

本文主要内容来自 [SpriCoder的博客](#)，更换了更清晰的图片并对原文的疏漏做了补充和修正。

本文提供了 pdf 版，以供打印：

1. 商务智能的起源

1.1 数据库系统的发展

- 1968 年 IBM 公司：层次 IMS
- 1969 年美 CODASY：网状 DBTG 标准
- 1970 年 IBM 的 E.F.Codd 提出关系模型
- 20 世纪 70 年代以层次、网状为主流
- 20 世纪 80 年代关系系统逐渐代替层次与网状模型
- 以关系型数据库为基础，建立大量业务系统和信息系统，累计大量数据

1.2 计算平台的发展

- 计算能力
- 存储能力
- 网络/非网络传输能力

1.3 采集手段的发展

- 基于光学的采集手段：二维码、图像/视频
- 传感器、移动设备
- RFID 等

1.4 数据爆炸

Table 1.2: Worldwide production of original information, if stored digitally, in terabytes circa 2002. Upper estimates assume information is digitally scanned, lower estimates assume digital content has been compressed.

Storage Medium	2002 Terabytes Upper Estimate	2002 Terabytes Lower Estimate	1999-2000 Upper Estimate	1999-2000 Lower Estimate	% Change Upper Estimates
Paper	1,634	327	1,200	240	36%
Film	420,254	76,69	431,690	58,209	-3%
Magnetic	5187130	3,416,230	2,779,760	2,073,760	87%
Optical	103	51	81	29	28%
TOTAL:	5,609,121	3,416,281	3,212,731	2,132,238	74.5%

Source: How much information 2003

1. 至 2010 年，仅以 Disk storage 存储的数据量就将达到 110,000,000 terabytes
2. 在 2007 年，人类能够存储 2.9×10^{20} (295 EXABYTES) 压缩字节，传输将近 2×10^{21} 字节，并在通用计算机上每秒执行 6.4×10^{18} 条指令。

EXABYTES：代表非常大量的数据

1.5 数据，信息，知识

改善信息访问，以信息片段的形式存在。



1.5.1 知识

1. 知识是头脑中的见识和背景，需要反思和综合，知识难以构造，难以在机器上捕获和转移，通常是默认信息；
2. 知识属于三者中的最高层的部分，是用来解释为什么信息会是这个样子；
3. 如果当前的数据分布仍然服从过去的规律，那么就证明过去的知识对未来仍然是有意义的，所以我们尝试构建出知识和预测，根据知识和信息做出决策。

1.5.2 信息

1. 信息是具有相关性和目的的数据，我们需要对信息的意义达成共识，必须是被人为调制处理后的。
2. 我们使用信息去查看情况，事务机制，查看数据块。

1.5.3 数据

1. 数据是有关世界的客观事实，易于组织和捕获、转移
2. 数据流中使用到什么样的数据，我就采集什么数据。
3. 了解实体的所有数据片段才能做出某些判断

1.6 知识价值链

1.6.1 知识价值链：数据

Data has in itself little semantic content:

For example: Price(\$)

5000

Data has more meaning when presented as part of a structure (context) ----> *Information*

For example:

Prodno	Type	Make	Model	Price(\$)	Code
4825	Wagon	Ford	1994	5000	Almad21

1. 数据使用 01 串的形式存储，以及很多的东西都是可以 01 串化的。
2. 5000 这个数据本身单独拿出来是没有任何意义的，数据本身仅仅只是数据，只有当数据被表示为了结构的一部分时才有了更多的含义。
3. 大量的信息系统中，我们在使用数据的时候，人为地添加了关于这个数据的解释。

1.6.2 知识价值链：信息

Information is relevant if it makes a difference to receive it or not

For example:

Prodno	Type	Make	Model	Price(\$)	Code
4826	Car	BMW 328	1997	5000	Almad21

Consumer Report
Prices of Cars: BMW 328

1996	\$12,000
1997	\$19,000
1998	\$27,000
1999	\$32,000

将数据放在一起，相互认证相互解释相互绑定，有一个作为整体的趋势，让原本没有联系的数据，赋予上下文，使得数据的意义更加的明确，就成为了信息。

信息基于数据，没有数据就没有信息。但信息不完全来自数据。

1.6.3 知识价值链：知识

Information is the source of knowledge

Information Example: Loan Risk Assessment

Name	Client Since	Credit Evaluation	Assets	Loan		Risk
Rich Harris	1985	Good	20,000	20,000	...	Medium
John Leggo	1997	Fair	100,000	130,000	...	High
Sandra Hilt	1992	Good	10,200	2000	...	Low
...

Knowledge (As a business rule)
"If we know the applicant, and his/her credit is good, and the applicant has assets for at least twice the amount of the loan, then the risk is low, and unless you find something really strange, the loan should be granted"

信息是知识的来源。

知识（作为业务规则）：“如果我们知道申请人，并且他/她的信用良好，并且申请人的资产至少是贷款额的两倍，那么风险就很小，除非您发现确实很奇怪的事情，否则应该给予贷款”。

信息的利用依赖知识的提取。

1.6.4 知识价值链：智慧

机器难以达到的水平。

举例：

通过分析商店销售情况，发现购买牛奶和购买饼干的顾客重叠度高。为了提高营业收入：

1. 把饼干和牛奶捆绑销售以提高销售额
2. 把饼干和牛奶放在商店的两侧以使顾客遍历整个商店

1.7 企业的新问题（区别于旧问题）

1. 企业决策应当源于企业和行业的真实状况，真实状况来源于真实完整的数据；
2. 解决的问题：如何有效将企业收集的数据转化为信息，并最终为企业决策所使用？
3. 在数据爆炸和信息贫乏的前提下，需要解决信息的可访问性、及时性、表达格式和完整性问题。

1.7.1 以零售业为例

经营规模	描述	日常操作	决策操作
杂货店	数十种商品，一两员工	记账本、单机版信息管理系统	人工方式
小规模零售连锁	三五家同质零售商店 信息的结构和复杂程度和杂货店相比没有变化，只是信息量有变化	网络版信息管理系统	报表、信息工人
复杂零售集团	不同地域，不同性质部门 各种不同领域产品的管理系统，进行横向纵向的复杂切分	各种信息管理系统、专门信息系统	?

1.7.2 早期解决方案

第一代：基于主机的查询与报表

1. 早期的商务信息系统使用批处理应用程序为商业用户提供它们所需的信息；
2. 第一代的商务信息系统只能被诸如业务分析人员之类的熟悉数据且有相当计算机经验的人员使用；
3. 管理人员很少能够使用这些早期的系统，他们必须依靠信息提供者来解答他们的问题，并给他们所需要的信息。

第二代：数据仓库

1. 第二代信息系统应用了数据仓库技术。从而使性能有了一个飞跃；
2. 数据仓库被设计用来满足业务用户的需求，而不是每天的操作型应用；
3. 数据仓库中的信息是干净的、一致的，并且这些信息以业务用户能理解的形式储存。

1.7.3 为什么需要商务智能

许多数据仓库解决方案共同的一个缺点就是系统的开发者总是专注于软件技术，而不是商业解决方案。因此，尽管他们所提供的产品能够很好地构建和访问数据仓库，但这些产品实施起来相当复杂：

1. 数据仓库产品通常很少针对特定的产业、应用领域提供解决特定的业务问题的软件包
2. 企业所需要的是应用程序和商务解决方案，而不仅仅是技术

数据仓库解决方案的另一个问题是：人们仍然过多地关心如何建立数据仓库，而不是如何对它进行访问，许多企业似乎觉得他们只需要建立起数据仓库并为用户提供合适的工具，问题就解决了。事实上，这只是个开始而已。

1.7.4 商务智能

1. 商务：Buying and selling; commerce; trade
2. 智能：人工智能，是夸大的“智能”。
 1. 图灵测试：商务智能需要通过图灵测试吗？商务本身就是人类最复杂的行为之一
 2. 专家系统：医疗领域等

BI 的目标：

1. 决策支持 (Decision-making Supporting)

2. 改善信息访问 (Information access-improving)

定义一：从数据到信息，BI 是将数据转换为有意义的内容的过程

1. BI 通过把来自于不同系统的数据汇聚成一个单一的可获取的信息源——数据仓库 (Data Warehouse, DW) 。
2. 基于用户的要求，使用各种工具来分析数据仓库中的数据，并可视化其结果

定义二：访问、钻研、分析和挖掘数据获得启发和了解，从而提供更完善和更全面考虑的决策支持——Gartner Group

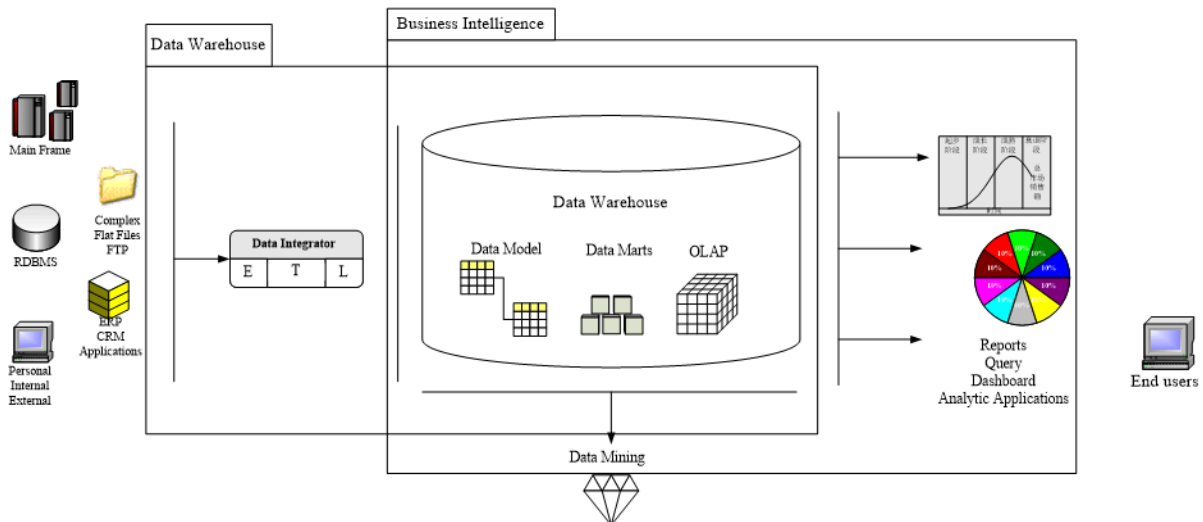
定义三：

商务智能是企业利用现代信息技术收集、管理和分析结构化和非结构化的商务数据和信息，创造和累计商务知识和见解，改善商务决策水平，采取有效的商务行动，完善各种商务流程，提升各方面商务绩效，增强综合竞争力的智慧和能力——《三位一体的商务智能：管理、技术与应用》

1. 结构化就是在数据库内的数据
2. 非结构化就是不在数据库中的数据
3. 商务智能系统不仅支持最新的 IT 技术，同时也提供了打包的应用解决方案
4. 商务智能系统着眼于终端用户对业务数据的访问和业务数据的传送，它同时为信息提供者和信息消费者提供了支持
5. 商务智能系统支持对所有格式商务信息的访问，而不仅仅是那些存储在数据仓库中的信息。
6. 然而，在本课程中，我们集中关注 IT 技术，略过应用解决方案
7. 推荐阅读：[企业数据的秘密：大数据时代商业规则](#)

2. 商务智能的构件

2.1 BI 的构件



2.1.1 信息与决策

1. 信息持久化：如果不持久化信息，则会导致相应的信息挥发掉。
2. 决策的时效性：
 1. 正常决策：正常按照公司运转需要作出决策
 2. 临时决策：突然发生了紧急事件，需要决策
3. 问题：有很多的信息和处理都是重复的，所以我们可以尝试将中间结果或者最终结果存储下来。
4. 多维数据可以等价于信息体。

2.1.2 访问与处理

1. 大量的数据库会存放在关系数据库中：用结构化查询语言对数据库进行查询，数据库本身不接触信息采集的人员。
2. 关系数据库是可以双重定制的，比如查询考勤率，因此应当提供查询工具。
3. 联机查询：Analytic Application、OLAP (Online Analytical Processing)
4. 数据挖掘：可以属于商务智能也可以不属于。

2.1.3 ETL：数据集成的代名词

因为我们需要将非结构化、半结构化和结构化文件解析并存储，这些文件可能编码不同，不仅仅是字符编码，还可以是我们的一些约定，他们是我们的大项文件的组成部分

ETL 的含义：

1. E (Extract)：理解、排查错误
2. T (Transform)：转换
3. L (Load)：装载

ETL 过程中可能出现的问题：

1. 错误的检查识别：有部分数据是显然违背了数据预定义或者常识，应该识别出来。
2. 同名问题：主键是否可能出现重名问题，比如同姓名，身份证号相同等等。
3. 点名问题：点名记录中，有的是按照时分秒点名、有的不点名、有的不精确到秒、有的是通过小测试点名（多包含一个小测试成绩）

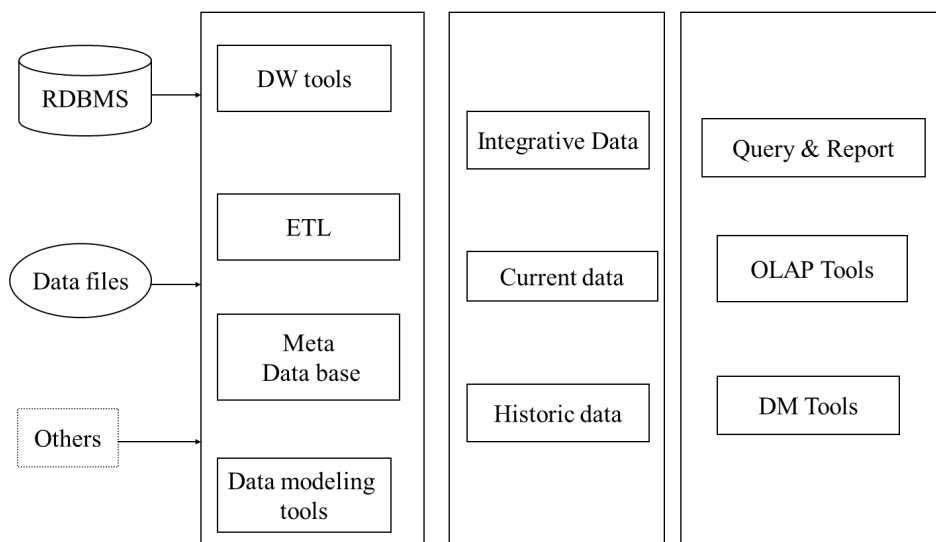
数据仓库会有很多的备份，而不是一个，互为冗余关系，底层数据计算可以得到上层的聚合数据

N 对 N 的关系

扩展阅读：

1. [\(推荐\) ETL 的基础知识，看完你就全明白了!](#)
2. [数据仓库 ODS、DW 和 DM 概念区分](#)
3. [ETL 讲解 \(很详细!!!\)](#)

2.2 BI 的构件图示



从左到右：

1. 第一部分：数据源
2. 第二部分：数据获取
 1. 最重要的 ETL
 2. 用数据建模工具和其他工具进行管理

3. 使用元数据类型来决定数据的形式
3. 第三部分：数据存储和管理：所有的数据都应该涵盖在这些数据中，一半的数据存储和管理 + 信息获取。
4. 第四部分：信息获取：这部分已经成为信息
 1. 和主题相关的数据项
 2. 主题：拥有共同上下文的综合体

2.3 数据源

商务智能体系中的所有数据均来源于数据源。

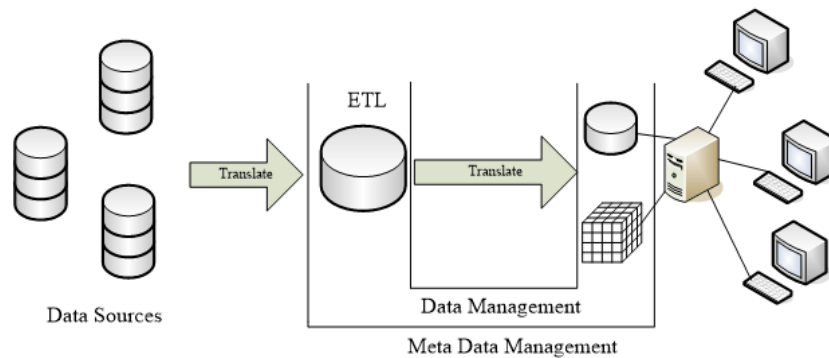
数据源包括：

1. 操作型数据库：大量数据都来源于这个之中，用来存放每一次操作产生的数据的备份的数据库
2. 历史数据：当前操作型数据被构建前（系统数据化前）的数据信息化后进入数据仓库，比如大量的数据在失效后转储到低效低成本的存储中去，比如五年前的交易数据。
3. 外部数据：企业或者组织的边界
 1. 操作型环境下绝大多数情况下是没有外部数据的
 2. 决策性环境下一般会用到外部数据。
 3. 外部数据被需要的情况：
 1. 计算销售情况、传统菜单的成本盈利
 2. 如果我们打算推出樱花虾套餐，我们需要调研竞品（其他公司有没有做过）用来决定要不要做。
 3. 某些产品或者行为可能会面临政策等问题的影响。
 4. 比如合作伙伴的相关信息也应该包含在其中。
4. 数据仓库中的信息：数据仓库本身是数据的上游
 1. 自身怎么成为自身的数据来源？这个数据仓库可能是外部的某些数据仓库。这个数据仓库还可能是自己的数据仓库（中间结果会被再次使用到）
 2. 这个数据仓库可能分析是用户自身数据的相关数据来分析相应的标签，这个标签可能也被后续很多方面用到。
5. 相关的数据库和数据结构：任何可以分析的数据

数据源存在于不同的平台，既可以是格式化的也可以是非格式化的

2.4 数据仓库：概要

1. 数据仓库是 BI 的重要组成部分和数据基础
2. 其典型工作是要对集成、清洗、聚集、预计算和查询任务所需要的大量数据进行批处理
3. 其主要技术包括
 1. 数据提取、转换和装载（extract, translation and load, ETL）
 2. 数据管理
 3. 数据访问
 4. 元数据



2.4.1 ETL

2.4.1.1 ETL 过程

1. 辨识与主题相关的原始数据：
 1. 有用的数据：数据源的数据是可靠的，有一定百分比可以保证有效性，比如 QQ 的年龄。
 2. 主题：是相对主观的评价，比如我们认为某一个字段是无用的，那么就不应该在这部分使用。
 3. 这个步骤是相对困难的。
2. 开发数据抽取策略：正确和完整的数据 (E)
3. 将原始数据转换为目标规格 (T)
4. 将原始数据加载到预定目标区域 (L)

2.4.1.2 数据目标

1. 原子层 (Atomic Layer) 和集成数据：
 1. 每一条操作的数据，详细程度。
 2. 计算资源不是无限的情况下，我们可以没有集成数据，集成数据是提前计算出来的需要用的数据，比如特定大小的图盘
2. 数据集市 (Data Market) :
 1. 数据仓库是一个大而全的数据备份，是所有分析人员的主观的分析集合
 2. 一般会超过个人的需要：不应该让所有的底层人员可以拿到所有的数据，因此大多数情况下不会使用到所有的数据仓库，用到数据集市即可，比如 4S 店和集团做决策的区别
 3. 所有的人都访问数据仓库会导致效率低。
3. 操作数据存储 (Operational Data Storage, ODS)
 1. 产生出来的大量的操作数据是不会直接进入数据仓库，因为没有预留资源，也没有办法做到准实时。
 2. 隔一段时间，完成一次 ETL，保证数据仓库的一致性。这就意味着当前的数据仓库的数据并不会和最新的操作数据一致。
 3. 操作数据存储：数据仓库的补充，是一个准实时的仓库，允许使用还未进入数据仓库的数据。
4. 缓冲区 (Staging Area) : 处于性能考虑，在 BI 中放置一定的缓冲区，保证数据更有效不丢失的处理

2.4.2 数据管理

1. 涉及到数据的存储，索引和备份
2. 管理的数据的范围
 1. 关系数据库中的数据
 2. 数据立方体 (Data Cube) 中的多维数据 (multi-dimensional data)

2.4.3 数据访问

1. 一种或多种访问数据仓库中的数据的方式。
 1. 需要结合需求进行分析，将客观的数据的处理流程全部信息化，比如收银员的收银流程。
 2. 所有的分析是没有客观的流程的，比如筛选流程可以是不同（筛选法，筛选方式不同）
2. 面向多种商业用户
 1. 行政管理人员
 2. 商务分析人员
 3. 操作经理
 4. 临时终端用户
 5. 其他
3. 需要支持多种访问方式以及分析和展现工具
 1. 查询和报告工具
 2. 桌面 OLAP
 3. 关系 OLAP
 4. 多维 OLAP
 5. 数据挖掘
 6. 基于仪表盘和基于代理技术等方面的客户决策支持界面
4. 程序员不应该再插手这部分。

2.4.4 元数据

1. 元数据是关于数据的数据，是如何管理数据仓库的重要数据。
 1. 描述了数据的结构、内容、编码、索引等内容
 2. 表名，系统名，索引，数据结构，编码，ID，.....
2. 种类
 1. 关于数据源的元数据
 2. 关于数据模型的元数据
 3. 关于数据仓库映射的元数据
 1. 数据仓库外部的映射关系，比如 1 映射给女，0 映射给男
 2. 数据仓库内部也存在映射关系：比如简单数据和聚集数据之间
 4. 关于数据仓库使用的元数据：每一个信息体中的每一个信息的含义
3. 元数据管理是针对元数据的管理，使用 DBMS（关系型、对象、对象-关系）进行管理

2.5 分析和展现层

针对信息的分析和展现

1. 报表和查询
2. 联机分析处理（OLAP）
3. 数据统计
4. 数据挖掘

2.5.1 报表和查询

1. 针对不同商业用户的报表和查询：尤其是非信息处理专业的商业用户
2. 需要具备定制报表和定制查询的能力
3. 包含非限定的报表和查询种类

2.5.2 联机分析处理 (OLAP)

联机分析处理提供一种快速的、交互式的、相互融合的信息访问方式，不仅可以回答 who、what 之类的问题，而且可以回答 how、why 之类的问题。

具有以下特点：

1. 对数据进行多维审视的能力：背景越复杂，会导致当前处理和决策就越复杂。
2. 精密计算能力：所有的视图需要在有效的时间内将所有图计算出来并显示给分析人员。
3. 时间智能

OLAP 的基本分析操作：

1. 切片
2. 切块
3. 旋转
4. 数据概括 (roll up)
5. 数据细化 (drill down)

2.5.3 数据统计

1. 数据统计是将数据中含有的信息概括为统计值：例如最大值、最小值、平均值等具有特殊含义的值
2. 典型的数据统计方法：
 1. 关系分析
 2. 要素分析
 3. 回归分析
3. 数据存储的是多维信息体，我们拿到的视图是多维的，我们也可以进行相应的对比分析。

2.5.4 数据挖掘

数据挖掘就是对数据库或数据仓库中蕴涵的、未知的、非平凡的、有潜在应用价值的模式或规则的提取：

1. 基于数据库和数据仓库是任意的，也可以不面向如上两个。一般是面向数据库和数据仓库，如果没有数据库和数据仓库，则需要将构建这二者的相应的操作都做一遍
2. 这些模式（规则）的提取：不要求是有效的模式，不要求规则是公理
3. 蕴涵：一定是有隐含对应的关系
 1. 不能告诉你解释：因为不支持对当前的数据集（未知）的解释
 2. 不能不告诉你解释，因为是根据之前的数据集的解释
4. 未知的：是没有发现的规律
5. 非平凡：有一定的概率是对的，有一定的概率是错的，在当前的数据集上有一定的百分比成立
6. 潜在应用价值：当前数据挖掘的结果，如果不能有一个所谓的未知性，那么是不存在潜在的应用价值的。

用于发掘数据中隐藏的模式：数据的规律是针对这一部分数据和相似数据是成立的，但是对于其数据不一定是成立的。

需要借鉴各种相关领域的理论和方法。

用于发现隐藏模式的算法既可以是自动进行，也可以在人工指导下完成。

2.5.4.1 数据挖掘方法

1. 特征化与区分：特征工程
2. 关联规则挖掘：若干个不同规律出现的频繁程度，A 在则 B 一定在。
3. 分类挖掘：根据已有分类标签来提取分类规律，建立一般数据和分类标签的映射。
4. 聚类挖掘：将用户分为几个小的类别，但是不确定会将同学们分为什么类别，根据特性拆分为小组。
5. 时序和挥发性数据挖掘：有些数据是具有时序性的，有些数据是没有必要进行持久化，挥发性数据可能需要我们进行一定的决策，但是有一些时序性数据需要是在线的。
6. 异常分析：上面五点是通过一般的数据进行分析，部分时候异常信息会包含有较多的信息量，比如有一位同学四处进行大额汇款（出现异常情况，和日常消费习惯不一样）

2.5.4.2 面向的数据类型（其他的数据挖掘方式）

1. 关系型数据
 1. 事务型数据库
 2. 数据仓库
2. 文本数据挖掘
3. Web 挖掘
4. 空间数据挖掘
5. 多媒体数据挖掘

3. 商务智能解决方案

3.1 IBM

DB2 Data Warehouse Enterprise Edition

1. 数据仓库引擎：一般架设在成熟的数据库上
 1. DB2 UDB Enterprise Server Edition - 通用海量并行数据仓库
 2. DB2 UDB Data Partitioning Feature - 数据分区部件
2. EII 与 ETL
 1. DB2 Information Integrator Standard Edition - 信息集成中间件
 2. DB2 Warehouse Manager Standard Edition - 数据仓库 ETL 工具
3. 查询和多维分析：OLAP
 1. DB2 Query Patroller - 基于成本的查询负载管理工具
 2. DB2 Cube Views - OLAP 元数据交换工具及物化查询表生成器
4. 数据挖掘
 1. DB2 Intelligent Miner Scoring - 数据挖掘评分工具
 2. DB2 Intelligent Miner Modeler - 数据挖掘建模工具
 3. DB2 Intelligent Miner Visualization - 数据挖掘模型图示化工具
5. 前端分析
 1. IBM Office Connect Analytical/Enterprise Web Edition - Excel 多维分析插件
 2. DB2 Alphablox 前端应用开发组件

3.2 Oracle

Oracle 业务智能企业增强版 (EE)

1. Oracle BI Server: 常见的企业业务模型和抽象层
2. Oracle BI Answers: 即席查询和报表
3. Oracle BI Interactive Dashboards: 高交互性信息板, 用于访问业务智能和应用程序内容
4. Oracle BI Delivers: 主动式业务活动监视和警报
5. Oracle BI Disconnected Analytics: 针对移动专业人员的完整分析功能
6. Oracle BI Publisher (以前称为 XML Publisher): 企业报表和“点对点”报表的分发
7. Oracle BI Briefing Books: 信息板页面快照, 用于在离线模式下查看和共享
8. Hyperion Interactive Reporting: 直观且高度互动的即席报表
9. Hyperion SQR Production Reporting: 批量生成高质量的格式化报表
10. Hyperion Financial Reporting: 图书级质量的格式化财务和管理报表
11. Hyperion Web Analysis: 基于 Web 的联机分析处理 (OLAP) 分析、演示和报告

3.3 其他

1. Microsoft

1. Microsoft SQL Server Business Intelligence (BI) platform

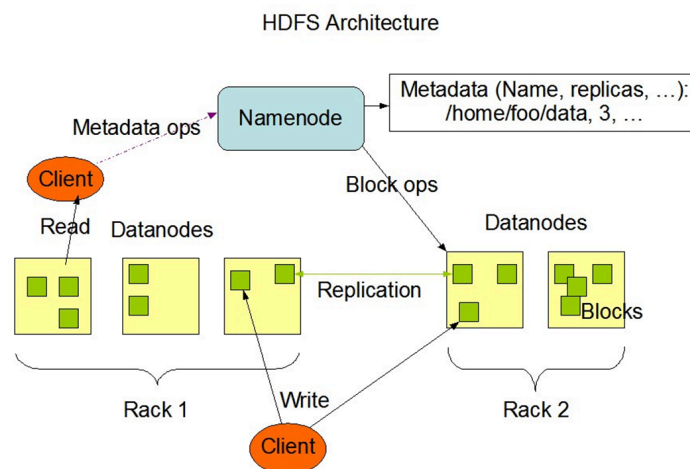
1. SQL Server Management studio, 数据库管理工具
2. SQL Server BI studio, 商务智能解决方案工具 (可以去看看)

2. Sybase: Sybase IQ 解决方案

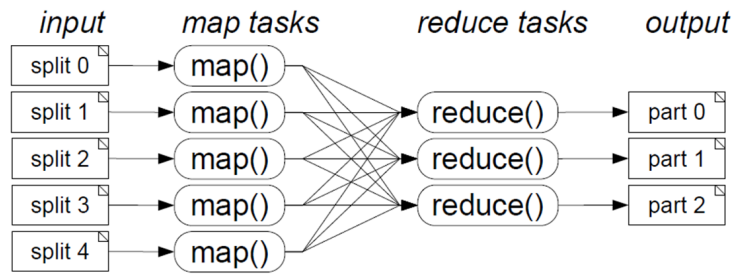
3. SAP: SAP NetWeaver Business Intelligence

4. 云平台: HIVE

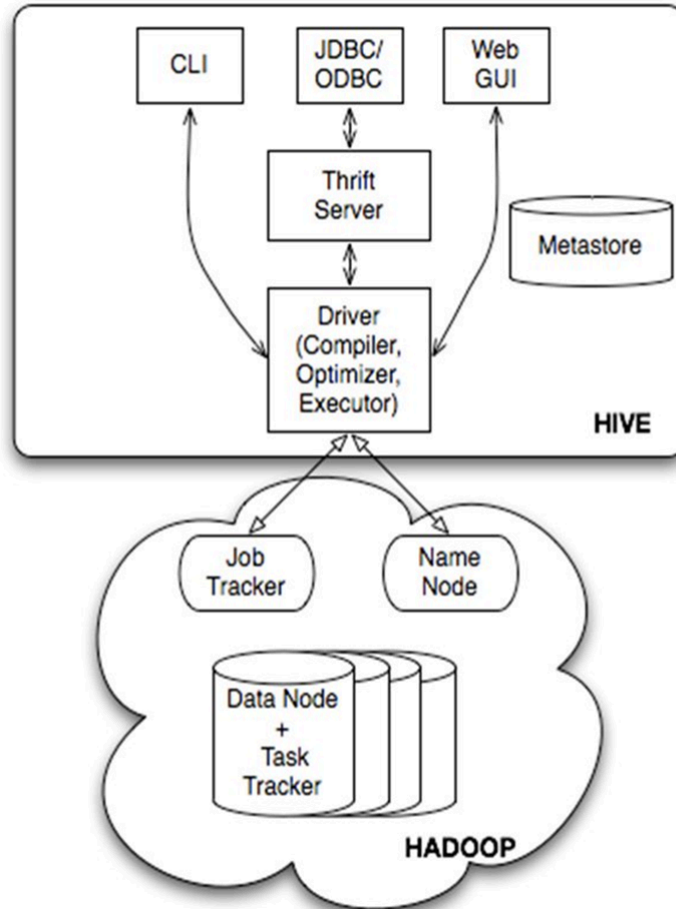
3.4 HDFS



3.5 Map Reduce

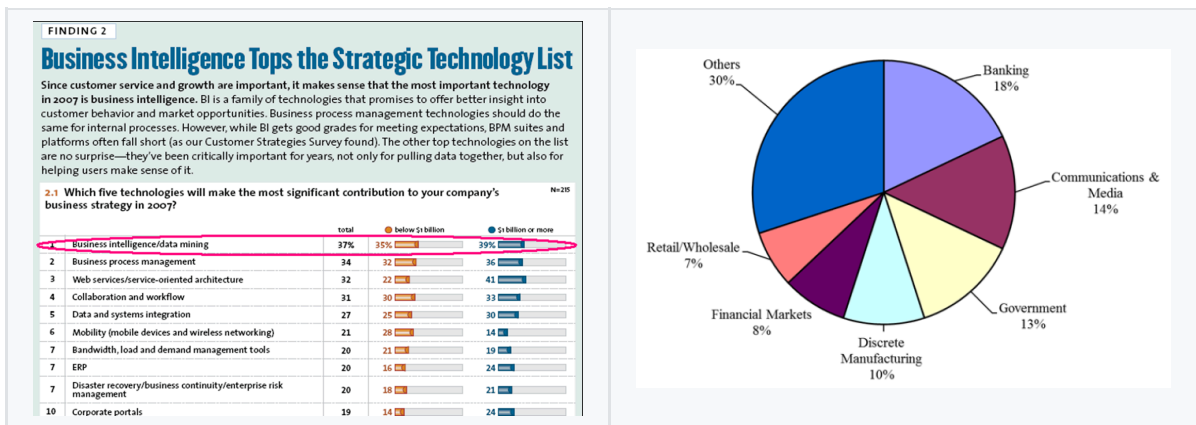


3.6 HIVE



4. 商务智能的发展

4.1 商务智能的重要地位



	排名	相关主题
2007	1	Business Intelligence/Data Ming
2009	9	Business Intelligence
2010	2	Advanced Analytics
2011	5	Next Generation Analytics
2012	6	Next-Generation Analytics
2012	7	BigData
2013	6	Strategic Big Data
2015	4	Advanced, Pervasive and Invisible Analytics

1. Banking | Communications & Media: 大型垄断制企业会垄断这部分，有商务智能的部分。
2. Government: 政府不以盈利为目的，可以收集到各个方面的数据，人口普查（操作型数据库收集），进行决策支持

4.2 商务智能的发展方向

1. 分布式的商务智能：逻辑意义上的分布式商务智能，希望有一个大而全的数据仓库。
2. 插件化集成的商务智能：尽力减少人工成本，没有办法解决的，以下为两个困难的事情
 1. 即使我们有两个一模一样的企业，由于分析人员的主观性不同，也会导致数据仓库中的主题是不同的；主题内的数据特征也可以是不一样的；两个企业之间的数据库可能也有一定差异。
 2. 每一个企业的数据环境可能是不一样的，但是对于数据仓库可能比较麻烦。
3. 大众化的商务智能：可能只有很少部分的人可能会用到商务智能，在有限资源下进行决策，大多数决策是不需要在线的，
 1. 战术决策的支持：汽车等
 2. 个人商务智能工具：根据个人情况进行决策
4. 在线商务智能：在线数据挖掘
 1. 考虑计算资源
 2. 数据结构也需要能够保证在线
5. 零滞后的商务智能：
 1. 时效性：数据仓库是定时刷新的。
 2. 做一年的决策时，半天的数据是无关紧要的（如果没有发生重大事件），但是

3. 个人做下半天的决策时，上半天的数据是很重要的，比如上午已经购买，但是没入库，这时候需要保证零滞后的商务智能（至少尽可能减少滞后性）

6. 非结构化数据基础上的商务智能