The Amazing Utility

of `ROW_NUMBER()`

SQL Saturday
Denver, CO
23 September 2022

Breanna Hansen
breanna@tf3604.com
@tf3604

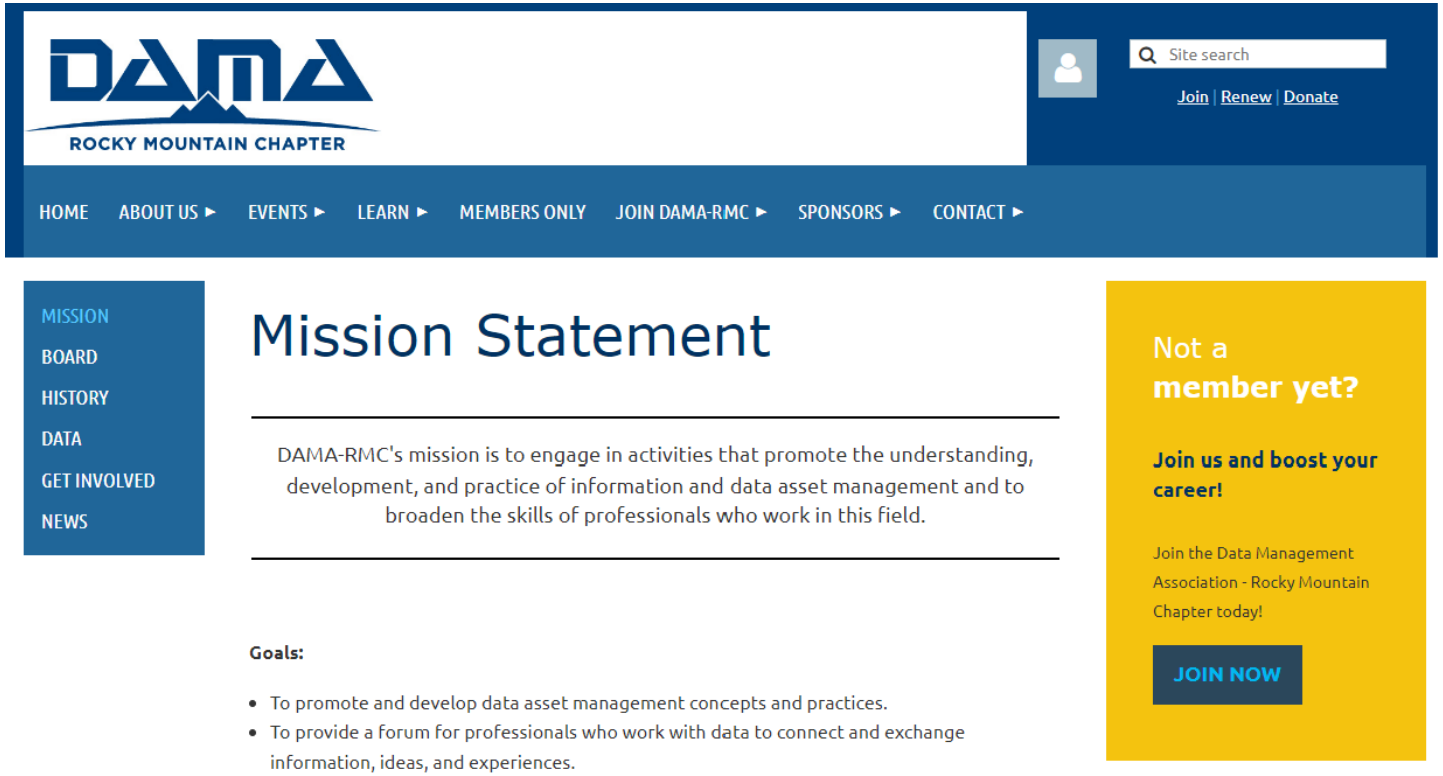
Thank you to our Host and Sponsors!

Global Sponsor

Summit Sponsors

Crest Sponsor

Visit our Friends at DAMA



[DAMA - Rocky Mountain Chapter - Home (wildapricot.org)]

Breanna Hansen


breanna@tf3604.com


@tf3604.com

children.org

- 20+ Years working with SQL Server
  - Development work since 7.0
  - Administration going back to 6.5
  - Fascinated with SQL internals

www.tf3604.com/rownumber

Agenda

- What is `ROW_NUMBER`?
- The Utility of `ROW_NUMBER`
- Performance
- Specific example

Ranking functions

- ROW_NUMBER
- RANK
- DENSE_RANK

ROW_NUMBER vs RANK vs DENSE_RANK

	TeamName	Wins	row_number	rank	dense_rank
1	Cleveland	22	1	1	1
2	Detroit	20	2	2	2
3	Minnesota	20	3	2	2
4	Kansas City	14	4	4	3
5	Chicago	13	5	5	4

That OVER() Clause

- Optional `PARTITION BY` clause
  - The "`GROUP BY`" experience
- Mandatory `ORDER BY` clause
  - Determine ranking sort order

Example Usage

```
select obj.name, obj.type_desc,
    row_number()
        over (order by obj.name) rn
from sys.objects obj
order by obj.name;
```

1	EventNotificationErrorsQueue	SERVICE_QUEUE	1
2	ImportantCustomers	VIEW	2
3	InternalSeparatedHexToParts	SQL_INLINE_TABLE_VALUED_FUNCTION	3
4	MSreplication_options	USER_TABLE	4
5	plan_persist_context_settings	INTERNAL_TABLE	5
6	plan_persist_plan	INTERNAL_TABLE	6
7	plan_persist_query	INTERNAL_TABLE	7
8	plan_persist_query_text	INTERNAL_TABLE	8
9	plan_persist_runtime_stats	INTERNAL_TABLE	9
10	plan_persist_runtime_stats_interval	INTERNAL_TABLE	10
11	QueryNotificationErrorsQueue	SERVICE_QUEUE	11

```
select obj.name, obj.type_desc,
    row_number()
        over (partition by obj.type_desc
        order by obj.name) rn
from sys.objects obj
order by obj.type_desc, obj.name;
```

10	sqlagent_job_history	INTERNAL_TABLE	10
11	sqlagent_jobs	INTERNAL_TABLE	11
12	sqlagent_jobsteps	INTERNAL_TABLE	12
13	sqlagent_jobsteps_logs	INTERNAL_TABLE	13
14	syscommittab	INTERNAL_TABLE	14
15	EventNotificationErrorsQueue	SERVICE_QUEUE	1
16	QueryNotificationErrorsQueue	SERVICE_QUEUE	2
17	ServiceBrokerQueue	SERVICE_QUEUE	3
18	InternalSeparatedHexToParts	SQL_INLINE_TABLE_VALUED_FUNCTION	1
19	sp_MScleanupmergepublisher	SQL_STORED_PROCEDURE	1
20	sp_MSrepl_startup	SQL_STORED_PROCEDURE	2
21	sp_who5	SQL_STORED_PROCEDURE	3

Standard SQL is some of the best SQL

- `ROW_NUMBER` is ANSI standard! (2003)
- Implemented by most RDBMS vendors
- First available in SQL Server in 2005

Side note: Where can row_number appear?

```
select row_number() over (...)
from TableA a join TableB b
on a.ID = b.row_number() over (...)
where a.Ranking = row_number() over (...)
group by row_number() over (...)
having row_number() over (...) = 1
order by row_number() over (...);
```

Bad Good

To use it elsewhere, put it in a subquery ...

```
select ID
from
(
   select a.ID, a.Ranking,
      row_number() over (...) rn
   from TableA a join TableB b
   on a.ID = b.ID
) ResultsetPlusRN
where Ranking = rn
order by rn;
```

… or in a CTE

```
with ResultsetPlusRN as
(
   select a.ID, row_number() over (...) rn
   from TableA a join TableB b
  on a.ID = b.ID
)
select rs.ID
from ResultsetPlusRN rs
where a.Ranking = rs.rn
order by rs.rn;
```

The Utility of `ROW_NUMBER()`

Some potential uses

- Tally table / function
- Top *n* within group
- Duplicate detection / deletion
- Gaps and islands
- String splitting
- Bulk “identity” generation

Tally table / function

- Tally table / number tables have many uses
  - Replace code that uses cursors & loops
- Use `ROW_NUMBER` to create tally table
- Or … use `ROW_NUMBER` to create tally function

Tally table

Many ways to create (see this Stack Overflow item)

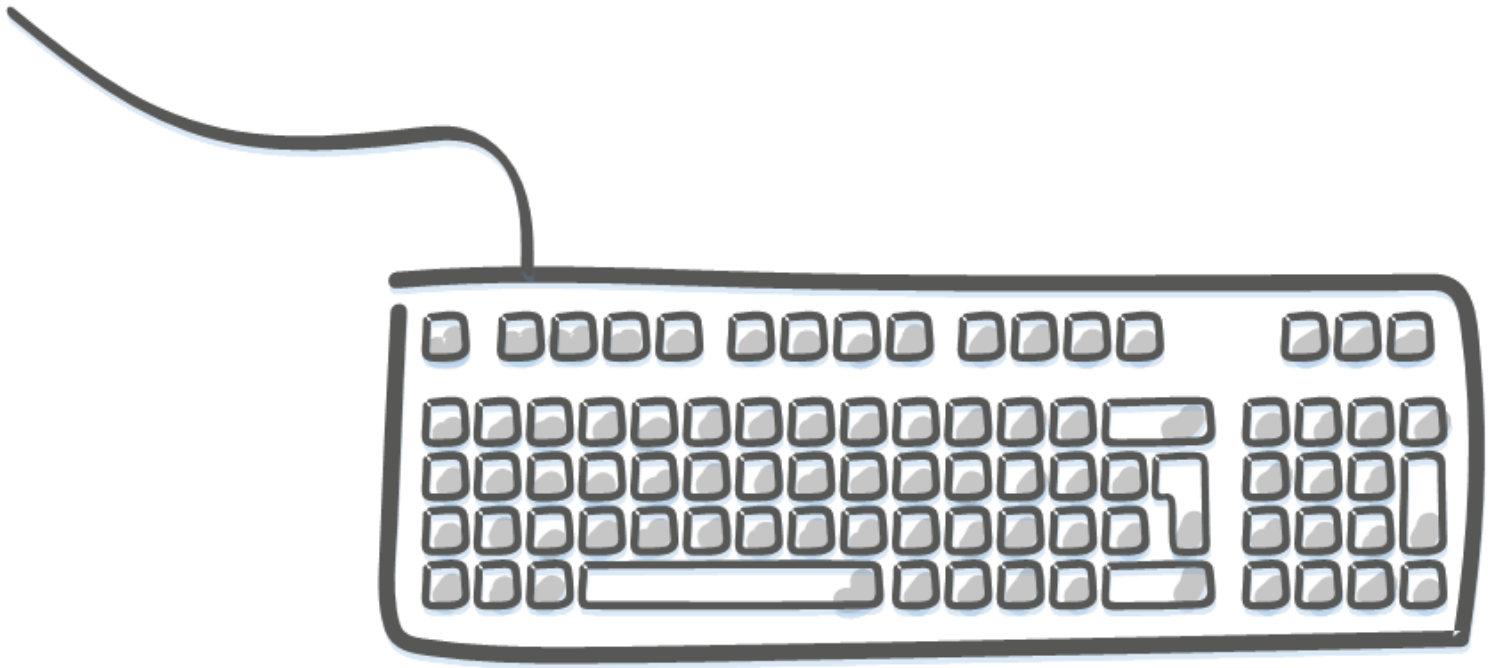
```
with l0 as (select 1 v union all select 1),
l1 as (select a.v from l0 a, l0),
l2 as (select a.v from l1 a, l1),
l3 as (select a.v from l2 a, l2),
l4 as (select a.v from l3 a, l3),
l5 as (select a.v from l4 a, l4),
nums as (select row_number()
   over (order by (select null)) n from l5)
insert Nums (n)
select n.n from nums n where n.n <= 1000000;
```

Tally function (table-valued)

```
create function dbo.fnGetNums(@upperLimit int)
returns table as return
with  l0 as (select 1 v union all select 1),
      l1 as (select a.v from l0 a, l0),
      l2 as (select a.v from l1 a, l1),
      l3 as (select a.v from l2 a, l2),
      l4 as (select a.v from l3 a, l3),
      l5 as (select a.v from l4 a, l4),
      Nums as (select row_number()
         over (order by (select null)) n from l5)
select top (@upperLimit) Nums.n
from Nums;
```

Demo

Tally function



Top *n* within group

- Examples
  - Top 3 highest paid employees per department
  - Cheapest 10 products per category
  - Highest paying *x* customers per salesperson
- Generally, `ROW_NUMBER()` is most performant option

Top *n* within group: APPLY operator

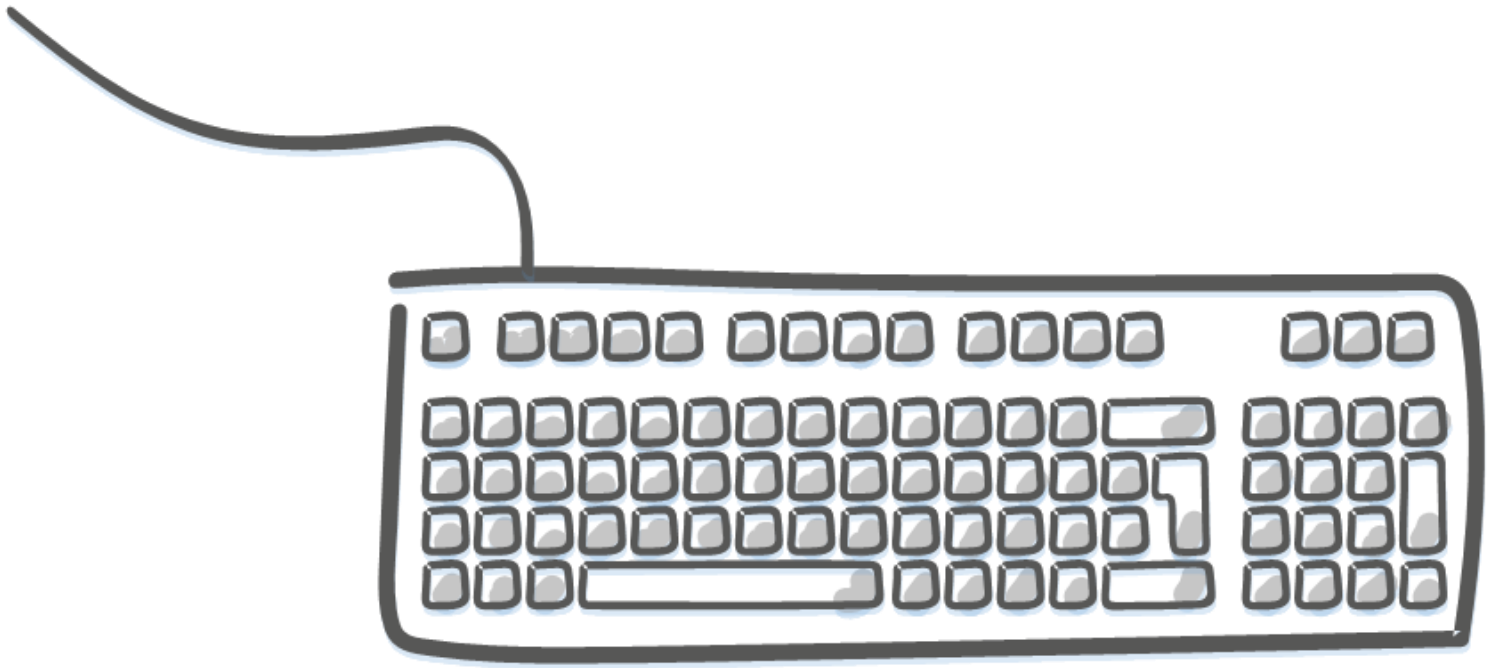
```
select e.EmployeeID, e.FirstName, e.LastName, d.DepartmentCode, e.Salary
from dbo.Department d
cross apply
(
    select top (3) e.EmployeeID, e.FirstName, e.LastName, e.Salary
    from dbo.Employee e
    where e.DepartmentCode = d.DepartmentCode
    order by e.Salary desc
) e
order by d.DepartmentCode, e.Salary desc;
```

Top *n* within group: ROW_NUMBER

```
with EmployeeWithRN as
(
    select e.EmployeeID, e.FirstName, e.LastName, e.DepartmentCode, e.Salary,
        row_number() over
            (partition by e.DepartmentCode
            order by e.Salary desc) rn
    from dbo.Employee e
)
select e.EmployeeID, e.FirstName, e.LastName, e.DepartmentCode, e.Salary
from EmployeeWithRN e
where e.rn <= 3
order by e.DepartmentCode, e.rn;
```

Demo

Top *n* within group

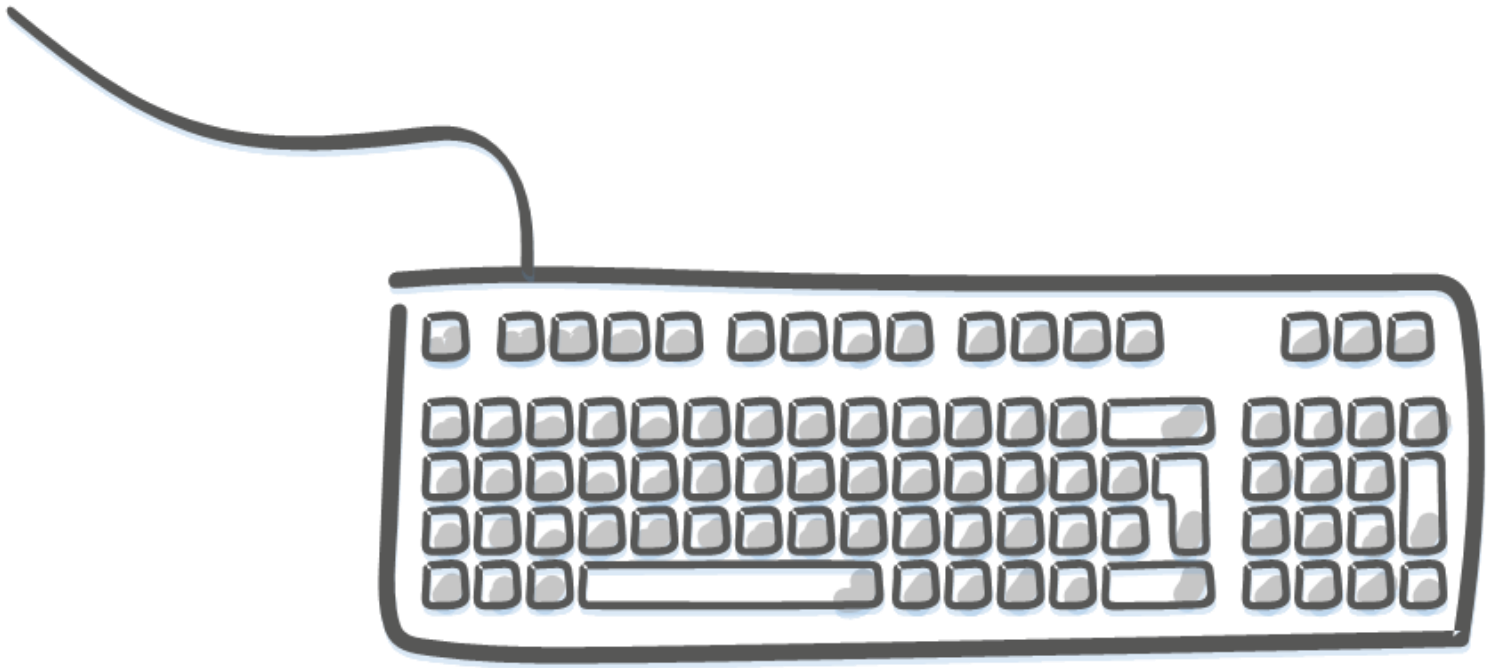


Duplicate detection / deletion

- By far the best way to delete duplicates from a heap
- Even for tables with clustered index can be very good
- Can be used for duplicate detection / elimination from a result set as well

Demo

Duplicate detection

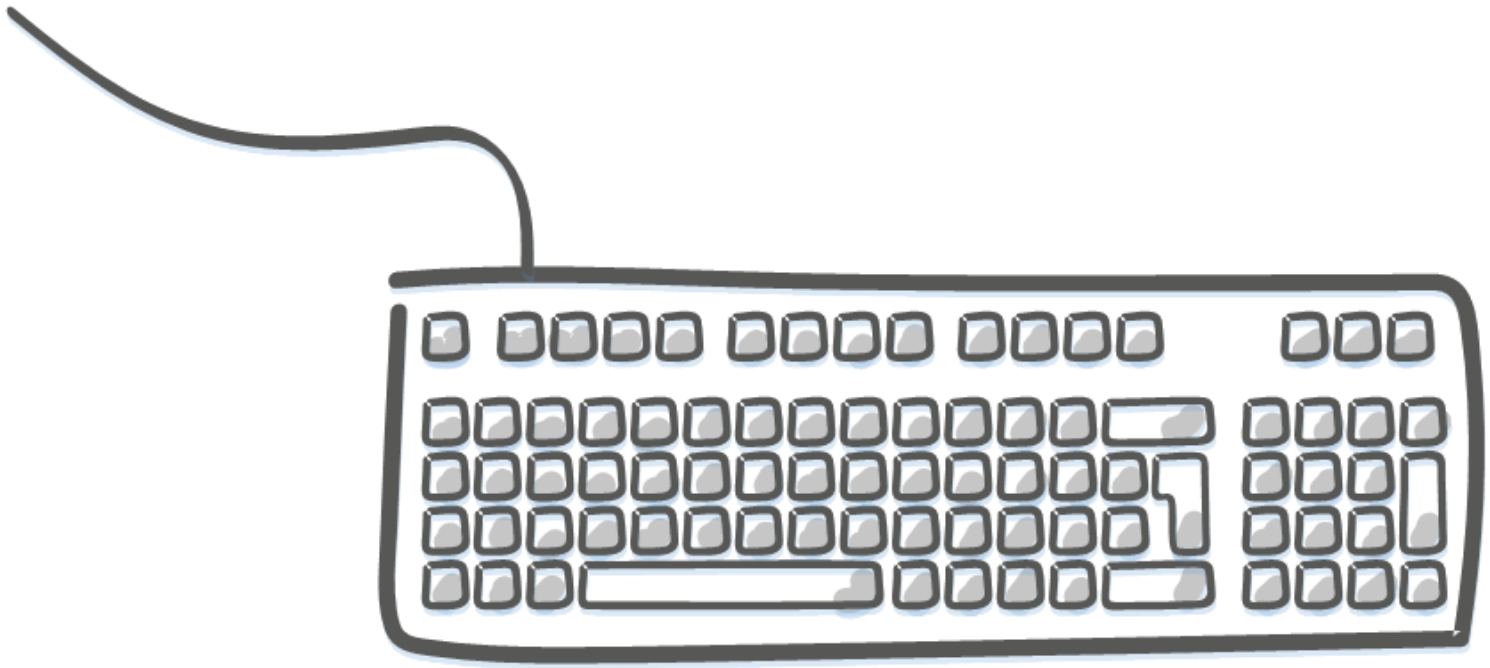


Gaps and islands

- Island: unbroken sequence of values
  - Example island of 4 through 11
  - 1, 4, 5, 6, 7, 8, 9, 10, 11, 17, …
- Gap: unbroken sequence of *missing* values
  - Example gap of 3 through 16
  - 1, 2, 17, 18, …
- Can be any continuous data type (integer, date)

Demo

Gaps and islands

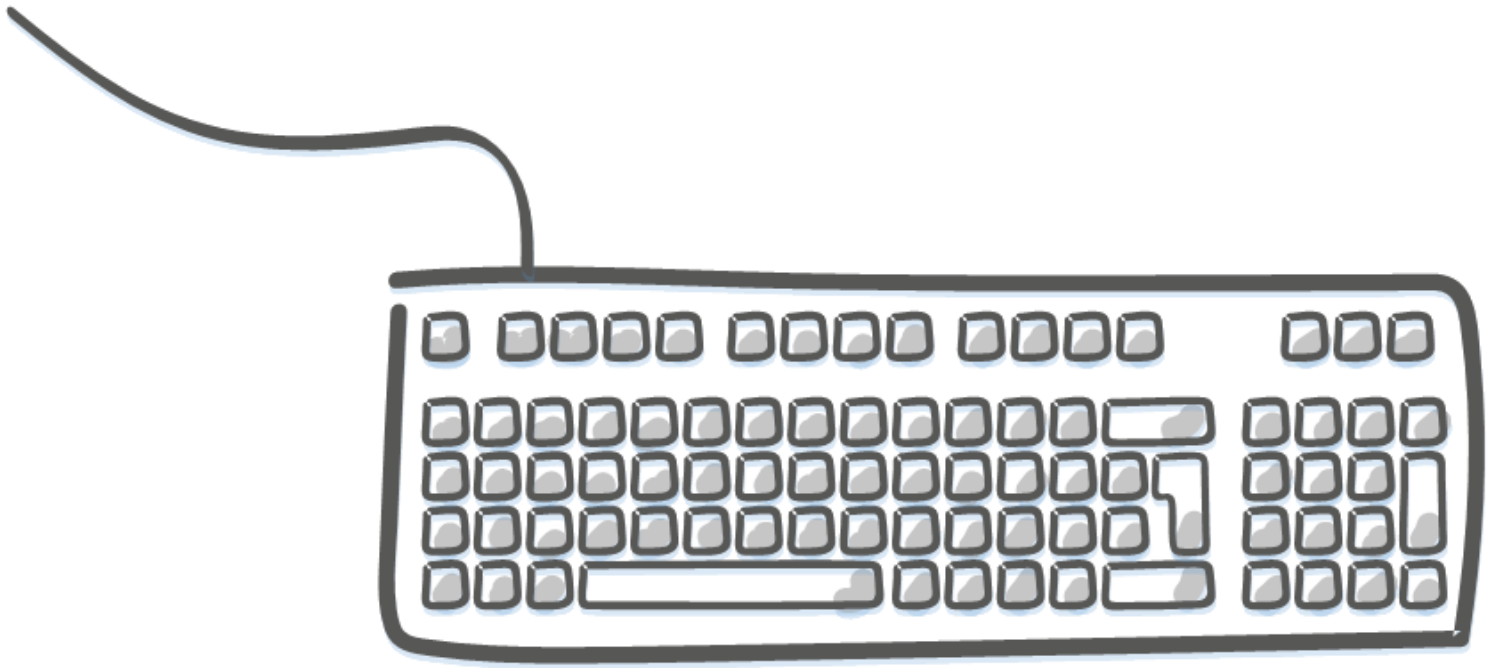


String splitting

- SQL 2016 introduced the `string_split` function
- Previously had to resort less straightforward methods
  - XML approach
  - ROW_NUMBER() approach

Demo

String splitting



Bulk “identity” generation

Scenario: We are merging data from two databases together (for simplicity we are assuming no duplicate records; also assuming a quiet system).

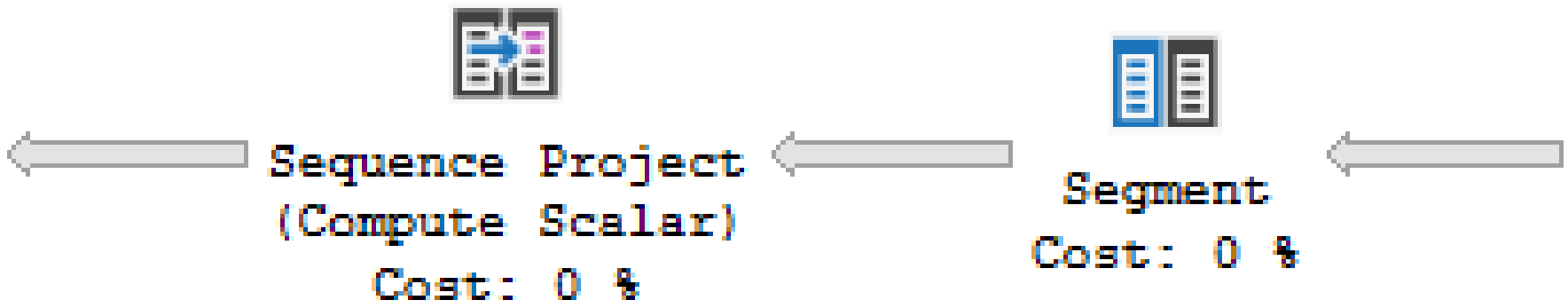
Bulk “identity” generation

```
declare @id int =
    (select max(PhoneNumberId) from dbo.PhoneNumber);
set identity_insert dbo.PhoneNumber on;

insert dbo.PhoneNumber
    (PhoneNumberId, AreaCode, Phone, PhoneExtension)
select   row_number()
    over (order by p.ID) + @id as PhoneNumberId,
         AreaCode, Phone, PhoneExtension
from  Database2.dbo.PhoneNumber p

set identity_insert dbo.PhoneNumber off;
```

Performance

- Segment and Sequence Project operators
- POC indexes

Segment & Sequence Project operators



Segment

- Divides input into partitions
  - Internal “GROUP BY”
- Requires sorted input
- Adds new column indicating if the “group” has changed
- Included even if there is no `PARTITION BY`

Sequence Project

- In relational algebra parlance, “Project” is similar to the SQL `select` clause
- A sequence project operator is similar to a regular projection but is order-preserving
- Generates increasing sequence of values, resetting back to 1 on partition change

POC indexes

- POC = Partition + Order + Covering

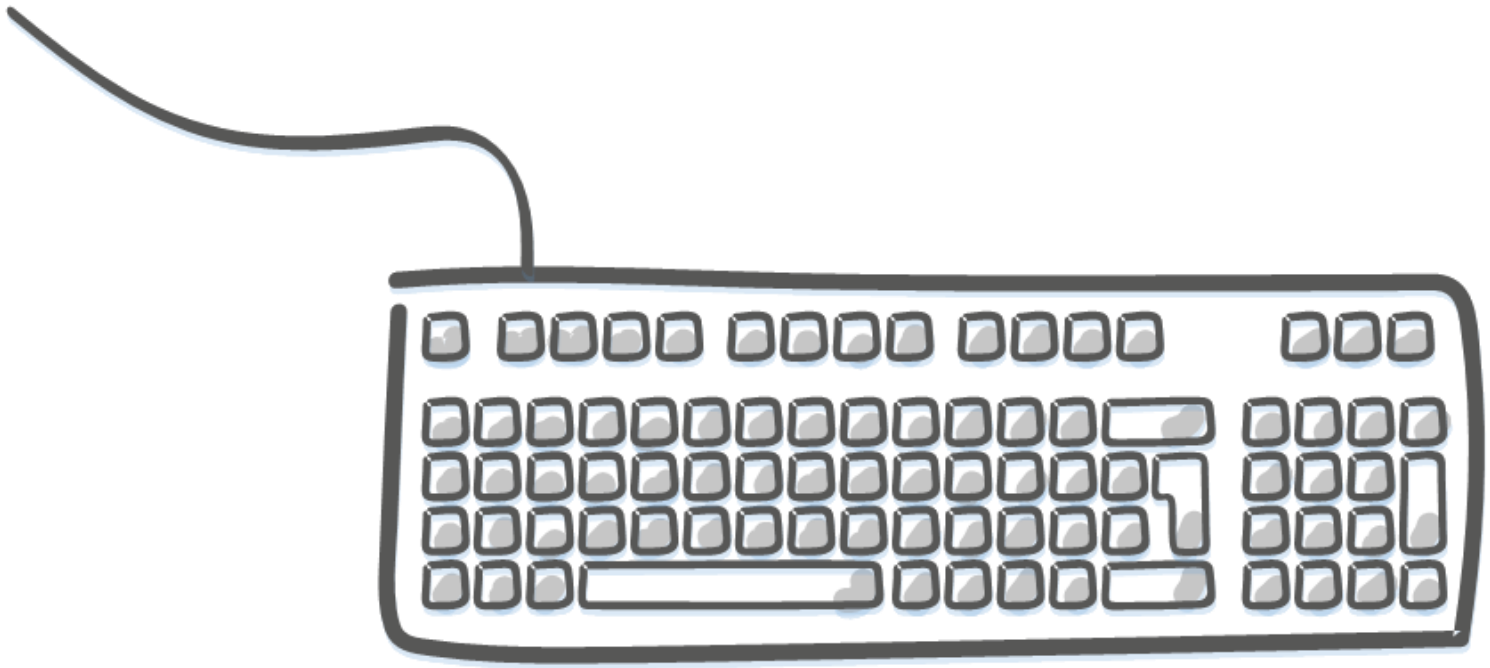
```
select ic1, ic2, row_number()
   over (partition by pc1, pc2
   order by obc1, obc2, obc3)
from tbl;

create index pocIndex on tbl
   (pc1, pc2, obc1, obc2, obc3)
   include (ic1, ic2);
```

Example: LSN Collector

- Scenario: You need to copy production backup files to a test server to be restored
- Job that collects Log Sequence Number information on all backup files
- Function that takes restore point time, then identifies
  - Most recent full backup
  - Most recent diff backup for the full
  - Subsequent log backups

Demo

LSN collector



Other uses?

How about you?

What are some other use cases for `ROW_NUMBER` that you have found?

Thank You

This presentation and supporting materials can be found at www.tf3604.com/rownumber.

- Slide deck
- Scripts

breanna@tf3604.com • @tf3604