

设施农业决策系统的数据仓库建设

杨云 兰岳奇 (陕西科技大学计算机与信息工程学院, 陕西咸阳712081)

摘要 介绍了数据仓库的主要技术及其在设施农业中的应用和实现方法, 以帮助种植者快速得到最优种植工艺。

关键词 设施农业; 数据仓库; ETL; 元数据

中图分类号 S126 文献标识码 A 文章编号 0517-6611(2006)21-5750-02

Data Warehouse Building of Establishment Agriculture Decision-making System

YANG Yun et al (Department of Computer and Information, Shaanxi University of Science & Technology, Xi'an, Shaanxi 712081)

Abstract In the process of agriculture development, because of the lack of sorts of technologies, the agriculture couldn't be completely industrialized. We analyzed that the causation was the lack of supporting technology of agricultural production from beginning to end. In order to obtain the best excellent technique of agricultural production, we introduced data warehouse of technology in installation agriculture through the successful application of data warehouse in Business Intelligence.

Key words Installation agriculture; Data warehouse; ETL; Metadata

20世纪80年代以来,我国设施农业发展迅猛。考虑到设施农业的工业生产采用许多现代工业技术,如机械技术、工程技术、计算机与自动控制技术等,这些技术和系统必然产生各自的数据,为整个生产过程服务,这就需要一套决策系统,管理协调这些各自独立的系统。这个系统利用了数据仓库技术,为其建立统一、全面、准确的数据,可快速准确地做出决策。笔者结合设施农业的实例,叙述数据仓库的设计、数据的装载、元数据的管理。

1 设施农业

设施农业也被称为工厂化农业,是运用现代工业技术成果和方法、用工程建设的手段为农产品生产提供可以人为进行控制和调节的环境和条件,使植物和动物处于最佳的生长状态,使光、热、土地等资源得到充分利用,形成农产品的工业化生产和周年生产,更加有效地保证农产品的供应,提高农产品质量、生产规模和经济效益,促进农业现代化。

2 设施农业数据仓库的建设

项目目标是建立一个能为种植者提供准确快速的决策和种植工艺,必须有大量工艺信息的数据仓库作为知识库,然后在此基础上按设施农业业务主题建立数据集市,最后在集市上对各自分类进行分析和展现。以小麦种植为例,系统中的数据仓库框架如图1。

2.1 确定主题 考虑到数据仓库建立的过程性和长期性,一般的数据仓库建立都是在所有业务中确立一个可实现的业务循环,作为实现的主题。根据设施农业自身的特点,各个分析主题不外乎都是为了通过分析不同种植用户的需求,从存储有专家知识的数据仓库中,选取出最优工艺,指导工业化农业生产。如选小麦种植为分析主题,这个主题的目的是:根据种植者的时间、地域、土壤、品种、水分等实际的种植条件,分析小麦种植工艺,选择合适的种子、种植方法,仓储技术。其主题模型如图2。

这个模型描述了一般农作物种植的相关部分以及各部分的关系。如一个农作物对应多个农作物工艺,一个具体的品种也对应多个工艺,一个工艺对应多种种植技术和仓储技

术,这样就可以从这个模型中层层细化筛选得到最优的种植技术和仓储技术。

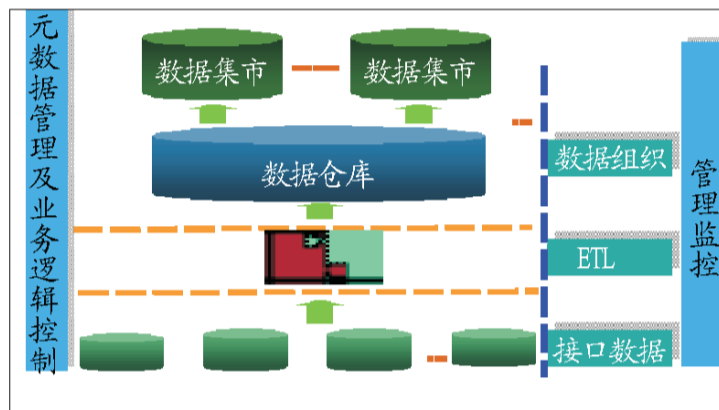


图1 数据仓库框架

2.2 建立数据模型 根据主题的逻辑模型,建立数据仓库数据模型。以主题模型中的实体为核心,结合种植小麦的工艺流程,建立工艺流程的业务逻辑。如小麦的选种、播种、种植、灌溉、防病虫害、施肥、收割、仓储等这些流程,对每个阶段进行细化建立业务数据模型;最后结合数据库理论建立各个层次的数据仓库数据模型,如接口层、事实层,集市层。

根据经验,在建立模型前,必须要有大量的相关工艺业务介入,考虑到整个设施农业生产,保证以后数据仓库能不断发展。还要考虑到不同层次的模型,对生产技术共性和差异性,考虑到这些因素建立的模型才能保证存储大量的农业种植技术,并且能减少数据库的冗余数据,提高数据性能。

笔者采用power designer建模工具,建立上面的数据模型,对接口层模型使用规范的处理,使模型满足第三范式,因为接口层有大量的细节数据,这样做可以减少数据冗余。对事实层和集市层采用星型结构,以主表为核心,维表为辅助,对主要技术指标进行汇总,以满足决策依据。

2.3 设施农业数据的装载(ETL) 由于设施农业的特点,必须包括许多农业科技知识才能支撑农业的自动化生产,而这些知识的存储方式又各式各样,必然使数据源多样化。笔者采用了CI1和DI2种实现方式将各种数据源整合成标准的数据接口,实现数据ETL装入到DB2数据库中。

CI1是一种访问数据库的接口,由一组API函数集组成,通过C++来封装这些函数,实现通过CI1对db2数据库数据的操作。使用C++高级语言丰富的数据结构和对内存的直接操作能力,就能满足复杂的种植工艺需求。包括从数

据仓库读取、转换、装入数据,异常处理。如:

```
Try// 连接数据库
{
rc = db2 - > Open( DBName , User Name , Password );
}
catch( ... )
{
WriteLog( 连接数据库出错. );
return ( - 1 );
}
```

```
if ( db2 - > IsOpen() )
{
WriteLog( 成功连接数据库 );
}
else
cout << not Connect . << endl ;
db2 - > ExecuteQuerySQL( sql_state , 4 ) ; // 执行SQL 语句
while ( db2 - > SQLBKFetch( recSet ) ) // 处理sql 语句返回
的结果集
{ 处理数据; }
```

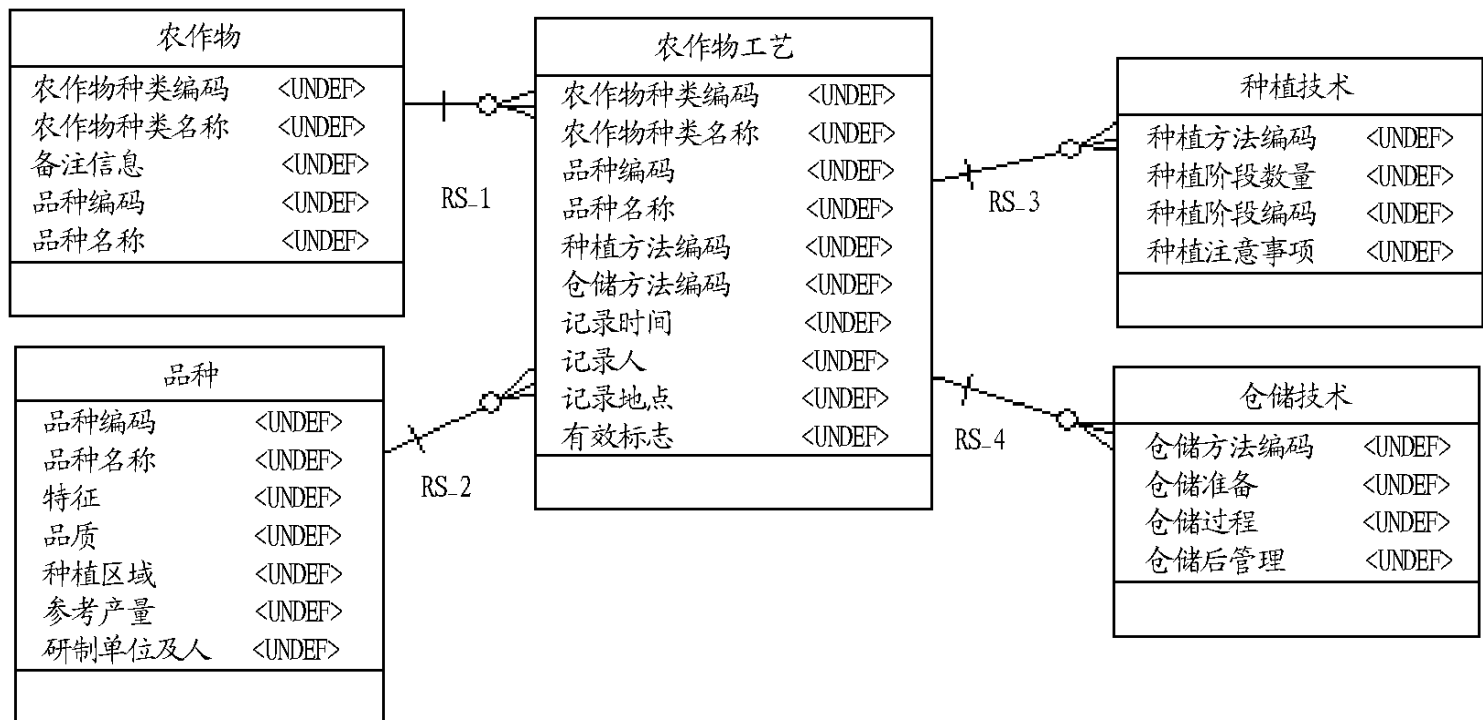


图2 主题模型

使用第3方的 EIL 工具 Business Objects Data Integrator, 建立连接数据源的 data store; 然后将业务需求分解到 workflow、job flow, 最后在 dataflow 中实现对数据的抽取和简单的转换; 并将最后处理结果写入目标数据库, 整个过程都是通过 DI 可视化开发工具实现的。

由于使用 C++ 语言 DB2 CLI 接口编程这种方式, 优点是编程语言强大的数据流程控制能力和丰富的数据结构, 再加上通过接口直接访问数据库, 致使这种方式能适用于复杂的处理流程和海量数据的处理。但其开发的效率差些, 所以笔者对复杂的业务数据抽取、转换使用这种方式, 如从接口表到数据集市和数据仓库的数据提取、转换。

第3方 EIL 工具 DI, 使用可视化的 DI 工具模块, 能快速地通过 ODBC 连接各种数据库和数据介质, 对数据源按批量处理, 并有专门的 web 界面管理 EIL 的调度和运行状态以及元数据信息, 使其开发效率比较高, 但是由于其使用 ODBC 来访问数据库和按批量处理, 使得这种方式只能处理不太复杂的业务数据需求。结合实际在从农业技术知识数据源到数据仓库的接口表的 EIL 使用这种方式。

2.4 元数据的管理 建立数据仓库一个重要的工作是元数据管理。元数据 (Metadata) 就是数据的数据, 用于建立、管理、维护和使用数据仓库。元数据管理是企业级数据仓库中的关键组件, 贯穿于建立数据仓库的整个过程。元数据使得用

户可以掌握数据的历史情况, 如数据从哪里来? 流通时间有多长? 更新频率是多大? 数据元素的含义是什么? 对它已经进行了哪些计算、转换和筛选等。通常把元数据分为技术元数据 (Technical Metadata) 和业务元数据 (Business Metadata)。技术元数据是描述关于数据仓库技术细节的数据, 这些元数据应用于开发、管理和维护数据仓库; 业务元数据在这里指的是种植工艺, 是从种植技术的角度描述数据仓库的数据, 并提供了良好的语义层定义, 业务元数据使业务人员能够更好地理解数据仓库分析出来的数据。

3 总结

以上介绍了数据仓库的主要技术和这些技术在设施农业中的应用及其实现方法。这个项目主要是为了用来帮助指导种植者, 快速得到最优种植工艺。该方法也可供其他行业建立决策系统参考。

参考文献

- [1] JOHN WILEY, SONS. 数据仓库 M. 北京: 机械工业出版社, 2003.
- [2] GEORGE BAKLARZ. DB2 UDB V8.1 for Linux, Unix, Windows 数据库管理 [M]. 北京: 机械工业出版社, 2003.
- [3] BRUCE ECKEL. C++ 编程思想 M. 北京: 机械工业出版社, 2004.
- [4] DAVID MARCO. 元数据仓储的构建与管理 M. 北京: 机械工业出版社, 2004.
- [5] IBM CO. Call level interface guide and reference [Z]. IBM CO, 1993.
- [5] IBM CO. IBM DB2 通用数据库 SQL 入门 [Z]. IBA CO, 1993.