Retail bank: 400 times faster

Jun Xie,PhD

LACTEC Information

jun.xie@lactec.net http://www.lactec.net

Query Big data using NIVDIA GPU
A case study for retail banks

Jun Xie,PhD (jun.xie@lactec.net)

LACTEC Information Service

http://www.lactec.net

Retail Bank : 400 times faster

A Customer story of application of GPU to Data base acceleration for big data

1 What is the big data means to a bank
2 Solution NVIDA K40 and K80 has been used for rebuild data base query engine
3 Real customer case study: A Customer behavior study 400 times faster than DB2 on RS6000 590 B Loan Analysis study : 1000 times faster than DB2 ; C Application for telecom industry: 400 times faster than Oracle
4 Products help people cut their cost and speedup the service

Who we are

Lactec Information Service Ltd is a software company focus on solution for banking Industry, we provide products from core banking , loan mgt , data warehouse /CRM and Financial engineering application. We have development team of size of 150 currently , located in Shanghai and US

HPC is our current concerns,GPU application research is Hot.

Now we switch to our topics... What problems we have and what solution we would like to share

THE

Problems and solution

02
03
04
Application

The pain if the data is big
How big the data

Banks have big and big data

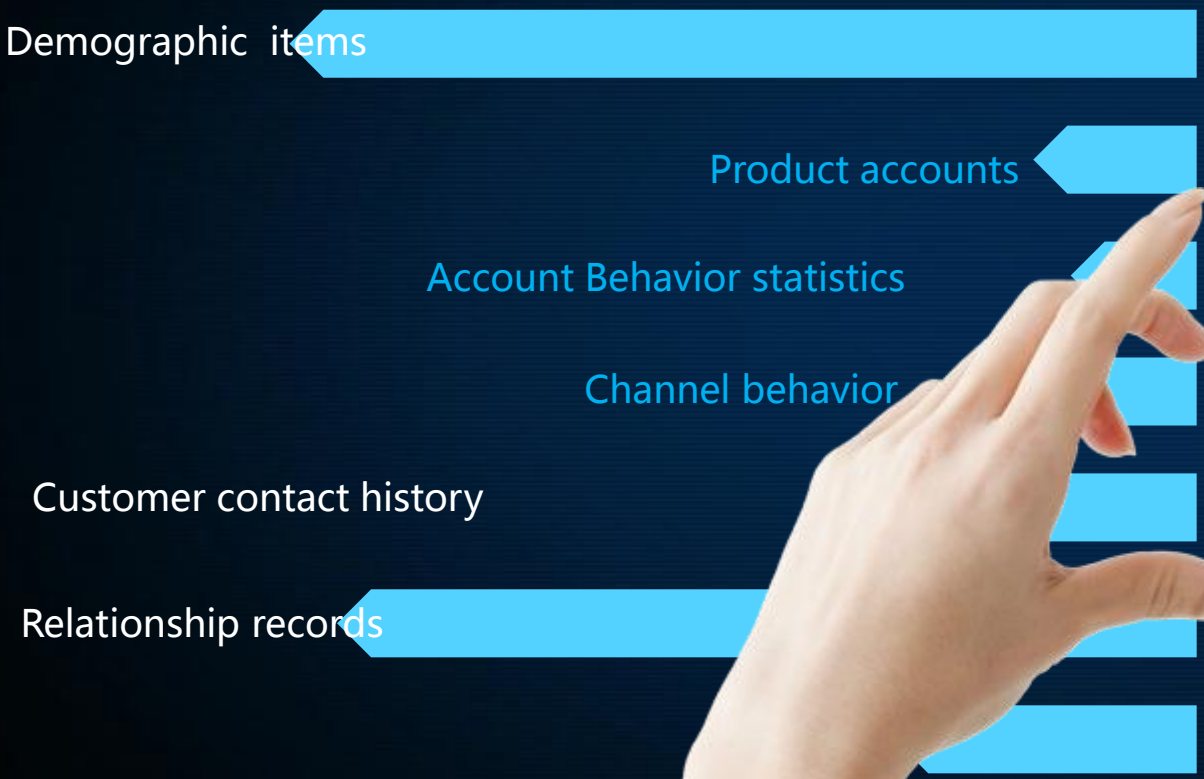A typical large bank (China has at least 300)

Have 30 M customers

50M accounts

1 Billion transaction a year (web querying transactions not included)

And even more …

The data a bank need to run the business?

| 100 | 90 | 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 | | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

Demographic items

A — Statistics index for a customer managers

Product accounts

B — Statistics index for a team

Account Behavior statistics

C — Market information

Channel behavior

D — Competive info

Customer contact history

E — Opportunity and thread

Relationship records

F — Risk management

G — Others

Customer view
~1500 columns

Business process and mgt

Last 20 years IBM, Oracle and Teradata have been educating customers to build a data warehouse, how
Life is harder when data accumulated, people in banking industry never have the ability to benefit from
their value of their big data except some statistics or reports

The Query is so slow

Simply querying a customer whole view may need 10 min, no say
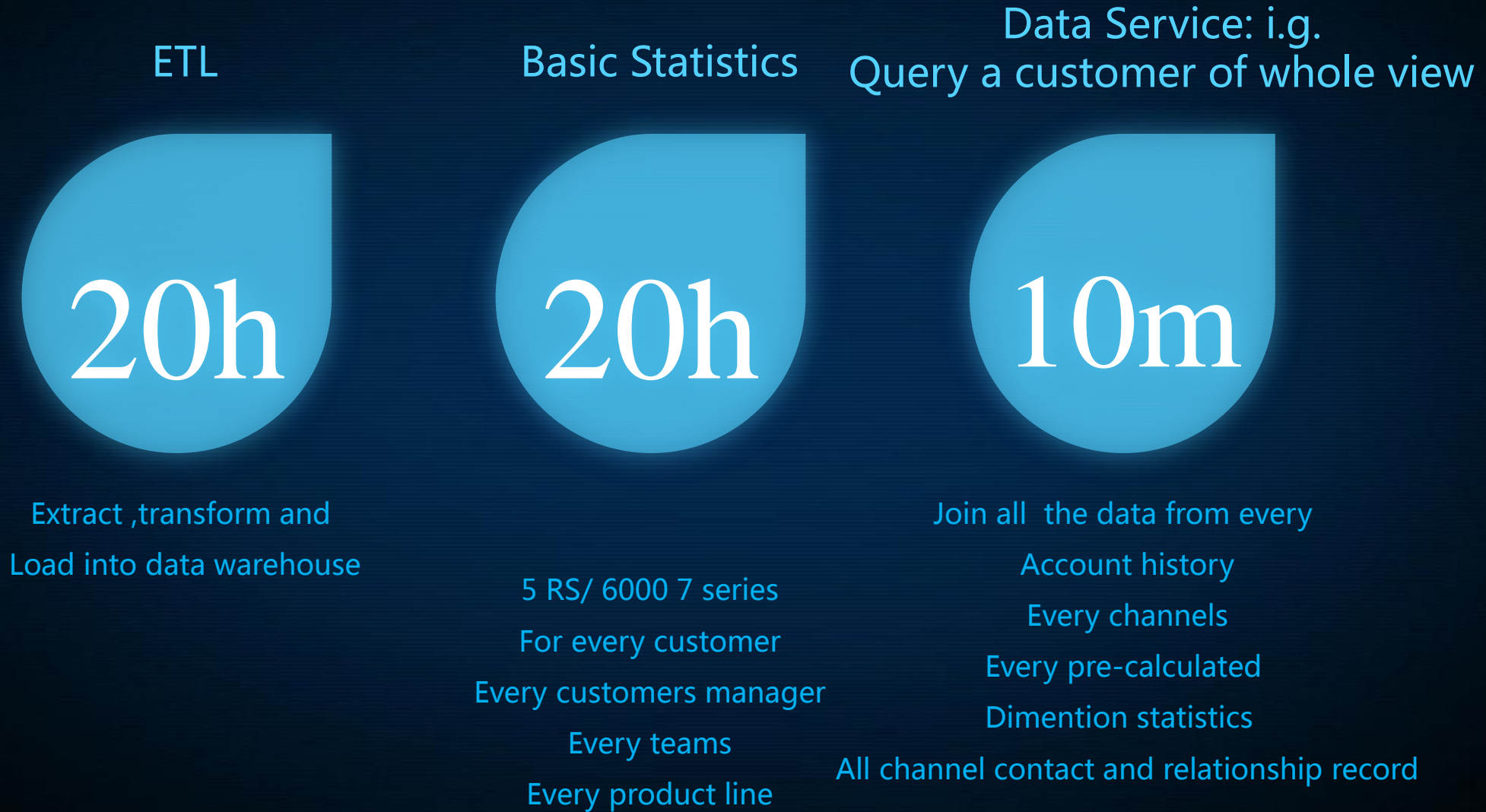customer behavior segmentation and pattern reorganization, etc

The cost is higher and higher

Our customer hate but people sell hardware and storage are happy。

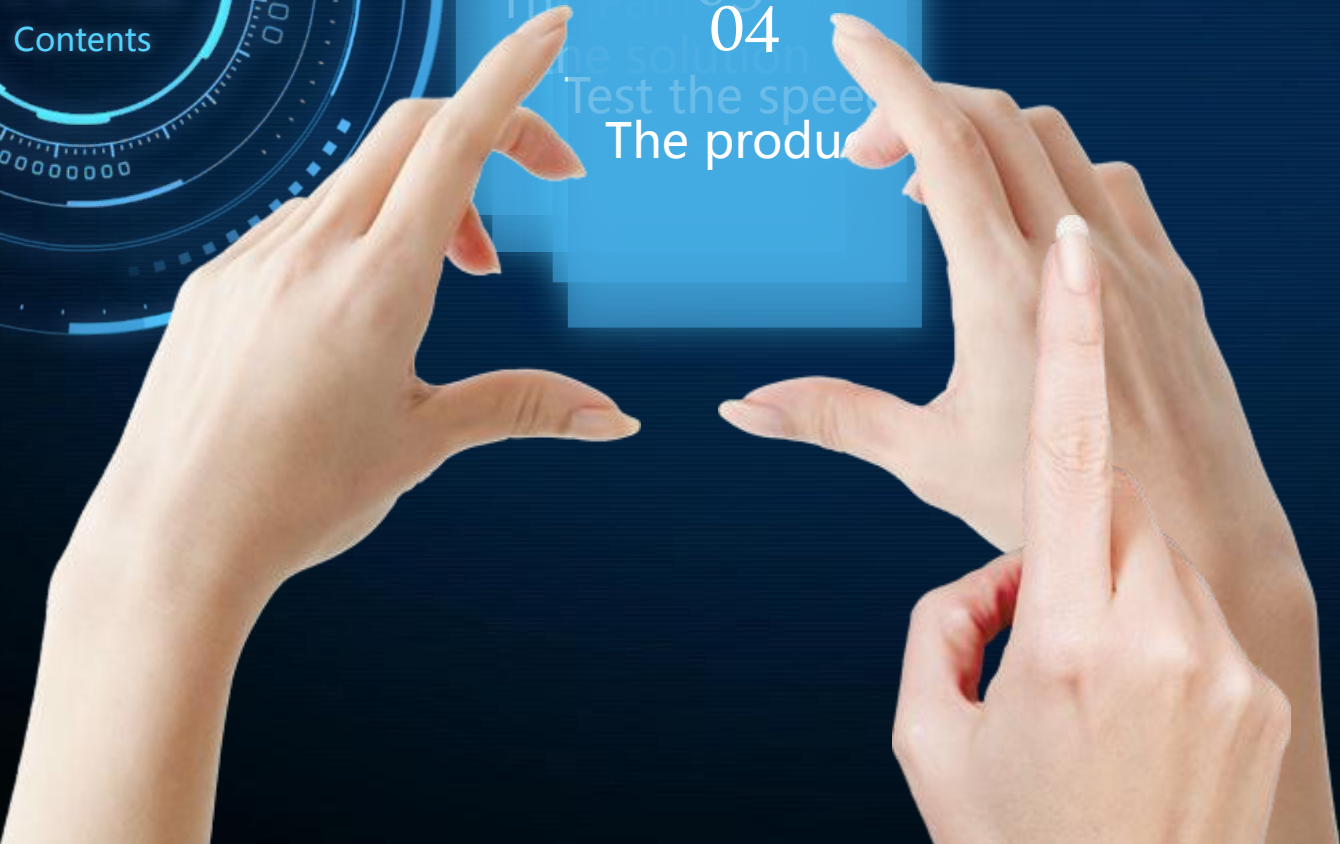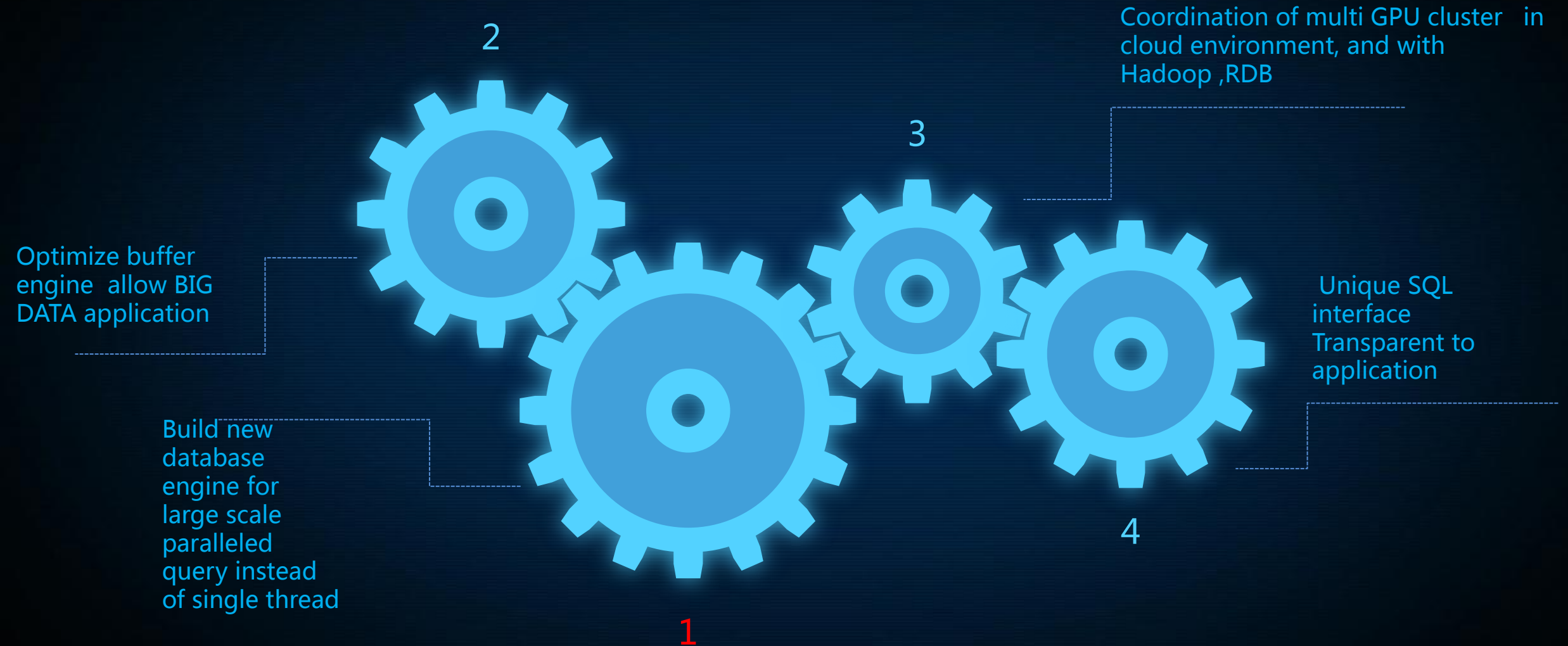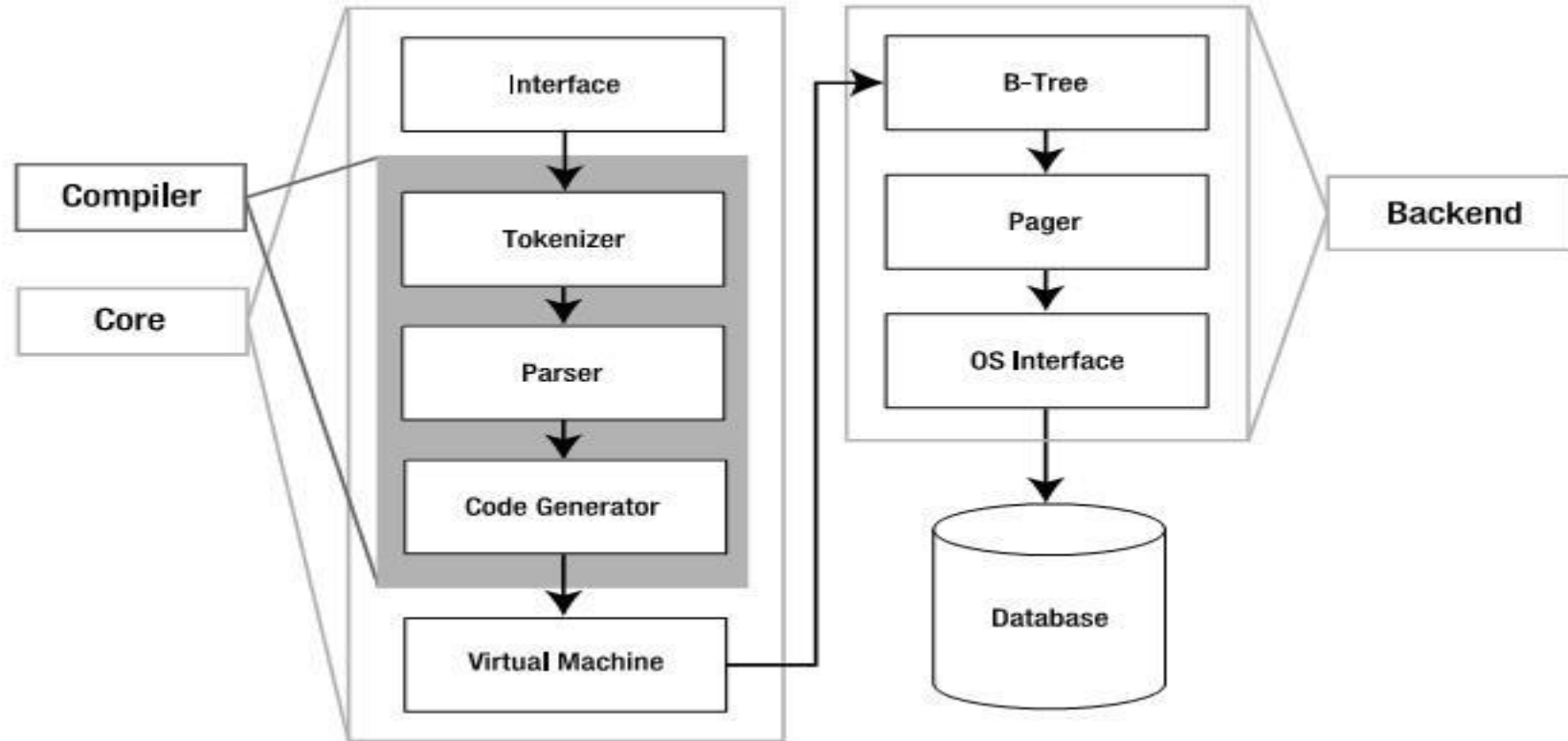# FACT: Every thing is slow even buy more expensive IBM RS6000 and DB2 new vers

ETL

Basic Statistics

Data Service: i.g.
Query a customer of whole view

20h

20h

10m

Extract ,transform and

Load into data warehouse

5 RS/ 6000 7 series

For every customer

Every customers manager

Every teams

Every product line

Join all the data from every

Account history

Every channels

Every pre-calculated

Dimention statistics

All channel contact and relationship record

Solution Summary
Rebuild data base query engine with NVIDA GPU technology and products

2

3

Coordination of multi GPU cluster in cloud environment, and with Hadoop ,RDB

Optimize buffer engine allow BIG DATA application

Unique SQL interface Transparent to application

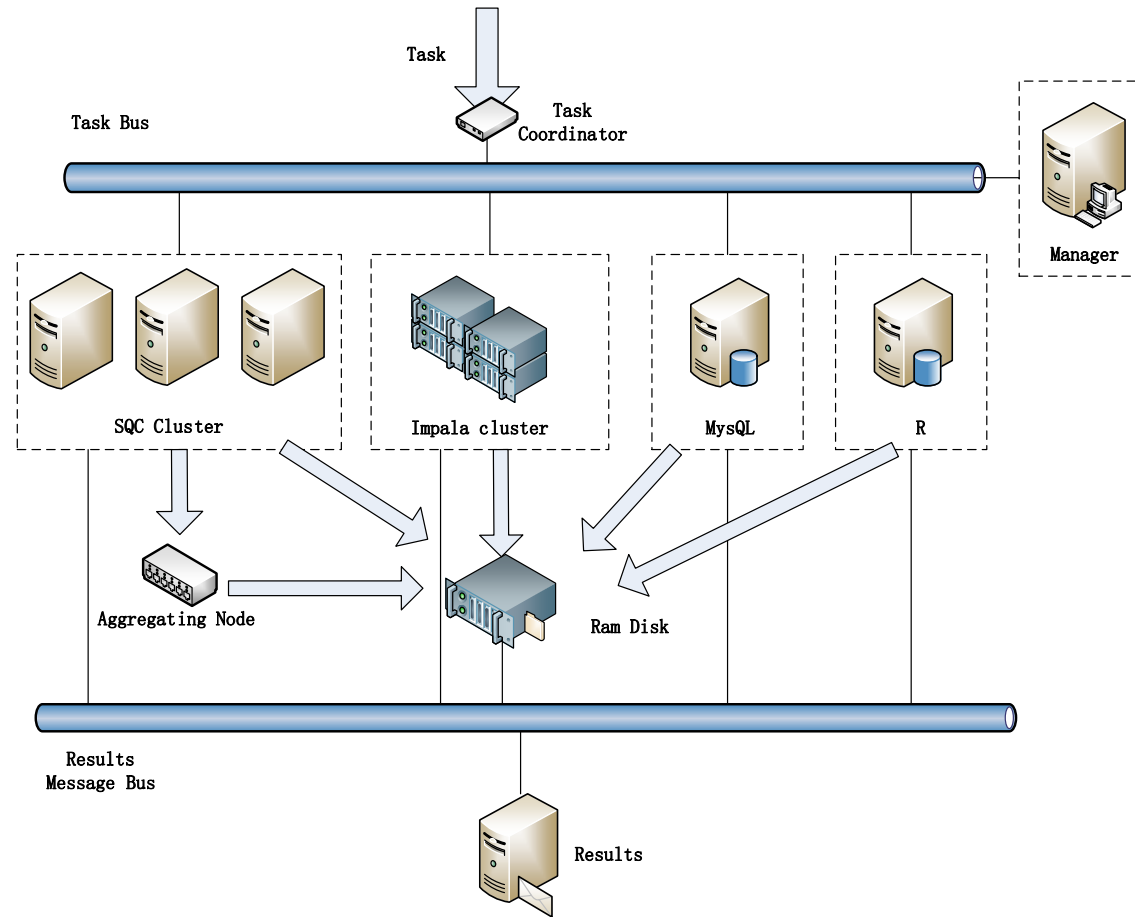Build new database engine for large scale paralleled query instead of single thread

4

1

Princple of the new query engine

- directs and executes query statements in GPUs

Cluster of GPUs,Hadoop and RDB

THE

Contents

01 02 03 04

The Produ

03

K40 and K80 was used to test the new data base engine

CASE1
300 M rows , 8 cols "long" table
CASE2
0.8M rows ,1076 cols "wide" table

With comparison to DB2 on RS6000/590
(8CPUs)

Acceleration Test
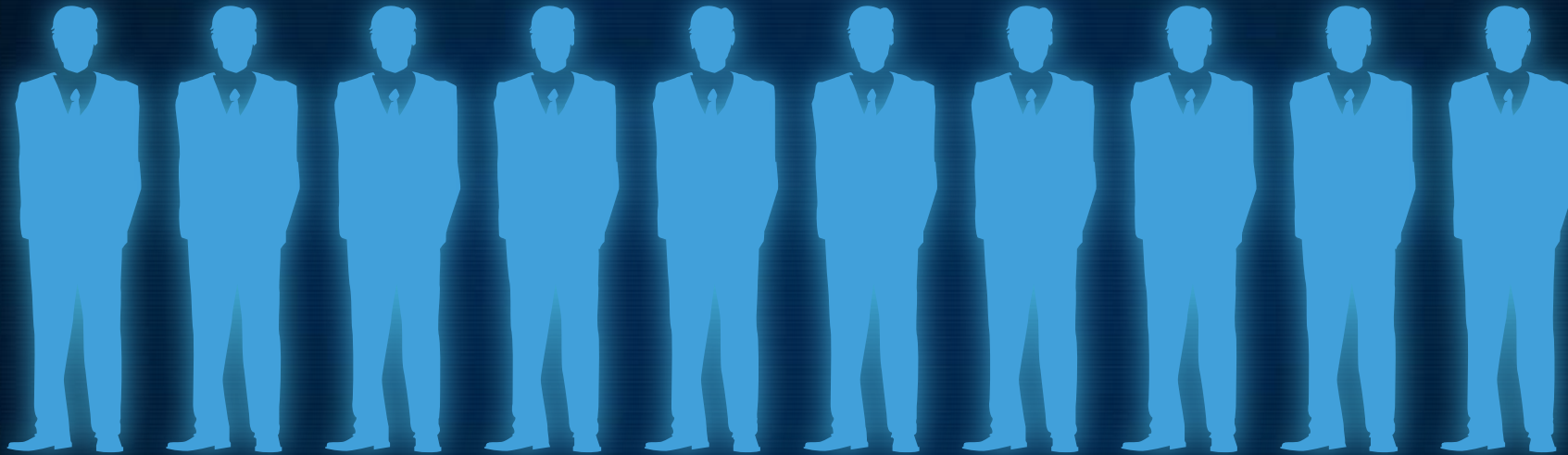## In real business application

CASE 1 LONG TABLE

Data mart of customer balance data set for 5 products
100 million rows by 5 columns
Query by means of RS/6000 590 cost 50-60 seconds, for example, find out all young women who shopping at downtown during evening, >$2000 a month
Query by GPU DB 0.1-0.2 seconnds

| 400 | 1 |

CASE 2 WIDE TABLE

Loan Mgt data of 840,000 rows，1075 cols，float and character mixed
1. Locate a customer
select custno,jbxx101,BRTCOD from C_MEASURE_ALL where custno = "0000000789 "
2. Find all customers of education level above undergraduated
select custno,jbxx101,BRTCOD,jbxx128_name from C_MEASURE_ALL where jbxx128_name like "%graduated% "
3. Find all customer with income range
select custno,jbxx101,BRTCOD,khckye441 from C_MEASURE_ALL where khckye441 > 1000 and khckye441 < 5000
4. Select all customer live in an area
select custno,jbxx101,BRTCOD,jbxx128_name, llxx103 from C_MEASURE_ALL where llxx103 like "%Villar% "
5. Sum up loan balance for all customer
select sum(khckye441) from C_MEASURE_ALL

1000   1

All queries tested in DB2 on RS6000 590 needs 10 min or more

K40 on a PC Server

SQL	Query time-k40（s）	Query time-k80(one device) (s)	Rows of results
1	0.043367	0.004363	2
2	0.087915	0.074285	706187
3	0.002414	0.002748	4222
4	0.059182	0.020931	10
5	0.005181	0.005439	1

DB2

您的内容打在这里，或者通过复制您的文本后，在此框中选择粘贴，并选择只保留文字。

K40

您的内容打在这里，或者通过复制您的文本后，在此框中选择粘贴，并选择只保留文字。

DB2

The max cols limited in 1000 columns, Ultra wide table cause effectiveness as the buffer structure failure , even locate a record needs 10 min ,

K40

您的内容打在这里，或者通过复制您的文本后，在此框中选择粘贴，并选择只保留文字。

The men's final

THE

Contents

02

03

04

The Product

Test in

Application

The Cost is 10%

Investment of hardware (servers, storage)and service。

The speed of data querying service is 400-1000 times

您的内容打在这里，或者通过复制您的文本后，在此框中选择粘贴，并选择只保留文字。您的内容打在这里，或者通过复制您的文本后您的内容打在这里。

The Products

Super Query Component: Single GPU Database software

DataTurbines : Cluster of GPU with unique interface of SQL with all in one option

Application API for integration of Hadoop, RDB(DB2,Oracle) and R

Thanks for coming

Happy good days in GTC 2015

GPU powers data base application , speed up more than 400 times ,evens 1000 times for Big Data. LACTEC welcome cooperation in technology or marketing activities

TEL : 0086 4008205021 **EMAIL** : jun.xie@lactec.net http://www.lactec.net