第一季 第二集

如何设计你自己的工作流程?



难度: 简单 时长: 35分钟

知识点

1、流程设计: Datist数据科学版安装和环境特点

2、节点用法: Datist数据科学版常用节点介绍

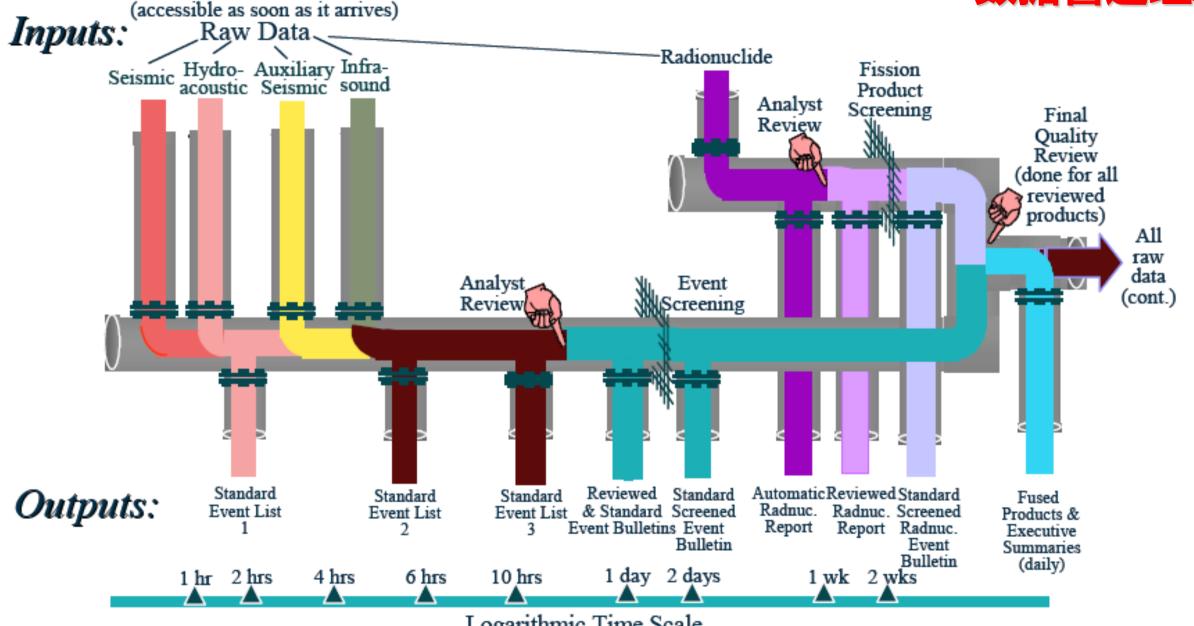
3、实例分析:数据绘图与可视化、数据统计与预处理、GIS与

地震目录分析



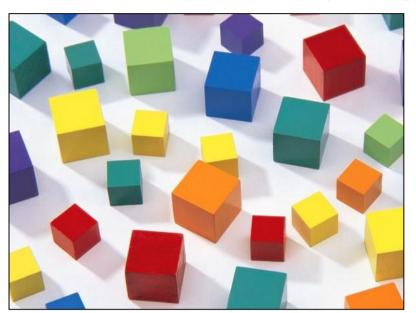


数据管道理念



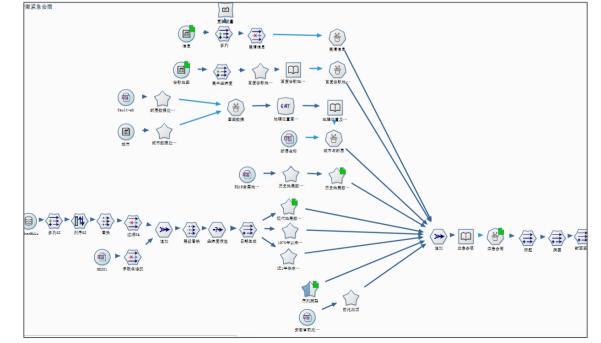
Logarithmic Time Scale

零编程、乐高式开发

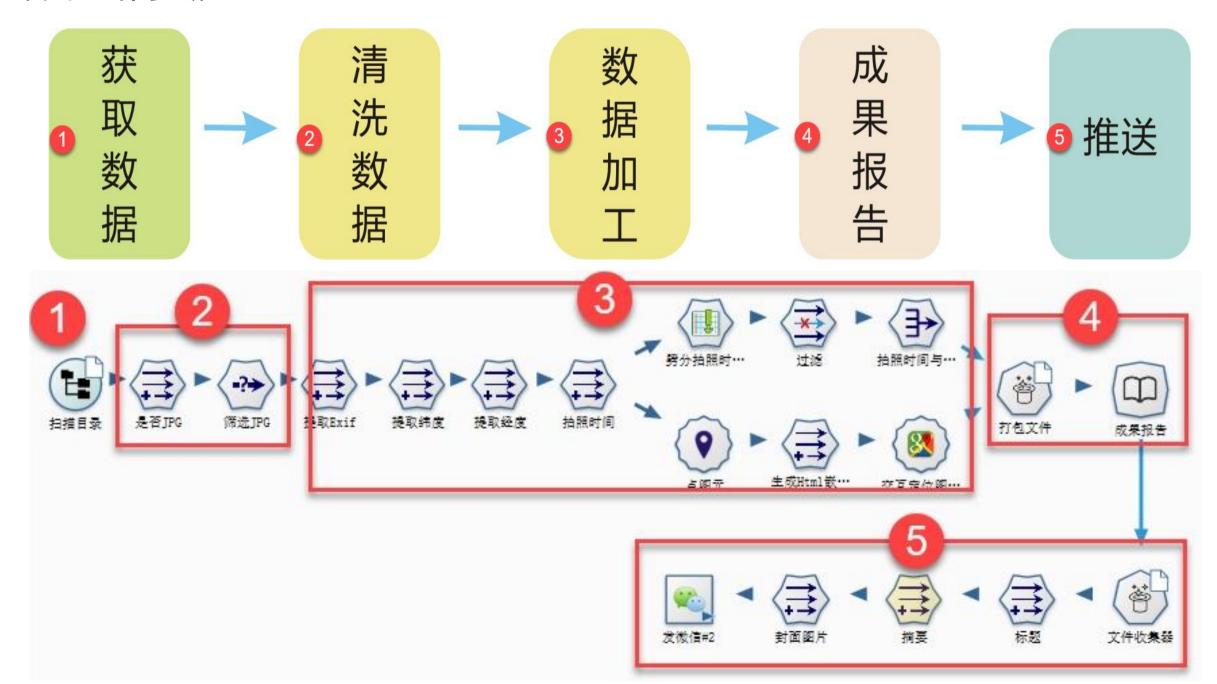








流程设计工作步骤



本集课程主要内容

| 常用节点 | 常用函数 | 实例分析 |
|--|--|---------------------------------|
| ■数据源、行列操作、 操作、数据库相关■图形、模型、输出、 软件接口、控制器■外部服务、场景设计 | ■数值计算类■字符处理类■逻辑运算类■数理统计类■日期与时间处理类■格式转换类 | ■数据绘图与可视化 ■数据统计与预处理 ■GIS与地震目录分析 |
| | | |

一、常用节点介绍

数据专家节点众多,这里主要讲解一些常用节点,同时在数据专家的工具箱中的每个常用节点我们都给出了相应的案例供模仿、学习,可以通过把鼠标放在节点上然后左键点击案例获取。





(一) 数据源节点

TXT节点

导入TXT文件,根据文本数据格式可选择单列读取,也可以通过设置分隔符号或固定宽度来劈分列数据。

- ① 与Excel相同,支持拖入。
- ② 同Excel。
- ③ 如有特殊编码格式可选择,一般问默认。
- ④ 将文本不分列, 读入到一个字段中。
- ⑤ 使用分隔符将文本分隔为多个字段。
- ⑥ 使用固定宽度分隔文本,支持自定义列宽。



(一) 数据源节点

数据源节点是数据专家获取数据的主要工具,对于不同的源数据,可选择数据库、Excel、TXT、Word、扫描目录、网页抓取及空间数据等节点连接外部数据源;也可以用数据录入、表格数据、流程穿越、先导流程等来连接内部数据源。节点位置在工具箱-数据源栏

接入数据库

该节点可接入各类常见的关系数据库,支持Oracle、MySql、Sql Server、Sqlite等15种类型。以MySql数据库为例,详细讲述数据库接入方法。

双击接入数据库节点,进入 节点属性界面,首先是连接界 面,用来设置连接数据库的参 数。

- ① 输入服务器名称
- ② 输入密码
- ③ 选择数据库



(一) 数据源节点

Excel节点

用于导入Excel数据表,可按序号或名称选择文件中的页签,同时可以在页签内截取某段数据。

- ① 选择文件路径,为使用方便数据专家支持文件拖入模式,把待读取的Excel文件直接拖入操作界面中。
- ② 根据页签的顺序读取指定页签,0为第一个页签。
- ③ 可直接根据页签名称选择指定页签。
- ④ 如第一行为字段标题则勾选,如第一行为数据不要勾选。
- ⑤ 根据需要截取Excel表中的部分数据。

同时可以在页签内截取某段数据。 把待读取的Excel文件直接拖入操作界面中。



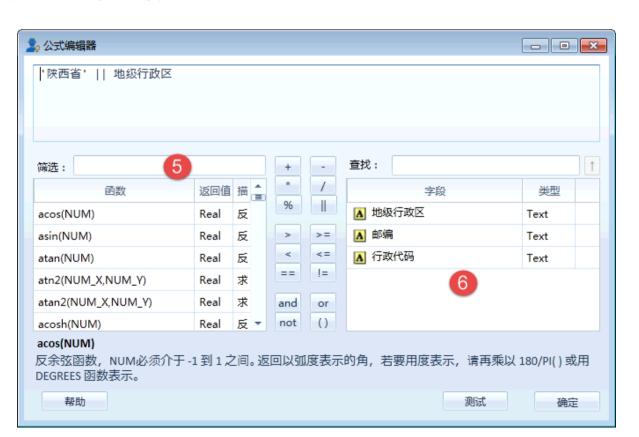
数据处理是数据专家的核心,虽然节点较多(包含列操作与行操作),但数据处理方式无非是节点+函数模式,充分利用这个模式就能把数据专家数据处理的功能无限放大。下面介绍一下常用节点,其他节点的使用方法可以参照这些常用节点。

▲ 列操作 ¢ 值偏离 A>B 字段累加 同义词 **₹** 补全列 向上取值 数据分栏 分栏合并 提取文本 打标签 (+;}∢ 解析Json



- 1、新列, 顾名思义就是增加一个新列, 可根据需求设定新列数据类型与值。
- ①设置新列的名称。
- ② 设置新列的赋值方式,有公式、条件及多条件三种方式,其中公式是无条件执行表达式,条件及多条件是在满足条件的情况下选择性的执行表达式。
 - ③ 设置新列数值类型。
 - ④表达式为新列赋值,其构成主要函数、前节点字段、流程变量、自定义值。点击E可弹出表达式编辑窗口。
 - ⑤ 函数查询出,支持字母及中文模糊查询(可根据功能关键字进行查询)。
 - ⑥添加前节点字段。





- 2、字段累加,增加新列,分组计算某个字段累加值或序号。
- ① 设置新字段名称。
- ② 设置计算类型,其中行号与序号为分组计算累加序号,字 多选。段累加为分组计算某字段的累加值。
 - ③ 选择分组字段名。
 - ④ 设置起始计算值。
 - ⑤ 选择进行累加计算的字段。



- 3、替换,在满足设定条件下,替换某字段的值。
- ① 选择待替换的字段,点击F进入字段选择界面,可 8选。
 - ② 输入判定条件,点击E进入公式编辑界面。
 - ③ 输入替换值,点击E进入编辑界面。

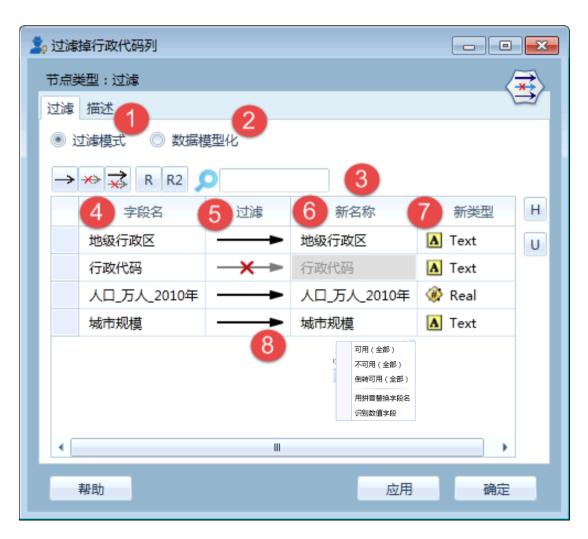


- 4、过滤,过滤掉多余的字段,修改字段名称和类型。
- ① 过滤模式,为常用模式,功能为过滤多余字段,修改字段名称和类型。
- ② 数据模型化,该模式是设定标准的字段名和类型,再设置已知字段与标准字段的对应关

系。

- ③ 批量选择过滤,查询待操作的字段。
- ④ 原字段名。
- ⑤ 选择过滤与否, 打叉的为过滤。
- ⑥ 设置新字段名,。
- ⑦设置新字段的类型。
- ⑧ 点击右键弹出选择菜单,

可根据需要进行批量操作。



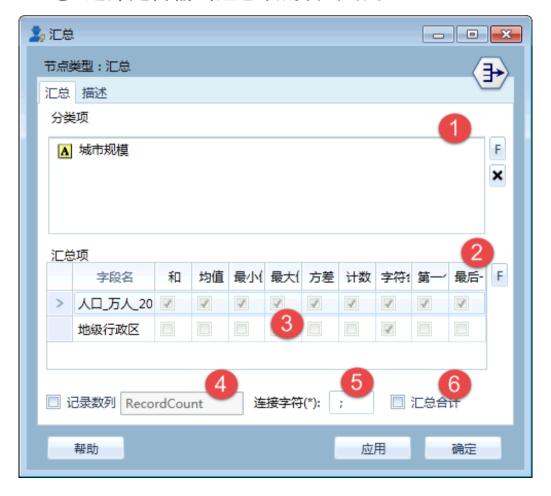
- 5、合并,将两个数据表按给定的字段值间规则合并为一个数据表。
- ① 设置输入项。
- ② 关键字设置,可选择同名模式和表达式模式。
- ③ 设置输出项,同过滤节点。
- ④ 通过鼠标左键拖动,选择输入节点顺序,第一个为主表,后面的为次表。
- ⑤ 同名字段模式,按相同字段名的字段值相等进行主表与次表的匹配。
- ⑥ 按自定义主表与次表关系讲行匹配。
- ⑦ 设置主表与次表匹配方式,完全匹配只保留能匹配的数据,左连接保留主表所有数据和次表匹配数据,排除是从主表中删除能匹配的数据。
 - ⑧ 表达式模式中设置匹配条件,点击E进入公式编辑模式 该节点比较复杂,使用前可参看案例。



- 6、追加,将多个数据表按给定的字段名合并为一个数据表。
- ① 设置输入项,同合并节点。
- ② 查看输出字段、主表与次表字段,设置输出字段。
- ③ 设置输出字段,可选择只保留主表列,也可选择包含主表与次表所有列。



- 7、汇总、分类统计多个字段的和、均值等统计数据。
- ① 选择分类字段,点击F,按字段对数据进行分类,可多选。
 - ② 点击F,选择汇总字段名。
 - ③ 选择汇总字段的统计项。
 - ④ 选择是否输出个分组的数据行数。
 - ⑤ 用于设置字符合并时的连着字符。
 - ⑥ 选择是否输出汇总项的合计结果。



(四) 地理信息节点

地理信息节点,是数据专家处理地理信息的工具,提供地理信息的可视化与图元运算。其中可视化节点包括:地理图、WebMap、高清影像;图元运算包括:投影变换、创建点图元、创建多边形、图元信息等等。

- 1、地理图,用于浏览和编辑各类地理信息图件。
- ①选取底图。
- ② 自定义图层名。
- ③ 用于图元分类。
- ④ 选择坐标类型,点坐标需加载经度和纬度坐标,任意图元加载图元字段。
- ⑤ 选择图元字段,如果是点坐标则选择经纬度字段。
- ⑥选择是否输出图元信息。
- ⑦选择是否打开新窗口。

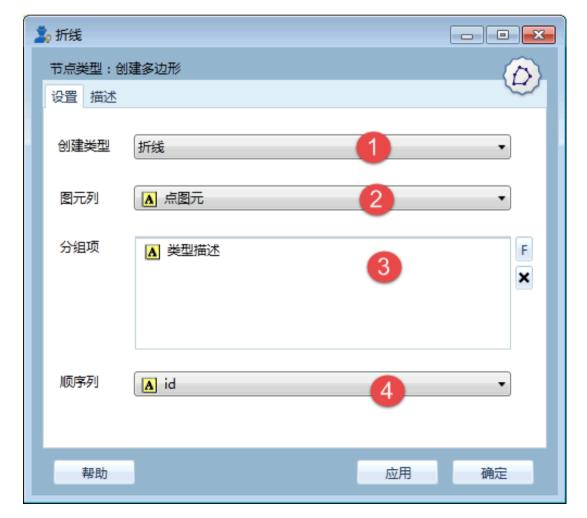


(四) 地理信息节点

- 2、创建点图元,根据经纬度数据创建一个点数 小凸多边形。据,point(xy)。 ② 导入点
 - ①设置新图元名称。
 - ② 导入原经纬度字段名。



- 3、创建多边形,由点图元创建线图元。
- ① 选择创建类型,创建类型包含: 折线, 多边形, 最 凸多边形。
- ② 导入点图元字段名。
- ③ 点击F, 选择分组字段, 每组创建一个图元, 可多选
- ④ 设置创建线图元的点之间的顺序。



(四)地理信息节点

- 4、缓冲区,以点为中心划一个预设半径的圆形区域。②选择图元名称(运行结果中可以查看)。
- ①选择点图元字段名。
- ② 设置距离值。
- ③根据个点的情况设定不同的距离。
- ④ 设置距离单位。



- 5、WebMap,将图元信息投影到百度或谷歌地图中。
- ① 地图类型,含百度地图、百度卫星等。
- ③选择投影图元数据。
- ④ 增加描述信息(运行结果中可以查看)。
- ⑤ 分组标记,方便标记图元颜色。
- ⑥ 根据图元面积大小确定百度、谷歌地图的显示级别



(四)地理信息节点

- 6、距离, 计算两点间的距离。
- ① 选择待计算的点图元字段名,主意点图元经纬度的顺序与格式。
 - ②设置距离单位及计算结果存放字段名。

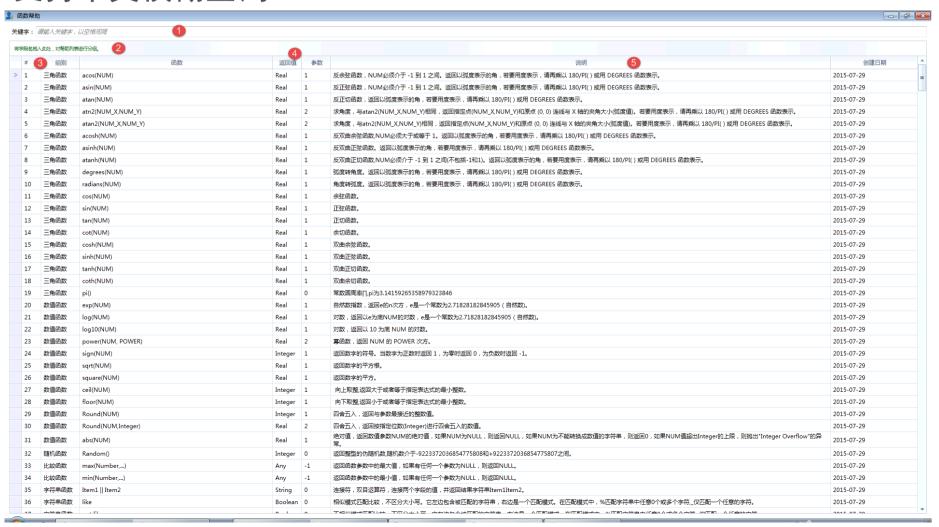


- 7、区块筛选, 筛选面积内是所有点数据。
- ①加载面图元。
- ② 加载点数据。



三、常用函数介绍

函数作为数据专家的重要组成部分,它以节点为载体,如果说节点是数据专家的骨架,那么函数就是数据专家的血液,它承载着数据处理与转换职责。数据专家函数数据已超过600个,如何选择合乎需求的函数是流程编制的关键。数据专家为开发者提供了函数查询功能,支持中文模糊查询。

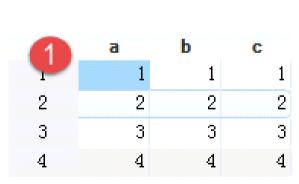


(二) 数值计算类函数

数值计算是数据处理软件最基础的功能,数据专家通过数值计算函数来实现数据的计算功能。

数值计算函数主要包含三角函数和数值函数两大类,主要功能为计算三角函数、加减乘除、余数、保留小数位、指数、对数、乘幂等等涵盖了所有数值基本运算需求。数值函数用法比较简单,下面以power()函数在新列节点中的编写为例,介绍函数在节点表达式中的编写方式。

- ① 原数据。
- ② 数值计算函数使用时要特别注意前节点字段的数据类型和计算结果储层字段的数据类型,如不清楚可以参考函数帮助。
 - ③ 编写表达式,按照函数格式,添加字段名、变量或数值。
 - ④ 函数计算结果。





| | a | Ь | С | 新列 |
|---|---|---|---|-------|
| | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 16 |
| 3 | 3 | 3 | 3 | 729 |
| 4 | 4 | 4 | 4 | 65536 |

(三) 逻辑运算类函数

广义的讲逻辑运算<mark>包含所有的条件语句函数和返回值为Boolean型的函数</mark>,它用于处理数据间的逻辑关系。首先介绍Boolean型函数,其功能通俗的讲就是判断一件事是真还是假,待判断的事件可以是数值类,也可以是字符串、地理信息、日期类等等。通常在涉及条件判断的节点中使用,例如:筛选、替换、新列(条件)等节点。

1、数值类

对于数值类的事件通常是要判断两个或多个数值之间的关系,比如大于、小于、相等、不相等、in、not in等等,通过函数帮助里的比较函数就能实现这些基本功能。

2、字符串类

字符串类Boolean函数种类较多,但功能大致可以分以下两种:

① 判断某个字符串是否具有某种特征。

判断字符串是否以某个子字符串结束: EndsWith、HasEndString。

判断字符串是否以某个子字符串开始: StartsWith、HasStartString。

判断字符串是包含某个子字符串: HasSubString、HasSubStringsOR、HasSubStringsAnd。

字符串定位函数:包含单词IndexOf,判断字符串某个位置是否为某个字符。

判断字符串是包含某个子字符串但不以它开始或结束: HasMidString。

② 判断两个或多个字符串之间是否存在某种关系。

判断两个字符串间的数字字母顺序: AlphaBefore。

判断两个或多个字符串的匹配关系: IsMatch。

(三) 逻辑运算类函数

3、地理信息

地理信息类Boolean函数主要是判断两个图元间是否存在某种关系,其中包括点是否在多边形内、面是否在多边形内以及两图元是否存在包含、相交、相离、重叠、接触等关系。

4、信息函数类

信息类Boolean函数主要是判断值类型,如判断值是否全为字母IsAlpha、是否全为汉字IsChinese、是否为某一固定的时间格式ISDatetime、是否为整数IsInteger、是否为数值型IsNumber、是否为实数IsReal。

5、日期与时间类

日期与时间类Boolean函数主要功能是判断两时间是否相等(DatetimeEqual)、一个时间是否在另一个之前或者之后(DatetimeBefore、DatetimeAfter、TimeBefore、TimeAfter)。

如何准确的找到所需的Boolean型函数,可以在公式编辑器筛选框输入该函数单词的开始部分,也可以根据功能关键字输入中文筛选,也可以进入函数帮助界面通过返回值类型分类后选择Boolean型函数进行查看。

(四) 字符处理类函数

字符处理函数种类繁多,这里只介绍一些常用的函数,其他的可以通过上述的查询功能选择应用。

- 1、字符串连接 | | , 连接多个字符串。
- ①为字段名,②为字符串,③为流程变量。下图展示了字符连接||应用时,字段、字符串、流程变量的连接方式。JionItems也具有连接字符串的功能。
- 2、字符串截取类

截取左边N个字符LeftStr。

截取右边N个字符RightStr。

截取字符串第N-M(两个子字符串)之间的字符串 SubStrBetween(S)。

3、字符串删除类

- ①以Remove开头的字符串函数都为从字符串中删除相应的子字符串,具体参考函数帮助。
 - ②Itrim、rtrim、rime删除空格的函数。

4、字符串插入类

向字符串指定位置插入字符串: InserString

- 5、字符串长度相关类
- 以包含Length字符的函数。
- 6、字符串替换函数

以replace开头的函数,替换字符串中的某个子串。

7、字符串补全函数

PadI和Padr为字符串补全函数,分别是从左边和右边把字符串的长度补全为N。



(五) 数理统计类函数

数理统计是大数据分析的基础,数据专家除了提供一些常用的统计函数如:平均值、最大值、最小值、 计数、求和、分位数、标准偏差、方差等等,还提供了峰度系数、相关系数、Skew扭曲等函数。具体使 用方法与功能请参看函数帮助中的统计类。

(六)日期与时间处理类函数

数据专家函数帮助中日期与时间函数,除了上面讲述的逻辑与转换函数外,还提供的大量的时间处理相关函数,概况为一下几类。

1、时间运算函数

① 时间与数值的运算

将指定的数加到时间字段上,它们都add开头的,函数后缀不同功能不一,例如AddDays将指定的天数加到Datetime上,AddYears将指定的年数加的Datetime上等等,其中指定数可以为负数。

② 计算两个时间点之间的间隔

这些函数都包含关键字difference,根据需求不同,可计算间隔的年、月、日、时、分、秒以及完整模式的间隔。 其中以Time开头的函数返回值为整数,使用时应注意返回值类型。

2、获取当前时间

Now函数为获取当前完整时间,包括年月日时分秒; year只获取当前年时间; month只获取当前月时间等等以此类推。

3、获取日期的序数

获取日期为该年、月、周中的第多少天,day、dayofyear、dayofmonth、dayofweek,同时可选择中英文的不同表达模式。

(七) 格式转换类函数

格式转换是数据处理中必不可少的环节、数据专家函数中格式转换大致分为以下几类。

1、数值型和字符型间的相互转换

字符型转数字型ToInteger、ToReal等,数值型转字符型常用ToString。

- 2、转换为List格式
- 一些函数可以把字符串转换为相应的数列形式,如ToIntegerlist,把字符串转换为整数数列;ToDoublelist把字符串转换为实数数列,等等这些包含To和list的函数都具备这样类似的功能,使用时可以通过这两个关键字精细搜索。
 - 3、大地坐标间的相互转换

包含经纬坐标数字型和度间的相互转换,类似投影转换节点中的各种坐标格式的转换,简单的图元间的相互转换。

4、字符串内部转换

主要是根据需要转换字符串的表达方式,如ToChineseMoney将数字转换为人民币大写; ToDBC字符串转换为全角; TOSBC全角转半角; Topinyin将汉字转换为拼音; ToPinyinFirstLetter汉字转换为拼音首字母。

5、日期格式间相互转换

ToChineseCalendar日期转换为农历、ToDatetime字符串转日期、还有ToOadate、toShortDate、toShortTime、To_UnixTime,以及儒略日的转换等等。

6、日期转换为字符串

除了ToString函数外,数据专家还提供了format开头的函数数将日期转换为固定形式的文本。

7、base64与其它格式间的相互转换

在关键字搜索框中输入base,根据需求选择所需的格式转换函数。

其它格式转换函数可以阅读数据专家函数帮助中转换函数字段,使用时主要被转换值是否与函数说明中的一致。

如何从网上获取帮助: datist.readthedocs.io



用流程使数据畅通,构建预报业务新生态

