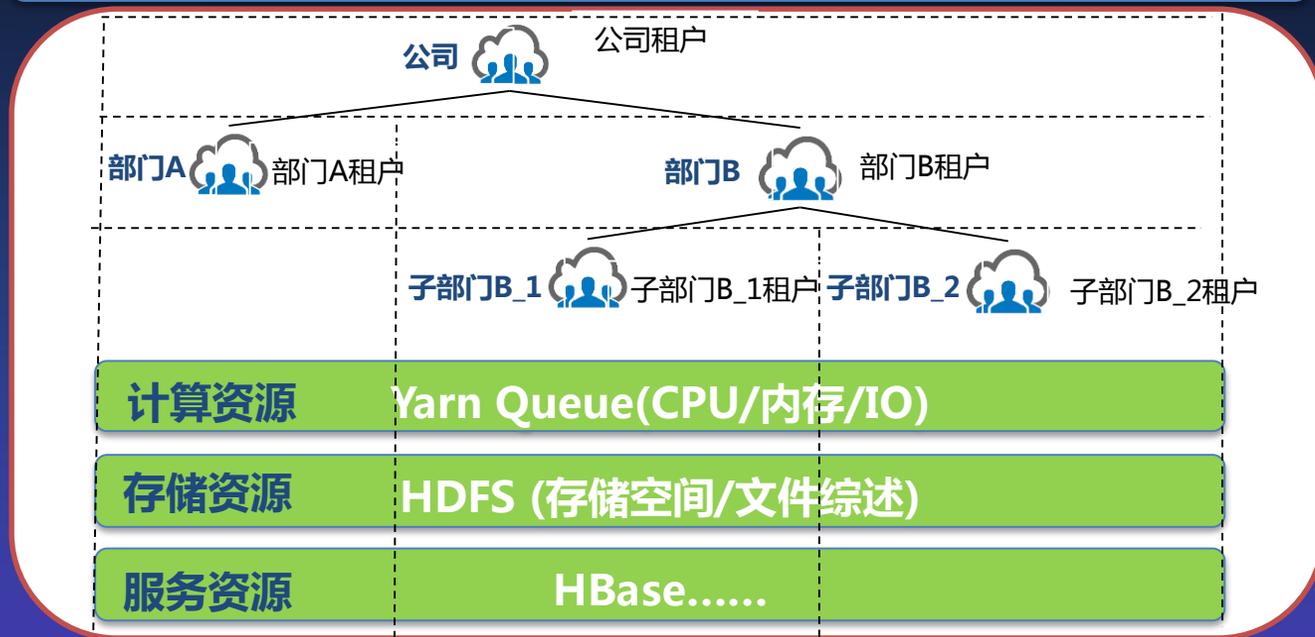
服务	健康状态	角色数
DBService	良好	DBServer: 2
HBase	良好	HMaster: 2 RegionServer: 3 ...
HDFS	良好	Zkfc: 2 DataNode: 3 Journa...
Hive	良好	MetaStore: 2 WebHCat: 2 H...
- Cluster Node Health (集群主机健康状态统计):** A donut chart showing the distribution of node health. A callout points to the '良好' (Good) segment, labeled '集群节点状态'.
- Network Speed (主机网络读写速率分布):** A bar chart showing network I/O rates. A callout points to the '0-200MB/s' bar, labeled '集群服务状态'.
- Disk Usage (主机磁盘占用率分布):** A bar chart showing disk usage percentages. A callout points to the '0-25%' bar, labeled '集群节点状态'.
- Memory Usage (主机内存占用率分布):** A bar chart showing memory usage percentages. A callout points to the '0-25%' bar, labeled '内存资源占用分布'.
- Alerts (告警管理):** A list of alerts with columns for ID, description, severity, time, and service name. A callout points to a '致命' (Critical) alert, labeled '集群异常告警，一目了然'.

可视化集群管理，一切尽在掌握

- ◆ 一键式集群安装部署和扩容
- ◆ 服务组件配置项可视化
- ◆ 全面的系统监控与告警管理，服务状态、节点状态（CPU、内存、硬盘、网络等）一目了然，
- ◆ SNMP/FTP/SYSLOG标准接口与企业已有管理系统无缝对接

MRS-可视化的多级租户管理，与企业组织结构

多层次租户管理



一站式租户管理

The screenshot shows the '添加租户' (Add Tenant) interface. The breadcrumb navigation is '租户管理 > 租户管理 > 添加租户'. The form includes the following fields and options:

- 名称:** test
- 租户类型:** 叶子租户 非叶子租户
- 动态资源:** Yarn
- 默认资源池容量:** 10 %
- 储存资源:** HDFS
- 存储空间配额:** 50000 MB
- 服务:** + 关联服务

Below the form, a table lists associated services:

服务名称	
HBase	独占

可视化的多级租户管理，与企业组织结构相匹配，简化系统资源分配与管理

- ◆ 多级的租户模型
- ◆ 一站式管理租户资源管理
- ◆ 租户资源使用情况实时监控

MRS使用技巧—按需使用 VS 包周期

MRS的计费模式：ECS的费用+EVS的费用+MRS的服务费（3.9分/小时/核，>=18核封顶）

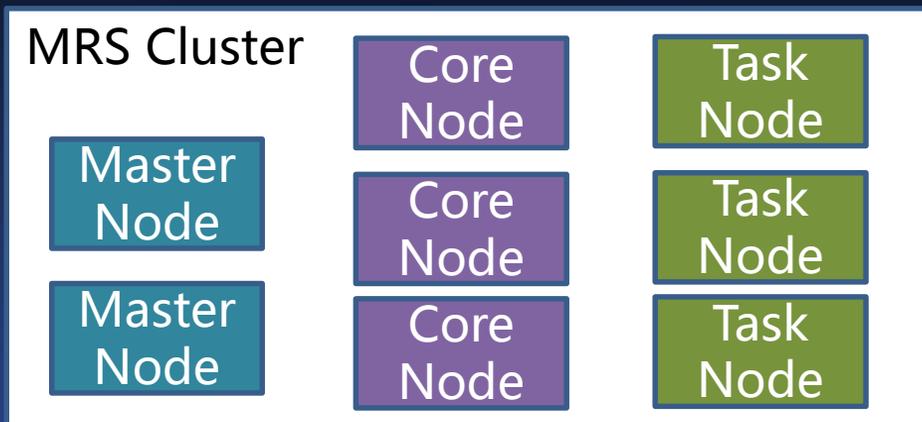
集群创建时间+所有任务Σ（数据加载时间+数据分析时间）+ 集群删除时间 < 24小时，可以考虑按需使用



VS



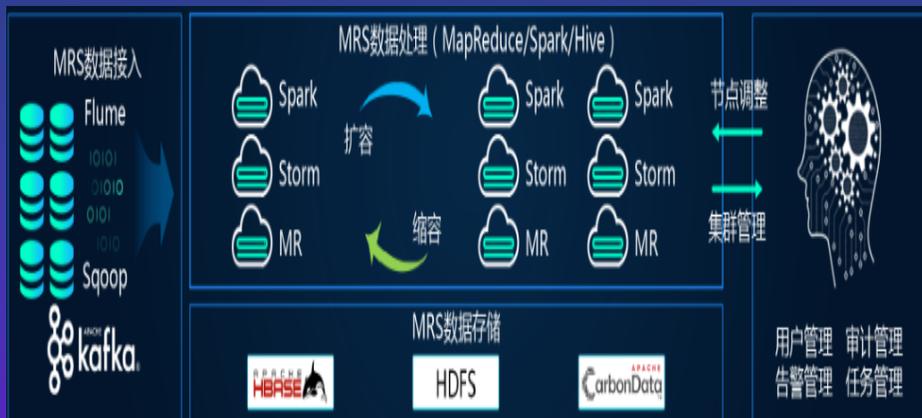
MRS使用技巧—TaskNode & 弹性伸缩



MasterNode：管理节点，存放集群元数据，故障则集群不可用

CoreNode：核心节点，计算和存储数据，最重要的节点

TaskNode：任务节点，不存放数据，只提供计算能力的节点



自动弹性伸缩，保护业务平稳运行

自动：通过实时监测和业务运行的预判，对计算节点自动弹性伸缩

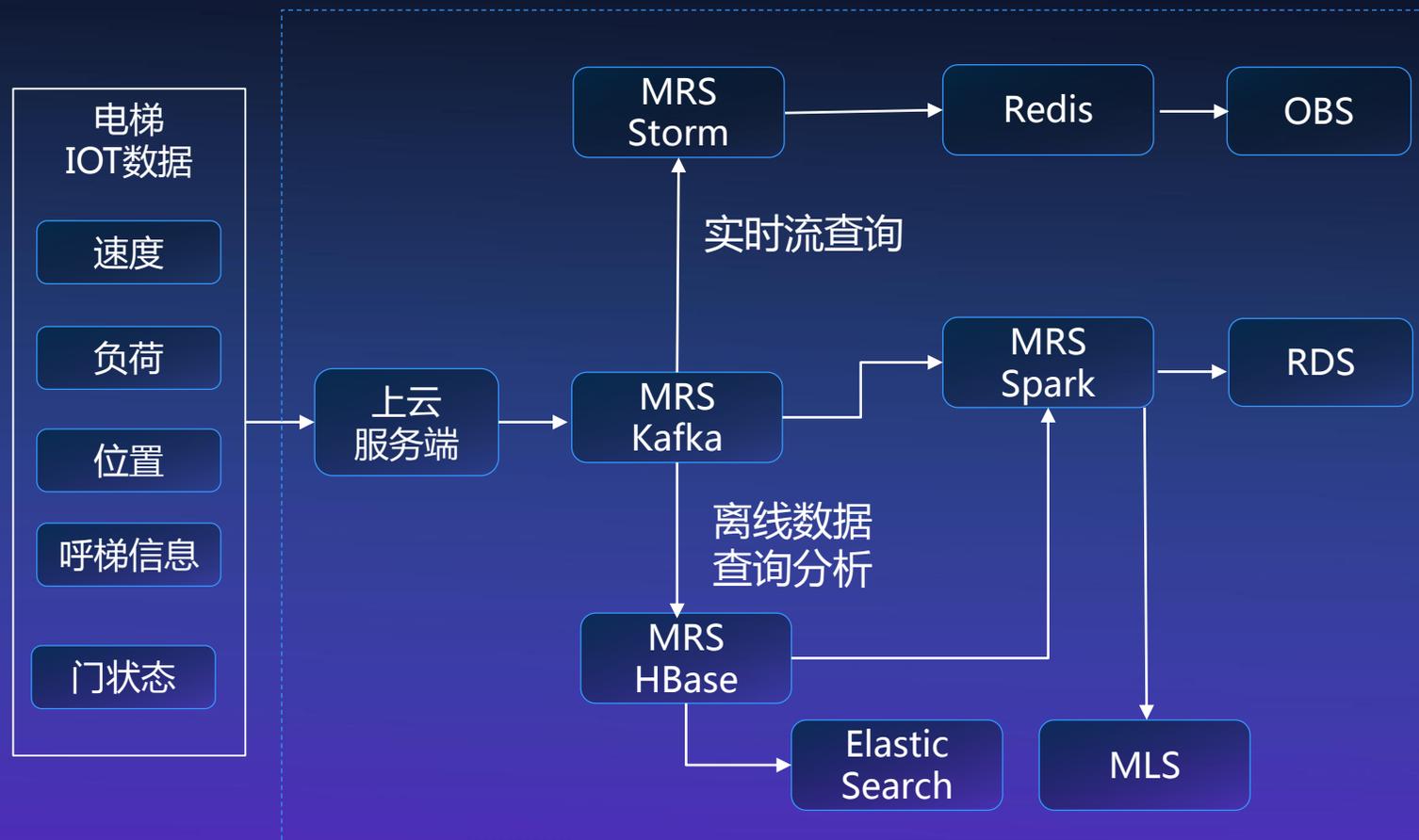
灵活：可自定义触发策略和多条件叠加，支持集群规模和频率保护

智能：自动计算和选择最优退服节点，自动对扩容节点均衡任务

MRS使用技巧—选用合适的虚拟机规格和规模

场景	说明	虚拟机规格	规模	成本
功能开发/测试	做功能开发联调测试，非商用	s2.xlarge.2(4U8G)	1M+1C	1.99元/小时
小规模商用	适合存储数据量在不大，计算业务量不大的场景	通用计算型： s2.xlarge.4(4U16G)以及更大	2M+nC (n>=3)	6.29元/小时起
大规模商用	数据存储量大，计算量大的场景	磁盘增强型： d2.4xlarge(16U128G)以及更大	2M+nC (n>=3)	33.98元/小时起
高性能计算	数据存储量小，但是计算量大的场景	计算优化型： hc2.2xlarge (8U16G) 以及更大	2M+nC (n>=3)	15.6元/小时起

助力杭州新再灵打造智能梯联网



业务场景

- 使用MRS Kafka组件，解决每秒数万条实时电梯监控数据的上云队列暂存。
- 使用MRS Storm组件，实时处理Kafka数据，并导入redis实现实时电梯数据监控。
- 使用MRS Hadoop组件，定期从kafka抽取数据，将海量数据入库MRS hbase。
- 借助MRS HBase极速点查及批量scan能力，配合ElasticSearch向业务层提供电梯数据极速检索查询服务。
- 使用MRS Spark组件，轻松构建批量计算应用，定时从HBase获取数据进行多维数据分析，生成云梯运行报表供上层业务部门使用；同时实现定时冷数据备份功能，全量数据备份至低成本OBS存储服务。
- 使用MLS机器学习对数据进行价值挖掘，包括电梯故障预测性维护。

客户声音：“通过新再灵与华为云MRS团队的努力，目前上线的新再灵大数据平台已通过实践检验，具备大规模推广的技术能力，MRS服务的技术能力值得点赞。”

华为云MRS的未来规划

- BootstrapAction
- Auto Healing
- Auto Performance Improvement
- Intelligent Maintained

实践说明：车联网驾驶行为大数据分析

1. 业务分析

原始数据为车主的驾驶行为信息，包括车主在日常的驾驶行为中，是否急加速、急减速、空挡滑行、超速、疲劳驾驶等信息，通过Spark组件的强大的分析能力，分析统计指定时间段内，车主急加速、急刹车、空挡滑行、超速、疲劳驾驶等违法行为的次数。



2. 数据理解

驾驶员ID	车牌号	纬度	经度	速度	方向	地点	记录上报时间	急加速	急减速	空挡滑行	空挡滑行结束	空挡滑行时长	超速	超速结束	超速时长	疲劳驾驶	怠速	漏油	
likun1000003	华AVM936	32.05644	118.7776	72	211	211	2017/1/1 8:00											1	
haowei1000008	华A709GB	30.6786	104.0708	143	115	115	2017/1/1 8:00							1					
haowei1000008	华A709GB	30.6786	104.0708	120	115	115	2017/1/1 8:00								1	0			
zouan1000007	华A58M83	28.21155	112.9795	77	177	177	2017/1/1 8:00												
haowei1000008	华A709GB	30.67837	104.0713	99	115	115	2017/1/1 8:00												
zengpeng1000000	华AZQ110	29.54317	106.5294	71	220	220	2017/1/1 8:00												
xiexiao1000001	华AEB132	29.54495	106.5337	105	83	83	2017/1/1 8:00												
shenxian1000004	华ADJ750	30.68204	104.0671	88	339	339	2017/1/1 8:00												
likun1000003	华AVM936	32.05554	118.777	82	211	211	2017/1/1 8:00												
panxian1000005	华AX542C	23.12174	113.3075	67	357	357	2017/1/1 8:00												
hanhui1000002	华AZI419	28.21583	112.98	97	16	16	2017/1/1 8:00												
duxu1000009	华AT75H8	38.50123	106.2092	104	118	118	2017/1/1 8:00												
zouan1000007	华A58M83	28.21086	112.9795	139	177	177	2017/1/1 8:00							1					
duxu1000009	华AT75H8	38.50068	106.2103	137	118	118	2017/1/1 8:00							1					
duxu1000009	华AT75H8	38.50052	106.2106	120	118	118	2017/1/1 8:00								1	1			

3. 输出结果样例

驾驶员ID	车牌号	急加速次数	急刹车次数	空挡滑行次数	空挡滑行时间总和	超速次数	超速时间总和	疲劳驾驶次数	怠速次数	漏油次数
hanhui1000002	华AZI419	401	444	327	2844	3349	31813	3997	433	371
zouan1000007	华A58M83	360	385	315	2997	3181	31248	3594	389	385
haowei1000008	华A709GB	321	314	255	2659	2639	25522	3204	312	318
xiezhil1000006	华A6CU11	255	310	254	2074	2535	23942	2931	312	279
panxian1000005	华AX542C	395	434	330	2930	3531	33946	4307	417	441
shenxian1000004	华ADJ750	374	356	297	2810	3126	31494	3767	383	366
likun1000003	华AVM936	341	354	291	3043	3044	28728	3552	347	376
zengpeng1000000	华AZQ110	340	344	272	2894	2763	25479	3274	284	337
xiexiao1000001	华AEB132	264	261	248	2525	2324	23434	2720	314	253
duxu1000009	华AT75H8	238	284	247	2632	2301	22338	2814	264	248

1

软件开发服务DevCloud

一站式云端DevOps平台，集成华为近30年研发实践和前沿理念

超低价格套餐供您选择 [了解详情>](#)

[免费体验](#) [价格计算器](#) [帮助文档](#)

 **精彩活动**
免费注册既享1000元大礼

 **新手入门&FAQ**
几分钟学会玩转DevCloud



2

NEW

车联网驾驶行为大数据分析

基于DevCloud和MRS服务 Spark程序开发实例

 2018.4.10-2018.5.31

[点击进入](#)

链接地址：https://activity.huaweicloud.com/devcloud_activitys/index.html

提醒：提前注册官网账户，并创建好MRS集群！

华为云
技术
私享会

THANK YOU