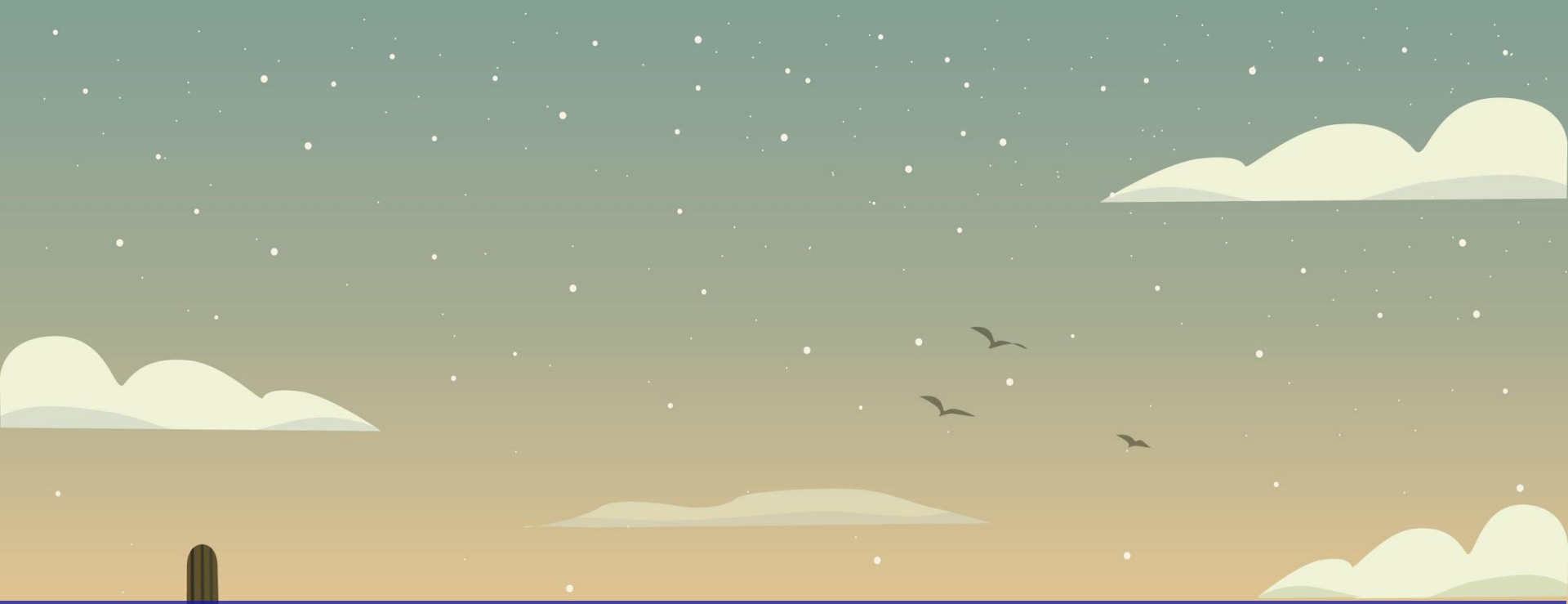


Identity at Scale

Debugging Reality in Salesforce Data 360

Tahsin Zulkarnine - 23 Jan 2026

Identity Resolution at Global Scale

- A global hospitality platform with millions of guest profiles
- Reservations created across web, call centers, and travel agents
- Data 360 selected to unify customer identity
- One morning: customer communication halted. Wrong emails, wrong profiles, wrong people!
- Identity Resolution in Data 360 failing silently for last few day!!!

Tahsin Zulkarnine

Sr. Digital Solution Architect,
NTT Data

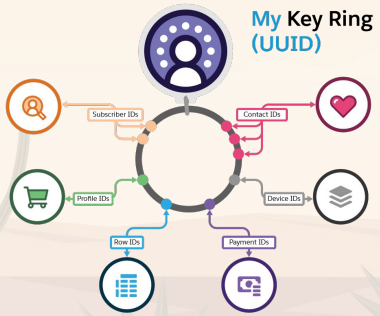
tahsinz@gmail.com
tahsin.zulkarnine@nttdata.com

Agenda

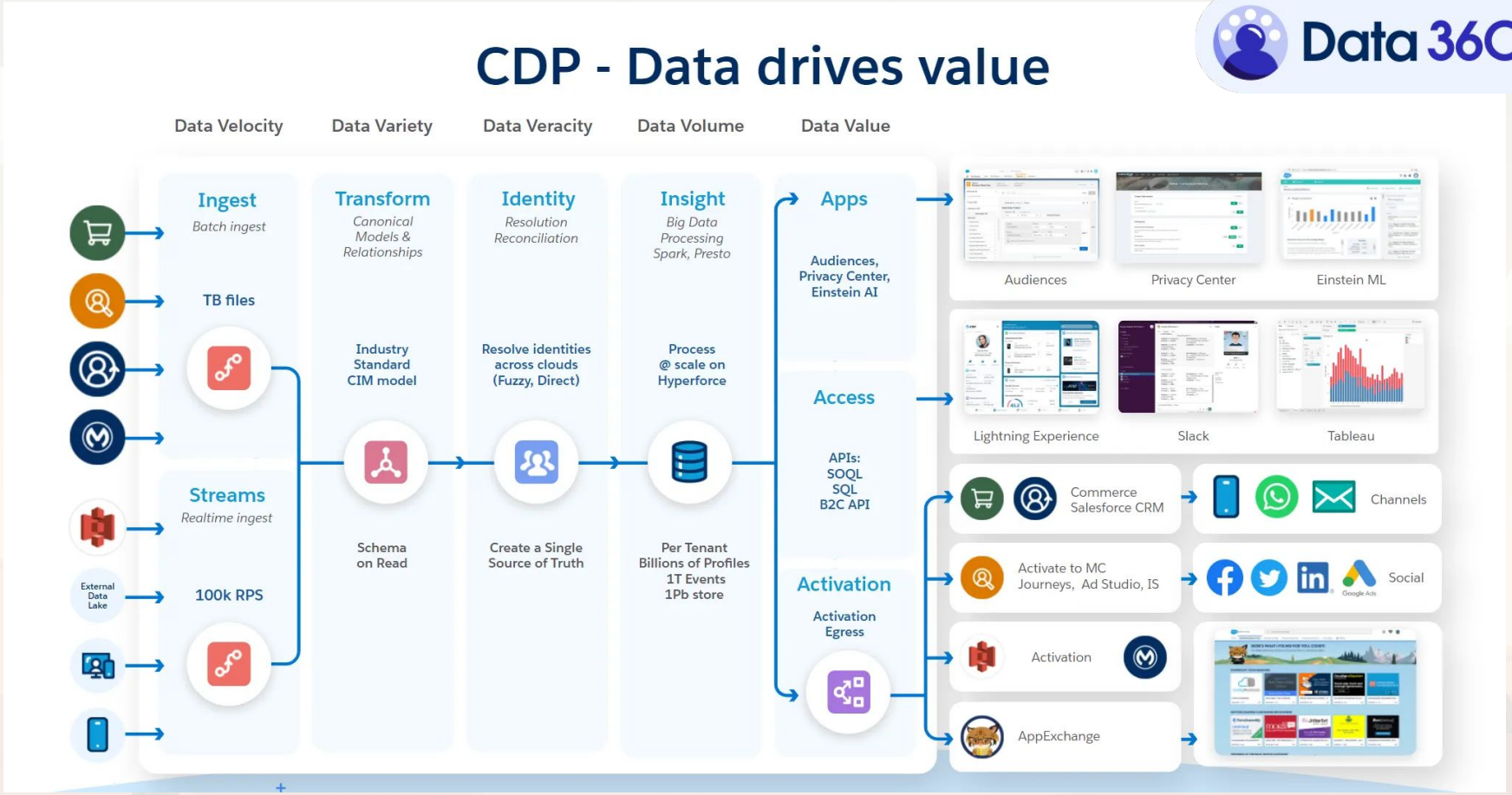
01	/	Overview: Data 360 & Identity Resolution	5′
02	/	A real-world Hospitality Use Case	6′
03	/	When Identity Resolution breaks in Scale	5′
04	/	Failure patterns we discovered	12′
05	/	Containment, Cleanup and Trade-offs	12′
06	/	Key Takeaways	5′
07	/	Q&A	5′

Purpose

- Understand Data Cloud Identity Resolution in practice
- See how silent failures emerge at scale
- Learn real failure patterns from production systems
- Walk away with prevention & recovery strategies



Overview: Salesforce Data 360



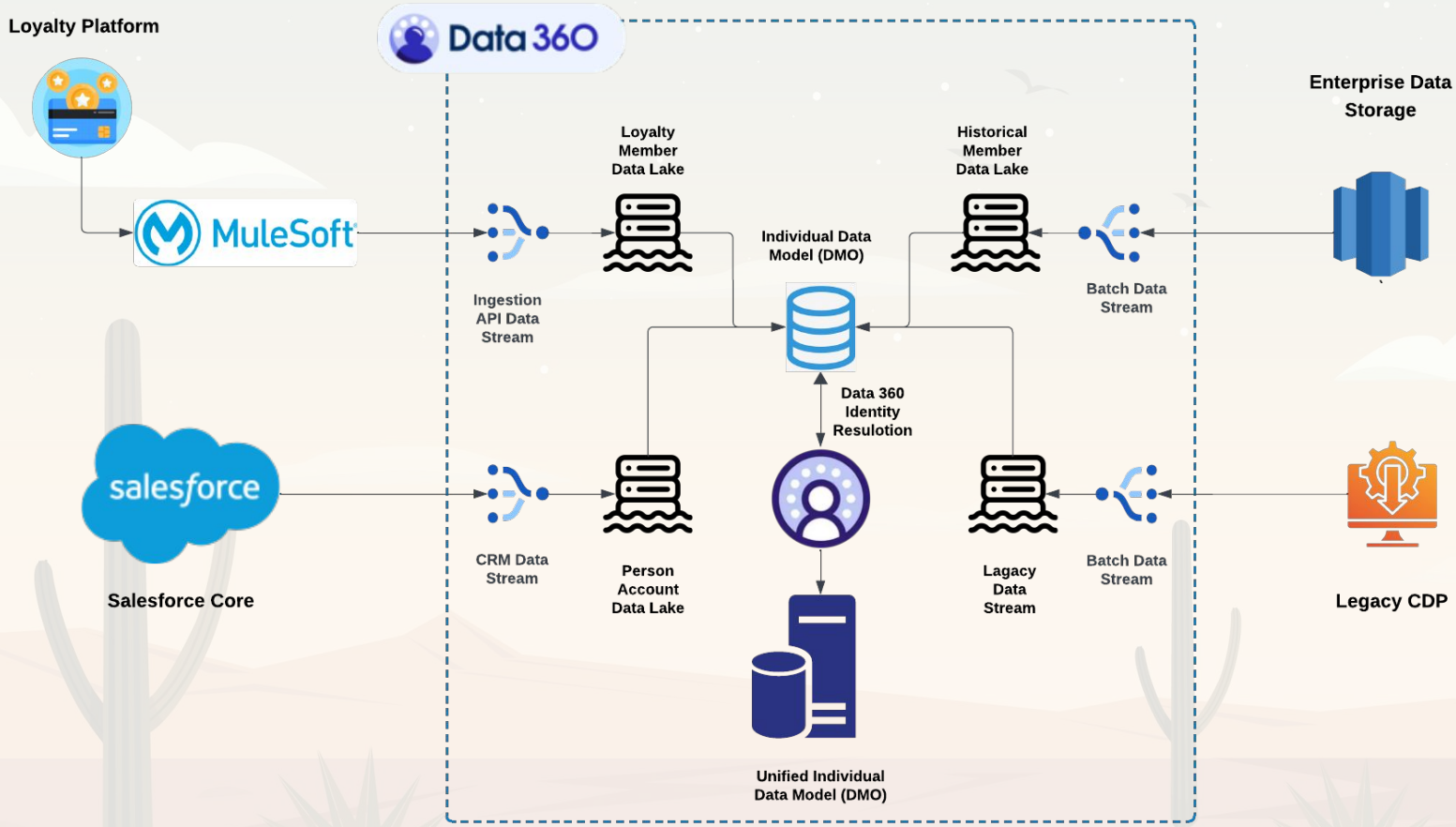
Why Identity Resolution Exists

Consumer Identity Resolution

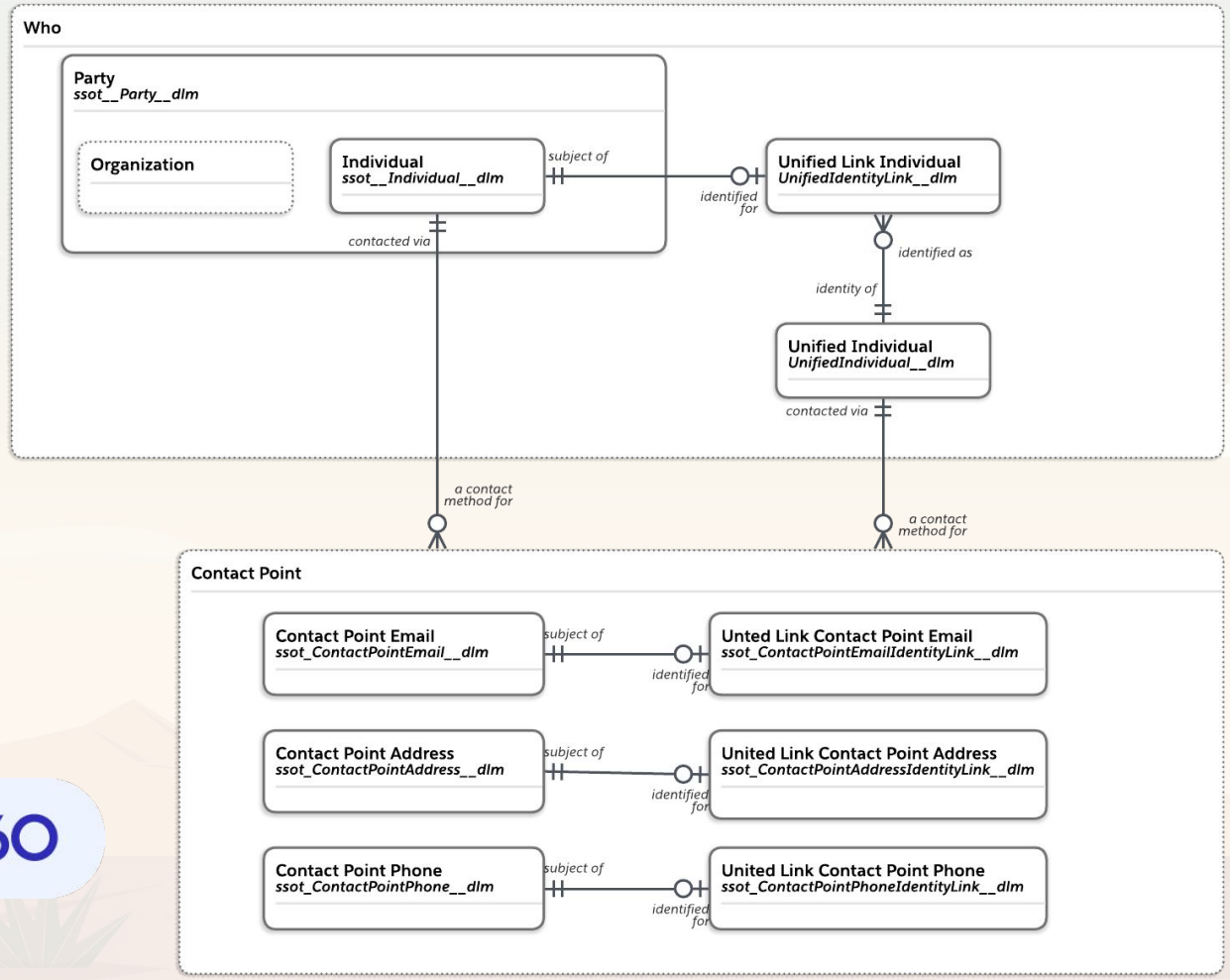
Mary Smith at Home	
Individual ID:	123123123123
Household ID:	456789101112
First Name:	Mary
Middle Name:	A
Last Name:	Smith
Age:	41
Date of Birth:	5/1/1979
Home Address:	2345 Ave C Papillion, NE 68046
Family:	4 Adults, 2 Children Married
Email(s):	Mary.Smith@gmail.com MSmith123@yahoo.com
Cell Phone:	(111) 123 – 4567
Registered Voter:	Y
Interests:	Shopping Cooking Traveling Reading Science

Mary Smith at Work	
Contact Title:	Vice President
Contact Email:	Mary.Smith@KarensFireside.com
Contact Gender:	Female
IG Business ID:	696115559
Contact ID:	123667894123
Business Name:	Karen's Fireside Inc.
Address:	1214 N Monroe St Papillion, NE 68046-20146
Phone:	(402) 592-4413
Hours:	Daily 9am -5pm
Website:	http://www.karensfireside.com
Social Media:	https://www.facebook.com/karens-fireside-event-venue-1774143466165459
SICs:	Banquet Rooms, Halls & Auditoriums, Wedding & Reception Sites, etc.
Employee Size:	9
Location Sales Volume:	$407,000

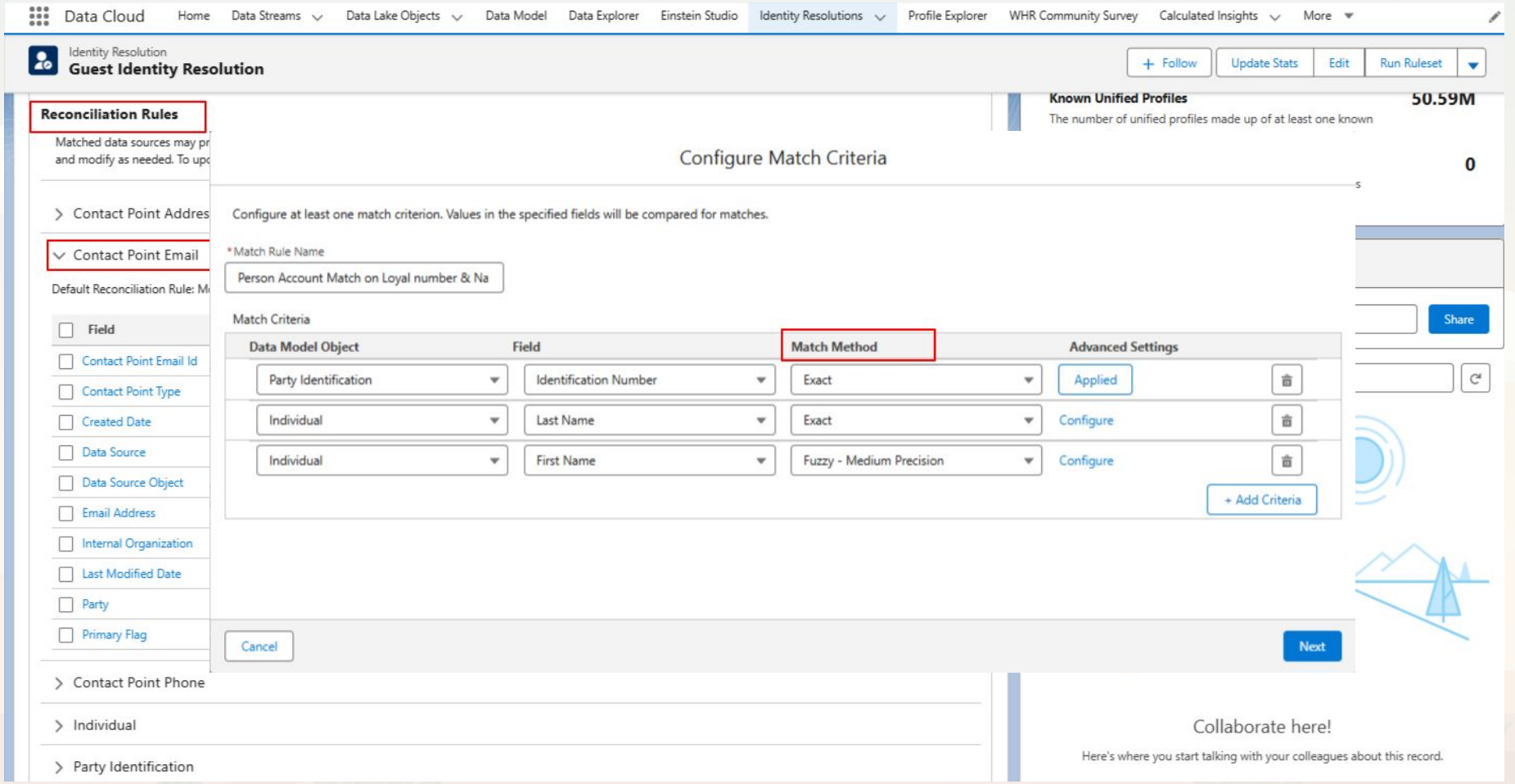
Use Case: Global Hospitality Brand



Data 360 - Underlying Data Model



Identity Resolution Configuration



When Identity Resolution Breaks at Scale

Observation

- Identity Resolution job failed.
- Unified profiles not generated for a subset of data
- Limited diagnostic details surfaced
- No automated alert triggered
- Salesforce Support Case Opened

50K Duplicate Cluster Limit

- Extremely large duplicate clusters formed
- Duplicate keys expanded to 50K+ Individuals
- Resolution stalled or failed for affected clusters
- Downstream systems stopped receiving updated identities

Data Analysis

- Large-scale ingestion of translated Non-Latin source data
- High overlap in matching keys
- Many Individuals sharing identical identity attributes

Detecting Duplicate Clusters

Identity Entropy

- a way to describe how much distinguishing information exists in your identity data.
  - High identity entropy: records are rich and distinct.
  - Low identity entropy: records look similar

Steps

- Start with Individuals, not Unified Profiles
- Normalize Contact Points
- Combine Name + Contact as Match Keys
- Look for Unusually Large Groups
- Large duplicate clusters are a data-shape problem, not a configuration problem.

```
WITH individuals AS (
  SELECT ssot__Id__c, ssot__LastName__c
  FROM ssot__Individual__dlm
),
contact_keys AS (
  SELECT
    i.ssot__Id__c, i.ssot__LastName__c,
    LOWER(e.ssot__EmailAddress__c)  AS email_norm,
     p.ssot__FormattedE164PhoneNumber__c AS phone_norm
  FROM individuals i
  LEFT JOIN ssot__ContactPointEmail__dlm e
  LEFT JOIN ssot__ContactPointPhone__dlm p
),
duplicate_keys AS (
  SELECT
    ssot__LastName__c,
    COALESCE(email_norm, phone_norm) AS match_key,
    COUNT(DISTINCT ssot__Id__c) AS cluster_size
  FROM contact_keys
  GROUP BY ssot__LastName__c, match_key
  HAVING COUNT(*) > 1000
)
```

Failure Pattern #1 - Regional Translation

Machine-Translated Regional Data

Use Case: Original source records are translated and normalized before ingestion into Data 360.

Name (Original / Native)	Name (Translated)	Phone (Normalized)
李伟	Li Wei	+86-000000
李 威	Li Wei	+86-000000

Issue Observed

- Distinct native records collapsed into identical translated values
- Normalization removed remaining differentiators
- Identity Resolution formed extreme duplicate clusters.

Diagnosis

- Translation reduced linguistic variance
- Normalization standardized contact points
- OR-based matching expanded clusters transitively
- Cluster size exceeded Identity Resolution safeguards

Failure Pattern #2 - Matching Rule Inconsistency

Incorrect Profile Returned in Service Lookup

Call-center agent searches by phone number.

Name	Phone	Email	Loyalty ID
Alex Chan	+1-5555-1111	alex.c@example.com	L-10021
Brian Chan	+1-5555-1111	brian.c@example.com	L-10098

Issue Observed

- Multiple Individuals unified under one identity
- Incorrect profile surfaced to agent

Diagnosis & Fix

- Shared phone number
- Fuzzy name (Medium Precision) treated “Alex Chen” and “Brian Chen” as similar
- Fizzy precision updated to High Precision
- Still Merging other profile with same last name

Failure Pattern #3 - Identity Propagation

Payment Method Cross-Attached

Use Case: Call-center booking with stored payment.

Name	Phone	Payment Token
Daniel Park	+1-555-2222	CARD-A-****4567
David Park	+1-555-2222	CARD-B-****6753

Issue Observed

- Payment method associated with incorrect Individual

Diagnosis and Fix

- Shared phone number and Fuzzy name (Medium Precision) unified profiles
- Transitive expansion collapsed household members.
- Once unified, reconciliation logic selected one payment context

Failure Pattern #4 - Query Model Mismatch

Search Returns Too Many Profiles

Agent searches by email.

Email	Profiles Returned
aviantravelagent@example.com	7 profiles

Issue Observed

- Multiple profiles returned for a single query
- Manual selection required by agent

Diagnosis

- Shared emails reused across bookings
- Identity Resolution preserved contact points correctly
- Retrieval logic assumed uniqueness
- Identity may be correct; query assumptions are not

Failure Pattern #5 - Data Governance Violation

Wrong Loyalty Identity Associated

Use Case: Agent searches using loyalty identifier.

Name	Email	Loyalty ID
Maria Lopez	maria.l@example.com	L-20344
Marta Lopez	marta.l@example.com	L-20344

Issue Observed

- Same loyalty ID associated with multiple Individuals
- Reduced confidence in identity accuracy

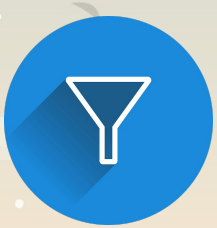
Diagnosis

- Loyalty ID reused across historical records
- System assumed loyalty ID = unique person
- This is a data governance issue, not a matching issue

Containment Strategies

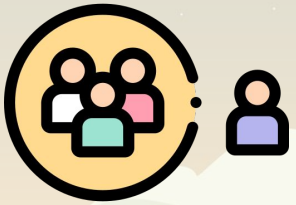
Data Spaces

Isolate High-risk Data



Filters & Rules

Stop know bad patterns

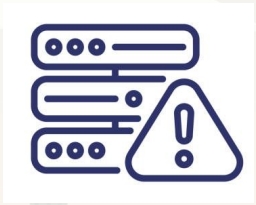


Source Isolation

Separate Problematic Systems

Data Isolation Isn't Enough

Downstream Systems Already Polluted



- Incorrect identities already activated
- Data persisted in downstream systems
- Referenced beyond Data 360

Cleanup Complexity

- No OOB Delete Feature
- Requires sequencing and safeguards
- Data Consumption
- Risk of recontamination

Cross-Team Coordination

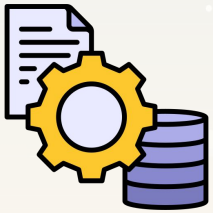
- Multiple teams impacted
- Shared ownership of cleanup decisions
- Alignment on timing and scope
- Operational risk if misaligned

Options 1-4: Salesforce Native Cleanup

Manual Export + File Deletes

- Manual effort.
- High cost
- 600K Limit per CSV file

Apex Batch Delete



- Automated.
- Can be handled within daily governor limits.
- API Usage Overhead

Apex + File Hybrid

- Automated.
- Multiple System Dependency

CI + Segment Export

- Data 360 native
- No Apex or APIs.
- Error handling requires effort.
- Data must be activated

Options 5-8: Advanced Cleanup Options

Query & Stream

- Query enterprise storage
- Prepare stream chunks to Buckets.
- Manual Effort

ETL-Driven Cleanup

- Use scheduled data jobs (i.e Appflow)
- Partition and push deletion files into Cloud Bucket for high-volume removal.

Serverless Pipeline

- Integration flows call serverless functions for deletes.
- Timeout consideration for large data

Third Party

Data360 Cleanser

Data Cloud Cleanup Utility

data360cleanser.tiemein.com

- Built-in Cleanup Configuration
- Archival & Audit Trail for Deleted Data
- Flexible SOQL or SQL-Driven Deletion Logic

Identity Debugging is Different

Traditional Debugging	Identity Debugging
Errors surface where they occur	Errors surface downstream
Failures are deterministic	Failures are probabilistic
Logs explain what broke	Logs show “success” or “failed” only
One bad record causes failure	Millions of valid records interact
Reproduces easily in lower envs	Often only appears in production
Fixing rules fixes behavior	Data distribution dominates outcomes

Key Insight

Identity bugs are emergent behaviors, not execution errors.

Key Takeaways

- Identity Resolution breaks at 50K duplicate cluster
- Silent failure is the most dangerous failure mode
- Precision tuning has hard limits and real trade-offs
- Identity failures compound once data leaves Data 360
- Cleanup is an architectural problem, not a feature checkbox

Thank you!

Questions?