Anatomy of an MDF File

Mark S. Rasmussen
improve.dk

C:\>whoami

- Technical Lead @ iPaper
- Developer
- DBA
- Sysadmin
- Project lead
- Comp. Sci. @ AU

Know your data

Know your workload

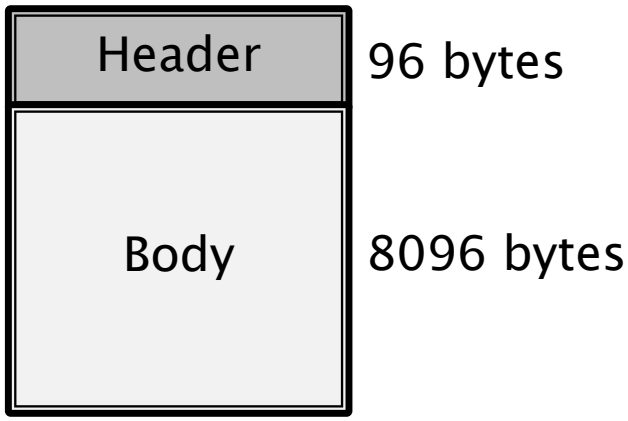
Know how SQL Server works

I'll get back to that...

Files

- MDF
- LDF
- NDF

Pages

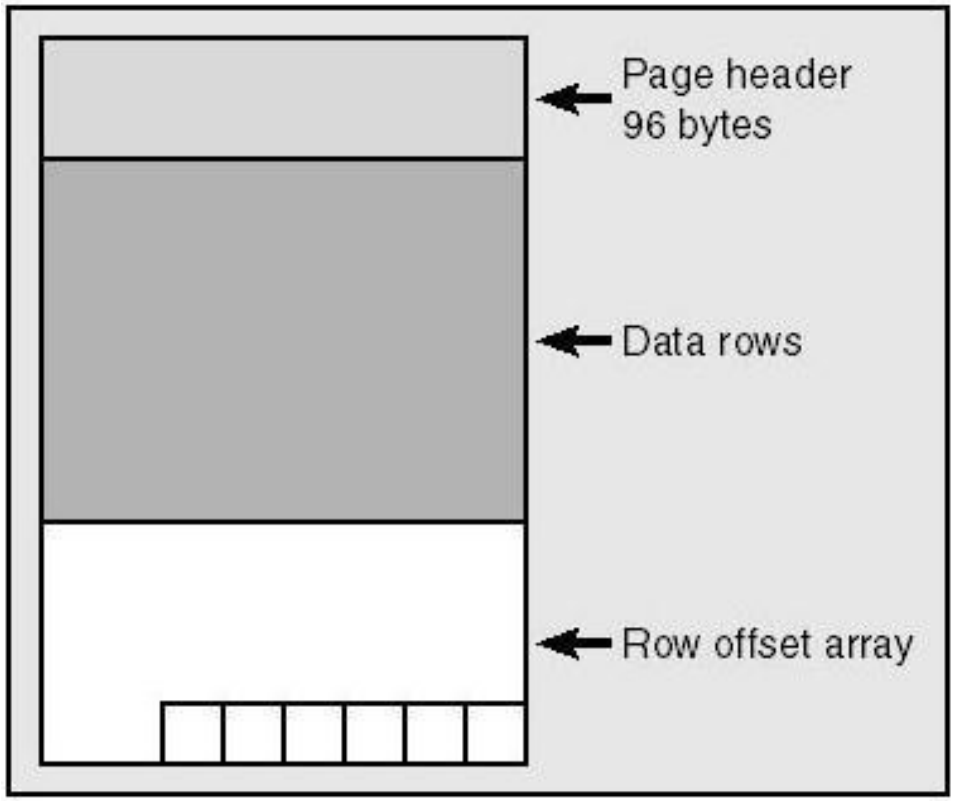


- Smallest datastructure
- 8192 bytes
- 96 byte header, id, linkedlist, lsn, etc
- 8096 byte body
- Contains all sorts of (meta)data

Page types

Page type	Description
1	Data
2	Index
3	BLOB data
4	Variable length data
8	GAM
9	SGAM
10	IAM
11	PFS

Data Pages



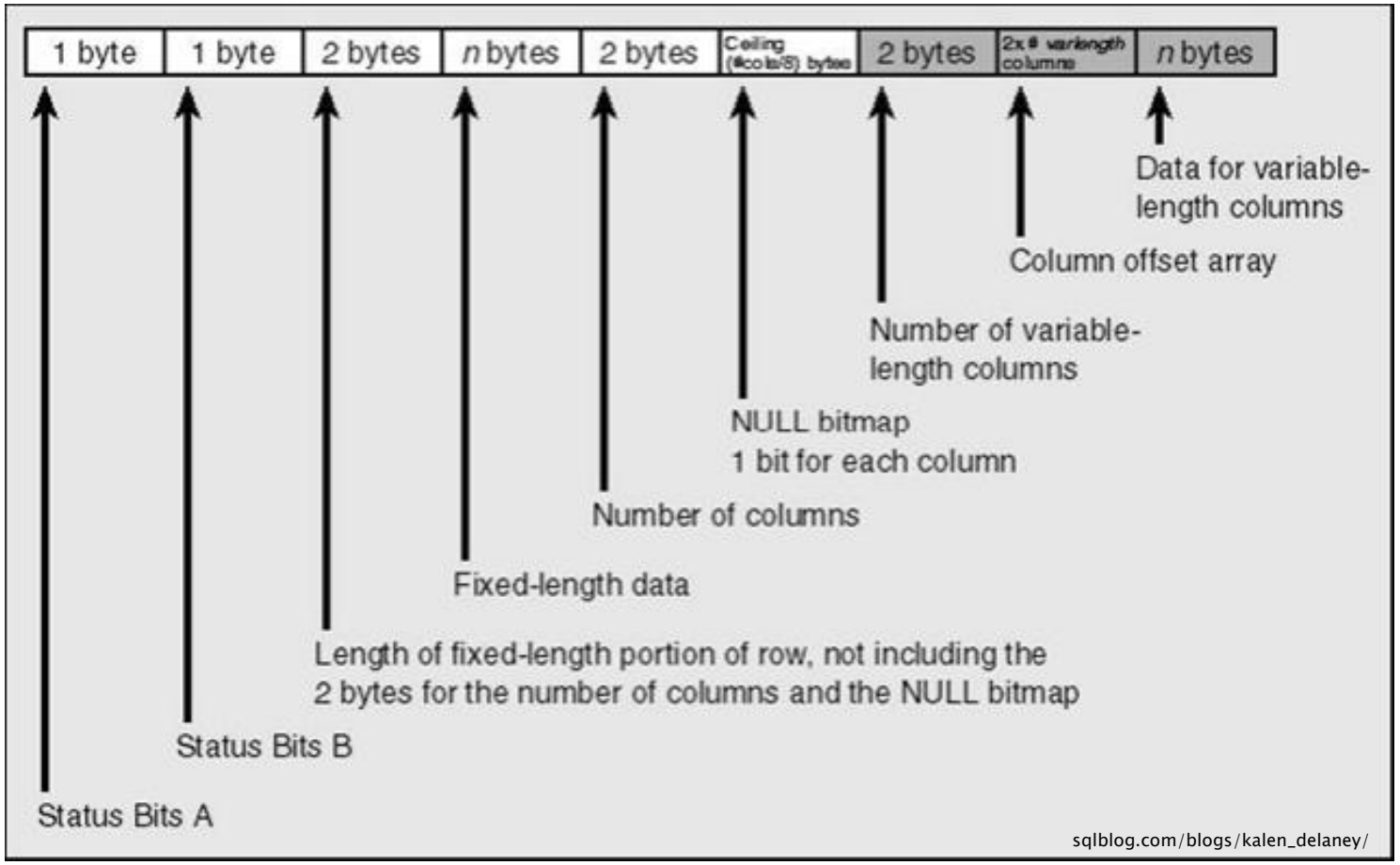
Demo

Page header

- Same for all page types
- Some fields described
  - http://www.sqlskills.com/BLOGS/PAUL/post/Inside-the-Storage-Engine-Anatomy-of-a-page.aspx
  - MSSQL 2008 Internals – Kalen Delaney
- That's all good, but what about the rest?
- How is it actually stored?

Demo

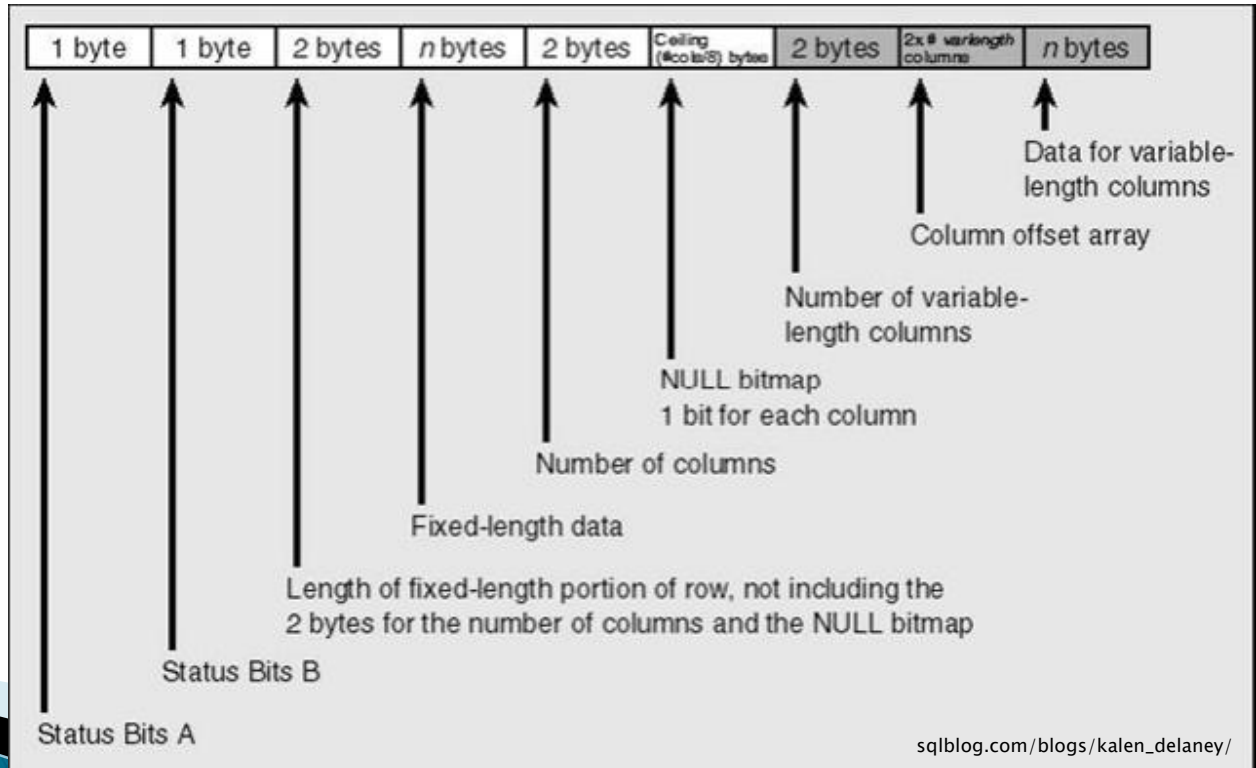
Data page records



An example row

```
00: 30001500 01000000 4d300000 006a1c11
16: 01c19e00 00060000 02002600 34004a6f
32: 686e2044 6f654400 65006e00 6d006100
48: 72006b00
```

```
create table Persons
(
    ID int identity(1,1),
    Name varchar(50),
    Sex char(1) NULL,
    Age int,
    Country nvarchar(50),
    Created datetime default(getDate())
)
```



Null bitmap

- One bit per nullable column
- Using integer math:
  - (numColumns + 7) / 8
- Fixed length data always present, even if null
- Varlength col offset array present, even if null
- Not always present

Demo

Off-row storage

- text, varchar(max), varbinary, (b)lobs.
- High-order bit in varlength offset array

- Example
  - Length = 39 793 / 1001101101110001
  - Length = 7025 / 0001101101110001

Off-row storage

- Instead of data, pointer pointing to relevant text record(s)

- Demo

Bit storage

- Number of bytes = (numBits + 7) / 8
- Up to eight columns stored in same byte
- More than eight bit columns = more bytes

```
create table BitTest
(
    Bit1 bit,
    Bit2 bit,
    Bit3 bit,
    Bit4 bit,
    Bit5 bit,
    Bit6 bit,
    PostalCode char(4),
    Bit7 bit,
    Bit8 bit,
    Bit9 bit
)
```

FEc19e00 0001

When are rebuilds required?

- Fixed length columns
- Column order

- No rebuilds:
  - Generally adding columns ”to the end”
  - Null changes

Page allocation

- Index insertion is simple
- Heaps require support
  - IAM
  - GAM
  - SGAM
  - PFS

Page allocation

- Extents
  - Mixed
  - Uniform
- First 8 table pages = mixed
- More than 8 table pages = uniform

Index Allocation Map

- Tracks object extent allocation across ~4GB
- Indexes, heaps, row-overflow, alloc units
- 44 byte IAM header (in body)
- 7988 bytes for extent bitmap
- 63904 extents tracked
- Values
  - 1 = Extent fully owned by object
  - 0 = Extent not fully owned by object

Demo

Global Allocation Map

- Tracks extent allocation across ~4GB
- 7988 bytes for extent bitmap
- 63904 extents tracked
- Third page of data file, reoccurs every 511232 pages
- Values
  - 1 = Unallocated extent
  - 0 = Allocated extent

Demo

Shared Global Allocation Map

- Tracks extent allocation across ~4GB
- 7988 bytes for extent bitmap
- 63904 extents tracked
- Fourth page of data file, reoccurs every 511232 pages
- Values
  - 1 = Mixed extent with at least one available page
  - 0 = Either uniform or no pages available

Demo

Extent Allocation Status

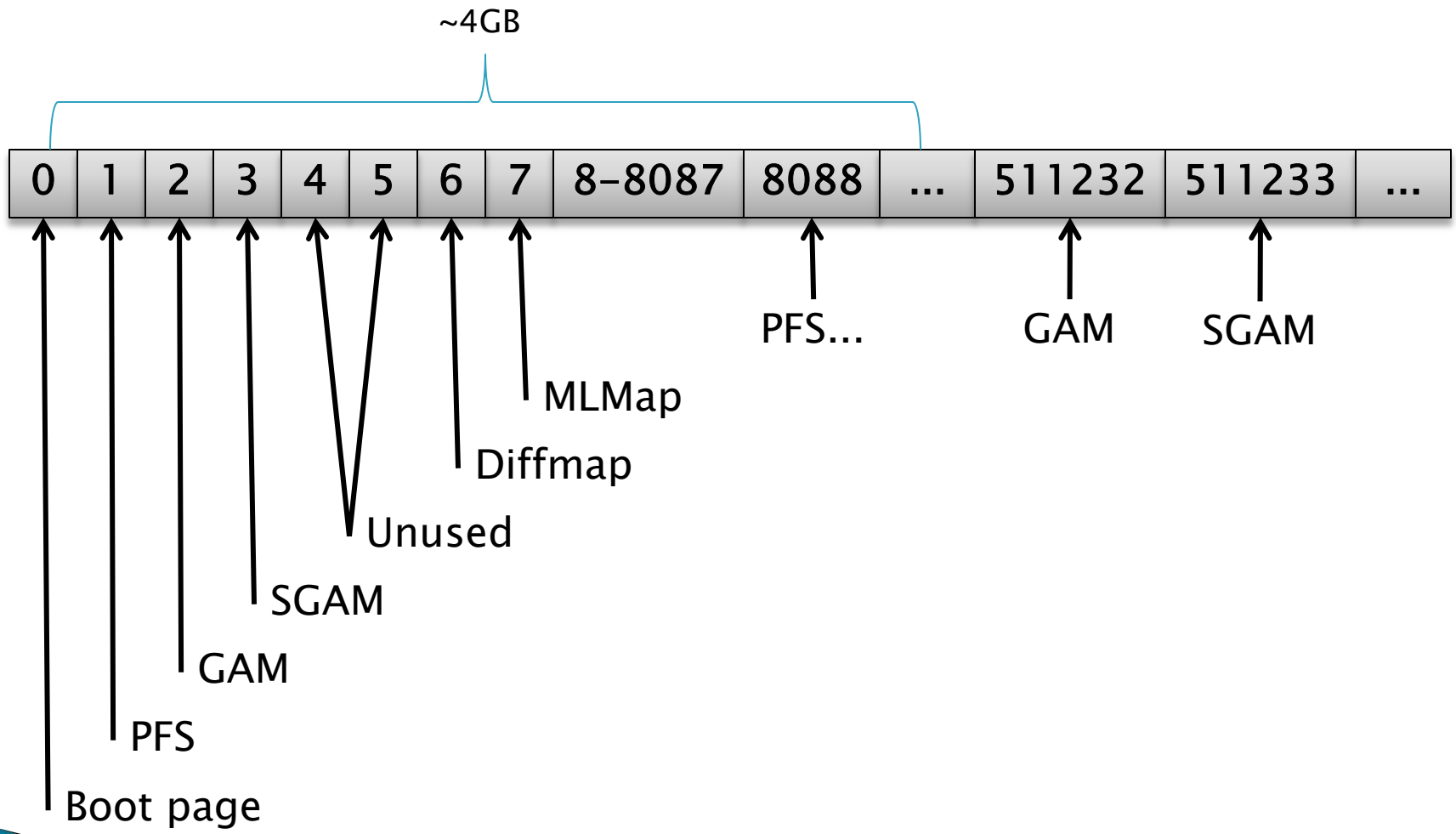
GAM	SGAM	Any IAM	Comments
0	0	0	Mixed extent with all pages allocated
0	0	1	Dedicated extent (must be allocated to only a single IAM page)
0	1	0	Mixed extent with >= 1 unallocated page
0	1	1	Invalid state
1	0	0	Unallocated extent
1	0	1	Invalid state
1	1	0	Invalid state
1	1	1	Invalid state

Page Free Space

- Tracks page free space across ~64MB
- 8088 bytes for bytemap
- Second page of data file, occurs every 8088 pages
- Only tracks heaps & lob/blob pages

Demo

MDF File at a glance



GLaDOS
A fatal exception E2 has occurred at E4E2:D7C5D5C4 in E4D 5E3(C9) .
D3C5C5C5. The current application will be terminated.
* Press any key to flood the facility with deadly neurotoxins.
* Press CTRL+ALT+DEL again to reinstate testing. You will lose any
non-vital personnel and their progress through the current test.
Press any key to continue _

Key takeaways

- Limited scope, but no documentation
- Know the internals => deduce the rest
- Make sound schema decisions
- Last measure data recovery
- Please be careful!

Mark S. Rasmussen
improve.dk
@improvedk