Achieving Data Privacy Compliance in Postgres or Greenplum

Data Protection versus Data Access Control . . . What's the difference and why you should care

April 20, 2018

PostgresConf US 2018

Jersey City / United States

April 16 - 20, 2018

Les McMonagle (CISSP, CISA, ITIL)

VP of Security Strategy

BlueTalon

Agenda

- Data Protection versus Data Access Control – What’s the Difference ?
- Why are these applicable to GDPR, HIPAA, COPPA, GLBA and other Regs
- Generally Accepted Privacy Principles – GDPR is not so new . . .
- Getting Postgres or Greenplum ready for processing regulated PII
- Q & A

Unified, Attribute Based Access Control (ABAC)

Versus

Data Protection using Encryption and Tokenization

Choosing one or the other or both

Typical Security Stack and Functionality

Security Functionality	Attribute Based Access Control	Data Protection Tools
Authentication	For security admins only	For security admins only
Audit	**Access auditing and lineage from capture of data usage**	When data is protected or unprotected
Authorization	**Row, column, filtering & attribute based policies across data platforms**	Typically only based on RBAC Must call UDF
Masking	**Dynamic masking based on any User or Data attributes**	Data Obfuscation tools, some others
Encryption	Seamless interoperation only	**Disk, Volume, File Level encryption and UDF-based field level encryption**

The Next Evolution of Data Access Control

1st Generation: Rights to database objects granted to Users (90's)

2nd Generation: Rights to database objects granted to Roles (RBAC)

- Then Role membership managed in shared external LDAP directories

3rd Generation: Attribute Based Access Control (ABAC)

- Complexity and granularity requirements exceeding existing capabilities
- Based on virtually any User or Data Attribute(s)
- Centralized management and control, consistently applied rules
- Layered with targeted Data Protection (Disk, Volume, File, Table, Column Level)
- Overlapping hierarchy and matrixed access rights compliance requirements

Gartner Agrees: Companies will insist on "centralized management platforms that can directly control data security policies across multiple data silos."

Data Protection

PROS:

- Critical data fields protected at-rest
- Independent access control to protected fields
- Independent audit trail of access to protected data fields
- Centralized key management
- Data can be moved in protected form (when there is no "T" in ETL)

CONS:

- Physical data model changes required
- Logical data model changes required
- For ETL, data must be unprotected, transformed and protected again
- Sensitivity to original Data Type
- Significant performance impact on any full column operations
- "Like" or "Range" operations limited
- Access control and audit trail on encrypted or tokenized fields only

Great fit when data protection & access control only required for a few critical data fields (PCI-DSS)

Enterprise Wide Attribute Based Data Access Control

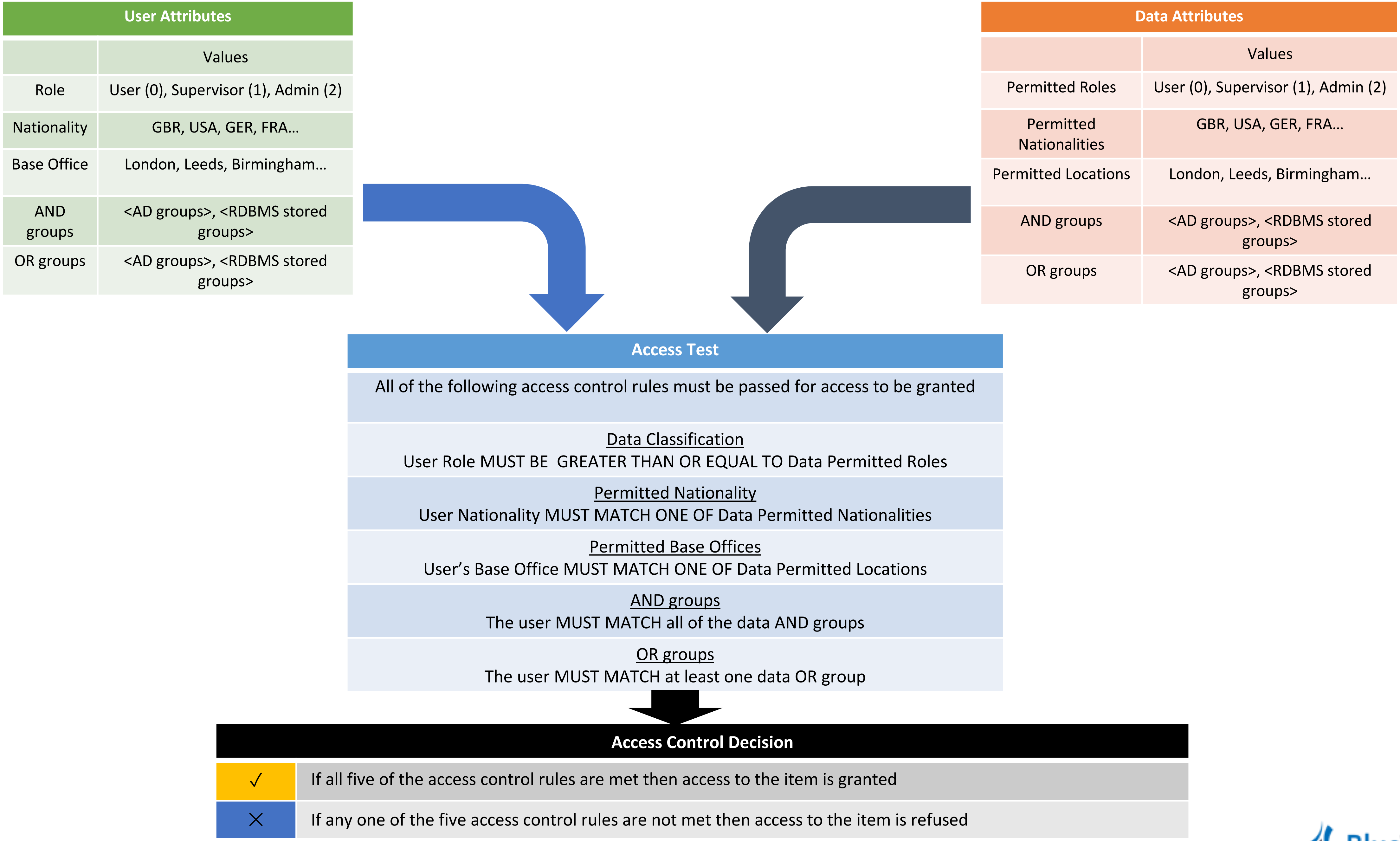
PROS:

- Unified, comprehensive, centrally managed, data access control consistently applied across all platforms
- ABAC using any user or data attribute(s)
- No changes to data/application/client layer
- No negative performance impact
- Transparent to users and applications
- Independent tamper-proof audit trail of all access to data
- Lower TCO through reduced DBA costs

CONS:

- Base table (at-rest) data in the clear
- Some added latency to queries

Typically the cornerstone of any enterprise wide data access control or data governance program



User Attributes	
	Values
Role	User (0), Supervisor (1), Admin (2)
Nationality	GBR, USA, GER, FRA…
Base Office	London, Leeds, Birmingham…
AND groups	<AD groups>, <RDBMS stored groups>
OR groups	<AD groups>, <RDBMS stored groups>

Data Attributes	
	Values
Permitted Roles	User (0), Supervisor (1), Admin (2)
Permitted Nationalities	GBR, USA, GER, FRA…
Permitted Locations	London, Leeds, Birmingham…
AND groups	<AD groups>, <RDBMS stored groups>
OR groups	<AD groups>, <RDBMS stored groups>

Access Test
All of the following access control rules must be passed for access to be granted
Data Classification User Role MUST BE GREATER THAN OR EQUAL TO Data Permitted Roles
Permitted Nationality User Nationality MUST MATCH ONE OF Data Permitted Nationalities
Permitted Base Offices User’s Base Office MUST MATCH ONE OF Data Permitted Locations
AND groups The user MUST MATCH all of the data AND groups
OR groups The user MUST MATCH at least one data OR group

Access Control Decision	
✓	If all five of the access control rules are met then access to the item is granted
✗	If any one of the five access control rules are not met then access to the item is refused

BlueTalon

Example Complex, Fine-Grained Access Control Decision Details

- User Role to Data Permitted Roles
  - Evaluated in the following logical manner:
    - IF User Role = Admin (2) THEN: user can access all data but certain fields are masked
    - ELSE IF User Role = Supervisor (1) THEN: user can only see User (0) and Supervisor (1) data

- User Nationality to Data Permitted Nationalities
  - The User's Nationality must be one of the Data Permitted Nationalities – Label Based, Mandatory Access Control (MAC)

- User Base Office to Data Permitted Locations
  - The User's Base Office must be one of the Data Permitted Locations

- User AND groups to Data AND groups
  - Evaluated in the following logical manner:
    - IF Data AND groups = *none* THEN: the Data AND groups have no effect on user data access
    - ELSE: the User's AND groups must include all the groups listed in the Data's AND groups

- User OR groups to Data OR groups
  - Evaluated in the following logical manner:
    - IF Data OR groups = *none* THEN: the Data OR groups have no effect on user data access
    - ELSE: one of the User's OR groups must match a group listed in the Data's OR groups

Why settle for one when you can use both (ABAC & Protection)

PROS:

- Independent access control on all PII
- At-Rest encryption or tokenization of critical fields
- Data often remains in protected form
- Maximum protection
- Minimal performance impact
- Minimal operational impact
- Minimized risk
- Easily manage both centrally when properly integrated

CONS:

- Added data security software costs
- Some additional administration work

Powerful combination gaining the benefits of both options and eliminating many of the cons

Generally Accepted Privacy Principles

Global Data Privacy Regulation Convergence

Long standing, internationally accepted, privacy principles include: 1/3

Accountability:
An organization should be held accountable for PII under its control.

Notice:
Notice must be provided to the Data Subject of the purpose for collecting PII.
Data Subject must be notified of applicable policies (Consent, Access, Disclosure).
Notice must be provided of any changes to the applicable privacy policies or the data collected is used for any reason other than the originally stated purpose.
Notice must be provided in clear and conspicuous language.
Notification should be sent at the time of collection or immediately before.

Consent:
Data Subjects must be informed of, and explicitly consent to, the collection, use and disclosure of sensitive information.
The Data Subject must provide informed consent to the collection of PII.
Consent required to use PII for purposes other than those originally stated.
Data Subjects must be made aware of the consequences of denying consent.

Long standing, internationally accepted, privacy principles include: 2/3

Collection Limitation:
Only PII relevant to the identified purpose may be collected.
Information must be collected by fair and lawful means.

Use Limitation:
PII may only be used for the purposes stated at the time of collection.
PII is retained no longer than necessary to complete the stated purpose.

Disclosure:
Consent from the Data Subject is required to disclose information to third parties.
Organizations must ensure that any third parties comply with their privacy policies.
Information may be disclosed if required by law or for health and safety reasons.

Access and Correction:
Data Subjects are able to able to access PII an organization maintains on them.
Data Subjects requesting information must supply sufficient proof of identity.
Requested PII is provided clearly, at reasonable cost and within a reasonable time.
If denied access, Data Subject is informed of reason, options to challenge denial.
Data Subjects are able to update or correct PII held by the organization.

Long standing, internationally accepted, privacy principles include: 3/3

Security and Safeguards:
Provide safeguards to prevent loss, misuse, unauthorized access, disclosure, alteration and destruction of PII.
Destroy or permanently obfuscate discarded or expired data.

Data Quality:
Organizations must ensure PII is accurate, complete and up-to-date.

Enforcement:
Implement processes to ensure compliance with the privacy policies.
Include a process for data subjects to file complaints and have them reviewed.

Openness:
Ensure privacy policies are clearly published and publicly available.
Have a means to establish the existence, nature and purpose of use, of PII.

HIPAA Compliance

HITECH, Omnibus Rule and the HIPAA 18

BlueTalon

HIPAA 18 - All encrypted/tokenized/obfuscated = Safe Harbor

(A) Names

(B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code

(C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, etc.

(D) Telephone numbers

(L) Vehicle identifiers, serial numbers, license plate #'s

(E) Fax numbers

(M) Device identifiers and serial numbers

(F) Email addresses

(N) Web Universal Resource Locators (URLs)

(G) Social security numbers

(O) Internet Protocol (IP) addresses

(H) Medical record numbers

(P) Biometric identifiers, including finger and voice prints

(I) Health plan beneficiary numbers

(Q) Full-face photographs and any comparable images

(J) Account numbers

(R) Any other unique identifying #, characteristic, or code

(K) Certificate/license numbers

Refer to reference document:

HIPAA 18 FIELDS FOR SAFE HARBOR OR EXPERT DETERMINATION

HIPAA - Safe Harbor versus Expert Determination

Replicability Consistently occur in relation to an individual (Blood Glucose level vs DOB)

Data Source Availability How prevalent is the data in other locations / sources (Lab results vs Name, Address)

Distinguishability Extent the data can uniquely identify an individual
(Age/Gender/3 Digit Zip – 0.04% vs DOB/Gender/5 Digit Zip – 50%)

Assess Risk Combined Risk of Identification based on the above 3 characteristics (Low vs High)

De-identified health information utilizing these methods is no longer PHI
not protected by the Privacy Rule because it does not fall within the definition of PHI

Refer to the following URL for more details:

https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/

COPPA Compliance

Children’s Online Privacy and Protection Act (COPPA)

More specific, globally enforced, access control requirements

Strict rules governing access to PII on minors

- Federal and State specific rules (eg. Maine considers a minor anyone under 18 vs 13)
- International rules (GDPR), PIPEDA (Canada) and many other countries
- Collection and maintenance of parental permission
- Mrs. Fields Cookies fined for collecting addresses of minors without parental permission
- Must require and track age and permission info for all minors
- More limiting rules on exactly what data can be collected and for what purposes

Unified Data Access Control and GDPR

BlueTalon

Unique GDPR Regulatory Compliance Challenges

GDPR is the farthest reaching impactful regulation ever created

88 pages with 99 specific compliance requirements

Quickly becoming the global standard for data privacy and access control

Data Residency (Physical, Logical location)

Achieving Anonymisation or Pseudonymisation

FUD Factor - Attention Getting Financial Penalties

Convergence of Generally Accepted Privacy Principles

on Consent, Data Retention, Use, Disclosure, Access Control, Accountability, Destruction

Unified Data Access Control Applicability For GDPR Compliance

- **Fine-Grained, Consistently Applied, Global, Local and Context Sensitive Data Access Controls**
  - Ability to manage RLS and CLS by generating dynamic views based on virtually any User or Data attributes.
- **Ability to define and manage separate Data Domains and External Attribute Domains**
  - Data access control rules can be defined and applied globally or locally to different Data Domains.
  - Attribute Domains are internally, or externally, managed authoritative source(s) for attribute information.
- **Independent, Tamper-Proof Audit Trail of All Data Access – No need for DAM products**
  - Independent accountability and audit trail of all access to GDPR data on all platforms.
- **Consent Preferences, Tracking/Management of Opt-In, Opt-Out & Right-to-be-Forgotten Choices**
  - Majority of data fields on some systems are dedicated to tracking Data Subject preferences.
- **Reduced DBA, System Admin and Equivalent Labor Costs**
  - Eliminate cost of maintaining thousands of static security Views and associated View logic on all platforms.
- **Reduced Risk of a Data Breach or Unauthorized Access to PII**
  - Centrally managed, consistently and automatically applied, unified data access controls on all platforms.
- **Ability to Provide CLS and RLS on Platforms with No Other Option(s)**
  - Many new, less mature or open source data repositories provide only limited CLS or RLS capabilities.

All with minimal operational impact & little or no measurable performance impact

GDPR Articles Where ABAC & Encryption Play A Significant Role

Article 3: Territorial Scope:

Article 6: Lawfulness of Processing:

Article 7: Conditions for Consent:

Article 8: Conditions applicable to child's consent in relation to information society services:

Article 9: Processing of special categories of personal data:

Article 10: Processing of personal data relating to criminal convictions and offences:

Article 11: Processing which does not require identification:

Article 24: Responsibility of the controller:

Article 25: Data protection by design and by default:

Article 28: Processor:

Article 30: Records of processing activities:

Article 32: Security of processing:

Article 45: Transfers on the basis of an adequacy decision:

Article 46: Transfers subject to appropriate safeguards:

Article 47: Binding corporate rules:

Article 48: Transfers or disclosures not authorised by Union law:

Article 49: Derogations for specific situations:

. . . and playing a partial role in 16 others with including: 12, 13, 14, 15, 18, 26, 33, 35, 39, 40, 41, 42, 44, 58, 77, 83

GDPR Compliance Summary

Unified data access control capability, and Attribute Based Access Control (ABAC) are
Powerful tool(s) in achieving compliance with GDPR, HIPAA and other current and emerging data privacy regulations.

Data privacy regulations impose data privacy protection requirements, based on a range of data and user attributes.

Centrally managed ABAC enables organizations to easily and efficiently **map these constantly changing data privacy requirements to actual policy-based, technical controls** governing access to the sensitive or regulated data.

Rules are managed in a single location and applied consistently across a wide range of data storage platforms.

Market demand for this type of data-centric protection and access control is clear according to Gartner.

Managing multiple disparate, proprietary solutions has become unmanageable.

Centrally managed ABAC enables cross-border data flows, leveraging of new, cost-effective, data hosting solutions, and broader, more open, access to PII in full compliance with GDPR.

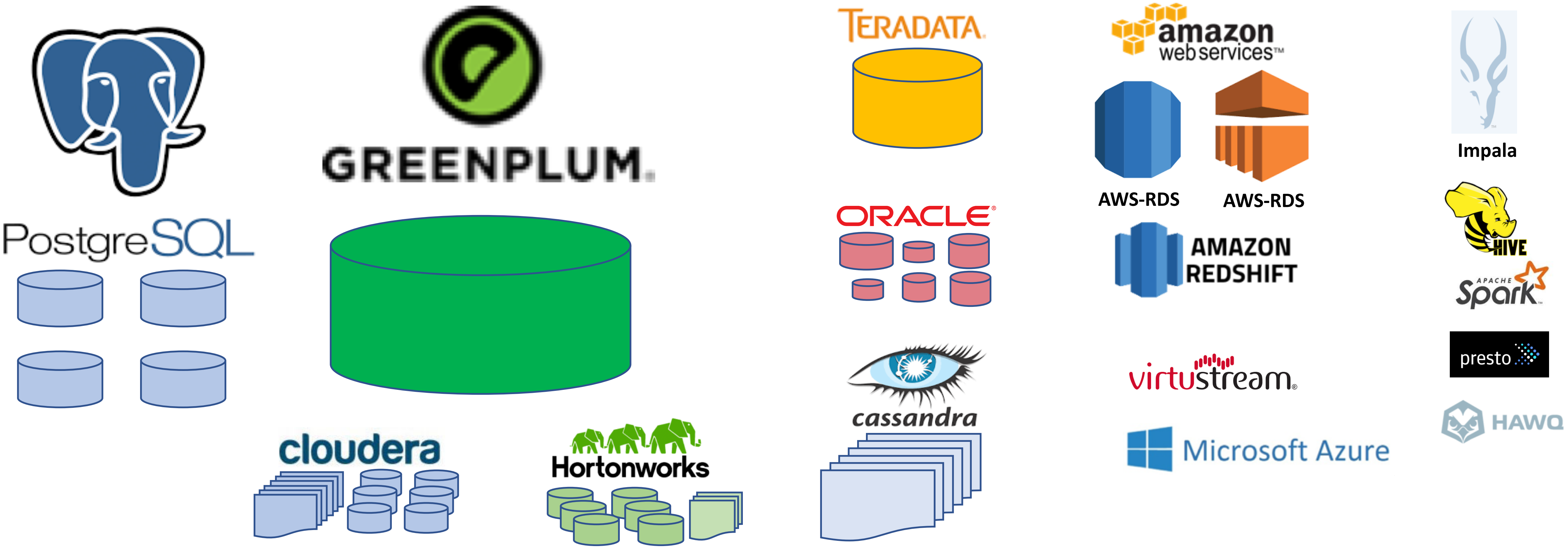
GDPR *WILL* impact any company with European Operations or Customers

Specific GDPR Compliance Checklist and Action Items

Basic steps any organization should follow

Data Discovery

Identify all GDPR regulated Personally Identifiable Information (PII).



Identify each data repository in scope for GDPR compliance and how it is accessed

Evaluate / Define adequate Pseudonymisation and Anonymisation

FName	LName	SSN	Address	City	State	Country	Phone	Email	DOB
Robert	Jones	333-22-9999	123 3rd Ave East	Atlanta	GA	USA	616-829-8091	robert@gmail.com	57/12/29
Carole	Smith	444-33-8888	75 Westfield Drive	Regina	SK	Canada	306-240-2241	csmith@telus.com	60/01/11
Timothy	Dalton	555-44-7777	43 Harris Road	Portland	ME	USA	207-442-1032	td2@twc.com	61/02/14
George	Jacobs	666-55-6666	70 Gartner Road	San Diego	CA	USA	858-694-4920	gjacobs@ibm.com	63/03/15
Alice	Winters	777-66-5555	1600 Penn Ave	Los Angeles	CA	USA	818-925-8888	alicew@hp.com	64/03/17
Cynthia	Smith	888-77-4444	7172 Coach Hill Road	Seattle	WA	USA	206-423-5479	c.smith@uw.edu	67/05/18
Mary	Magden	999-88-3333	83 Evergreen Rise	Calgary	AB	Canada	403-240-7094	marym3@shaw.ca	69/07/21
Michael	Jordon	111-99-2222	1048 26th Street	New York	NY	USA	212-430-1627	airj@knicks.com	72/09/23
John	Roche	222-00-1111	427 Main Street	Denver	CO	USA	303-942-1836	jr@hughes.com	73/11/25
Lynne	Evans	333-11-0000	396 Detroit Lane	Dearborn	MI	USA	810-563-6249	LynneE@ford.com	75/12/26

Evaluate / Define adequate Pseudonymisation and Anonymisation

Which fields must be removed or masked to achieve ***Pseudonymisation . . .***

FName	LName	SSN	Address	City	State	Country	Phone	Email	DOB
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	Atlanta	GA	USA	616-999-9999	XXXXX@gmail.com	57/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	Regina	SK	Canada	306-999-9999	XXXXX@telus.com	60/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	Portland	ME	USA	207-999-9999	XXXXX@TWC.com	61/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	San Diego	CA	USA	858-999-9999	XXXXX@ibm.com	63/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	Los Angeles	CA	USA	818-999-9999	XXXXX@hp.com	64/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	Seattle	WA	USA	206-999-9999	XXXXX@UW.edu	67/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	Calgary	AB	Canada	403-999-9999	XXXXX@shaw.ca	69/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	New York	NY	USA	212-999-9999	XXXXX@Knicks.com	72/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	Denver	CO	USA	303-999-9999	XXXXX@hughes.com	73/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	Dearborn	MI	USA	810-999-9999	XXXXX@ford.com	75/11/11

Evaluate / Define adequate Pseudonymisation and Anonymisation

Which fields must be removed or masked to achieve ***Anonymisation . . .***

FName	LName	SSN	Address	City	State	Country	Phone	Email	DOB
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	GA	USA	616-999-9999	XXXXX@XXXXX.com	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	SK	Canada	306-999-9999	XXXXX@XXXXX.com	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	ME	USA	207-999-9999	XXXXX@XXXXX.com	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	CA	USA	858-999-9999	XXXXX@XXXXX.com	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	CA	USA	818-999-9999	XXXXX@XXXXX.com	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	WA	USA	206-999-9999	XXXXX@XXXXX.edu	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	AB	Canada	403-999-9999	XXXXX@XXXXX.ca	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	NY	USA	212-999-9999	XXXXX@XXXXX.com	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	CO	USA	303-999-9999	XXXXX@XXXXX.com	11/11/11
XXXXXXXX	XXXXXXXXXX	999-99-9999	XXXXXXXXXXXXXXXXXX	XXXXXXXX	MI	USA	810-999-9999	XXXXX@XXXXX.com	11/11/11

Analysis/Scoring should follow defensible generally accepted method or process

GDPR Scope Reduction

Decide which data sets or systems can be taken out of scope

. . . through advance, on ingest, Anonymisation of base table data

All Base Tables in data repository contain only Obfuscated Data

Static masking of PII fields on ingest to keep system out of GDPR scope

NOTE:

Anonymised data has ***NO*** GDPR data privacy requirement

Anonymised data is ***NOT*** subject to "Right-to-be-Forgotten" rule

By definition, Anonymised data cannot be associated with a Data Subject

When GDPR regulated data must be accessed in the clear

- **Data Owner (Data Controller) dictates / authorizes who can view PII in the clear**
- **Data Subjects are notified of collection, use, retention and sharing of their data**
- **Data Subject’s Consent is solicited, collected and tracked**
- **Decide which fields must be dynamically masked at runtime for all other authorized users / applications**
- **Will likely require different rules, based on different User and Data Attributes to decide which fields must be obfuscated or masked to achieve Pseudonymisation or Anonymisation of any given data set containing PII.**

Review & Gain Approval from Data Protection Officer (DPO) for all data access rules

Other GDPR data processing and management best practices

Decide on method(s) for tracking Data Subject preferences
(opt-in, opt-out), right to be forgotten, contact method(s), etc.

Establish single authoritative source as system of record
(Trusted Attribute Domain Source)

Define global policy (rules) for GDPR compliance

Define any additional local system or application policy (rules) for GDPR compliance

Incorporate preference tracking / management in data access rules
This can often require as many fields as the data itself.

Getting Postgres or Greenplum ready for processing regulated PII

Data Discovery. Identify all regulated Personally Identifiable Information (PII).

Whether each platform is in various regulatory compliance scope or not

Evaluate each data set to define adequate level of Pseudonymisation and Anonymisation.

Which fields must be removed or masked to achieve Anonymisation ro Pseudonymisation..

Analysis/Scoring should follow defensible generally accepted method or process.

Decide which data sets or systems can be taken out of scope through Anonymisation of base table data.

Using static masking of enough PII fields on ingest to keep system out of scope for GDPR.

Anonymised data has NO GDPR data privacy requirement.

Decide which fields must be dynamically masked at runtime when Anonymisation is not an option.

Decide on method(s) for tracking Data Subject preferences (opt-in, opt-out) and right to be forgotten.

Single authoritative source or system of record is best.

Define global policy (rules) for GDPR compliance.

Incorporate preference tracking / management.

This can often require as many fields as the data itself.

BlueTalon® & Pivotal Greenplum Partnership

BlueTalon & Greenplum Partnership

Partnership founded on core BlueTalon role in Greemplum Security Ecosystem

- BlueTalon provides Fine-Grained Data Access Control to Greenplum
- BlueTalon Unified Data Access Control Enables Data Mobility

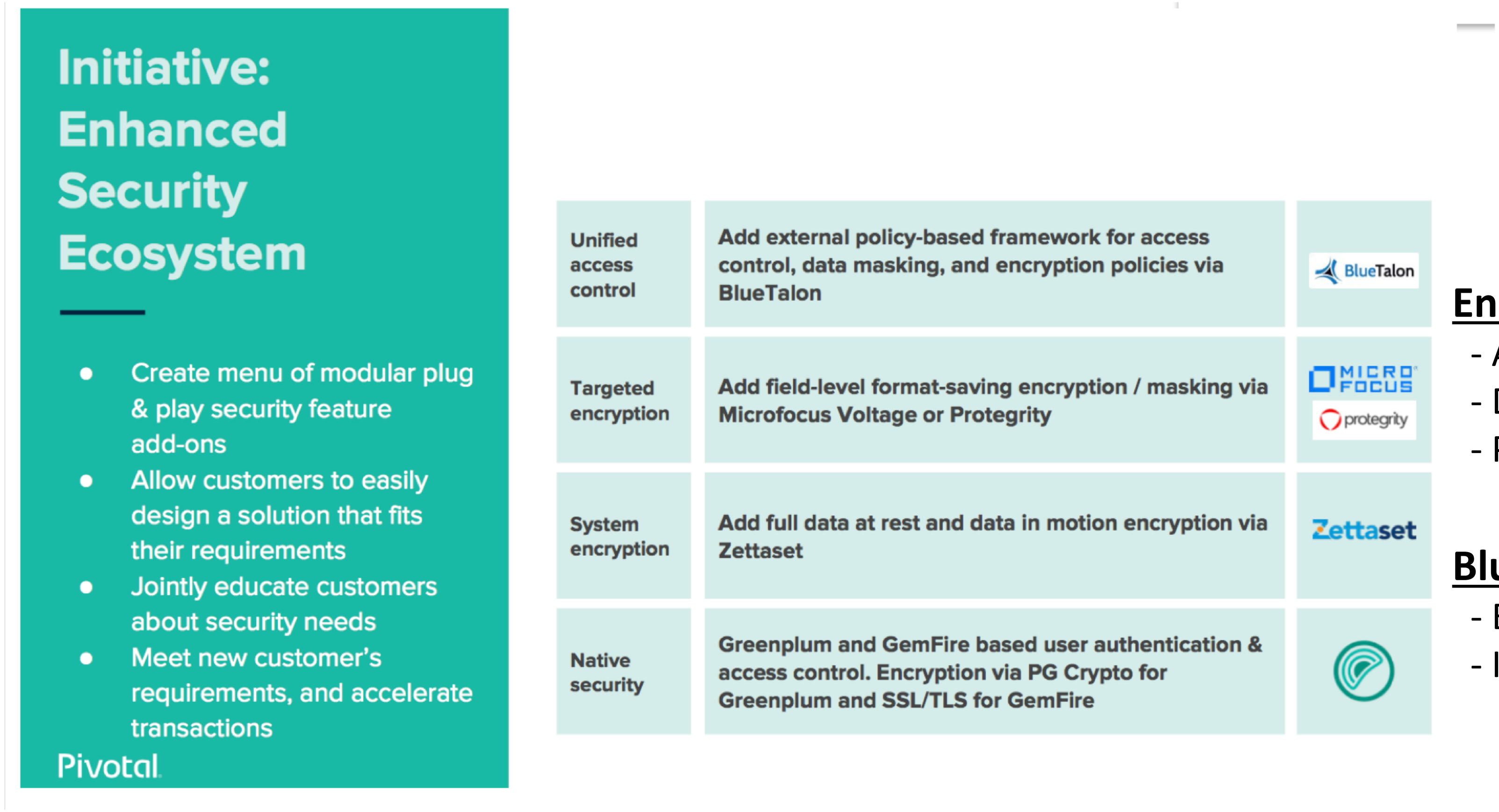
Important Joint Customer Success with large global Manufacturer Finance Datalake

- Leverages BlueTalon for Secure Greenplum
- $Millions saved through BlueTalon Enabling Dynamic Views

Significant Activity with Secure Data Lakes involving Greenplum platform

- Telecoms
- Large US Airport
- International Banks
- Retailers

BlueTalon & Pivotal Partnership: Greenplum Security Play



Initiative: Enhanced Security Ecosystem

- Create menu of modular plug & play security feature add-ons
- Allow customers to easily design a solution that fits their requirements
- Jointly educate customers about security needs
- Meet new customer's requirements, and accelerate transactions

Pivotal

Unified access control	Add external policy-based framework for access control, data masking, and encryption policies via BlueTalon	BlueTalon
Targeted encryption	Add field-level format-saving encryption / masking via Microfocus Voltage or Protegrity	MICRO FOCUS protegrity
System encryption	Add full data at rest and data in motion encryption via Zettaset	Zettaset
Native security	Greenplum and GemFire based user authentication & access control. Encryption via PG Crypto for Greenplum and SSL/TLS for GemFire	

Enhanced Security Ecosystem

- Announced Greenplum Partner Summit (Dec '17)
- Debut in Greenplum Sales Kickoff (Feb'18)
- Presentations at Postgres Conf (April '18)

BlueTalon Complements Security Ecosystem

- Enables Unified Data Access Control (cross platform)
- Interoperation with Encryption Vendors

ABAC vs Data Protection, GDPR and HIPAA Reference Information

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

GDPR Regulation – Final Version
http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Gartner Annual Market Guide for Data Centric Audit and Protection (DCAP) Report – March 2017
https://www.gartner.com/doc/3645326/market-guide-datacentric-audit-protection

BlueTalon Whitepapers, Product Data Sheets and GDPR Lexicon / Glossary

http://bluetalon.com/resources/ . . .

BlueTalon GDPR Compliance Reference Document v8 - FINAL - March 2018.docx
BlueTalon GDPR Lexicon Glossary v1 Feb 2018.docx
HIPAA 18 FIELDS FOR SAFE HARBOR OR EXPERT DETERMINATION.docx
Choosing Between Data Protection and Unified Data Access Control v4.docx

Q & A - Open Discussion

Les McMonagle
VP of Security Strategy
BlueTalon, Inc.
Les@BlueTalon.com
(617) 501-7144 Cell

BlueTalon

Appendix

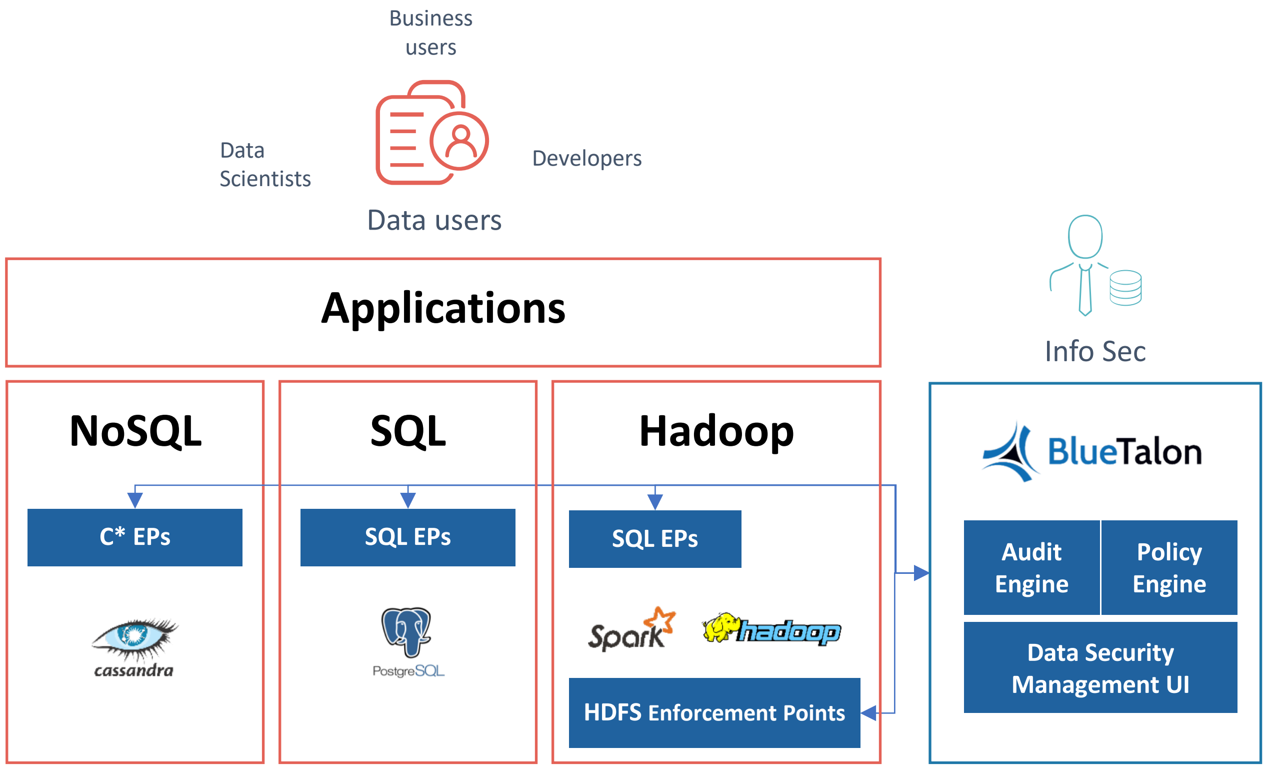
Additional information

BlueTalon® Summary and Business Benefits

Business, Delivery and Technology Integration

Benefits of BlueTalon® approach

- Coordinated Across Platforms (Hadoop, NoSQL & SQL)
  - Meta-data of tables, files and rules are reusable
  - Persistent security for transient Hadoop clusters
  - Centralized, Consistently Applied Access Control
- Transparent Enforcement
  - No Changes to Data, Application or Client Layers
  - Minimal performance overhead for security
- Contextual Auditing
  - Tagged with policy, role and actions
- Granular Authorization and Access Control
  - Granular: file, row, column, cell, sub-cell
  - Decisions based on business data
  - Any User or Data Attributes Govern Access
- Dynamic masking
  - Selective by user or role without duplicating data



Return on Investment (ROI)

Data Security Controls Administration Costs

Duplication of administrative effort on each data repository
View creation and maintenance (100’s or 1000’s of Views)

Eliminate tedious repetitive tasks

Lower security admin costs, lower DBA costs

Managing access rights independently on each platform

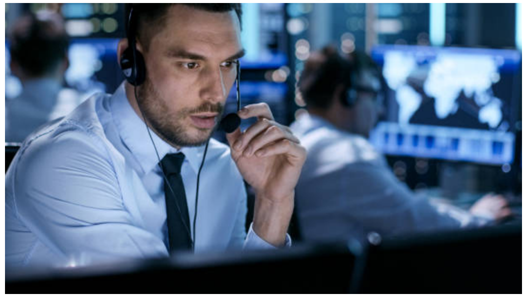
Inconsistently applied rules
Security gaps and oversights
Multiple integration points for centralized User Provisioning and Entitlement Management applications

Manage all access rights once in one location

Consistently applied rules
No security gaps
Single integration point for managing ***all*** data access

Performance Impact of Security Controls

Adding Where Predicate – 30% to 50%
Adding Security Table Join – 50% to 100%
Data encryption and decryption
Activity Log data collection and storage
Excessive data access and retrieval overhead



Positive impact on performance

Add minimal (under 100ms???) query latency only
No changes to client, application or data layers
Often shorter query execution from reduced data volumes

Regulatory and Policy Compliance Audit Costs

Building in data privacy by design work efforts across platforms
SOX, GDPR, HIPAA, ITAR and other compliance audits
Third Party Compliance Audits (SOX, PCI, HIPAA)

Regulatory and Policy Compliance Audit Costs

Privacy by design built in
Simplify audits and data access reporting and accountability

Gartner Agrees: Companies will insist on “centralized management platforms that can directly control data security policies across multiple data silos.”

Centralized management of security eliminates complexity

Gartner
Market
Guides
2017

“By 2020,
data-centric audit and protection products
will replace disparate siloed data security tools
in 40% of large enterprises…” (up from 5% today)

Gartner Market Guide for Data-Centric Audit and Protection, 2017

Gartner
Cool
Vendor
2016
TM

Winner
Audience Favorite
Strata Hadoop 2015

Finalist Start Up of the Year
InfoSec Products Guide’s 2016
Global Excellence Awards

Info Security
Products Guide
2016
GLOBAL
EXCELLENCE
FINALISTS
★★★★★

CYBERSECURITY
Excellence Awards
Data-centric
Security
WINNER
★ 2016 ★

EDITOR’S CHOICE
DATA CENTRIC
SECURITY
SOLUTION 2016
★ CDM ★

Dozen Reasons to Recommend BlueTalon

1. BlueTalon is certified with Cloudera
2. Cloudera + BlueTalon together exceed functionality of Hortonworks + Ranger
3. BlueTalon scales as environment grows
4. BlueTalon works across the Enterprise, not limited to localized implementations
5. BlueTalon can help Cloudera enter into environments with mixed Hadoop and non-Hadoop data sources
6. Land and expand in mixed environments
7. Remove security blockers on deals
8. BlueTalon is simple to deploy, integrates nicely with existing Authentication and Encryption technologies
9. BlueTalon has very low overhead and may actually increase database query processing performance
10. BlueTalon provides complete visibility for auditing data access
11. BlueTalon reduces duplication of data for different use cases and audiences
12. BlueTalon enables dynamic generation of data views based on access policy

BlueTalon®

Customer Case Study: General Electric’s Finance Data Lake

About GE and Finance Data Lake

- GE is the world’s Digital Industrial Company, transforming industry with software-defined machines and solutions that are connected, responsive and predictive.

- Finance Data Lake - Multi-platform environment used by finance teams of different GE businesses for financial reporting
  - Storage and computation in RDBMS (Greenplum) and Hadoop (Hortonworks)
  - Hosted both on-premises and in cloud
  - Land, standardize and roll up reporting on data from 50+ ERP source systems
  - Standardized model contains 10,000+ tables and 150,000+ columns
  - Move data using middleware, ETL (Tableau) and data federation tools
  - Consume data using COTS (Tableau, OBIEE) or custom home-grown apps
  - Data and reports from the lake are consumed by 3,000+ End-Users

Requirements & Challenges

- Access Control Requirements
  - End-User’s finance functional area (e.g. Account Payables) or the business they belong to (e.g. Aviation) determines their access to subset of data
  - data as landed in mirror contains restricted columns and rows that must be excluded from access

- Challenge #1: Duplicated effort from custom views and database grants
  - Required maintaining ~300 views with joins to security tables with ~10m rows
  - Required duplicating views in each data platform with slightly different syntax
  - Required maintaining grants on views for End-User that accessed the data

- Challenge #2: Consistency problem from duplicated security tables
  - Change management required making updates in multiple systems
  - Incomplete updates led to inconsistent access to same data in different platforms

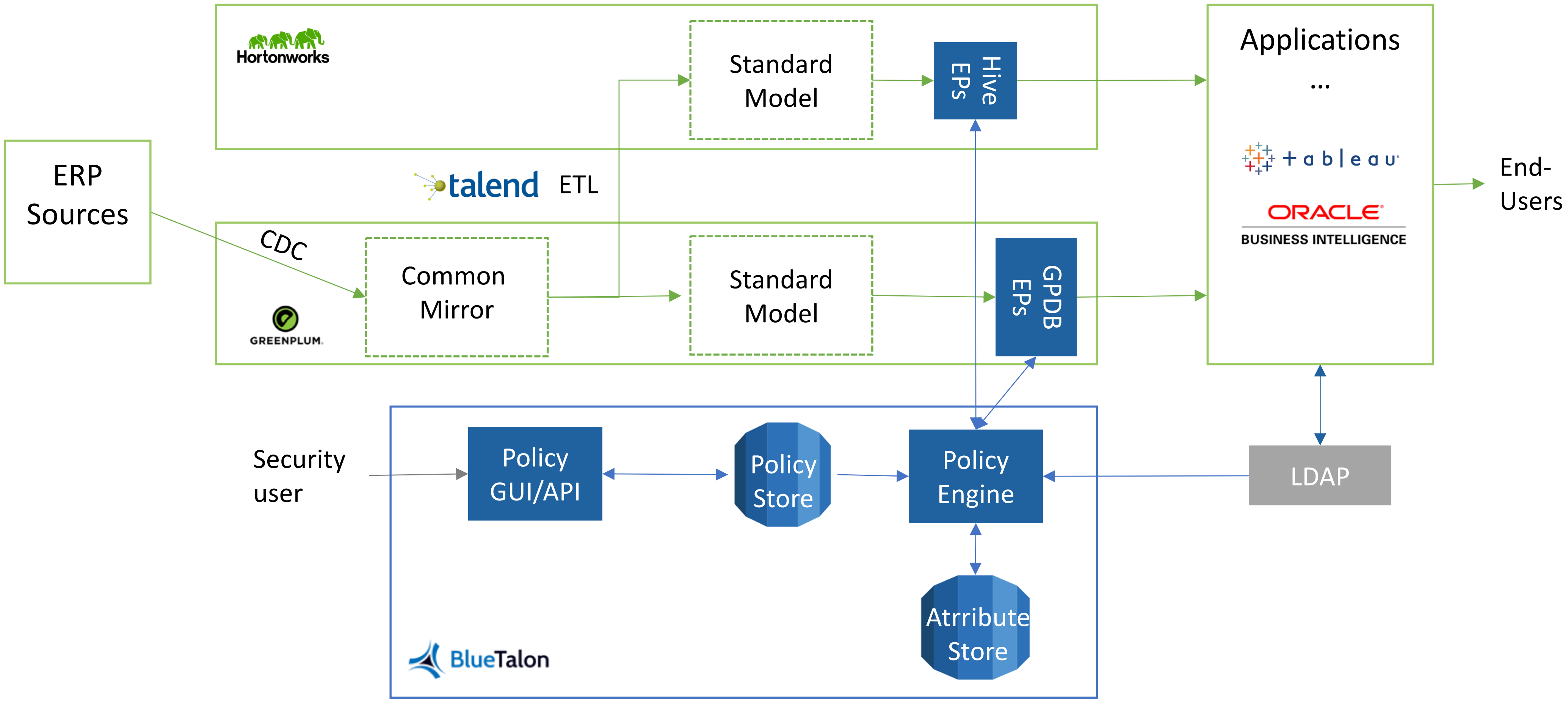
- Challenge #3: Performance problems from joins with large security tables
  - Same queries perform differently due to different join performance on different platforms
  - Caching security tables in applications led to duplicated controls in yet another system

How BlueTalon® Helped

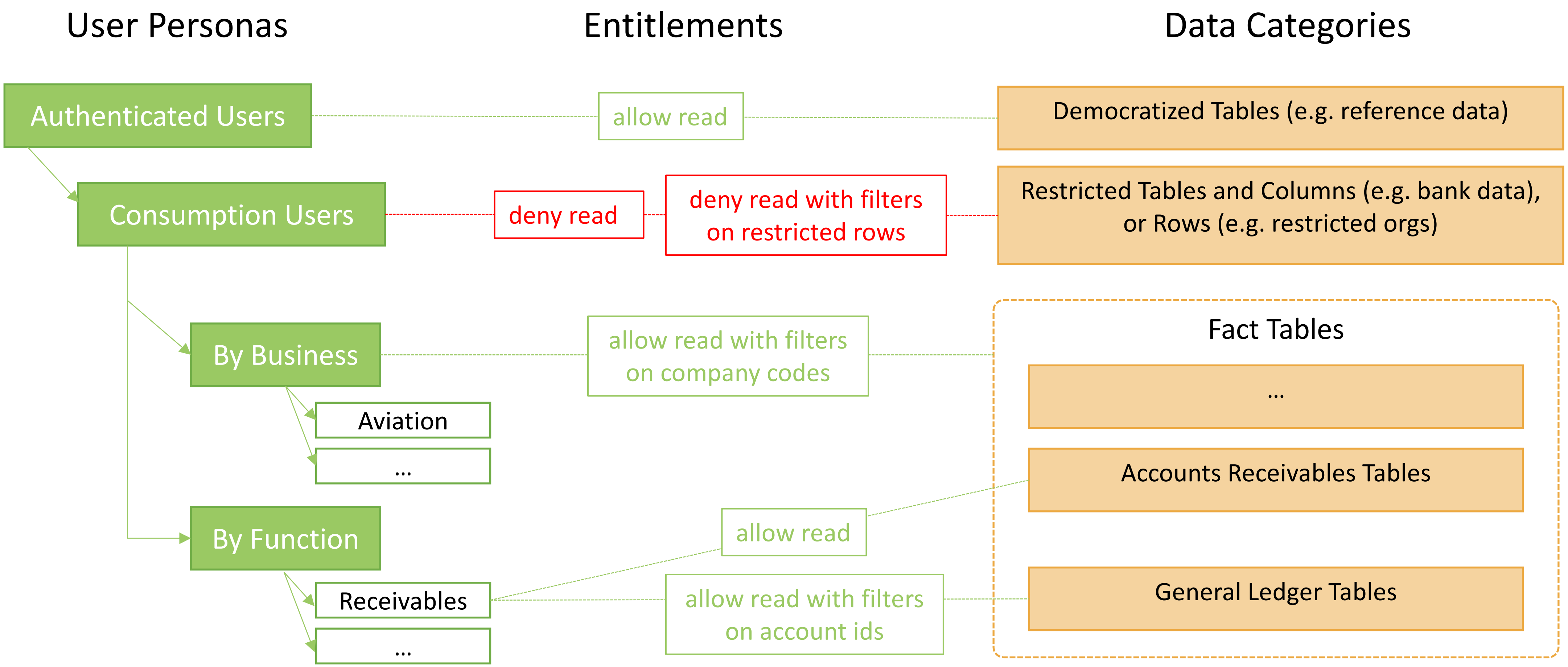
- BlueTalon's solution included
  - Definition of policies in BlueTalon® Policy GUI
  - Enforcement of policies via enforcement points on Greenplum and Hive
  - Dynamic resolution with caching of user attributes by Policy Engine
  - Authentication via Kerberos on Hadoop and OpenLDAP on Greenplum
  - Automation of user to role assignment or table rule creation via REST API

- Benefits
  - Eliminated custom views in different platforms (Hive, Greenplum)
    - Replaced ~300 views with ~20 simpler policies defined in one place, in one way
    - Changes to policies were applied consistently across platforms
  - Eliminated duplication of security tables across platforms
    - Replaced with single source lookup from Policy Engine
    - User attribute lookup & caching eliminated variability from join performance
  - Eliminated the need to add pre-filters in applications

How BlueTalon Helped: Eliminated duplication of security logic



How BlueTalon Helped: Simplified Fine-Grained Access Control


User Personas
Entitlements
Data Categories
Authenticated Users
allow read
Democratized Tables (e.g. reference data)
Consumption Users
deny read
deny read with filters on restricted rows
Restricted Tables and Columns (e.g. bank data), or Rows (e.g. restricted orgs)
By Business
allow read with filters on company codes
Aviation
…
Fact Tables
…
Accounts Receivables Tables
By Function
allow read
Receivables
allow read with filters on account ids
General Ledger Tables
…

BlueTalon