


SQL Server Index Internals


Sebastian Meine

SQL Stylist with sqlity.net
sebastian@sqlity.net

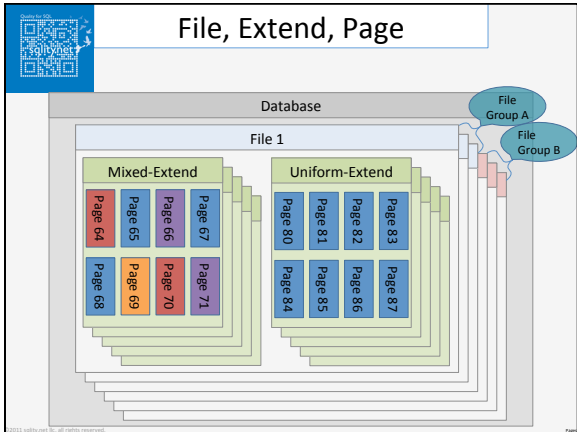


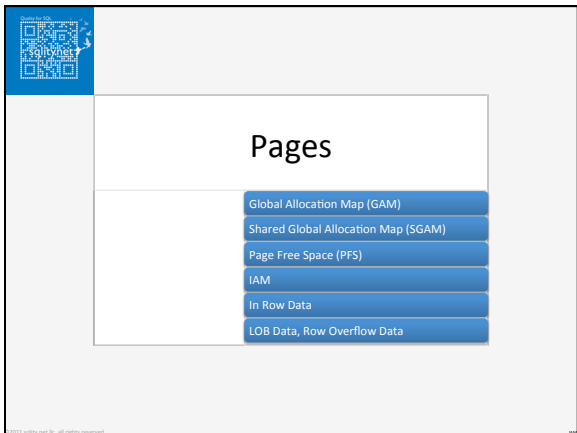
Outline

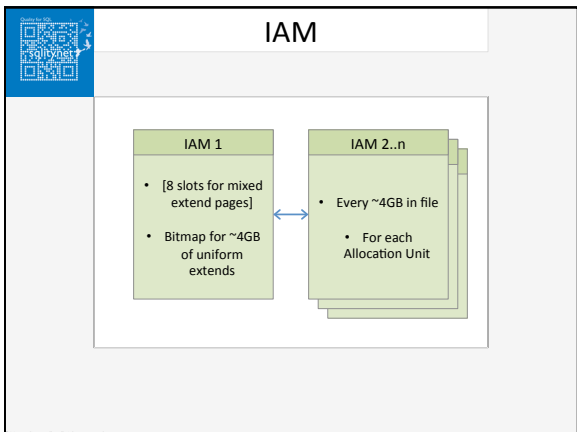
- Data Storage
- Data Access

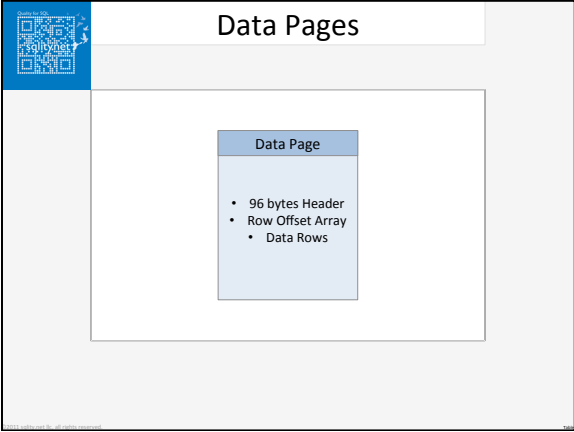


Data Storage





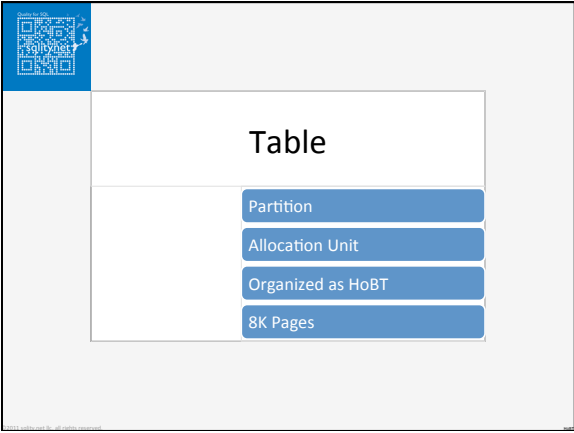




Data Pages

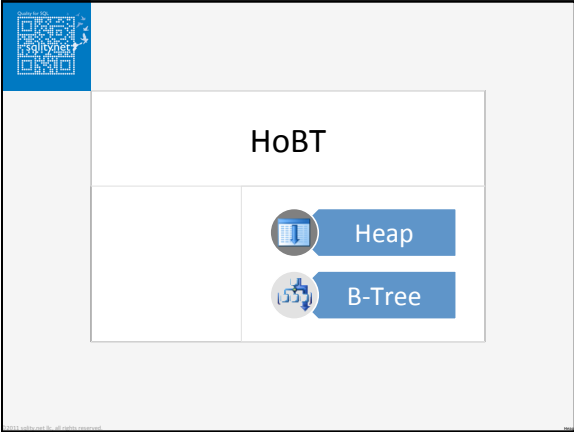
Data Page

- 96 bytes Header
- Row Offset Array
 - Data Rows



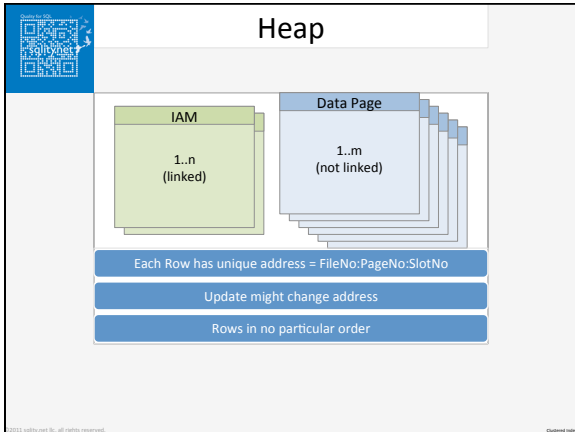
Table

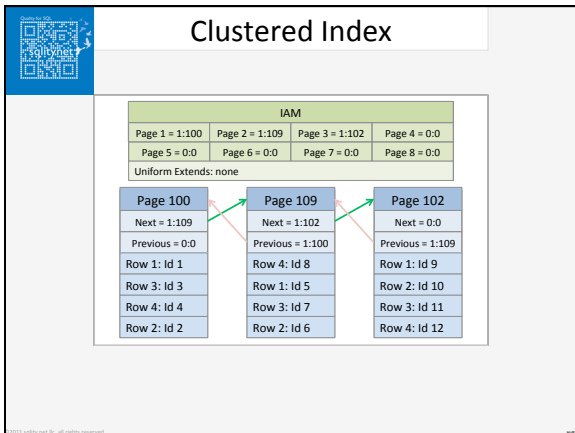
- Partition
- Allocation Unit
- Organized as HoBT
- 8K Pages



HoBT

- Heap
- B-Tree





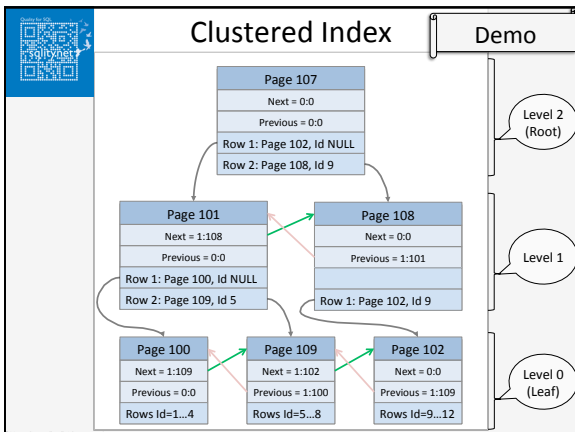


B-Tree

- NOT a Binary Tree
- 1971
- Rudolf Bayer & Ed McCreight
- "B" not explained
- (balanced, bushy, Boeing, Bayer)

B+Tree

- Sorted Data
- Records identified by "Keys"
- Efficient for inserts and deletes
- Only Key values in non-leaf levels



Non-Leaf Pages

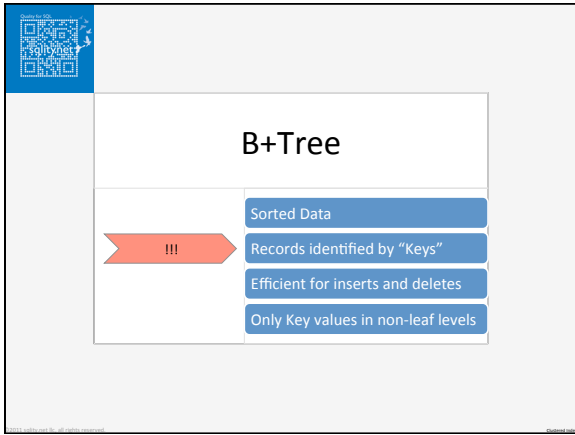
- Store pointer to each Child Page
- Store first Key value for each Child Page

Clustered Index Leaf Pages

- Store all columns
- (might store 24 byte pointer to LOB)

Nonclustered Index Leaf Pages

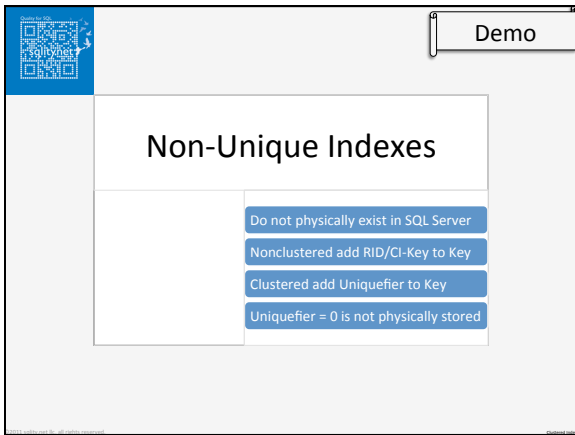
- Store all Index Key columns
- Store RID / Clustered Key
- Store included Columns



B+Tree

- Sorted Data
- Records identified by "Keys"
- Efficient for inserts and deletes
- Only Key values in non-leaf levels

!!!



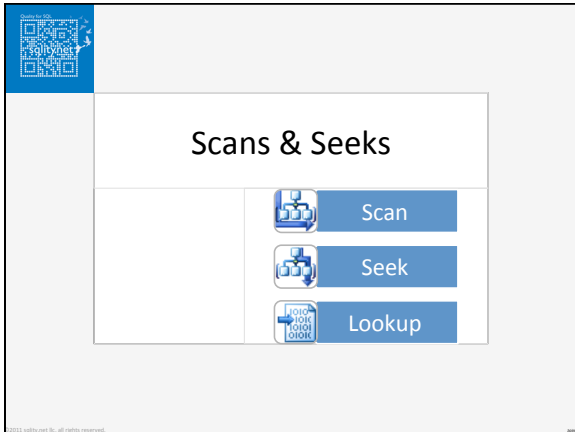
Non-Unique Indexes

- Do not physically exist in SQL Server
- Nonclustered add RID/CI-Key to Key
- Clustered add Uniquefier to Key
- Uniquefier = 0 is not physically stored

Demo



Data Access



Scans & Seeks

- Scan
- Seek
- Lookup

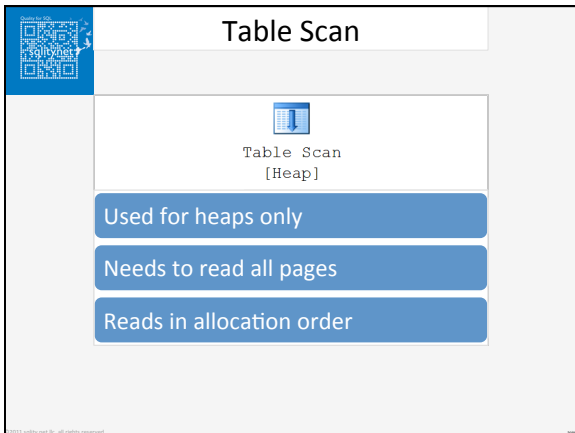
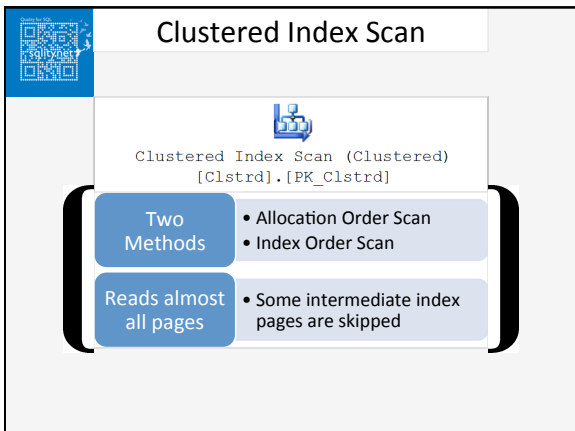


Table Scan
[Heap]


- Used for heaps only
- Needs to read all pages
- Reads in allocation order



Clustered Index Scan (Clustered)
[Clstrd].[PK_Clstrd]

- Two Methods
 - Allocation Order Scan
 - Index Order Scan
- Reads almost all pages
 - Some intermediate index pages are skipped


Clustered Index Seek


Clustered Index Seek (Clustered)
[Clstrd].[PK Clstrd]

Traverses B-Tree to find row

Number of pages read dependent on index depth

Index Scan


Index Scan (NonClustered)
[Clstrd].[UNCI Clstrd]


Two Methods

- Allocation Order Scan
- Index Order Scan

Reads almost all pages

- Some intermediate index pages are skipped


Index Seek


Index Seek (NonClustered)
[Clstrd].[UNCI Clstrd]

Traverses B-Tree to find row

Number of pages read dependent on index depth


RID Lookup



RID Lookup (Heap)
[Heap]

- Used for heaps only
- Retrieve single row using RID
- Might encounter forwarding record


Key Lookup



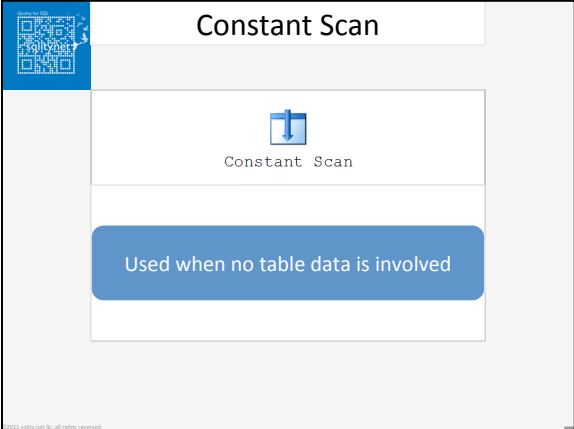
Key Lookup (Clustered)
[Clstrd] . [PKC_Clstrd]

- Used for Clustered Indexes
- Retrieves row using Clustered Index Seek
- No forwarding records in Clustered Indexes

Scan vs. Seek



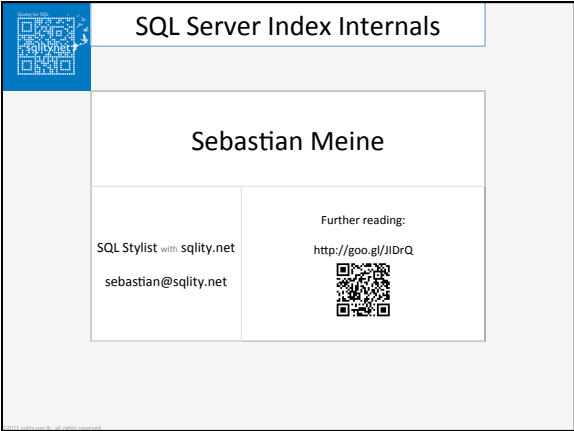
- Is a Seek better? – It depends
- Sometimes a seek is a scan
- Sometimes a scan is a seek



Constant Scan

Constant Scan

Used when no table data is involved



SQL Server Index Internals

Sebastian Meine

SQL Stylist with sqlity.net
sebastian@sqlity.net

Further reading:
<http://goo.gl/JIDrQ>
