

ORACLE

Oracle 跨平台查询Big Data SQL

邹中凡

资深大数据架构师



议程

1. 什么是Big Data SQL
2. Big Data SQL架构与特性
3. Big Data SQL案例

Oracle Big Data SQL是什么

可以在Oracle数据库中:



简单访问任何Hadoop中数据



提供统一的安全机制访问所有数据源



使用Oracle 的SQL分析任何数据



有非常好的性能表现使用Big Data SQL的智能扫描

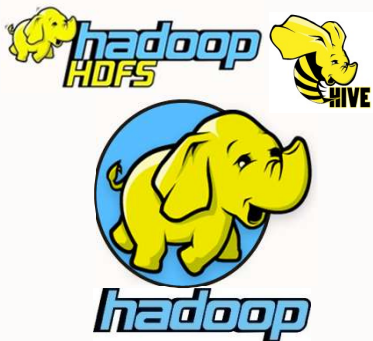


Big Data SQL目标：跨平台统一访问不同数据源

同时支持Hadoop + Kafka + NoSQL + RDBMS + 对象存储统一查询



Oracle Big Data SQL



Big Data SQL支持的平台

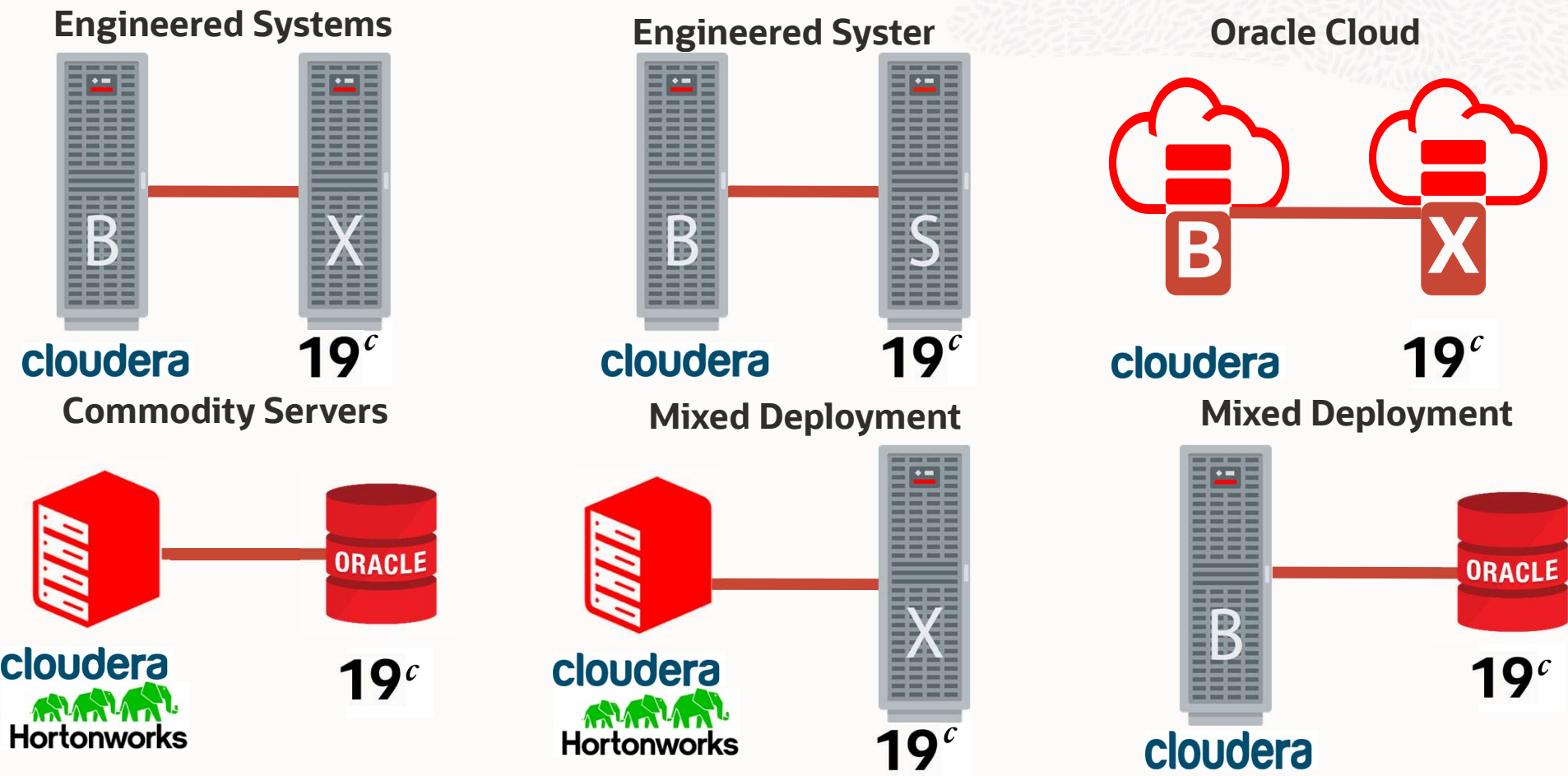
- Oracle数据库版本：12.1.0.2以上
- Hadoop版本：CDH 5.5以上 或 HDP 2.3以_

Oracle Database Version	Database Hardware	Hadoop Cluster Hardware	Hadoop Distribution and Version
12.1.0.2 or higher	Any Intel x86 64-bit system (Linux OL6, RHEL6)	Any Intel x86 64-bit system (Linux OL6, RHEL6)	• CDH* 5.5 and higher • HDP† 2.3 and higher
12.1.0.2 or higher	Oracle Exadata (Linux OL6)	Oracle Big Data Appliance (Linux OL6)	CDH 5.5 and higher

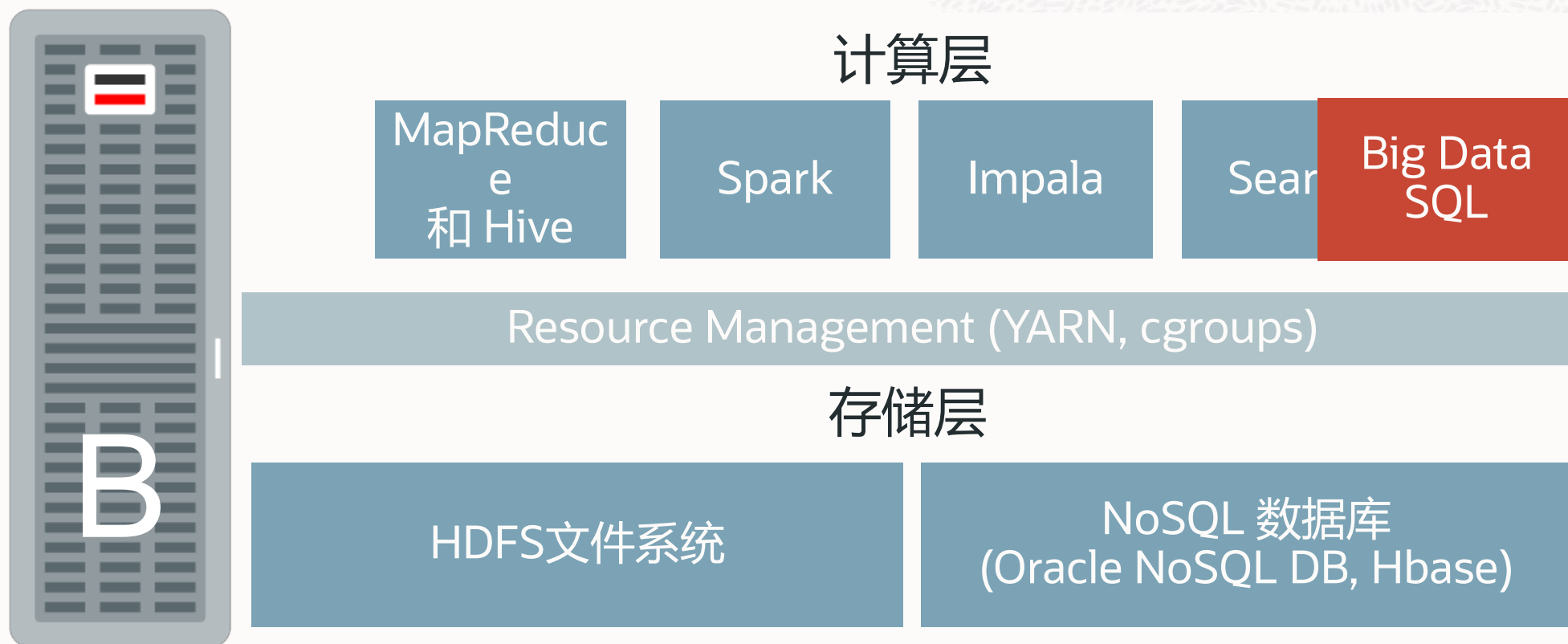
* CDH: Cloudera's Distribution Including Apache Hadoop

†HDP: Hortonworks Data Platform

Big Data SQL配置搭配



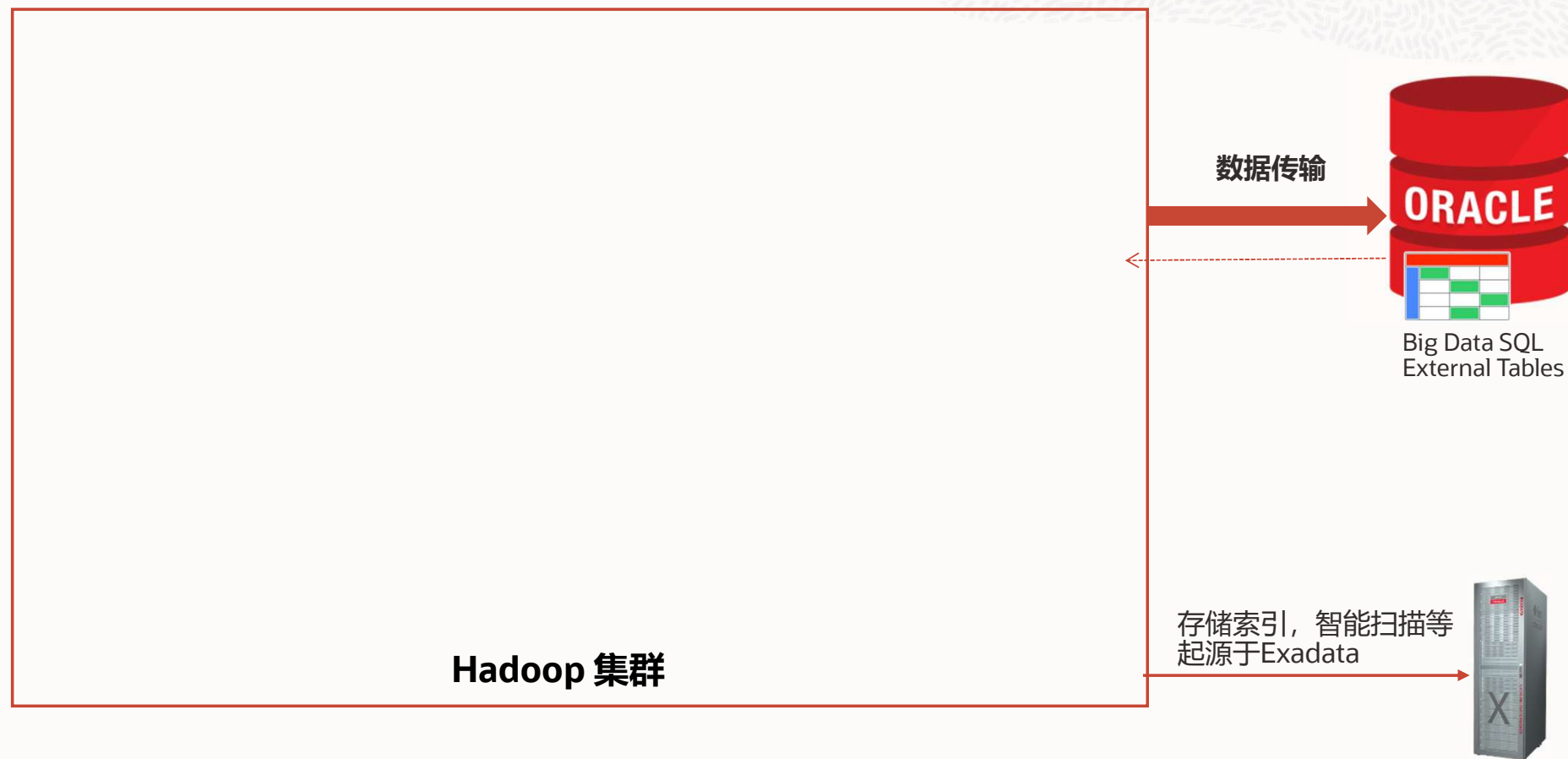
Big Data SQL是一种新的Hadoop计算引擎



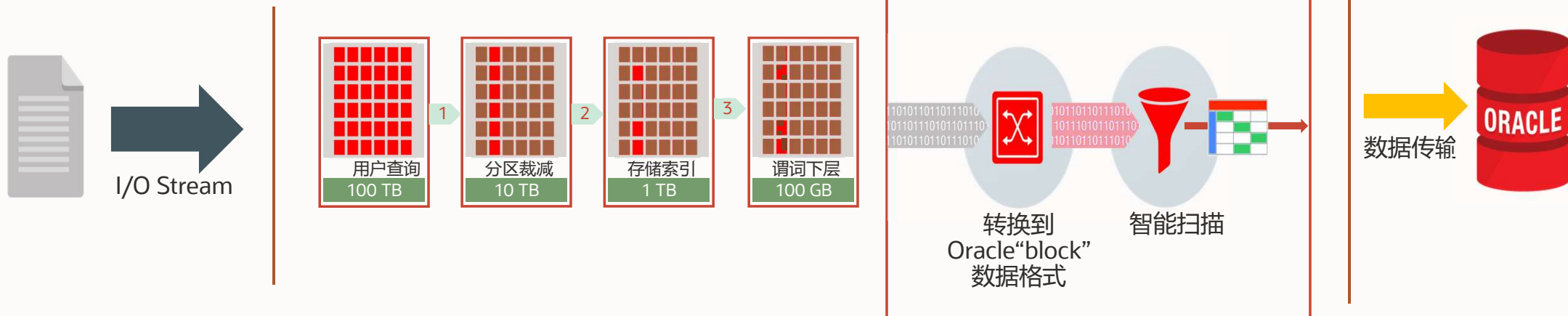
议程

1. 什么是Big Data SQL
2. Big Data SQL架构与特性
3. Big Data SQL案例

Big Data SQL架构



Big Data SQL查询执行路径



Big Data SQL的实现：外部表定义

```
CREATE TABLE movielog (  
  click VARCHAR2(4000))  
ORGANIZATION EXTERNAL (  
  TYPE ORACLE_HIVE  
  DEFAULT DIRECTORY DEFAULT_DIR  
  ACCESS PARAMETERS  
  (  
    com.oracle.bigdata.tablename=logs  
    com.oracle.bigdata.cluster=mycluster  
  )  
  REJECT LIMIT UNLIMITED;
```

3种新的外部表

- ORACLE_HIVE (从hive metadata中继承)
- ORACLE_HDFS (手工指定数据访问方式)
- ORACLE_BIGDATA (支持对象存储)

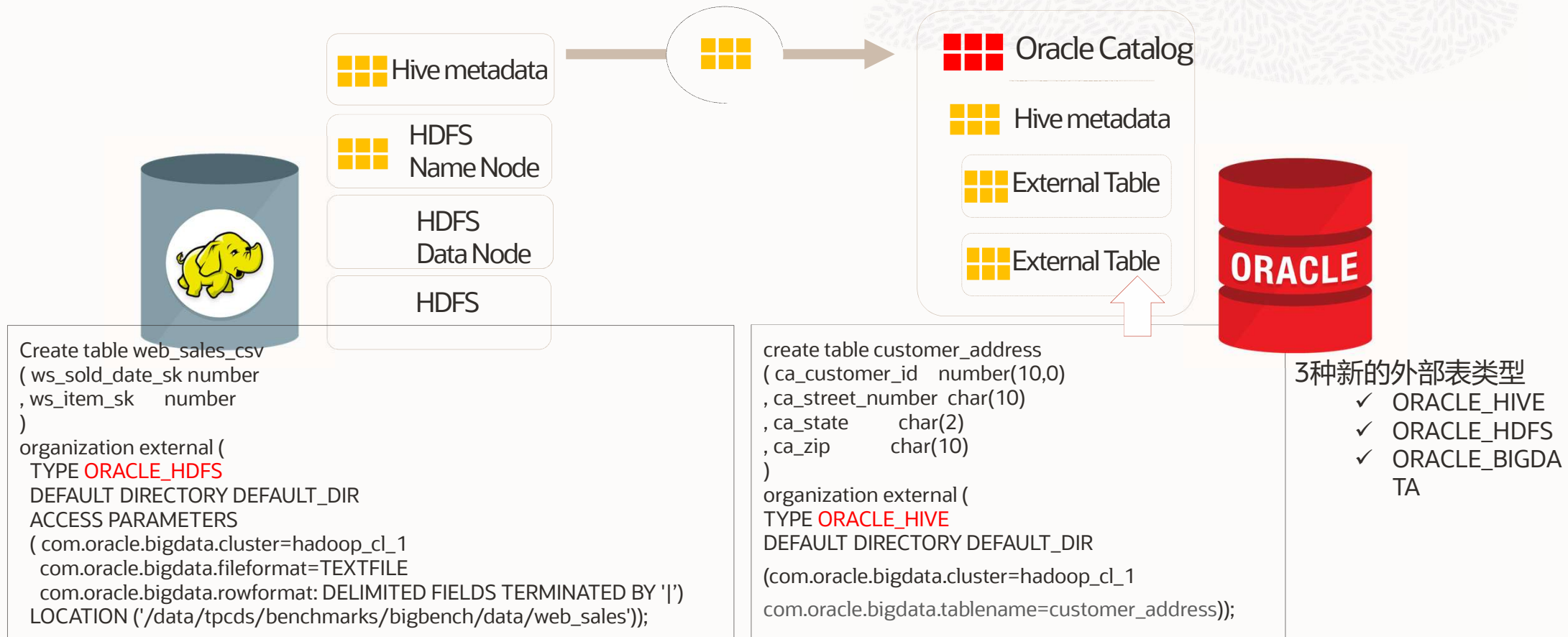
外部表需要访问：

- Hadoop cluster
- Remote Hive database/table

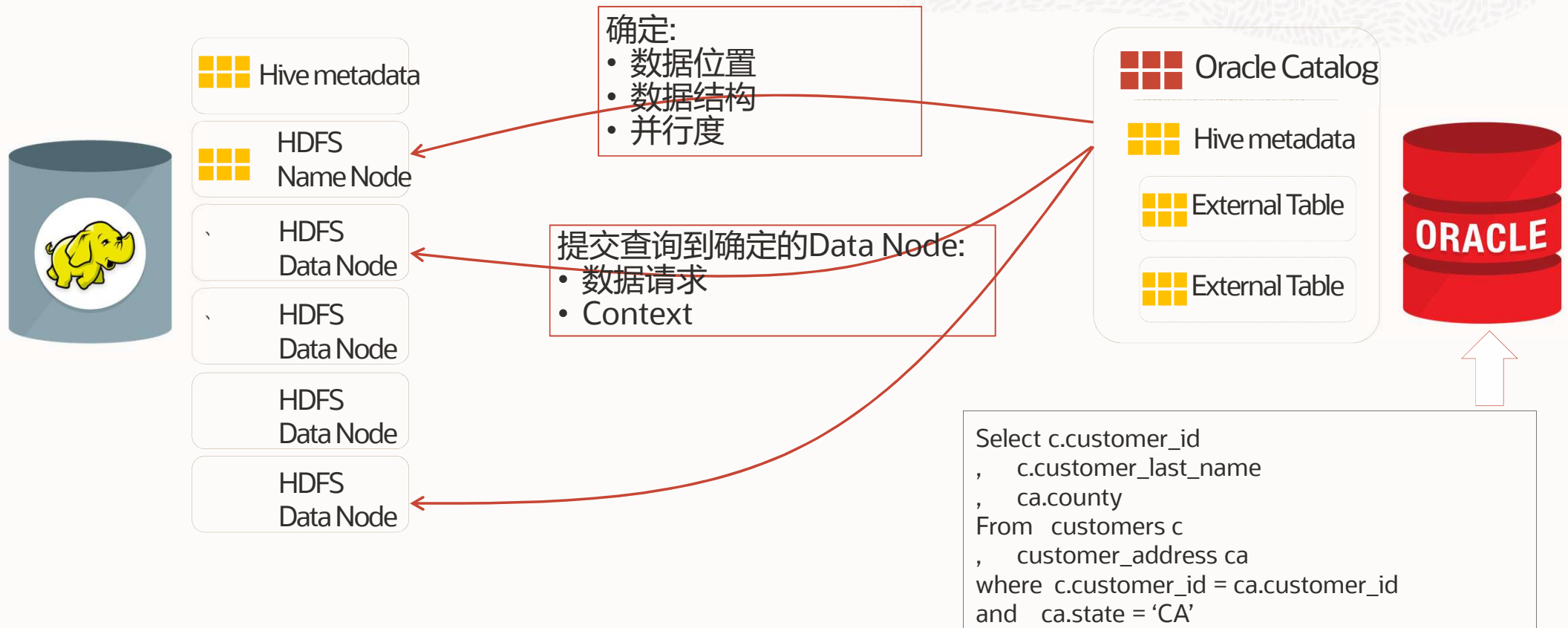
自动生成Table结构

- SQL Developer & DBMS_HADOOP package

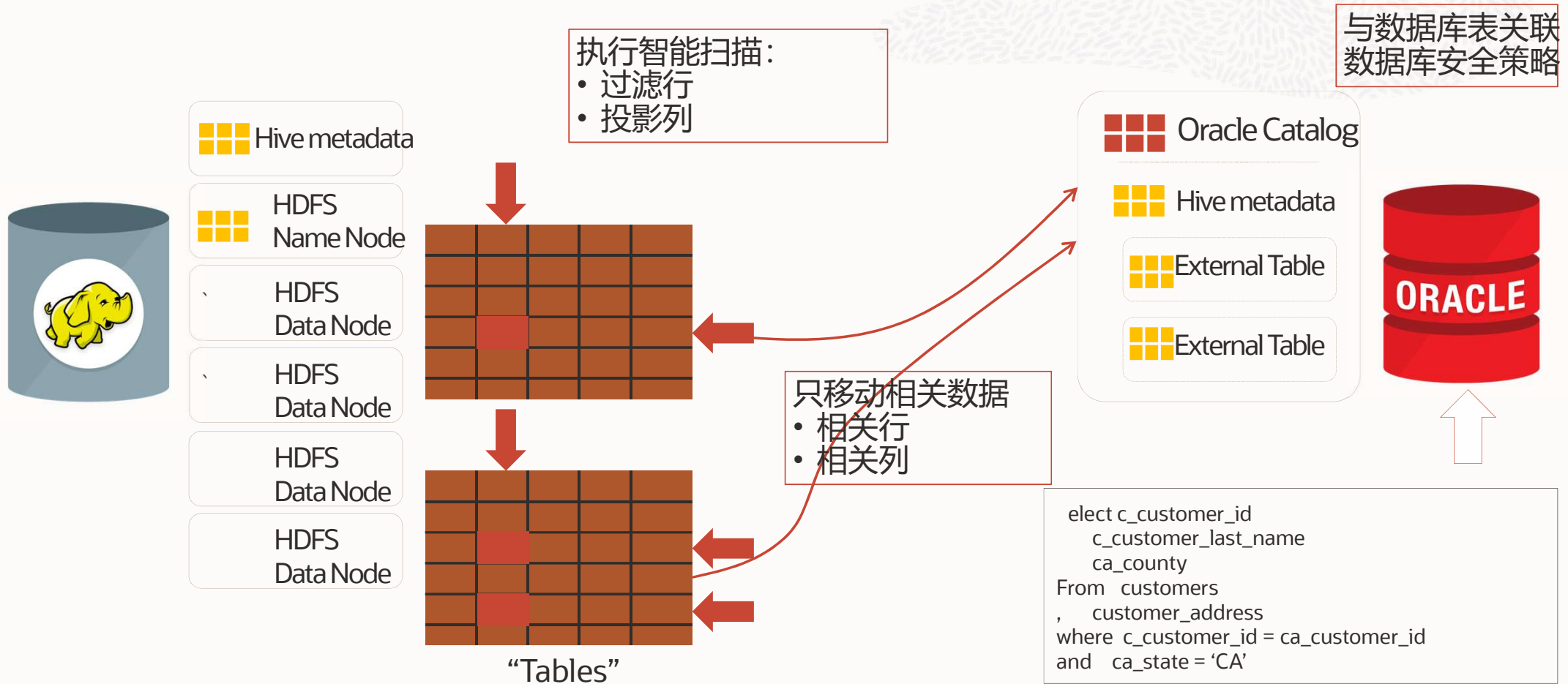
Big Data SQL执行步骤 1: 同步Metadata, 获得DDL



Big Data SQL执行步骤 2：在Hadoop中执行查询



Big Data SQL执行步骤 3：在Hadoop中执行智能扫描



Big Data SQL关键特性

(Hive) 分区裁剪

- 读取Hive元数据并做分区裁剪
- 目标：减少IO

存储索引

- 维护元数据中标记数据块跳过IO
- 目标：减少IO

智能扫描

- 计算下沉到Data node节点，确保返回符合条件的行和列到数据库中
- 目标：减少数据移动

布隆过滤

- 生成布隆过滤并下推到Data node节点
- 目标：Join优化

谓词下沉

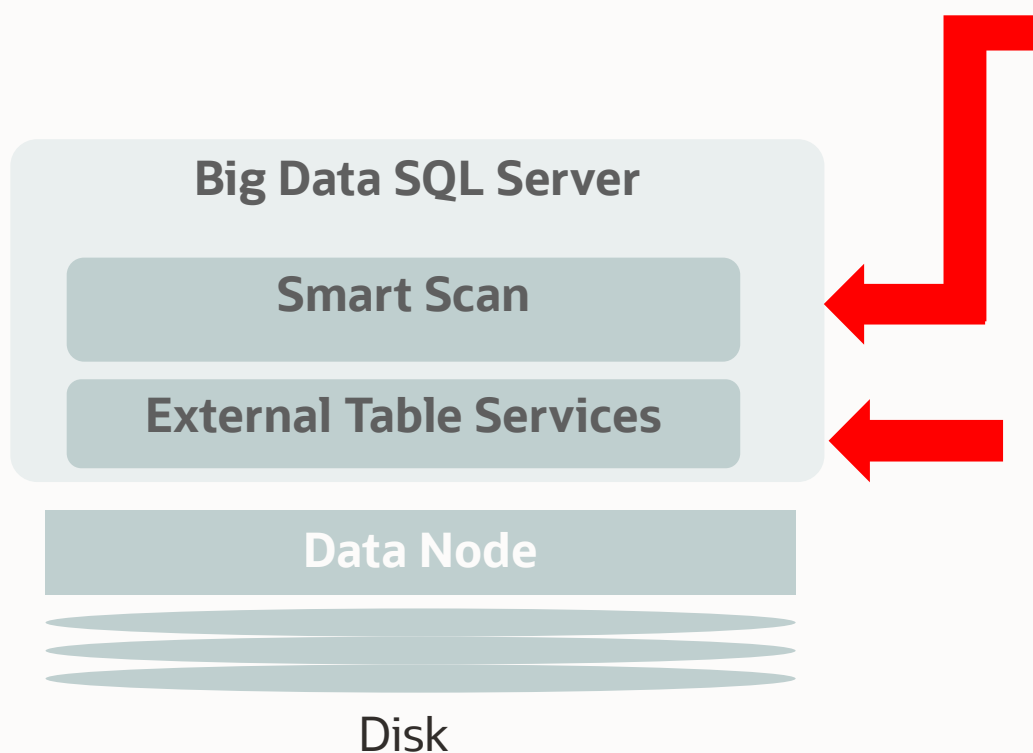
- 为Parquet、ORC数据进行行和列的裁剪
- 目标：减少IO

安全

- 应用数据库的安全策略到Hadoop、Nosql中
- 目标：增强数据安全



BDS特性之一： Hadoop智能扫描(Smart Scan for Hadoop)



“上层Oracle数据库”

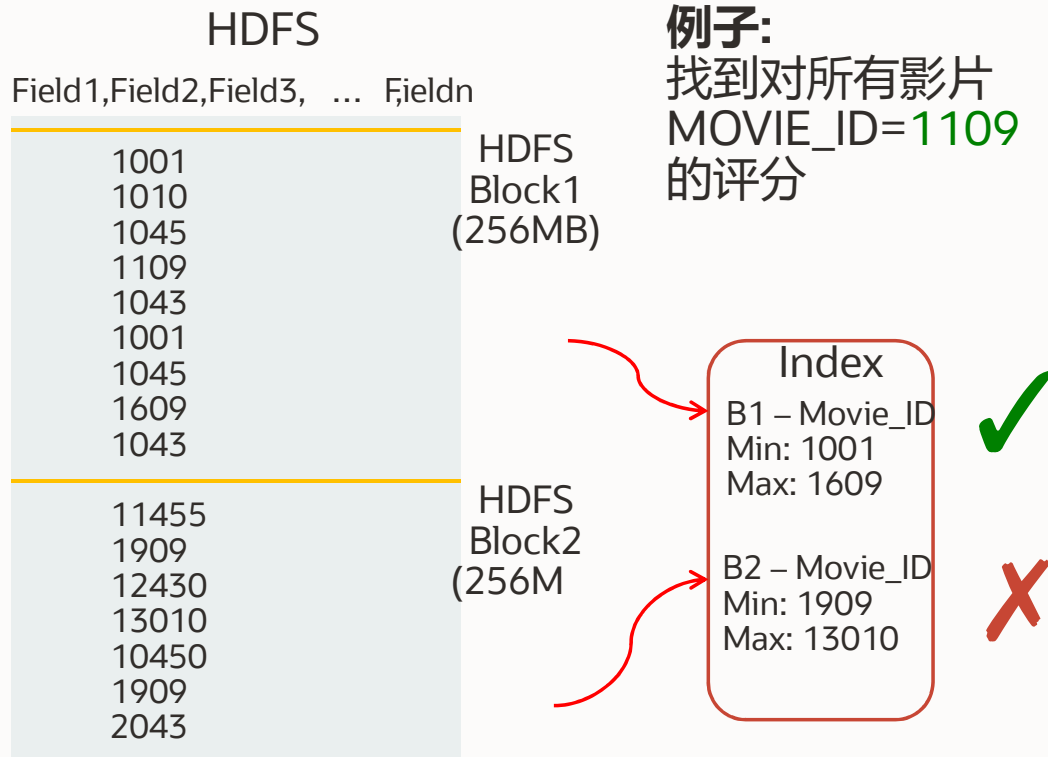
- 条件过滤
- 列投影
- 解析半结构化数据

“下层Hadoop存储引擎”

- 让计算靠近数据
- Schema-on-read的hadoop数据访问模式
- 转换成Oracle数据流



BDS特性之二：存储索引(Storage Index)



- 存储索引通过减少HDFS IO的扫描提升查询性能
- 存储索引记录每个HDFS block的最大、最小值(最大32个字段)

BDS特性之三：条件下沉



```
SELECT ticker, AVG(price)
OVER (
  PARTITION BY ticker
  ORDER BY mnth
) as avg_price
FROM stock_prices
WHERE mnth < :x
AND mnth > :y
AND stx_xchange = :z
ORDER BY mnth
```

```
SELECT ticker, price, mnth
WHERE mnth < :x
AND mnth > :y
AND stx_xchange = :z
```

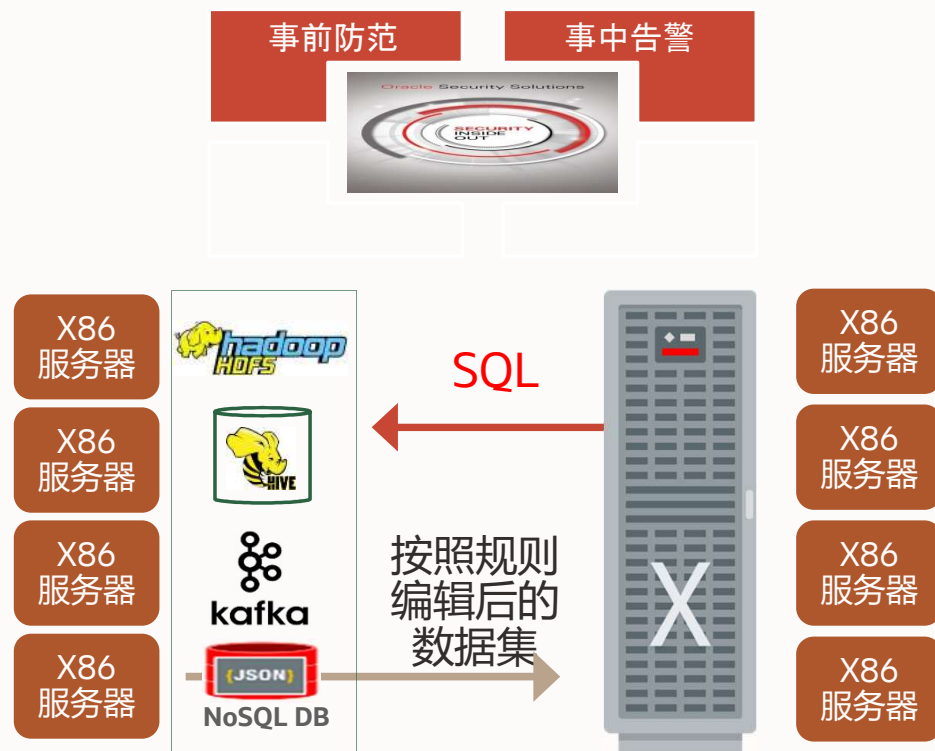
```
WHERE mnth < :x
AND mnth > :y
AND stx_xchange = :z
```

 Parquet

 Parquet

- 将条件推送到存储层评估执行(包括绑定变量)
- 支持下列过滤方式
 - Parquet/ORC过滤(自描述格式)
 - Hive分区裁剪
 - Hbase子扫描
- 余下记录仍可实现SmartScan

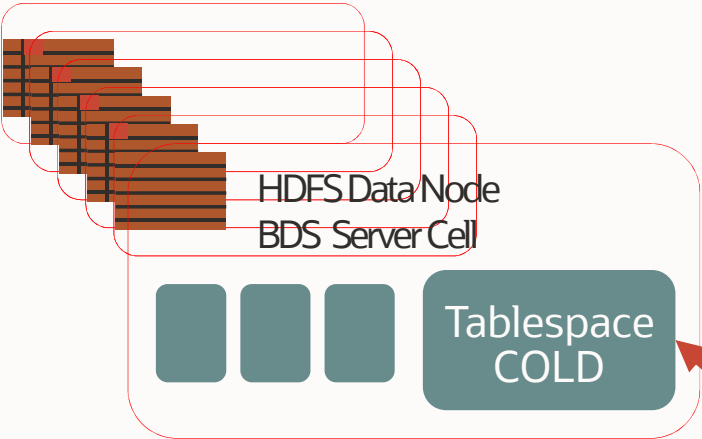
BDS特性之四：安全



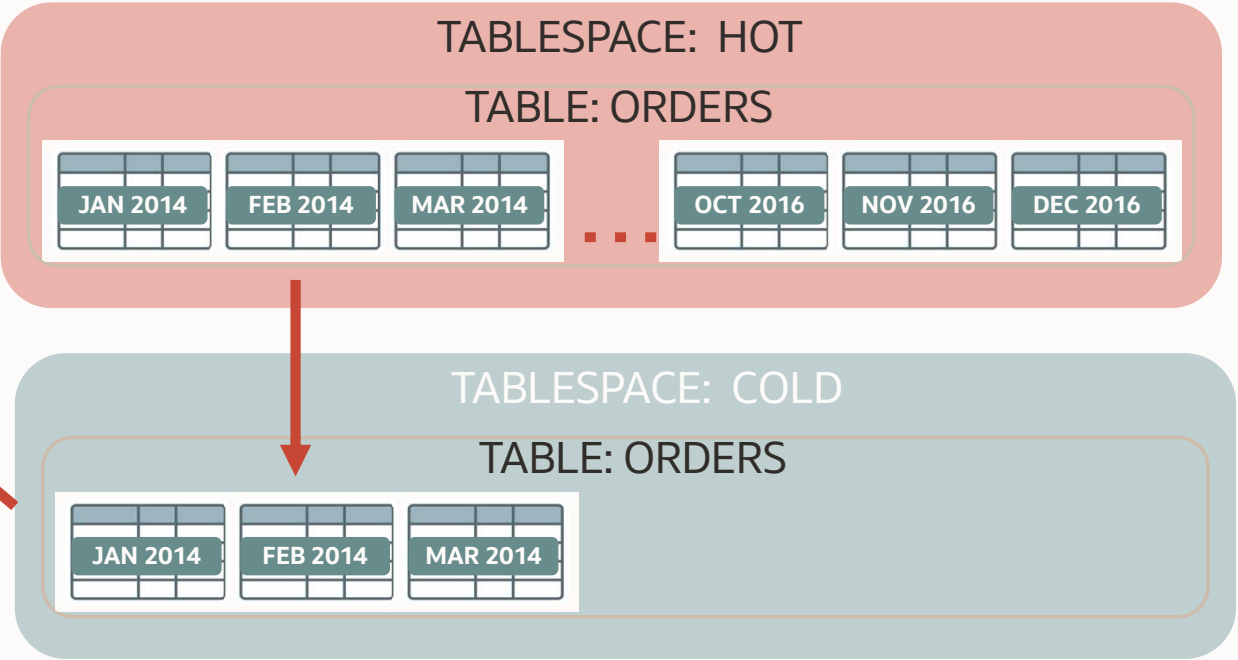
- Oracle数据库提供全面数据安全保护
- Hadoop平台安全能力
 - 认证
 - 授权
- Big Data SQL在Hadoop & NoSQL上应用Oracle数据库的安全管控能力
 - 通过Oracle Audit Vault审计
 - 细粒度的访问控制
 - 数据屏蔽和编辑
 - 虚拟数据库
 - 其他安全能力

BDS特性之五：数据归档

HDFS:



Oracle Database 12c:





Demo

议程

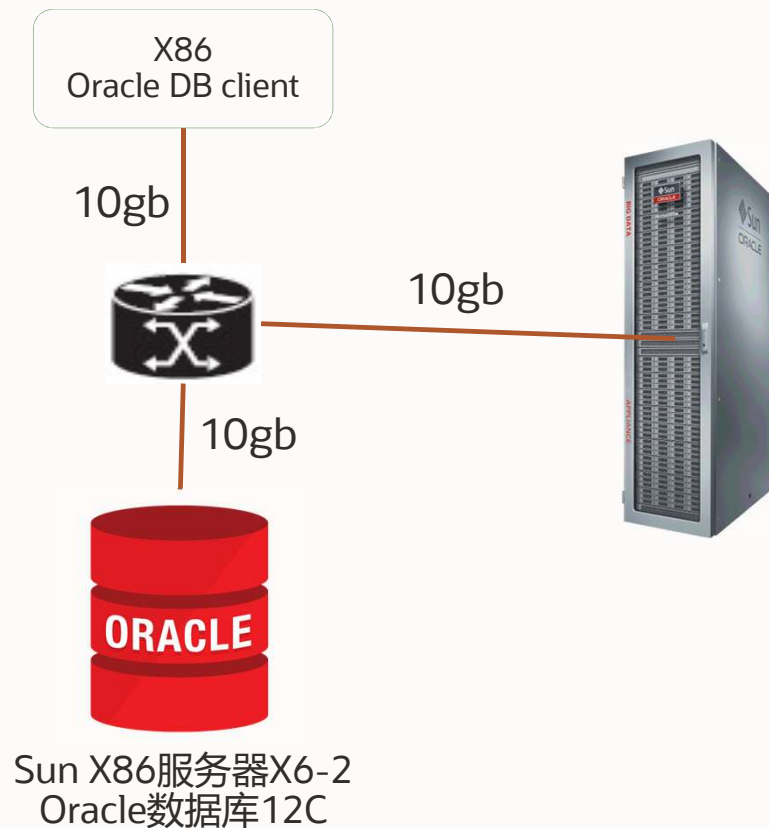
1. 什么是Big Data SQL
2. Big Data SQL架构与特性
3. Big Data SQL案例

某省电网BDS案例：项目背景

- ✓ 某省电网公司构建大数据平台，一期已经基于Cloudera Hadoop搭建了28个节点的环境，每天收集变压器数据，基于Spark进行计算，生成60万条的变压器结果数据，提供给外部进行查询和访问
- ✓ Oracle针对某省电网大数据分析需求，提出了相关产品解决方案，主要是基于Oracle Big Data SQL实现丰富的SQL支持，保证重用已有的技能和工具，考虑到变压器数据量巨大，因此希望通过测试了解和评估变压器计算结果的查询性能，为将来产品的部署提供参考。



某省电网BDS案例：环境部署



相关产品：

- Oracle Big Data Appliance
初始化版本6节点，由于其中某节点部署其他软件，没加入集群，实际使用5个节点
- Oracle Big Data SQL 3.1
- Oracle Database 12C

某省电网BDS案例：测试环境明细

测试环境			
测试环境	软件版本		配置说明
Oracle数据库	Oracle	12.1.0.2.0	服务器型号：ORACLE SERVER X6-2 RAM：256GB CPU：Intel(R) Xeon(R) CPU E5-2699 v4 @ 2.20GHz,2个，一共44个核心/88个线程。
BDA 4.5	BDA	4.5	一共5个Datanode节点。每个节点配置信息如下： 服务器型号：ORACLE SERVER X5-2L RAM：128GB CPU：Intel(R) Xeon(R) CPU E5-2699 v3 @ 2.30GHz，2个，一共36个核心/72个线程
	CDH	5.5.1	
	Hadoop	2.6.0+cdh5.5.1+924	
	Hive	1.1.0+cdh5.5.1+329	
	Impala	2.3.0+cdh5.5.1+0	
	Spark	1.5.0+cdh5.5.1+94	
BDS	BDS	3.1	BDS在BDA和Oracle数据库中都需要安装。



某省电网BDS案例：测试数据

测试数据				
数据表	记录数	存储方式	存储大小	测试数据说明
t_line_info	5万条	Oracle	13 MB	配网线路基础信息表（台账），数据规模：全省50000条记录
		Hive TEXTFILE	10.5 MB	
t_line_day	9千135百万	Oracle	14.4 GB	配网线路日统计信息表（日最大负荷、日最大负载率等等），此表按年分区，一共5年数据。
		Hive TEXTFILE	22.1 GB	
		Hive ORC Zlib	5.9 GB	
		Hive ORC SNAPPY	8.2 GB	
		Hive ORC NONE	14.6 GB	
		Hive PARUQET	7.4 GB	
t_tran_info	10万条	Oracle	184 MB	配电变压器基础信息台（台账），数据规模：全省60万条记录。
		Hive TEXTFILE	183.9 MB	
t_tran_day	11亿条	Oracle	1.25 TB	配电变压器日统计信息表（日最大负荷、日最大负载率等等），此表按年分区，一共5年数据。
		Hive TEXTFILE	1.5 TB	
		Hive ORC Zlib	402.0 GB	
		Hive ORC SNAPPY	631.5 GB	
		Hive ORC NONE	1.3 TB	
		Hive PARUQET	444.6 GB	

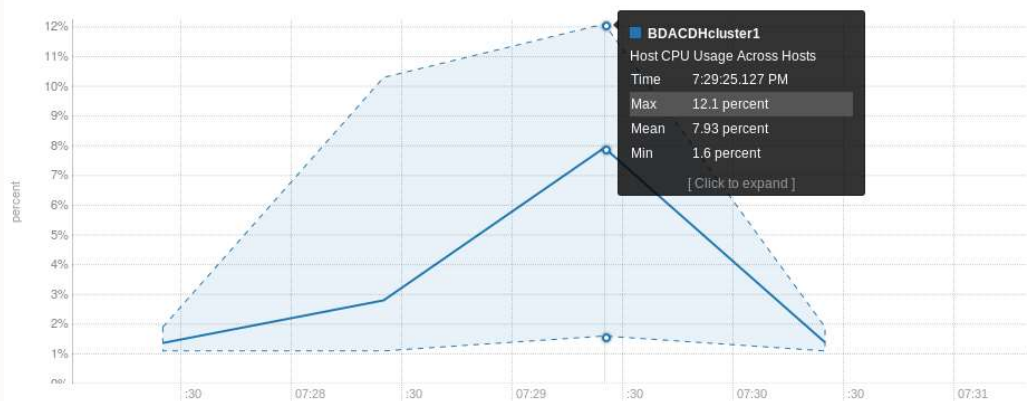
某省电网BDS案例：测试结果

测试场景	并发数	BDS			Hive on MR			Hive on Spark			备注
		TEXTFILE	ORC	PARQUET	TEXTFILE	ORC	PARQUET	TEXTFILE	ORC	PARQUET	
场景1：查询全省某一天（如：2017-01-09）重过载线路明细（TOP100），按“最大负载率”降序排列。	1	6.46	32.56	12.47	37.15	34.06	38.01	30.3	26.57	31.46	联合查询：t_line_day_text(某一天，5万条)和t_line_info(全量5万条)
	10	7.83	45	30	110	59	80	81	53	71	
	20	12	49	54	185	76	119	108	66	98	
场景2：统计全省各地市某一段时间（如：2017年1月）线路重过载情况。（注意数量要去重，长周期统计采用“存在即算”的原则）	1	6.89	29	10.11	61	33	39.3	39.8	36.5	34.91	联合查询：t_line_day_text(某一月，150万条)和t_line_info(全量5万条)
	10	8	38	22	480	149	184	86	57	81	
	20	18	43	42	1219	328	522	121	68	102	
场景3：统计并分析全省各地市某一段时间（如2017年1月）段线路重过载情况。	1	8.99	31.79	12.35	57.85	34.32	38.29	39.54	33.64	38.76	联合查询：t_line_day_text(某两年中某两月，300万条)和t_line_info(全量5万条)
	10	10	41	23	600	145	187	87	56	85	
	20	14	46	46	1123	425	381	120	69	106	
场景4：查询全省某一天（如：2017-01-09）重过载配变明细（轻载不算），按“最大负载率”降序排列。	1	2.83	2.08	1.88	459.56	152.38	147.1	287.73	125.19	68.38	单表查询t_tran_day_text(某一天，60万条)
	10	5.67	10	7	4343	1133	1330	1297	241	198	
	20	10	9	8	9789	2268	2664	2321	449	392	
场景5：统计全省各地市某一段时间（如：2017年1月）配变重过载情况。（注意数量要去重，长周期统计采用“存在即算”的原则）	1	7.13	8.17	7.16	871.62	314.7	319	292.6	129.48	55.53	单表查询t_tran_day_text(某月，1800万条)
	10	14.53	14	14	9232	2693	2840	1267	198	166	
	20	22.59	22	23	32683	9436	10795	2408	483	284	
场景6：统计并分析全省各地市某一段时间（如2017年1月）段线路重过载情况。	1	18.42	60	60.1	917.42	318.57	302	297.3	121.35	61.88	单表查询t_tran_day_text(某两年中某两月，3600万条)
	10	23	59	62	9698	2799	2889	1383	289	189	
	20	32.77	62	65	36002	8954	9881	2030	530	346	
场景7：统计全省各地市某一段时间（如：2017年1月）线路/配变重过载情况。（注意数量要去重，长周期统计采用“存在即算”的原则） 署其	1	13.29	19.76	17.44	1009	354	361	296	133.76	63.25	联合查询t_tran_day_text(某两年中某两月，3600万条),t_line_day_text(某一月，150万条)和t_line_info(全量5万条)
	10	24.9	48	35	10195	3086	3034	1614	357	182	
	20	31.48	63	51	33423	9889	10882	2868	522	319	



某省电网BDS案例：测试过程资源使用情况

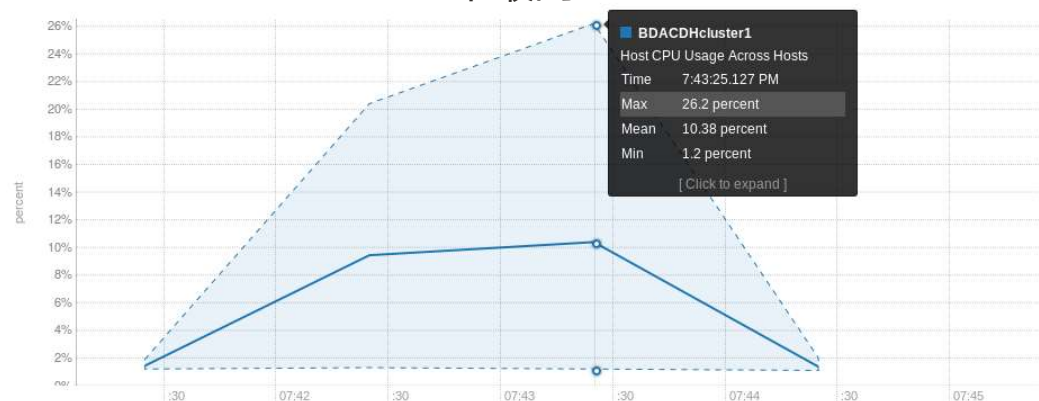
场景3并发用户20：
BDA CPU使用率平均8%，峰值12.1%；DB CPU使用率最高10%



```
[oracle@x62bj200 ~]$ mpstat 2  
Linux 4.1.12-37.4.1.el6uek.x86_64 (x62bj200.oscbj.com) 02/24/2017 _x86_64_ (88 CPU)
```

	CPU	%usr	%nice	%sys	%iowait	%irq	%soft	%steal	%guest	%idle
05:04:27 PM										
05:04:29 PM	all	1.76	0.00	0.36	0.05	0.00	0.07	0.00	0.00	97.76
05:04:31 PM	all	7.94	0.00	1.13	0.04	0.00	0.09	0.00	0.00	90.81
05:04:33 PM	all	1.93	0.00	0.44	0.04	0.00	0.08	0.00	0.00	97.51
05:04:35 PM	all	3.07	0.00	0.78	0.05	0.01	0.14	0.00	0.00	95.96
05:04:37 PM	all	4.62	0.00	1.03	0.05	0.00	0.19	0.00	0.00	94.12
05:04:39 PM	all	9.36	0.00	2.31	0.04	0.00	0.49	0.00	0.00	87.80
05:04:41 PM	all	10.28	0.00	2.46	0.04	0.00	0.53	0.00	0.00	86.70
05:04:43 PM	all	6.35	0.00	1.46	0.05	0.00	0.23	0.00	0.00	91.91
05:04:45 PM	all	2.24	0.00	0.60	0.04	0.00	0.09	0.00	0.00	97.02
05:04:47 PM	all	1.77	0.00	0.35	0.05	0.00	0.05	0.00	0.00	97.78
05:04:49 PM	all	1.72	0.00	0.39	0.04	0.00	0.07	0.00	0.00	97.78
05:04:51 PM	all	1.74	0.00	0.39	0.05	0.00	0.06	0.00	0.00	97.77
05:04:53 PM	all	1.74	0.00	0.35	0.05	0.00	0.08	0.00	0.00	97.79
05:04:55 PM	all	2.43	0.00	0.45	0.04	0.00	0.06	0.00	0.00	97.02
05:04:57 PM	all	1.92	0.00	0.37	0.05	0.00	0.06	0.00	0.00	97.60
05:04:59 PM	all	1.74	0.00	0.36	0.05	0.00	0.06	0.00	0.00	97.80
05:05:01 PM	all	1.72	0.00	0.37	0.04	0.00	0.07	0.00	0.00	97.80
05:05:03 PM	all	1.74	0.00	0.35	0.05	0.00	0.07	0.00	0.00	97.79

场景7并发用户20：
BDA CPU使用率平均10.4%，峰值26.2%；DB CPU使用率最高15%



```
[oracle@x62bj200 ~]$ mpstat 2  
Linux 4.1.12-37.4.1.el6uek.x86_64 (x62bj200.oscbj.com) 02/24/2017 _x86_64_ (88 CPU)
```

05:01:47 PM	CPU	%usr	%nice	%sys	%iowait	%irq	%soft	%steal	%guest	%idle
05:01:49 PM	all	1.51	0.00	0.24	0.05	0.00	0.03	0.00	0.00	98.17
05:01:51 PM	all	1.53	0.00	0.30	0.04	0.00	0.03	0.00	0.00	98.10
05:01:53 PM	all	5.58	0.00	0.76	0.05	0.00	0.03	0.00	0.00	93.57
05:01:55 PM	all	9.62	0.00	0.83	0.05	0.00	0.05	0.00	0.00	89.45
05:01:57 PM	all	5.53	0.00	1.01	0.04	0.00	0.04	0.00	0.00	93.38
05:01:59 PM	all	4.16	0.00	1.11	0.04	0.00	0.05	0.00	0.00	94.64
05:02:01 PM	all	6.29	0.00	1.13	0.05	0.00	0.05	0.00	0.00	92.47
05:02:03 PM	all	1.62	0.00	0.35	0.04	0.00	0.03	0.00	0.00	97.97
05:02:05 PM	all	2.64	0.00	0.70	0.05	0.00	0.07	0.00	0.00	96.54
05:02:07 PM	all	6.08	0.00	1.71	0.04	0.01	0.35	0.00	0.00	91.81
05:02:09 PM	all	5.31	0.00	1.39	0.05	0.00	0.23	0.00	0.00	93.02
05:02:11 PM	all	6.68	0.00	2.54	0.04	0.00	0.30	0.00	0.00	90.44
05:02:13 PM	all	5.77	0.00	1.89	0.04	0.00	0.21	0.00	0.00	92.09
05:02:15 PM	all	4.40	0.00	1.40	0.05	0.00	0.17	0.00	0.00	93.99
05:02:17 PM	all	7.97	0.00	2.31	0.04	0.00	0.41	0.00	0.00	89.27
05:02:19 PM	all	10.98	0.00	3.26	0.03	0.00	0.63	0.00	0.00	85.06
05:02:21 PM	all	14.39	0.00	4.27	0.04	0.00	0.91	0.00	0.00	80.39
05:02:23 PM	all	15.50	0.00	4.60	0.04	0.00	1.01	0.00	0.00	78.85
05:02:25 PM	all	10.39	0.00	2.83	0.04	0.00	0.64	0.00	0.00	86.10
05:02:27 PM	all	5.29	0.00	1.37	0.04	0.00	0.21	0.00	0.00	93.09
05:02:29 PM	all	2.01	0.00	0.45	0.05	0.00	0.07	0.00	0.00	97.43
05:02:31 PM	all	1.81	0.00	0.33	0.05	0.00	0.05	0.00	0.00	97.77
05:02:33 PM	all	1.73	0.00	0.33	0.04	0.00	0.03	0.00	0.00	97.86
05:02:35 PM	all	1.76	0.00	0.32	0.04	0.00	0.05	0.00	0.00	97.82
05:02:37 PM	all	1.73	0.00	0.33	0.05	0.00	0.04	0.00	0.00	97.86
05:02:39 PM	all	1.70	0.00	0.32	0.05	0.00	0.05	0.00	0.00	97.89
05:02:41 PM	all	1.81	0.00	0.31	0.04	0.00	0.05	0.00	0.00	97.79



某省电网BDS案例：测试结论

- ✓ Big Data SQL无论是在单用户还是并发用户，脚本完全支持Oracle SQL语法，不需要进行
- ✓ 在20用户并发情况下，BDA CPU和数据库服务器CPU的使用率都比较小，能够胜任目前工作和未来扩展需求，
- ✓ 随着并发数增加，Big data SQL略有延迟，表现了良好的可扩展性
- ✓ 当并发用户数到了120（相当于1200个用户在访问）时，响应时间增加，但不会出现问题，考虑到CPU等资源使用率饱和导致下降
- ✓ 与其他SQL on Hadoop相比，Oracle Big Data SQL无论在SQL语法支持，并发数支持，还是扩展性方面，都远远优于其他产品

捷克Home Credit: 项目背景



客户：Home Credit是一家贷款提供商，在全球很多国家（白俄罗斯、捷克共和国、中国、印度、印度尼西亚、哈萨克斯坦、菲律宾、俄罗斯联邦、斯洛伐克共和国、美国、越南）提供服务。借助Big Data SQL构建湖仓一体平台。

测试场景：

✓ 场景1：

- DB端数据量：1570万行应用程序记录和1.5亿行客户记录
- Hadoop数据量：6300万行应用程序记录和6亿行客户记录
- 成功标准：20分钟返回结果

✓ 场景2：

- DB端数据量：2.63亿行应用程序记录和25.71亿行客户记录
- Hadoop数据量：10.25亿行应用程序记录和100亿行客户记录
- 成功标准：40分钟返回结果

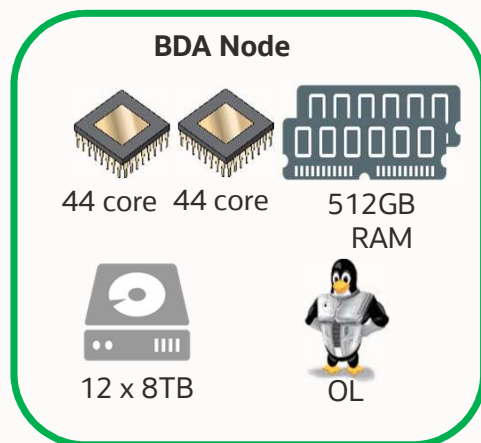


捷克Home Credit: 测试环境

HOME
CREDIT

BDA X6-2 Starter Rack

6 Nodes, each with:



cloudera
CDH 5.11

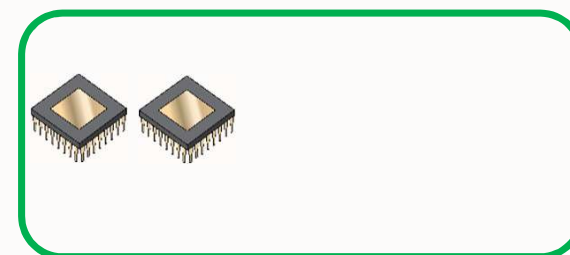
Infiniband



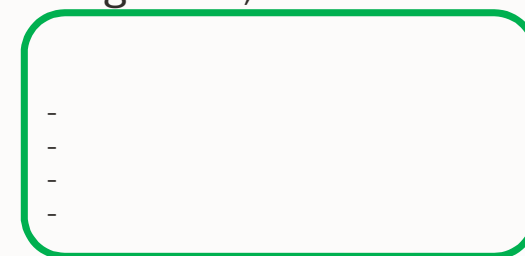
12^c

Exadata X6-2 Eighth Rack

2 database nodes, each with:



3 storage cells, each with:



Big Data SQL 3



捷克Home Credit：测试场景1—数据分布

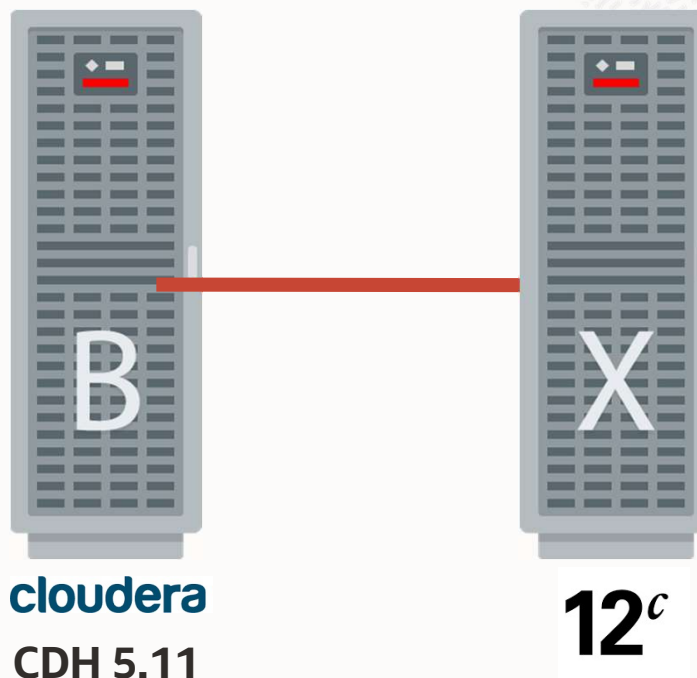
HOME
CREDIT

Hadoop中表 (BDS)

- APPLICATION_EXT_1
 - 63,145,296 Rows
 - Textfile/134.6 GB
- CLIENT_EXT_1
 - 604,114,320 Rows
 - Textfile/1370 GB

Hadoop中表 (Impala)

- APPLICATION_TEST_1
 - 15,786,324 Rows
 - Parquet Snappy/5.2 GB
- CLIENT_TEST_1
 - 151,028,580 Rows
 - Parquet Snappy/4.7 GB
- APPLICATION_EXT_1
 - 63,145,296 Rows
 - Parquet Snappy/20.8 GB
- CLIENT_EXT_1
 - 604,114,320 Rows
 - Parquet Snappy/218.8 GB



Exadata中表

- APPLICATION_TEST_1
 - 15,786,324 Rows
 - 29.6 GB
- CLIENT_TEST_1
 - 151,028,580 Rows
 - 289.22 GB

捷克Home Credit：测试场景1—单个SQL测试



```
select /*+ PARALLEL(88) */ * from (
select /*+ PARALLEL(88) */
t1.NAME_ADDR_CURRENT_REGION NAME_ADDR_CURRENT_REGION_CL
,t1.NAME_ADDR_CURRENT_DISTRICT NAME_ADDR_CURRENT_DISTRICT_CL
,sum(case t1.FLAG_APPROVED when 1 then 1 Else null end) as count_FLAG_APPROVED_CL
,sum(case t1.FLAG_CANCELLED when 1 then 1 Else null end) as count_FLAG_CANCELLED_CL
,sum(case t1.FLAG_REJECTED when 1 then 1 Else null end) as count_FLAG_REJECTED_CL
,round(VARIANCE(t1.num_CLIENT_AGE),2) var_age_CL
,round(AVG(t1.amt_income_main),2) amt_income_main_CL
,median(t1.RATE_INTEREST) med_RATE_INT_CL
,t2.NAME_ADDR_CURRENT_REGION NAME_ADDR_CURRENT_REGION_AP
,t2.NAME_ADDR_CURRENT_DISTRICT NAME_ADDR_CURRENT_DISTRICT_AP
,sum(case t2.FLAG_APPROVED when 1 then 1 Else null end) as count_FLAG_APPROVED_AP
,sum(case t2.FLAG_CANCELLED when 1 then 1 Else null end) as count_FLAG_CANCELLED_AP
,sum(case t2.FLAG_REJECTED when 1 then 1 Else null end) as count_FLAG_REJECTED_AP
,round(VARIANCE(t2.num_CLIENT_AGE),2) var_age_AP
,round(AVG(t2.amt_income_main),2) amt_income_main_AP
,median(t2.RATE_INTEREST) med_RATE_INT_AP
,t3.NAME_ADDR_CURRENT_REGION NAME_ADDR_CURRENT_REGION_EC
,t3.NAME_ADDR_CURRENT_DISTRICT NAME_ADDR_CURRENT_DISTRICT_EX
,sum(case t3.FLAG_APPROVED when 1 then 1 Else null end) as count_FLAG_APPROVED_EX
,sum(case t3.FLAG_CANCELLED when 1 then 1 Else null end) as count_FLAG_CANCELLED_EX
,sum(case t3.FLAG_REJECTED when 1 then 1 Else null end) as count_FLAG_REJECTED_EX
,round(VARIANCE(t3.num_CLIENT_AGE),2) var_age_EX
,round(avg(t3.amt_income_main),2) amt_income_main_EX
,median(t3.RATE_INTEREST) med_RATE_INT_EX
,t4.NAME_ADDR_CURRENT_REGION NAME_ADDR_CURRENT_REGION_EA
,t4.NAME_ADDR_CURRENT_DISTRICT NAME_ADDR_CURRENT_DISTRICT_EA
,sum(case t4.FLAG_APPROVED when 1 then 1 Else null end) as count_FLAG_APPROVED_EA
,sum(case t4.FLAG_CANCELLED when 1 then 1 Else null end) as count_FLAG_CANCELLED_EA
,sum(case t4.FLAG_REJECTED when 1 then 1 Else null end) as count_FLAG_REJECTED_EA
,round(VARIANCE(t4.num_CLIENT_AGE),2) var_age_EA
,round(avg(t4.amt_income_main),2) amt_income_main_EA
,median(t4.RATE_INTEREST) med_RATE_INT_EA
FROM CLIENT_TEST_1 t1
join APPLICATION_TEST_1 t2
on t2.SKP_APPLICATION = t1.SKP_APPLICATION
join H_CLIENT_EXT_1 t3
on t3.SNAPSHOT_DATE = t1.SNAPSHOT_DATE and t3.SKP_APPLICATION = t1.SKP_APPLICATION
join H_APPLICATION_EXT_1 t4
on t4.SKP_APPLICATION = t1.SKP_APPLICATION
WHERE t1.SNAPSHOT_DATE = 1
group by t1.NAME_ADDR_CURRENT_REGION, t1.NAME_ADDR_CURRENT_DISTRICT
,t2.NAME_ADDR_CURRENT_REGION, t2.NAME_ADDR_CURRENT_DISTRICT
,t3.NAME_ADDR_CURRENT_REGION, t3.NAME_ADDR_CURRENT_DISTRICT
,t4.NAME_ADDR_CURRENT_REGION, t4.NAME_ADDR_CURRENT_DISTRICT
)t5 where rownum <= 20;
```

对比	耗时 (秒)
Big Data SQL	18
Impala	59

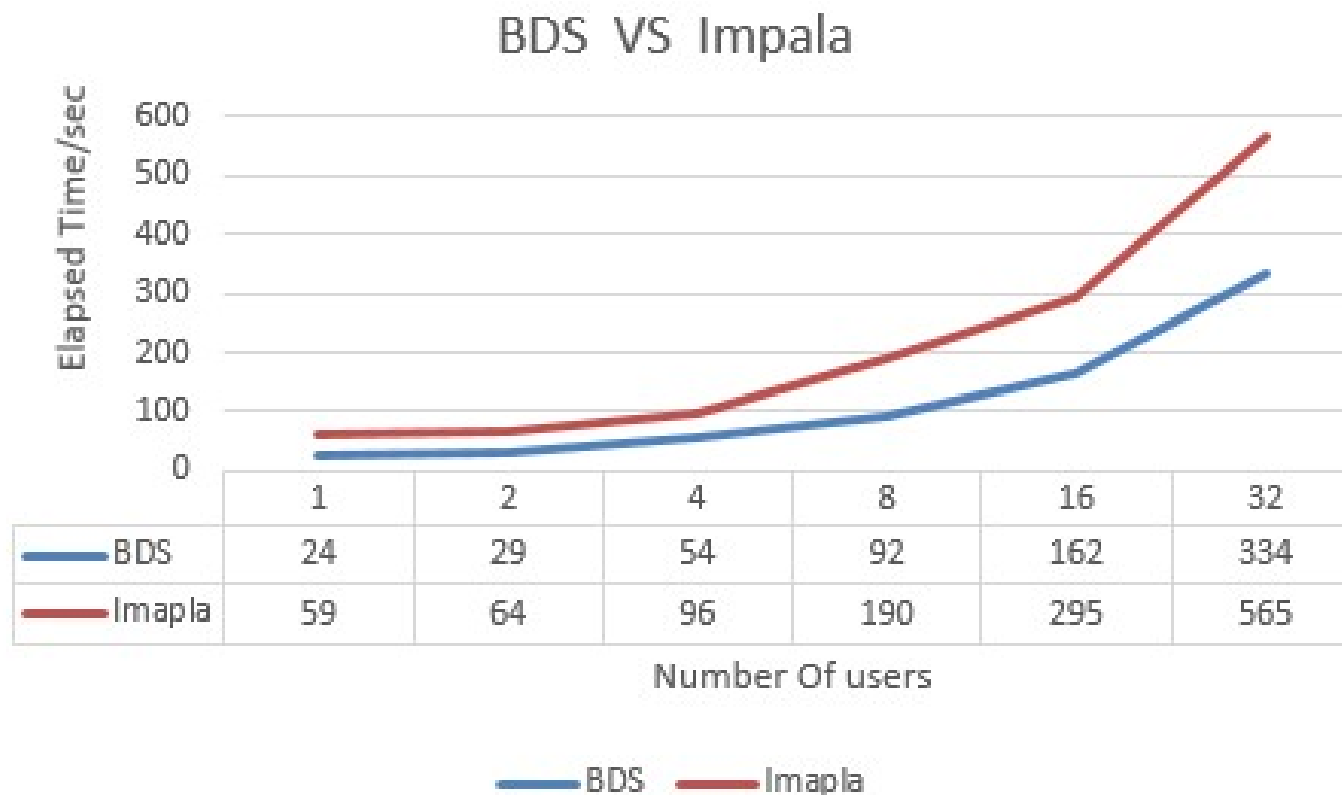
Big Data SQL测试统计信息	
统计指标	统计值 (GB)
cell XT granule bytes requested for predicate offload 显示在HDFS中有多少字节(分区裁剪之后)	1479
cell interconnect bytes returned by XT smart scan 显示返回给Database的字节数	8
cell XT granule IO bytes saved by storage index 显示被存储索引过滤掉的字节数	1263

注：BDS 使用88个并行度



捷克Home Credit: 测试场景1—并行度

HOME
CREDIT



注: BDS使用16个并行度

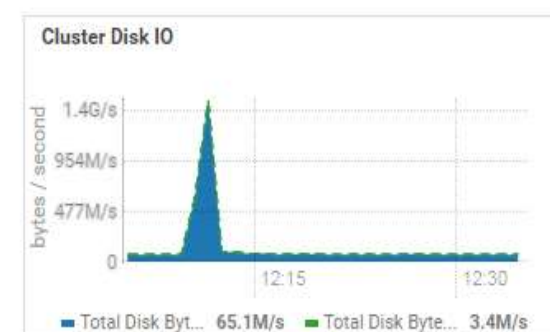
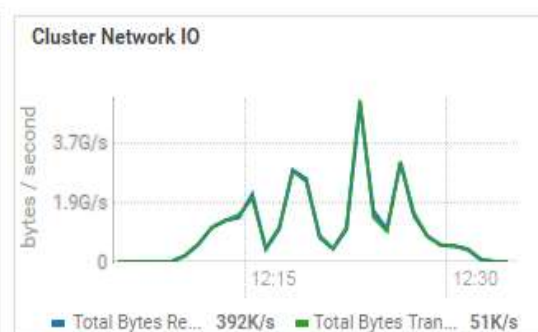
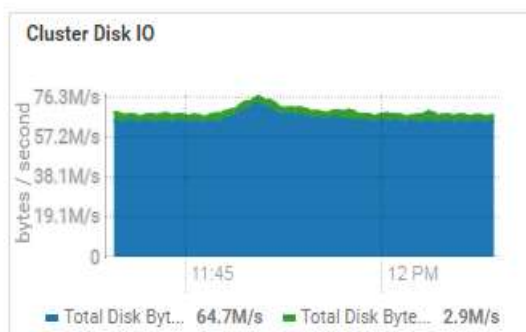
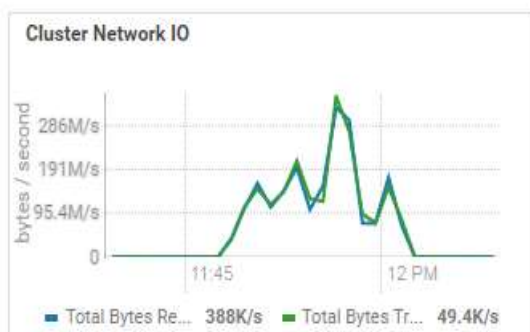
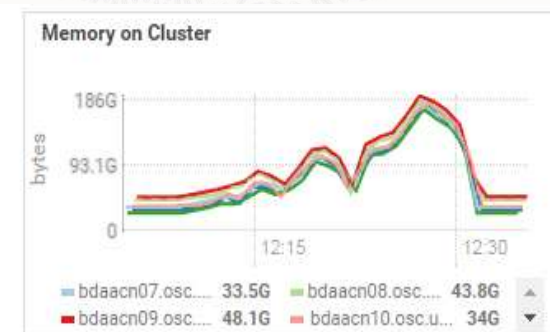
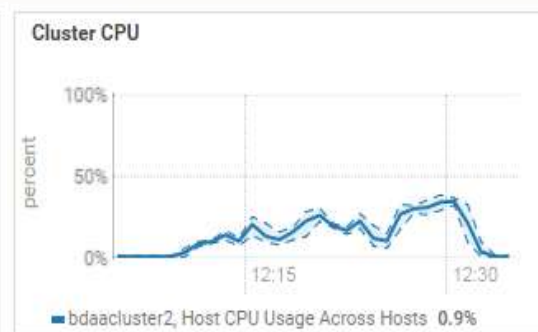
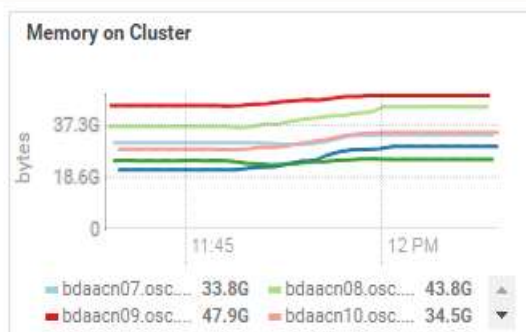
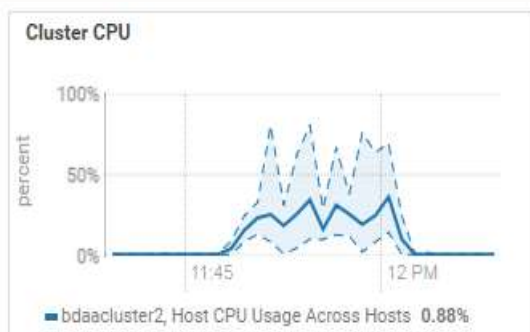


捷克Home Credit: 测试场景1——并行

HOME
CREDIT

BDS 1个到32个并发数

Impala1个到32个并发



BDS占用的CPU资源更多，而内存更少

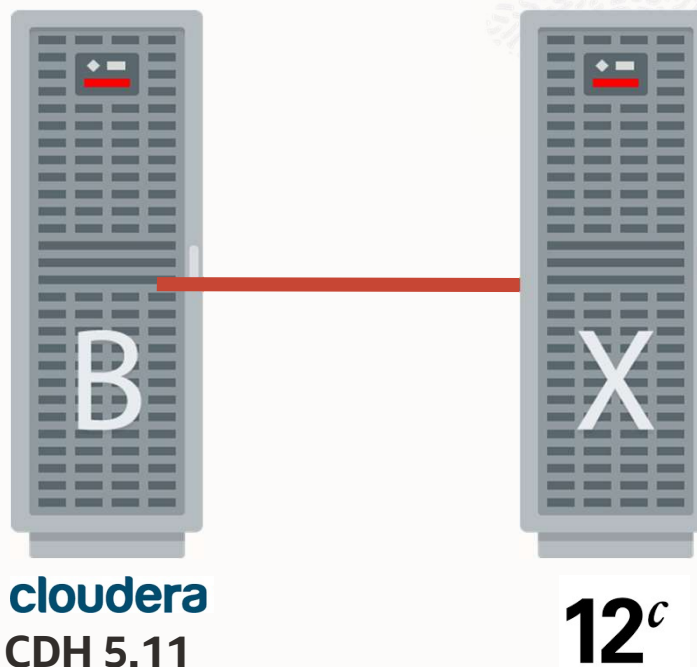


捷克Home Credit: 测试场景2—数据分布

HOME
CREDIT

在Hadoop中 (Impala)

- APPLICATION_TEST_2
 - 263,105,400 Rows
 - Parquet Snappy/86.6 GB
- CLIENT_TEST_2
 - 2,517,143,000 Rows
 - Partitioned by SNAPSHOT_DATE,20 partitions
 - Parquet Snappy/885.3 GB
- APPLICATION_EXT_2
 - 1,052,421,600 Rows
 - Parquet Snappy/347.1 GB
- CLIENT_EXT_2
 - 10,068,572,000 Rows
 - Partitioned by SNAPSHOT_DATE,20 partitions
 - Parquet Snappy/3584 GB



在Exadata中 (BDS)

- APPLICATION_TEST_2
 - 263,105,400 Rows
- CLIENT_TEST_2
 - 2,517,143,000 Rows
 - Partitioned by SNAPSHOT_DATE,20 partitions
 - 4823.65 GB

在Hadoop中 (BDS)

- APPLICATION_EXT_2
 - 1,052,421,600 Rows
 - Textfile/2252 GB
- CLIENT_EXT_2
 - 10,068,572,000 Rows
 - Partitioned by SNAPSHOT_DATE,20 partitions
 - Textfile/21321 GB

注: Impala分配了 400GB内存, 每个节点总内存是512GB
DB端, SGA分配了130GB, PGA分配了90GB, 每个DB节点的内存总共是256GB。



捷克Home Credit： 测试场景2—单个SQL测试



```
select /*+ PARALLEL(88) */ * from (
select /*+ PARALLEL(88) */
t1.NAME_ADDR_CURRENT_REGION NAME_ADDR_CURRENT_REGION_CL
,t1.NAME_ADDR_CURRENT_DISTRICT NAME_ADDR_CURRENT_DISTRICT_CL
,sum(case t1.FLAG_APPROVED when 1 then 1 Else null end) as count_FLAG_APPROVED_CL
,sum(case t1.FLAG_CANCELLED when 1 then 1 Else null end) as count_FLAG_CANCELLED_CL
,sum(case t1.FLAG_REJECTED when 1 then 1 Else null end) as count_FLAG_REJECTED_CL
,round(VARIANCE(t1.num_CLIENT_AGE),2) var_age_CL
,round(AVG(t1.amt_income_main),2) amt_income_main_CL
,median(t1.RATE_INTEREST) med_RATE_INT_CL
,t2.NAME_ADDR_CURRENT_REGION NAME_ADDR_CURRENT_REGION_AP
,t2.NAME_ADDR_CURRENT_DISTRICT NAME_ADDR_CURRENT_DISTRICT_AP
,sum(case t2.FLAG_APPROVED when 1 then 1 Else null end) as count_FLAG_APPROVED_AP
,sum(case t2.FLAG_CANCELLED when 1 then 1 Else null end) as count_FLAG_CANCELLED_AP
,sum(case t2.FLAG_REJECTED when 1 then 1 Else null end) as count_FLAG_REJECTED_AP
,round(VARIANCE(t2.num_CLIENT_AGE),2) var_age_AP
,round(AVG(t2.amt_income_main),2) amt_income_main_AP
,median(t2.RATE_INTEREST) med_RATE_INT_AP
,t3.NAME_ADDR_CURRENT_REGION NAME_ADDR_CURRENT_REGION_EC
,t3.NAME_ADDR_CURRENT_DISTRICT NAME_ADDR_CURRENT_DISTRICT_EX
,sum(case t3.FLAG_APPROVED when 1 then 1 Else null end) as count_FLAG_APPROVED_EX
,sum(case t3.FLAG_CANCELLED when 1 then 1 Else null end) as count_FLAG_CANCELLED_EX
,sum(case t3.FLAG_REJECTED when 1 then 1 Else null end) as count_FLAG_REJECTED_EX
,round(VARIANCE(t3.num_CLIENT_AGE),2) var_age_EX
,round(avg(t3.amt_income_main),2) amt_income_main_EX
,median(t3.RATE_INTEREST) med_RATE_INT_EX
,t4.NAME_ADDR_CURRENT_REGION NAME_ADDR_CURRENT_REGION_EA
,t4.NAME_ADDR_CURRENT_DISTRICT NAME_ADDR_CURRENT_DISTRICT_EA
,sum(case t4.FLAG_APPROVED when 1 then 1 Else null end) as count_FLAG_APPROVED_EA
,sum(case t4.FLAG_CANCELLED when 1 then 1 Else null end) as count_FLAG_CANCELLED_EA
,sum(case t4.FLAG_REJECTED when 1 then 1 Else null end) as count_FLAG_REJECTED_EA
,round(VARIANCE(t4.num_CLIENT_AGE),2) var_age_EA
,round(avg(t4.amt_income_main),2) amt_income_main_EA
,median(t4.RATE_INTEREST) med_RATE_INT_EA
FROM CLIENT_TEST_1 t1
join APPLICATION_TEST_1 t2
on t2.SKP_APPLICATION = t1.SKP_APPLICATION
join H_CLIENT_EXT_1 t3
on t3.SNAPSHOT_DATE = t1.SNAPSHOT_DATE and t3.SKP_APPLICATION = t1.SKP_APPLICATION
join H_APPLICATION_EXT_1 t4
on t4.SKP_APPLICATION = t1.SKP_APPLICATION
WHERE t1.SNAPSHOT_DATE = 1
group by t1.NAME_ADDR_CURRENT_REGION, t1.NAME_ADDR_CURRENT_DISTRICT
,t2.NAME_ADDR_CURRENT_REGION, t2.NAME_ADDR_CURRENT_DISTRICT
,t3.NAME_ADDR_CURRENT_REGION, t3.NAME_ADDR_CURRENT_DISTRICT
,t4.NAME_ADDR_CURRENT_REGION, t4.NAME_ADDR_CURRENT_DISTRICT
)t5 where rownum <= 20;
```

对比	耗时 (秒)
Big Data SQL	891
Impala	5818

Big Data SQL测试统计信息	
统计指标	统计值 (GB)
cell XT granule bytes requested for predicate offload 显示在HDFS中有多少字节(分区裁剪之后)	3476
cell interconnect bytes returned by XT smart scan 显示返回给Database的字节数	136
cell XT granule IO bytes saved by storage index 显示被存储索引过滤掉的字节数	0

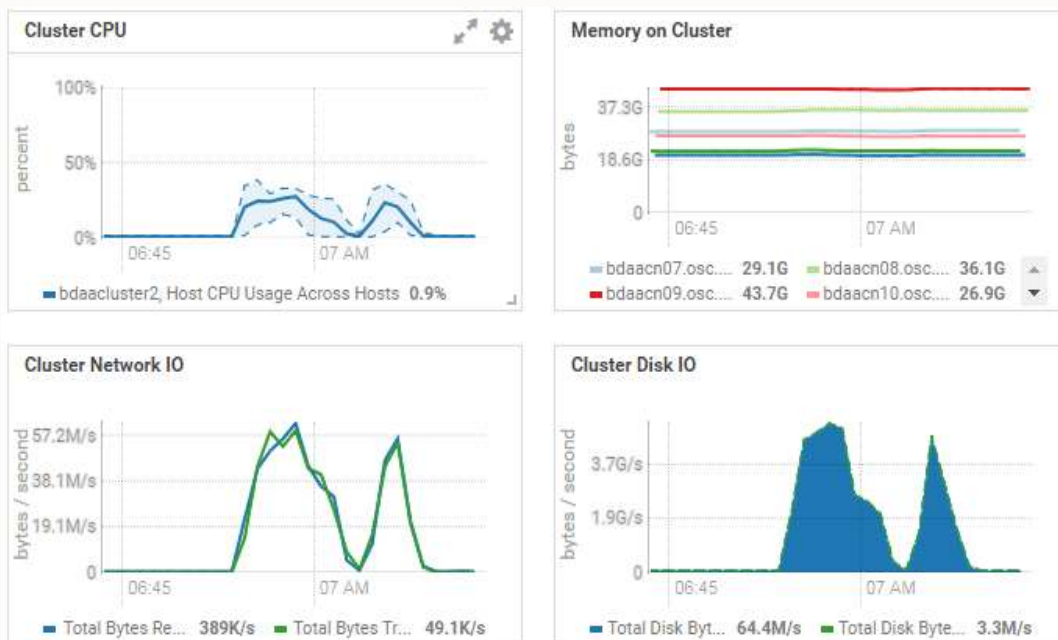
注：BDS使用176个并行度



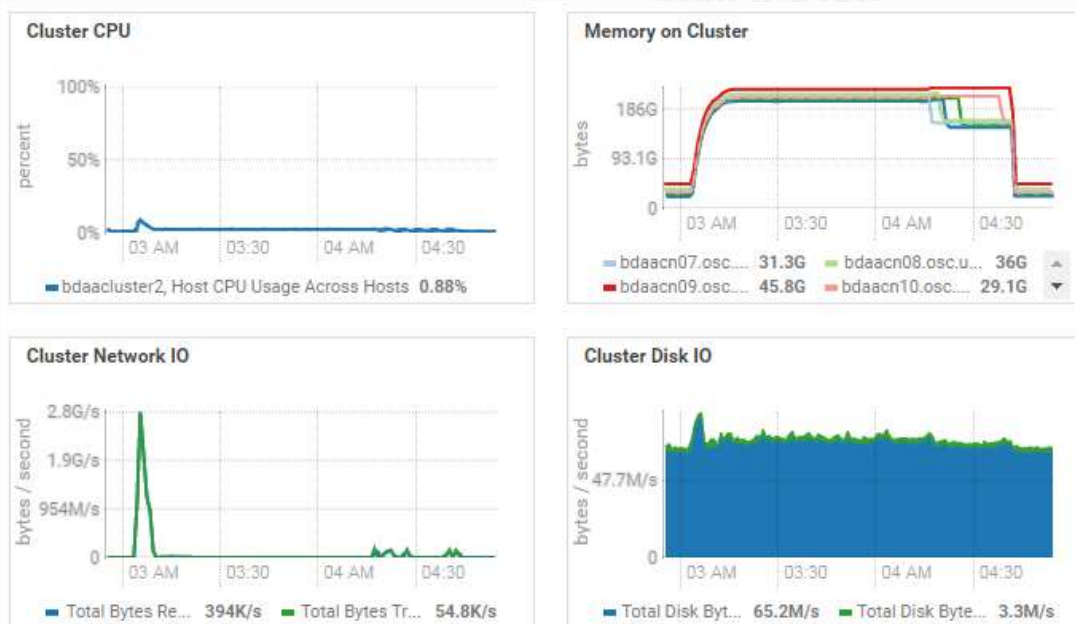
捷克Home Credit：测试场景2—单个SQL测试

HOME
CREDIT

BDS 单用户跑



Impala 单用户跑



1.Impala使用大量的内存，基本把内存占满了。

2.BDS使用更多的CPU资源

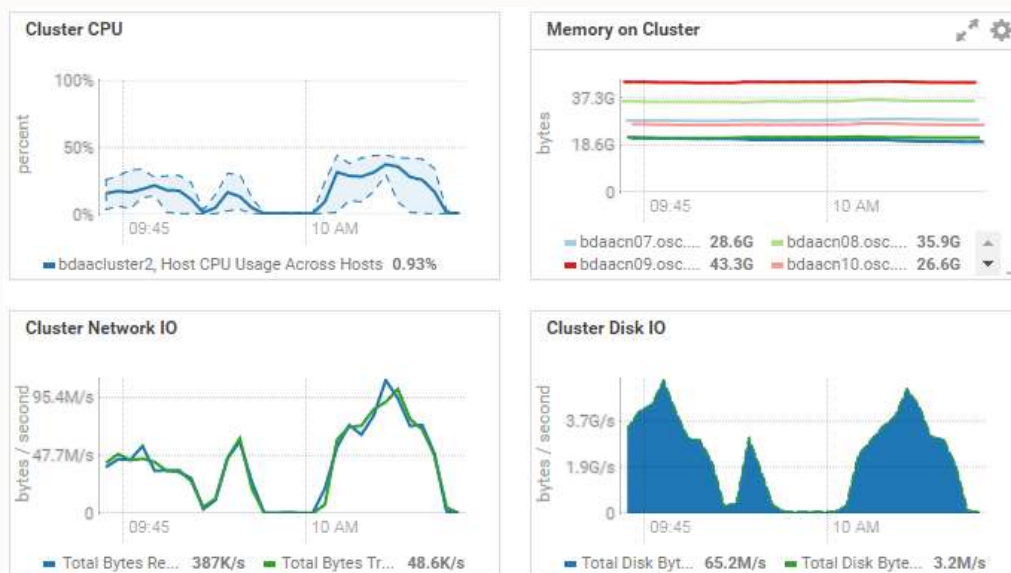


捷克Home Credit: 测试场景2—并发测试

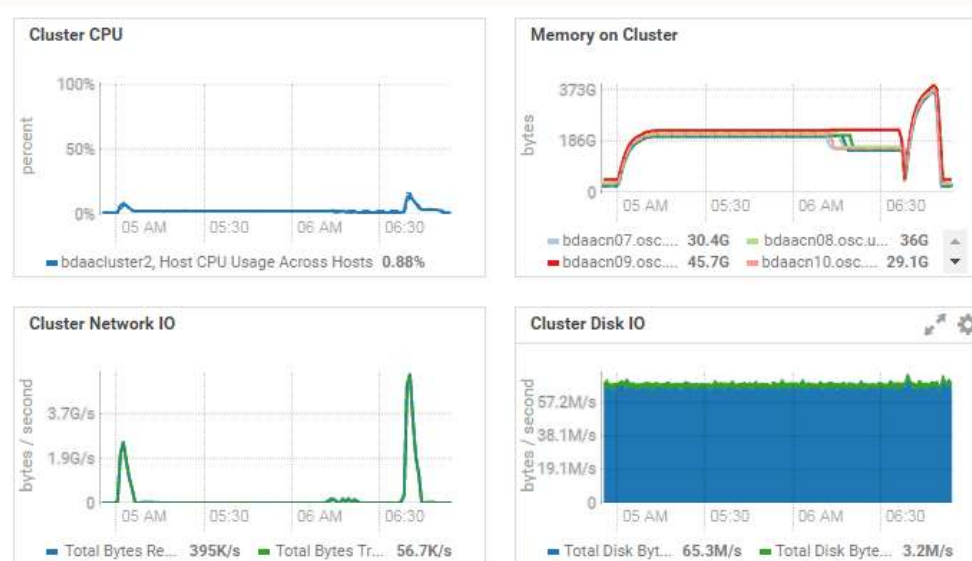
HOME
CREDIT

并发数	Big Data SQL耗时	Impala (Parquet格式) 耗时
1	1081	5818秒
2	PGA 内存不够	Cluster Crashed (OOM)

BDS 单个到2个用户跑



Impala 单个到2个用户跑



With 2 users Impala uses more than 400GB memory on CDH cluster, it exceeded the Impala limit in CDH. While for 2 users BDS uses more than 90 GB memory in PGA of database, it exceeded the PGA limit.

注: BDS使用88个并行度



捷克Home Credit: 总结

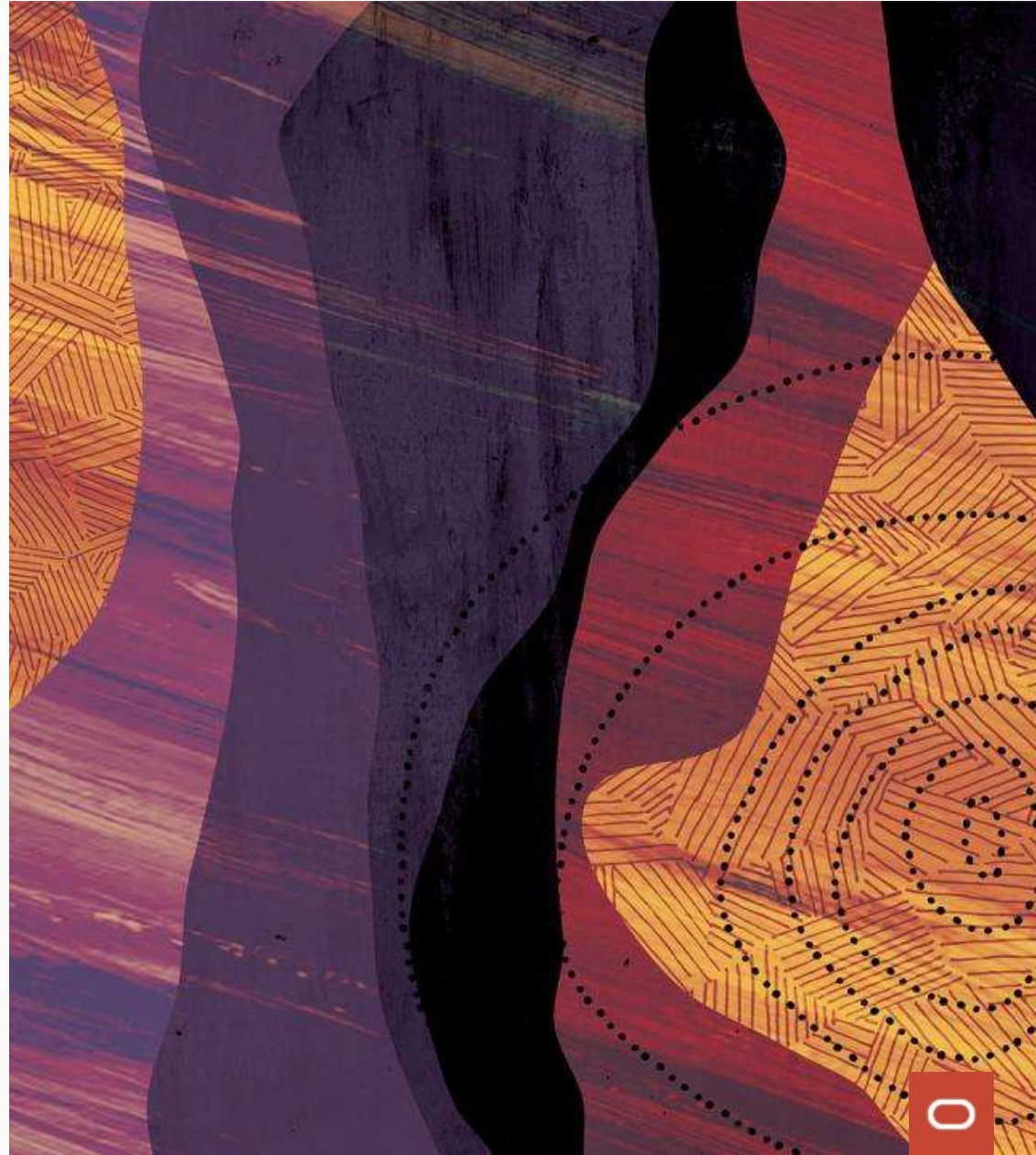
HOME
CREDIT

- ✓ Big Data SQL具有强劲的性能表现，由于智能扫描和存储索引特性。

BDS测试场景	耗时	客户期望时间
场景1	18秒	20分钟
场景2	891秒	40分钟

- ✓ Big Data SQL比Impala有更优的性能。

Q&A



ORACLE

