Database design

Advanced Topics

NIFR - 2024

Cardinality: Questions

Agenda

Learning objectives

- What happens when a relational database is queried?
  - Query stats & execution plan
- Indexing
  - Binary search trees
- Normalisation
- Views

HOW?
WHAT?

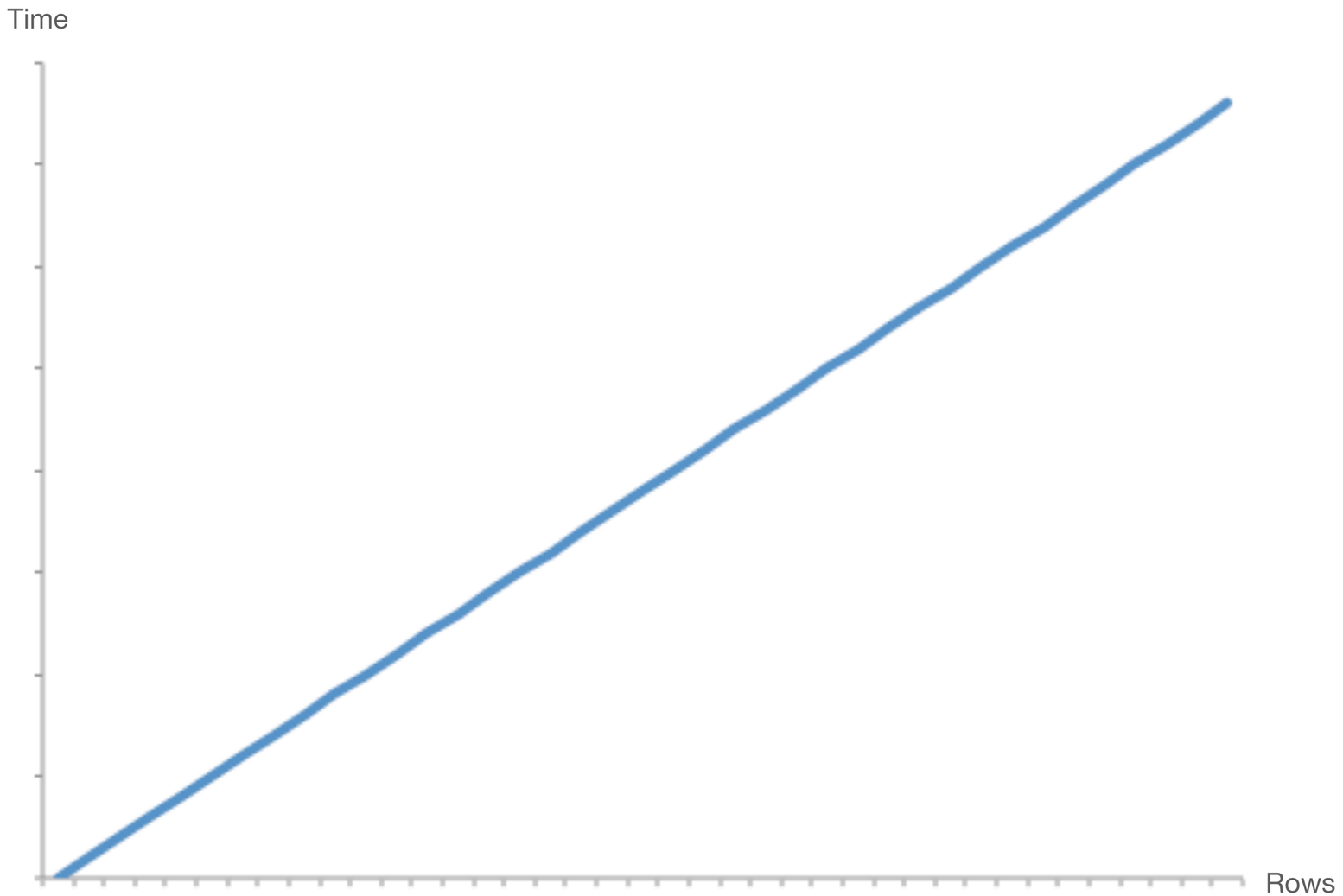
SELECT *
FROM spotify
WHERE artist =
‘Melanie C’

artist	song	duration_ms	explicit	year
Britney Spears	Oops!...I Did It Again	211160	0	2000
blink-182	All The Small Things	167066	0	1999
Faith Hill	Breathe	250546	0	1999
Bon Jovi	It's My Life	224493	0	2000
*NSYNC	Bye Bye Bye	200560	0	2000
Sisqo	Thong Song	253733	1	1999
Eminem	The Real Slim Shady	284200	1	2000
Robbie Williams	Rock DJ	258560	0	2000
Destiny's Child	Say My Name	271333	0	1999
Modjo	Lady - Hear Me Tonight	307153	0	2001
Gigi D'Agostino	L'Amour Toujours	238759	0	2011
Eiffel 65	Move Your Body - Ga…	268863	0	1999
Bomfunk MC's	Freestyler	306333	0	2000
Sting	Desert Rose	285960	0	1999
Melanie C	Never Be The Same…	294200	0	1999
Aaliyah	Try Again	284000	0	2002

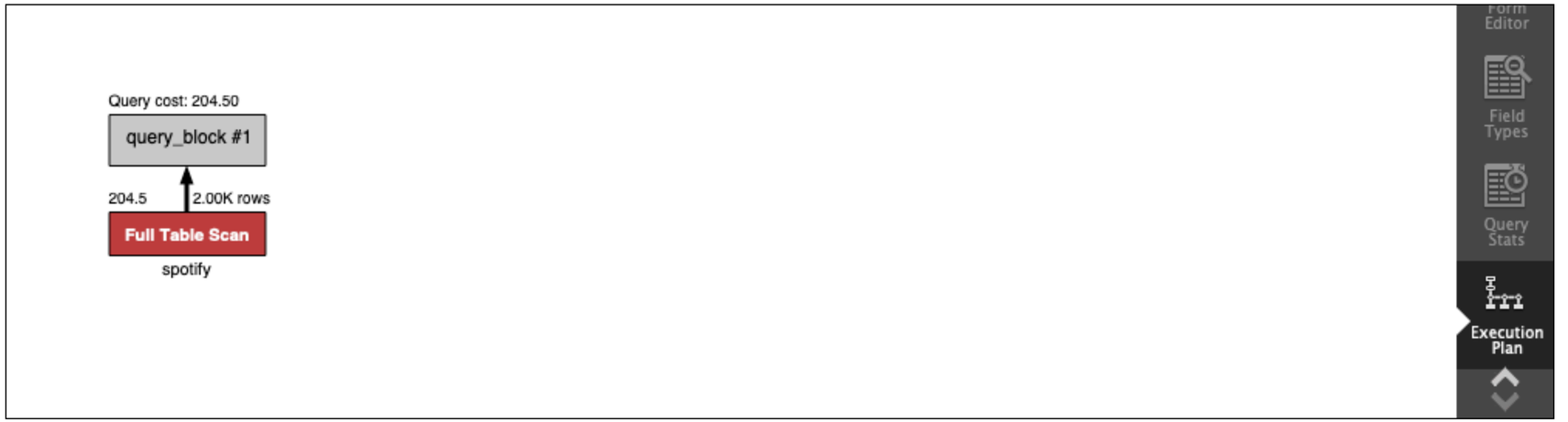
Iterating through 16 rows
Full table scan
Brute force

n = 16

artist	song	duration_ms	explicit	year
Britney Spears	Oops!...I Did It Again	211160	0	2000
blink-182	All The Small Things	167066	0	1999
Faith Hill	Breathe	250546	0	1999
Bon Jovi	It's My Life	224493	0	2000
*NSYNC	Bye Bye Bye	200560	0	2000
Sisqo	Thong Song	253733	1	1999
Eminem	The Real Slim Shady	284200	1	2000
Robbie Williams	Rock DJ	258560	0	2000
Destiny's Child	Say My Name	271333	0	1999
Modjo	Lady - Hear Me Tonight	307153	0	2001
Gigi D'Agostino	L'Amour Toujours	238759	0	2011
Eiffel 65	Move Your Body - Ga…	268863	0	1999
Bomfunk MC's	Freestyler	306333	0	2000
Sting	Desert Rose	285960	0	1999
Melanie C	Never Be The Same…	294200	0	1999
Aaliyah	Try Again	284000	0	2002


Time
Rows

```sql
SELECT *
FROM spotify
WHERE artist LIKE "C%";
```



Demo: Indexing artist names

Query plan is the DBEngine (InnoDB) at work

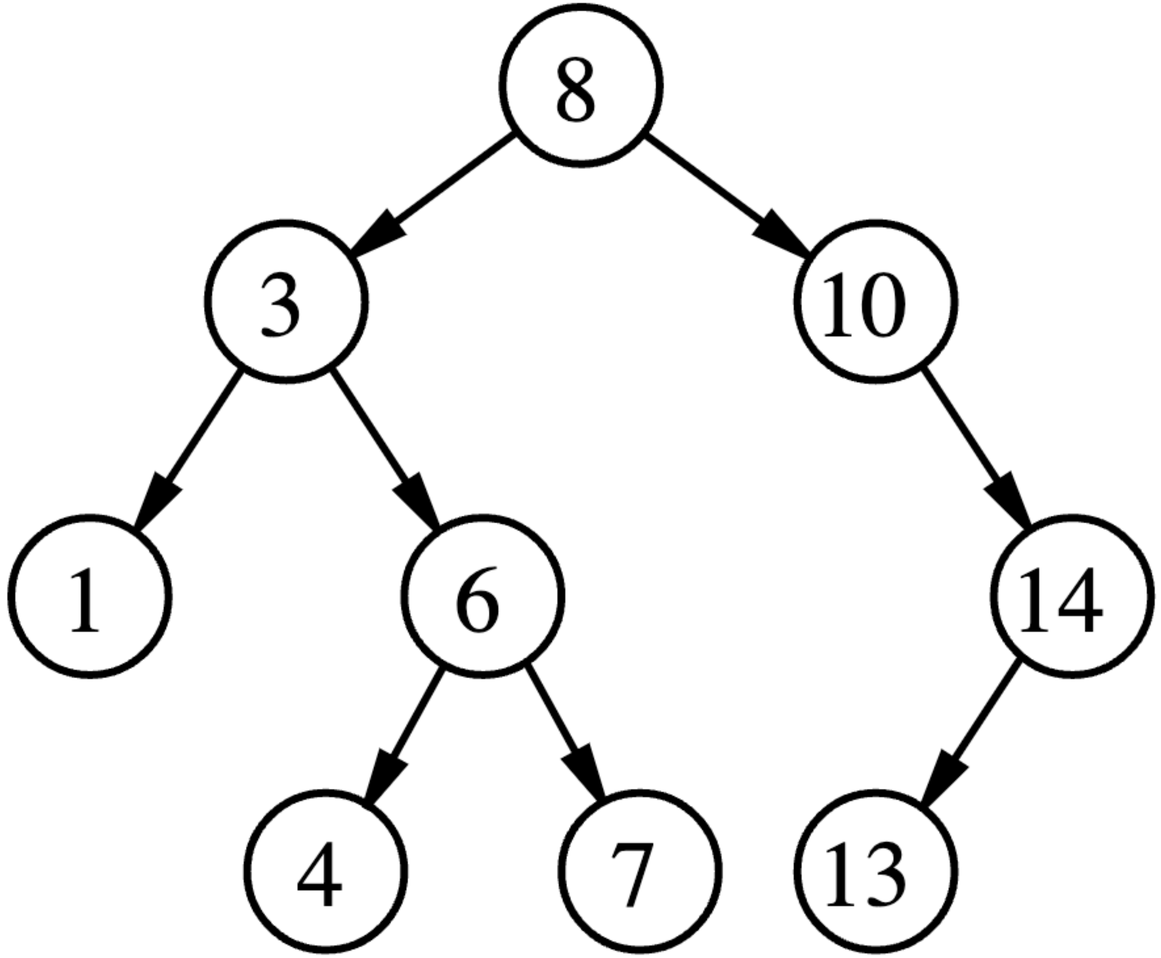
Declarative nature of MySQL

Iterating through 16 rows
Full table scan
Brute force

n = 16

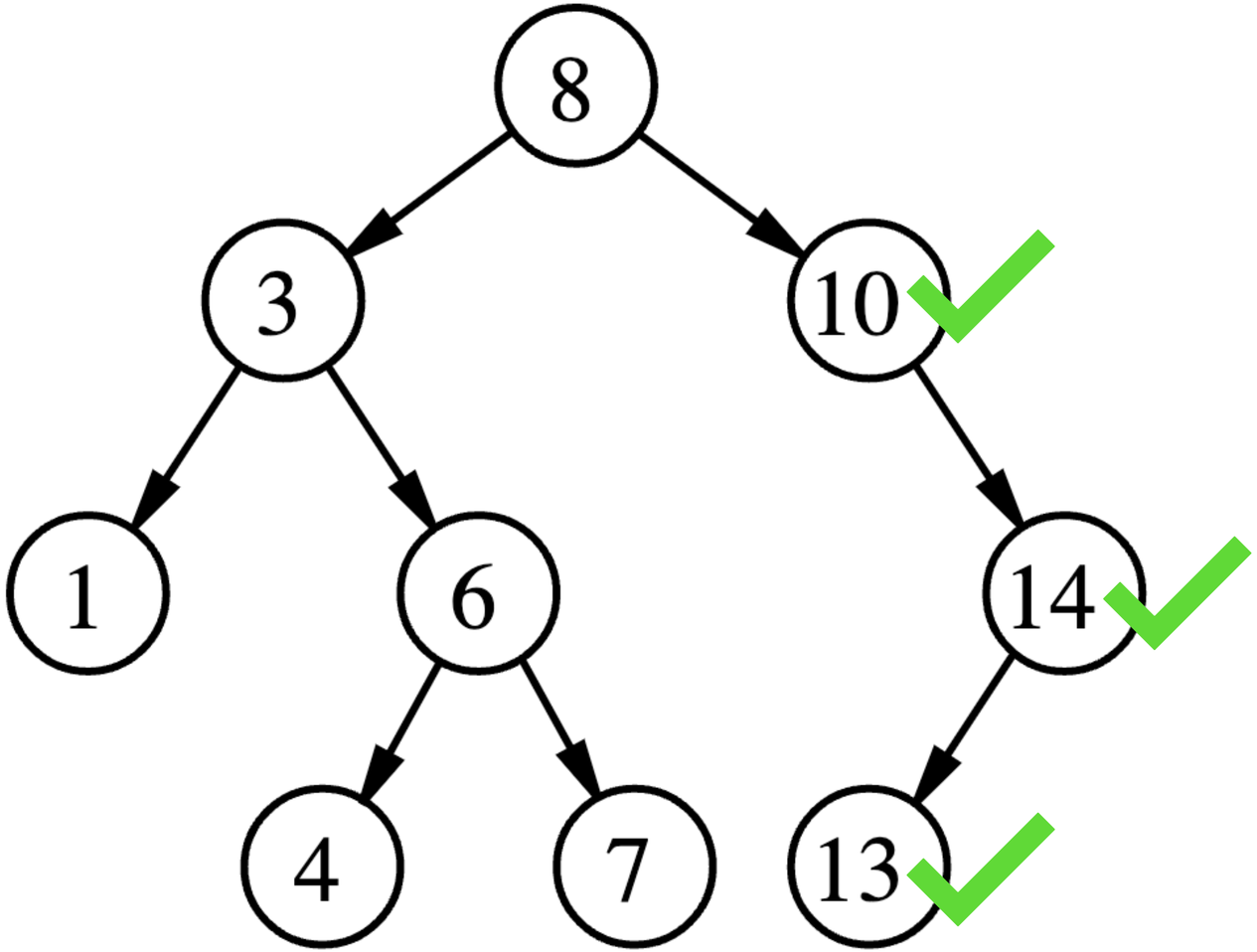
artist	song	duration_ms	explicit	year
Britney Spears	Oops!...I Did It Again	211160	0	2000
blink-182	All The Small Things	167066	0	1999
Faith Hill	Breathe	250546	0	1999
Bon Jovi	It's My Life	224493	0	2000
*NSYNC	Bye Bye Bye	200560	0	2000
Sisqo	Thong Song	253733	1	1999
Eminem	The Real Slim Shady	284200	1	2000
Robbie Williams	Rock DJ	258560	0	2000
Destiny's Child	Say My Name	271333	0	1999
Modjo	Lady - Hear Me Tonight	307153	0	2001
Gigi D'Agostino	L'Amour Toujours	238759	0	2011
Eiffel 65	Move Your Body - Ga…	268863	0	1999
Bomfunk MC's	Freestyler	306333	0	2000
Sting	Desert Rose	285960	0	1999
Melanie C	Never Be The Same…	294200	0	1999
Aaliyah	Try Again	284000	0	2002

Binary Search Tree



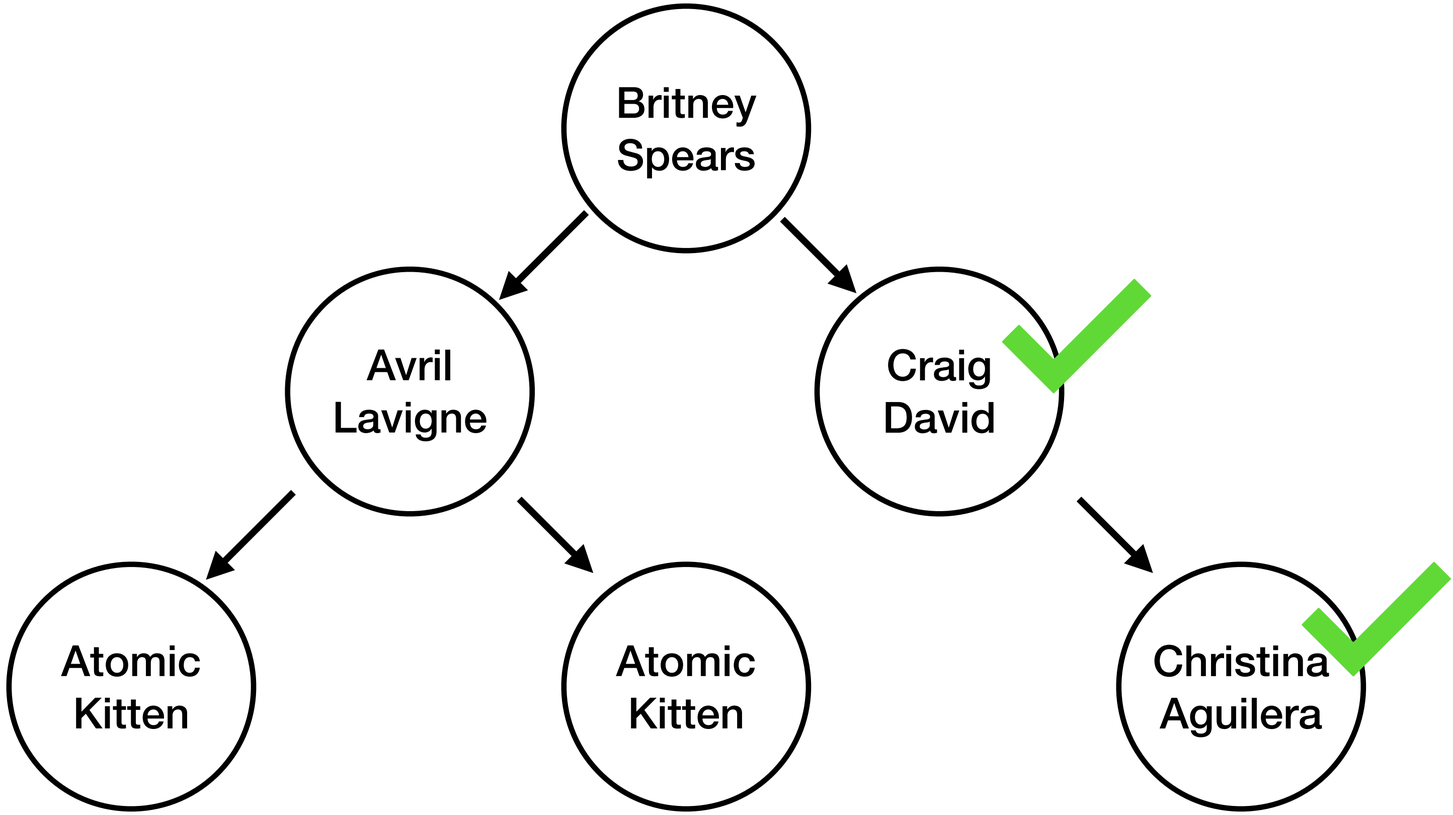
Binary Search Tree

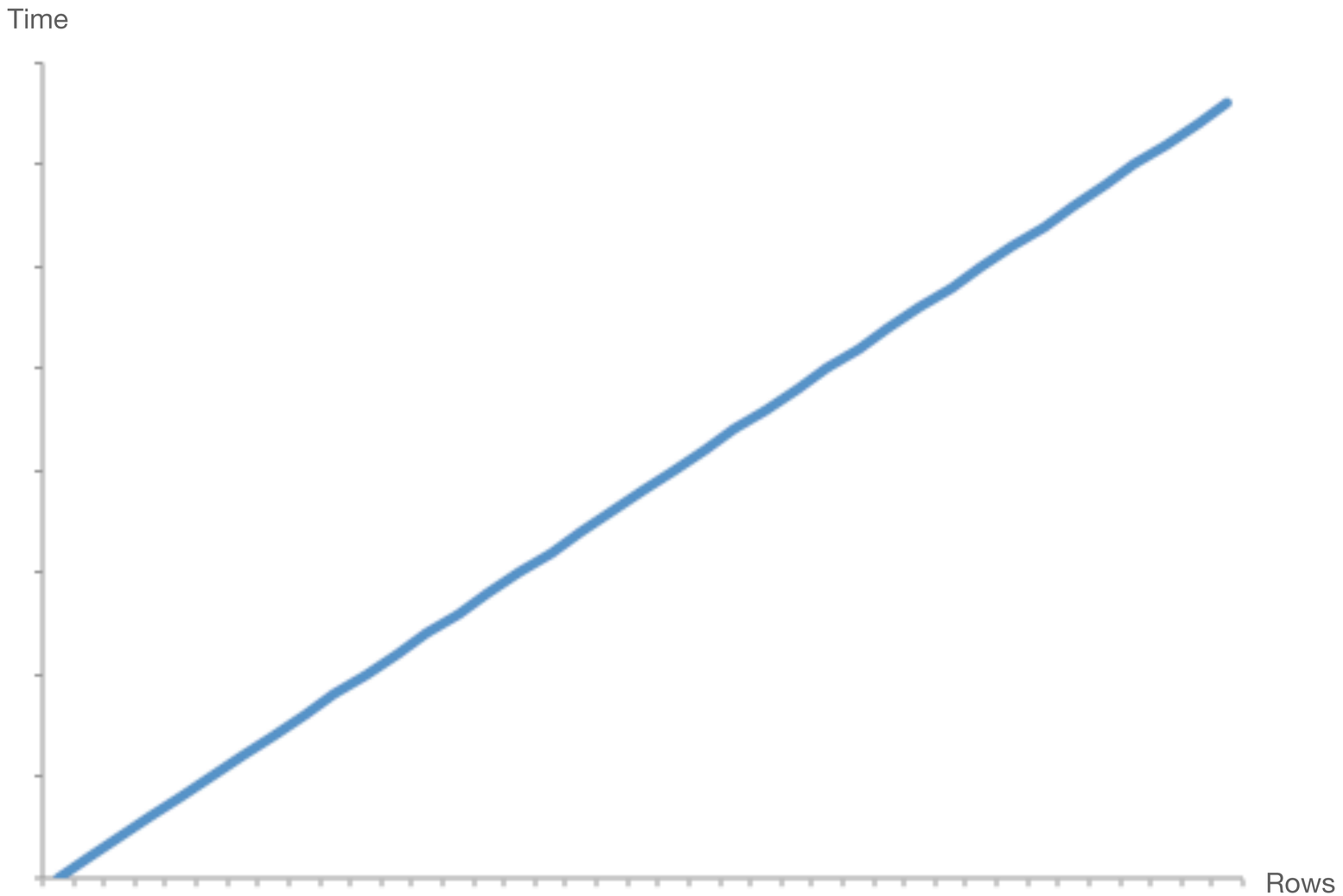
SELECT *
FROM tree
WHERE number > 8

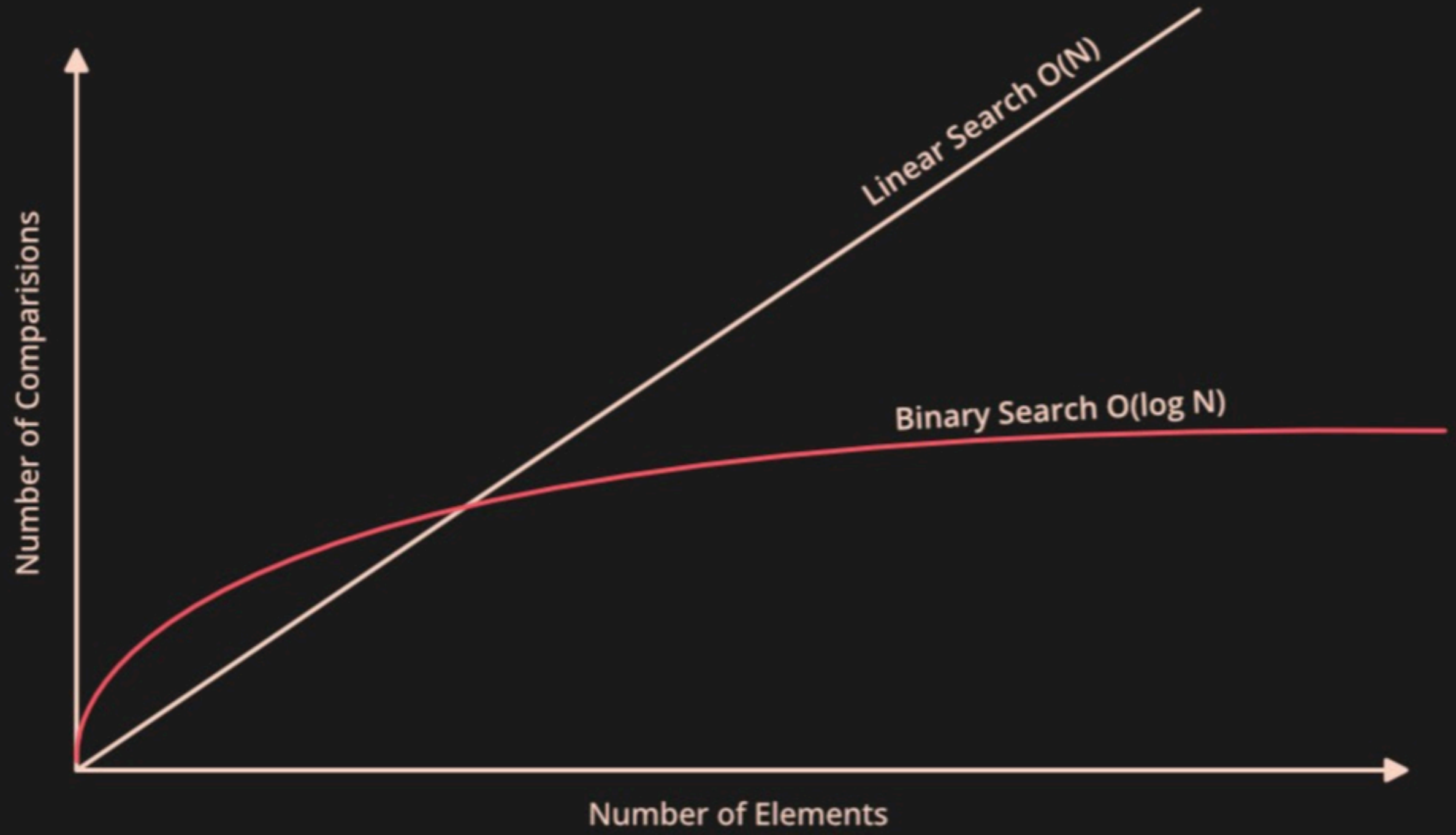


Binary Search Tree

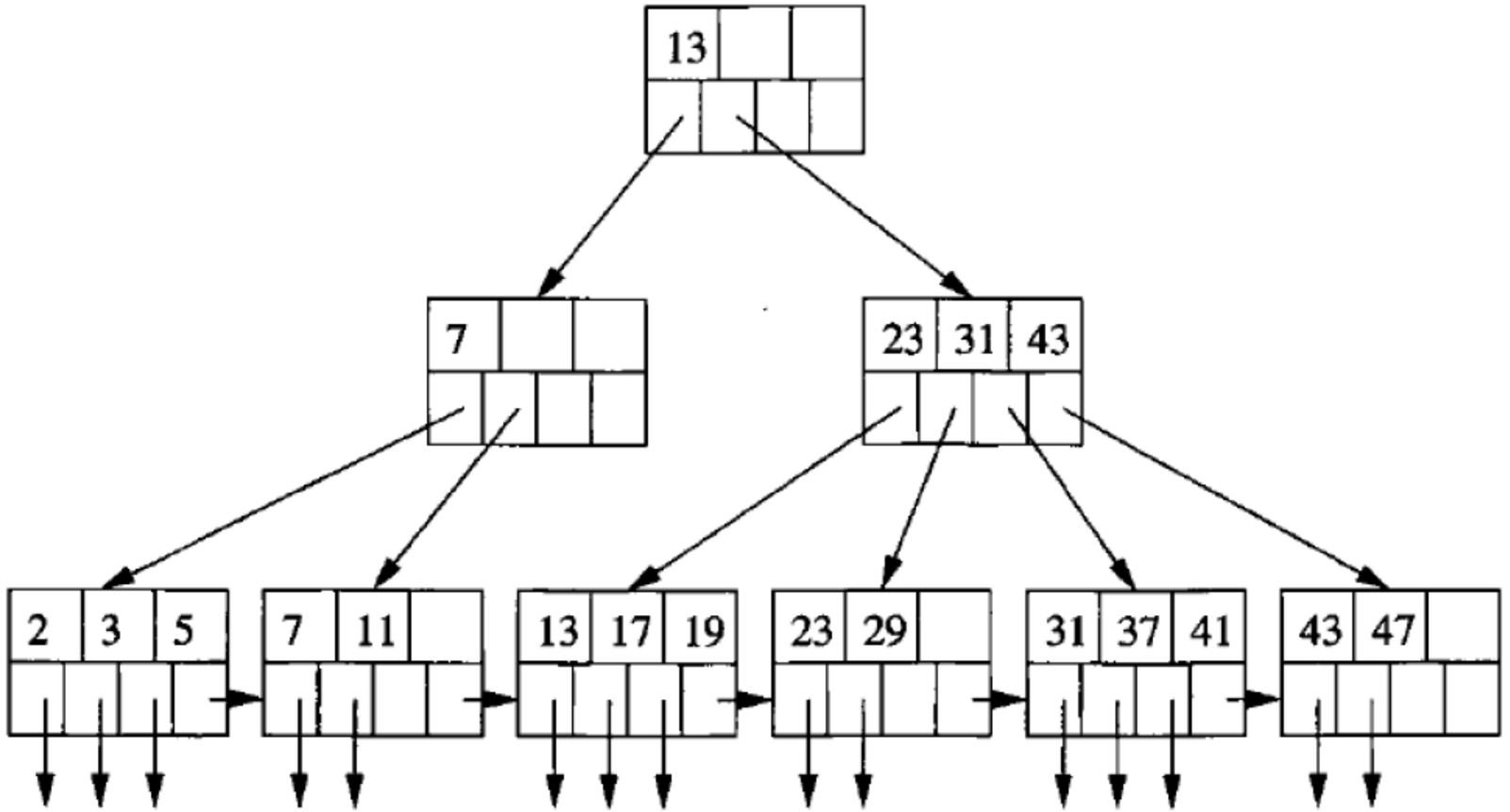
```
SELECT *
FROM spotify
WHERE artist LIKE ‘c%’
```




Time
Rows

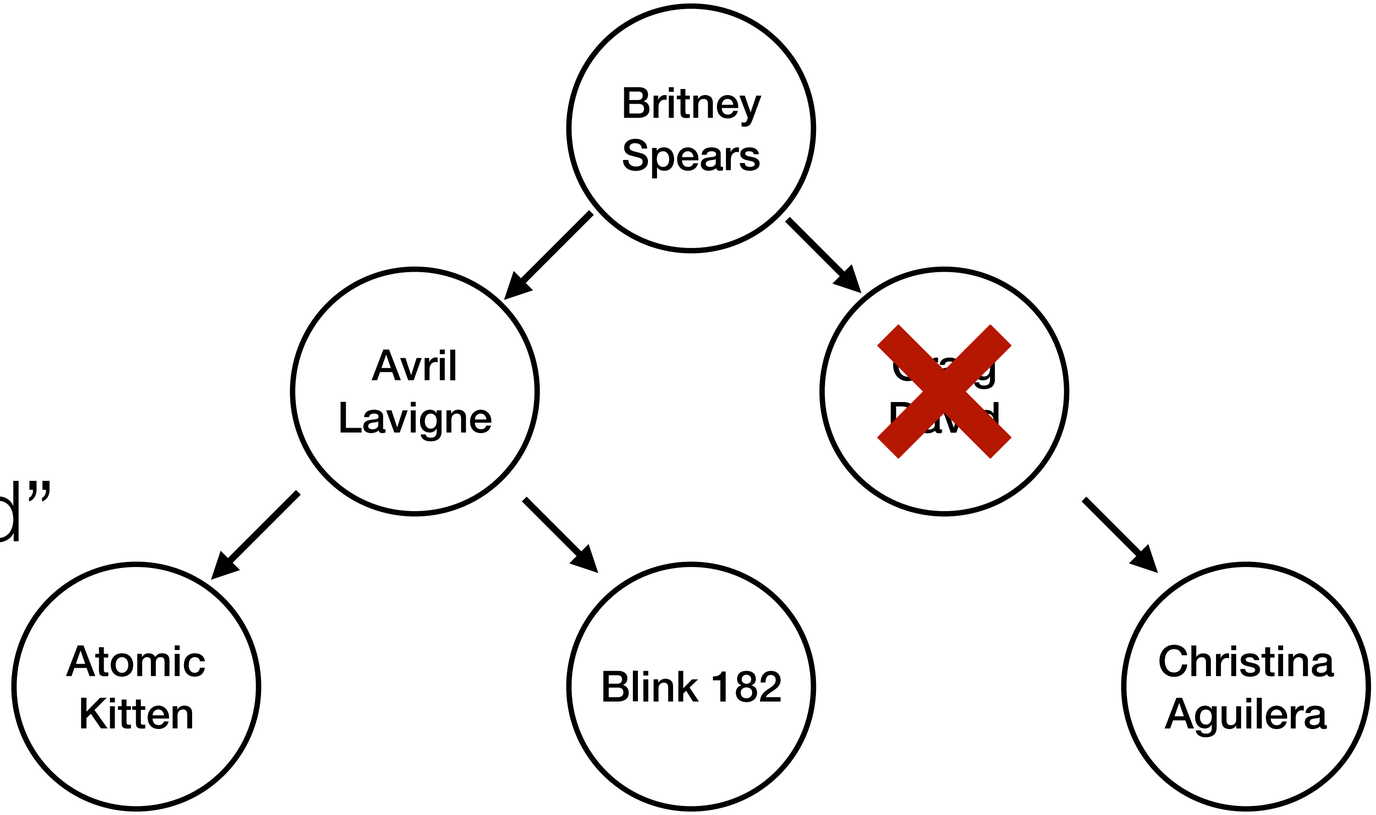

Linear Search O(N)
Binary Search O(log N)
Number of Comparisions
Number of Elements

The “ugly” reality B+ Trees



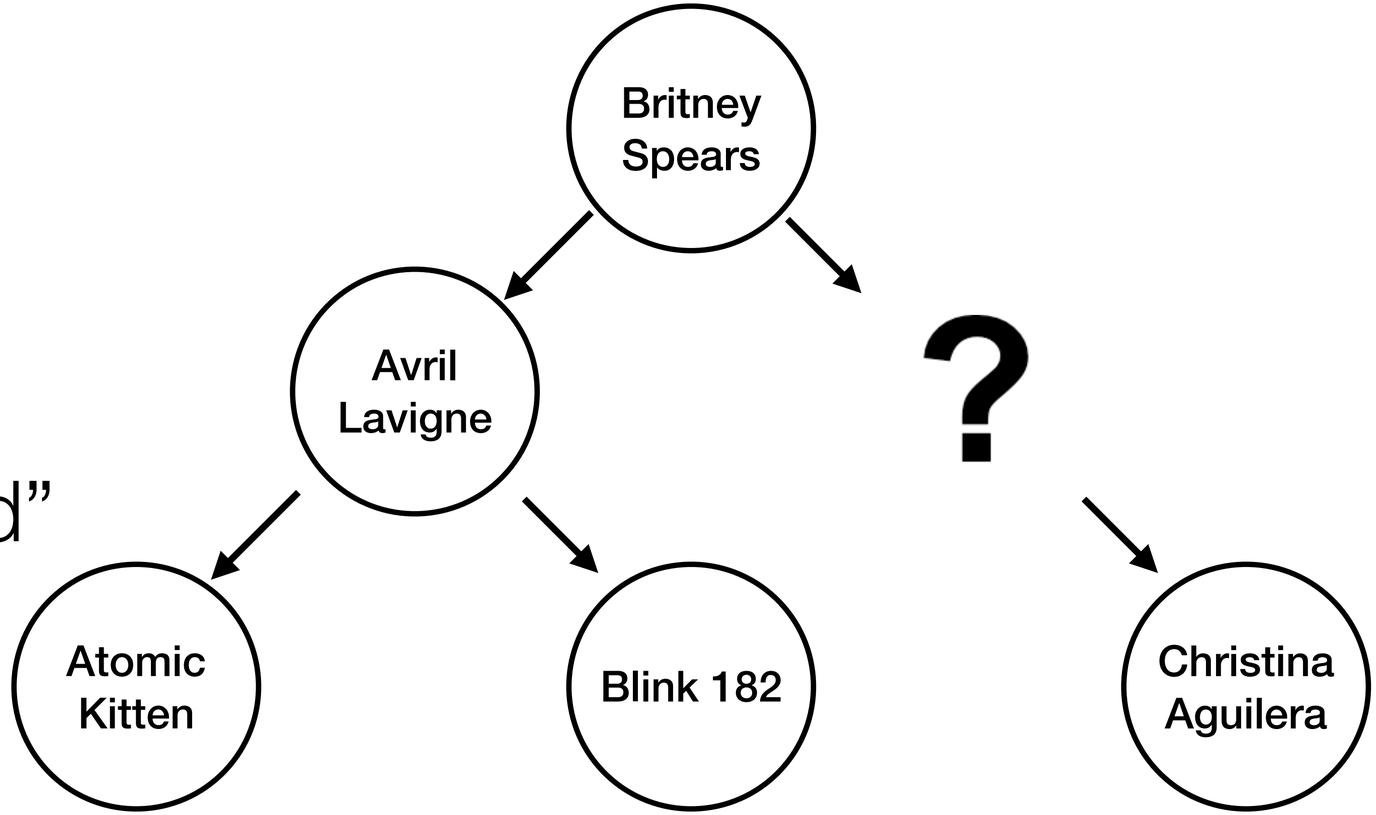
Binary Search Tree - Downside

DELETE
FROM spotify
WHERE artist = “Craig David”



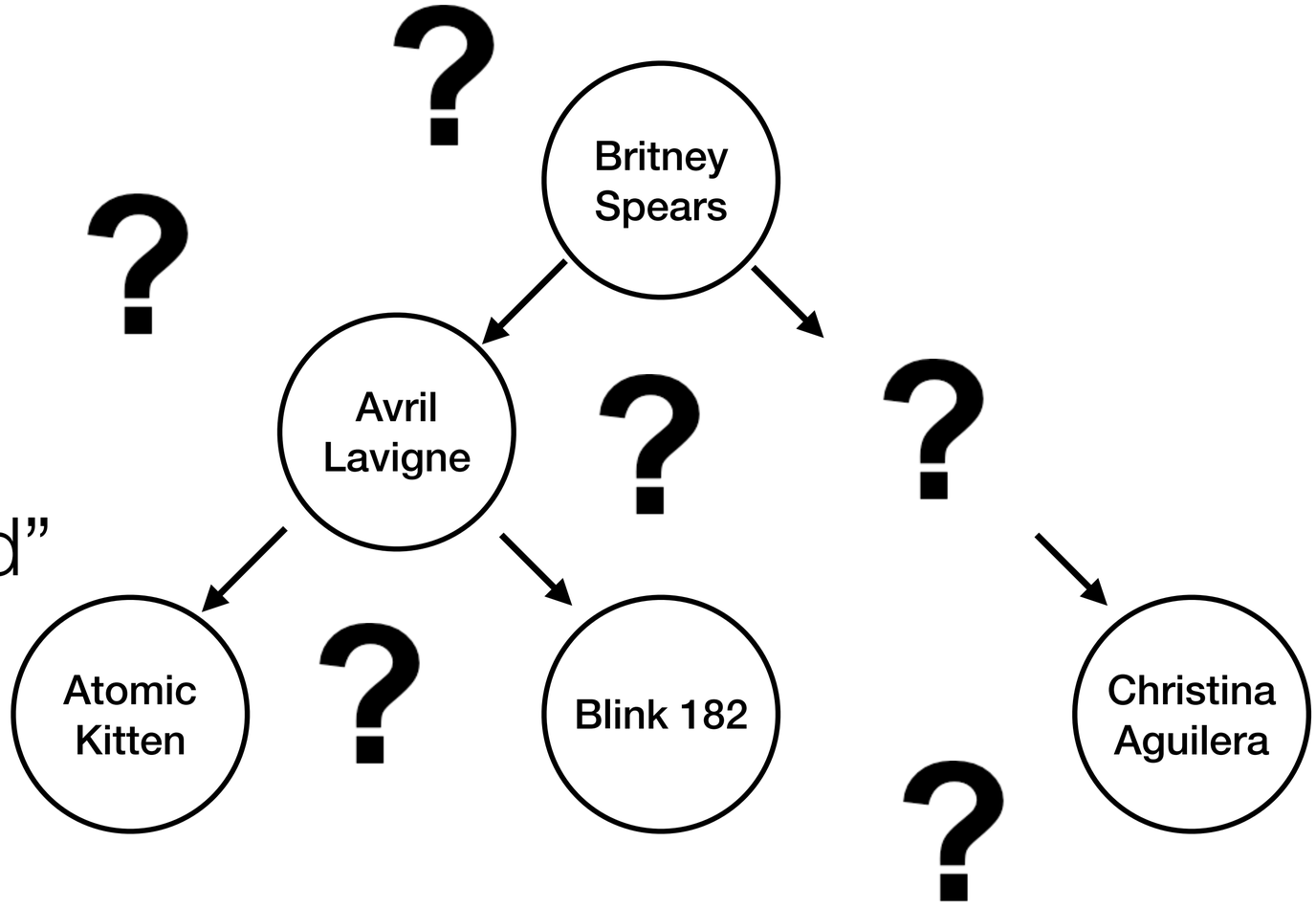
Binary Search Tree - Downside

```
DELETE
FROM spotify
WHERE artist = “Craig David”
```



Binary Search Tree - Downside

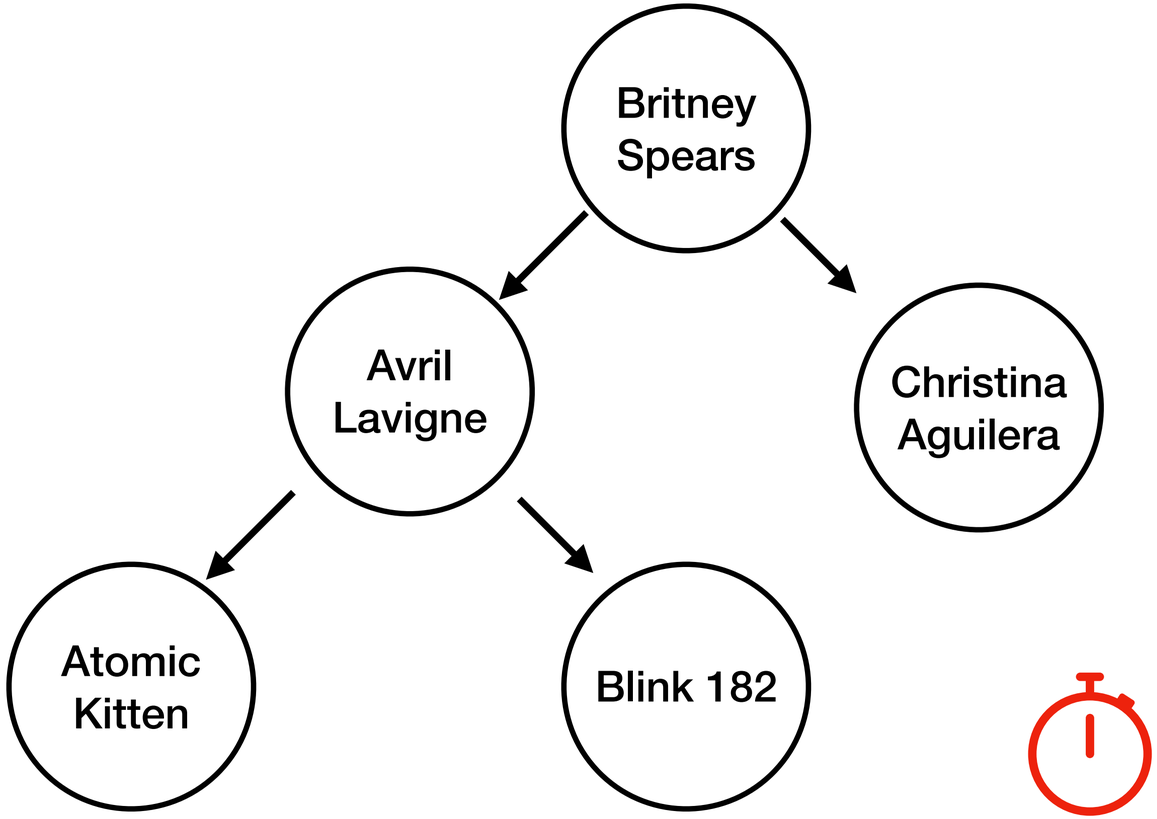
DELETE
FROM spotify
WHERE artist = “Craig David”



Binary Search Tree - Downside

Selecting is faster

Removing/adding/updating data is slower

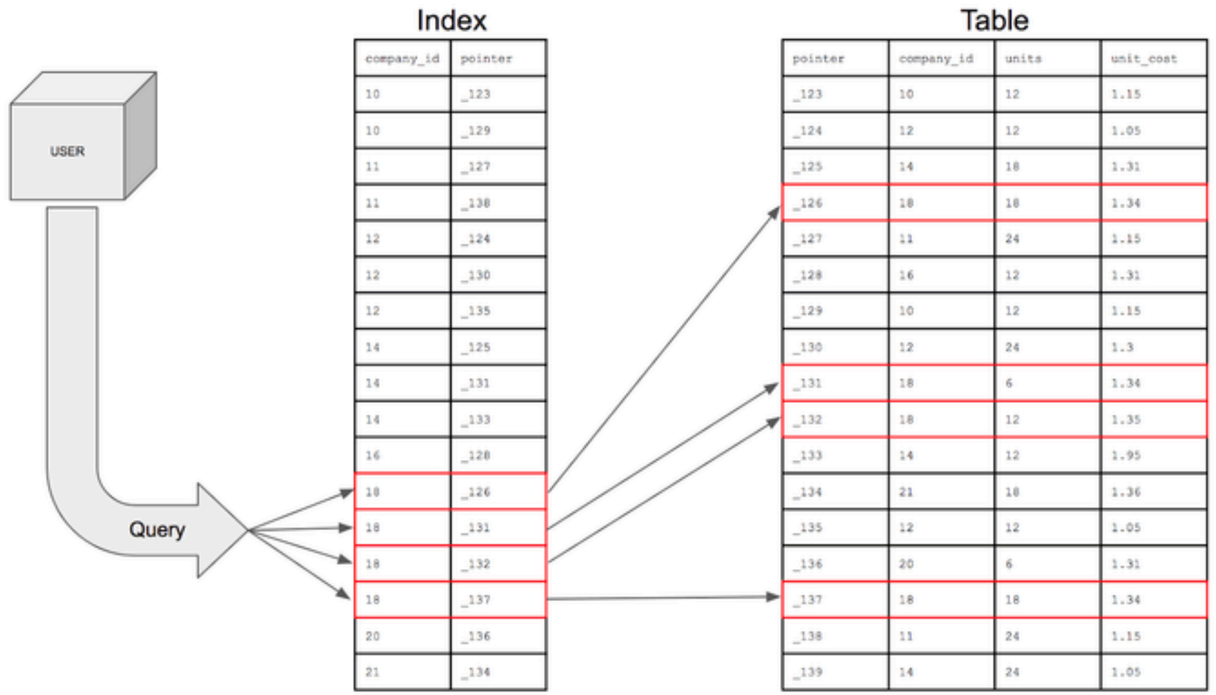


Indexing

```
CREATE INDEX artist_idx
ON spotify
```

artist	song	duration_ms	explicit	year
Britney Spears	Oops!...I Did It Again	211160	0	2000
blink-182	All The Small Things	167066	0	1999
Faith Hill	Breathe	250546	0	1999
Bon Jovi	It's My Life	224493	0	2000
*NSYNC	Bye Bye Bye	200560	0	2000
Sisqo	Thong Song	253733	1	1999
Eminem	The Real Slim Shady	284200	1	2000
Robbie Williams	Rock DJ	258560	0	2000
Destiny's Child	Say My Name	271333	0	1999
Modjo	Lady - Hear Me Tonight	307153	0	2001
Gigi D'Agostino	L'Amour Toujours	238759	0	2011
Eiffel 65	Move Your Body - Ga…	268863	0	1999
Bomfunk MC's	Freestyler	306333	0	2000
Sting	Desert Rose	285960	0	1999
Melanie C	Never Be The Same…	294200	0	1999
Aaliyah	Try Again	284000	0	2002

Binary Search Tree - Implementation detail



Demo

Full table scan or index range scan?

```
SELECT *
FROM spotify
WHERE year = 2001;
```

Indexing

Database design - advanced topics

- Indexes organises column entities to be stored as B+ trees
  - Allows for faster retrieval
  - Slower updates, inserts & deletes
- Primary Keys are by default indexed
- Database tuning
- Relevant for columns that receives a lot of retrieval and search
  - Can be detected by usage

Indexing exercises

Agenda

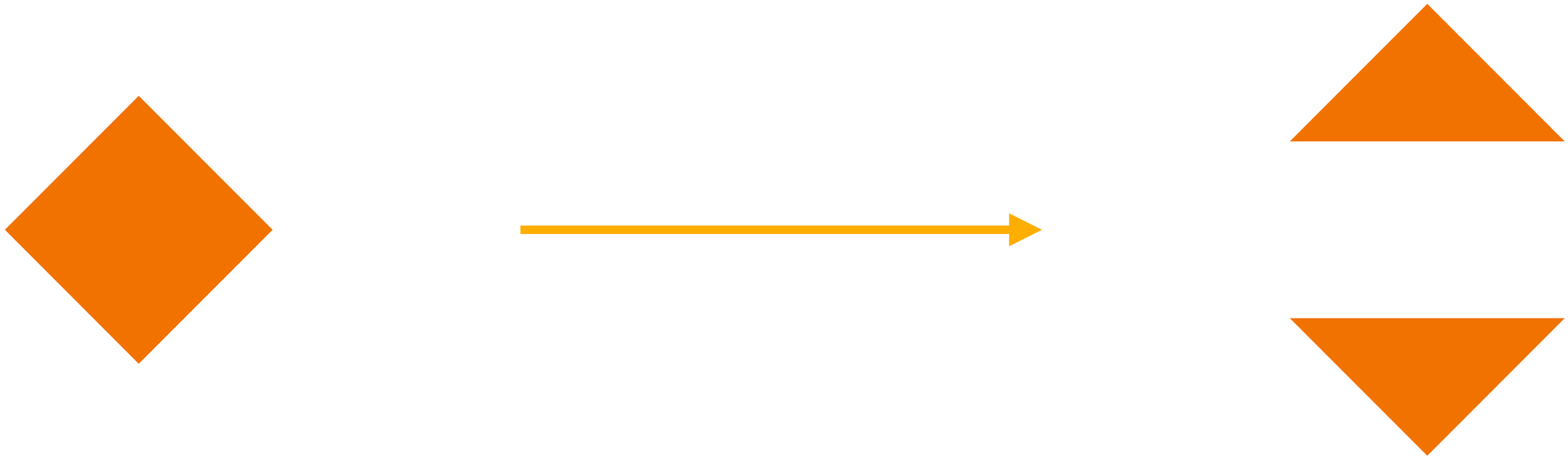
Learning objectives

- ~~What happens when a relational database is queried?~~
  - ~~Query stats & execution plan~~
- ~~Indexing~~
  - ~~Binary search trees -- conceptually~~
- Normalisation
- Views

Normalisation = Decomposition

Decomposition != Intuitive

Decomposition



Why normalise ?

Database design

- Prevent anomalies - ensure integrity
  - Insertions anomalies
  - Deletion anomalies
  - Updates anomalies
- Reduce data redundancy

3 Normal forms

Definition - Formal

"First normal form (1NF) is a property of a relation in a relational database. A relation is in first normal form if and only if no attribute domain has relations as elements.”

Codd, E.F (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM. Classics. 13 (6): 377–87. p. 380-381

Informally

1 Normal Form

- No tables can have tables as values
- Only a single element within a value
- No repeating columns (artist_1, artist_2, artist_3)

Spotify - Why does this violate 1NF?

Artist	Nationality	Album 1	Album 2	…
Eminem	U.S.A	Slim Shady LP	Marshall Mathers LP	…
Major Lazer	U.S.A, United Kingdom	Guns don’t kill people… lasers do	Free the universe	…
Queen	United Kingdom	Queen	Queen II	…

Spotify - Decomposition

Artist	…
Eminem	…
Major Lazer	…
Queen	…

Country
U.S.A
United Kingdom

Artist	Album
Eminem	Slim Shady LP
Eminem	Marshall Mathers LP
Major Lazer	Guns don’t kill people… lasers do
Major Lazer	Free the universe
Queen	Queen
Queen	Queen II

Spotify - Decomposition

Primary Key	Artist	…
1	Eminem	…
2	Major Lazer	…
3	Queen	…

Many-to-many junction table

Artist	Nationality
1	U.S.A
3	United Kingdom
2	U.S.A
2	United Kingdom

Artist	Album
1	Slim Shady LP
1	Marshall Mathers LP
2	Guns don’t kill people… lasers do
2	Free the universe
3	Queen
3	Queen II

From

Artist	Nationality	Album 1	Album 2	…
Eminem	U.S.A	Slim Shady LP	Marshall Mathers LP	…
Major Later	U.S.A, United Kingdom	Guns don't kill people… lasers do	Free the universe	…
Queen	United Kingdom	Queen	Queen II	…

To

Primary Key	Artist	…
1	Eminem	…
2	Major Lazer	…
3	Queen	…

Artist	Nationality
1	U.S.A
3	United Kingdom
2	U.S.A
2	United Kingdom

Artist	Album
1	Slim Shady LP
1	Marshall Mathers LP
2	Guns don't kill people… lasers do
2	Free the universe
3	Queen
3	Queen II

Definition - Formal

[2NF] It does not have any non-prime attribute that is functionally dependent on any proper subset of any candidate key of the relation. A non-prime attribute of a relation is an attribute that is not a part of any candidate key of the relation.

Codd, E.F (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM. Classics. 13 (6): 377–87. p. 380-381

Informally

2 Normal Form

- Is in 1st normal form
- A relation must not contain any partial dependencies
- Every non-key attribute is functionally dependant on the primary key

Or

- Every non-key attribute tells something about the entity of the row

- This form is relevant for tables with composite primary keys

Functional dependant on the full primary key

2 Normal Form

Functional dependant on the full row entity

2 Normal Form

Britney Spears	Oops!...I Did It Again	211160	0	2000
blink-182	All The Small Things	167066	0	1999
Faith Hill	Breathe	250546	0	1999
Bon Jovi	It's My Life	224493	0	2000
*NSYNC	Bye Bye Bye	200560	0	2000
Sisqo	Thong Song	253733	1	1999
Eminem	The Real Slim Shady	284200	1	2000
Robbie Williams	Rock DJ	258560	0	2000

A song is a row entity

Primary key represents a song

A pokemon is a row entity

Primary key represents a pokemon

1	Bulbasaur	45	65	65	49	49	45	Grass
2	Ivysaur	60	80	80	63	62	60	Grass
3	Venusaur	80	100	100	83	82	80	Grass
4	Charmander	65	50	60	43	52	39	Fire
5	Charmeleon	80	65	80	58	64	58	Fire
6	Charizard	100	85	109	78	84	78	Fire
7	Squirtle	43	64	50	65	48	44	Water
8	Wartortle	58	80	65	80	63	59	Water

Britney Spears	Oops!...I Did It Again	211160	0	2000
blink-182	All The Small Things	167066	0	1999
Faith Hill	Breathe	250546	0	1999
Bon Jovi	It's My Life	224493	0	2000
*NSYNC	Bye Bye Bye	200560	0	2000
Sisqo	Thong Song	253733	1	1999
Eminem	The Real Slim Shady	284200	1	2000
Robbie Williams	Rock DJ	258560	0	2000

All non-key attributes provides data about a unique song

All non-key attributes provides data about a unique pokemon

1	Bulbasaur	45	65	65	49	49	45	Grass
2	Ivysaur	60	80	80	63	62	60	Grass
3	Venusaur	80	100	100	83	82	80	Grass
4	Charmander	65	50	60	43	52	39	Fire
5	Charmeleon	80	65	80	58	64	58	Fire
6	Charizard	100	85	109	78	84	78	Fire
7	Squirtle	43	64	50	65	48	44	Water
8	Wartortle	58	80	65	80	63	59	Water

Functional dependant on the full primary key

2 Normal Form

Definition - Formal

[3NF] Every non-prime attribute of R is non-transitively dependent on every key of R.

Codd, E.F (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM. Classics. 13 (6): 377–87. p. 380-381

Transitivity

$$\text{If } A > B \text{ and } B > C \Rightarrow A > C$$

Transitivity

$$\text{If } A > B \text{ and } B > C \Rightarrow A > C$$

$$A = 9$$
$$B = 5$$
$$C = ?$$

Bookstore - What is the odd one out?

Books

Id	Book	Series	Publication Year	Author	Author_nationality
1	Harry Potter & The Philosophers Stone	Harry Potter	1997	J.K Rowling	United Kingdom
2	Harry Potter & Chamber of secrets	Harry Potter	1998	J.K Rowling	United Kingdom
3	Harry Potter & The prisoner of Azkaban	Harry Potter	1999	J.K Rowling	United Kingdom
4	War and Peace	NULL	1869	Leo Tolstoy	Russian

Bookstore - Author nationality

Books

Id	Book	Series	Publication Year	Author	Author_nationality
1	Harry Potter & The Philosophers Stone	Harry Potter	1997	J.K Rowling	United Kingdom
2	Harry Potter & Chamber of secrets	Harry Potter	1998	J.K Rowling	United Kingdom
3	Harry Potter & The prisoner of Azkaban	Harry Potter	1999	J.K Rowling	United Kingdom
4	War and Peace	NULL	1869	Leo Tolstoy	Russian

Author nationality is only transitively dependant on book id

Books

Id	Book	Series	Publication Year	Author	Author_nationality
1	Harry Potter & The Philosophers Stone	Harry Potter	1997	J.K Rowling	United Kingdom
2	Harry Potter & Chamber of secrets	Harry Potter	1998	J.K Rowling	United Kingdom
3	Harry Potter & The prisoner of Azkaban	Harry Potter	1999	J.K Rowling	United Kingdom
4	War and Peace	NULL	1869	Leo Tolstoy	Russian

Bookstore - Non composite key

Books

Id	Book	Series	Publication Year	Author	Author_nationality
1	Harry Potter & The Philosophers Stone	Harry Potter	1997	J.K Rowling	United Kingdom
2	Harry Potter & Chamber of secrets	Harry Potter	1998	J.K Rowling	United Kingdom
3	Harry Potter & The prisoner of Azkaban	Harry Potter	1999	J.K Rowling	United Kingdom
4	War and Peace	NULL	1869	Leo Tolstoy	Russian

Bookstore - Non composite key

Books

Id	Book	Series	Publication Year	Author	Author_nationality
1	Harry Potter & The Philosophers Stone	Harry Potter	1997	J.K Rowling	United Kingdom
2	Harry Potter & Chamber of secrets	Harry Potter	1998	J.K Rowling	United Kingdom
3	Harry Potter & The prisoner of Azkaban	Harry Potter	1999	J.K Rowling	United Kingdom
4	War and Peace	NULL	1869	Leo Tolstoy	Russian

Bookstore - 3NF

Books

Id	Book	Series	Publication Year	Author_id
1	Harry Potter & The	Harry Potter	1997	1
2	Harry Potter & Chamber of	Harry Potter	1998	1
3	Harry Potter & The prisoner of	Harry Potter	1999	1
4	War and Peace	NULL	1869	2

Authors

Author_id	Author_name	Author_nationality
1	J.K Rowling	United Kingdom
2	Leo Tolstoy	Russian

Eliminating data redundancy

Id	Book	Series	Publication Year	Author	Author_nationality
1	Harry Potter & The Philosophers Stone	Harry Potter	1997	J.K Rowling	United Kingdom
2	Harry Potter & Chamber of secrets	Harry Potter	1998	J.K Rowling	United Kingdom
3	Harry Potter & The prisoner of Azkaban	Harry Potter	1999	J.K Rowling	United Kingdom
4	War and Peace	NULL	1869	Leo Tolstoy	Russian

Id	Book	Series	Publication Year	Author_id
1	Harry Potter & The	Harry Potter	1997	1
2	Harry Potter & Chamber of	Harry Potter	1998	1
3	Harry Potter & The prisoner of	Harry Potter	1999	1
4	War and Peace	NULL	1869	2

Author_id	Author_name	Author_nationality
1	J.K Rowling	United Kingdom
2	Leo Tolstoy	Russian

Minimal scheme notation:

Student(cpr, name, email, active)

Normalisation exercises

Views

Advanced Database topic

- Virtual table
  - “Stored query”
- Only relevant data
- Reusable
- Security mechanism -> Provides access to certain users
- Hide complexity

Creating a view