

GROUPBY 2020 | MAY 12-13

Let's Dive into SQL Server I/O
To Improve T-SQL Performance

Andy Yun
Principal Solutions Engineer - SentryOne





Andy Yun

Principal Solutions Engineer

- SQL Server DBA & DB Developer
- Chicago Suburban User Group Chapter Leader
- Chicago SQL Association – Director-at-Large
- Working with SQL Server since 2001
- Speaking since Early 2014
- Microsoft MVP (2017-2018)



@SQLBek - sqlbek@gmail.com

<https://blogs.sentryone.com/andyyun/>

<https://www.github.com/sqlbek/>

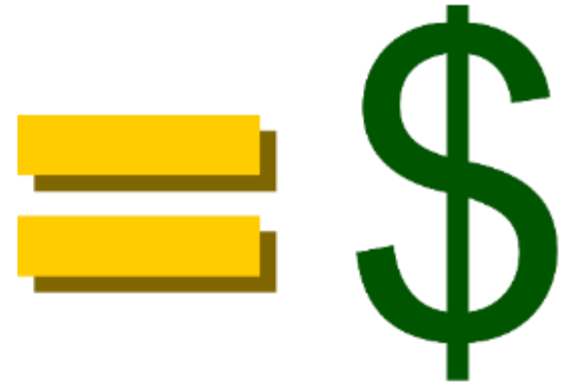
Setting the Stage



Pack It Up!



Are We There Yet?



Have You Ever...

- Had a query only returns 5 records, why is it slow?
- Had a query typically runs in 5 seconds, but is now running slow?
- Wondered what your query is really doing behind the scenes?

Today's Agenda

1. Internals: Query Lifecycle
2. Internals: Data Structures
3. Internals: Navigating
Heaps & Indexes
4. T-SQL & I/O Scenarios

Goal:

To explore SQL Server I/O, to better understand query performance issues

Session Level:

- 200 - 300
- Splash of 400

Internals: Query Lifecycle

I Like Analogies



T-SQL Query



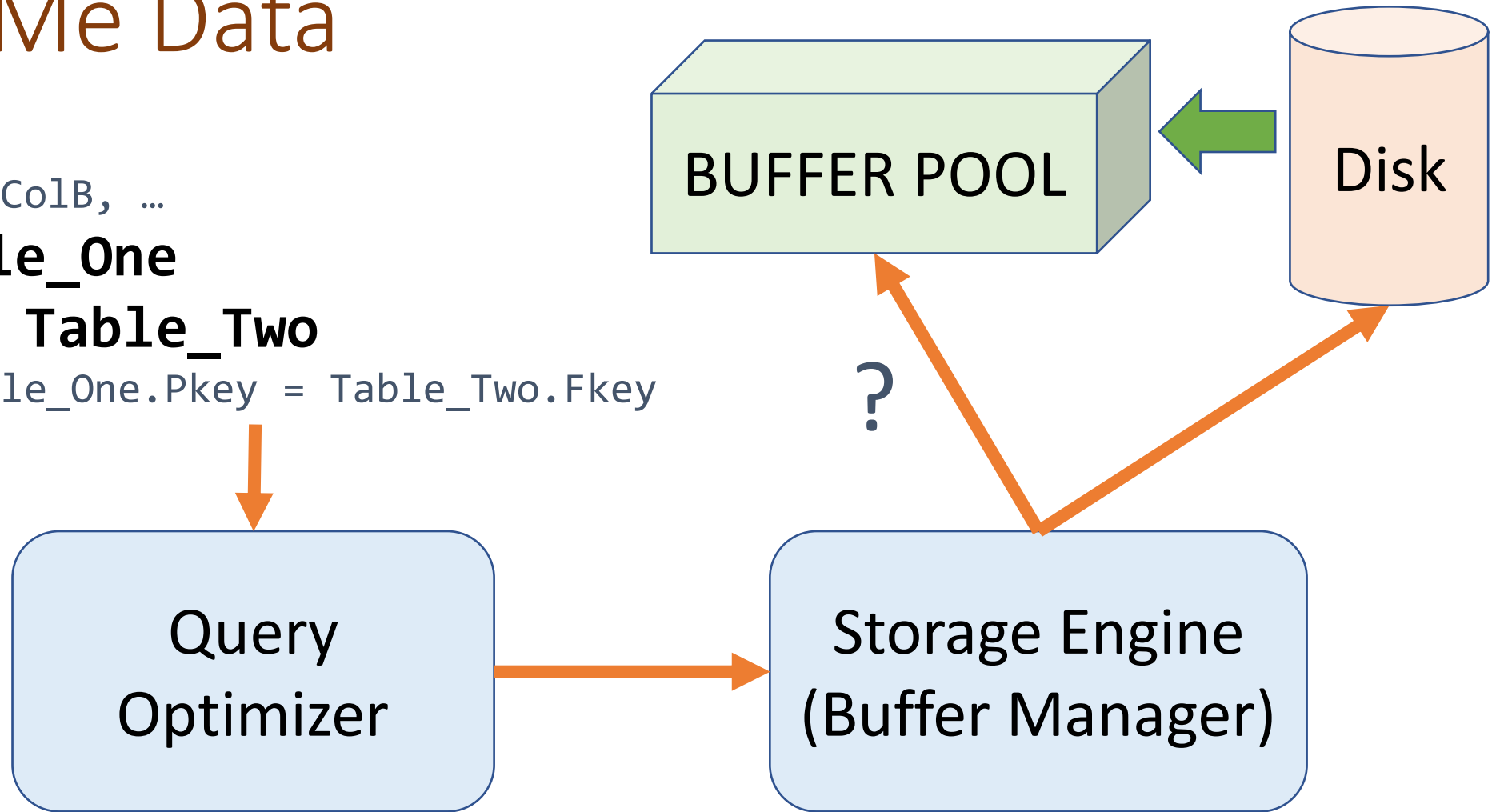
Query
Optimizer



Storage
Engine

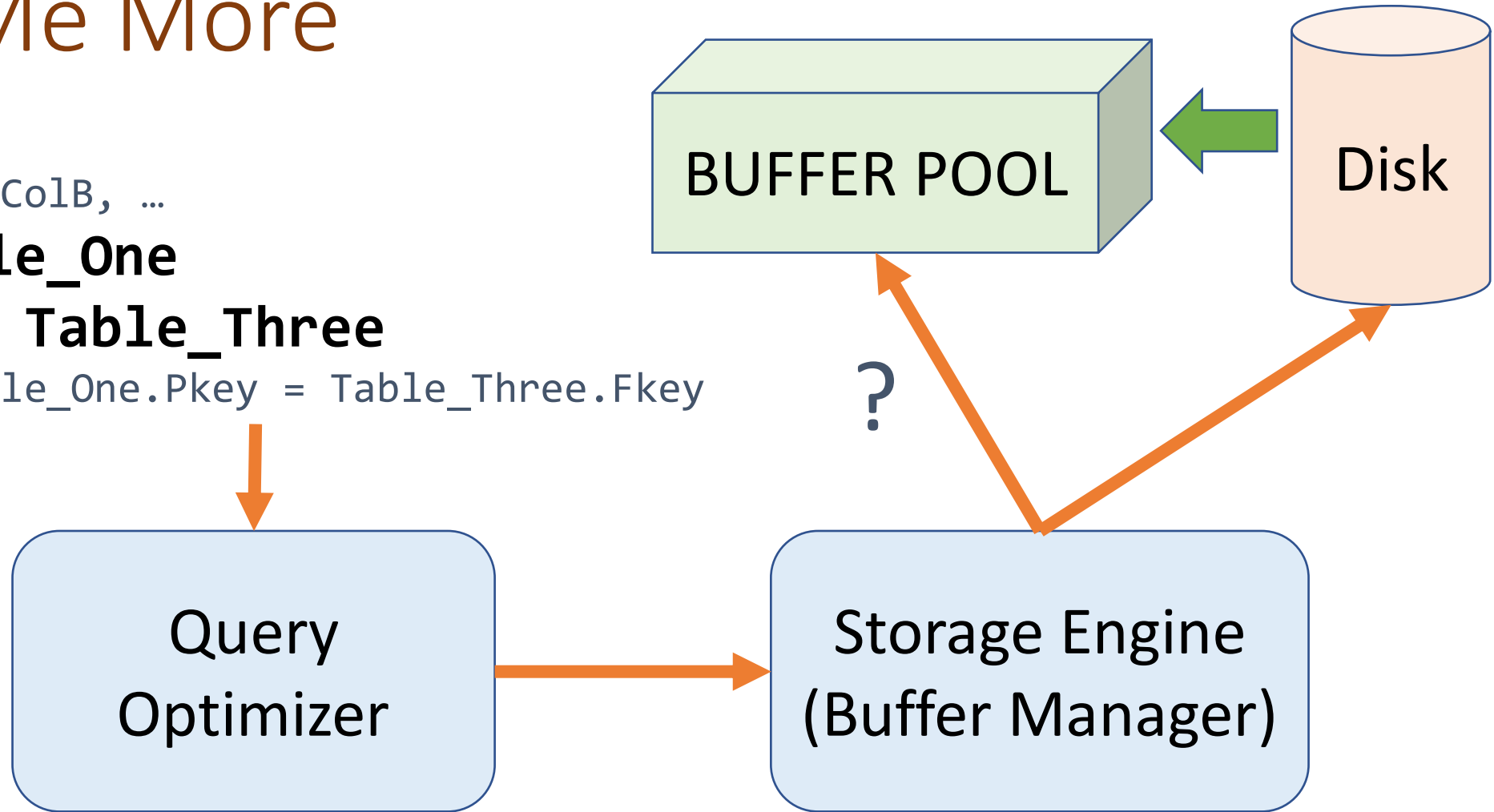
Give Me Data

```
SELECT  
    ColA, ColB, ...  
FROM Table_One  
INNER JOIN Table_Two  
    ON Table_One.Pkey = Table_Two.Fkey
```



Get Me More

```
SELECT  
    ColA, ColB, ...  
FROM Table_One  
INNER JOIN Table_Three  
    ON Table_One.Pkey = Table_Three.Fkey
```



Logical Reads vs Physical Reads

Which is more important?

Disregard logical reads &
only focus on physical?

How do the two relate?



Carrying a box
to be examined



Examining the
contents of the box

Where Did We Pack That Widget?!

1. Room = empty



2. Bring in 5 boxes to examine
= 5 physical reads



3. Open & examine 5 boxes
= 5 logical reads

1. Room = 5 boxes



2. Bring in 3 boxes to examine
= 3 physical reads



3. Open & examine 8 boxes
= 8 logical reads

Key Takeaways

Disk I/O is slower than Buffer Pool I/O

All work is executed against data pages in the Buffer Pool

Internals: Data Structures

Just Get Me One Record

```
SELECT TOP 1  
    ColA, ColB  
FROM Table_One
```

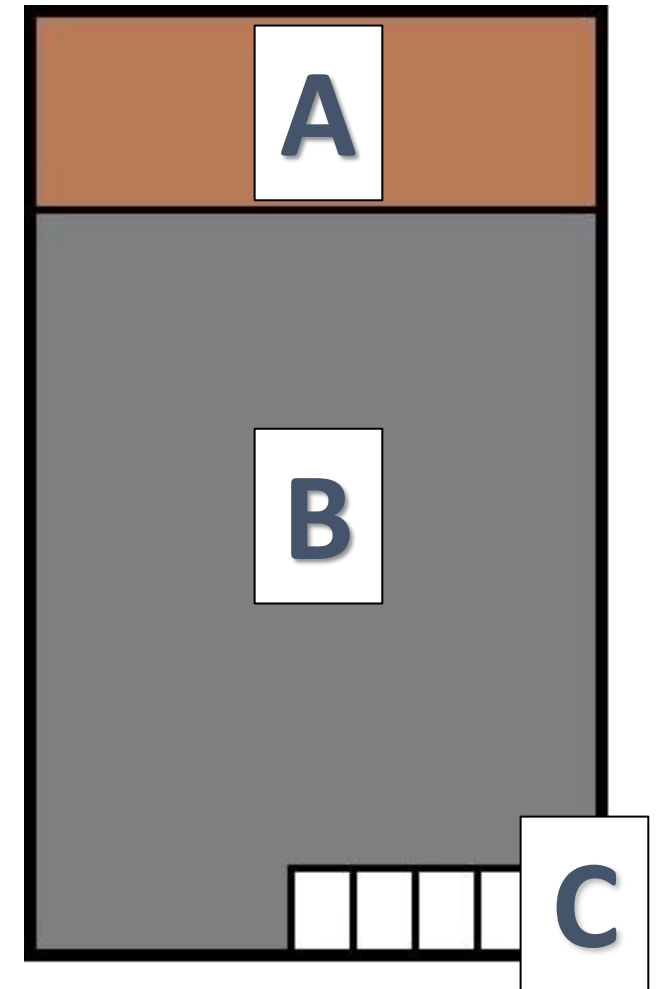
?

Data Page: Overview

Comprised of 3 components:

- A. Page Header: 96 bytes
- B. Data Records
- C. Record or Slot Array:
2 bytes per record

All Pages have a Fixed Size of 8KB or 8,192 bytes

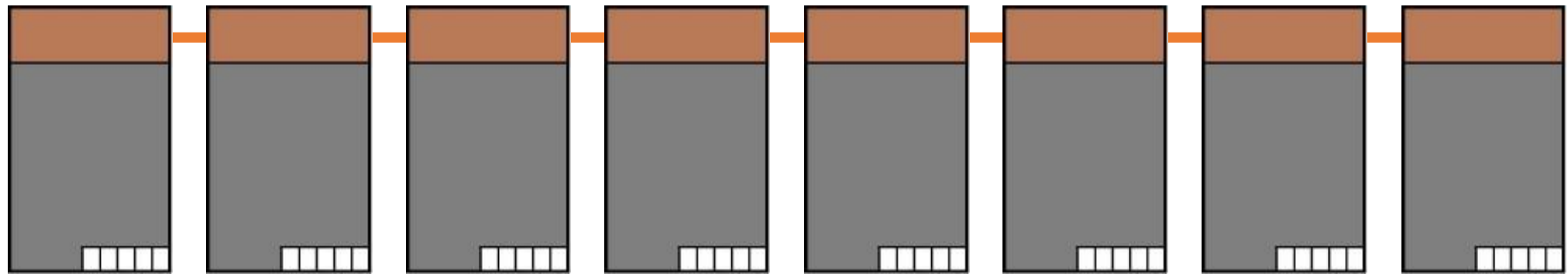


A collection of data pages is...

```
CREATE TABLE MyTable (  
    MyTableID INT,  
    ColumnOne VARCHAR(15),  
    DateCreated DATE  
);
```

?

Heap

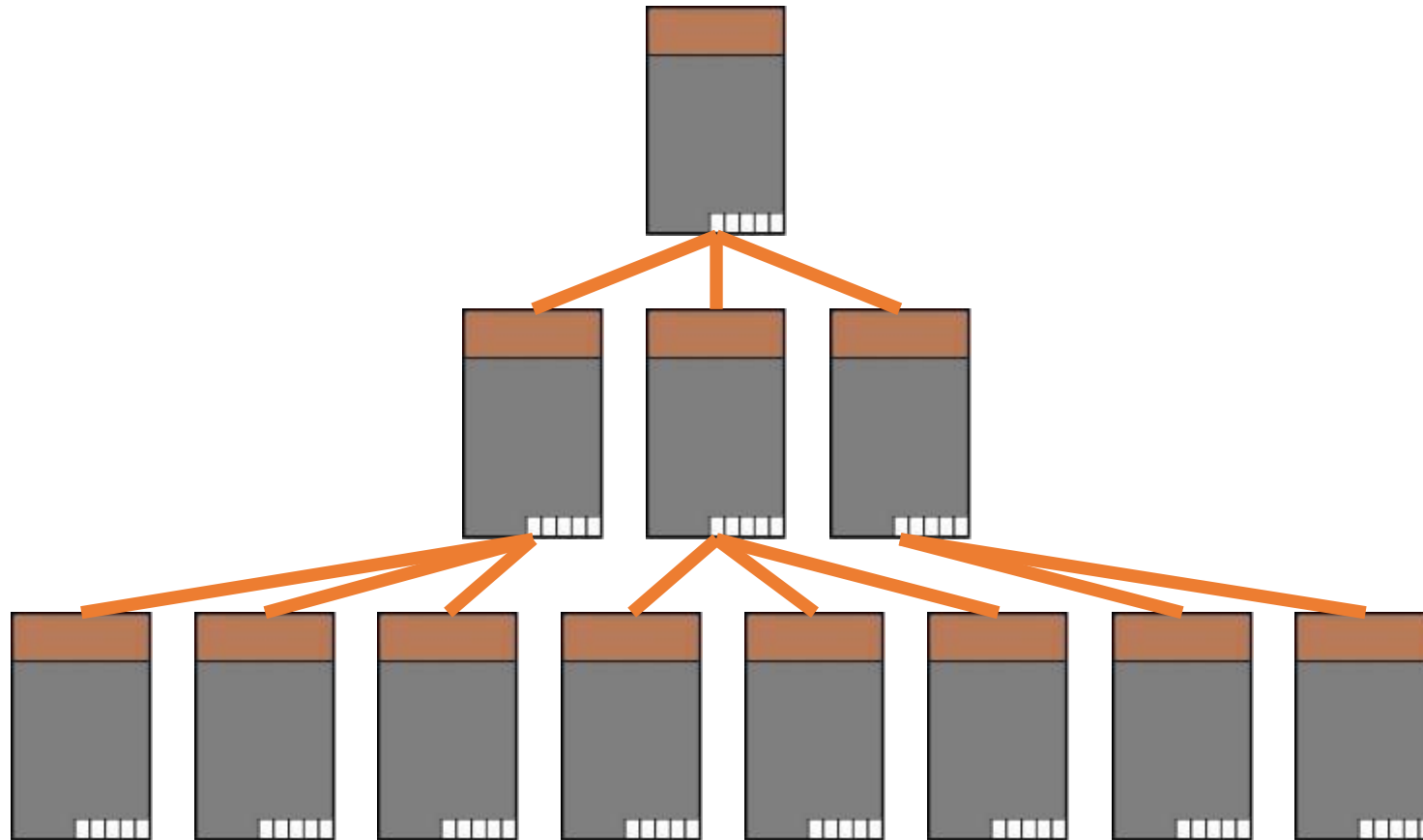


I Want Order!

```
CREATE CLUSTERED INDEX  
    CK_MyTable  
ON MyTable (  
    MyTableID INT  
);
```

?

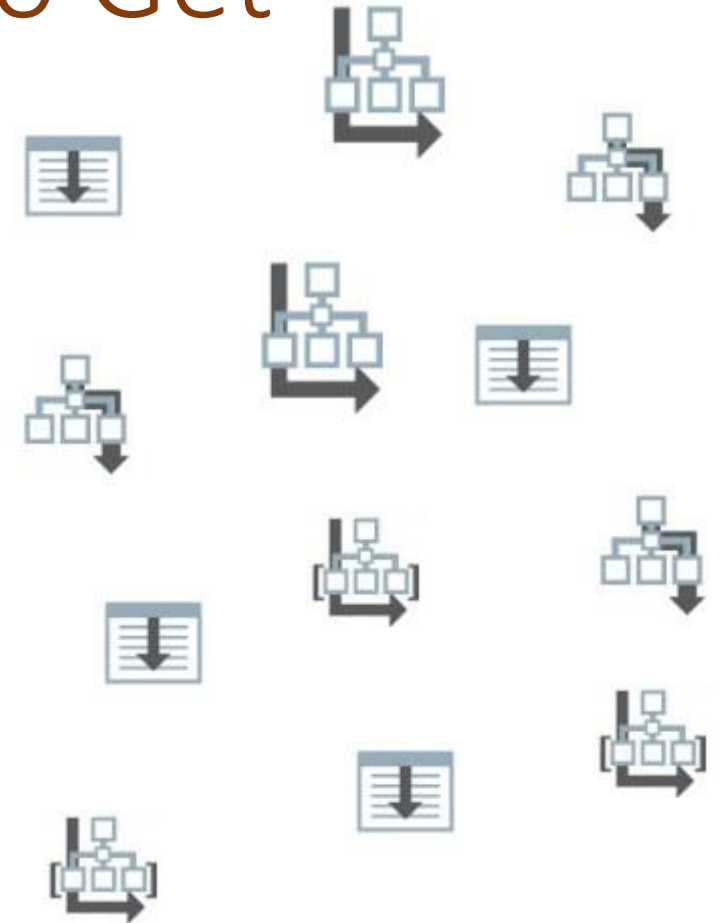
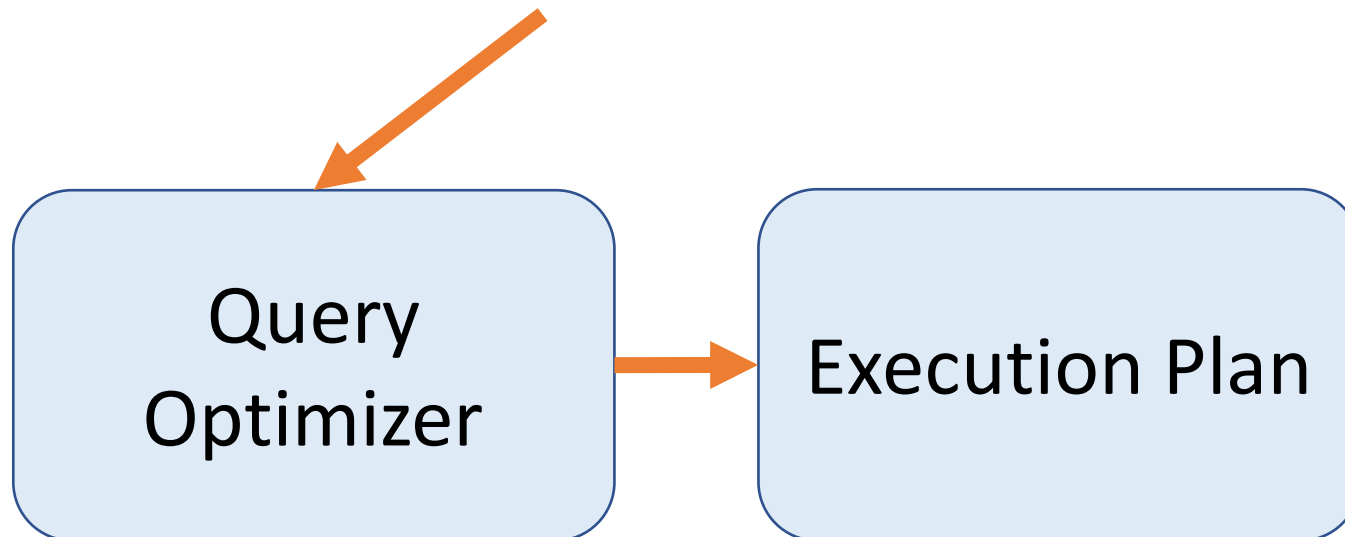
B-Tree – Clustered & Non-Clustered Idx



Internals: Navigating Heaps & Indexes

T-SQL: How To Get vs What To Get

```
SELECT  
    ColA, ColB, ...  
FROM Table_One  
INNER JOIN Table_Two  
    ON Table_One.Pkey = Table_Two.Fkey
```



Customer Table

```
CREATE TABLE dbo.Customer (  
    FirstName          VARCHAR(50),  
    LastName           VARCHAR(50),  
    Address             VARCHAR(50),  
    City               VARCHAR(50),  
    State              CHAR(2),  
    ZipCode            CHAR(5),  
    Email              VARCHAR(50),  
    FirstVisit         DATE,  
    RepeatCustomer     BIT  
);
```


Heap

Simpson, Mary,...
.,2003-01-08,0
Chavez, Esme,...,
2003-09-26,0
Ramirez, Lori,...,
2002-01-17,0

Martinez, Luis,...
,2003-08-19,0
Simon, Paul,...,2
002-06-28,1
Hu, Li,...,2002-
01-28,0

Raji, Bo,...,2004-
04-20,1
Agbonile, Lisa,...,
2004-01-01,0
Sandberg, Ryne,.
...,2003-11-06,1

Chow, Matt,...,20
01-11-12,0
Umeda, Umesh,.
...,2003-10-13,1
Howard, Kate,...,
2002-02-14,1

McDon, Jynell,...,
2004-03-09,0
Sanchez, Ray,...,2
002-01-17,0
Adams, James,...
,2002-02-10,1

Truempy, Nick,...
,2001-07-28,0
Mehta, Rit,...,20
04-05-19,1
Raman, Arthur,...
,2001-12-20,1

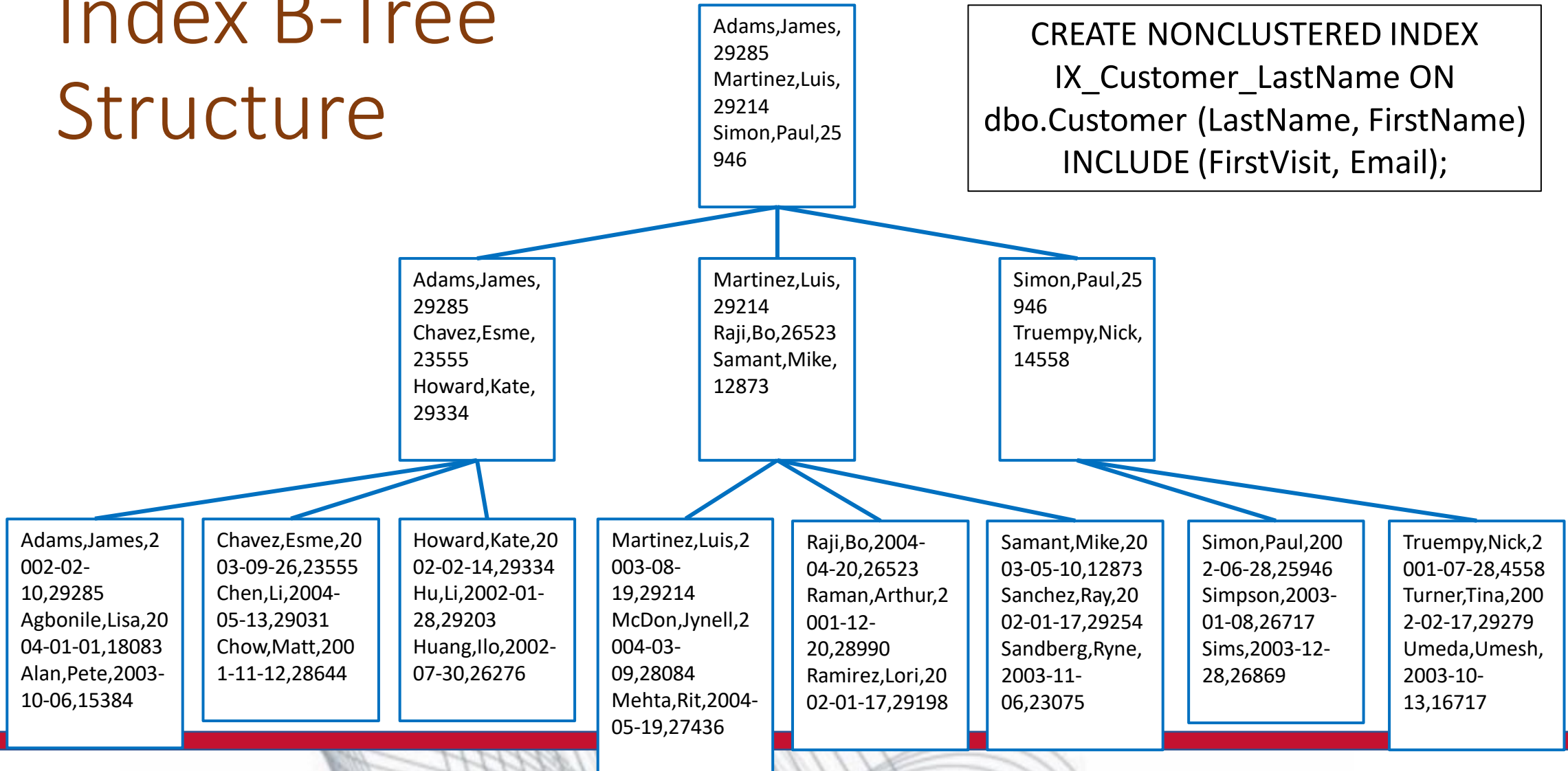
Huang, Ilo,...,200
2-07-30,1
Samant, Mike,...,
2003-05-10,1
Alan, Pete,...,200
3-10-06,1

Sims, Beth,...,20
03-12-28,0
Chen, Li,...,2004-
05-13,1
Turner, Tina,...,2
002-02-17,1

```
CREATE TABLE dbo.Customer (  
    CustomerID          INT IDENTITY(1, 1) PRIMARY KEY CLUSTERED,  
    FirstName           VARCHAR(50),  
    LastName            VARCHAR(50),  
    Address             VARCHAR(50),  
    City                VARCHAR(50),  
    State               CHAR(2),  
    ZipCode             CHAR(5),  
    Email               VARCHAR(50),  
    FirstVisit          DATE,  
    RepeatCustomer      BIT  
);  
  
CREATE NONCLUSTERED INDEX IX_Customer_LastName  
    ON dbo.Customer (LastName, FirstName)  
    INCLUDE (FirstVisit, Email);
```

Index B-Tree Structure

```
CREATE NONCLUSTERED INDEX  
IX_Customer_LastName ON  
dbo.Customer (LastName, FirstName)  
INCLUDE (FirstVisit, Email);
```



Root Level

```
CREATE NONCLUSTERED INDEX  
    IX_Customer_LastName  
    ON dbo.Customer  
    (LastName, FirstName)  
    INCLUDE (FirstVisit, Email);
```

- Adams,James,29285
- Martinez,Luis,29214
- Simon,Paul,25946

Intermediate Level

```
CREATE NONCLUSTERED INDEX  
IX_Customer_LastName  
ON dbo.Customer  
(LastName, FirstName)  
INCLUDE (FirstVisit, Email);
```

- Adams,James,29285
- Chavez,Esme,23555
- Howard,Kate,29334

- Adams,James,29285
- Martinez,Luis,29214
- Simon,Paul,25946

- Martinez,Luis,29214
- Raji,Bo,26523
- Samant,Mike,12873

Leaf Level

```
CREATE NONCLUSTERED INDEX  
IX_Customer_LastName  
ON dbo.Customer  
(LastName, FirstName)  
INCLUDE (FirstVisit, Email);
```

- Adams,James,2002-02-10,jada@x.com,29285
- Agbonile,Lisa,2004-01-01,lagb@x.com,18083
- Alan,Pete,2003-10-06,pala@x.com15384

- Adams,James,29285
- Chavez,Esme,23555
- Howard,Kate,29334

- Chavez,Esme,2003-09-26,echa@x.com,23555
- Chen,Li,2004-05-13,lche@x.com29031
- Chow,Matt,2001-11-12,mcho@x.com28644

Querying a Heap

```
SELECT
  FirstName, LastName, FirstVisit
FROM Customer
WHERE LastName = 'Dean'
```

Simpson,Mary,...
.,2003-01-08,0
Chavez,Esme,...,
2003-09-26,0
Ramirez,Lori,...,
2002-01-17,0

Martinez,Luis,...
,2003-08-19,0
Simon,Paul,...,2
002-06-28,1
Hu,Li,...,2002-
01-28,0

Raji,Bo,...,2004-
04-20,1
Agbonile,Lisa,...,
2004-01-01,0
Sandberg,Ryne,.
...,2003-11-06,1

Chow,Matt,...,20
01-11-12,0
Umeda,Umesh,.
...,2003-10-13,1
Howard,Kate,...,
2002-02-14,1

McDon,Jynell,...,
2004-03-09,0
Sanchez,Ray,...,2
002-01-17,0
Adams,James,...
,2002-02-10,1

Truempy,Nick,...
,2001-07-28,0
Mehta,Rit,...,20
04-05-19,1
Raman,Arthur,...
,2001-12-20,1

Huang,Ilo,...,200
2-07-30,1
Samant,Mike,...,
2003-05-10,1
Alan,Pete,...,200
3-10-06,1

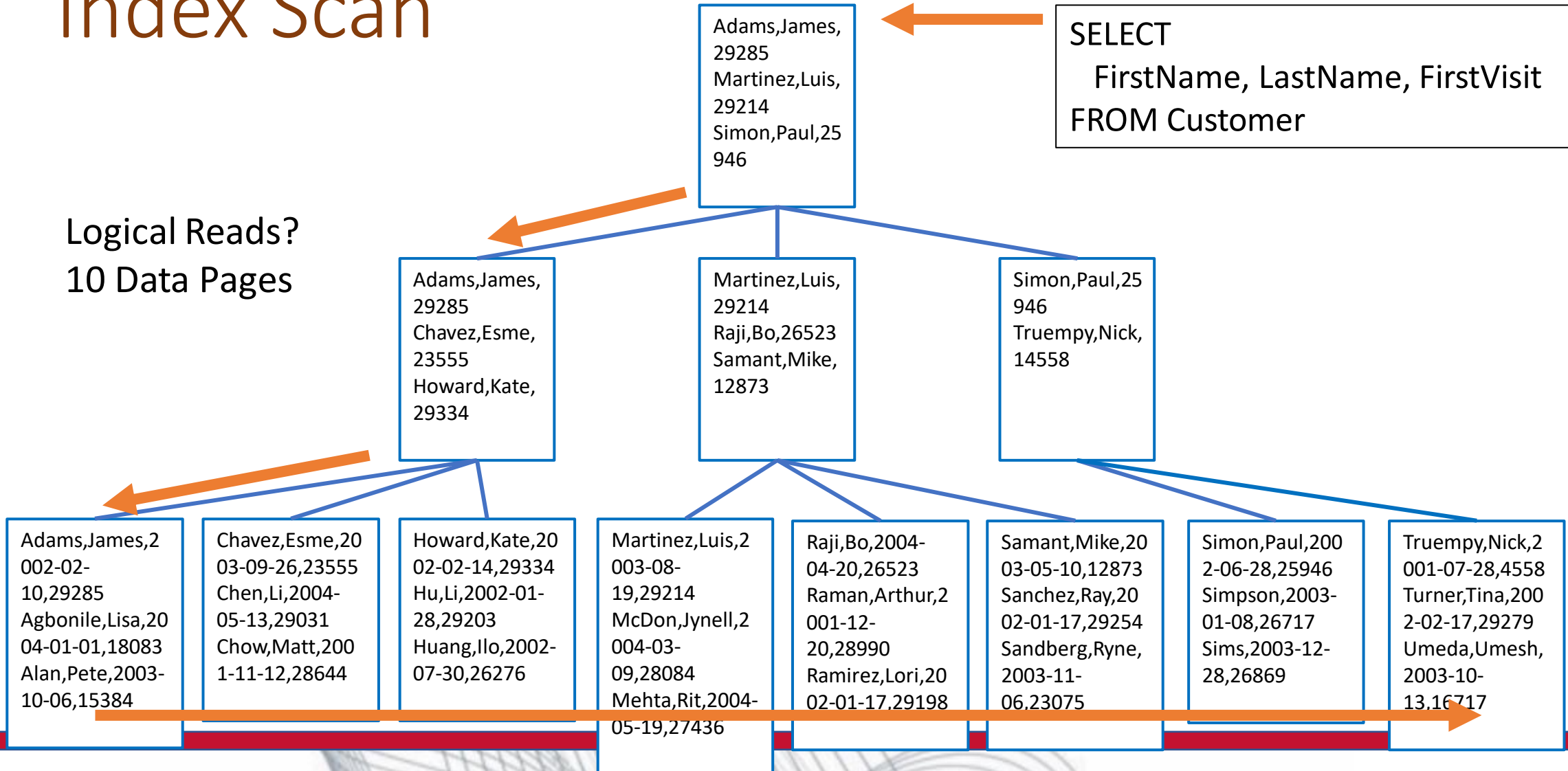
Sims,Beth,...,20
03-12-28,0
Chen,Li,...,2004-
05-13,1
Turner,Tina,...,2
002-02-17,1

Logical Reads?
8 Data Pages

Index Scan

```
SELECT  
  FirstName, LastName, FirstVisit  
FROM Customer
```

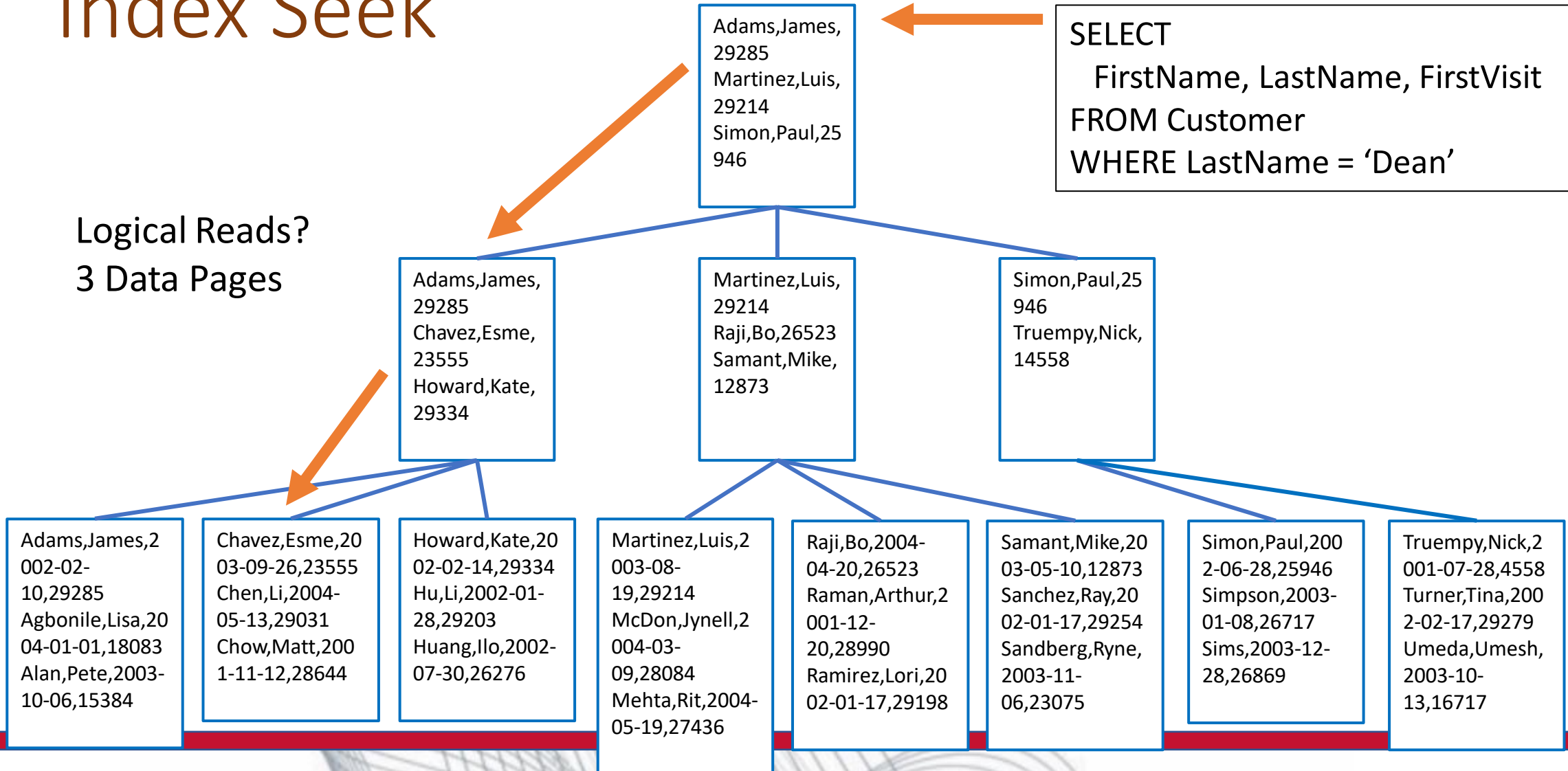
Logical Reads?
10 Data Pages



Index Seek

```
SELECT  
  FirstName, LastName, FirstVisit  
FROM Customer  
WHERE LastName = 'Dean'
```

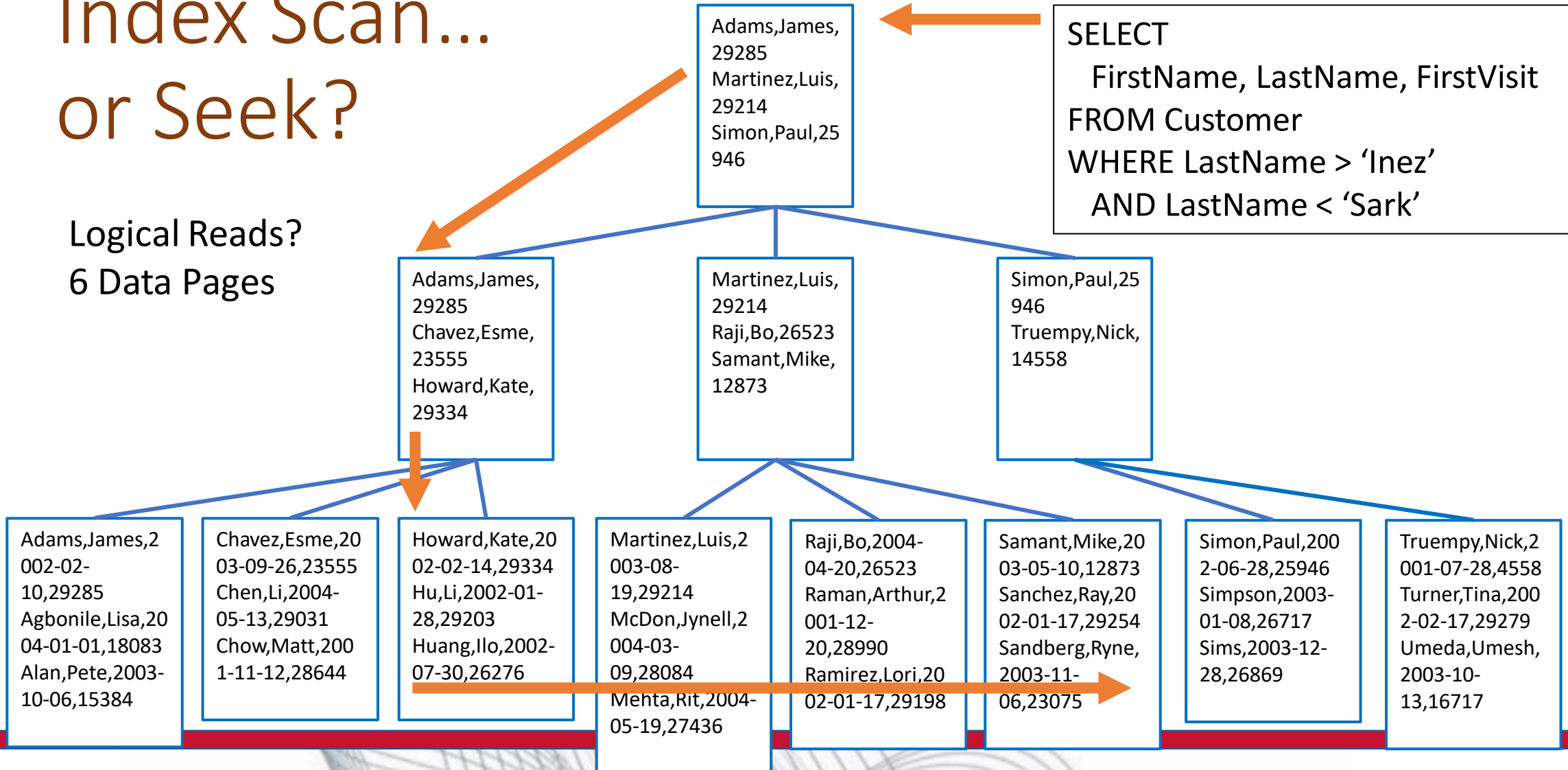
Logical Reads?
3 Data Pages



Index Scan... or Seek?

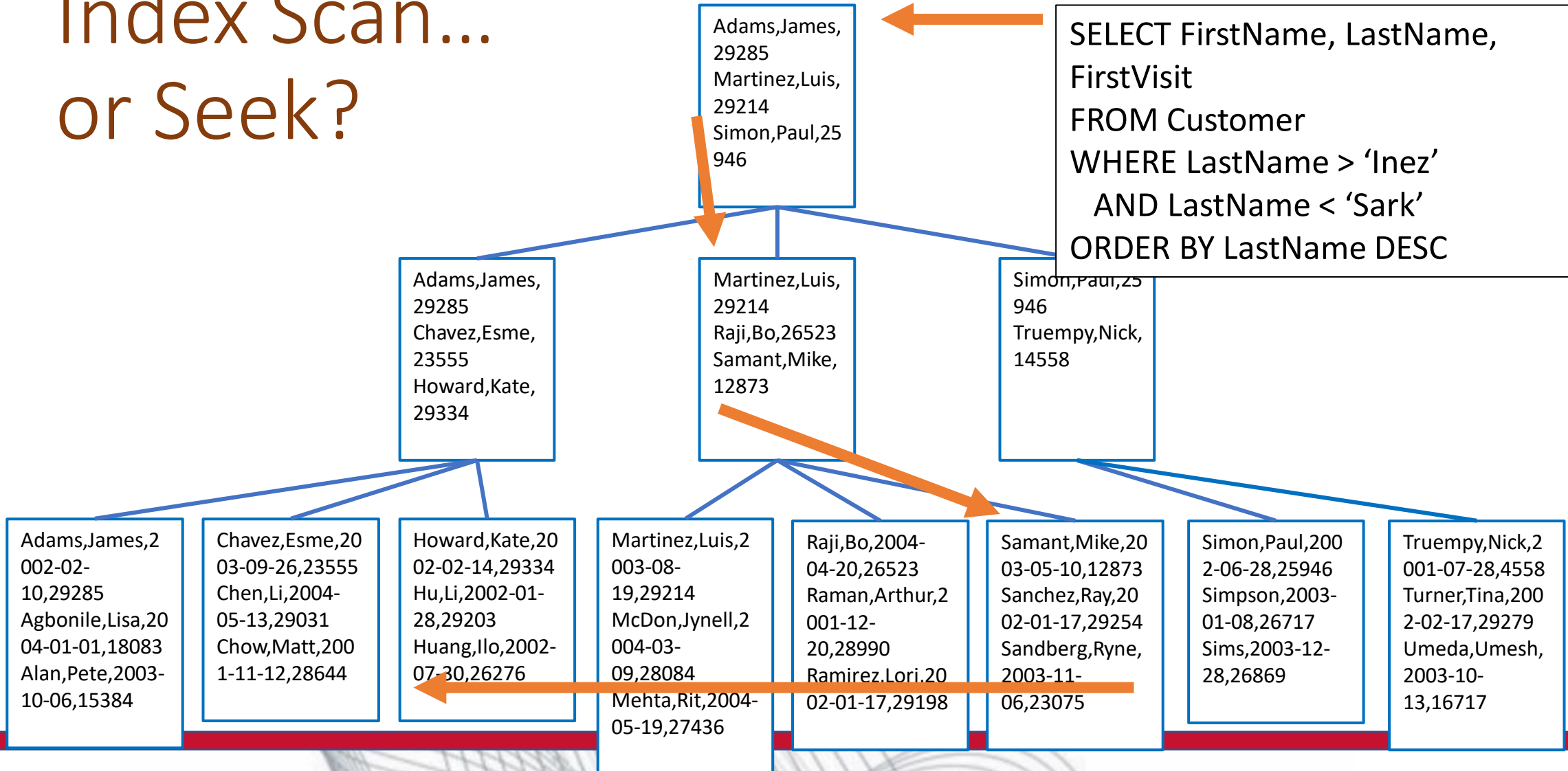
Logical Reads?
6 Data Pages

```
SELECT
  FirstName, LastName, FirstVisit
FROM Customer
WHERE LastName > 'Inez'
  AND LastName < 'Sark'
```



Index Scan... or Seek?

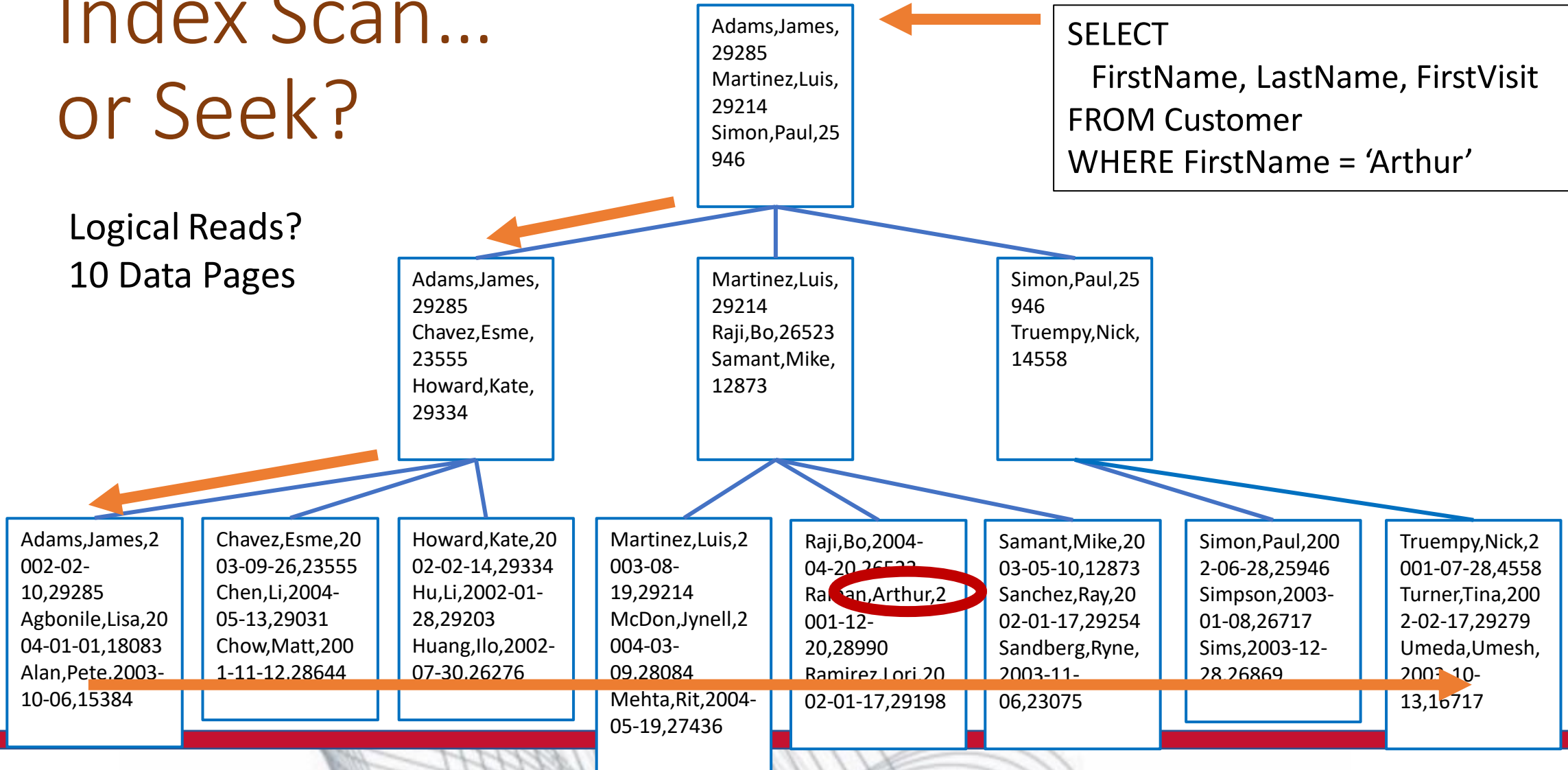
```
SELECT FirstName, LastName,  
FirstVisit  
FROM Customer  
WHERE LastName > 'Inez'  
AND LastName < 'Sark'  
ORDER BY LastName DESC
```



Index Scan... or Seek?

Logical Reads?
10 Data Pages

```
SELECT  
  FirstName, LastName, FirstVisit  
FROM Customer  
WHERE FirstName = 'Arthur'
```

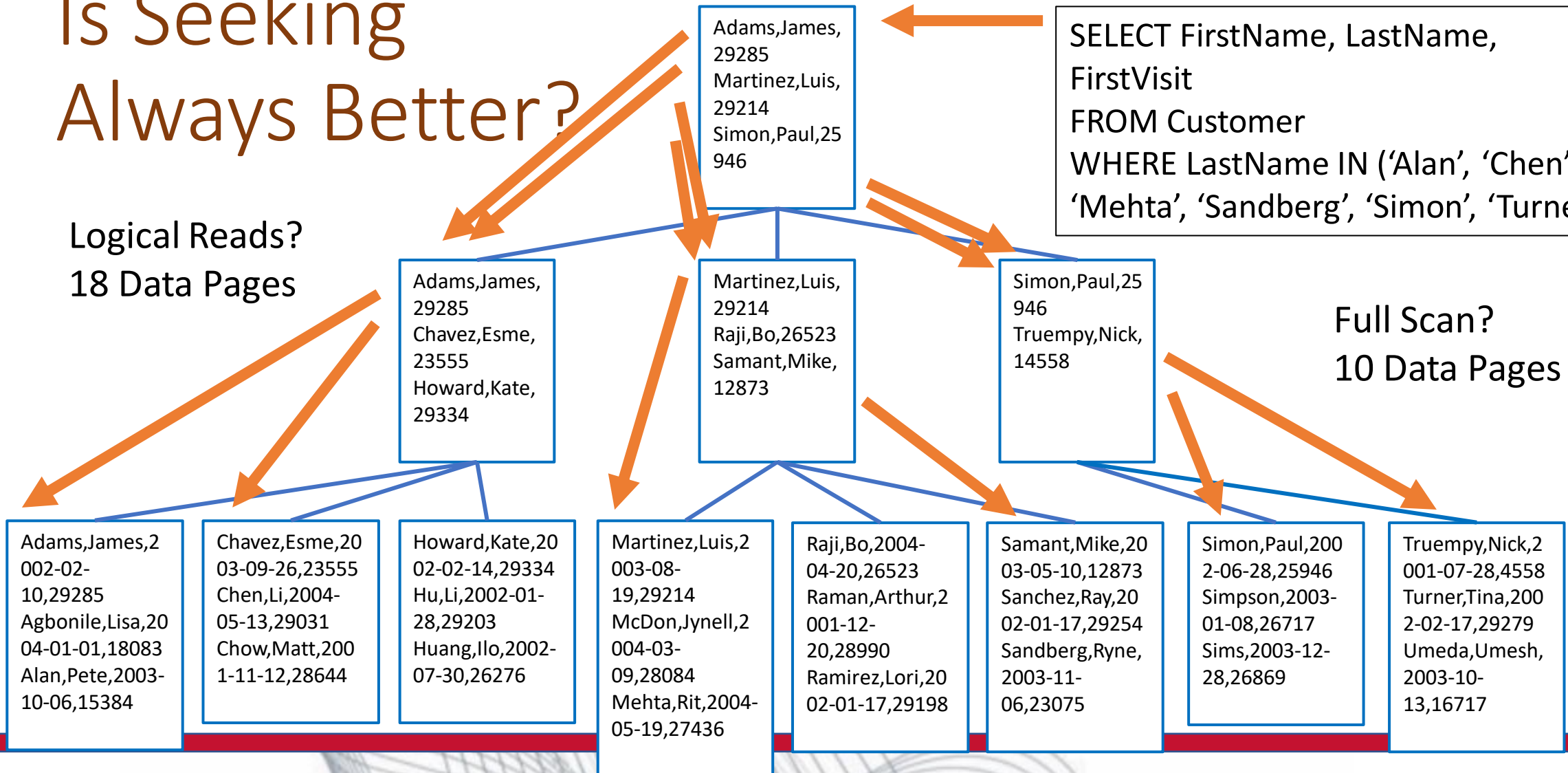


Is Seeking Always Better?

Logical Reads?
18 Data Pages

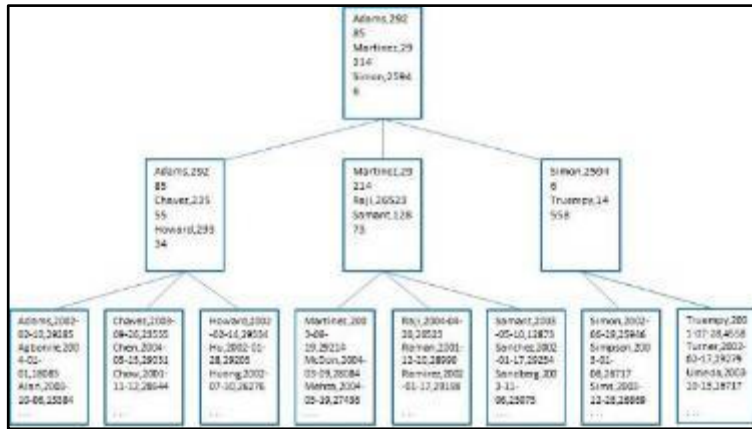
```
SELECT FirstName, LastName,  
FirstVisit  
FROM Customer  
WHERE LastName IN ('Alan', 'Chen',  
'Mehta', 'Sandberg', 'Simon', 'Turner')
```

Full Scan?
10 Data Pages



Key Lookup

```
SELECT  
  FirstName, LastName,  
  FirstVisit, City, State  
FROM Customer  
WHERE LastName = 'Huang'
```



Adams,James,
29285
Martinez,Luis,
29214
Simon,Paul,25
946

Martinez,Luis,
29214
Raji,Bo,26523
Samant,Mike,
12873

Simon,Paul,25
946
Truempy,Nick,
14558

Adams,James,2
002-02-
10,29285
Agbonile,Lisa,20
04-01-01,18083
Alan,Pete,2003-
10-06,15384

Chavez,Esme,20
03-09-26,25555
Chen,Li,2004-
05-13,29031
Chow,Matt,200
1-11-12,28644

Howard,Kate,20
02-02-14,29334
Hu,Li,2002-01-
28,29203
Huang,Ilo,2002-
07-30,26276

Martinez,Luis,2
003-08-
19,29214
McDon,Jynell,2
004-03-
09,28084
Mehta,Rit,2004-
05-19,27436

Raji,Bo,2004-
04-20,26523
Raman,Arthur,2
001-12-
20,28990
Ramirez,Lori,20
02-01-17,29198

Samant,Mike,20
03-05-10,12873
Sanchez,Ray,20
02-01-17,29254
Sandberg,Ryne,
2003-11-
06,23075

Simon,Paul,200
2-06-28,25946
Simpson,2003-
01-08,26717
Sims,2003-12-
28,26869

Truempy,Nick,2
001-07-28,4558
Turner,Tina,200
2-02-17,29279
Umeda,Umesh,
2003-10-
13,16717

Index Operation Recap

- Index Key Order Matters. Can only seek on leading key column
- Add Appropriate Columns as Included Columns to “Cover a Query”
- Seeks are not necessarily better than Scans – Check Properties in Execution Plan

DEMO: T-SQL & I/O Scenarios

Parting Thoughts

Widen your perspective

Server resources are among your entire workload

Be mindful of “unseen” work being done to fulfil a query

Thank you!

<https://github.com/SQLBek>

Andy Yun | @SQLBek

ayun@sentryone.com | SQLBek@gmail.com

<http://blogs.sentryone.com/andyyun/>

<http://sqlbek.wordpress.com>

Want To Learn More About SentryOne?

Book a demo with me!

<http://www.sentryone.com/BookAndy/>