**通用数据挖掘系统V3.0**

**概要设计说明书**

**Version 3.0**

**文档编号 TTS\*\*\***

**2012-3-31**

**达内IT 集团**

**文档历史**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **版本** | **标题** | **内容** | **作者** | **时间** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

编写目的

此概要设计文档描述DMS3.0项目的总体设计，在明确系统需求的基础上划分系统的功能模块，进行系统开发的分工，明确各模块间的接口，为进行后面的详细设计和实现作准备。文档的预期读者包括：项目设计人员，项目经理，程序员，测试人员，部署人员，项目技术支持人员等。

1. 参考文档
2. 《DMS3.0需求说明书》；文档编号：TTS\*\*\*；版本号v1.0
3. 名称解释

**DMS**：数据挖掘系统（Data Mining System）

**业务流程：**描述系统工作的总体过程。

**物理结构：**描述系统各功能模块与物理设备之间的联系。

**系统模块：**描述系统的功能模块。

**接口设计：**设计出系统与用户，系统与外部，系统各功能模块之间的依赖关系。

**用户接口：**描述系统与用户交互的依赖关系。

**外部接口：**描述系统与外部其他系统间的依赖关系。

**内部接口：描述系统内部各功能模块间的依赖关系。**

**用户：**需要使用Unix实验室出租业务的客户，用户通过电信公司提供的帐号和密码登录指定的Unix机器，并使用Unix服务。

**终端机器：**用户通过一台本地终端机器来访问我们的Unix机器，并使用Unix服务，该本地终端就是我们所说的终端机器。在我们的程序中要记录这一终端的IP，做为向用户收费的依据之一。

**登录日志文件：**在Unix机器中自动记录用户登入/登出日志的文件，该文件为：/var/adm/wtmpx；在我们的程序中该文件也可以是程序外部指定的另一个文件。我们的程序需要提取该文件的数据来匹配出用户登入/登出的记录，并将其作为将来的收费依据。

**用户登入记录：**在登录日志文件中记录的用户登入Unix系统的日志记录，该日志记录包含用户的登录名，用户登入Unix的时刻，登录的终端机器IP等信息。

**用户登出记录：**在登录日志文件中记录的用户登出Unix系统的日志记录，该日志记录包含用户的登录名，用户登出Unix的时刻，登录的终端机器IP等信息。我们需要将用户的登入记录和登出记录匹配起来，得到一条条完整的登录信息，并计算出用户本次登录的时长。

**用户的登录记录：**将用户的登入记录和登出记录按照对应关系匹配起来，形成一条完整的登录记录，登录记录中包含以下信息：用户登录名，登入时刻，登出时刻，登录时长，登录终端机器IP等。

**匹配的登入/登出记录：**等同于**用户的登录记录**。

**未匹配的登入记录：**在登录日志文件中理论上一条登入记录跟一条登出记录匹配，但有可能在我们采集数据时用户还没有登出，所以日志文件中可能出现没有登出记录与之匹配的登入记录，这些登入记录我们将其保存在一个指定的文件中，在下次采集时再读取。

1. 总体设计
   1. 业务流程

这个是项目的大体流程图，还没涉及到具体的细节。大体的流程如图所示，先读取数据即把数据从日志文件中读取出来，之后把放入内存集合中的数据进行匹配成需要的数据，把匹配好的集合发送到服务端。服务端接收数据，保存数据到数据库，数据进入数据库进行整合。



**说明：**

1，业务流程中几个存储文件的命名规则：

* 1. 系统日志文件由外部传递给程序，程序需要判定是否存在然后才能采集文件名为wtmpx。备份时添加时间后缀，如wtmpx.20170628085212
  2. 未匹配登入数据项集合的存储路径由外部传递给程序,文件名采用logins.dat。
  3. 未发送出去的匹配好的数据集合的存储路径由外部传递程序，文件名采用fail.dat。
  4. 使用文件DAO时，文件名称采用dms.dat。

1. 业务规则说明：

2.1日志登入/登出数据项读取规则

系统日志文件中的数据项有很多，但并不是所有数据项都需要采集的，只有满足如下条件的数据项才是需要采集的：

1. 登录名不能以点号（.）开头，以点号开头的一般是登录程序本身的日志数据项，不需要采集。
2. 登录类型必须是7(登入)或8(登出)，登录类型具体说明见下面列表：

|  |  |  |
| --- | --- | --- |
| **登录Type值** | **登录type值说明** | **C定义的宏** |
| 0 |  | #define EMPTY |
| 1 |  | #define RUN\_LVL |
| 2 |  | #define BOOT\_TIME |
| 3 |  | #define OLD\_TIME |
| 4 |  | #define NEW\_TIME |
| 5 | --初始化进程 | #define INIT\_PROCESS |
| 6 | --登录进程 | #define LOGIN\_PROCESS |
| 7 | --用户进程(登入) | #define USER\_PROCESS |
| 8 | --终止进程(登出) | #define DEAD\_PROCESS |

表2-1：登录类型定义

* 1. 系统模块(需要补充)

物理上系统分为采集客户端（被采集机器）和采集服务器两部分，在采集客户端包含数据采集模块，数据匹配分析模块和数据发送模块；在采集服务器端包含数据接收模块，数据存储模块和数据显示模块，各模块的主要功能和职责如下(需要完善)：

1. 数据采集模块：

该功能模块的主要任务是DMS客户端程序读取用户登录数据文件,对文件进行备份，并清空原始文件，读取上次未匹配的登入记录，将文件的信息解析为程序对象，既将用户登入/登出记录分别存在对应的集合中，以方便匹配模块的匹配。

1. 数据匹配和分析模块：

该模块的主要任务是将数据读取模块读取到的数据按照登入/登出的关系匹配成一条完整的登录记录。若本次匹配是上次未匹配的登入记录，则将其在文件中删除，若仍有有登入记录未匹配则将其存储在指定的文件中。

1. 数据发送模块：

该模块的主要任务是建立客户端网络连接，将匹配好的用户登录记录以字符串的形式发送到服务器。

1. 数据接收模块：

该模块的主要任务是建立服务器端网络连接，从客户端读取匹配好的用户登录记录字符串。

1. 数据存储模块：

该模块的主要任务是将数据接收模块接收到的数据交给DAO，由DAO将数据保存到指定的数据库表中。【DAO (Data Access Objects) 数据访问对象是第一个面向对象的接口】

1. 接口设计
   1. 用户接口

用户：执行命令，启动DMS3.0服务端程序。

系统：服务端系统运行，等待接收数据。

用户：执行命令，启动DMS3.0客户端程序。

系统：客户端系统运行，采集并发送数据。

* 1. 外部接口

系统读取用户登录记录文件，获得用户登录信息。

系统读取上次未匹配的登入记录文件，获取未匹配的登入记录数据。

系统将匹配好的数据写入指定数据库表中。

系统整合出月报表和年报表。

* 1. 内部接口

数据读取模块与数据匹配模块间的接口：

输入：登入/登出数据集

输出：匹配好的日志记录集

数据匹配模块与数据发送模块间的接口：

输入：匹配好的日志记录集

输出：若网络中断，产生发送失败的日志记录文件

数据发送模块与数据接收模块间的接口：

输入：日志记录数据

输出：服务器日志队列中产生数据

数据接收模块与数据保存模块间的接口：

输入：服务器日志队列中的数据

输出：DAO获取到的数据

数据保存模块与数据库间接口：

输入：DAO获取到的数据

输出：数据库表中的数据

数据整合模块与数据库间接口：

输入：数据库表中的日记录表

输出：形成日报表、周报表。月报表

1. 数据结构设计

根据系统需求，设计出如下数据实体：

1. LogRec：登入/登出记录类，用来表示一条原始的用户登入/登出记录。
2. MLogRec：匹配的登录记录类，用来表示一条匹配的用户登录记录。

以下分别对这两个实体类的设计作出说明：

* 1. 登入/登出记录数据类
     1. LogRec类图：



* + 1. LogRec类成员变量说明：

|  |  |  |
| --- | --- | --- |
| **成员变量定义** | **成员变量说明** | **备注** |
| **logname** | 用户登录名 | 匹配同一次登录会话的必须数据之一 |
| **pid** | 进程ID | 匹配同一次登录会话的必须数据之二 |
| **type** | 登录类型7-登入，8登出 | type的值在1-8之间，但只处理7与8两种情况 |
| **logtime** | 登录时刻\*//\*单位是秒 | 要采集的数据，logtime是登入或登出时刻 |
| **logip** | 登录IP | 要采集的数据 |

* 1. 匹配的用户登录记录类
     1. MLogRec类图：



* + 1. MLogRec类成员变量说明：

|  |  |  |
| --- | --- | --- |
| **成员变量定义** | **成员变量说明** | **备注** |
| **logname** | 用户登录名 |  |
| **logintime** | 登入时刻 |  |
| **logouttime** | 登出时刻 |  |
| **durations** | 登录时间长度 |  |
| **logip** | 登录终端的ip地址 |  |

1. 数据库设计

根据需求，设计出如下数据库表结构：

表结构：

用户登录记录明细表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **字段** | **类型** | **长度** | **约束** | **备注** |
| 登录记录编号 | Number |  | PRIMARY KEY | 系统自动生成 |
| 用户ID | Char | 32 | NOT NULL |  |
| 用户终端IP | Char | 32 | NOT NULL |  |
| 登入时间 | Char | 32 | NOT NULL |  |
| 登出时间 | Char | 32 | NOT NULL |  |
| 登陆时常 | Number | 32 | NOT NULL |  |