\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

xml tags(nodes) with parameters.  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[training@localhost ~]$ cat xml6  
<tr><cid>101</cid><pr pid="p1">1000</pr><pr pid="p3">3000</pr></tr>  
<tr><cid>102</cid><pr pid="p1">1200</pr><pr pid="p2">2000</pr></tr>  
<tr><cid>103</cid><pr pid="p2">5000</pr><pr pid="p3">3000</pr></tr>  
[training@localhost ~]$

[training@localhost ~]$ cat customers  
<rec><cid>101</cid><name>Mani</name><city>Hyd</city></rec>  
<rec><cid>102</cid><name>Vani</name><city>Pune</city></rec>  
<rec><cid>103</cid><name>Veni</name><city>Hyd</city></rec>  
[training@localhost ~]$

[training@localhost ~]$ cat products  
<rec><pid>p1</pid><brand>Samsung</brand><pname>Mobile</pname><mrp>5000</mrp></rec>  
<rec><pid>p2</pid><brand>Lg</brand><pname>Mobile</pname><mrp>7000</mrp></rec>  
<rec><pid>p3</pid><brand>Lg</brand><pname>Laptop</pname><mrp>6000</mrp></rec>  
[training@localhost ~]$

hive> create database xml10;  
OK  
Time taken: 0.034 seconds  
hive> use xml10;  
OK  
Time taken: 0.027 seconds  
hive> create table transraw(line string);  
OK  
Time taken: 0.057 seconds  
hive> create table custraw(line string);  
OK  
Time taken: 0.034 seconds  
hive> create table prodraw(line string);  
OK  
Time taken: 0.038 seconds  
hive>

hive> load data local inpath 'customers' into table custraw;  
Copying data from file:/home/training/customers  
Copying file: file:/home/training/customers  
Loading data to table xml10.custraw  
OK  
Time taken: 0.155 seconds  
hive> select \* from custraw;  
OK  
<rec><cid>101</cid><name>Mani</name><city>Hyd</city></rec>  
<rec><cid>102</cid><name>Vani</name><city>Pune</city></rec>  
<rec><cid>103</cid><name>Veni</name><city>Hyd</city></rec>  
Time taken: 0.065 seconds  
hive>

hive> create table customers(cid string,   
    >   name string, city string);  
OK  
Time taken: 0.043 seconds  
hive> insert overwrite table customers   
    >   select xpath\_string(line,'rec/cid'),  
    >    xpath\_string(line,'rec/name'),       
    >    xpath\_string(line,'rec/city')   
    > from custraw;

hive> select \* from customers;  
OK  
101     Mani    Hyd  
102     Vani    Pune  
103     Veni    Hyd  
Time taken: 0.059 seconds  
hive>

hive> load data local inpath 'products'  
    >  into table prodraw;  
Copying data from file:/home/training/products  
Copying file: file:/home/training/products  
Loading data to table xml10.prodraw  
OK  
Time taken: 0.109 seconds  
hive> select \* from prodraw;  
OK  
<rec><pid>p1</pid><brand>Samsung</brand><pname>Mobile</pname><mrp>5000</mrp></rec>  
<rec><pid>p2</pid><brand>Lg</brand><pname>Mobile</pname><mrp>7000</mrp></rec>  
<rec><pid>p3</pid><brand>Lg</brand><pname>Laptop</pname><mrp>6000</mrp></rec>  
Time taken: 0.057 seconds  
hive>   
hive> create table products(pid string,   
    >  pname string, brand string,   
    >   mrp int);  
OK  
Time taken: 0.036 seconds  
hive> insert overwrite table products  
    >   select xpath\_string(line,'rec/pid'),  
    >    xpath\_string(line,'rec/pname'),      
    >    xpath\_string(line,'rec/brand'),  
    >    xpath\_int(line,'rec/mrp')            
    > from prodraw;

hive> select \* from products;  
OK  
p1      Mobile  Samsung 5000  
p2      Mobile  Lg      7000  
p3      Laptop  Lg      6000  
Time taken: 0.06 seconds  
hive>

hive> select \* from transraw;  
OK  
Time taken: 0.052 seconds  
hive> load data local inpath 'xml6'  
    >  into table transraw;  
Copying data from file:/home/training/xml6  
Copying file: file:/home/training/xml6  
Loading data to table xml10.transraw  
OK  
Time taken: 0.094 seconds  
hive> select \* from transraw;  
OK  
<tr><cid>101</cid><pr pid="p1">1000</pr><pr pid="p3">3000</pr></tr>  
<tr><cid>102</cid><pr pid="p1">1200</pr><pr pid="p2">2000</pr></tr>  
<tr><cid>103</cid><pr pid="p2">5000</pr><pr pid="p3">3000</pr></tr>  
Time taken: 0.056 seconds  
hive>

hive> create table transraw2(cid string,   
    >   pid array<string>,  
    >  pr array<string>);  
OK  
Time taken: 0.034 seconds  
hive> insert overwrite table transraw2  
    >   select xpath\_string(line,'tr/cid'),  
    >  xpath(line,'tr/pr/@pid'),   
    >  xpath(line,'tr/pr/text()') from transraw;

hive> select \* from transraw2;  
OK  
101     ["p1","p3"]     ["1000","3000"]  
102     ["p1","p2"]     ["1200","2000"]  
103     ["p2","p3"]     ["5000","3000"]  
Time taken: 0.058 seconds  
hive> create table cidpid(cid string, pid string);  
OK  
Time taken: 0.041 seconds  
hive> insert overwrite table cidpid  
    >   select cid, mypid from transraw2  
    >    lateral view explode(pid) p as mypid;

hive> select \* from cidpid;  
OK  
101     p1  
101     p3  
102     p1  
102     p2  
103     p2  
103     p3  
Time taken: 0.054 seconds  
hive> create table cidpr(cid string, pr int);  
OK  
Time taken: 0.072 seconds  
hive> insert overwrite table cidpr  
    >   select cid, mypy from transraw2  
    >     lateral view explode(pr) p as  mypy;

hive> select \* from cidpr;  
OK  
101     1000  
101     3000  
102     1200  
102     2000  
103     5000  
103     3000  
Time taken: 0.069 seconds  
hive> alter table cidpid add columns(n int);  
OK  
Time taken: 0.056 seconds  
hive>

add jar Desktop/hivejars.jar;  
create temporary function auto   
   as 'hive.analytics.SeqNumber';

hive> insert overwrite table cidpid  
    >   select cid, pid, auto() from cidpid;

hive> select \* from cidpid;  
OK  
101     p1      1  
101     p3      2  
102     p1      3  
102     p2      4  
103     p2      5  
103     p3      6  
Time taken: 0.066 seconds  
hive>

hive> alter table cidpr add columns(n int);  
OK  
Time taken: 0.057 seconds  
hive> insert overwrite table cidpr  
    >   select cid, pr, auto() from cidpr;

hive> select \* from cidpr;  
OK  
101     1000    1  
101     3000    2  
102     1200    3  
102     2000    4  
103     5000    5  
103     3000    6  
Time taken: 0.061 seconds  
hive>

hive> create table trans(cid string, pid string,  
    >  pr int);  
OK  
Time taken: 0.031 seconds  
hive> insert overwrite table trans  
    >   select l.cid, pid, pr   
    >   from   
    >  cidpid l join cidpr r   
    >  on (l.n=r.n);

hive> select \* from trans;  
OK  
101     p1      1000  
101     p3      3000  
102     p1      1200  
102     p2      2000  
103     p2      5000  
103     p3      3000  
Time taken: 0.063 seconds  
hive>

hive> select \* from products;  
OK  
p1      Mobile  Samsung 5000  
p2      Mobile  Lg      7000  
p3      Laptop  Lg      6000  
Time taken: 0.052 seconds  
hive> create table  info1(cid string,   
    >   pid string, pname string,   
    >   brand string, mrp int, pr int);  
OK  
Time taken: 0.075 seconds  
hive> insert overwrite table info1  
    >    select cid, l.pid, pname, brand,  
    >    mrp , pr    
    >   from trans l join products r  
    >  on (l.pid = r.pid);  
hive> select \* from info1;  
OK  
101     p1      Mobile  Samsung 5000    1000  
102     p1      Mobile  Samsung 5000    1200  
102     p2      Mobile  Lg      7000    2000  
103     p2      Mobile  Lg      7000    5000  
101     p3      Laptop  Lg      6000    3000  
103     p3      Laptop  Lg      6000    3000  
Time taken: 0.057 seconds  
hive> select \* from customers;  
OK  
101     Mani    Hyd  
102     Vani    Pune  
103     Veni    Hyd  
Time taken: 0.059 seconds  
hive>   
hive> create table info(cid string,   
    >  name string, pid string,   
    >  pname string, brand string,   
    >  mrp int, pr int, city string)  
    > row format delimited   
    >   fields terminated by ',';  
OK  
Time taken: 0.082 seconds  
hive>

hive> insert overwrite table info          
    >   select l.cid, name, pid, pname, brand,  
    >  mrp , pr , city                          
    > from info1 l join customers r  
    >  on (l.cid = r.cid);

hive> select \* from info;  
OK  
101     Mani    p1      Mobile  Samsung 5000    1000      Hyd  
101     Mani    p3      Laptop  Lg      6000    3000      Hyd  
102     Vani    p1      Mobile  Samsung 5000    1200      Pune  
102     Vani    p2      Mobile  Lg      7000    2000      Pune  
103     Veni    p2      Mobile  Lg      7000    5000      Hyd  
103     Veni    p3      Laptop  Lg      6000    3000      Hyd  
Time taken: 0.053 seconds  
hive>