大数据分析系统的建设

ruyue.ma@gmail.com

推销两个观点

- 数据系统
- 分层设计
- 为什么推销这两个观点？

数据系统

What is a data system? A system that manages the storage and querying of data.

Query = Function(All data).
Sometimes you retrieve what you stored
Oftentimes you do transformations, aggregations, etc.

MapReduce is a framework for computing arbitrary functions on arbitrary data.
@nathanmarz, the author of Storm

Mysql是数据系统，包括查询层(SQL)和存储层。

HBase、HDFS、 NoSQL都至多算是存储层。

cloudera

Cloudera Impala: HyperTable

Cloudera Impala: How is Hypertable compared to Impala?

Hypertable: http://www.hypertable.com

2 Answers

Doug Judd

17 votes by Mark Johnson, Rebecca Ritter, Christoph Rupp, (more)

Hypertable is a direct technology competitor to Impala. The goal of the Hypertable project, from its inception, has been to build a scalable database to store, analyze, and serve massive datasets. As part of this effort, we are building all of the powerful query capabilities that you find in a traditional relational database, which means full SQL support. Todd was misinformed when he made his response. The job of Hypertable is **not** to act as a simple storage component, but to be a full-featured, high performance, scalable database.

2+ Comments · 1:54 on Tue Nov 06 2012

Todd Lipcon, Employee since Feb '09

11 votes by Denis Arnaud, Patrick Angeles, Alon Amit, (more)

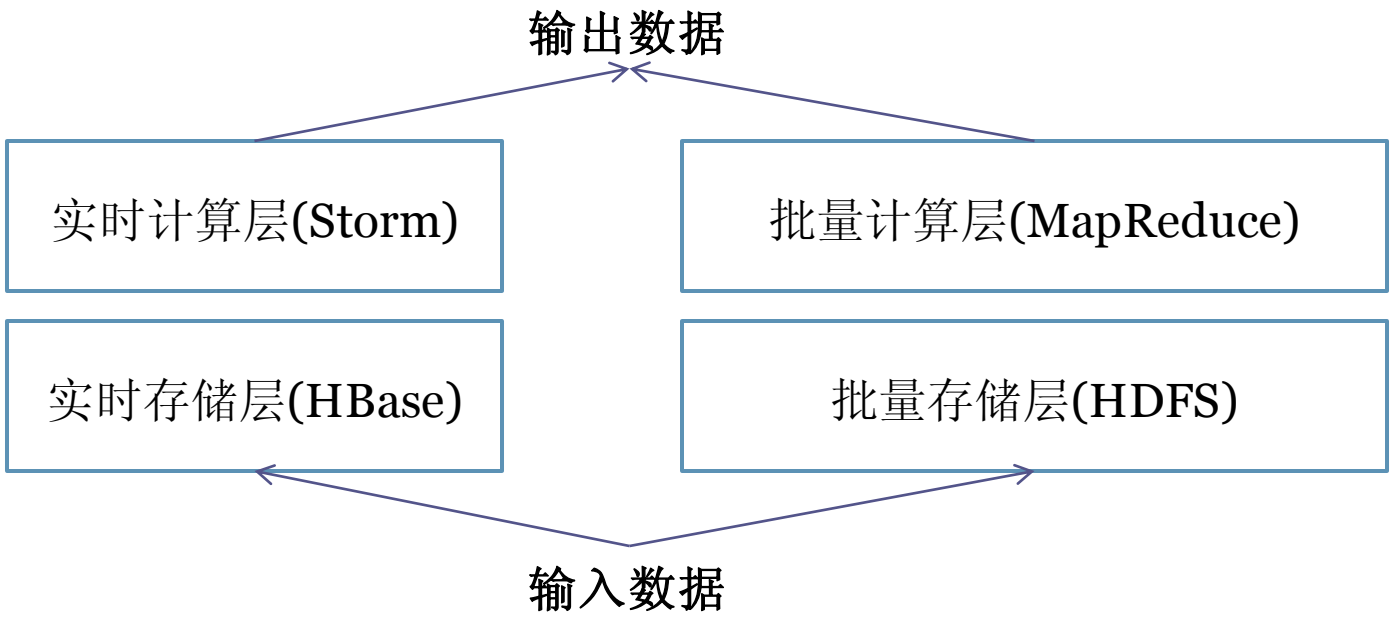
HyperTable is a storage component, whereas Impala is a query engine.

分层设计

- 垂直分层
  - 时效性库，小时库，天级库，周库
- 水平分层
  - OLTP，OLAP
  - SQL，NoSQL
- 为什么分层？
  - 减少设计复杂性
  - 减少使用运维复杂性
  - 资源效率使用最高

为什么推销这两个观点

- 小数据->大数据
- 大数据系统也是storage+query
- 大数据需要分层考虑



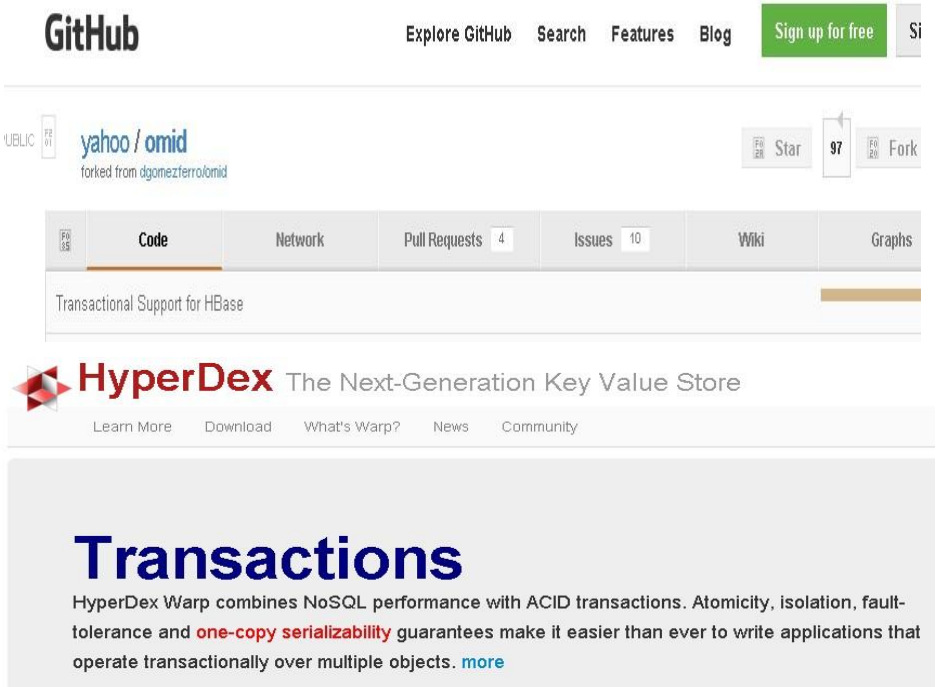
最近两个趋势

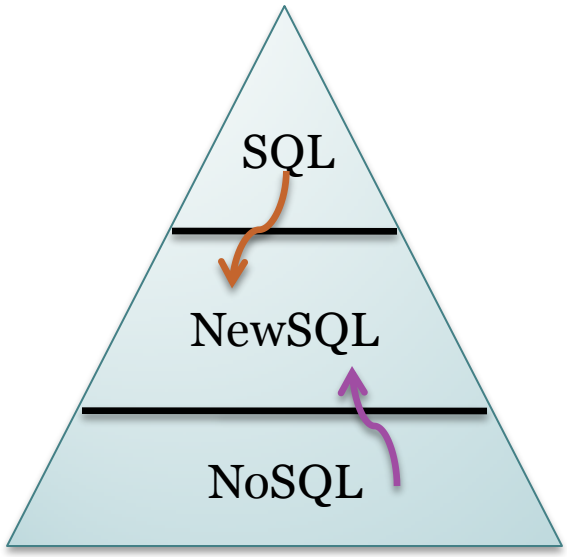
- NewSQL
- Interactive Analysis
- 说明了什么

NewSQL

- NoSQL太过原始，SQL容量性能有限

Megastore: Providing Scalable, Highly Available Storage for Interactive Services (Google 2011)
F1 - The Fault-Tolerant Distributed RDBMS Supporting Google's Ad Business (Google 2012)





Interactive Analysis

- Hive响应太慢，数据库容量性能有限

Dremel: Interactive Analysis of Web-Scale Datasets (Google 2010)
Tenzing : A SQL Implementation On The MapReduce Framework *(Google 2011)*
PowerDrill: Processing a Trillion Cells per Mouse Click (Google 2012)

cloudera
Ask Bigger Questions

WHY CLOUDERA | PRODUCTS | SOLUTIONS | PARTNERS | RESOURCES | SUPPORT | ABOUT

Hadoop & Big Data

Our Customers

Cloudera Impala: Real-Time Queries in Apache Hadoop, For Real

by Marcel Kornacker & Justin Erickson | October 24, 2012 | 14 comments | Tweet

Apache Drill Distributed system for interactive analysis.

Apache Drill (incubating) is a distributed system for interactive analysis of large-scale datasets, based on Google's Dremel. Its goal is to efficiently process nested data. It is a design goal to scale to 10,000 servers or more and to be able to process petabyes of data and trillions of records in seconds.

MemSQL, The Real-Time Analytics Platform.

MemSQL's real-time analytics platform is built on the world's fastest, most scalable in-memory database, capable of simultaneously handling real-time transactions and analytic workloads. MemSQL unleashes the full potential of Big Data by consuming and returning data instantly.

Making Data Work

Shark: Real-time queries and analytics for big data

Shark is 100X faster than Hive for SQL, and 100X faster than Hadoop for machine-learning

Print

Listen

by Ben Lorica | @bigdata | Comment | November 27, 2012

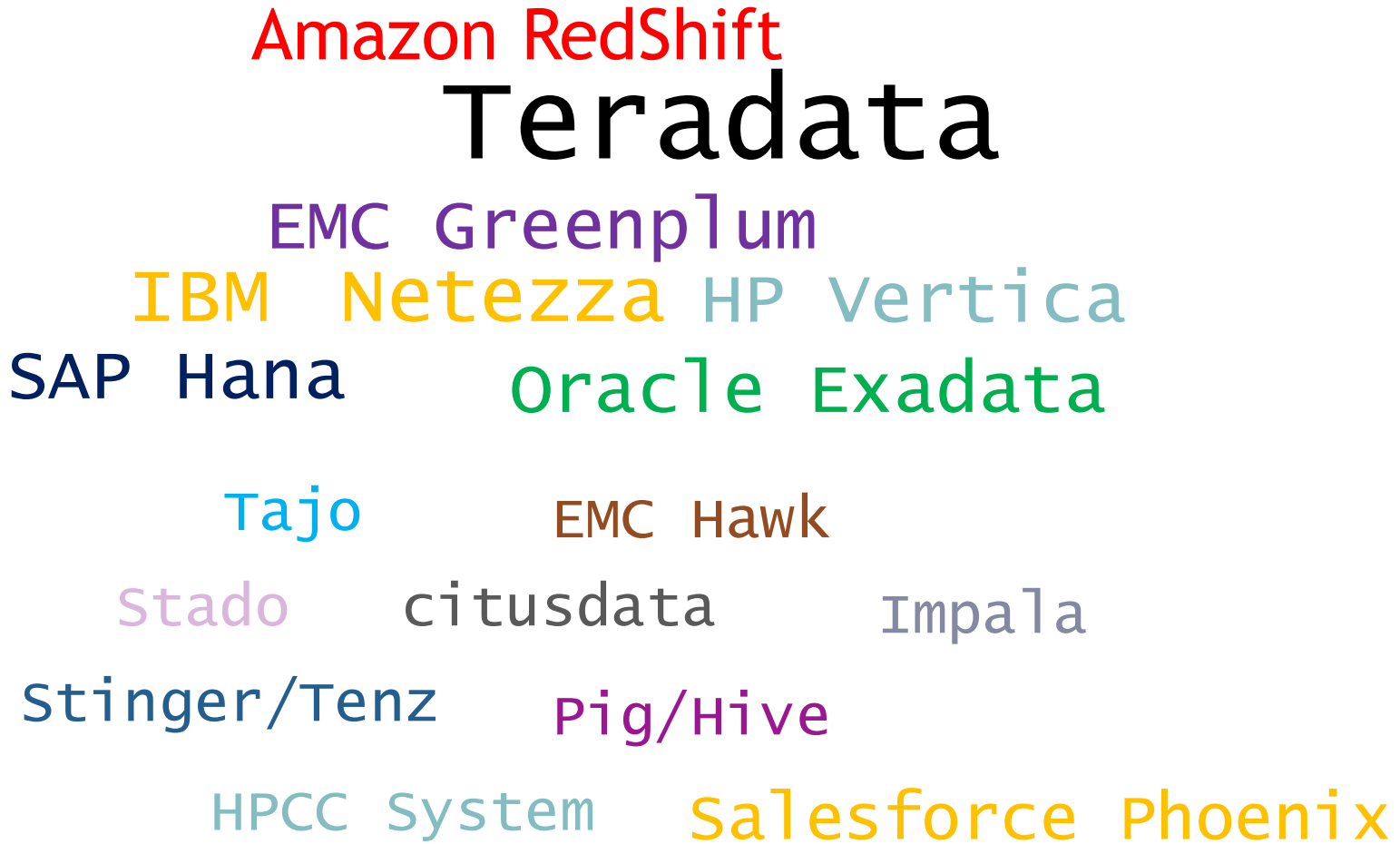
说明了什么

- 大数据的开源社区正在向数据库厂商发起挑战
- OLTP: 难度会稍大，撬动更多的是mysql、postgresql的领地
- OLAP: 很有希望
  - 成本昂贵
  - 稳定性要求低
  - 数据量大
  - 时效性低
  - 不是不可缺少的组件

今天的重点：

大数据分析

热词榜


Amazon RedShift
Teradata
EMC Greenplum
IBM Netezza
HP Vertica
SAP Hana
Oracle Exadata
Tajo
EMC Hawk
Stado
citusdata
Impala
Stinger/Tenz
Pig/Hive
HPCC System
Salesforce Phoenix

大数据分析架构

Business Intelligence

Reporting and Analysis

Static Report | Query Report | OLAP Report and Analysis

Datamart

RDBMS | RDBMS | OLAP DB

Other

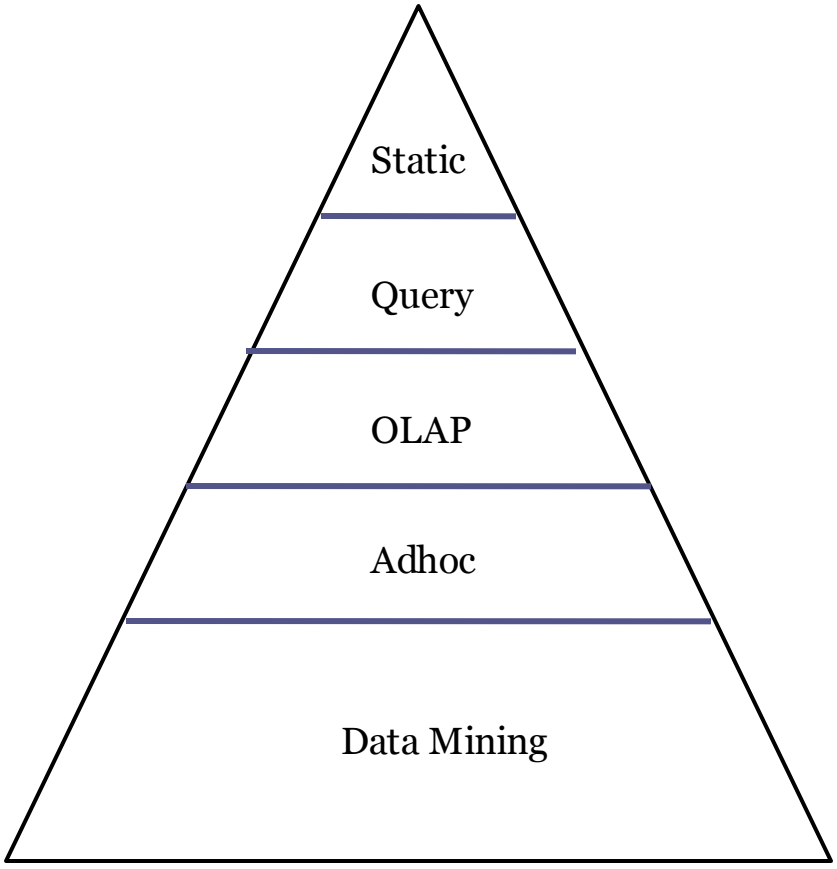
文本分析 | Data Mining

ETL-2

Dataware House

ETL-1

大数据分析发展趋势



从上往下：

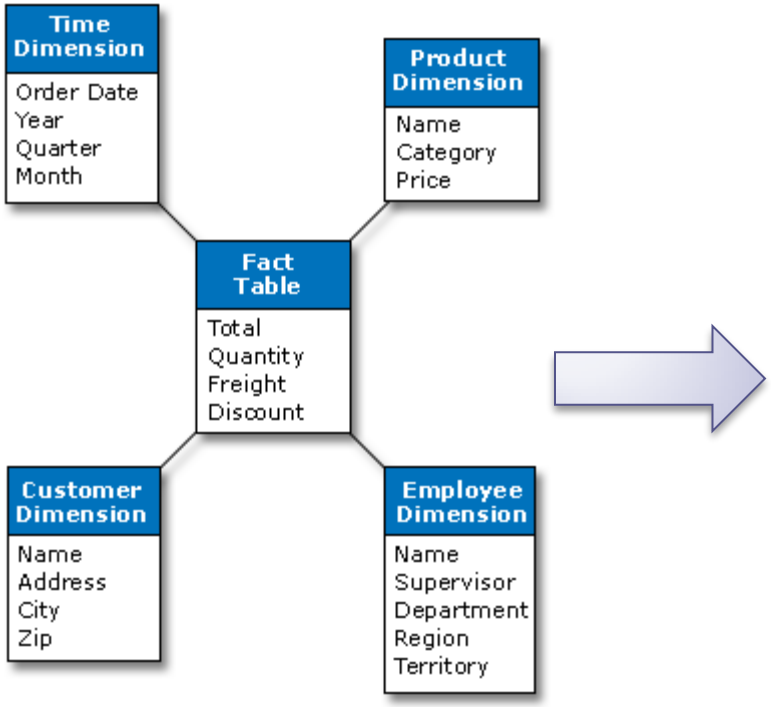
1. 数据量越来越大，维度越来越多
2. 交互性越来越难做
3. 技术难度越来越大
4. 以人为主->以机器为主
5. 用户专业程度越来越高，越来越少

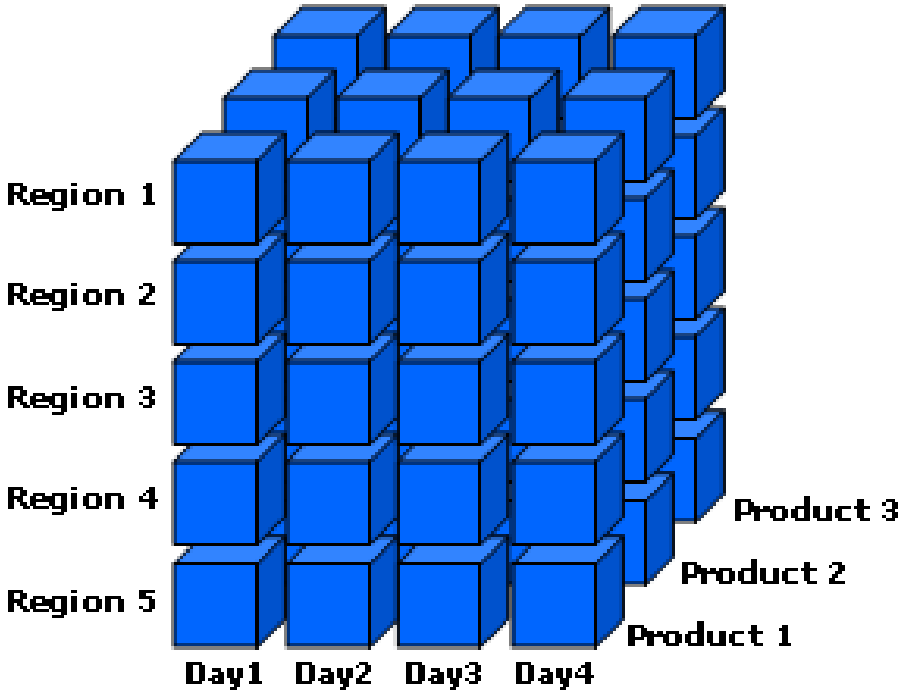
非结构化数据

- 非结构化数据进行结构化后，利用原有技术分析
- 直接文本分析
  - 百度热搜词 – static report
  - 用户query分析 – query report
  - 搜索引擎 – OLAP多维分析
  - MapReduce上的调研作业 – adhoc
  - 新闻聚类 – data mining

OLAP技术难点

- 多维分析: rollup, drill-down, slicing和dicing
- 各类维度组合，并提供交互式响应





OLAP grid

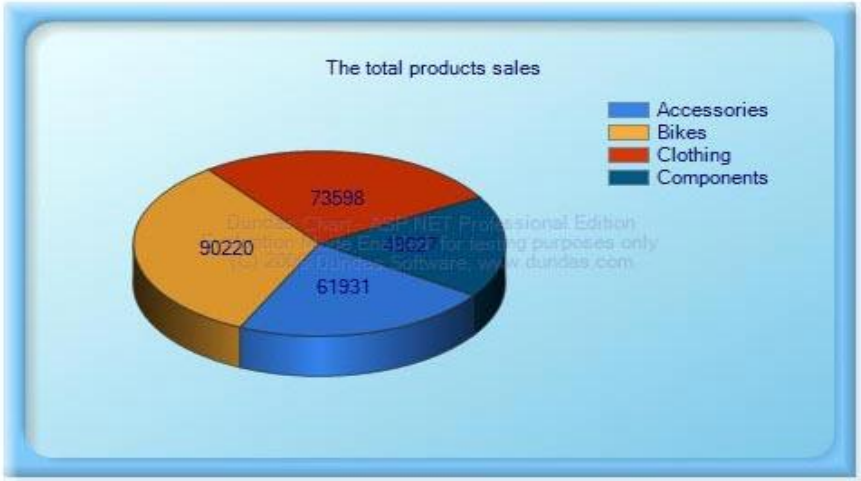
- Measures
- KPIs
- Customer
- Date
- Delivery Date
- Employee
- Geography
- Internet Sales Order Details
- Product

Filter area | Columns area: Date.Fiscal

Rows area: Product Categories | Values: Order Quantity, Product Gross Profit Margin

Measures position
(•) Columns area ( ) Before all
( ) Rows area (•) After all

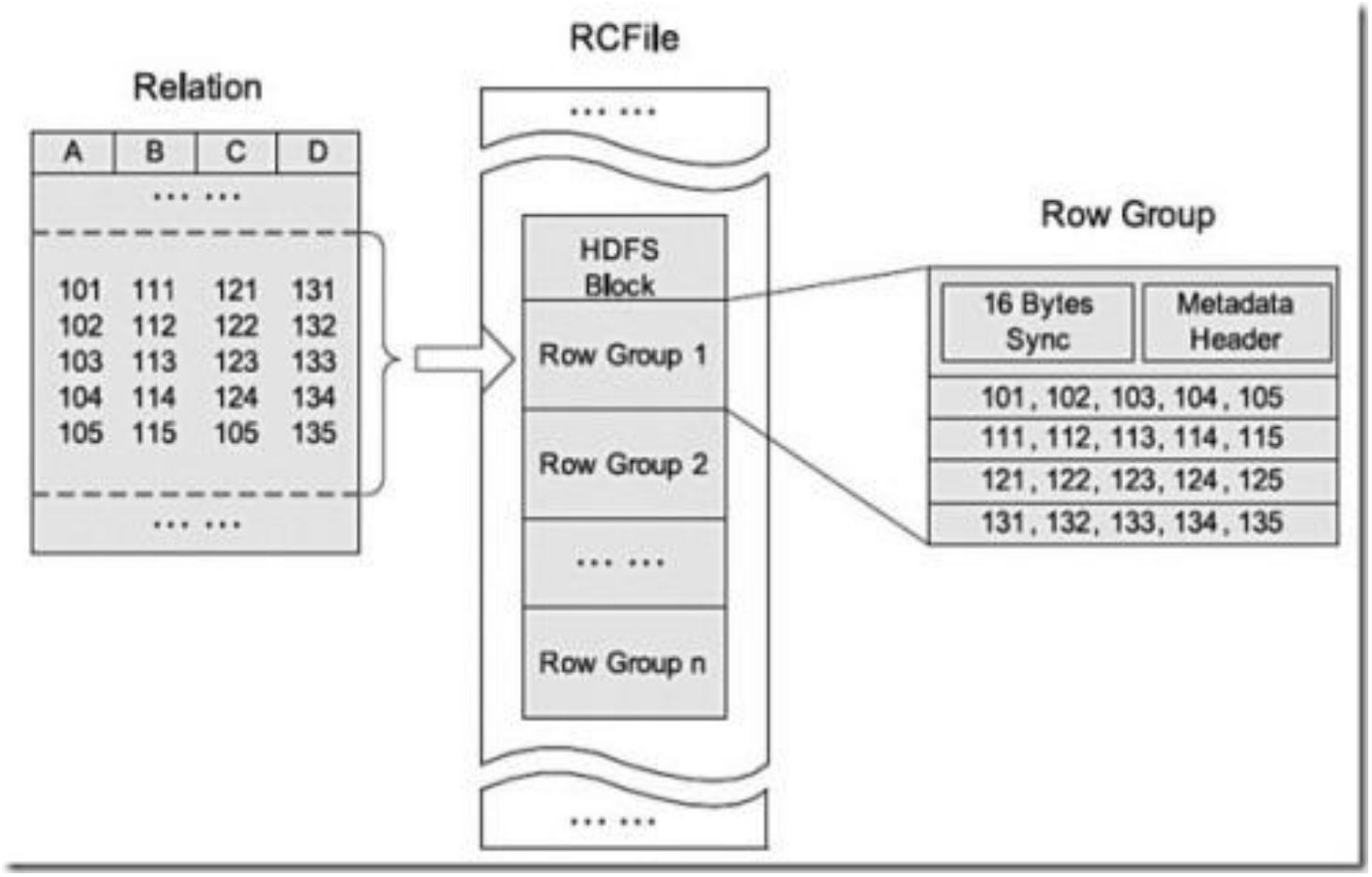
Fiscal Year		FY 2002					Total			
Measures		Order Quantity	Product Gross Profit Margin				Order Quantity	Product Gross Pr		
Category	Subcategory		Value	Goal	Status	Trend		Value	Goal	S
Accessories		1 825	40,38%	40,00%		0,00%	1 825	40,38%	40,00%	
Bikes	Mountain Bikes	4 951	4,84%	12,00%		0,00%	4 951	4,84%	12,00%	
	Road Bikes	10 331	18,38%	12,00%		0,00%	10 331	18,38%	12,00%	
	Touring Bikes			12,00%		0,00%			12,00%	
	Bikes	**15 282**	**12,31%**	**12,00%**		**0,00%**	**15 282**	**12,31%**	**12,00%**	
Clothing		4 084	-5,57%	20,00%		0,00%	4 084	-5,57%	20,00%	
Components		2 905	8,94%	10,00%		0,00%	2 905	8,94%	10,00%	
Total		**24 096**	**12,13%**	**12,00%**		**0,00%**	**24 096**	**12,13%**	**12,00%**	



OLAP技术难点 - 解决手段

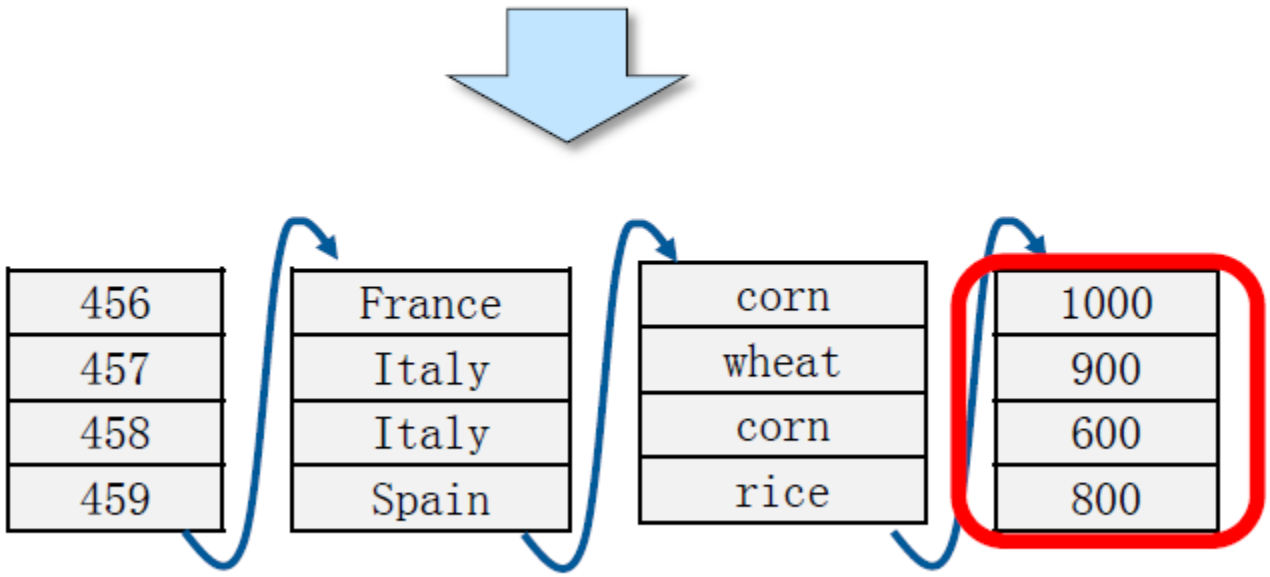
- 减少不必要的列读写
  - 行列混合
  - 列式存储
- 减少不必要的行读写
  - hyperdex – 多维hash
  - infobright – knowledge grid
- 压缩
- 预先计算
  - 块级别的
  - 物化视图

减少不必要的列读写 - 行列混合



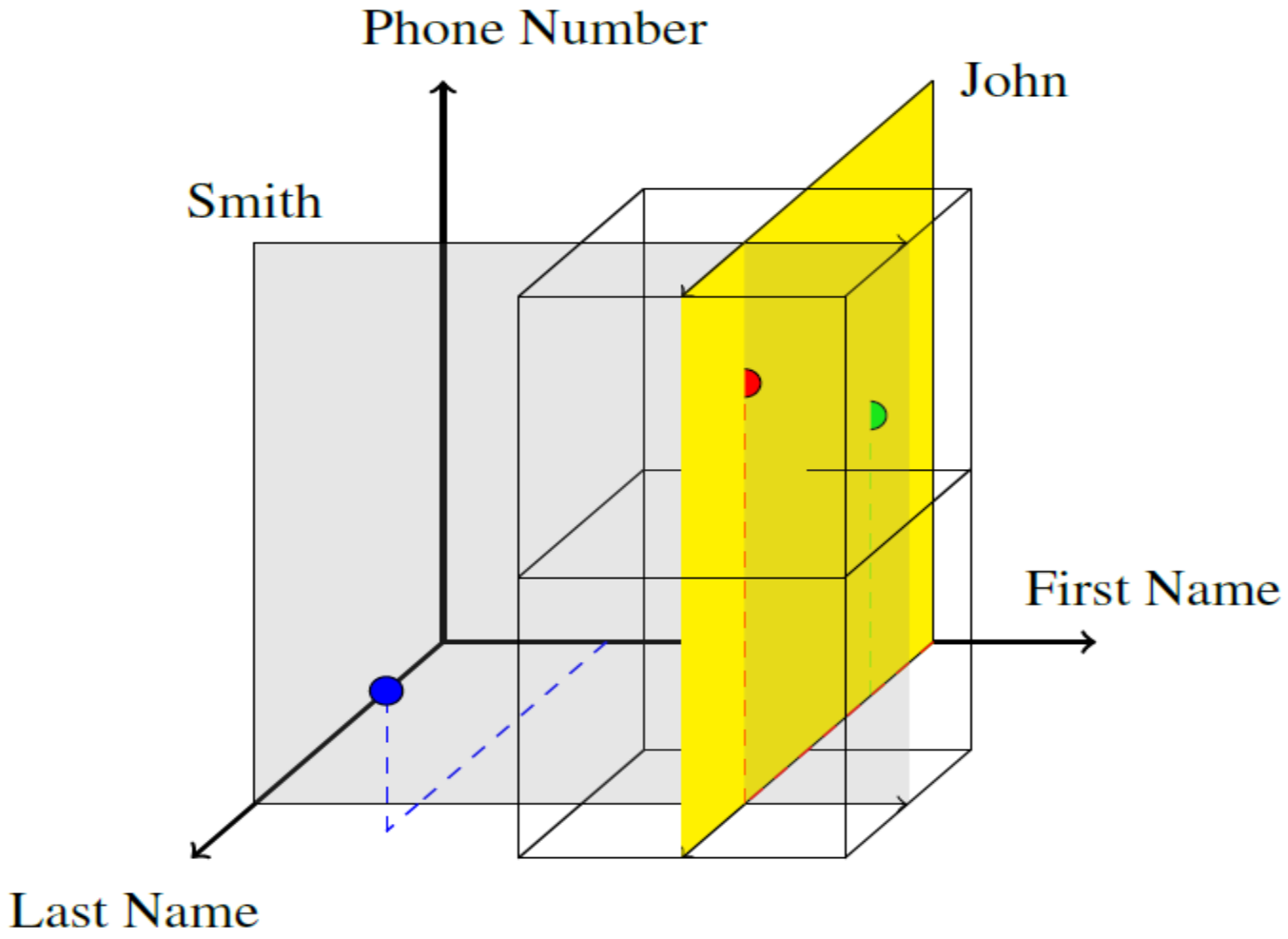
减少不必要的列读写 – 列式存储

Order	Country	Product	Sales
456	France	corn	1000
457	Italy	wheat	900
458	Italy	corn	600
459	Spain	rice	800



Column order organization

减少不必要的行读写 – 多维hash

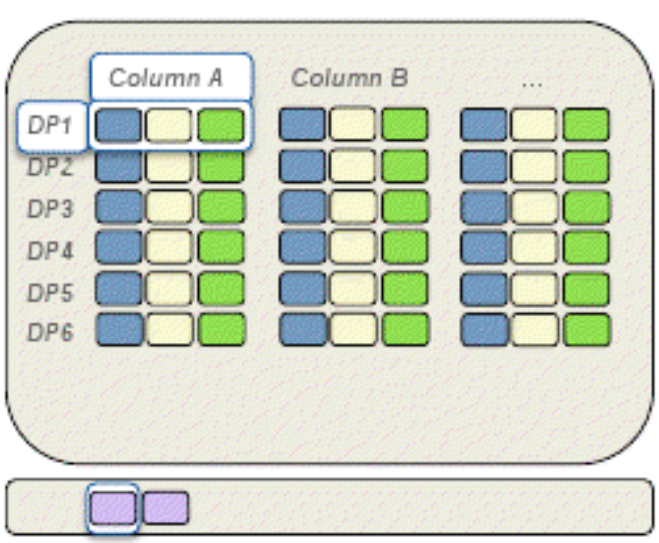


减少不必要的行读写 – infobright knowledge grid

Knowledge Grid
applies to the whole table

Knowledge Nodes
built for each Data Pack

Information about the data



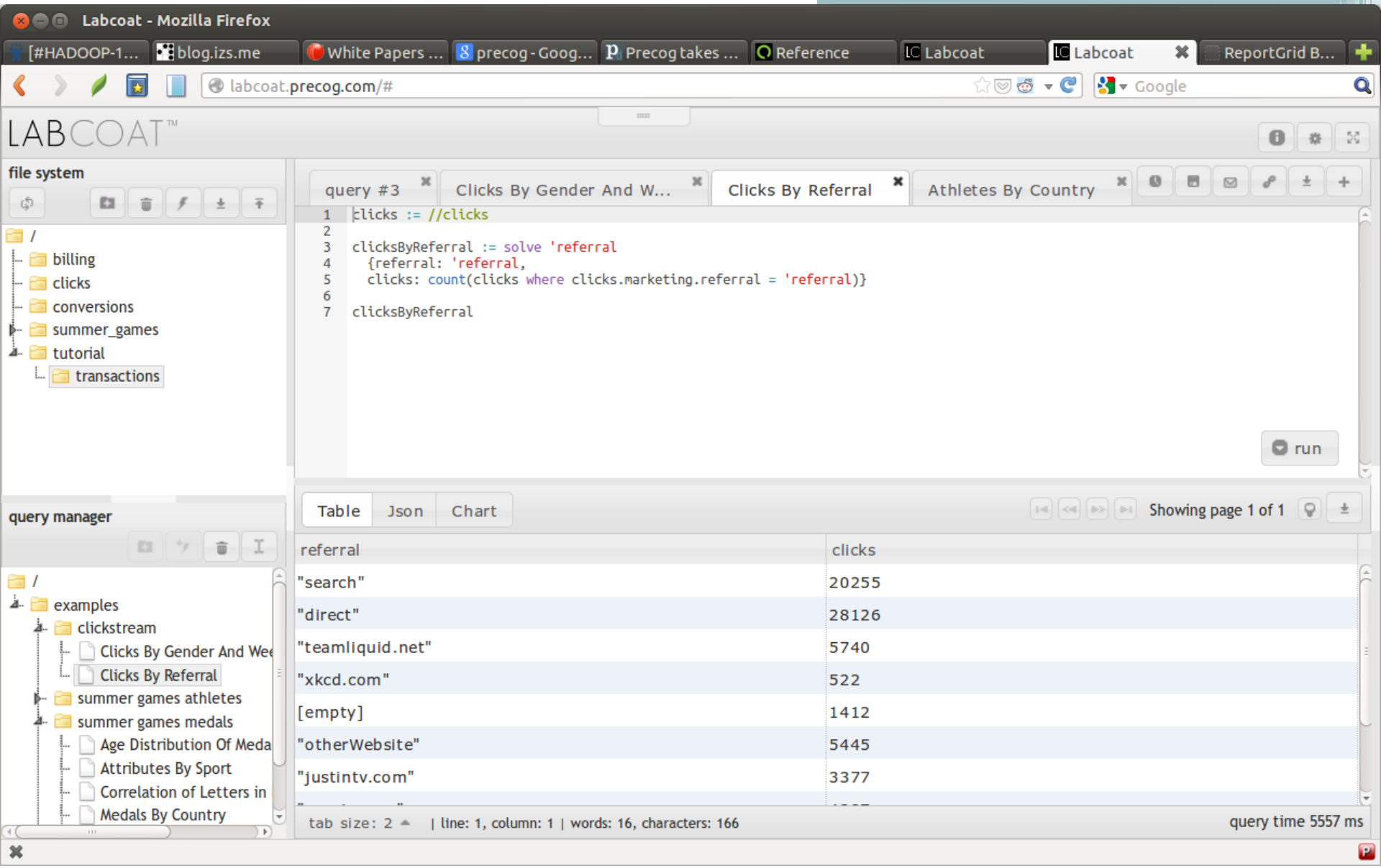
DPN	Data Pack Node	Built during LOAD
Histogram	Numerical Histogram	
CMAP	Character Map	
P-2-P	Pack-to-Pack	Built using JOIN

预先计算

- 块级别的
  - 对每一个数据块，提前计算好其max, min, sum, count等。
- 物化视图
  - 提前计算好需要的几个维度的rollup表

Adhoc技术难点

- 任意维度分析: 存储优化，等同OLAP
- 交互式响应
  - MapReduce太慢
  - Impala
- 任意分析: 简单的SQL可能并不好用
- 方便的查询分析编写环境和展现工具
  - 数据可能导入OLAP做进一步分析


Labcoat - Mozilla Firefox
[#HADOOP-1...
blog.izs.me
White Papers ...
precog - Goog...
Precog takes ...
Reference
Labcoat
Labcoat
ReportGrid B...
labcoat.precog.com/#
Google
LABCOAT™
file system
/
billing
clicks
conversions
summer_games
tutorial
transactions
query manager
/
examples
clickstream
Clicks By Gender And Wee
Clicks By Referral
summer games athletes
summer games medals
Age Distribution Of Meda
Attributes By Sport
Correlation of Letters in
Medals By Country
query #3
Clicks By Gender And W...
Clicks By Referral
Athletes By Country
1 clicks := //clicks
2
3 clicksByReferral := solve 'referral
4 {referral: 'referral,
5 clicks: count(clicks where clicks.marketing.referral = 'referral')}
6
7 clicksByReferral
run
Table
Json
Chart
Showing page 1 of 1
referral
clicks
"search"
20255
"direct"
28126
"teamliquid.net"
5740
"xkcd.com"
522
[empty]
1412
"otherWebsite"
5445
"justintv.com"
3377
tab size: 2
| line: 1, column: 1 | words: 16, characters: 166
query time 5557 ms

Labcoat - Mozilla Firefox

LABCOAT™

file system

- /
  - billing
  - clicks
  - conversions
  - summer_games
    - athletes
    - historic_medals
    - london_medals
  - tutorial

query #3 | Clicks By Gender And W... | Clicks By Referral | Athletes By Country | *Percentage of Female A...

```
1  import std::time::*
2
3   clicks := //clicks
4   clicks' := clicks with {week: weekOfYear(clicks.timeStamp)}
5
6  clicksByGenderAndWeek := solve 'gender, 'week
7    {week: 'week,
8    gender: 'gender,
9    clicks: count(clicks' where clicks'.customer.gender = 'gender & clicks'.week = 'week)}
10  clicksByGenderAndWeek
```

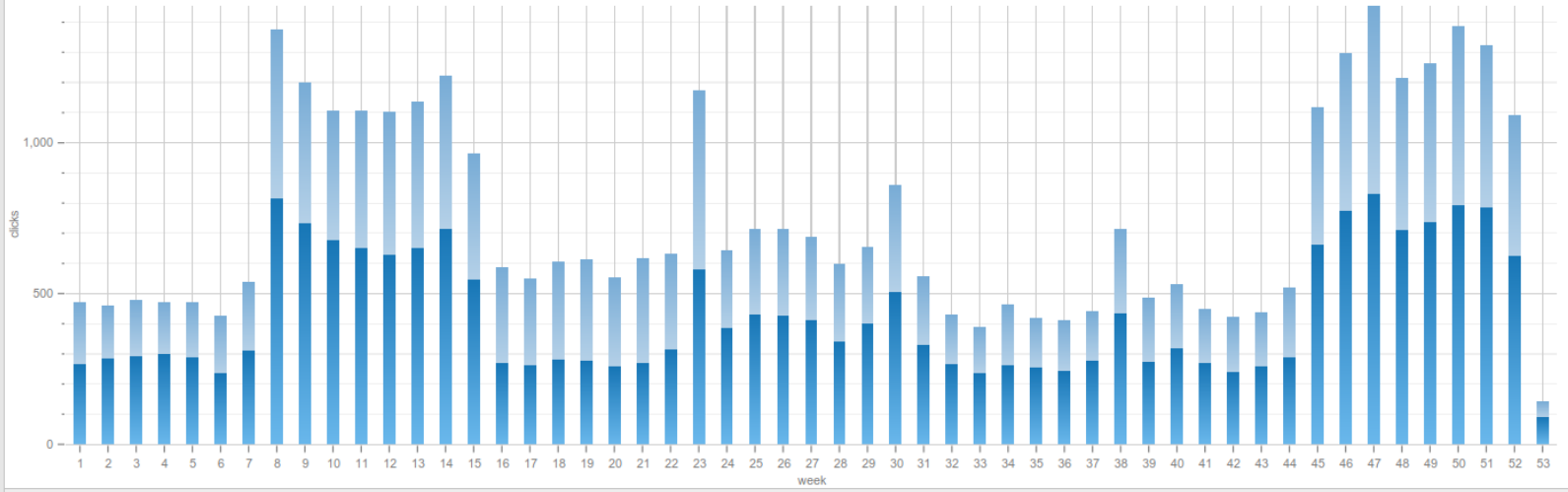
run

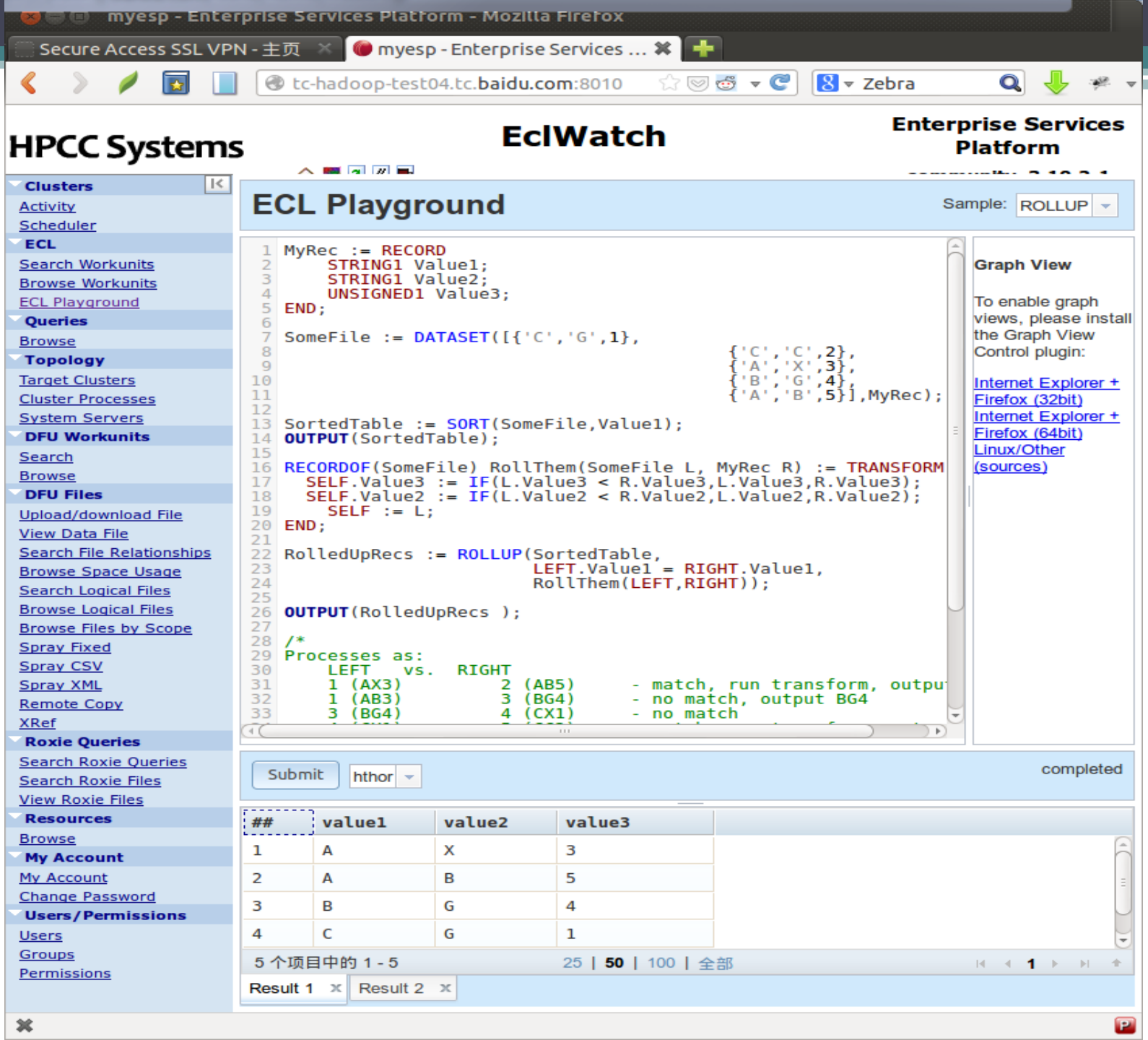
query manager

- /
  - examples
    - clickstream
      - Clicks By Gender And Week
      - Clicks By Referral
    - summer games athletes
      - Athletes By Country
      - Athletes By Sport
      - Top US Sports - Men
      - Top US Sports - Women
    - summer games medals
    - tutorial

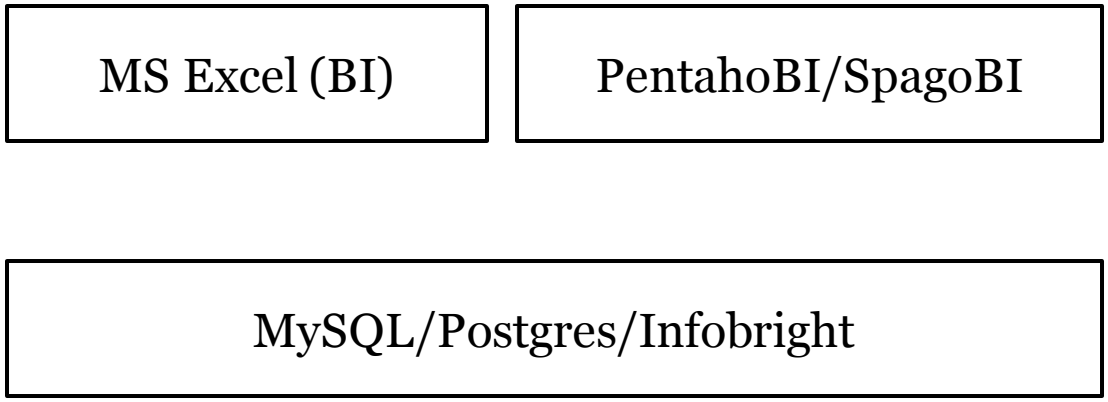
Table | Json | Chart

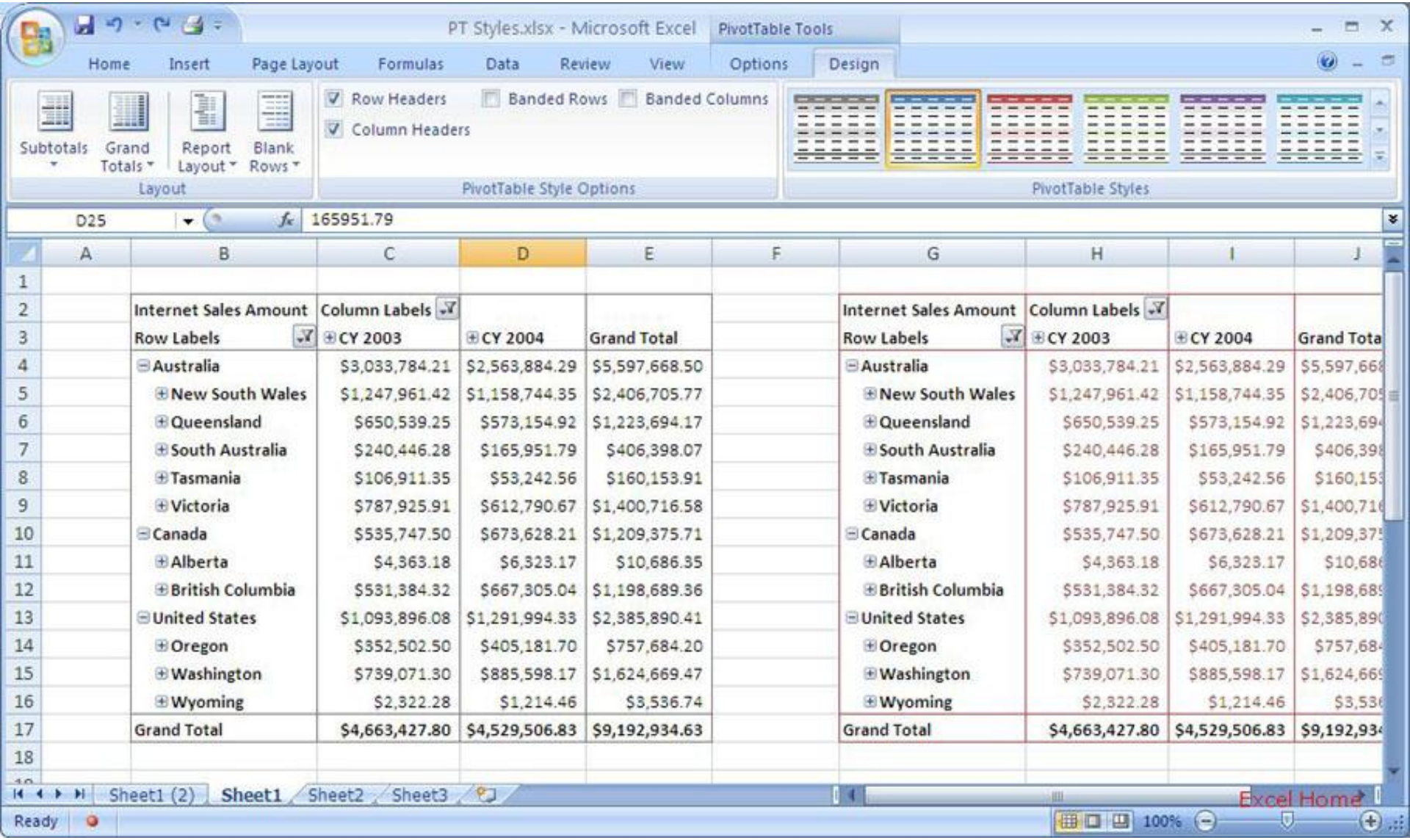
export to builder | options

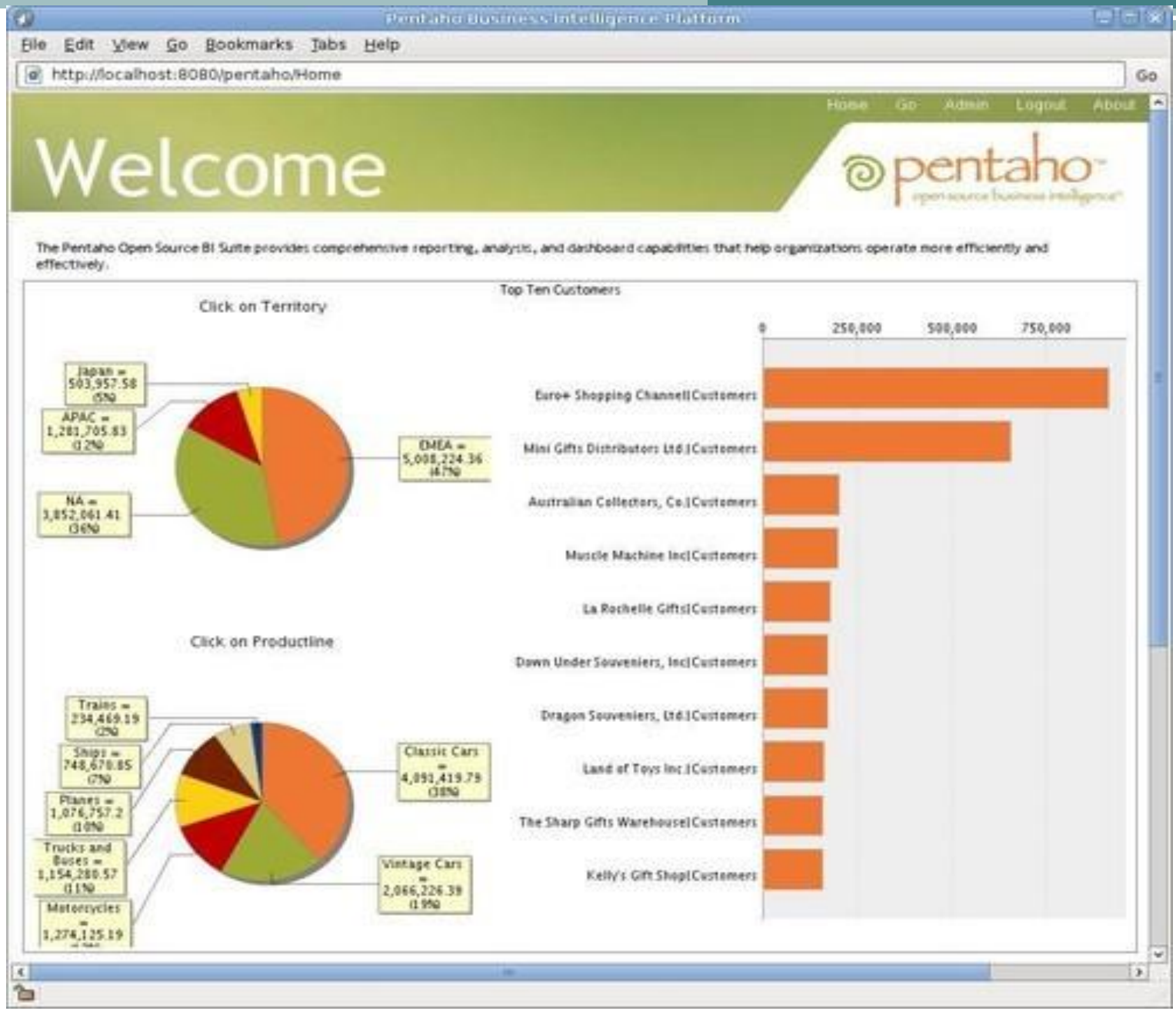



myesp - Enterprise Services Platform - Mozilla Firefox
Secure Access SSL VPN - 主页
myesp - Enterprise Services ...
tc-hadoop-test04.tc.baidu.com:8010
Zebra
HPCC Systems
EclWatch
Enterprise Services Platform
Clusters
Activity
Scheduler
ECL
Search Workunits
Browse Workunits
ECL Playground
Queries
Browse
Topology
Target Clusters
Cluster Processes
System Servers
DFU Workunits
Search
Browse
DFU Files
Upload/download File
View Data File
Search File Relationships
Browse Space Usage
Search Logical Files
Browse Logical Files
Browse Files by Scope
Spray Fixed
Spray CSV
Spray XML
Remote Copy
XRef
Roxie Queries
Search Roxie Queries
Search Roxie Files
View Roxie Files
Resources
Browse
My Account
My Account
Change Password
Users/Permissions
Users
Groups
Permissions
ECL Playground
Sample: ROLLUP
MyRec := RECORD
STRING1 Value1;
STRING1 Value2;
UNSIGNED1 Value3;
END;
SomeFile := DATASET([{'C','G',1},
{'C','C',2},
{'A','X',3},
{'B','G',4},
{'A','B',5}],MyRec);
SortedTable := SORT(SomeFile,Value1);
OUTPUT(SortedTable);
RECORDOF(SomeFile) RollThem(SomeFile L, MyRec R) := TRANSFORM
SELF.Value3 := IF(L.Value3 < R.Value3,L.Value3,R.Value3);
SELF.Value2 := IF(L.Value2 < R.Value2,L.Value2,R.Value2);
SELF := L;
END;
RolledUpRecs := ROLLUP(SortedTable,
LEFT.Value1 = RIGHT.Value1,
RollThem(LEFT,RIGHT));
OUTPUT(RolledUpRecs);
/*
Processes as:
LEFT vs. RIGHT
1 (AX3) 2 (AB5) - match, run transform, outpu
1 (AB3) 3 (BG4) - no match, output BG4
3 (BG4) 4 (CX1) - no match
Graph View
To enable graph views, please install the Graph View Control plugin:
Internet Explorer + Firefox (32bit)
Internet Explorer + Firefox (64bit)
Linux/Other (sources)
Submit
hthor
completed
value1 value2 value3
1 A X 3
2 A B 5
3 B G 4
4 C G 1
5 个项目中的 1 - 5
25 | 50 | 100 | 全部
1
Result 1
Result 2

数据分析系统搭建 - 小系统

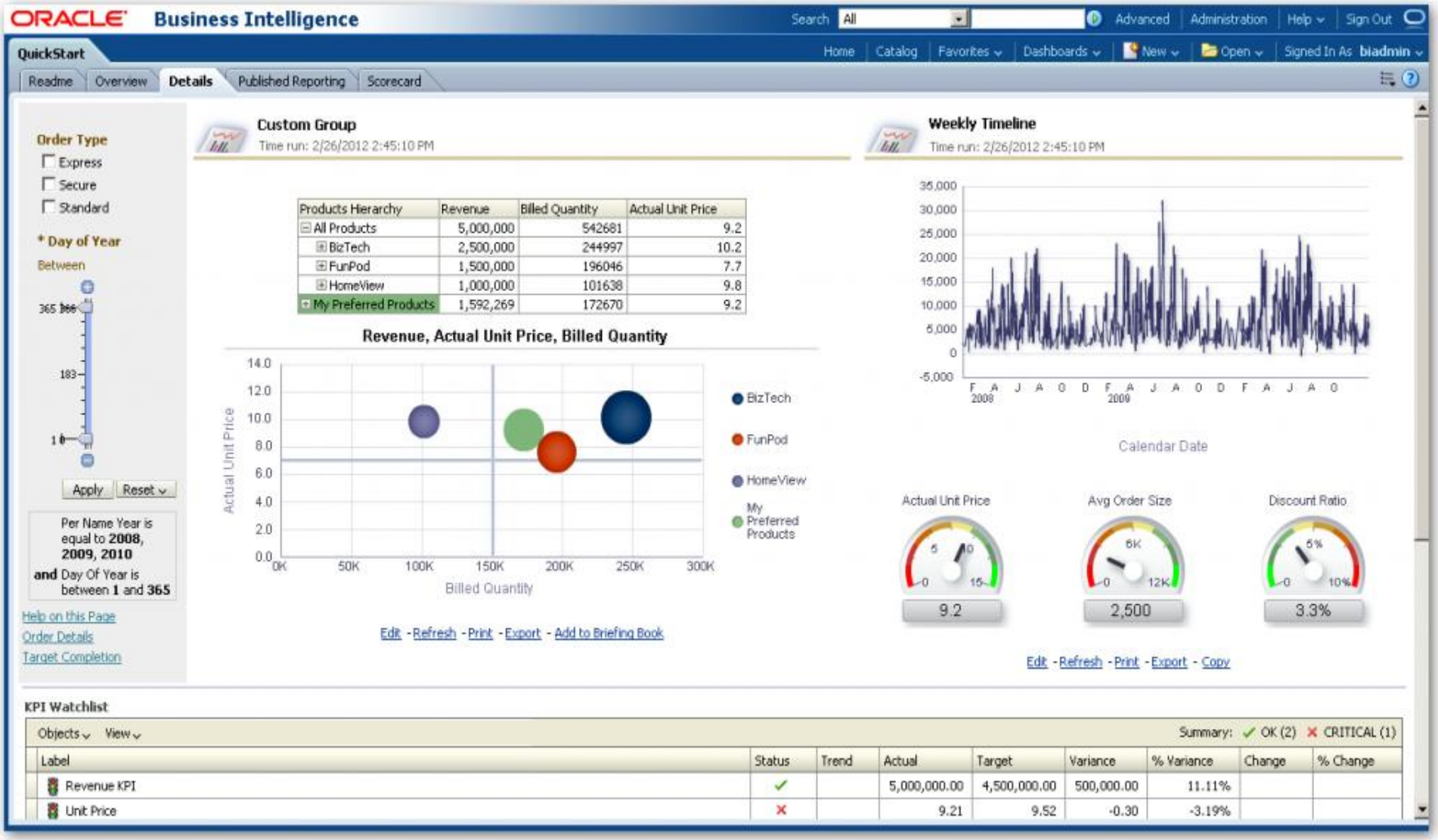



PT Styles.xlsx - Microsoft Excel
PivotTable Tools
Home
Insert
Page Layout
Formulas
Data
Review
View
Options
Design
Subtotals
Grand Totals
Report Layout
Blank Rows
Layout
Row Headers
Banded Rows
Banded Columns
Column Headers
PivotTable Style Options
PivotTable Styles
D25
165951.79
Internet Sales Amount
Column Labels
Row Labels
CY 2003
CY 2004
Grand Total
Australia $3,033,784.21 $2,563,884.29 $5,597,668.50
New South Wales $1,247,961.42 $1,158,744.35 $2,406,705.77
Queensland $650,539.25 $573,154.92 $1,223,694.17
South Australia $240,446.28 $165,951.79 $406,398.07
Tasmania $106,911.35 $53,242.56 $160,153.91
Victoria $787,925.91 $612,790.67 $1,400,716.58
Canada $535,747.50 $673,628.21 $1,209,375.71
Alberta $4,363.18 $6,323.17 $10,686.35
British Columbia $531,384.32 $667,305.04 $1,198,689.36
United States $1,093,896.08 $1,291,994.33 $2,385,890.41
Oregon $352,502.50 $405,181.70 $757,684.20
Washington $739,071.30 $885,598.17 $1,624,669.47
Wyoming $2,322.28 $1,214.46 $3,536.74
Grand Total $4,663,427.80 $4,529,506.83 $9,192,934.63
Sheet1 (2)
Sheet1
Sheet2
Sheet3
Excel Home
Ready
100%


Pentaho Business Intelligence Platform
File Edit View Go Bookmarks Tabs Help
http://localhost:8080/pentaho/Home
Go
Home Go Admin Logout About
Welcome
pentaho
open source business intelligence
The Pentaho Open Source BI Suite provides comprehensive reporting, analysis, and dashboard capabilities that help organizations operate more efficiently and effectively.
Click on Territory
Japan = 503,957.58 (5%)
APAC = 1,281,705.83 (12%)
NA = 3,852,061.41 (36%)
EMEA = 5,008,224.36 (47%)
Click on Productline
Trains = 234,469.19 (2%)
Ships = 748,670.85 (7%)
Planes = 1,076,757.2 (10%)
Trucks and Buses = 1,154,280.57 (11%)
Motorcycles = 1,274,125.19
Classic Cars = 4,091,419.79 (38%)
Vintage Cars = 2,066,226.39 (19%)
Top Ten Customers
0
250,000
500,000
750,000
Euro+ Shopping Channel|Customers
Mini Gifts Distributors Ltd.|Customers
Australian Collectors, Co.|Customers
Muscle Machine Inc|Customers
La Rochelle Gifts|Customers
Down Under Souveniers, Inc|Customers
Dragon Souveniers, Ltd.|Customers
Land of Toys Inc.|Customers
The Sharp Gifts Warehouse|Customers
Kelly's Gift Shop|Customers

大数据分析系统搭建 - 商业版

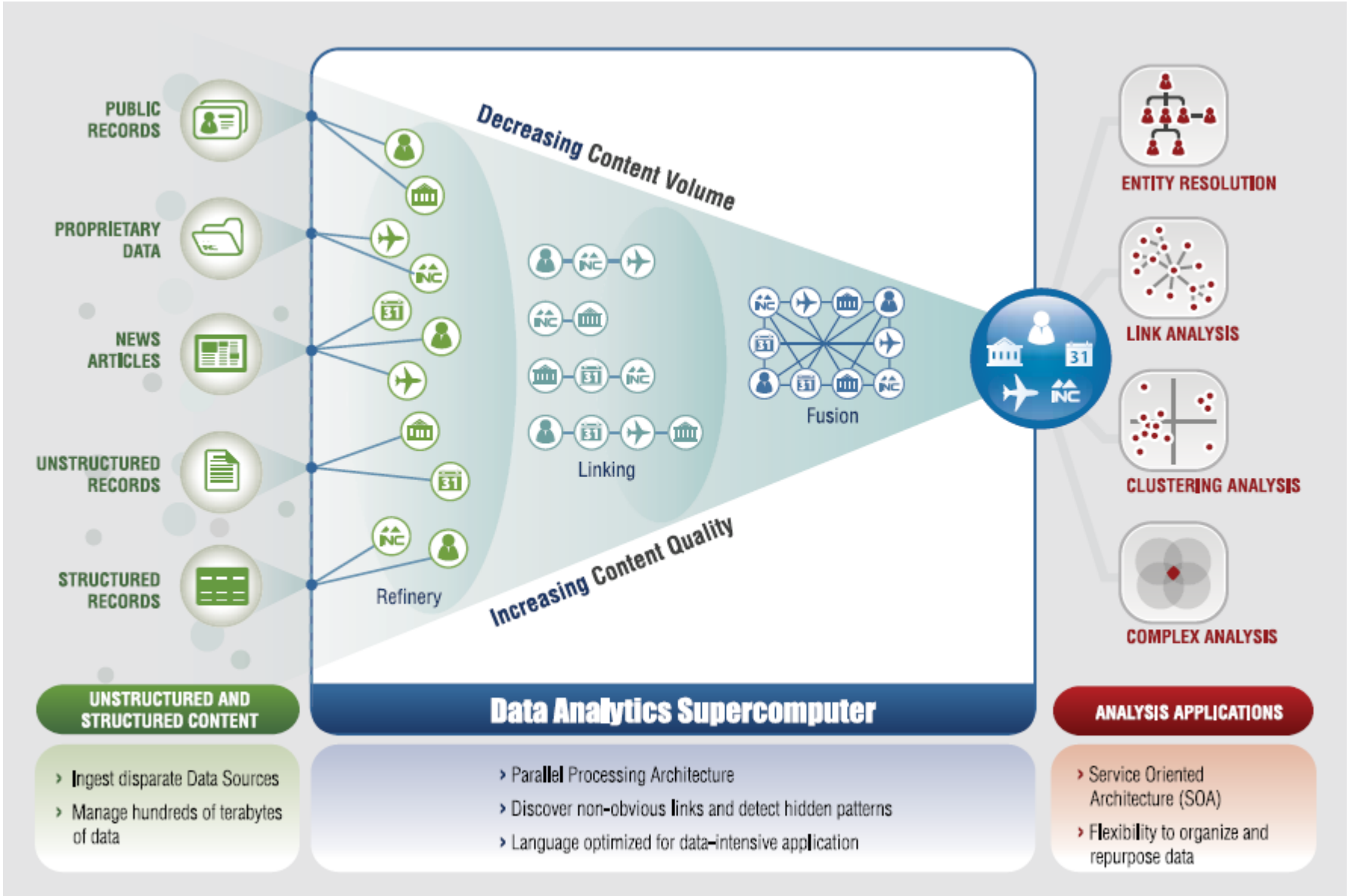
- Oracle BIEE + Oracle Exadata
- 其它产品
  - Greenplum
  - SAP HANA
  - Netteza
  - … …

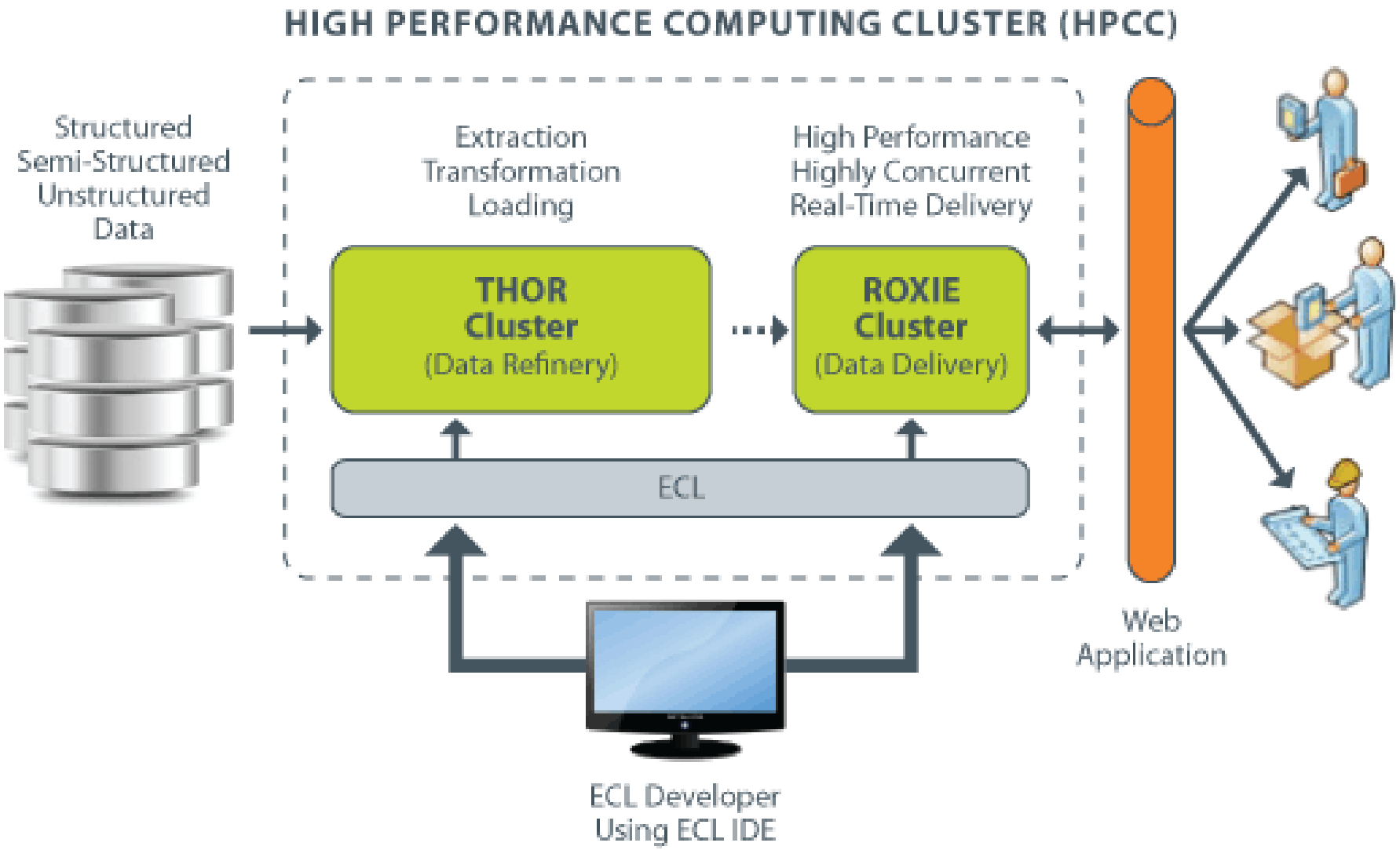

ORACLE Business Intelligence
Search All
Advanced
Administration
Help
Sign Out
QuickStart
Home
Catalog
Favorites
Dashboards
New
Open
Signed In As biadmin
Readme
Overview
Details
Published Reporting
Scorecard
Order Type
Express
Secure
Standard
* Day of Year
Between
365
183
1
Apply
Reset
Per Name Year is equal to 2008, 2009, 2010 and Day Of Year is between 1 and 365
Help on this Page
Order Details
Target Completion
Custom Group
Time run: 2/26/2012 2:45:10 PM
Products Hierarchy
Revenue
Billed Quantity
Actual Unit Price
All Products 5,000,000 542681 9.2
BizTech 2,500,000 244997 10.2
FunPod 1,500,000 196046 7.7
HomeView 1,000,000 101638 9.8
My Preferred Products 1,592,269 172670 9.2
Revenue, Actual Unit Price, Billed Quantity
Actual Unit Price
Billed Quantity
BizTech
FunPod
HomeView
My Preferred Products
Edit - Refresh - Print - Export - Add to Briefing Book
Weekly Timeline
Time run: 2/26/2012 2:45:10 PM
Calendar Date
Actual Unit Price
9.2
Avg Order Size
2,500
Discount Ratio
3.3%
Edit - Refresh - Print - Export - Copy
KPI Watchlist
Objects
View
Summary: OK (2) CRITICAL (1)
Label
Status
Trend
Actual
Target
Variance
% Variance
Change
% Change
Revenue KPI 5,000,000.00 4,500,000.00 500,000.00 11.11%
Unit Price 9.21 9.52 -0.30 -3.19%

大数据分析系统搭建 - 开源版

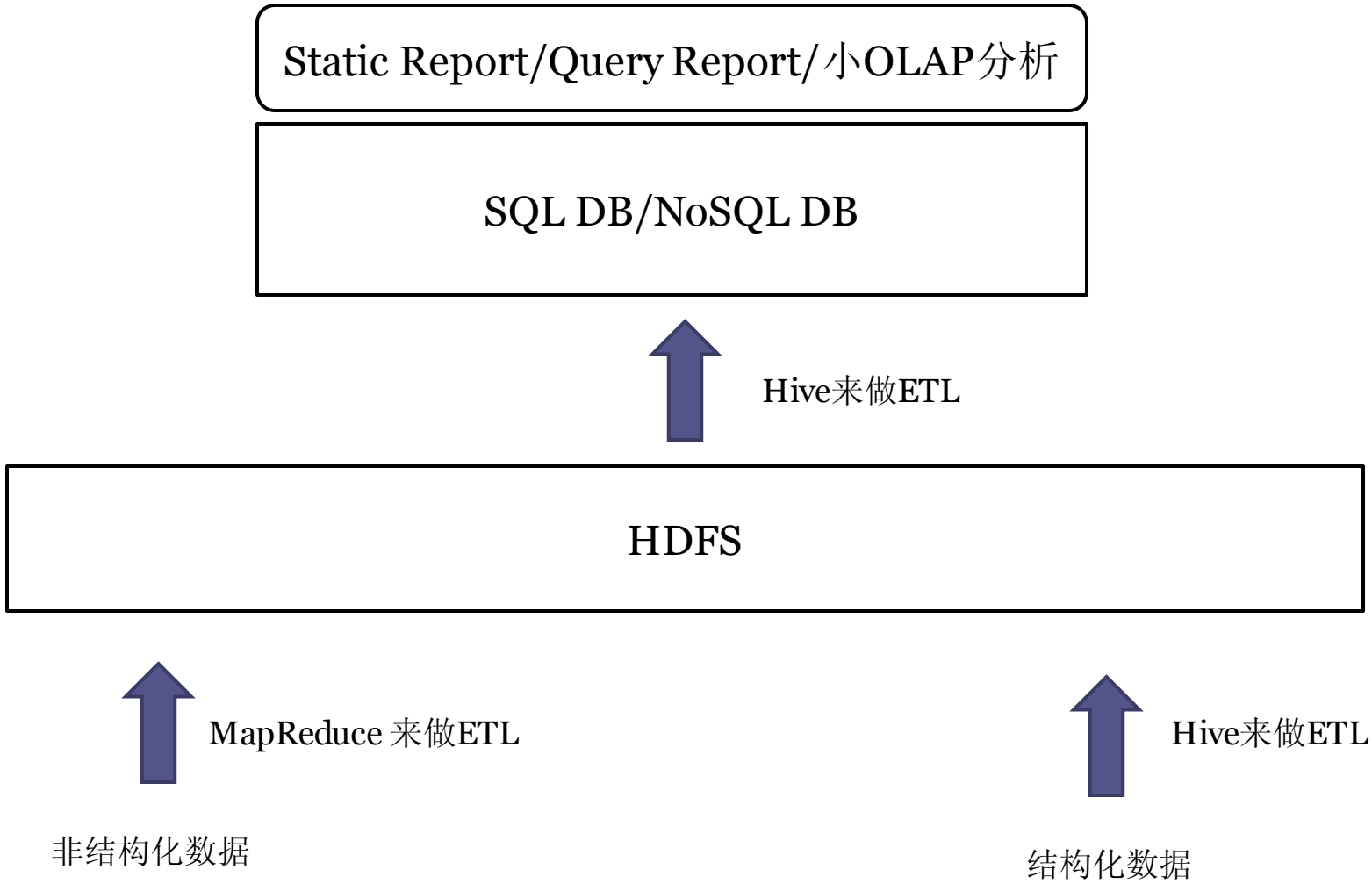
- 开源还没有很成熟的产品来构建大数据下的OLAP
- 短期解决
  - 商业产品
  - 交互性强，访问量大：转为查询请求放入SQL或NoSQL中查询
  - 交互性要求不高，访问量少的: 转为利用Hive/Impala来做

中等规模分析方案 – HPCC Systems




HIGH PERFORMANCE COMPUTING CLUSTER (HPCC)
Structured
Semi-Structured
Unstructured
Data
Extraction
Transformation
Loading
High Performance
Highly Concurrent
Real-Time Delivery
THOR
Cluster
(Data Refinery)
ROXIE
Cluster
(Data Delivery)
ECL
Web
Application
ECL Developer
Using ECL IDE

大规模分析方案 - Hadoop



谢　　谢

Q&A