

大数据技术发展的两大方向和最新进展

孙元浩

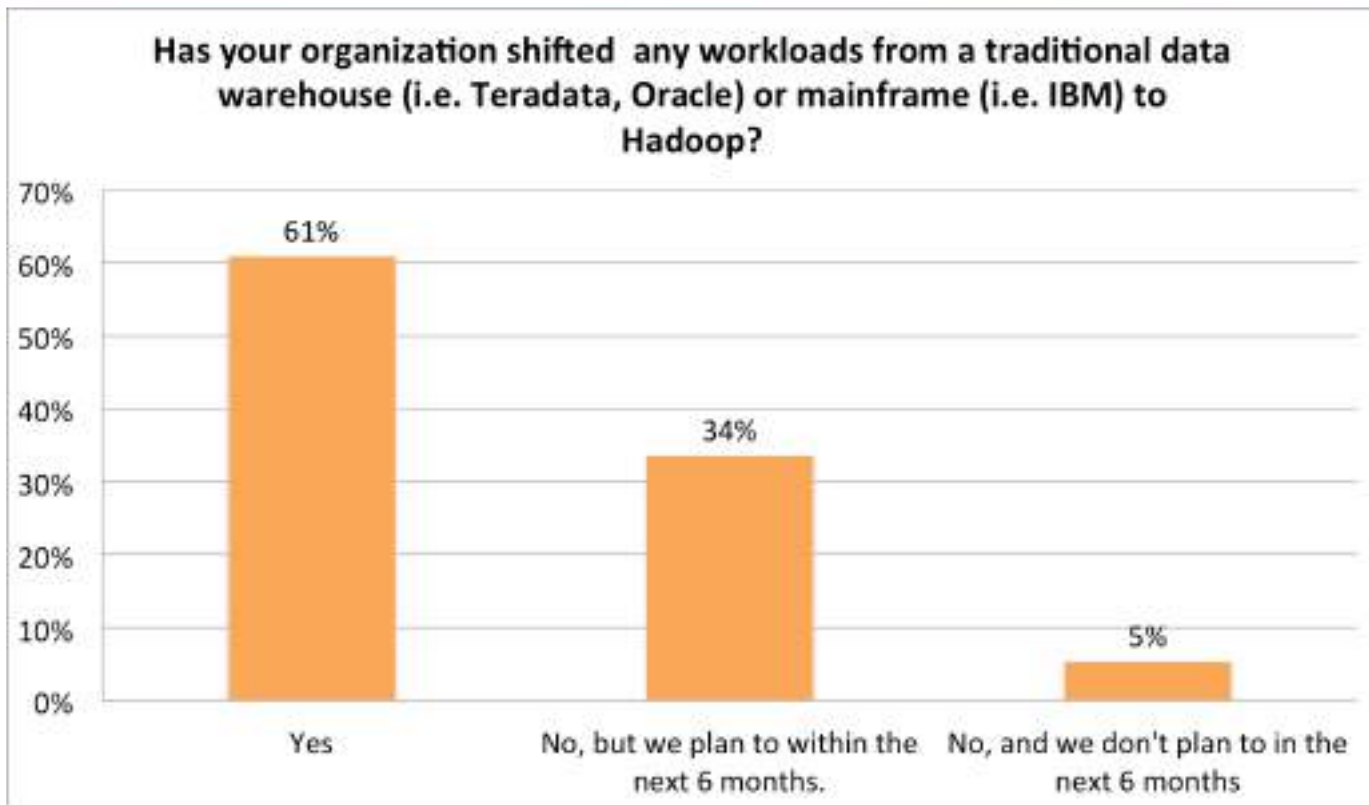
星环科技 联合创始人兼CTO
www.transwarp.io
2015年4月16日



方向一：SQL on Hadoop是Killer App

对SQL支持程度的制约阻碍了企业应用Hadoop技术

60%的Hadoop应用在SQL统计

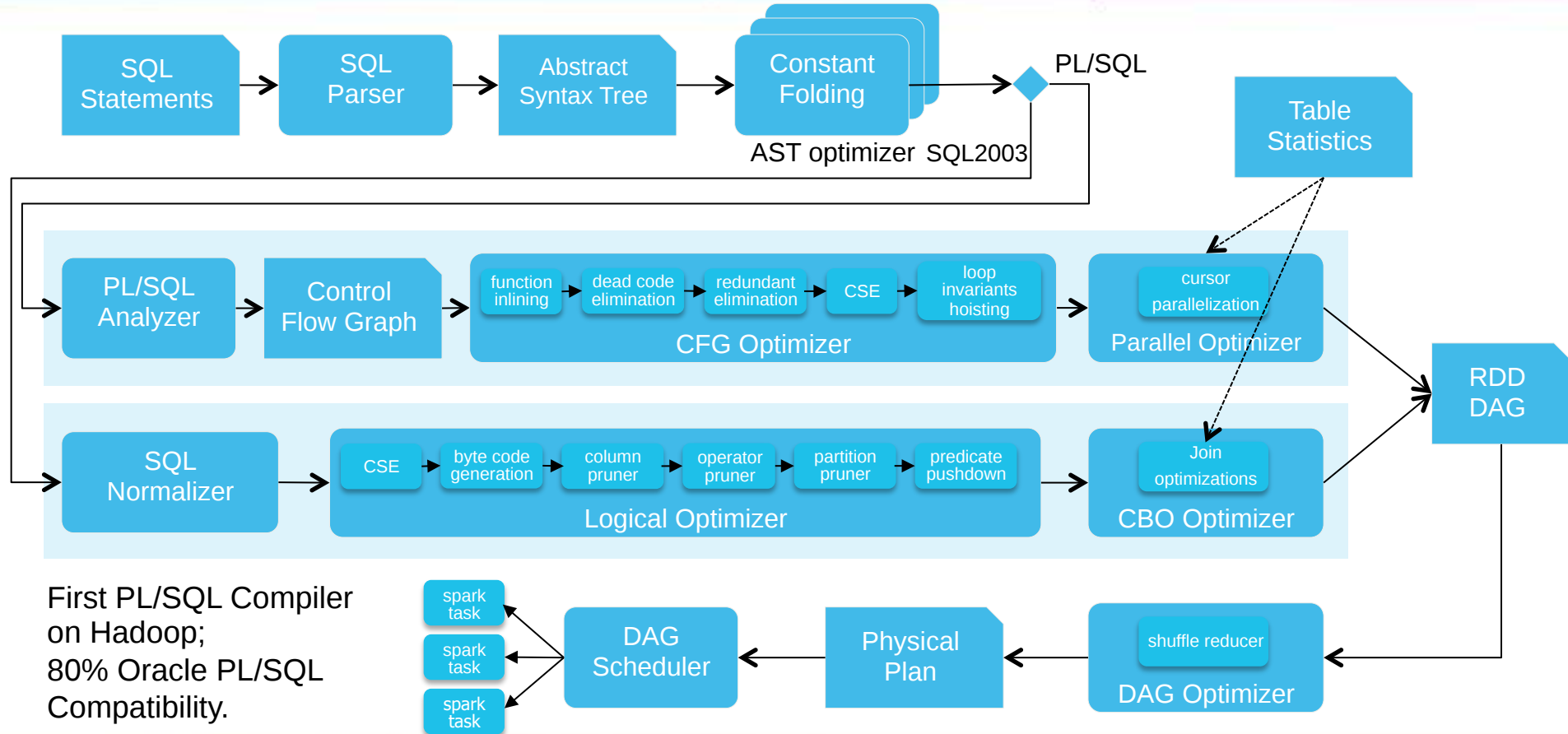


Source: wikibon.org

SQL on Hadoop 技术

名称	计算引擎	ANSI SQL支持程度	存储过程	第一个版本发布时间
Cloudera Impala	类Dremel, 类MPP引擎	SQL92子集+SQL2003扩展	不支持	2011/10
Hortonworks Tez/Stinger	Map/Reduce改进	SQL92子集+SQL2003扩展	不支持	2012/5
Transwarp Inceptor	Spark	SQL99 + SQL2003	Oracle Compatible PL/SQL	2013/11
Databricks SparkSQL	Spark	HiveQL (SQL92子集)	不支持	2014/6
MapR Drill	改进自OpenDremel	SQL92子集	不支持	2012/6立项, 2014/11发布
IBM BigSQL v3	DB2/DPF like MPP Engine over HDFS	SQL 2003	不支持	2014/6
Pivotal HAWQ	Greenplum like MPP Engine over HDFS	SQL 2003	部分支持 (Postgres like)	2013/2
Splice Machine	Apache Derby + HBase	SQL 1999	不支持	2015 GA
Action Vortex	MPP Engine over HDFS	SQL 2003	不支持	2014

Inceptor PL/SQL Compiler



First PL/SQL Compiler
on Hadoop;
80% Oracle PL/SQL
Compatibility.

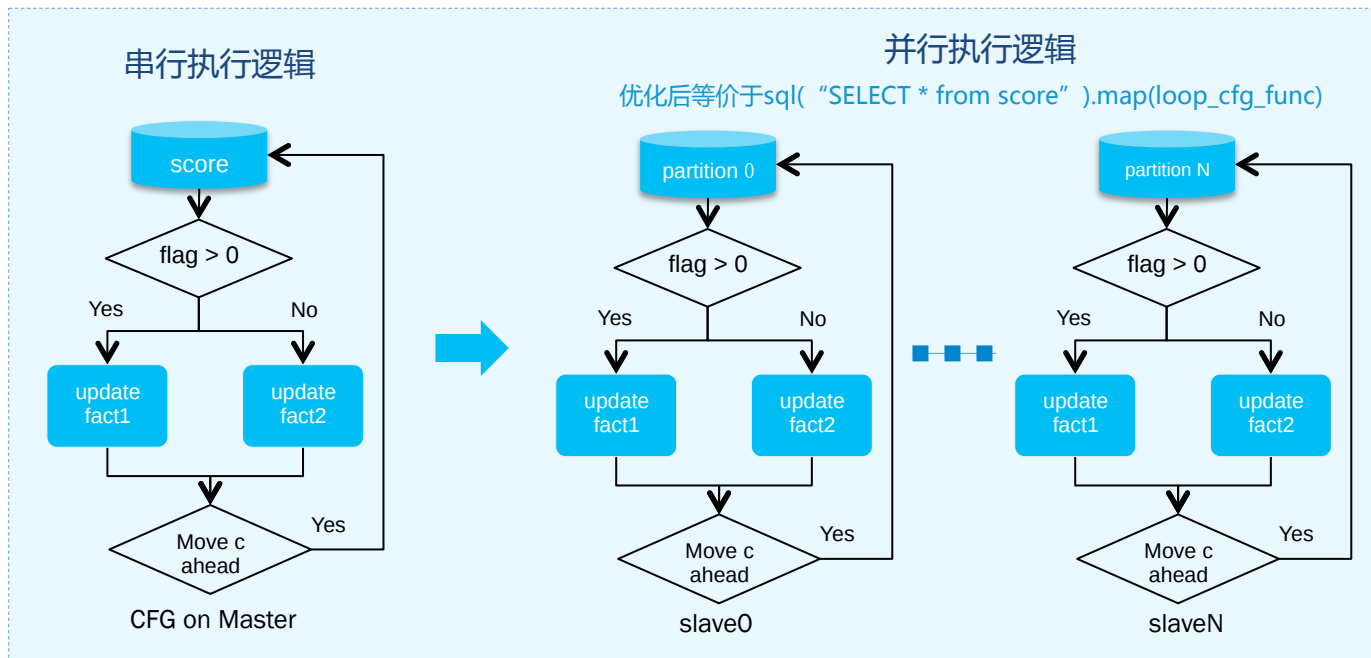
Parallel Query Optimizer

- partition parallelism
- control flow parallelism
- pipeline parallelism

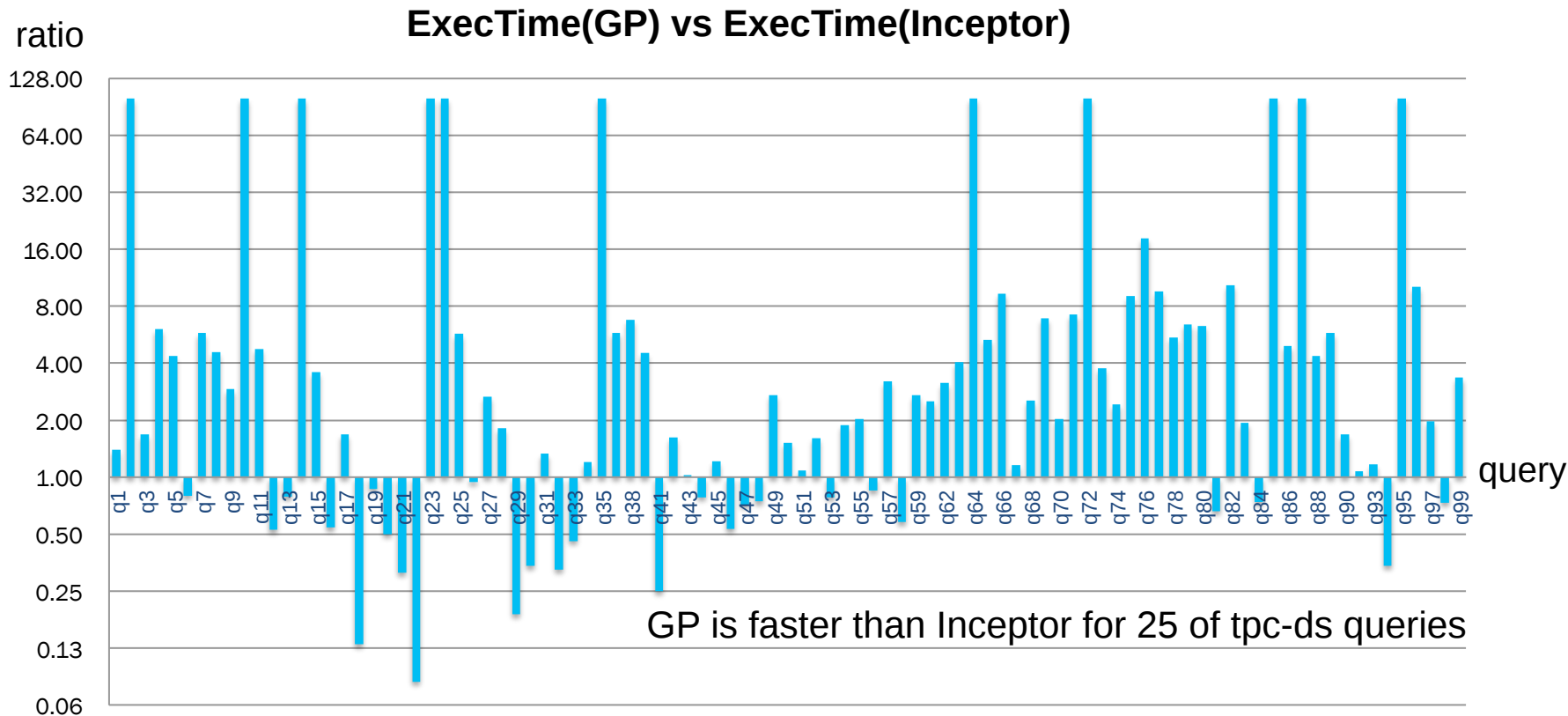
游标示例程序

```
CURSOR c IS SELECT * from  
score  
OPEN c  
FOR v_rec IN c LOOP  
  IF v_rec.flag > 0 THEN  
    UPDATE fact1 SET ...  
  ELSE  
    UPDATE fact2 SET ...  
  END IF  
END LOOP
```

A cursor can be parallelized if there is no loop-carried dependence or the dependence is inductive.



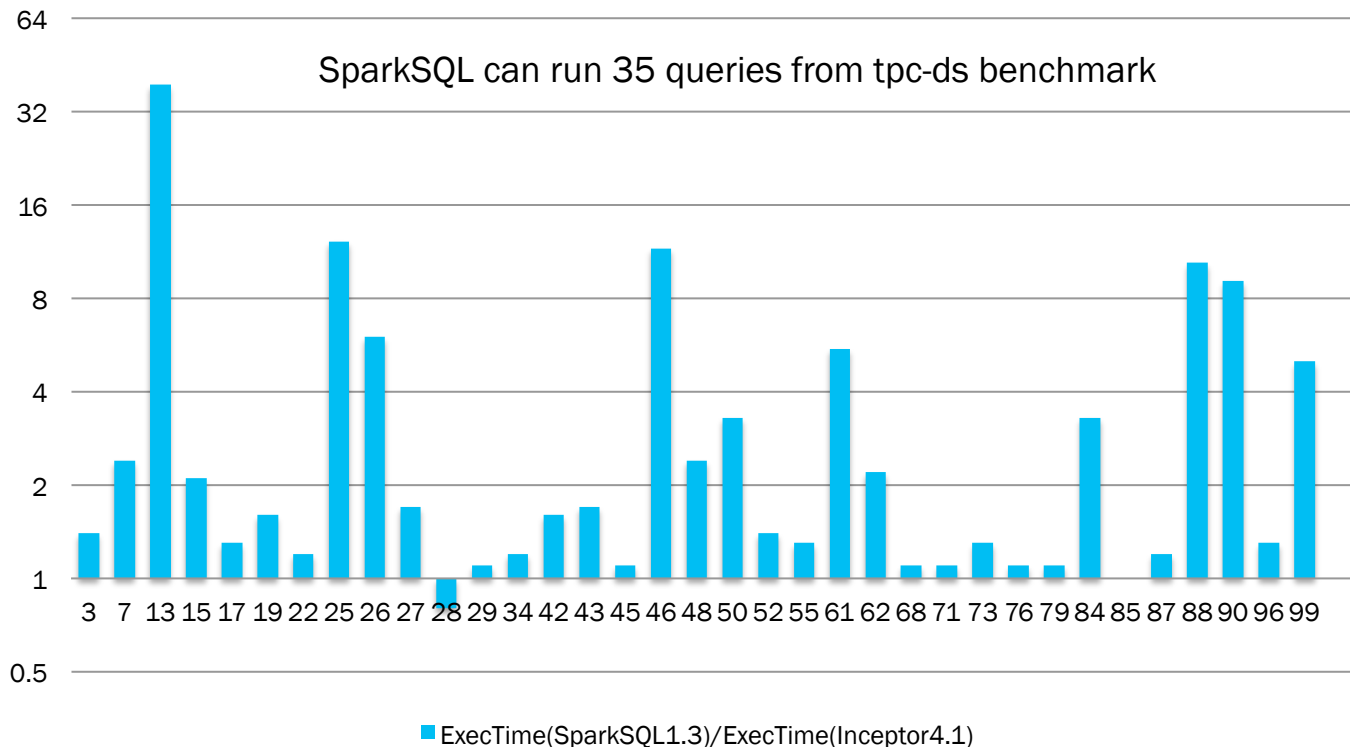
Transwarp Inceptor vs Greenplum DB



Transwarp Inceptor 4.1 vs Spark SQL 1.3

Inceptor性能
比SparkSQL的
加速比例

Inceptor4.1 vs SparkSQL1.3 Speedup

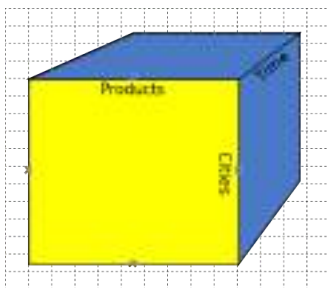


TPC-DS
Query

交互式OLAP分析：Distributed Cube

- Cube是OLAP分析的常用技术

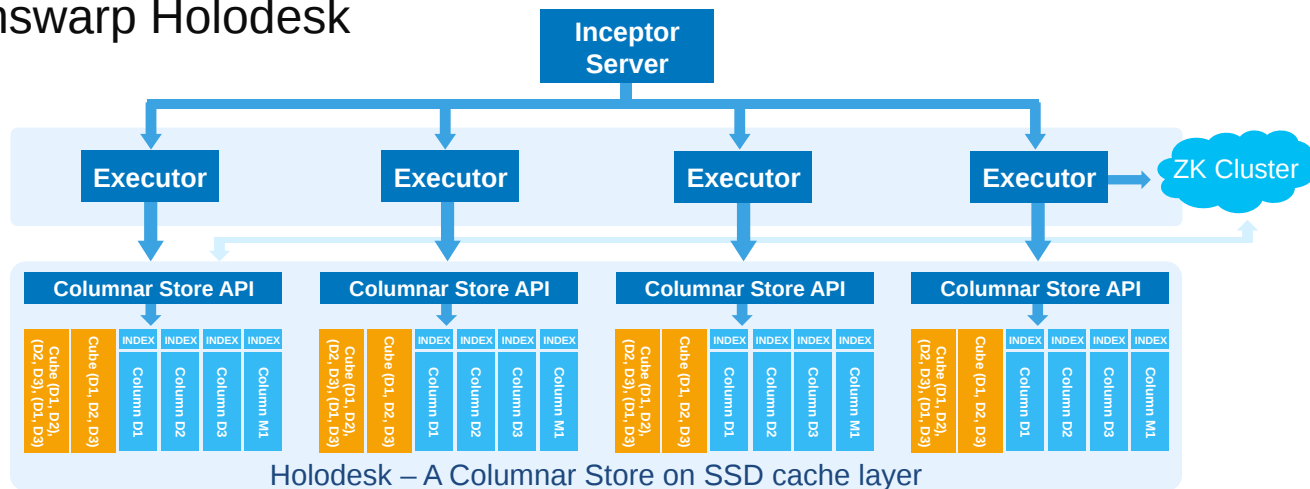
- Slicing
- Dicing
- Rollup
- Drill Up/Down
- Pivot



如何定义一个Cube？

```
create table store_sales tblproperties(  
  'cache'='ram',  
  'holodesk.dimensions'='product, cities, time'  
) as select * from store_sales;00000
```

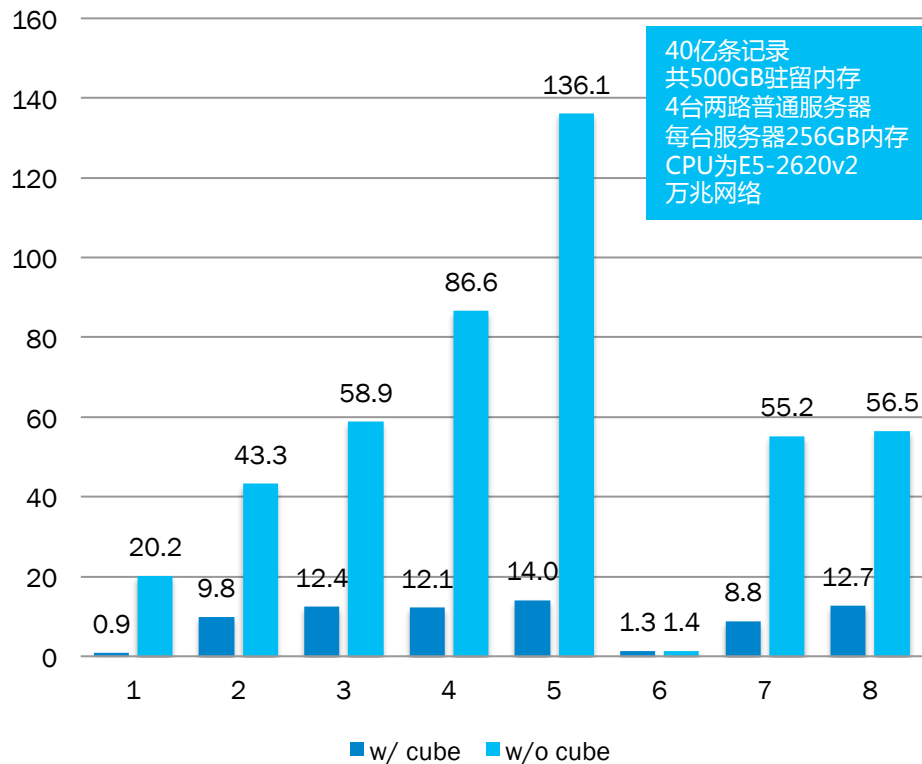
- Cube on Transwarp Holodesk



Holodesk Cube带来的性能加速

Operation		SQL query
q1	count	select count(*) from store_sales
q2	measure	select sum(ss_sales_price) from store_sales
q3	aggregation	select sum(ss_sales_price) from store_sales group by ss_customer_sk
q4	drill down	select sum(ss_sales_price) from store_sales group by ss_sold_date_sk
q5	drill down	select sum(ss_sales_price) from store_sales group by ss_customer_sk, ss_sold_date_sk
q6	slice	select sum(ss_sales_price) from store_sales_r where ss_customer_sk=5000 group by ss_customer_sk, ss_sold_date_sk
q7	dice	select sum(ss_sales_price) from store_sales where ss_sold_date_sk between 2450629 and 2451816 group by ss_customer_sk
q8	pivot	select sum(ss_sales_price) from store_sales where ss_customer_sk > 5000 and ss_sold_date_sk between 2450629 and 2451816 group by ss_customer_sk, ss_sold_date_sk

执行时间 (秒)



方向二：Hadoop加速Docker化

企业对多租户资源管控和弹性计算的需求促使Hadoop发生变革



统一的企业大数据平台 (Data Hub)

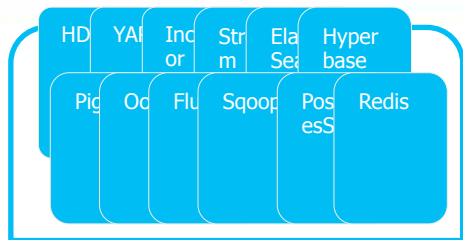
需求一：资源弹性共享 - 提高资源利用率

- 灵活部署：Big Data + Application
- 资源调度：Auto-scaling + Self-healing
- 服务发现：Central Repository

需求二：隔离性 - 保障服务质量和安全性

- 数据隔离：Data Sources, Access Pattern, Confidential Levels
- 计算隔离：CPU、Memory、I/O
- 应用隔离

Announcing Transwarp Operating System



Service Repository

Transwarp Operating System

Ring 0: Docker/Container

Ring 1: Resource scheduler

Ring 2: Built-in system services

Ring 3: Central service repository (docker images)

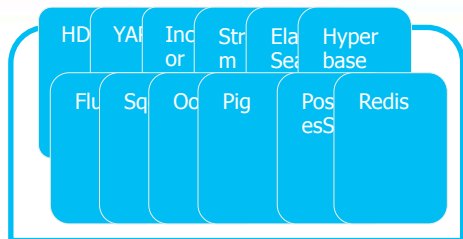
可运行在裸机组成的集群中, 或者是公有云上



Transwarp Operating System

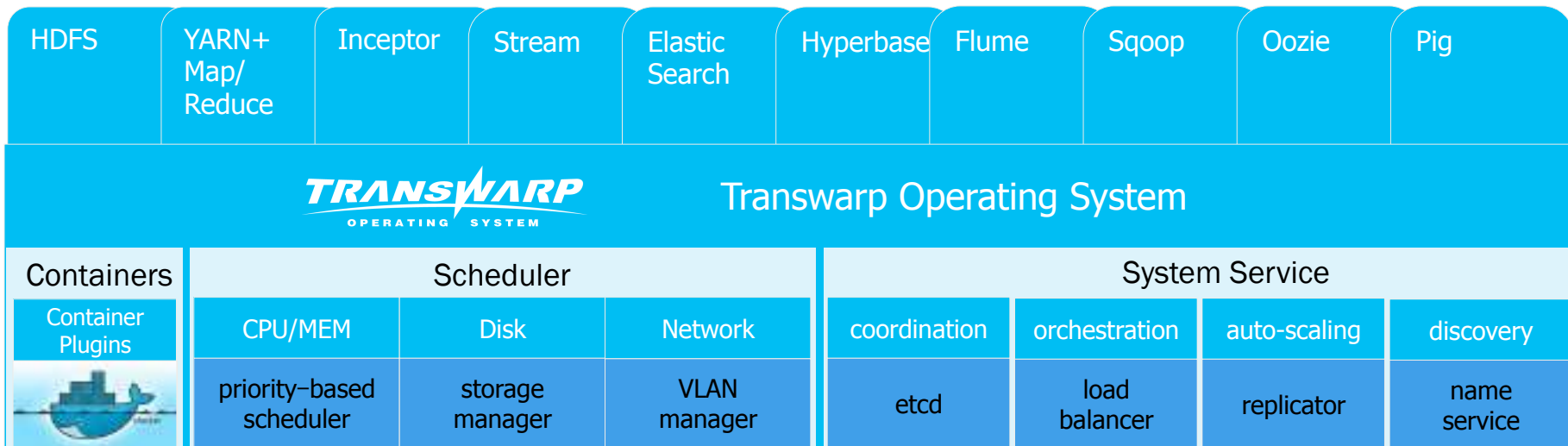
Containers	Scheduler			System Service			
Container Plugins	CPU/MEM	Disk	Network	coordination	orchestration	auto-scaling	discovery
	priority-based scheduler	storage manager	VLAN manager	etcd	load balancer	replicator	name service

TOS – automated deployment



Service Repository

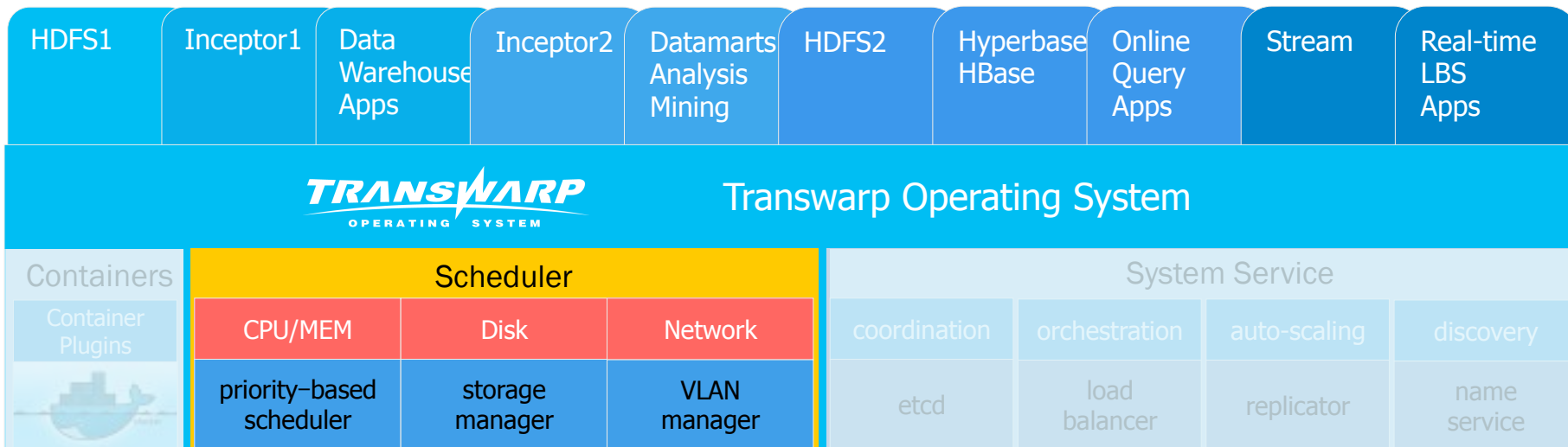
通过Web、REST API or CLI 一键瞬间安装和部署集群
自动根据服务的依赖性安装所需的其它服务组件



TOS – Better Scheduler for Isolation

为什么要重写资源管理框架来代替YARN？

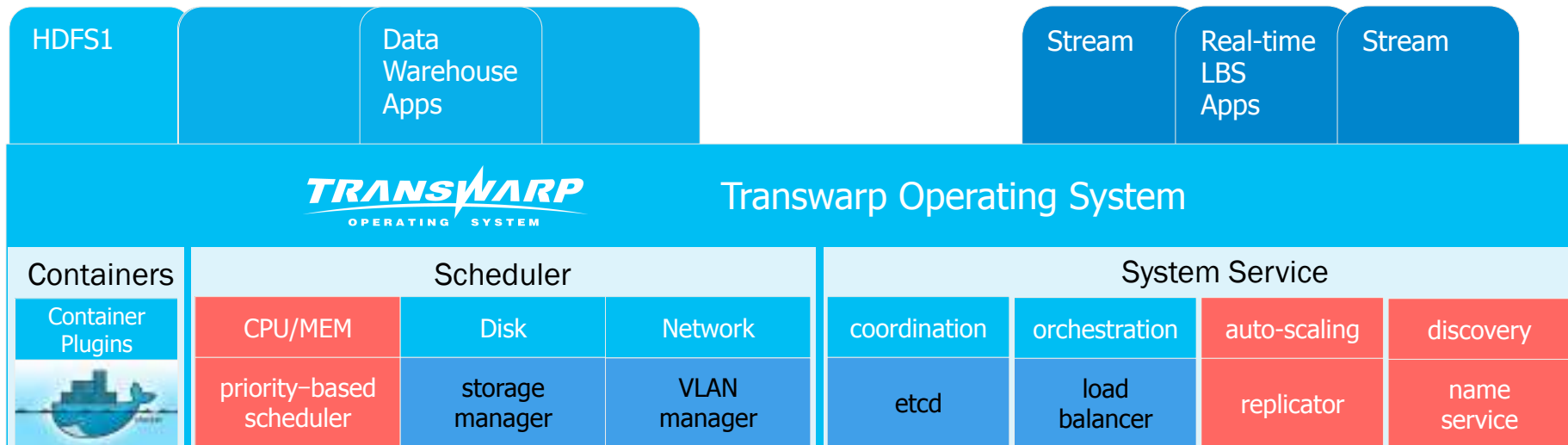
	资源粒度	隔离程度	依赖性	通用性
YARN	CPU/MEM	进程级别、不精确	依赖某个HDFS	支持少量计算引擎
Kubernetes	CPU/MEM	Container	不依赖Hadoop	支持通用Linux负载
TOS	CPU,MEM, DISK,NETWORK	Container + Quota + VLAN	不依赖Hadoop	支持大数据及通用应用



TOS- auto-scaling & self-healing

动态扩容/收缩集群：Capacity Scheduler + Priority/Price-based Bidding (支持抢占)

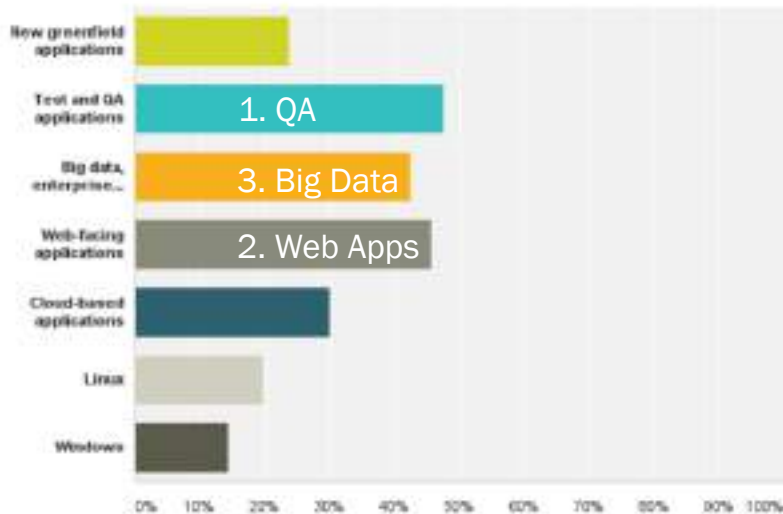
自动修复集群：Replicator监测集群规模并保持该规模



Big Data将成为Docker的主要应用之一

Q7: Where are you using or planning to use Docker over the next year?
(Check all that apply)

Answered: 745 Skipped: 0



Transwarp Operating System

- ✓ automated hadoop deployment
- ✓ run any docker images
- ✓ better isolation
- ✓ auto-scaling & self-healing

Transwarp Operating System 将在2015Q2二季度末发布!

The image features the word "TRANSWARP" in a bold, italicized, red sans-serif font. A red swoosh underline is positioned beneath the letters, with a sharp, upward-pointing triangular element integrated into the middle of the underline, just below the 'W'. The background is a dark space scene showing the curved horizon of the Earth with a bright, glowing sun or star on the left, creating a lens flare effect. The Earth's surface shows blue oceans and green landmasses.

TRANSWARP