

# Securing & Protecting Data in DevOps...or any way else

Karen Lopez  
Data Evangelist  
InfoAdvisors  
[www.datamodel.com](http://www.datamodel.com)

1

2

[www.spotthestation.nasa.gov](http://www.spotthestation.nasa.gov)



3

3

KL4



**KAREN LOPEZ**  
SENIOR CONSULTANT  
DIGITAL SMARTIESKIRT, INFOADVISORS



**Microsoft**  
Most Valuable Professional

 @datachick

 facebook.com/lopezk

 linkedin.com/in/karenlopez

### Karen Lopez

- Karen has 20+ years of data and information architecture experience on large, multi-project programs
- She is a space fan
- She wants you to love your data





4

KL5

POLL: Who Are You?



5

5

## Slide 4

---

**KL4** Update needed  
Karen Lopez, 11/1/2018

## Slide 5

---

**KL5** Need to verify the primary speaker throughout.  
Karen Lopez, 11/1/2018



### Why this topic?

- **Because**
- We
- Love
- Our
- Data

6

### Let's Go!



7

About this session


- Interactive\*
- Some Questions
- Several Answers
- I'm a DATA person

- "At another company"
- Sharing data tools & approaches
- SQL Server as example
- Love Your Data

8

8

Protecting data




OVERVIEW



DISCOVER




CATEGORIZE



PROTECT



MONITOR & ASSESS



MORE  
THOUGHTFUL  
STUFF

9

9

Let's Chat

How did the most recent data breaches happen?

Which are the most embarrassing for the the IT Profession?

...for the developer profession?

10

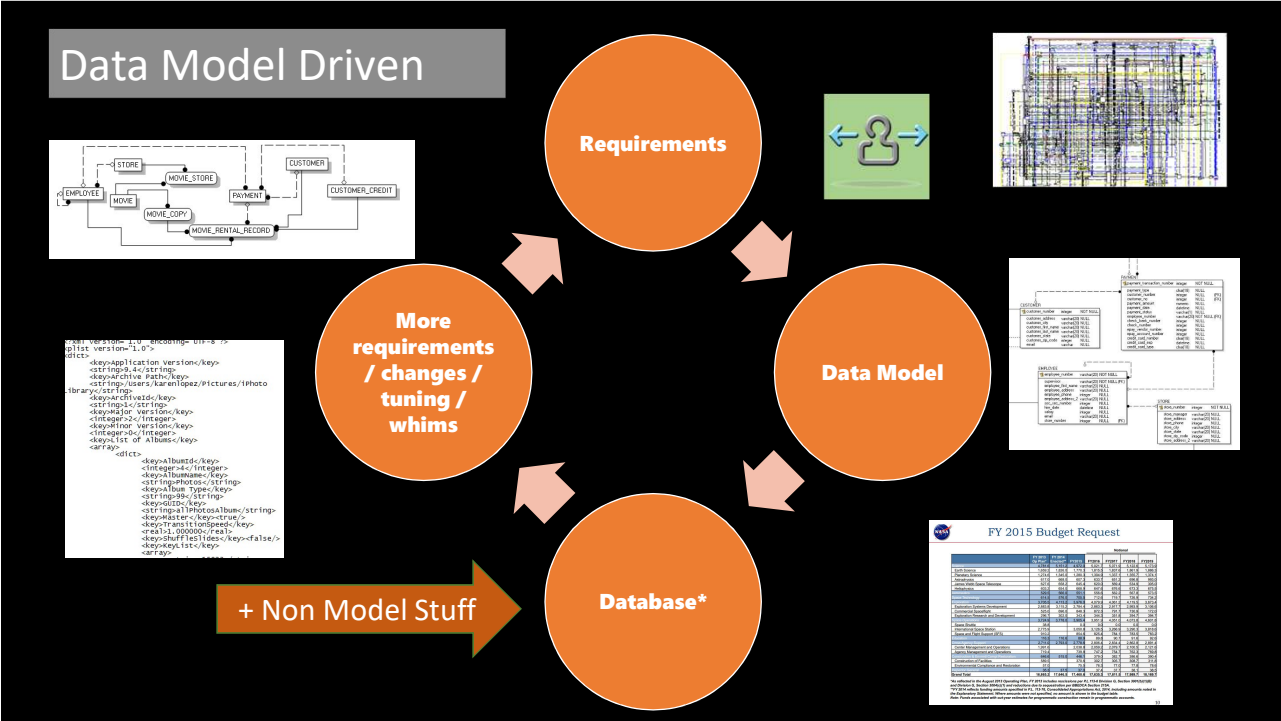
10

Is it after 25 May?  
What year?

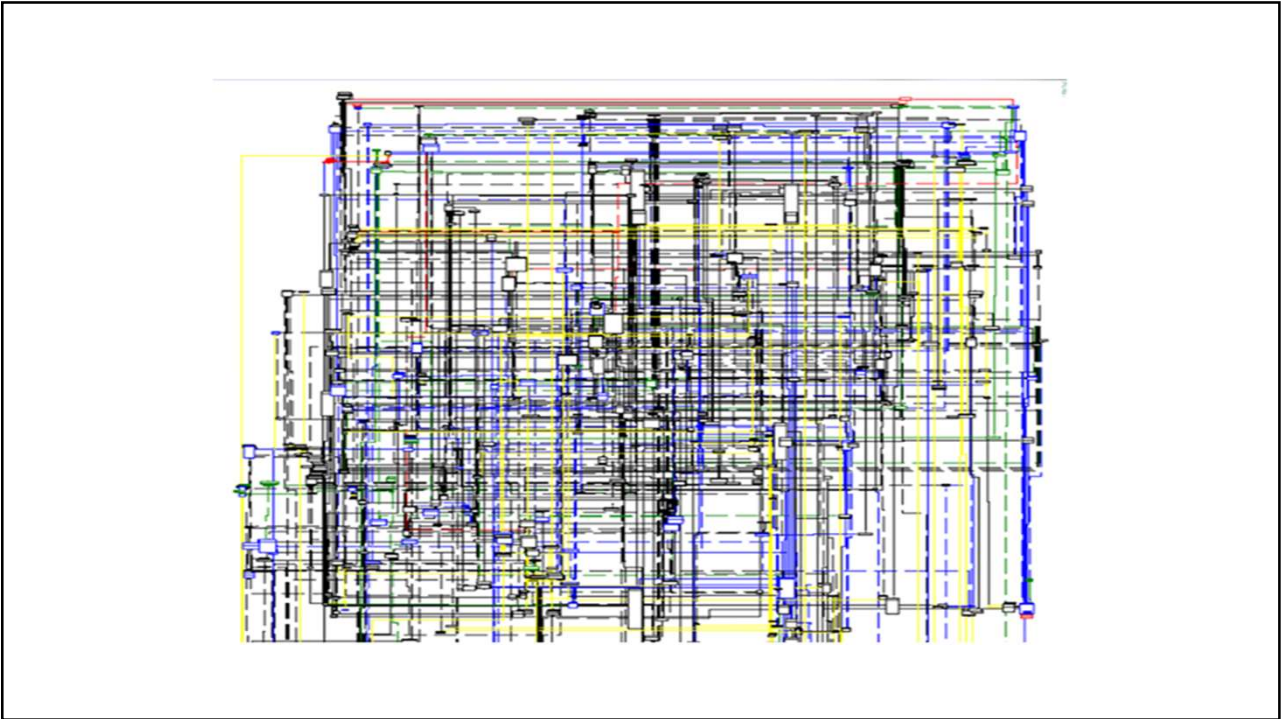
What about Dev, Ops, Data,  
and Security?

11

11



12



13



Data Models

- Karen’s Preference
- Track all kinds of metadata
- Advanced Compare features
- Support DevOps and Iterative development
- Support Conceptual, Logical and Physical design

Employee

Employee ID	INTEGER	NOT NULL
@TwitterName	CHAR(140)	NULL
National ID Number	CHAR(15)	NOT NULL
ContactID (FK)	INTEGER	NOT NULL
Login ID	CHAR(256)	NOT NULL
ManagerID (FK)	INTEGER	NULL
Title	CHAR(50)	NOT NULL
Birth Date	DATE	NOT NULL
Marital Status	BIT	NOT NULL
Gender	CHAR(1)	NOT NULL
Hire Date	DATETIME	NOT NULL
Salaried Flag	Flag	NOT NULL
VacationHours	INTEGER	NOT NULL
SickLeaveHours	SMALLINT	NOT NULL
CurrentFlag	Flag	NOT NULL
Modified Date	DATETIME	NOT NULL

Quality Steward: Betty Case

Design Steward: Karen Lopez

Executive Steward: Woodward

Privacy Level: Highly Confidential

Security Impact: High

Compliance Mapping: EU Data Privacy

is managed by

was paid

14

Normal Conflicts..

Developers vs. Data Quality

Data Professionals vs. Development Speed

Data vs. Code

Data vs. Metadata

Software Defined vs. Data Defined

15



## Karen's Data Governance Position

---

**Data security at the **data** level**

---

**Models & catalogs capture security/privacy needs**

---

**Design security from the start**

---

**Measurement & monitoring**

---

**In other words, governance**

16

## Typical DevOps Security Focuses

---



Code Reviews



Auditing



Key and Secrets  
Management



Repositories



...what else?

17

Typical  
DevOps Data  
Security  
Misses

What’s actually in that database?

What’s actually in that JSON/XML?

Where did the test data come from?

What’s actually in that \_\_\_\_\_?

What did I just post to Github?

18

18


Discovery


What do we have?  
Where is it? How do we  
know?


19


19

Data  
Classification  
/Categorization

 Syntax-based


 Sematic-based


 AI-based


 Data Profiling vs. Data Naming


21

Data Curation

 Related to Data Stewardship


 Covers more than Data Categorization

 Important part of Data Governance

 New-ish term going into GDPR and other protection concepts

22

One more time...



A photograph of Karen Lopez, a woman with short reddish-brown hair and glasses, wearing a black blazer. She is standing behind a podium with a microphone. A red and white name tag is pinned to her blazer that reads: **@DATACHICK**  
**KAREN LOPEZ**  
SENIOR CONSULTANT  
DIGITAL SMARTIESKIRT, INFOADVISORS

Every Design Decision must be based on **Cost**, **Benefit** and **Risk**


23

Data Curation


- Takes time, but...
- Builds on other efforts
- Contributes to future efforts

24


# Catalog Data Assets




Every compliance effort starts with inventory



Capture the hard work of every project



Build incrementally



Start with what exists physically

25

25

## Welcome!

Microsoft Azure Data Catalog is a fully managed cloud service that enables users to discover, understand, and consume data sources.

This application extracts data source metadata such as object names, attribute names, and data types, and registers that metadata in Azure Data Catalog.

Please sign in to Azure Data Catalog to continue.

Sign in >

Modify proxy settings.

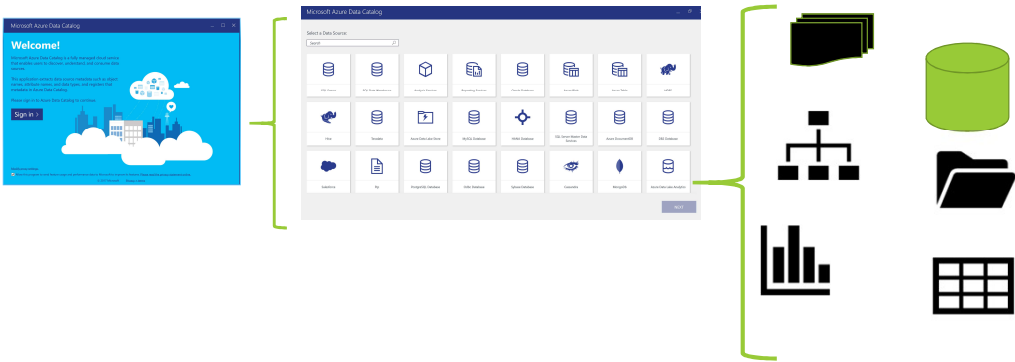
☒ Allow this program to send feature usage and performance data to Microsoft to improve its features. Please read the privacy statement online.

# Azure Data Catalog

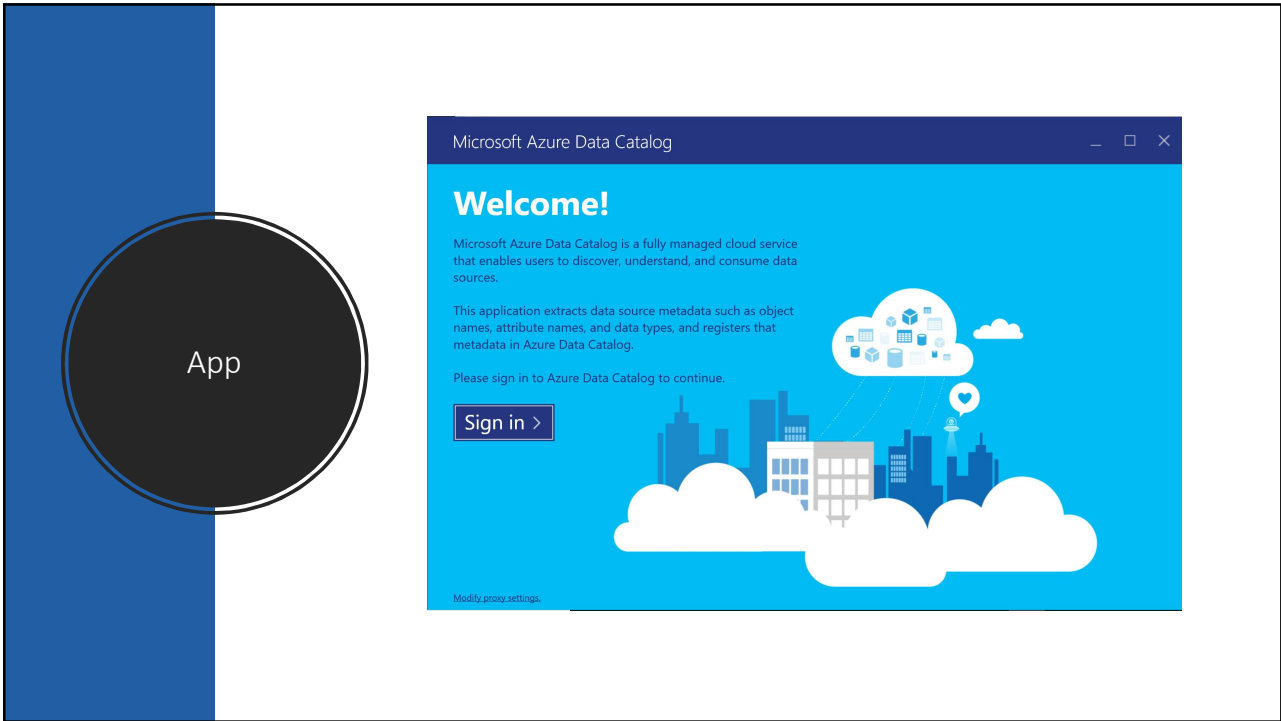
Azure Data Catalog is a fully managed **cloud service** whose users can **discover the data sources** they need and **understand the data sources** they find. At the same time, Data Catalog helps organizations get more value from their existing investments.

26

# Azure Data Catalog



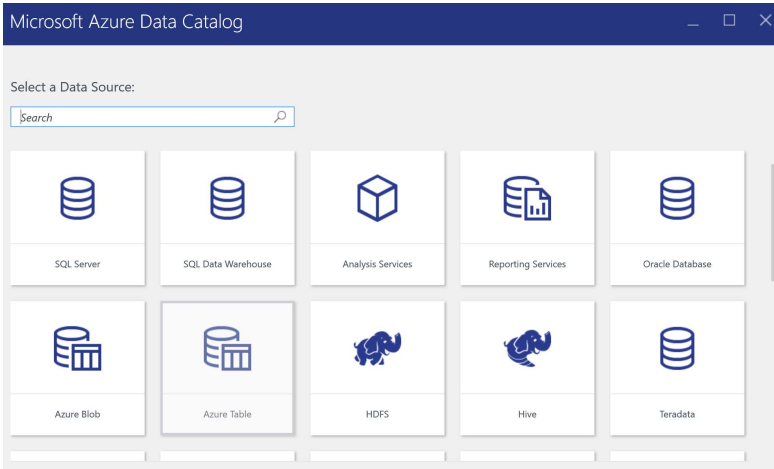
27



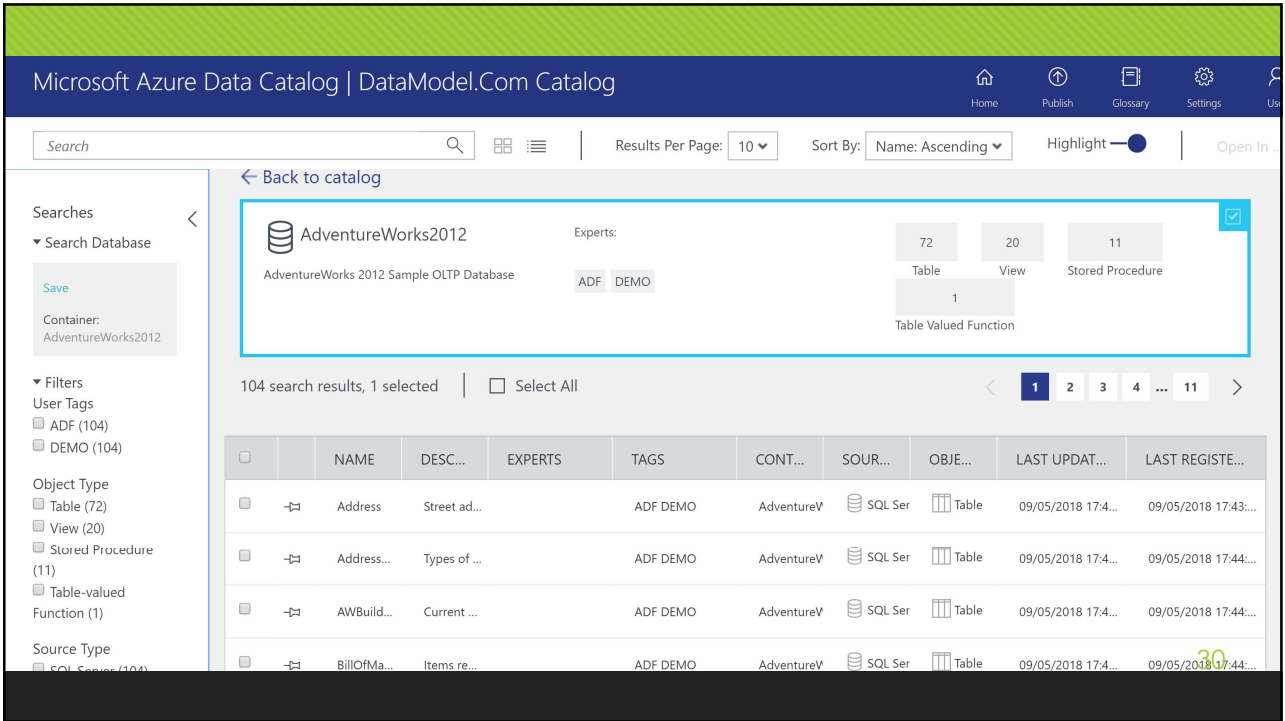
28

Data Source

- Microsoft
- Oracle
- Hadoop
- DB2
- Teradata
- MySQL
- HANA
- Salesforce
- ..and more

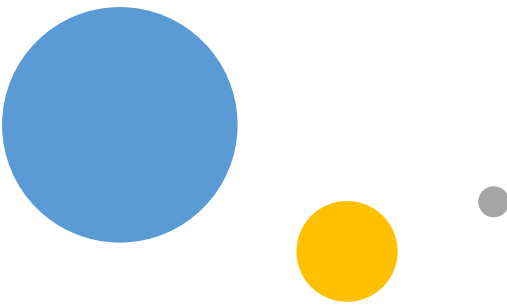


29



30





# Categorization

Sensitive, Confidential,  
PII and Special Data

32

32



## But really, who?

- End Users
- Self-Serve BI Users
- DBAs
- Developers
- Ops
- Data Architects

33

Other Options

Informatica

IBM Watson

Erwin Data Governance

Data Modeling Tool Portal

???

34

34

Assess

What sorts of data do we steward? How should we protect it?

36

36

Sensitivity Label

Confidential

Public

General

Confidential

Confidential - GDPR

Highly Confidential

Highly Confidential - GDPR

[n/a]

ptions (click to minimize)

	Schema	Table	Column	Information Type	Sensitivity Label
<input type="checkbox"/>	Sales	CreditCard	CreditCardID	Credit Card	Confidential
<input type="checkbox"/>	Sales	CreditCard	ExpMonth	Credit Card	Confidential
<input type="checkbox"/>	Sales	CreditCard	ExpYear	Credit Card	Confidential
<input type="checkbox"/>	Sales	Tarjeta	ExpMonth	Credit Card	Confidential
<input type="checkbox"/>	Sales	Tarjeta	ExpYear	Credit Card	Confidential
<input type="checkbox"/>	Sales	Tarjeta	TarjetaCreditoID	Credit Card	Confidential

37

SQL Data Classification Report

Microsoft

Server name: PORC\SQL2012  
Database name: AdventureWorks2008

Report generated on: 2/28/2018 6:40:17 AM

Classified columns  
39 / 502

Tables containing sensitive data  
19 / 72

Unique information types  
6

Label distribution

Confidential

Confidential - GDPR

26

12

Information Type distribution

11

9

4

2

2

3

Contact Info

Credentials

Credit Card

Financial

Name

National ID

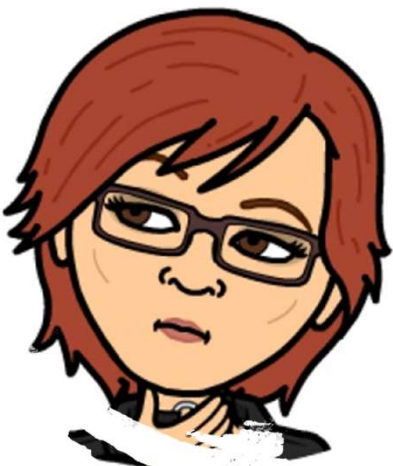
Schema	Table	Column	Information Type	Sensitivity Label
dbo	ErrorLog	UserName	Credentials	Confidential
HumanResources	Employee	NationalIDNumber	National ID	Confidential - GDPR
Person	Address	AddressLine1	Contact Info	Confidential - GDPR
Production	ProductReview	EmailAddress	Contact Info	Confidential - GDPR
Purchasing	Vendor	AccountNumber	Credentials	Confidential
Sales	CountryRegionCurrency	CurrencyCode	Financial	Confidential

38

38

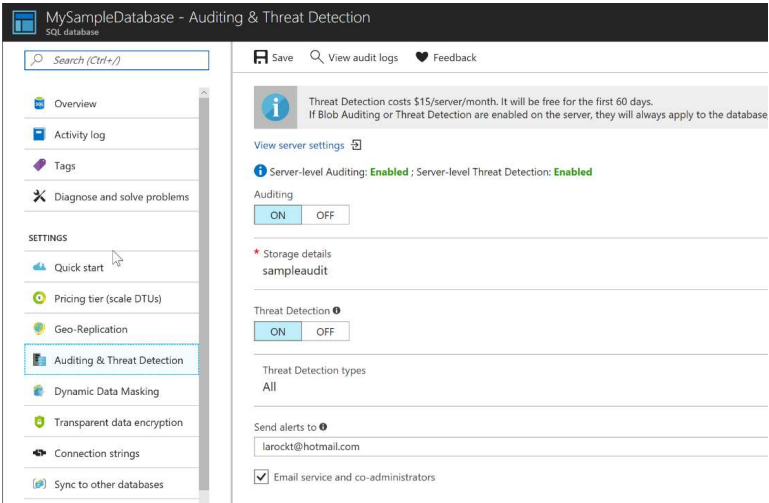
# Issues

- Data Pros spend 80% of their time sourcing, prepping and cleansing data
- Likely everyone else has these issues
- We are lousy at documenting data and meta data
- This makes Karen sad

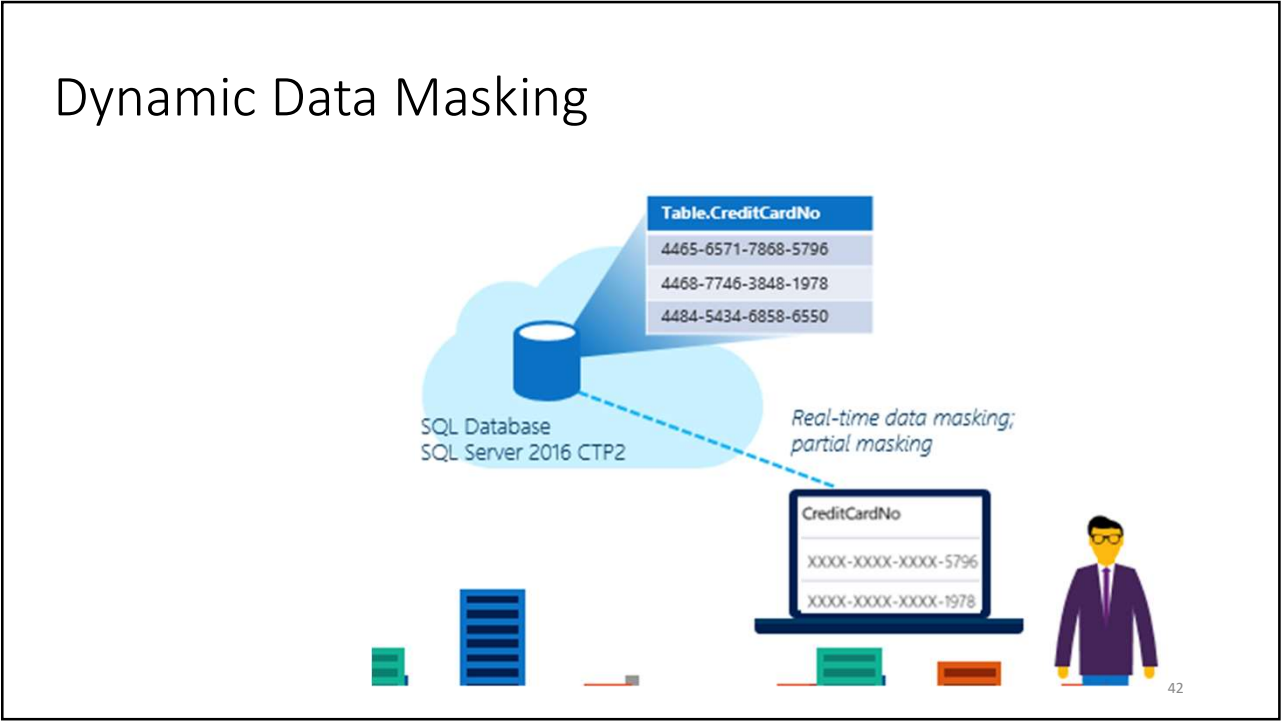


39

# Auditing and Threat Detection



40



42

## Data Masking Examples

XXXX XXXX XXXX 1234

[kxxxxxxx@ixxxxx.com](#)

\$99,9999

June, 99, 9999

KXXXXXX Lopez

43

43

# Privacy - Dynamic Data Masking

```
CREATE TABLE Membership(  
    MemberID int IDENTITY PRIMARY KEY,  
    FirstName varchar(100) MASKED WITH (FUNCTION =  
'partial(1,"XXXXXXX",0)') NULL,  
    LastName varchar(100) NOT NULL,  
    Phone# varchar(12) MASKED WITH (FUNCTION = 'default()') NULL,  
    Email varchar(100) MASKED WITH (FUNCTION = 'email()') NULL);  
  
INSERT Membership (FirstName, LastName, Phone#, Email) VALUES  
( 'Roberto', 'Tamburello', '555.123.4567', 'RTamburello@contoso.com'),  
( 'Janice', 'Galvin', '555.123.4568', 'JGalvin@contoso.com.co'),  
( 'Zheng', 'Mu', '555.123.4569', 'ZMu@contoso.net');
```

44

# Dynamic Data Masking



Column level



Data in the database, at rest, is not masked



Meant to *complement* other methods



Performed at the end of a database query right before data returned



Performance impact small

45

Security –  
Dynamic Data  
Masking in  
SQL Server

4  
functions  
available.  
today

- Default
- Email
- Custom String
- Random

46

46

Dynamic Data Masking

01  
Data in database is  
not changed

02  
Ad-hoc queries  
\*can\* expose data

03  
Does not aim to  
prevent users from  
exposing pieces of  
sensitive data


48

48



# Why would a Data Pro love it?

- Allows central, reusable design for standard masking
- Offers more reliable masking and more usable masking
- Applies across applications
- Removes whining about “we can do that later”



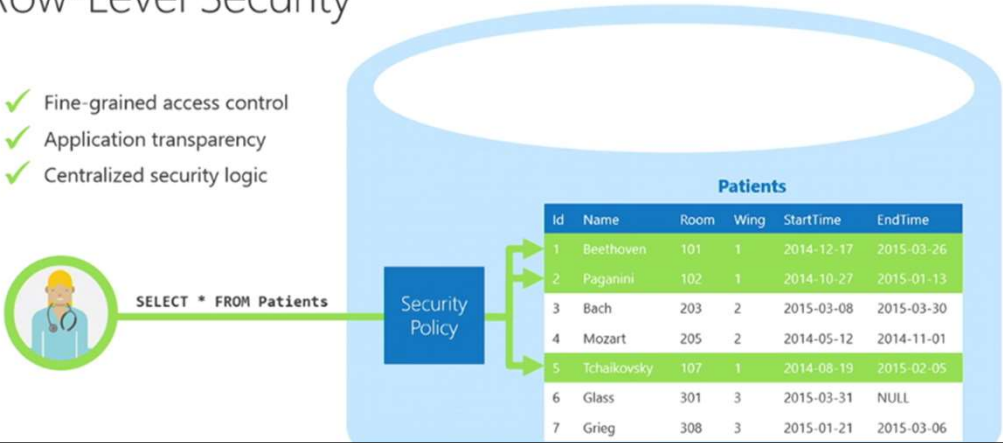
50

50

# Security – Row Level Security

## Row-Level Security

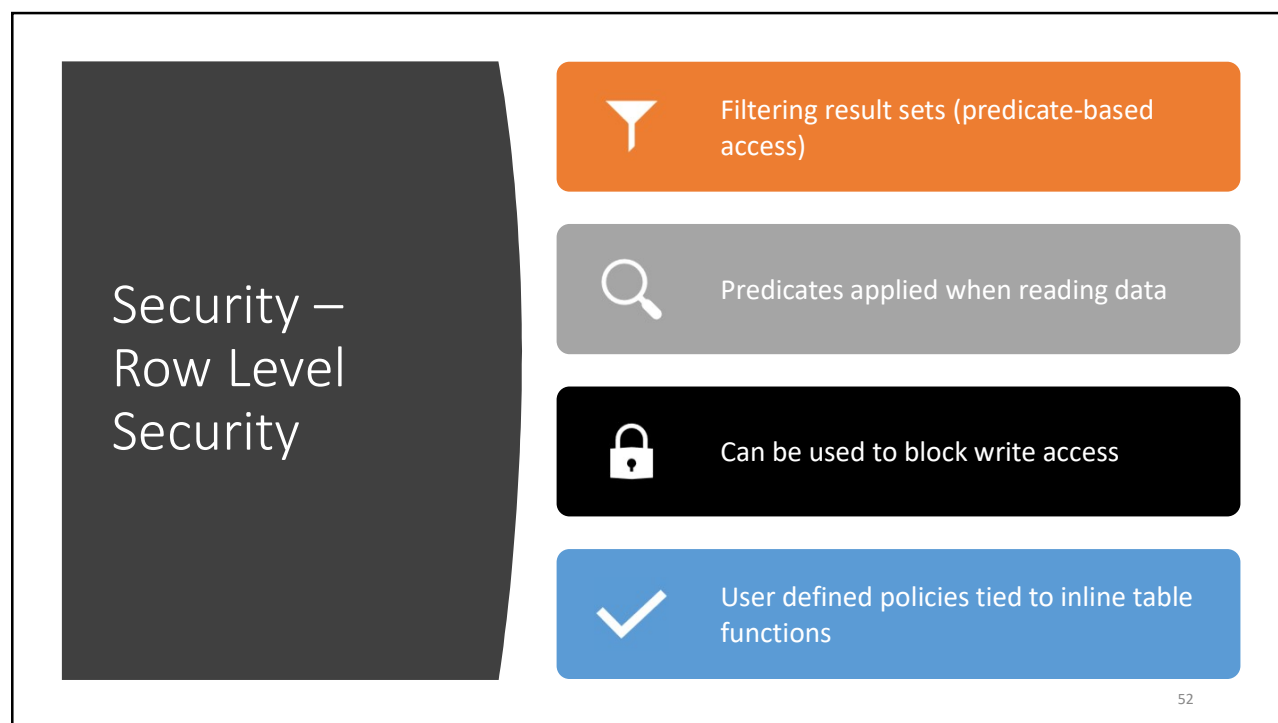
- ✓ Fine-grained access control
- ✓ Application transparency
- ✓ Centralized security logic



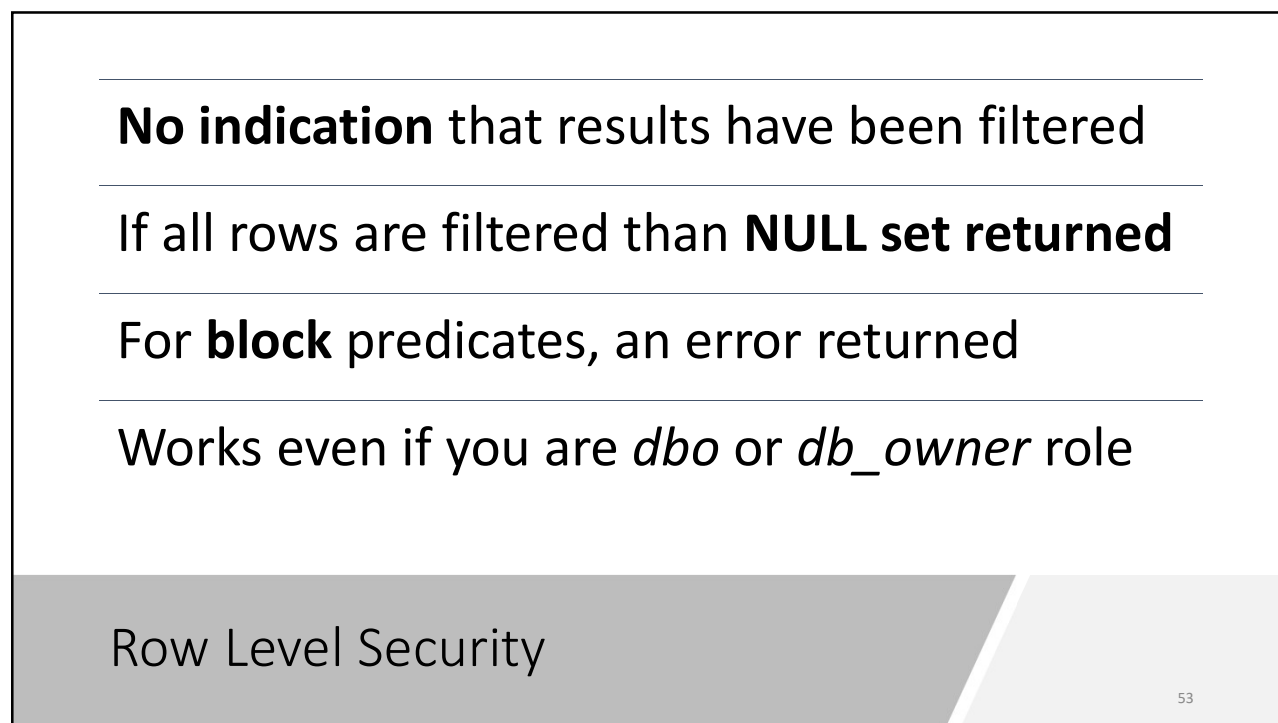
Patients					
Id	Name	Room	Wing	StartTime	EndTime
1	Beethoven	101	1	2014-12-17	2015-03-26
2	Paganini	102	1	2014-10-27	2015-01-13
3	Bach	203	2	2015-03-08	2015-03-30
4	Mozart	205	2	2014-05-12	2014-11-01
5	Tchaikovsky	107	1	2014-08-19	2015-02-05
6	Glass	301	3	2015-03-31	NULL
7	Grieg	308	3	2015-01-21	2015-03-06

51

51




52



53

# Why would a Data Pro love it?

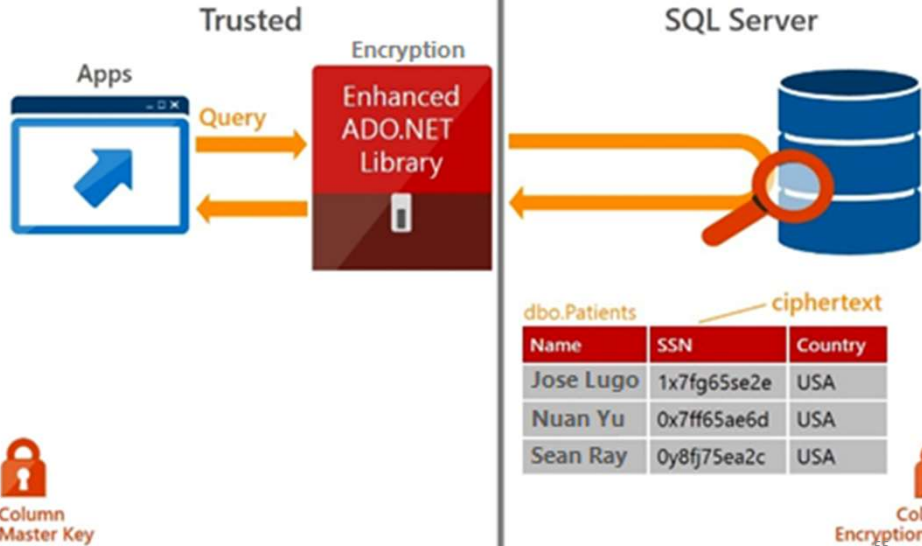
- Allows a designer to do this sort of data protection IN THE DATABASE, not just rely on code.
- Many, many pieces of code
- Applies across applications



54

# Security – Always Encrypted

Always!



The diagram illustrates the Always Encrypted architecture. On the left, 'Apps' send a 'Query' to an 'Enhanced ADO.NET Library'. This library then interacts with the 'SQL Server' database. The database contains a table named 'dbo.Patients' with columns 'Name', 'SSN', and 'Country'. The 'SSN' values are shown as ciphertext. A 'Column Master Key' is used for encryption, and a 'Column Encryption Key' is used for decryption.

Name	SSN	Country
Jose Lugo	1x7fg65se2e	USA
Nuan Yu	0x7ff65ae6d	USA
Sean Ray	0y8fj75ea2c	USA

55

## Security – Always Encrypted



ENABLED AT COLUMN LEVEL



PROTECTS DATA AT REST  
\*AND\* IN MEMORY



USES COLUMN MASTER KEY  
(CLIENT) AND COLUMN  
ENCRYPTION KEY (SERVER)

56

56


## Always Encrypted

```
-----
CREATE COLUMN ENCRYPTION KEY MyCEK
WITH VALUES
(
    COLUMN_MASTER_KEY = MyCMK,
    ALGORITHM = 'RSA_OAEP',
    ENCRYPTED_VALUE = 0x01700000016C006F00630061006C006D0061006300680069006E0065002F00
);
-----
CREATE TABLE Customers (
    CustName nvarchar(60)
        COLLATE Latin1_General_BIN2 ENCRYPTED WITH (COLUMN_ENCRYPTION_KEY = MyCEK,
        ENCRYPTION_TYPE = RANDOMIZED,
        ALGORITHM = 'AEAD_AES_256_CBC_HMAC_SHA_256'),
    SSN varchar(11)
        COLLATE Latin1_General_BIN2 ENCRYPTED WITH (COLUMN_ENCRYPTION_KEY = MyCEK,
        ENCRYPTION_TYPE = DETERMINISTIC,
        ALGORITHM = 'AEAD_AES_256_CBC_HMAC_SHA_256'),
    Age int NULL
);
GO
```


57

57

Security –  
Always  
Encrypted



Foreign keys must match encryption types



Client code needs to support AE (currently this means .NET 4.x)

58

58

Security –  
Always  
Encrypted

Wizard

Always Encrypted

Column Selection





Introduction  
Column Selection  
Master Key Configuration  
Validation  
Summary  
Results

Search column name...

☐ Apply one key to all checked columns: CEK\_Auto2 (New)

Encryption Type

Encryption Key

Name	State	Encryption Type	Encryption Key
Sales.CurrencyRate			
Sales.Customer			
Sales.CustomerPII			
CustomerID			
FirstName			
LastName			
SSN		Deterministic	CEK_Auto1
CreditCardNum...		Deterministic	CEK_Auto1
EmailAddress			
PhoneNumber			
TerritoryID			
Sales.OrderTracking			
Sales.PersonCreditC...			
Sales.SalesOrder_json			
Sales.SalesOrderDetail			
Sales.SalesOrderDet...			
Sales.SalesOrderDet...			

☐ Show affected columns only

< Previous

Next >

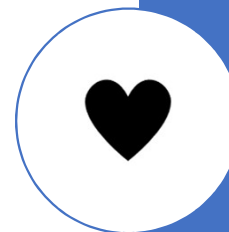
Cancel

59

59

## Why would a Data Pro love it?

- Always Encrypted, yeah.
- Allows designers to not only specify which columns need to be protected, but how.
- Parameters are encrypted as well
- Built in to the engine, easier for Devs




60

## What should we STOP doing?

Nobody ever talks about this....

61

61




## SQL Injection


- WE ARE STILL DOING THIS!
- IT'S STILL THE #1 (but unsecured storage is getting more popular)
- TEST. TEST SOME MORE
- Automated Testing
- Governance is important

63


## Unprotected “buckets”



“I’LL DELETE IT ONCE YOU GRAB THAT FILE”



SHARING WITH 2<sup>ND</sup> PARTIES



SHARING WITH 3<sup>RD</sup> PARTIES

64



---

Good people don't always stay that way

---

People mess up

---

Monitoring

---

Checking

---

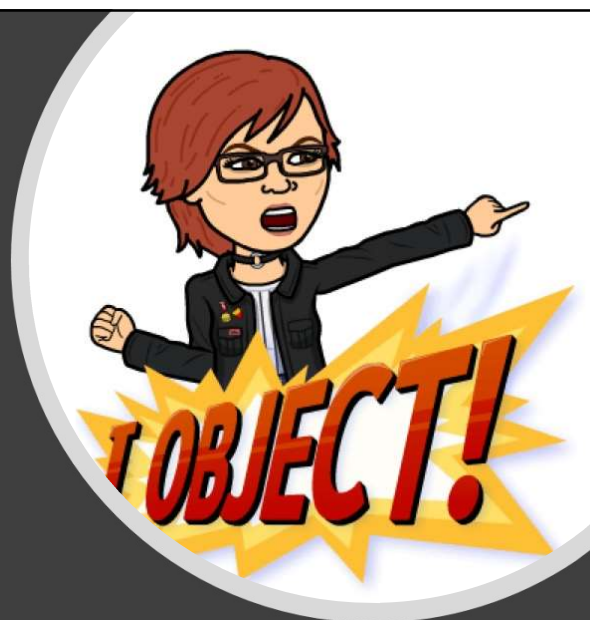
Automatic alerting

---

Trusting good people

65

Karen's Rant Topic for  
2019



66

66

## Test Data



- Restoring Production to Development
- Restoring Production, with Masking
- Restoring Production, with Randomizing
- Restoring Production...anywhere



- Design Test Data
- *Lorem Ipsum* for Data
- Really, Design Test Data

67

67

## Building a Culture of Data Security & Privacy

- **Reward** identification of threats
  - **Reward** identification of risks
  - Trust, but always cut the deck
  - Monitor, test, monitor, test, monitor...
- Be a customer with your data in there
  - Don't use production data for anything other than production and support

68

68

## Thank You

- @DataChick
- karenlopez@infoadvisors.com



75

75