**3.7** **软硬件方案**

商业智能在电子商务行业中实施面临着较多的挑战和问题，如何选用合适的 软硬件产品来支持企业的 BI实施是一个重要的问题。

首先是处理海量数据的问题，随着业务的发展，电子商务企业每天都会产生巨 大的数据量，比如网站用户信息、网站访问日志等记录，这些数据少则数十万数百 万，多则上亿；如何选择合理的软件硬件方案来高效准确地处理这些数据是商业智 能成功实施的一个关键。商业智能在电子商务行业中会面对不同层次的用户，从



**第** **三** **章** **商** **业** **智** **能** **在** **电** **子** **商** **务** **的** **解** **决** **方** **案**

底层的操作人员到公司高层领导，还包括网站用户。不同层次和类型的用户会产 生大量的差异化的需求，用户对实时性、数据粒度、用户终端等各方面的需求各不 相同，因此BI 实施过程中需要选择合适的软件和硬件，才能保证数据顺利的从业 务系统、数据仓库、数据集市一直到最后的报表展现都能满足不同用户的需求。

电子商务行业的 BI 涉及的关键技术有数据仓库及 ETL 技术、OLAP 技术和数 据挖掘技术，市场上已经存在着许多厂商的BI产品以及一些开源的项目来支持这些 技术(表3.1)。BI产品种类繁多，每家厂商几乎都会宣称自己的产品是如何的好，在 构建企业自身的 BI系统时应该根据企业自身的需求考虑软件硬件产品的选择。

**表3.1** **主流** **BI 工具**

|  |  |  |  |
| --- | --- | --- | --- |
| 数据库 | ETL工具 | O LAP工具 | 数据挖掘工具 |
| DB2 | DataStage | BO | SAS EM |
| Oracle | PowerCenter | Cognos | Clementine |
| SQL server | SSIS | QilikVicw | Weka |
| Teradata  Greenplum | Datalntegrator |  |  |

BI产品主要涉及数据库、ETL 工具、OLAP 工具及数据挖掘工具，有的厂商有 很全的产品线，从数据库、ETL 到前端展现及数据挖掘都有涉及，能够提供全面的 解决方案，比如Cognos 和 Oracle;用户可以购买厂商的整体解决方案实现 BI。很 多BI 产品都是很优秀的，但是对于企业来说要考虑各方面问题，好的产品不一定 会适合自身。

39

数据库方面 DB2 、Oracle都是很好的产品，Teradata 和 EMC 公司的 Green- plum 在大数据处理方面有着很强的能力。根据 Gantner 在2012年2月发布的数 据仓库魔力象限[]显示 Teradata在数据仓库方面处于领导者地位， Teradata 在实 际的应用中也显示出其强大的数据处理能力，另外 Oracle 、IBM也处于领导者象 限，EMC 公司的 Greenplum 凭借其云计算方面的能力也进入领导者象限(图 3.21)。ETL 工具方面比较主流的产品有 DataStage 和 PowerCenter, 这两者占据 了国内主要市场份额，DataStage 在被 IBM 收购之后显得后劲十足。分布式部署 云计算是解决大数据最好的方法，Greenplum 在云计算方面表现出色，但是这是需 要付费的，当前主流的 Hadoop 也很流行，可以让企业以较低的硬件投入实现分布 式，当然这对开发人员的要求更高。

主流的 OLAP 工具有 BO 、Cognos 以及QlikView 等，要评价工具好不好除了 看数据准不准这类开发问题，更重要的就是要在考虑功能强大、开发简单之外，速 度、美观、简单必不可少；很多厂商在性能上有自己解决办法，但大部分对开发者设

labilily to execute H

 **商** **业** **智** **能** **在** **电** **子** **商** **务** **中** **的** **实** **践** **与** **应** **用**

challengers leaders

|  |  |
| --- | --- |
| 1010data | Teradat  看  \*Oracle  IBM  ●  ·EMC/Greenplum  Sybase an SAP Company  Microsoft |
| ParAccel Kognitio  SAND Technology  Infobright  \*Actian  ●Exaso | Vertica |

niche players visionaries

—completeness of vision I—

As of February 2012

**图3.21** **Gantner 数据仓库魔力象限**

计能力要求较高，比如Cognos 产品就要求编程实现，这样就需要开发者对产品非 常熟悉才可行。Gartner 前端报表工具最具魔力象限连续几年纳入了 QlikView 这 个产品，这个产品代表了基于内存的报表分析技术，美观简单，用户体验极佳。

数据挖掘产品领域，有 SAS 、SPSS 等两大厂家，而像 IBM 、Teradata 也都有

40

自己的挖掘工具。相对而言，SAS EM功能更全面，性能更加优秀，算法也更全面，

具有图形化界面，可视化操作。SPSS Clementine 开发更简便，功能上集成度更高， 其中 SAS 需要编程，对开发者的要求更高一些。做好数据挖掘仅仅有工具是不够 的，更需要数据挖掘人员对数据挖掘的理解和认识。