



# Gonymizer

A Tool to Anonymize Sensitive PostgreSQL Data Tables  
for Use in QA and Testing

Levi Junkert

## Agenda

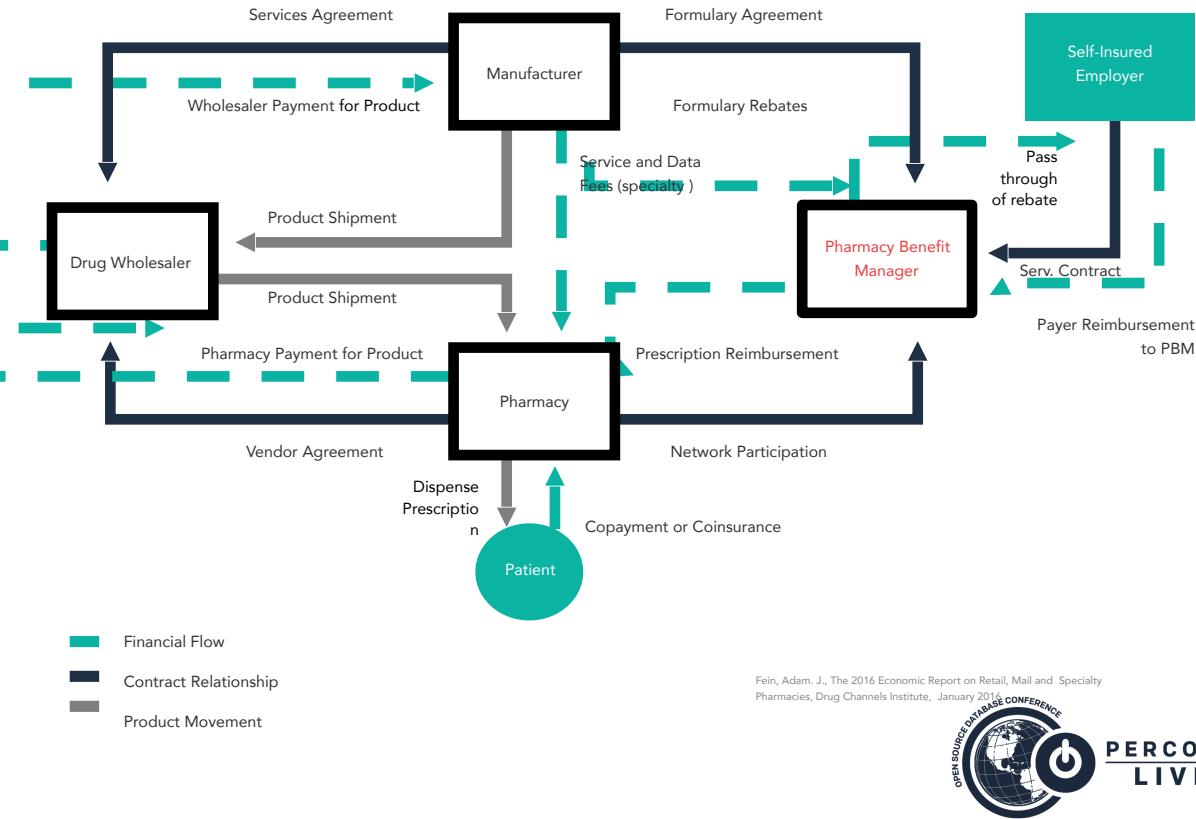
- Introductions
- HIPAA Regulations
- SmithRx Solution: Gonymizer
- Gonymizing multiple environments using Kubernetes
- Demo
- Wrap-up

## Who am I?

- Database systems 18 years
  - MySQL
  - PostgreSQL
- Software Engineering 11 years
  - Product, Infrastructure, and Automation
- DB and SE Experience
  - Facebook
  - RightNow ( Oracle)
  - OpenDNS (Cisco)
  - Get Better



PBMs sit in the middle of a large and complex ecosystem





Smi+  
hRx

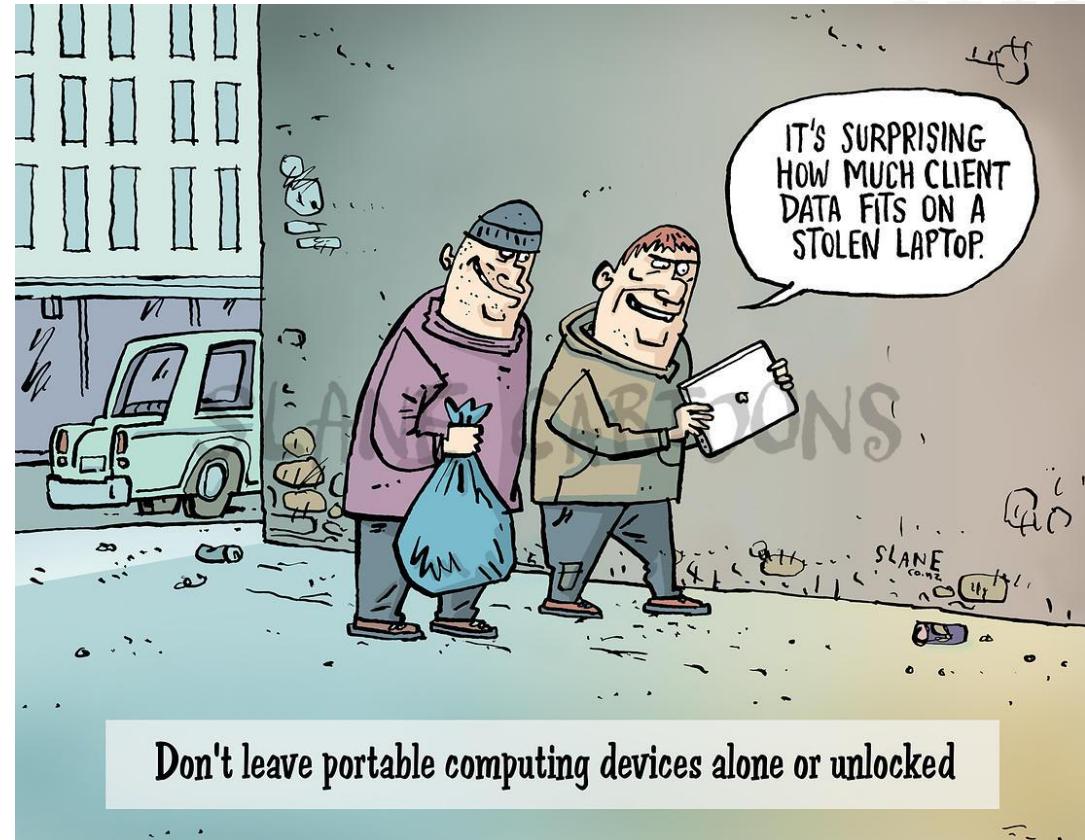


# HIPAA Regulations

Health Insurance Portability and Accountability Act of 1996  
United States Legislation

# HIPAA Regulations

- Examples
- Importance
- Solutions



© All Rights Reserved - Chris Slane

# HIPAA Regulations

- Examples
- Importance
- Solutions

Smi+  
hRx

## Police release video of Alameda Starbucks laptop thieves

By Alyssa Pereira, SFGATE Published 11:09 am PDT, Wednesday, October 4, 2017

The screenshot shows a news article from the "SECURITY" section of Government Technology magazine. The headline reads "Lost Laptop Triggers Security Concerns for Seattle University". A subtext below the headline states: "An unencrypted laptop with access to sensitive information about some 2,000 employees and dependents was left on a city bus in late March. The university has notified at-risk individuals." The article is by Asia Fields, published on May 1, 2019.

The screenshot shows a news article from Modern Healthcare. The headline reads "Medical imaging company to pay \$3 million HIPAA fine". The author is Jessica Kim Cohen. The article is dated May 06, 2019, at 12:21 PM.

**Court rules man can sue Costco for disclosing his erectile-dysfunction prescription**

Kate Bernot  
5/02/19 11:32am • Filed to: COSTCO

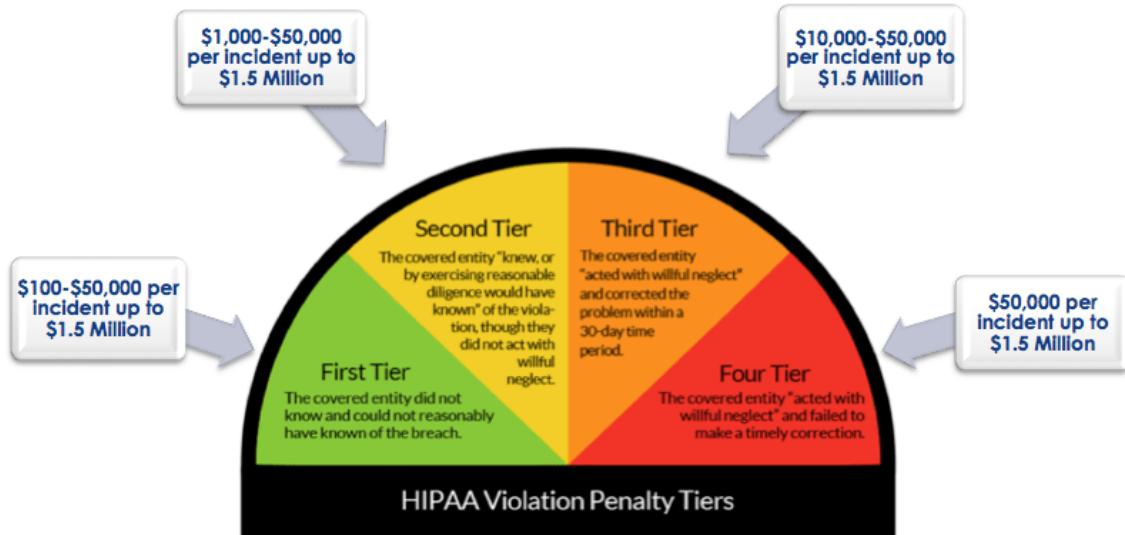
42 1 f t e



# HIPAA Regulations

- Examples
- Importance
- Solutions

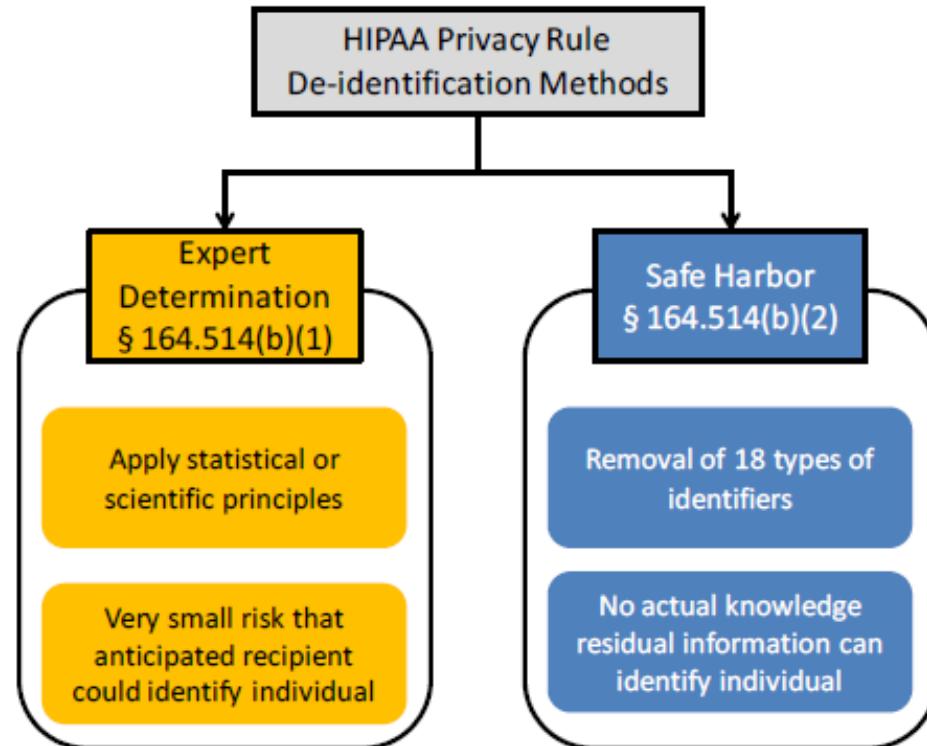
## Penalties



855.85HIPAA  
[www.compliancygroup.com](http://www.compliancygroup.com)

# HIPAA Regulations

- Examples
- Importance
- Solutions



# HIPAA Regulations

- Examples
- Importance
- Solutions

Smi+  
hRx

## Safe Harbor Method (164 somethin b2?)

- Name
- Street
- All elements of dates but year
- Telephone Number
- Fax Number
- Email Address
- Driver Licenses and IDS
- Vehicle IDs
- Biometric IDs
- Full face photos
  - Other comparable images
- URL addresses
- Social Security Number
- Account numbers
  - CC/Debit
- Passwords

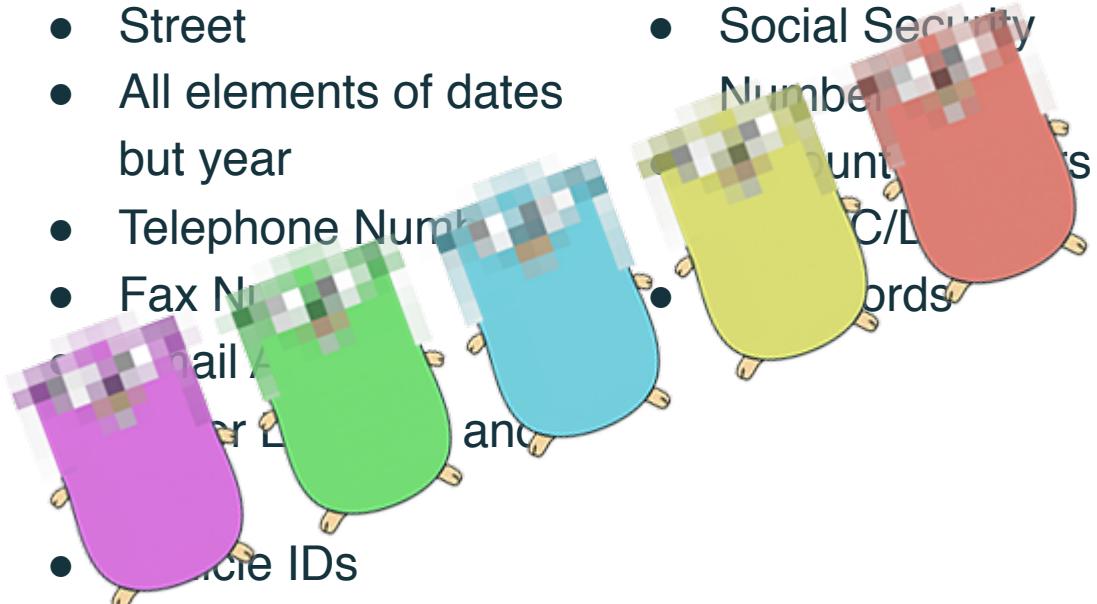
# HIPAA Regulations

- Examples
- Importance
- Solutions

Smi+hRx

## Safe Harbour Method (164 somethin b2?)

- Name
- Street
- All elements of dates but year
- Telephone Number
- Fax Number
- Email Address
- URL addresses
- Social Security Number
- Driver's License IDs
- Biometric IDs
- Full face photos
  - Other comparable images





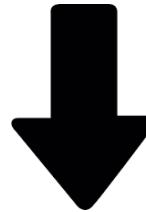
# SmithRx Solution: Gonymizer

# Gonymizer

- Requirements
- Design
- Mapping

Smi+  
hRx

Person		
Column	Value	Type
ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String



Person (Anonymized)		
Column	Value	Type
ID	2012e67c-d744-4da8-b24e-a35604439653	UUID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String



PERCONA  
LIVE

# Gonymizer

- Requirements
- Design
- Mapping

Smi+  
hRx

Person		
Column	Value	Type
ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String

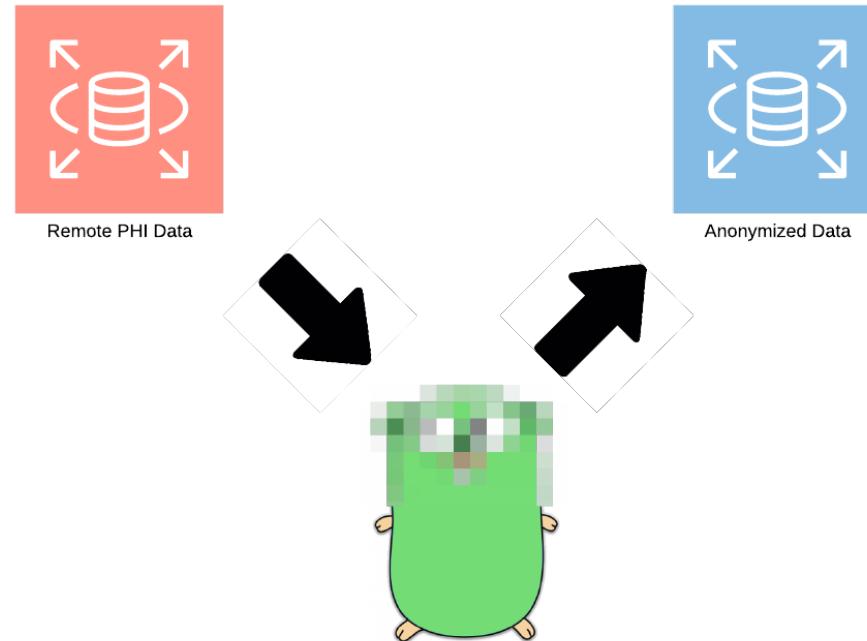
$$\downarrow \\ f(x) \\ \downarrow$$

Person (Anonymized)		
Column	Value	Type
ID	2012e67c-d744-4da8-b24e-a35604439653	UUID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String

# Gonymizer

- Requirements
- Design
- Mapping

Smi+  
hRx



# Gonymizer

- Requirements
- Design
- Mapping

- Dump of the original PHI dataset
- Load the data set into a PSQL host
- Assignable scrambler function(s)
- A map to know which columns to anonymize
- Must keep relationships intact
- Must be able to be automated

# Gonymizer

- Requirements
- Design
- Mapping

Smi+  
hRx

## Separated Commands

- **Dump**
  - Downloads SQL dump file
- **Map**
  - Build and maintain an anonymization map for
    - Schemas
    - Tables
    - Columns
- **Process**
  - Anonymizes columns according to map
- **Load**
  - Loads the anonymized file

# Gonymizer

- Requirements
- Design
- Mapping

Smi+hRx

PHI/PII Database Host



DUMP

pg\_dump



phi\_dump.sql



LOAD

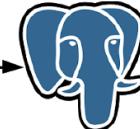


anonymized\_data.sql

Anonymized Database Host



psql -f



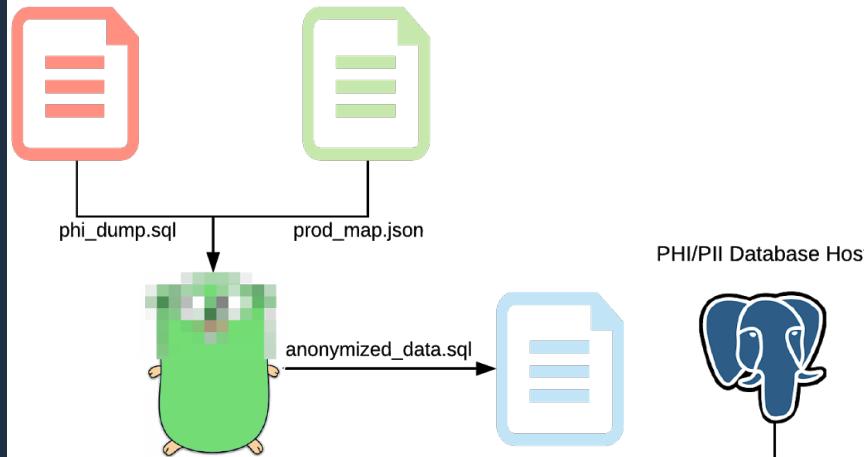
PERCONA  
LIVE

# Gonymizer

- Requirements
- Design
- Mapping

Smi+  
hRx

## PROCESS



PHI/PII Database Host

## MAP



Information Schema  
Queries



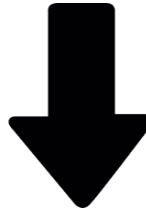
prod\_map.json

# Gonymizer



- Requirements
- Design
- Mapping

Person		
Column	Value	Type
ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String

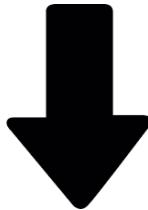


Person (Anonymized)		
Column	Value	Type
ID	2012e67c-d744-4da8-b24e-a35604439653	UUID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String

# Gonymizer

- Requirements
- Design
- Mapping

Person		
Column	Value	Type
ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String



Claim		
Column	Value	Type
ID	14ce84f8-17fa-4c5c-8eba-2dd490a7385b	UUID
Person ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID



Person (Anonymized)		
Column	Value	Type
ID	2012e67c-d744-4da8-b24e-a35604439653	UUID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String

Claim (Anonymized)		
Column	Value	Type
ID	415c134c-9d0e-4847-b307-48208af3bcc7	UUID
Person ID	2012e67c-d744-4da8-b24e-a35604439653	UUID

# Gonymizer

- Requirements
- Design
- Mapping

Person		
Column	Value	Type
ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String

Claim		
Column	Value	Type
ID	14ce84f8-17fa-4c5c-8eba-2dd490a7385b	UUID
Person ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID



UUID Hash Map	
Original Value	Anonymized Value
8ebc0420-883b-4149-9065-c5dd9ebd9d75	2012e67c-d744-4da8-b24e-a35604439653



Person (Anonymized)		
Column	Value	Type
ID	2012e67c-d744-4da8-b24e-a35604439653	UUID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String

Claim (Anonymized)		
Column	Value	Type
ID	415c134c-9d0e-4847-b307-48208af3bcc7	UUID
Person ID	2012e67c-d744-4da8-b24e-a35604439653	UUID

# Gonymizer

- Requirements
- Design
- Mapping

Person		
Column	Value	Type
ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String

Claim		
Column	Value	Type
ID	14ce84f8-17fa-4c5c-8eba-2dd490a7385b	UUID
Person ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID



UUID Hash Map	
Original Value	Anonymized Value
8ebc0420-883b-4149-9065-c5dd9ebd9d75	2012e67c-d744-4da8-b24e-a35604439653



Person (Anonymized)		
Column	Value	Type
ID	2012e67c-d744-4da8-b24e-a35604439653	UUID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String

Claim (Anonymized)		
Column	Value	Type
ID	415c134c-9d0e-4847-b307-48208af3bcc7	UUID
Person ID	2012e67c-d744-4da8-b24e-a35604439653	UUID

# Gonymizer

- Requirements
- Design
- Mapping

Person		
Column	Value	Type
ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String

Claim		
Column	Value	Type
ID	14ce84f8-17fa-4c5c-8eba-2dd490a7385b	UUID
Person ID	8ebc0420-883b-4149-9065-c5dd9ebd9d75	UUID
Doctor ID	90da8217-4ccb-4613-a054-f5e7daa3d260	UUID
Claim Details ID	3c71b5da-446d-4c6b-bb30-3f7923956a21	UUID

The diagram illustrates the mapping process. It shows two large downward-pointing arrows originating from the 'Person' and 'Claim' tables above, pointing towards a central 'UUID Hash Map' table below. This map contains four rows, each showing an original value and its corresponding anonymized value.

UUID Hash Map	
Original Value	Anonymized Value
8ebc0420-883b-4149-9065-c5dd9ebd9d75	2012e67c-d744-4da8-b24e-a35604439653
14ce84f8-17fa-4c5c-8eba-2dd490a7385b	415c134c-9d0e-4847-b307-48208af3bcc7
90da8217-4ccb-4613-a054-f5e7daa3d260	919dea1b-2d1c-41a6-8785-1c84f420344c
3c71b5da-446d-4c6b-bb30-3f7923956a21	3ec8eccb-3d8e-40fe-8706-10d02982415c

The diagram shows two large downward-pointing arrows originating from the 'Person' and 'Claim' tables above, pointing towards their respective 'Anonymized' tables below. These tables show the same structure as the originals but with different values.

Person (Anonymized)		
Column	Value	Type
ID	2012e67c-d744-4da8-b24e-a35604439653	UUID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String

Claim (Anonymized)		
Column	Value	Type
ID	415c134c-9d0e-4847-b307-48208af3bcc7	UUID
Person ID	2012e67c-d744-4da8-b24e-a35604439653	UUID
Doctor ID	919dea1b-2d1c-41a6-8785-1c84f420344c	UUID
Claim Details ID	3ec8eccb-3d8e-40fe-8706-10d02982415c	UUID

# Gonymizer

- Requirements
- Design
- Mapping

Smi+  
hRx

```
{  
  "DBName": "gonymizer_example_db",  
  "Seed": 0,  
  "SchemaPrefix": "",  
  "ColumnMaps": [  
    {  
      "Comment": "",  
      "TableSchema": "public",  
      "TableName": "person",  
      "ColumnName": "id",  
      "DataType": "uuid",  
      "ParentSchema": "",  
      "ParentTable": "",  
      "ParentColumn": "",  
      "OrdinalPosition": 2,  
      "IsNullable": false,  
      "Processors": [  
        {  
          "Name": "RandomUUID",  
          "Max": 0,  
          "Min": 0,  
          "Variance": 0,  
          "Comment": ""  
        }  
      ]  
    },  
    {  
      "Comment": "",  
      "TableSchema": "public",  
      "TableName": "claim",  
      "ColumnName": "person_id",  
      "DataType": "uuid",  
      "ParentSchema": "",  
      "ParentTable": "",  
      "ParentColumn": "",  
      "OrdinalPosition": 2,  
      "IsNullable": false,  
      "Processors": [  
        {  
          "Name": "RandomUUID",  
          "Max": 0,  
          "Min": 0,  
          "Variance": 0,  
          "Comment": ""  
        }  
      ]  
    }  
  ]  
}
```

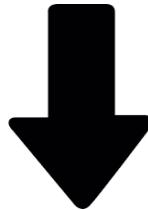
UUID Hash Map	
Original Value	Anonymized Value
8beb0420-883b-4149-9065-c5dd9eb9d75	2012e67c-d744-4da8-b24e-a35604439653
14ce84f8-17fa-4c5c-8eba-2dd490a7385b	415c134c-9d0e-4847-b307-48208af3bcc7
90da8217-4ccb-4613-a054-f5e7daa3d260	919dea1b-2d1c-41a6-8785-1c84f420344c
3c71b5da-446d-4c6b-bb30-3f7923956a21	3ec8eccb-3d8e-40fe-8706-10d02982415c

# Gonymizer



- Requirements
- Design
- Mapping

Person		
Column	Value	Type
ID	1	ID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String



Person (Anonymized)		
Column	Value	Type
ID	5234	ID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String

Smi+hRx

# Gonymizer

- Requirements
- Design
- Mapping

Person		
Column	Value	Type
ID	1	ID
Name	Levi Junkert	String
Address	1 Embarcadaro Way, San Francisco, CA 94117	String
SSN	123-12-1234	String

Claim		
Column	Value	Type
ID	54	INT
Person ID	1	INT
Doctor ID	23	INT
Claim Details ID	22	INT



Table Hash Map		
Key	Original Value	Anonymized Value
public.person.id	1	5234
public.claim.id	54	22144629



Person (Anonymized)		
Column	Value	Type
ID	5234	ID
Name	Gonzo T. Muppet	String
Address	155 E 28th St, New York, NY 10016	String
SSN	935-84-7755	String

Claim (Anonymized)		
Column	Value	Type
ID	22144629	INT
Person ID	5234	INT
Doctor ID	876945678	INT
Claim Details ID	32345235	INT

# Gonymizer

- Requirements
- Design
- Mapping

```
{  
  "DBName": "gonymizer_example_db",  
  "Seed": 0,  
  "SchemaPrefix": "",  
  "ColumnMaps": [  
    {  
      "Comment": "",  
      "TableSchema": "public",  
      "TableName": "person",  
      "ColumnName": "id",  
      "DataType": "integer",  
      "ParentSchema": "",  
      "ParentTable": "",  
      "ParentColumn": "",  
      "OrdinalPosition": 2,  
      "IsNullable": false,  
      "Processors": [  
        {  
          "Name": "AlphaNumericScrambler",  
          "Max": 0,  
          "Min": 0,  
          "Variance": 0,  
          "Comment": ""  
        }  
      ]  
    },  
    {  
      "Comment": "",  
      "TableSchema": "public",  
      "TableName": "claim",  
      "ColumnName": "person_id",  
      "DataType": "integer",  
      "ParentSchema": "public",  
      "ParentTable": "claim",  
      "ParentColumn": "id",  
      "OrdinalPosition": 2,  
      "IsNullable": false,  
      "Processors": [  
        {  
          "Name": "AlphaNumericScrambler",  
          "Max": 0,  
          "Min": 0,  
          "Variance": 0,  
          "Comment": ""  
        }  
      ]  
    }  
  ]  
}
```

Table Hash Map			
Key	Original Value	Anonymized Value	
public.person.id	1	5234	
public.claim.id	54	22144629	

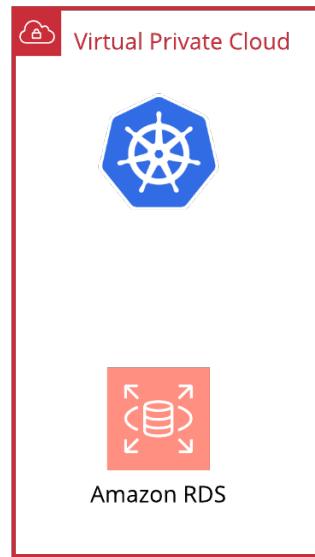


# Gonymizing Environments

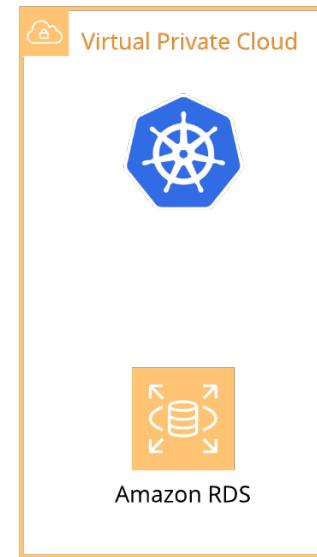
# Automation & Kubernetes

- Dump
- Process
- Load

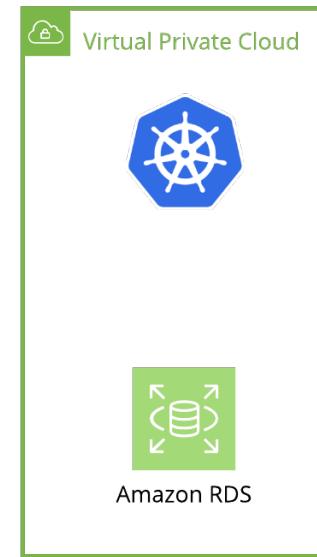
## Production



## Staging



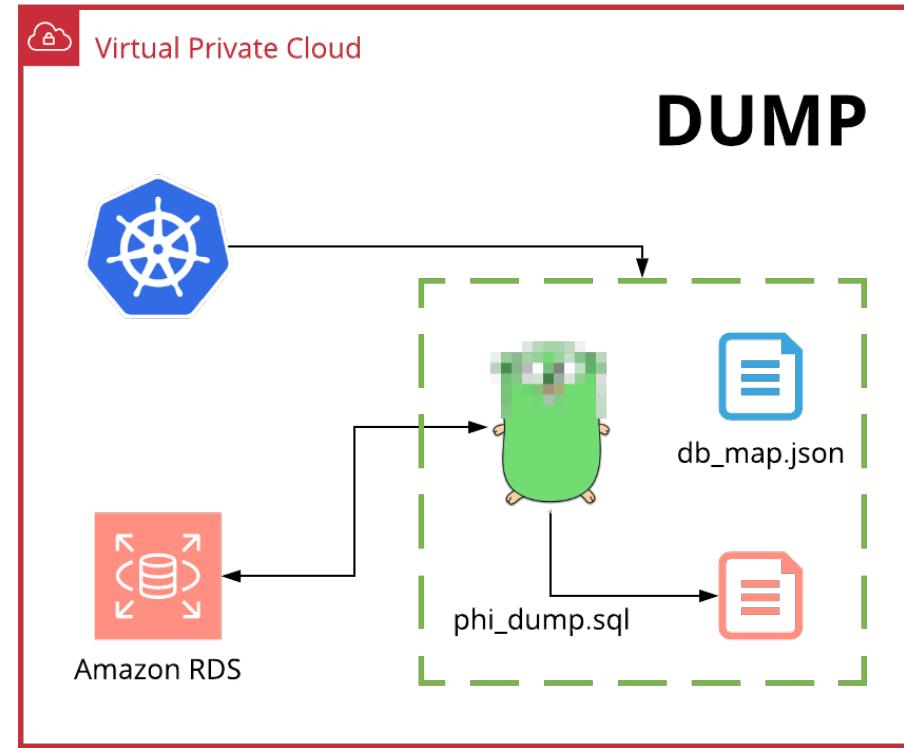
## Development



# Automation & Kubernetes

- Dump
- Process
- Load

# PRODUCTION



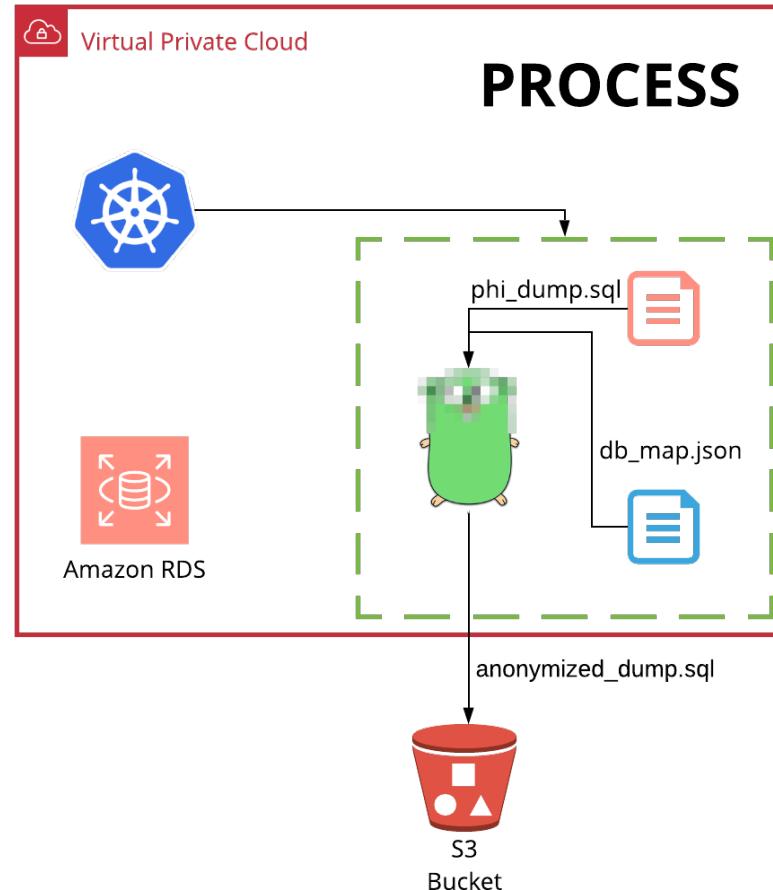
Smi+hRx

# Automation & Kubernetes

- Dump
- Process
- Load

Smi+hRx

## PRODUCTION

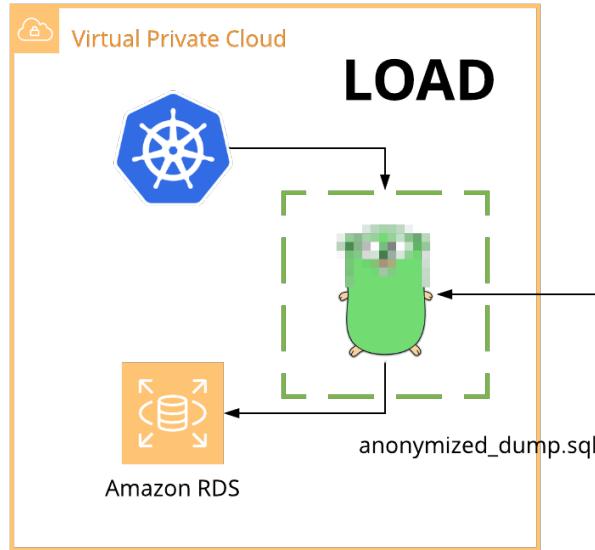


# Automation & Kubernetes

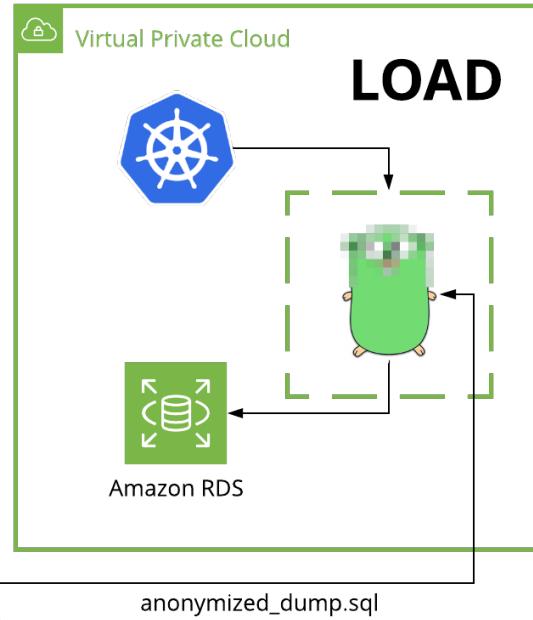
- Dump
- Process
- Load

Smi+hRx

## Staging



## Development





# Now a demo

The demo database comes from the Dell-Store 2 collection found at: <http://pgfoundry.org/projects/dbsamples/>

We are hiring  
mission-driven devs  
and PMs passionate  
about fixing U.S.  
healthcare

Smi+hRx

## Office Location

San Francisco, CA

## Current Positions Open

Software Engineer (mid-level)

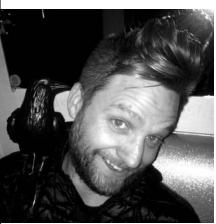
Senior Software Engineer

Product Manager (mid-level)

Senior Product Manager

## Send Resume To

[recruiting@smithrx.com](mailto:recruiting@smithrx.com)





# Call for Contributors

<https://github.com/smithoss/gonymizer>

# Gonymizer



PASSSED codecov 61% go report A+ godoc reference

## Weird name, what does it do?

The Gonymizer project (Go + Anonymizer) is a project that was built at Smit. Gonymizer is written in Go lang and is meant to help database administrator production database dumps before loading this data into a QA environment.

We have built in support, and examples, for:

Author ▾ Labels ▾ Projects ▾ Milestones ▾ Assigned ▾

20 Open ✓ 9 Closed

- ⓘ Create Country Processor enhancement #36 opened 8 days ago by junkert
- ⓘ Create the English Honorifics Scrambler enhancement #35 opened 8 days ago by junkert
- ⓘ Create Processor for TSRange enhancement #34 opened 12 days ago by junkert
- ⓘ Help screen examples are incorrect bug docs #33 opened 13 days ago by junkert
- ⓘ Load Command Contains --skip-procedures is this valid? question #32 opened 13 days ago by junkert
- ⓘ Investigate why system wide stored procedures are included in dump file enhancement #30 opened on Apr 4 by junkert
- ⓘ Create Docker Hub Images enhancement #27 opened on Mar 27 by junkert
- ⓘ Create CRONJOB BASH script examples to use local disk as storage #25 opened on Mar 19 by junkert
- ⓘ Create Lambda Examples enhancement #24 opened on Mar 19 by junkert



Thank you