Log Data Privacy

Techniques for data privacy in application logs

Ryan Guest, Software Architect

rguest@salesforce.com, @ryanguest

GENERAL TECHNIQUES

Dictionary Based

• Common Names (Top 100 Census Names)

• Location Identifier (US States)

DOMAIN SPECIFIC TECHNIQUES

Strings unique to your domain

- Custom ID format

  {005…………}

- Custom identifiers

Techniques for Anonymizing Personal Data

{... Ryan ... } ⟹ {... ...}

DATA MASKING

{… Ryan … } ⟹ {… ***** …}

AGGREGATION

{... Ryan ... } ⟹ {Count: 300 }

GENERALIZATION

{... 67.115.155.9 ... } ⟹ {... United States ...}

CATEGORICAL

{... 67.115.155.9 ... } ⟹ {... 50-100 MILES ...}

TOKENIZATION

{... Ryan ... } ⟹ {... Abcd ...}

DIFFERENTIAL PRIVACY

$\{... 2.145 ... \}$ $\Longrightarrow$ $\{... 2.389 ...\}$

ENCRYPTION (WITH ACCESS CONTROLS)

{... Ryan ... } ⟹ {... 0010111 ...}