

按列存储的配电网监测数据包区间编码正规化压缩处理

屈志坚¹,徐振清²,周锐霖¹,朱 丹¹

(1. 华东交通大学 电气工程学院,江西 南昌 330013;

2. 中车青岛四方机车车辆股份有限公司,山东 青岛 266000)

摘要:针对不断增长的配电网按行存储数据的模式造成数据体量持续增大、压缩率难以提高的问题,利用按列存储的数据包架构和扩展区间编码的正规化处理技巧,提出了一种大体量配电网监测信息的压缩处理新方法。以列数据包为底层存储、知识网格为组织架构,将区间编码和正规化处理融入底层监测数据包的并行压缩处理进程中。以北京某动车段的配电网调度监测工程数据为算例,进行列数据包区间编码的正规化压缩测试。仿真结果表明:扩展区间编码的正规化处理技术可使监测信息压缩率提高至94%,获得了比已有云计算集群更好的压缩效果,通过动态调整映射区间,即可方便实现在无穷区间中对海量监测信息编码压缩的工程应用需求。

关键词:配电网;监测数据;列压缩;区间编码;正规化处理;数据处理;数据压缩;信息处理

中图分类号:TM 761

文献标识码:A

DOI:10.16081/j.issn.1006-6047.2018.03.008

0 引言

随着智能配电网技术的发展和大量配电自动化装置的接入,配电调度自动化系统的容量及规模不断扩大,使配电自动化海量历史监测信息达到了前所未有的数量级^[1]。该类型数据对处理速度要求不高,却给数据库存储带来极大的压力,海量监控信息处理的I/O密集,使数据库服务器工作负载重^[2]。因此,如何实现海量配电网监测信息大比率压缩,降低配电监测数据存储体量,从而减轻服务器存储压力,已成为智能调度监测海量信息处理领域最为重要的研究方向^[3]。

目前使用较广泛的开放内存数据库 fastdb 和非关系型 NoSQL (Not only SQL) 内存数据库 mongodb 均以 2 GByte 容量为限^[4-5],SQLite 内存数据库则限制在 100 GByte 内^[6-9],在存储海量配电网监测信息时受到一定的制约。磁盘数据库按照内部数据存储架构主要分为行式 SQL 数据库和列式 NoSQL 数据库,行式数据库以记录为存储单元进行空间分配和数据处理,记录数据作为一个整体,更适用于在线交易性的 OLTP 应用^[10],以 Oracle、MySQL 为代表的关系数据库通过创建索引进行快速查询^[11]。然而,对

于海量配电网监控数据场景而言,维护索引的成本开销也在成倍增长,维护索引占用的磁盘空间往往比数据本身要多3倍以上,与关系数据库存储容量在 TByte 级以内的现实情况相矛盾^[12]。基于行存储的有损压缩的压缩率虽然较高,但解压后数据恢复精度受限,可能造成关键信息的漏报和误报。基于云计算集群的分布式无损压缩解决了有损压缩导致的数据失真问题,并在一定程度上提高了无损压缩的压缩比,但该方法的性能优化空间有限,亟需寻找一种新的无损压缩替代方案^[13]。伴随 NoSQL 的兴起,以列为单位组织数据存储和处理的列式数据库逐步成为新型的海量数据管理方式^[14],按列存储处理的明显优势是:同一列属性具有相同的数据类型和值域、规律性强、数据重复度高的特点,原生利于高效压缩的实现,适合海量存储、数据分析和商业智能等应用^[15-16]。本文基于列式存储和数据包 DP (Data Pack) 存储模式,对配电自动化系统中海量监测数据进行了压缩存储研究,提出一种扩展区间编码的正规化处理海量配电监测 DP 的按列压缩存储技术^[17-19],关于此领域的研究目前尚鲜有文献报道。

通过搭建海量配电网监控数据列式存储平台,在 DP 知识网格主架构下,以实际配电网监测工程系统数据为算例,对配电网监测信息进行列式压缩处理实验,获得的海量配电网监测大数据集的整体压缩比率可达约94%。列式压缩处理的存储架构保证各存储单元的监测数据有很高的相似性和重复度,对发挥 DP 区间编码和正规化压缩处理的效率相当有效。

1 配电调度监控信息的列存储模式

1.1 按列存储监控信息的模式

按照 DP、数据包节点 DPN(Data Pack Node)、知识节点 KN(Knowledge Node)三层结构架构配电监

收稿日期:2017-10-07;修回日期:2018-01-17

基金项目:国家自然科学基金资助项目(51567008);江西省杰出青年人才计划项目(20162BCB23045);江西省自然科学基金资助项目(20161BAB206156,20171BAB206044);江西省教育厅科技研究项目(GJJ160471)

Project supported by the National Natural Science Foundation of China(51567008), the Foundation Plan for Distinguished Young Scholars in Jiangxi Province(20162BCB23045), the Natural Science Foundation of Jiangxi Province(20161BAB206156, 20171BAB206044) and the Science and Technology Research Project of Jiangxi Provincial Education Department(GJJ160471)

测数据存储模式。由于按列分割的 DP 数据具有规律性强、相似性大的特点,原生便于压缩,因此将 DP 存储于最底层,以监控信息每列数据中的 64 KByte 列值为单元。DPN 和 DP 之间建立一一对应的关系,在监控数据加载时随 DP 创建,DPN 一对一记录监控信息的量测最值、均值、单元总计数等,对应 DP 内存储、压缩统计及聚集数据等的元数据。利用 KN 存储描述 DP 之间的依赖关系、分布状态统计信息、列间信息和表间关系的元数据集合,例如监控量测信息的数值范围、字符位置以及列 DP 之间的关联。大量 DPN 和 KN 一起构成知识网格 KG (Knowledge Grid)。

监控信息的知识网格列存储模式如图 1 所示,其中监测对象映射表如表 1 所示。将某 10 kV 配电网低压侧母线 1 和 2 负荷的监测信息融入三层列存储结构,将电压互感器和电流互感器信号经远程终端单元 (RTU) 采集传输处理后的遥测值按 V001、V002、…、V00M 列组织存储。

表 1 监测对象映射表

Table 1 Mapping of monitoring objects

对象 ID	对象属性	采样值	阈值/%	归属站所
V001	电压	220 V	5	BJNZ
V002	电流	7 A	5	BJNZ
V003	电流	12 A	5	BJNZ
⋮	⋮	⋮	⋮	⋮
V006	电压	220 V	5	BJNZ
⋮	⋮	⋮	⋮	⋮
V00M	电流	9 A	5	BJNZ

经量测值前置处理模块,由对象映射表按照时间序列加载进程,调通接口类与列处理引擎的 SQL 层连接。启动数据引擎和执行器,将配电网监测信息按对象类,以每列对象 64 KByte 为单位分割成多块 DP,生成与每块 DP 对应的 DPN 用于记录 DP 统计的元信息,记录包含 DP 间依赖、列信息等高级元数据的 KN,构建以 DP 为 RTU 监测信息底层存储、DPN 和 KN 为表信息元数据集合的配电网遥测值三层列存储架构。对 DP 划分区间,进行并行压缩编码。

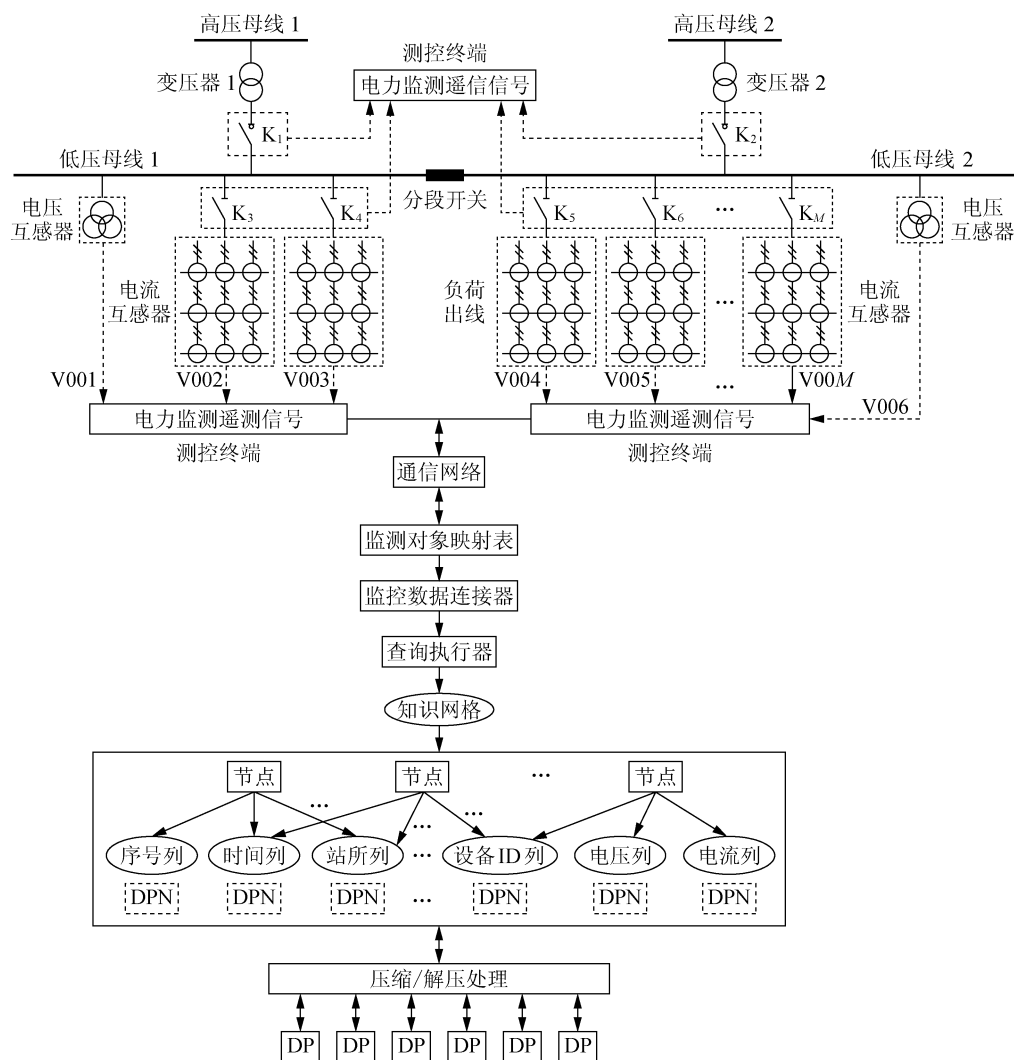


图 1 监控信息的知识网格列存储模式

Fig.1 KG-column storage mode of monitoring information

1.2 按列存储 DP

高数据压缩比主要依赖列存储方式和强大的压缩方法,每列数据分块压缩存放,每个压缩数据块由知识网格节点记录块内统计信息,将 RTU 遥测信号按三层知识网格架构存储,以配电自动化调度监测信息为例,每个监测点配置序列号、采样时间、站所、设备标识、电压值、电流值等不同类型的列存储信息,在底层存储上,按照以列为属性的方式,每列均以 64 KByte 为单位分解的 DP 形式单独存储数据。时间序列采集的配电网监测数据的按列存储及其底层 DP 压缩存储方式如图 2 所示。

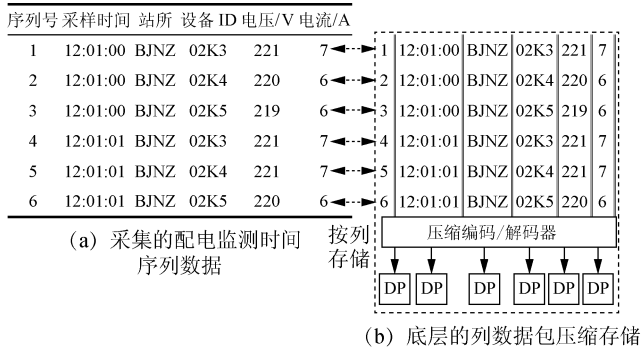


图 2 按列存储的配电网监测数据及其底层 DP 压缩存储方式

Fig.2 Distribution network monitoring data with column storage and its compress storage of bottom DPs

按列存储是指将归属于同一属性类的信息连续存放,首先将所有的采样时间类连续存放,然后依次将站所、设备标识、电压类等连续存放,依此类推,直至监测信息全部存放完毕为止。传统的行存储数据库的压缩是对监测数据表和记录进行统一编码,各属性间数据类型存在差异,决定了其压缩率很难提高,而将监控信息按列存储,使各属性类的数据重复度非常高,因而有望大幅提高整体的压缩率。

2 按列压缩的区间编码正规化处理

2.1 监测数据的区间压缩编码原理

首先将输入的监测数据信源进行区间化分割,分为高、低位部分^[20],对分割数据进行概率统计,将高、低位数据进行区间编码和合并存储,如图 3 所示。

相比于一般算术编码按比特位定义模型符号范围,区间编码模型的符号范围是任意的,可动态调整区间范围,适用于大规模数据压缩。将输入的监控信息列数据分割为高、低位数据块,并行执行区间编码,产生更少的固定比特位、数据最大值的统计符号个数以及更低的编码区间划分频率,从而提高压缩性能。

通过统计待编码序列的概率分布,给定一个足

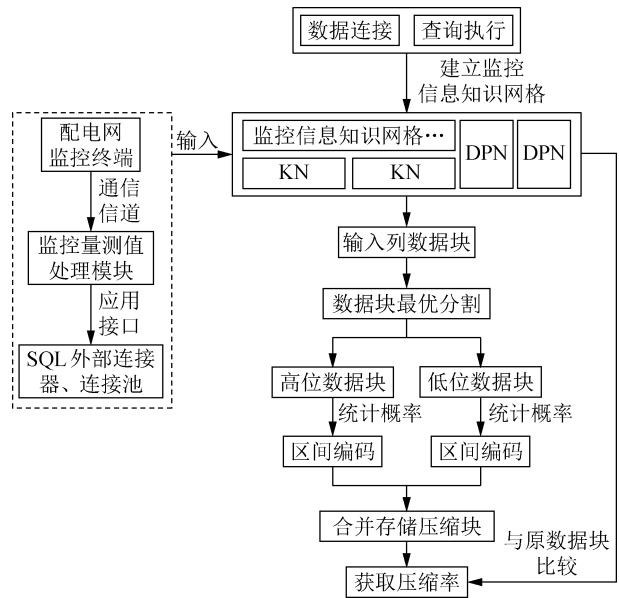


图 3 监控信息列式区间编码压缩处理流程

Fig.3 Compression process of monitoring information with column range encoding

够大的范围,在此范围内划分按符号概率分布的子范围,不断地迭代子范围,即编码当前符号后,将当前符号所处子范围作为下一个符号编码的总范围,最后用子范围内的某个数表示整个码流。整体而言,区间压缩编码是将所有数据映射到某个整数区间内,最后以属于该区间的某个整数作为编码输出。监控信息压缩编码原理如图 4 所示。

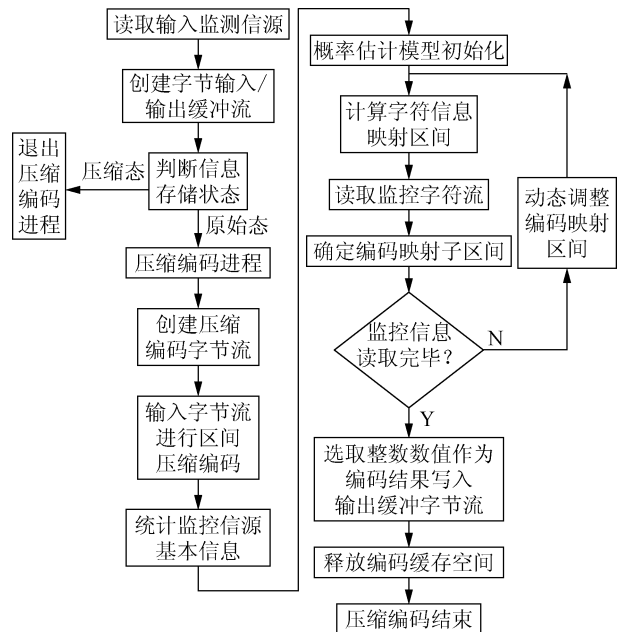


图 4 监控信息区间编码流程

Fig.4 Flowchart of monitoring information range encoding

以字节为单位读取输入监控信息字节流,即监控信源符号 S ,统计待压缩信源字符种类数 N ,以 i

表示 S 中的第 i 种字符, $1 \leq i \leq N$ 。编码区间设为 $[X, Y]$, 以 b^w 表示任意整数, X 与 Y 分别为区间的下沿和上沿, 区间范围 $R = Y - X + 1$ 。

令 f_s 为信源符号 S 的频度, T 为所有符号的总统计频度, 则 $T = \sum_{i=1}^N f_i$ 。 F_s 为符号 S 的累积频度, 满足累积分布函数 $F_s = \sum_{x \leq S} f_s$ 。

令所有符号初始频度都为 1, 初始区间的范围为 $R = b^w$, 然后根据每个符号的频度 f_s 、累积频度 F_s 及总频度 T , 按区间映射计算公式(1)~(3)计算符号信息在区间 $[X, Y]$ 中的映射子区间范围 R' 及映射子区间 $[X', Y']$ 。

$$R' = (R \div T) \times f_s \quad (1)$$

$$X' = X + (R \div T) \times F_s \quad (2)$$

$$Y' = X + (R \div T) \times (F_s + f_s) - 1 = X' + R' - 1 \quad (3)$$

其中, $\div$ 表示整除运算。

将当前字符频度加 1, 以 $[X', Y']$ 作为下一个符号的编码区间, 重复上述步骤计算下一符号映射子区间。不断地更新概率模型, 自适应地动态调整映射子区间, 编码后输出某个位于区间 $[X', Y']$ 内的整数值 V , 由于任意整数可以分解为基数幂的形式。因此将 V 分解成 $V = V' \times b^n$ 的形式, 将 b^n 部分省略 (其中 b, n 为实数), 只输出 V' 作为编码结果, 从而进一步减少冗余。

2.2 正规化处理策略

编码处理海量配电网监测信息时, 随着编码长度不断增加, 映射子区间范围会越来越小, 最终趋于 0。而若要满足海量监控字符对映射子区间范围的需求, 初始映射区间的范围需趋于无穷大, 正规化处理是指在一段有限长度的区间段内, 实现无限长区间的编码运算过程, 关键技巧是利用区间扩展策略, 当区间状态满足一定条件时, 利用区间扩展方法对映射区间进行重定义。

首先将有限区间 $[X, Y]$ 的运算限制在一个较小的区间 $[R_{\min}, R_{\max}]$ 内, 定义 $R_{\min} = b^{w-1}$, $R_{\max} = b^w$ 。将该区间中的数字分为 3 个部分: 不变数字 c 部分表示相同的区间上下沿最高位、 $(d+1) \times b^{n-1} - 1$ 部分和 $(d+1) \times b^{n-1}$ 部分, 其中 d 为延迟数字, n 为延迟长度, c 和 d 都可移出区间。

区间扩展的实现思路是将区间 $[X, Y]$ 转换为 $\{c, (d, n), [X'', Y'']\}$ 形式, 在区间 $[X'', Y'']$ 中模拟无限区间内的运算, 使用技巧是每当区间满足特定条件时, 便将数字从区间中移出, 在区间压缩编码过程中对海量配电网监控信息融合使用正规化处理的流程, 如图 5 所示。

正规化压缩编码监测信息的步骤如下。

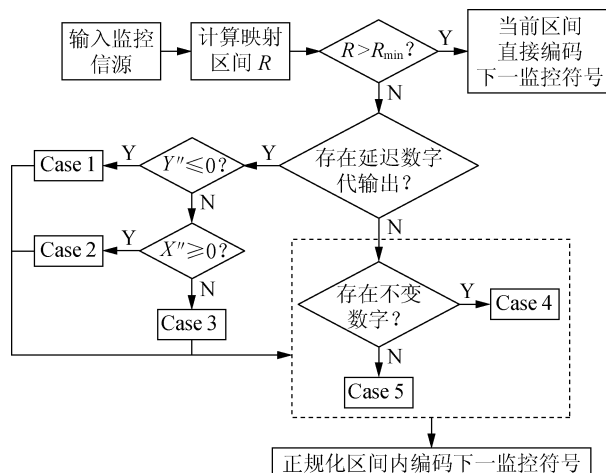


图5 监控信息区间压缩编码的正规化

Fig.5 Normalization of monitoring information range compression encoding

- 按列读取配电网 SCADA 系统的监测信息。
- 统计监测信息基本参数并初始化概率模型。
- 利用区间映射计算公式确定该监测符号的映射区间。

d. 判断 $R > R_{\min}$ 是否为真, 若为真则无需正规化, 在当前 R 区间内编码下一监测符号; 若为假, 则开始正规化处理, 执行步骤 e。

e. 若当前延迟长度 $n \neq 0$, 则存在延迟数字待输出, 执行步骤 f; 否则, 执行步骤 g。

f. 判断当前映射区间上沿 $Y'' \leq 0$ 是否为真, 若为真则执行 Case 1; 否则, 判断当前映射区间下沿 $X'' \geq 0$ 是否为真, 若为真则执行 Case 2, 否则执行 Case 3。执行步骤 f 处理完待延迟数字后, 同样需执行步骤 g 的流程步骤。

g. 判断是否有不变数字, 若有则执行 Case 4; 否则执行 Case 5。

Case 1: $Y'' \leq 0$, 先输出 d , 然后输出 $n-1$ 个数字 $b-1$ 。

Case 2: $X'' \geq 0$, 先输出 $d+1$, 然后输出 $n-1$ 个数字 0。

Case 3: 继续保留延迟数字, 延迟长度变为 $n+1$, 抛弃上、下沿最高位, 执行 Case 4 扩展公式。

Case 4: 将不变数字移出区间并输出, 同时利用区间扩展公式重新定义映射区间:

$$X'' = (X'' \times b) \bmod R_{\max} \quad (4)$$

$$Y'' = (Y'' \times b + b - 1) \bmod R_{\max} \quad (5)$$

其中, $\bmod$ 为取余运算。

Case 5: 记录下沿最高位数字延迟待输出, 待下次正规化时执行步骤 f, 同时依照式(6)计算扩展压缩编码区间。

$$X'' = [(X'' - R_{\min}) \times b] \bmod R_{\max} \quad (6)$$

每编码完一个监测信源符号,均进行正规化判断,正规化处理时先处理延迟数字,再检查是否有不变数字。

3 区间编码的正规化压缩处理过程

为了详述融合正规化处理的区间编码法在调度监控压缩中的应用实现,以配电网监控遥控(YK)、遥信(YX)、遥测(YC)、遥调(YT)类型及其资源标识符“YC03IaYC03IaYX00YC03IaYX00”为示例,阐述区间编码正规化压缩的全过程。为了便于后续计算和十进制直观表示,压缩前设 $b=10, w=3, R_{\max}=1\,000, R_{\min}=100$, 初始区间范围 $R=R_{\max}$ 。监控信息区间压缩融合正规化的映射子区间动态变化及编码处理过程,如表 2 所示。

当压缩编码区间较小时,按公式步骤输出不变数字和延迟数字,并扩展压缩编码区间。其中扩展区间输出序列结果如表 2 中正规化调整栏所示的 110834,由编码规则可知,其中编码区间内选取 900 为代表,即 $900=9\times 10^2$,因此,最后可输出编码结果 1108349,这样便方便地实现了以短整数表示长监测信源,可大幅减少存储空间的消耗。

4 工程算例的测试

4.1 列式压缩编码验证

以北京某动车段 10 kV 配电网 SCADA 实测监控信息为对象输入信源,进行列式存储压缩率的对比测试。监控采样信息由序列号、采样时间、站所 ID、监控设备 ID 以及包括电压、电流、功率等电气量测值组成。执行 SQL 命令“LOAD DATA INFILE”依次将 SCADA 记录集 $\{5\times 10^5, 10^6, 2\times 10^6, 5\times 10^6, 8\times 10^6, 10^7\}$ 加载进行区间编码和正规化压缩处理,观测千万级监控信息的关键代码为:

```
USE information_schema; // 选元信息数据库
```

```
SELECT ENGINE, ROW_FORMAT, TABLE_ROWS,
TABLE_COMMENT FROM TABLES WHERE TABLE_NAME =
'test1000w'; // 取测试信息
```

统计实测监控信息列式区间编码压缩处理测试结果,如表 3 所示。

结果表明:用原始态数据与压缩态数据大小之比表征的列式区间编码正规化处理监控信息的压缩率可达到 17:1 左右。

表 3 监测信息列式编码压缩存储测试

Table 3 Compression storage test of monitoring information with column range encoding

存储模式	存储状态	记录数统计	压缩率
列式	压缩态	500 000	17.082 : 1
列式	压缩态	1 000 000	16.868 : 1
列式	压缩态	2 000 000	16.849 : 1
列式	压缩态	5 000 000	16.703 : 1
列式	压缩态	8 000 000	16.957 : 1
列式	压缩态	10 000 000	17.053 : 1

4.2 与文献方法的对比测试

与处理大数据的云计算集群进行对比^[21],云计算集群支持 Gzip、LZO、Snappy 和 LZ4 这 4 种压缩方法。设监控信息第 i 序列的列属性原码元字节总长为 $L_{\text{length}}(i)_a$,压缩编码后字节总长为 $L_{\text{length}}(i)_b$,其中 $i \in [1, N]$, N 为列属性总数,压缩率为 c_r 。对相同的配电网实测信息进行压缩处理,压缩率记为:

$$c_r = \frac{1}{N} \sum_{i=1}^N \left(\frac{L_{\text{length}}(i)_a - L_{\text{length}}(i)_b}{L_{\text{length}}(i)_a} \right) \times 100\% \quad (7)$$

式(7)表征的是被压缩消除的冗余监控信息相对于监控信源数据量的占比,压缩率越大表示被消除的冗余信息量越多,压缩效果越好。

在同样的数据集条件下,按照本文列区间编码正规化压缩存储方法和已有文献中按列存储的 HBase 云计算集群压缩方法相比,得到的列式压缩监控信息的压缩率如图 6 所示。

表 2 正规化监控信息区间编码压缩示例

Table 2 Example of normalization of monitoring information range compression encoding

输入信源	读信源	编码映射	区间范围 1	正规化调整	区间范围 2
初始化	—	[000,99]	1 000	0,(0,0),[000,999]	1 000
YC	YC	[000,249]	250	0,(0,0),[000,249]	250
YC03Ia	03Ia	[100,149]	50	1,(0,0),[100,499]	500
YC03IaYC	YC	[000,165]	166	1,(0,0),[000,165]	166
YC03IaYC03Ia	03Ia	[069,114]	46	1,(0,1),[-310,149]	460
YC03IaYC03IaYX	YX	[032,088]	57	110,(0,0),[320,889]	570
YC03IaYC03IaYX00	00	[824,886]	63	1108,(0,0),[240,869]	630
YC03IaYC03IaYX00YC	YC	[240,428]	189	1108,(0,0),[240,28]	189
YC03IaYC03IaYX00YC03Ia	03Ia	[308,358]	51	11083,(0,0),[80,589]	510
YC03IaYC03IaYX00YC03IaYX	YX	[416,499]	84	110834,(0,0),[160,999]	840
YC03IaYC03IaYX00YC03IaYX00	00	[864,991]	128	110834,(0,0),[864,991]	128

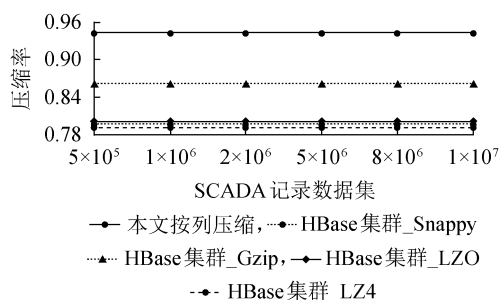


图6 列式压缩与云计算集群压缩的压缩率比较

Fig.6 Compression ratio comparison between column compression and cloud computing cluster compression

对比测试结果表明:本文按列对监控信息进行区间编码的正规化压缩处理,其压缩率可稳定地达到94%左右,高于在HBase云计算中压缩率最高的集群Gzip获得的压缩率86.3%,验证了在处理配电网监测信息时,由区间编码压缩融合正规化处理,并在列式存储中按列属性分类、以列DP为单位的压缩,可在配电自动化的海量工程应用数据中获得相当高的压缩率。

5 结论

a. 利用正规化处理方法在区间编码压缩监控信息时,通过动态调整映射区间,可实现在无穷区间中对海量调度监控信息编码压缩的工程应用需求。

b. 根据某动车段配电网的SCADA调度监控实例信息进行具体工程算例测试,结果表明:对监控信息的压缩处理达到约94%的高压缩率,可以获得比已有的云计算集群压缩更好的效果。

参考文献:

- [1] 张少敏,张帆,王保义. 基于REST和IEC61970的智能电网数据集成方法[J]. 电力自动化设备,2012,32(8):124-129.
ZHANG Shaomin,ZHANG Fan,WANG Baoyi. Data integration based on REST and IEC61970 for smart grid[J]. Electric Power Automation Equipment,2012,32(8):124-129.
- [2] 赵腾,张焰,张东霞. 智能配电网大数据应用技术与前景分析[J]. 电网技术,2014,38(12):3305-3312.
ZHAO Teng,ZHANG Yan,ZHANG Dongxia. Application technology of big data in smart distribution grid and its prospect analysis[J]. Power System Technology,2014,38(12):3305-3312.
- [3] 王广辉,李保卫,胡泽春,等. 未来智能电网控制中心面临的挑战和形态演变[J]. 电网技术,2011,35(8):1-5.
WANG Guanghui,LI Baowei,HU Zechun,et al. Challenges and future evolution of control center under smart grid environment[J]. Power System Technology,2011,35(8):1-5.
- [4] 彭小圣,邓迪元,程时杰,等. 面向智能电网应用的电力大数据关键技术[J]. 中国电机工程学报,2015,35(3):503-511.
PENG Xiaosheng,DENG Diyu,CHENG Shijie,et al. Key technologies of electric power big data and its application prospects in smart grid[J]. Proceedings of the CSEE,2015,35(3):503-511.
- [5] 屈志坚,郭亮,刘明光,等. 智能配电网量测信息变断面柔性压缩新算法[J]. 中国电机工程学报,2013,33(19):191-199.

- QU Zhijian, GUO Liang, LIU Mingguang, et al. New variable section flexible compression algorithm for measurement information in intelligent distribution network[J]. Proceedings of the CSEE, 2013, 33(19):191-199.
- [6] 吴丹丹,王松. 内存数据库及其应用综述[J]. 软件导刊,2016,15(6):168-170.
WU Dandan,WANG Song. Main memory database and its application;a survey[J]. Software Guide,2016,15(6):168-170.
- [7] 胡爽,周欢,钱卫宁. 内存数据库事务提交的关键技术与挑战[J]. 华东师范大学学报(自然科学版),2016(5):18-26.
HU Shuang,ZHOU Huan,QIAN Weining. Key techniques and challenges of transaction commit in main-memory database systems[J]. Journal of East China Normal University (Natural Science), 2016(5):18-26.
- [8] 曹筠宜. 内存数据库的研究及其应用[D]. 北京:北京邮电大学,2011.
CAO Yixuan. Study and application of main memory database[D]. Beijing:Beijing University of Posts and Telecommunications,2011.
- [9] LINDSTROM J, RAATIKKA V, RUUTH J, et al. IBM solid-DB: in-memory database optimized for extreme speed and availability[J]. IEEE Data Eng Bull, 2013, 36(2):14-20.
- [10] LI N, RAO J, SHEKITA E, et al. Leveraging a scalable row store to build a distributed text index[C]//Proceedings of the first international workshop on Cloud Data Management. [S.l.]: ACM, 2009: 29-36.
- [11] 涂振发. 云计算环境下海量空间数据高效存储关键技术研究[D]. 武汉:武汉大学,2012.
TU Zhenfa. Research on the technology of efficient mass spatial data storage in the cloud computing environment[D]. Wuhan: Wuhan University, 2012.
- [12] 葛磊蛟,王守相,瞿海妮. 智能配用电大数据存储架构设计[J]. 电力自动化设备,2016,36(6):194-202.
GE Leijiao,WANG Shouxiang,QU Haini. Design of storage framework for big data of SPDU[J]. Electric Power Automation Equipment,2016,36(6):194-202.
- [13] 邵振国,吴瑾樱,苏文博. 面向海量历史监测数据的谐波污染用户统计建模方法[J]. 电力自动化设备,2016,36(8):110-114.
SHAO Zhenguo,WU Jinying,SU Wenbo. Statistical modeling based on massive historical monitoring data for harmonic pollution customer[J]. Electric Power Automation Equipment,2016,36(8):110-114.
- [14] YOON Y, JEONG C, LEE J, et al. Hybrid database table stored as both row and column store: U.S. patent application 13/334,711[P]. 2011-12-22.
- [15] IVANOVA M G, KERSTEN M L, NES N J, et al. An architecture for recycling intermediates in a column-store[J]. ACM Transactions on Database Systems, 2010, 35(4):24.
- [16] ABADI D J, BONCZ P A, HARIZOPOULOS S. Column-oriented database systems[J]. Proceedings of the VLDB Endowment, 2009, 2(2):1664-1665.
- [17] 沈来信,王伟. 基于Tree-lib的大数据实时分析研究[J]. 计算机科学,2013,40(6):192-195,237.
SHEN Laixin,WANG Wei. Real time analytics study of big data based on Tree-lib[J]. Computer Science, 2013, 40(6):192-195, 237.
- [18] SLEZAK D, SYNAK P, WROBLEWSKI J, et al. Infobright analytic database engine using rough sets and granular computing[C]//IEEE International Conference on Granular Computing. San Jose,

CA, USA; IEEE, 2010: 432-437.

- [19] David Salomon. 数据压缩原理与应用[M]. 吴乐南, 译. 2 版. 北京: 电子工业出版社, 2003: 92-103.
- [20] CONRAD K J, WILSON P R. Grammatical ziv-lempel compression: achieving PPM-class text compression ratios with LZ-class decompression speed[C]//Data Compression Conference. Snowbird, UT, USA; IEEE, 2016: 586-586.
- [21] 屈志坚, 郭亮, 陈秋琳, 等. Hadoop 云构架的智能调度无损集群压缩技术[J]. 电力系统自动化, 2013, 37(18): 93-98.
- QU Zhijian, GUO liang, CHEN Qiulin, et al. Intelligent dispatching lossless cluster compression technology based on Hadoop cloud framework[J]. Automation of Electric Power Systems, 2013, 37(18): 93-98.

作者简介:



屈志坚

屈志坚(1978—), 男, 江西南昌人, 副教授, 博士, 主要研究方向为智能监控理论与信息处理技术(E-mail: 08117324@bjtu.edu.cn);

徐振清(1992—), 男, 山东菏泽人, 硕士, 研究方向为电力自动化智能监控技术;

周锐霖(1994—), 男, 湖南衡阳人, 硕士研究生, 研究方向为配电自动化监控信息技术;

朱 丹(1995—), 男, 贵州安顺人, 硕士研究生, 研究方向为配电网调度自动化技术。

Range encoding normalization compression processing for distribution network monitoring data pack with column storage

QU Zhijian¹, XU Zhenqing², ZHOU Ruilin¹, ZHU Dan¹

(1. School of Electrical Engineering, East China Jiaotong University, Nanchang 330013, China;

2. CRRC Qingdao Sifang Co., Ltd., Qingdao 266000, China)

Abstract: Aiming at the continually increasing large data sets and the difficulty of improving compression ratio caused by the increasing data storage mode by rows in distribution network, a new compression method for large distribution network monitoring data is proposed based on the data pack structure of column storage and the normalization processing of extended range encoding. Taking column data pack as underlying storage and knowledge grid as organization structure, the range encoding and normalization are integrated into the parallel compression process of the underlying monitoring data pack. Taking the distribution network monitoring data of Beijing train set depot section as an example, the column data pack range encoding normalization compression test is carried out. Simulative results show that the normalization of extension range encoding technology can increase the compression ratio of monitoring information to 94%, the compression effect is better than the existing cloud computing cluster, and the engineering application requirements of encoding the massive monitoring information in the infinite interval can be realized by dynamically adjusting the mapping range.

Key words: distribution network; monitoring data; column compression; range encoding; normalization; data processing; data compression; information handling

(上接第 55 页 continued from page 55)

Regional opportunistic maintenance model for distribution network considering improvement of total supply capability

MU Bin, LIU Yan

(School of Electrical and Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: Aiming at the contradiction between the performance requirements of equipments and the reliability requirements of the system during the condition-based maintenance, a regional opportunistic maintenance model for the distribution network is proposed based on the traditional condition-based maintenance, which considers the improvement of TSC(Total Supply Capability). A regional opportunistic maintenance model is built to determine the maintenance area, which comprehensively considers the key factors, i.e. individual performance of equipment, reliability requirement of system, correlation between equipments, system network structure, etc. On this basis, a load transfer model with the improvement of TSC as its objective function is built, which optimizes the network structure by the algorithm combining NDE(Node-Depth Encoding) and GA(Genetic Algorithm) to obtain the satisfactory reconstruction network. IEEE 33-bus system is taken as an example to verify the effectiveness of the proposed algorithm in improving TSC, and the effectiveness, practicality and scientificity of the proposed model is verified by RBTS Bus2 reliability test system.

Key words: regional opportunistic maintenance; TSC; NDE-GA; condition-based maintenance; distribution network; models