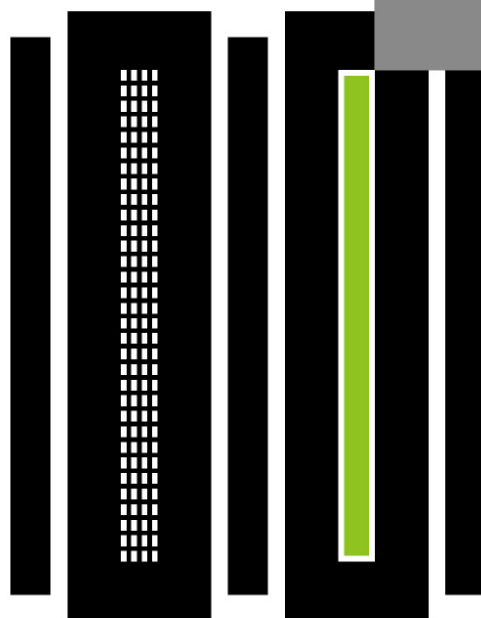
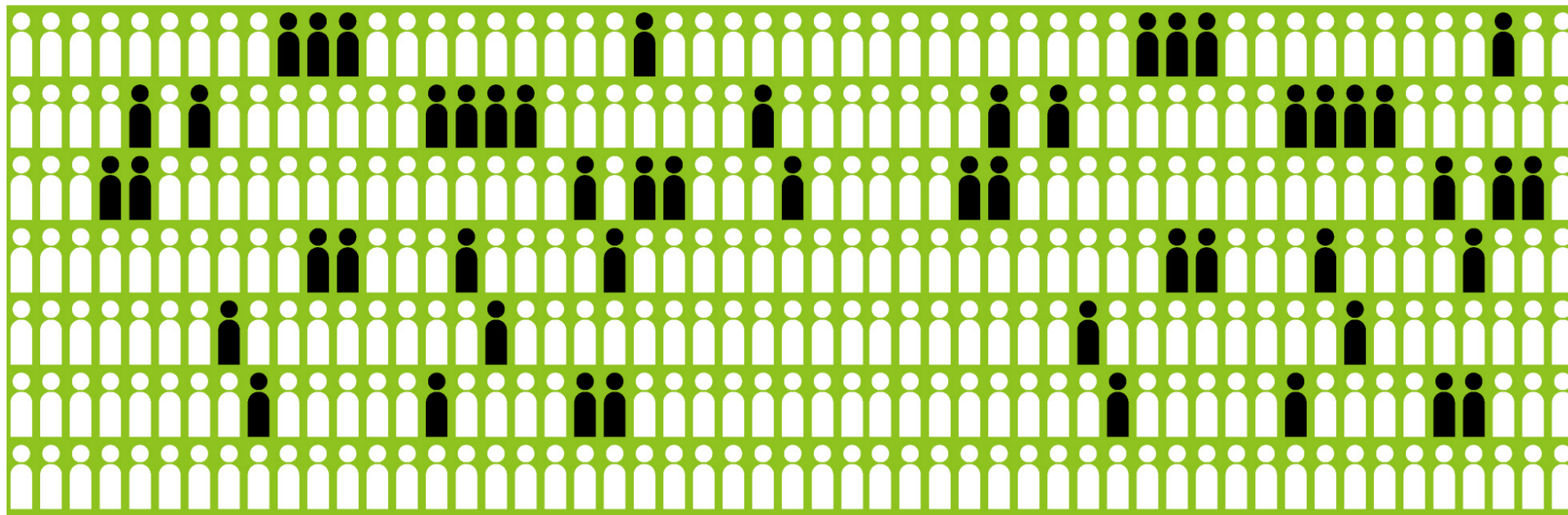


分析为道 **Z** 者见智
IBM主机商业分析(BA)高峰论坛




主机BI/DW及ISA0的业务价值

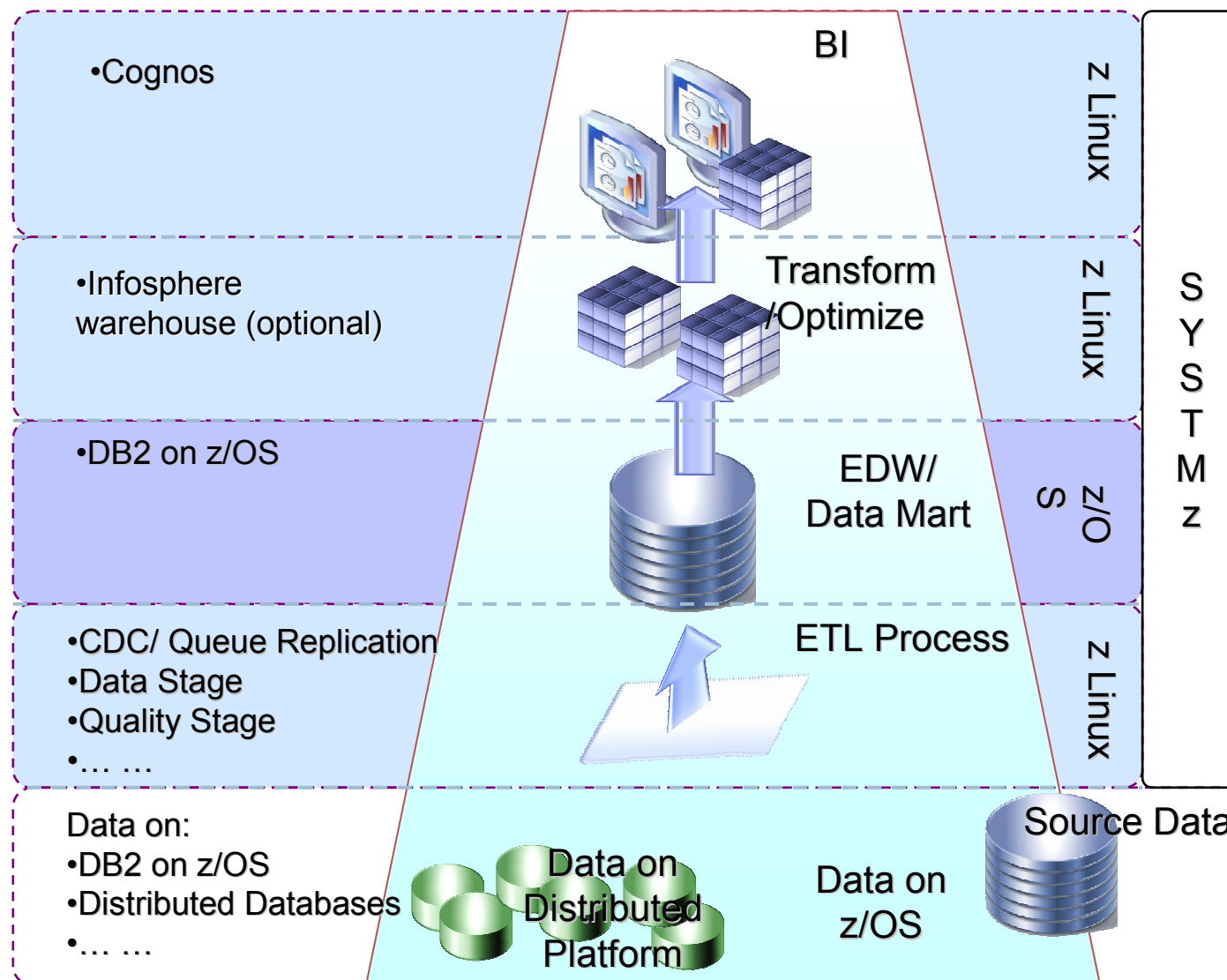
蒋渝



议程

- 分析类应用基于主机的解决方案 
- 查询优化工具：ISAO
 - IBM Smart Analytics Optimizer (ISAO)概要
 - 不同应用类型需要不同的数据存储格式
 - 在DB2 内定义适用于ISAO的查询
 - 在使用ISAO情况下，查询操作的执行过程

主机数据仓库解决方案概要



DB2 on z/OS 不断加强对DW 的支持

- Usability, Availability & Scalability
 - Online Partitioning Changes, More Partitions, Rotate
 - Schema Evolution:
 - VSCR with System z 64-bit Architecture
- Data Warehousing Support
 - Star Join Improvements
 - Materialized Query Tables
- Overall System & Query Performance
 - Locking Improvements
 - Multi-row INSERT & FETCH
 - More index Access for VARCHAR, data type mismatch, ...
 - DDF Performance Improvements
 - zIIP for improved cost of ownership

DB2 V8

- Usability, Availability & Scalability
 - Universal & Partition by Growth Table Space
 - Clone Tables (Online Load Replace)
 - XML
- Data Warehousing Support
 - Star Join Improvements (Dynamic Index ANDing)
 - SQL Enhancements (RANK, DENSE_RANK, Text, XML, INTERSECT, ...)
- Overall System & Query Performance
 - TRUNCATE Statement for fast delete, MERGE for insert or update
 - Index Compression
 - Significant CPU Reduction for DB2 Utilities & many queries
 - Insert and update scaling

DB2 V9

ISAO

DB2 V9/V10

议程

- 分析类应用基于主机的解决方案
- 查询优化工具：ISAO
 - IBM Smart Analytics Optimizer (ISAO)概要
 - 不同应用类型需要不同的数据存储格式
 - 在DB2 内定义适用于ISAO的查询
 - 在使用ISAO情况下，查询操作的执行过程



IBM Smart Analytics Optimizer(ISAO)

Capitalizing on the best of relational and the best of columnar databases

什么是ISAO?

The IBM Smart Analytics Optimizer(ISAO)是IBM最新推出的针对分析类查询的优化器,在不改变原有应用程序的情况下, ISAO可以极大缩短分析类查询的响应时间,满足决策系统对性能的要求。

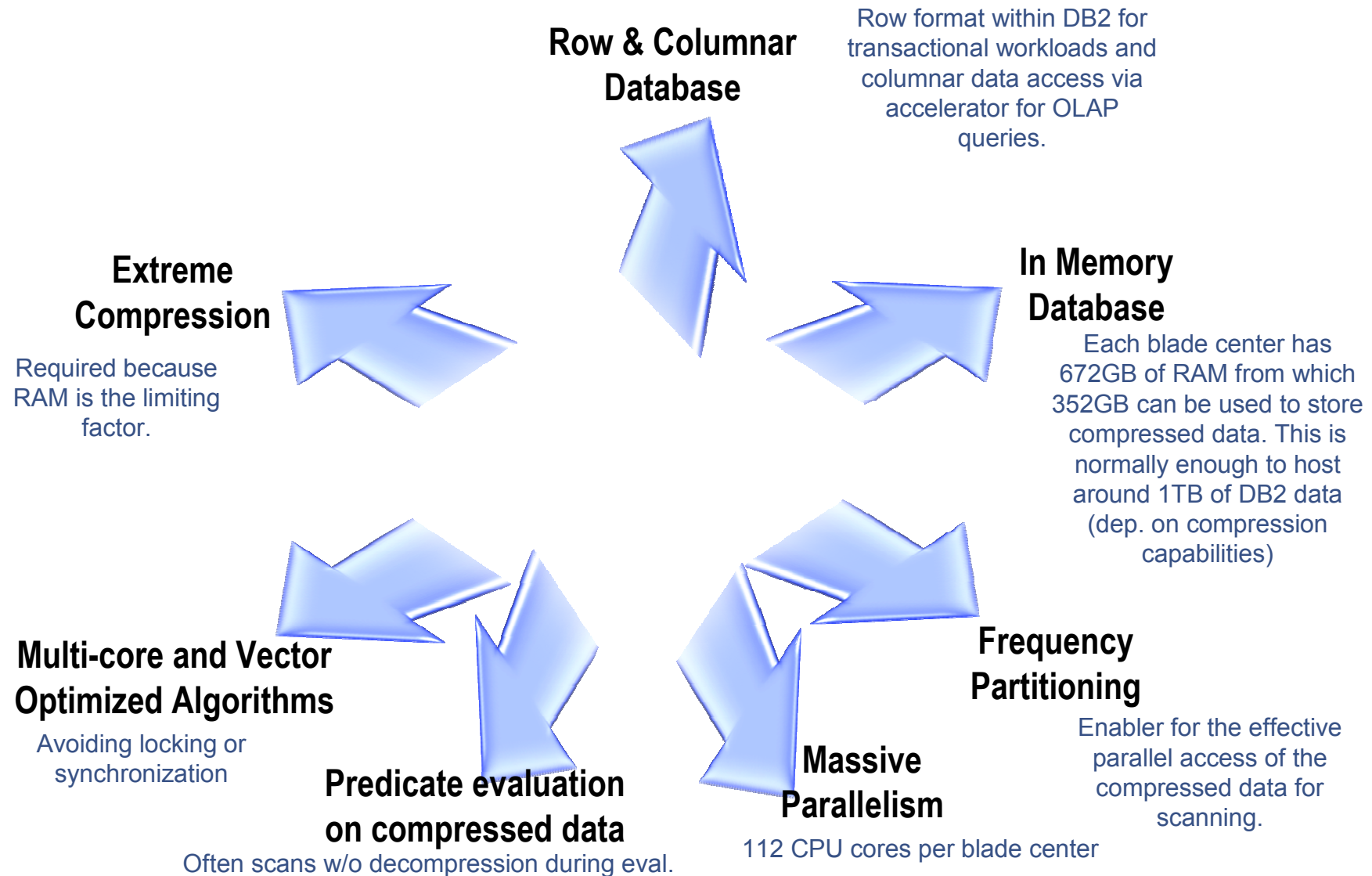


ISAO 的特点

- **性能:** 极大缩短查询的响应时间,使早期因为性能低下而不能投产的系统真正发挥作用
- **集成:** 和DB2 on z/OS无缝集成,访问数据库的应用程序无需改动就可以利用到加速器所提供的加速功能
- **自主优化能力:** DB2 on z/OS优化器可以保证查询以最优的方式执行,而不需要DBA的过多干预
- **对应用透明:** 应用程序无需改动就可以利用到ISAO提供的加速功能

Breakthrough Technology Enabling New Opportunities

在新技术的支持下ISAO性能有突破性的提高



IBM Smart Analytics Optimizer – 测试效果

Orders of magnitude faster for queries within Sweetspot⁺

下面类型的查询会因为ISAO而受益:

- 复杂的 OLAP 类查询
- 扫描大量的数据
- 趋势分析
- 商务决策



... and its acceleration factor :

Runtime of queries w/o ISAO

163 s

2311 s

25 s

1593 s

35 s

5435 s

Factor

48

511

12

206

4

1424

with ISAO

3s

5s

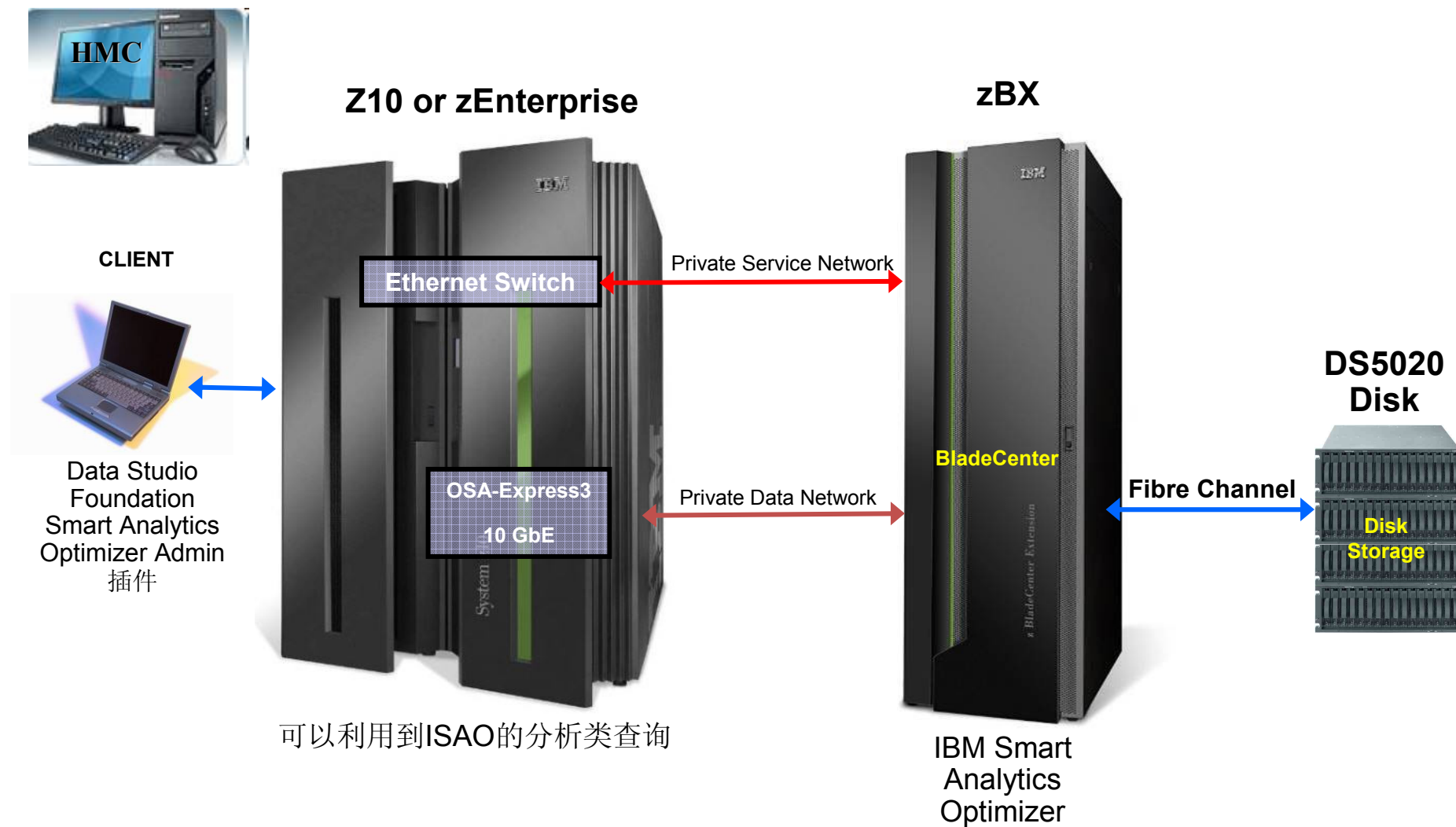
2s

8s

8s

4s

IBM Smart Analytics Optimizer 模块组成



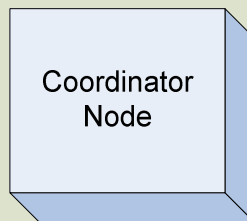
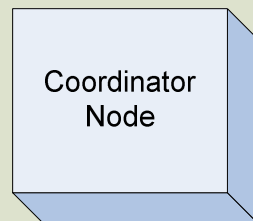
ISAO每个模块的作用



Query parsing and matching.

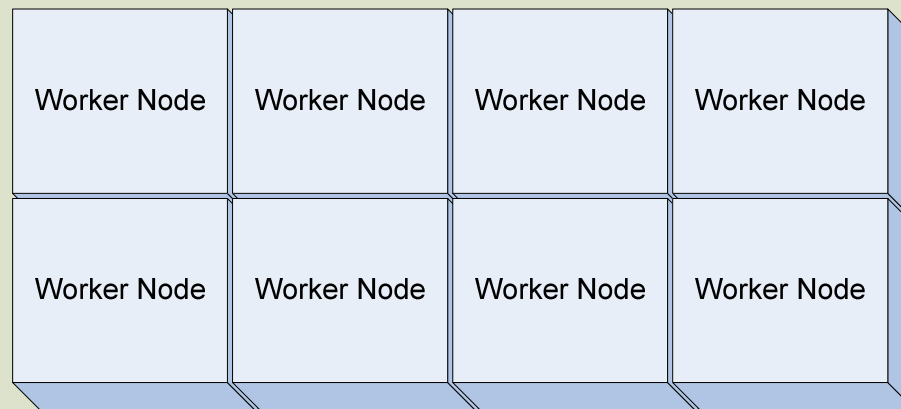
Is routing single query blocks to the optimizer.

Remains the „owner“ of all data.



Orchestrating the distributed tasks like LOAD or QUERY execution.

Communication peers to DB2

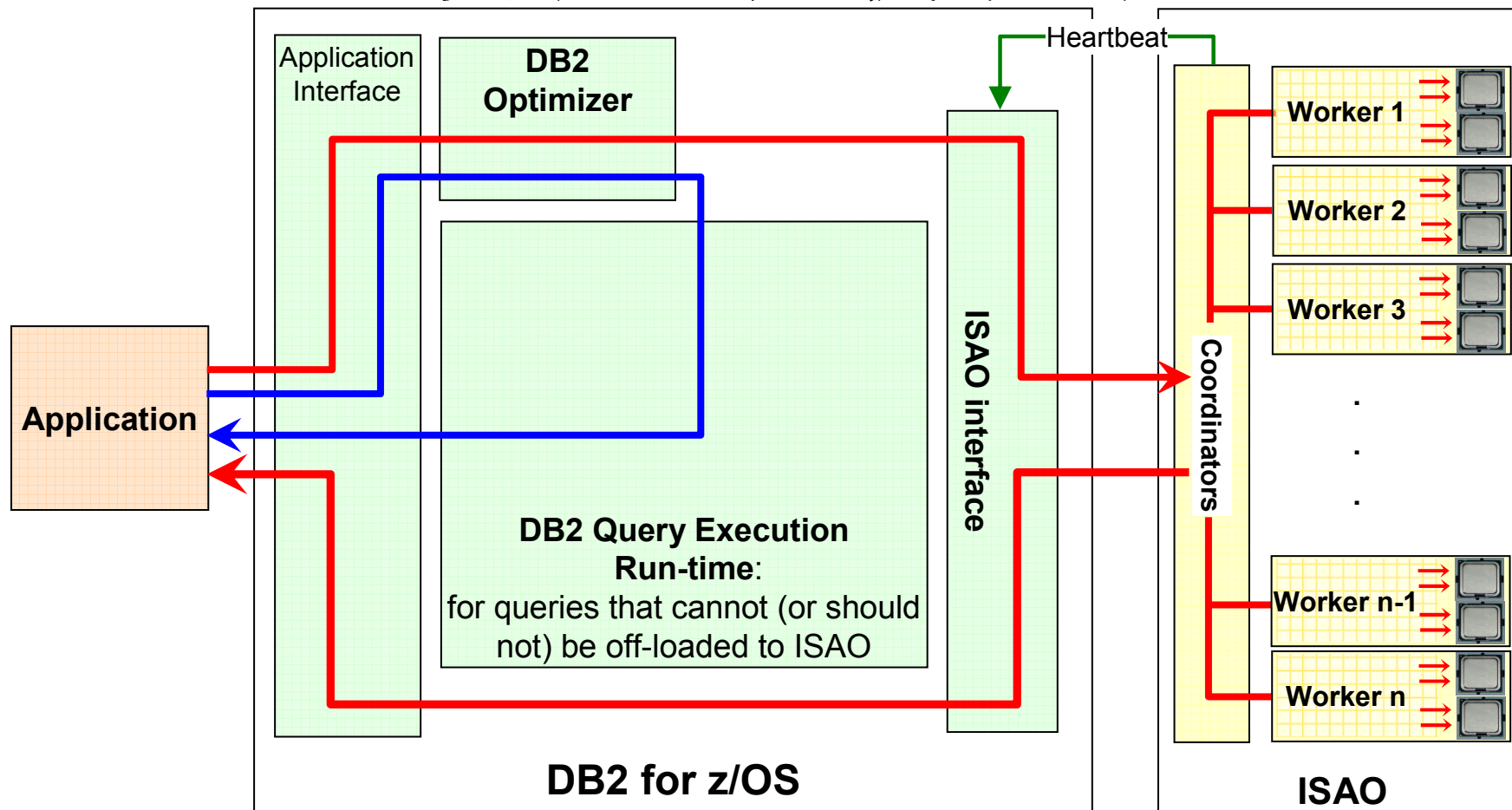


Have all the data in main memory (spread across all worker nodes).

Do the compression and query execution.

Always work together on a distributed task.

复杂查询的执行流程



总结： IBM Smart Analytics Optimizer 的特点

- 专用于查询加速的软硬件：通过网络连接的刀片系统
- 对应用程序透明
 - 应用程序继续连接到DB2 on z/OS.
 - DB2 on z/OS 优化器决定哪些查询可以使用到优化器，并由DB2 协调ISAO完成优化工作，整个过程不需要应用程序参与
 - 即使ISAO发生故障，DB2 on z/OS 也可以保证查询顺利完成
- 极大提高了查询性能
- 较少对DB2的调优需求 (MQTs, indexes, etc.)
- 极大提高了分析类应用的性价比：
 - 减少了主机CPU的负载
 - 使用到ISAO的查询在性能上得到很大提高
 - 较少了DBA的调优工作
- 混合的解决方案
 - 使Dynamic DW 和 Operational BI 成为可能
 - 保持了DB2 on z/os的服务水平
 - 同时支持交易类应用和分析类应用

议程

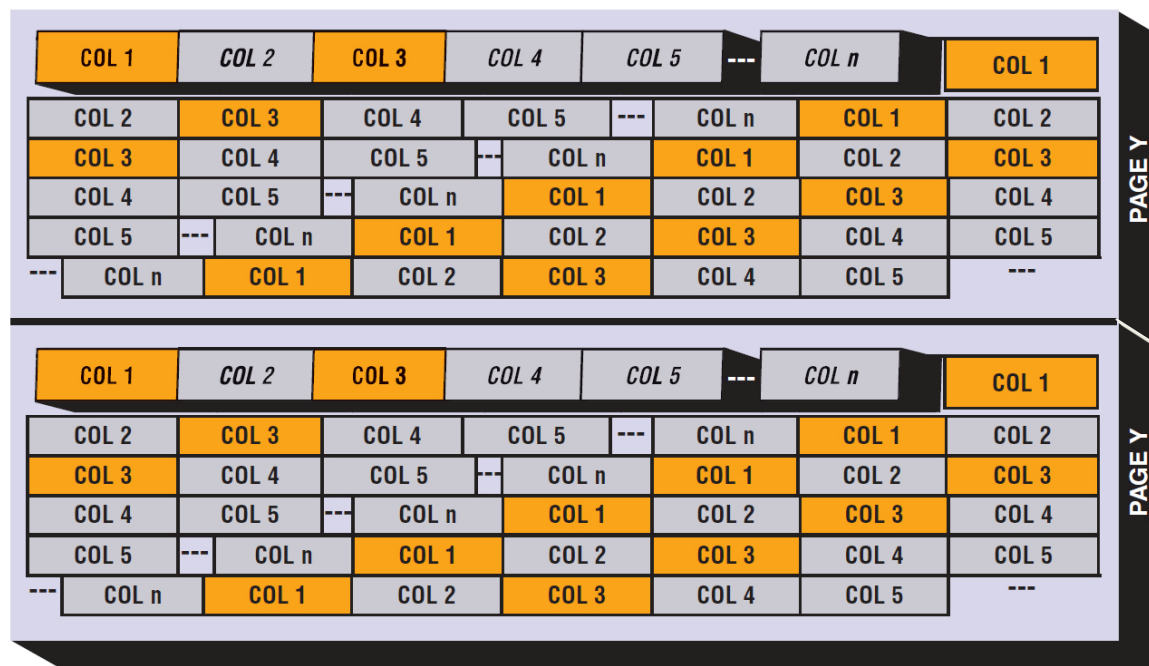
- 分析类应用基于主机的解决方案
- 查询优化工具：ISAO
 - IBM Smart Analytics Optimizer (ISAO)概要
 - 不同应用类型需要不同的数据存储格式
 - 在DB2 内定义适用于ISAO的查询
 - 在使用ISAO情况下，查询操作的执行过程



行存储

Each row stored sequentially

- 针对整条记录的 I/O 有很好性能
- 每次都需要获取并且解压缩每条记录
- 结果 –
 - 适合交易类应用
 - 不适合分析类应用



If only few columns are required the complete row is still fetched and uncompressed

列存储

Data is stored sequentially by column

- 数据按列存储并按列压缩:

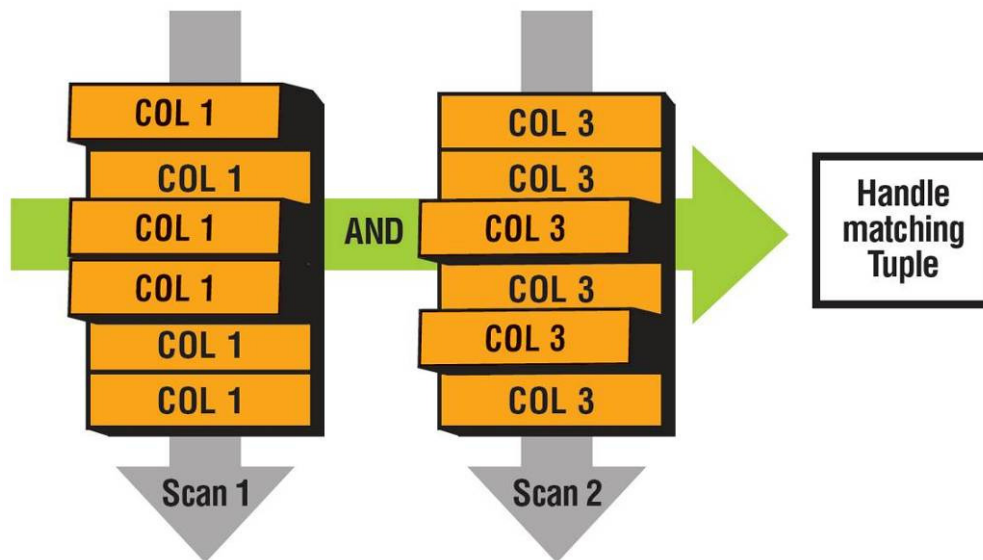
- 有利于列的顺序扫描
- 不利于随机查询

COL 1	COL 1	COL 1	COL 1	COL 1	COL 1
COL 1	COL 1	COL 1	COL 1	COL 2	COL 2
COL 2	COL 2	COL 2	COL 2	COL 2	COL 2
COL 2	COL 2	COL 3	COL 3	COL 3	COL 3
COL 3	COL 3	COL 3	COL 3	COL 3	COL 3
COL 4	COL 4	COL 4	COL 4	COL 4	COL 4
COL 4	COL 4	COL 4	COL 5	COL 5	COL 5
COL 5	COL 5	COL 5	COL 5	COL 5	COL 5
COL 5	COL 5	---	COL n	COL n	COL n
COL n	COL n	COL n	COL n	COL n	COL n
COL n					

If attributes are not required for a specific query execution, they are skipped completely.

列存储不适合于所有的场景

Individual record joins require many extra steps



多记录联合(join)效率不高

- Predicates processed separately
- Results from each column needs to be ANDed to determine a match
 - Significant additional processing

随机访问效率不高

Random record access is not performed well on pure columns stores.

在压缩的情况下处理数据 (Register-Store)

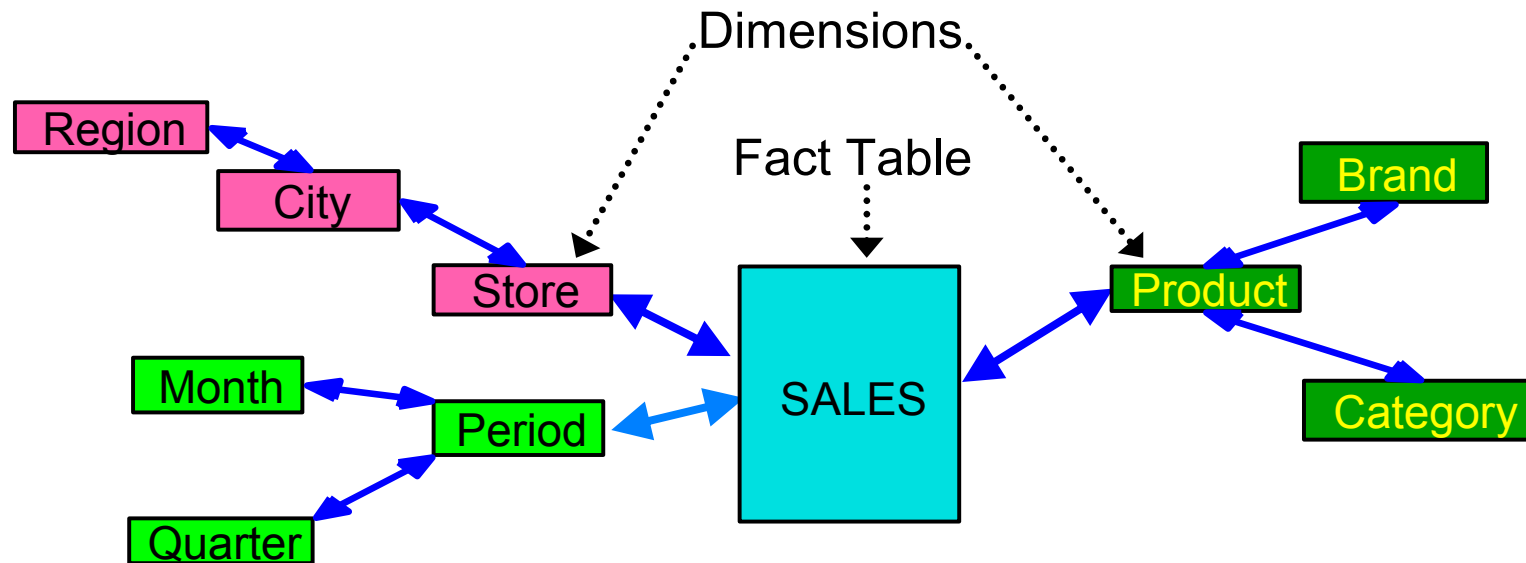
议程

- 分析类应用基于主机的解决方案
- 查询优化工具：ISAO
 - IBM Smart Analytics Optimizer (ISAO)概要
 - 不同应用类型需要不同的数据存储格式
 - 在DB2 内定义适用于ISAO的查询
 - 在使用ISAO情况下，查询操作的执行过程



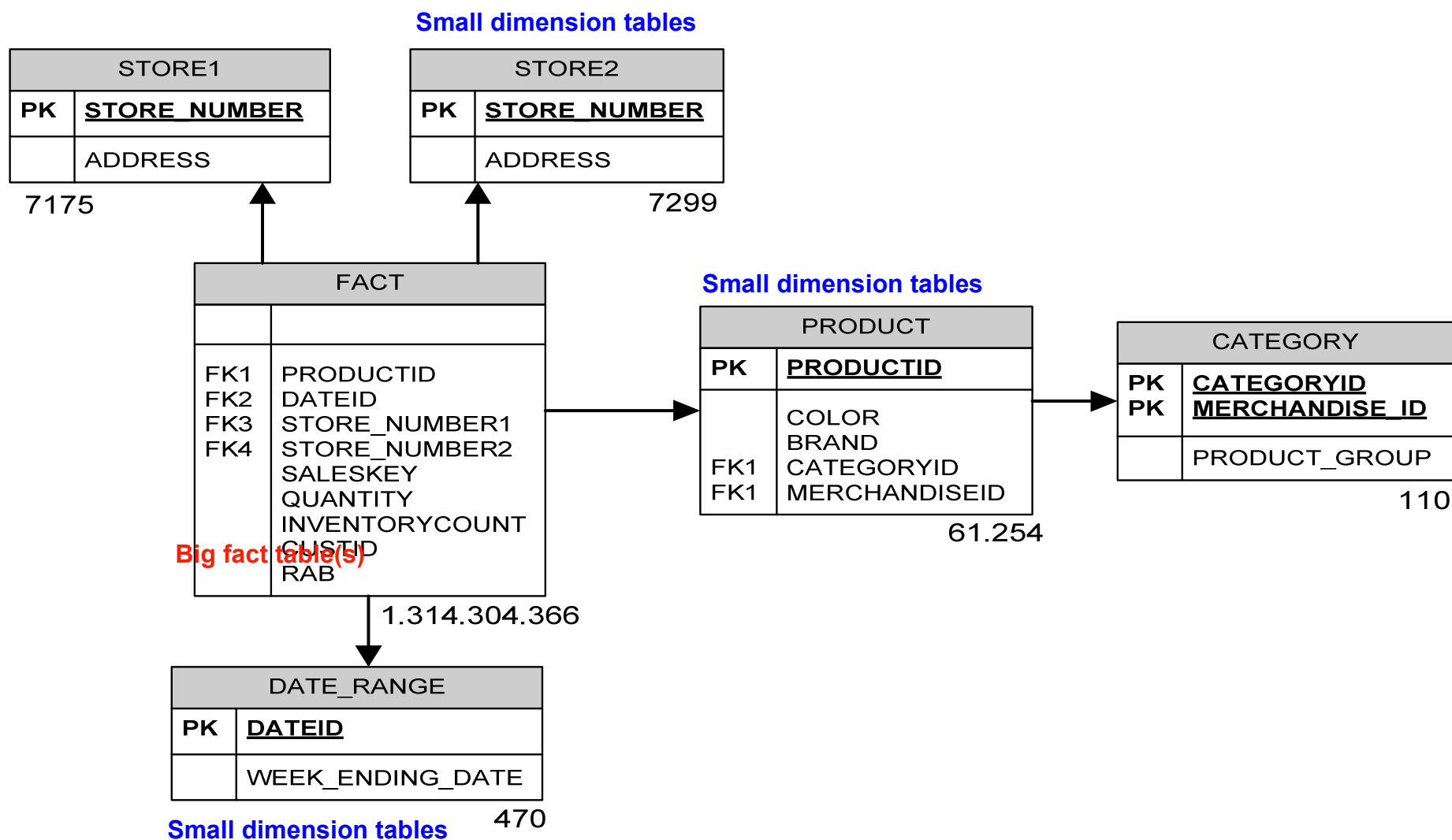
典型的分析类应用 (BI) 所需的表设计

- 数据及查询特点:
 - 星状或者雪花状数据结构:



- 通常是复杂的，涉及大量数据的查询
 - 趋势分析，决策分析
 - 涉及大量的数据 (这点与OLTP有很大区别)
 - 大多涉及汇聚操作 (e.g., COUNT, SUM, AVG,...)
 - **可以利用到ISAO!**

例子：最能发挥ISAO优势的数据结构



定义需要加速的数据

- 对DB2而言，数据集市(MART)是一组逻辑上相关联的表。例如：一个星状结构里所有的表属于同一个数据集市。
- 系统管理员通过客户端工具定义数据集市，组成数据集市的表以及这些表之间的关系。
- DB2 for z/OS 在数据字典(Catalog)里记录每个数据集市的信息。之后，DB2会把数据集市所涉及到的数据拷贝到ISAO。
- ISAO 压缩数据，并把数据装载到worker node的内存里



操作步骤: 定义并发布数据集市(mart)

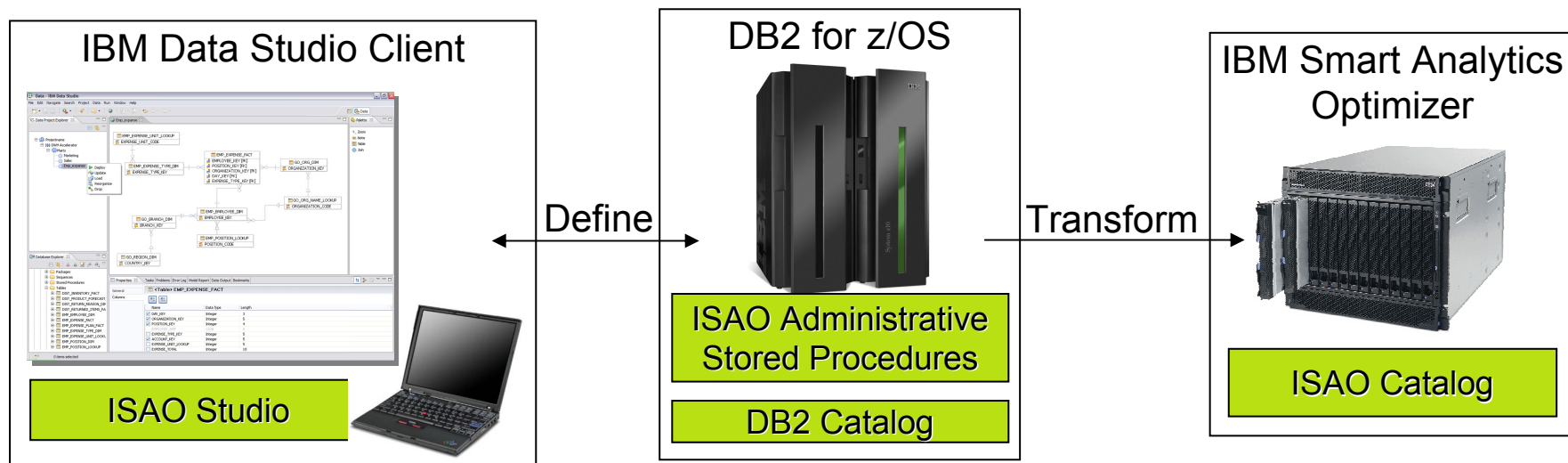
A. DBA 定义数据集市 (Data to Accelerate)

- 对DB2而言, 数据集市(MART) 是一组逻辑上相关联的表。
 - 例如: 一个星状结构里所有的表属于同一个数据集市。
- 系统管理员通过客户端工具定义数据集市, 组成数据集市的表以及这些表之间的关系。

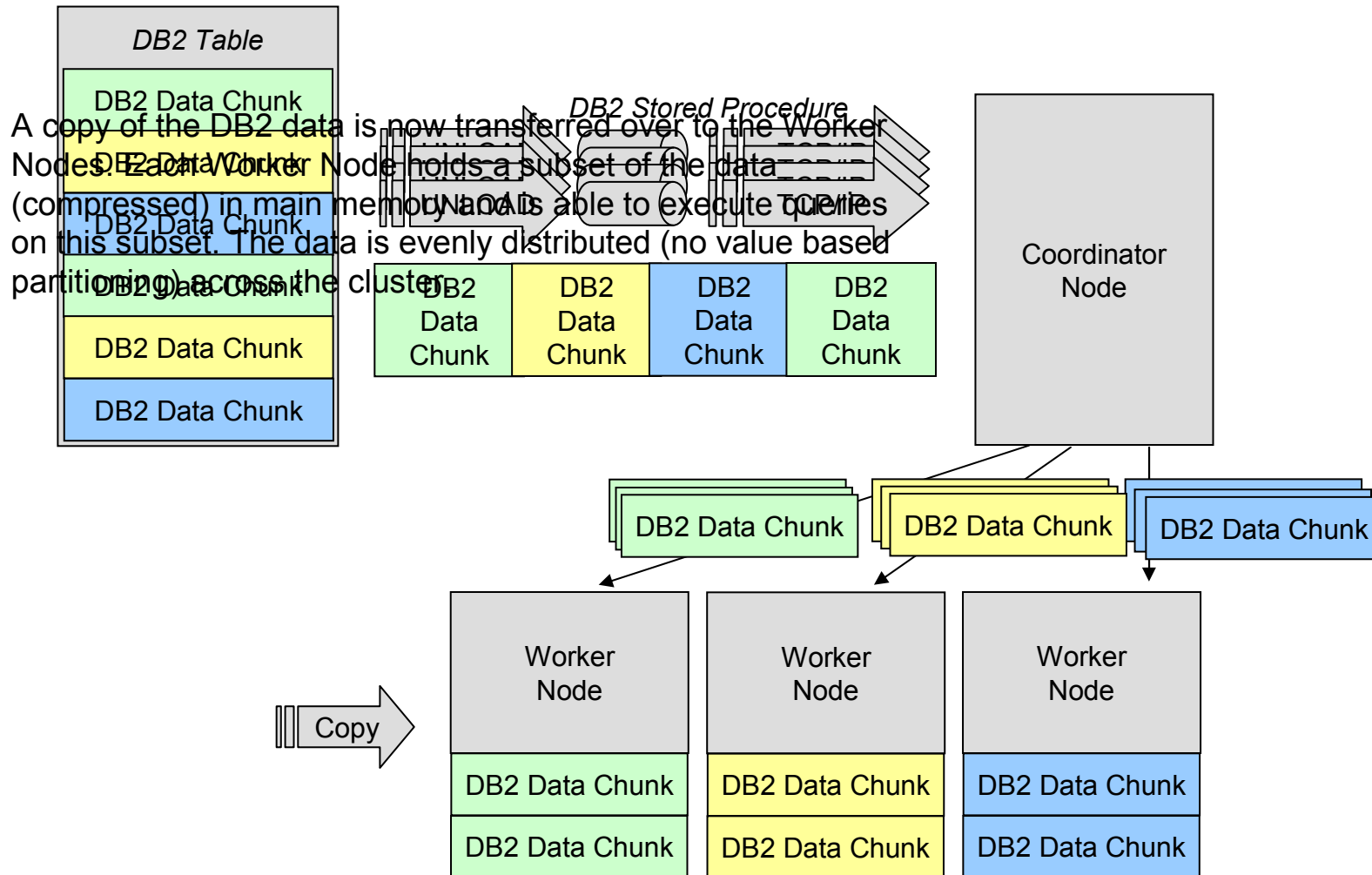
B. 通过存储过程, DB2 自动把mart里面的数据拷贝到ISAO

- DB2 for z/OS 在数据字典(Catalog)里记录每个数据集市的信息。

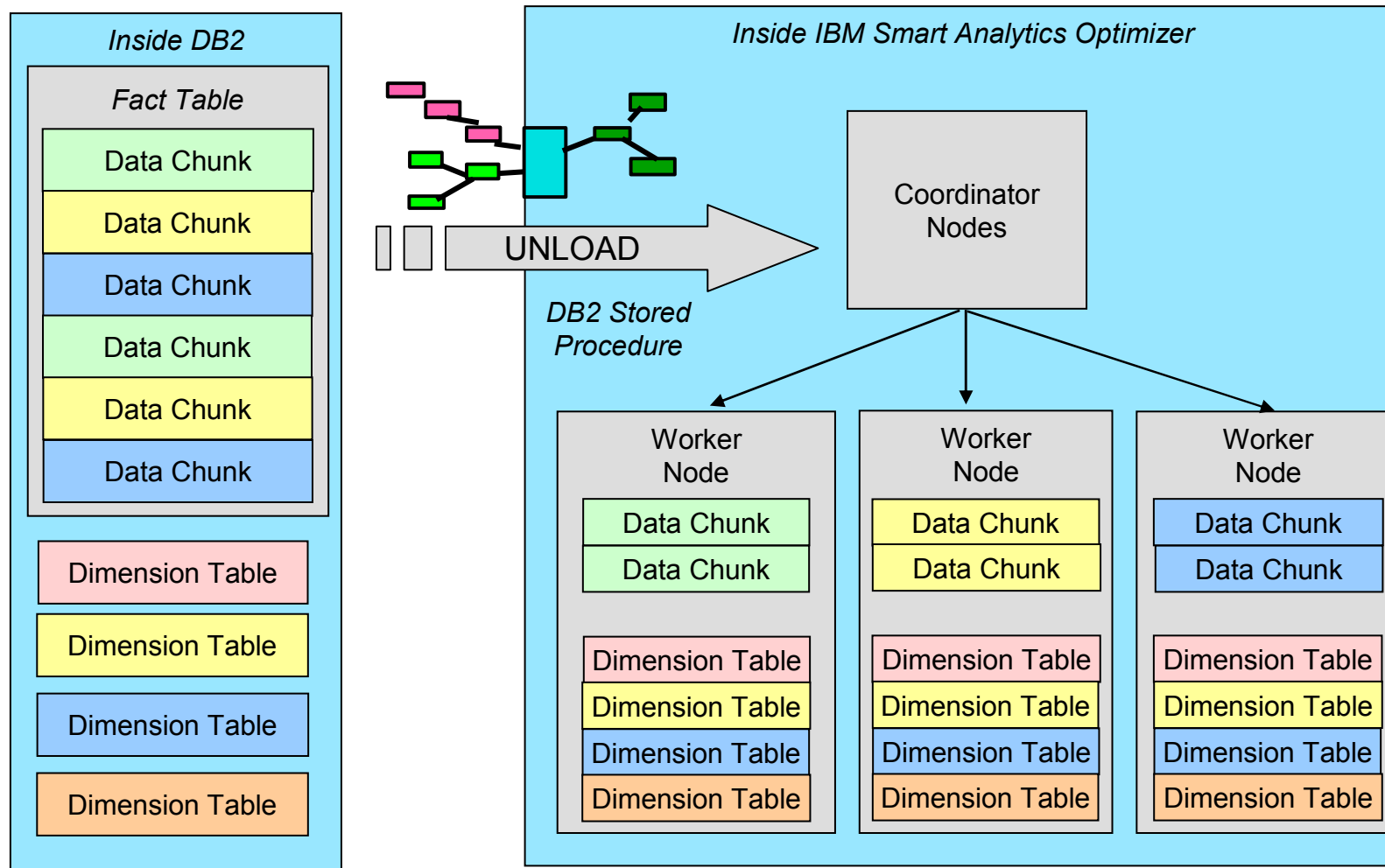
C. ISAO 自动压缩数据, 并把数据加载到worker node的内存里



C. IBM Smart Analytics Optimizer – 发布数据

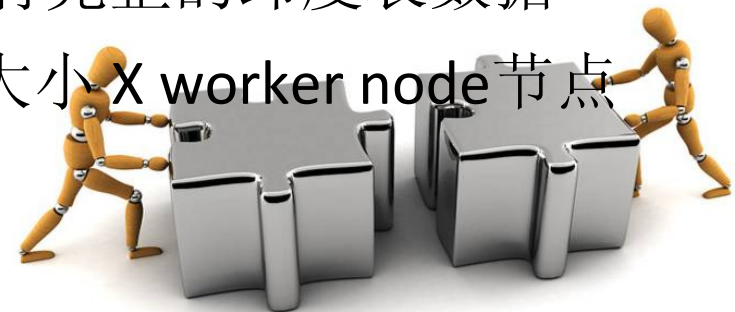


发布数据(Continue)



数据关联(Join)策略

- 根据ISAO集群内worker node的数量，事实表的数据被平均拆分，并把拆分后的数据发布到各个worker node。
 - 如果事实表很大，需要更多的worker node，以便worker node有足够的内存容纳事实表数据
- 事实表和纬度表之间的关联在worker node内部完成，不需要垮worker node。
 - 在每个worker node内存里都保存有完整的纬度表数据
 - 某纬度表所需的总内存空间为其大小 \times worker node节点的数量



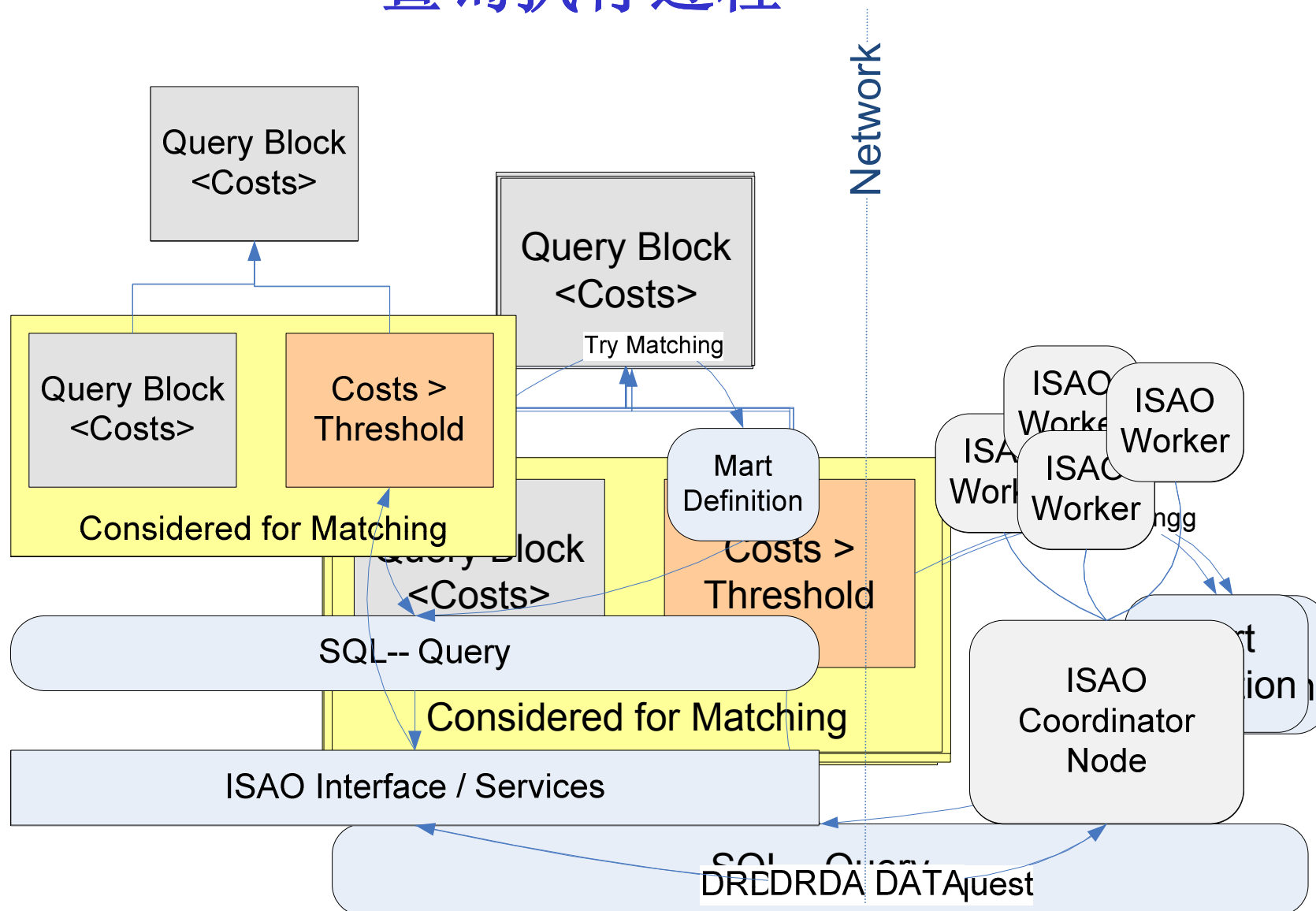
议程

- 分析类应用基于主机的解决方案
- 查询优化工具：ISAO
 - IBM Smart Analytics Optimizer (ISAO)概要
 - 不同应用类型需要不同的数据存储格式
 - 在DB2 内定义适用于ISAO的查询
 - 在使用ISAO情况下，查询操作的执行过程 ←

ISAO 的目标：加快分析系统的查询

- 快速扫描大量数据
- 操作数据时，通常带有聚合操作

```
SELECT PRODUCT_DEPARTMENT, REGION, SUM(REVENUE)
FROM FACT_SALES F
      INNER JOIN DIM_PRODUCT P ON F.FKP = P.PK
      INNER JOIN DIM_REGION R ON F.FKR = R.PK
      LEFT OUTER JOIN DIM_TIME T ON F.FKT = T.PK
WHERE T.YEAR = 2007
      AND P.TYPE = 'CAPIT207'
GROUP BY PRODUCT_DEPARTMENT, REGION
```



IBM Smart Analytics Optimizer – 适用范围及限制

- ISAO 只能加速查询操作.
- 所有的数据首先必须存储在DB2 for z/OS first, 然后再传输到ISAO
- 数据集市的所有数据都必须传输到ISAO, 而不仅仅是某个查询涉及到的数据.
- 不是所有的星状结构的数据都会使用到ISAO, DB2 优化器会选择最有效的执行路径.
- ISAO 的处理单元为一个query block
 - Queries which consist of several Query Blocks, are not seen as whole query by the accelerator but only as single, independent blocks. Outer query block containing inner query blocks is not routed
 - The results of subselects can not be used by Smart Analytics Optimizer in the outer query
(DB2 would need to pass a subselect result to Smart Analytics Optimizer)
- 除了以下函数, ISAO 支持大部分的 DB2 函数:
 - Mathematical functions like sin, cos, tan, exp, correlation
 - User defined functions
 - Advanced string functions like locate, left, like, overlay, position
 - Advanced OLAP functions like rank, dense_rank, row_number, rollup, cube

Many of these, as well as new functionality, will be addressed in future releases ...

Thank you
