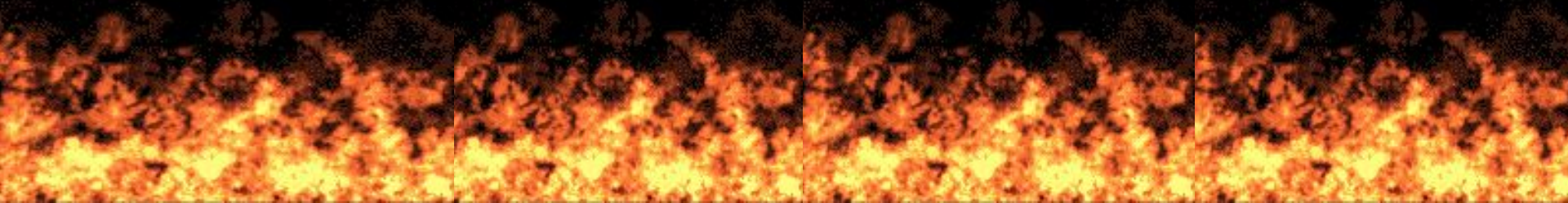
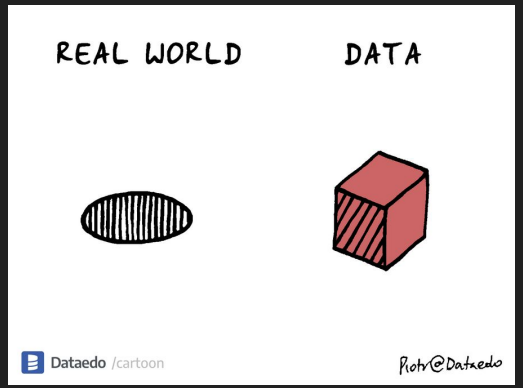
Controversy in DB Keys




REAL WORLD
DATA
Dataedo /cartoon
Piotr@Dataedo


WHEN YOU START WORKING ON REAL-WORLD DATA
DATA


No one:
Real World:
I am going to create
data that is so problematic

Real World Data* – More Interesting + More Complicated

name text	agency text	agencysubelement text	occupation text	payplan text	grade text
MASON JACK	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	7404-COOKING	WS	7
MAYFIELD JASON	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	7404-COOKING	WS	7
PAULA JENN	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	5432-EDUCATION	WS	8
TOBIN ELIAS	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	3542-PSYCHOLOGY	WS	9
ALLAN JANNETTE	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	4234-CORRECTIONAL	WS	8

dutysta bigint	state text	county text	city text	basicpay numeric	adjustbasicpay numeric	totalcompensation text	filedate bigint
184840167	INDIANA	VIGO	TERRE HAUTE	85587.87	85587.87	######	201201
184840167	INDIANA	VIGO	TERRE HAUTE	60925.06	60925.06	######	201201
189098367	INDIANA	VIGO	TERRE HAUTE	75925.06	75925.06	######	201201
167876000	ARIZONA	TUCSON	TUCSON	90725.00	90725.00	######	201201
543765167	TEXAS	JEFFERSON	BEAUMONT	90725.00	90725.00	######	201201

What should our primary key be?

* Structure is identical to real-world data. Data itself is made up

Solution = Surrogate Keys!

+automatically generated +unique +immutable +non-descriptive.

- Integer with Autoincrement
- UUID (128-bit value)
- Sequences

```sql
-- Add a new UUID column to your table
ALTER TABLE employees ADD COLUMN uuid_surrogate_keys UUID;
UPDATE employees SET uuid_surrogate_keys = uuid_generate_v4();

-- Make it the primary key
ALTER TABLE employees ADD PRIMARY KEY (uuid_surrogate_keys);
```

name text	agency text	agencysubelement text	occupation text	payplan text	grade text
MASON JACK	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	7404-COOKING	WS	7
MAYFIELD JASON	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	7404-COOKING	WS	7
PAULA JENN	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	5432-EDUCATION	WS	8
TOBIN ELIAS	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	3542-PSYCHOLOGY	WS	9
ALLAN JANNETTE	DJ-DEPARTMENT OF JUSTICE	DJ03-BUREAU OF PRISONS/FEDERAL PRISON SYSTEM	4234-CORRECTIONAL	WS	8

dutysta bigint	state text	county text	city text	basicpay numeric	adjustbasicpay numeric	totalcompensation text	filedate bigint	uuid_surrogate_keys [PK] uuid
184840167	INDIANA	VIGO	TERRE HAUTE	85587.87	85587.87	######	201201	8a2b24ec-33fd-4ba1-a4f8-6d91bc9701…
184840167	INDIANA	VIGO	TERRE HAUTE	60925.06	60925.06	######	201201	95395105-fb9f-4f18-9d20-bfac0da5110a
189098367	INDIANA	VIGO	TERRE HAUTE	75925.06	75925.06	######	201201	deb85886-e89f-4a35-9638-6cf4af793f4f
167876000	ARIZONA	TUCSON	TUCSON	90725.00	90725.00	######	201201	d233a6ea-749b-41ab-9975-d8664c0d3…
543765167	TEXAS	JEFFERSON	BEAUMONT	90725.00	90725.00	######	201201	6a4fdafb-1768-42e8-8f9f-300c1d416ae7

UKC

Ultimate Key Champion

Surrogate Keys

VS

Natural Keys

ROUND 1: Security

ROUND 2: Data Migration

ROUND 3: Performance

ROUND 4: Business Logic

Arguing in Favor of Natural Keys:

Benefits of Natural Keys

Business Logic Alignment: Improve the consistency between the database and the real-world domain.

Data Migrations and Integration: When integrating data from external sources or systems, natural keys can make the process smoother because they often match key attributes in those sources.

Disk Space: Primary key index already exist so no disk extra space is required for the extra column/index that would be used by a surrogate key column.

Data Lookup: Fewer table joins since join columns have meaning. For example, this can reduce disk IO by not having to perform extra reads on a lookup table.

Arguing in Favor of Surrogate Keys: Prince

Disadvantages of Natural Keys

- Adaptability Issues: becomes unstable when requirements change
- Poorer performance: since key value is usually larger. Requires more IO both when inserting/updating data as well as when you query.
- Limits the database –Cannot enter records when key value is unknown
- Uniqueness: Does not guarantee uniqueness.
- Simplicity: Difficult to be pick good keys

Arguing in Favor of Surrogate Keys: Prince

Benefits of Surrogate Keys

- Uniqueness: Surrogate keys guarantee uniqueness.
- Security: Surrogate keys can be more secure - UUIDs.
- Stability: Even if the natural key changes, the surrogate key remains constant.
- Performance: Less indexing and integer-based keys are more efficient for indexing.
- Simplicity: No need to manage complex natural keys.

Arguing in Favor of Natural Keys:

Disadvantages of Surrogate Keys

- Data Migrations and Integration: Surrogate keys are not directly tied to the data, so changes in the data, such as name updates, might necessitate changes in the key, potentially affecting data integrity.Surrogate keys might require complex mapping during data integration.
- Surrogate keys might make the data model less intuitive for non-technical users and might require additional training.
- Adaptability Issues: becomes unstable and overly complex when requirements change
- Requires more table joins to child tables since data has no meaning on its own.
- Disk space: surrogate key will require extra disk space.
- Joins: more table joins to child tables since data has no meaning on its own
- Key value has no relation to data so technically design breaks 3NF (i.e. normalization)

VOTE

Conclusion