UNLEASH THE POWER OF LIMITLESS CONNECTIVITY
Operational Transformation

Helm: Self-Service Customer Data Platform

Sriharsha Gangam
Principal Architect
Comcast Cable
Agenda

- Why Customer Data Platforms (CDPs)?
- Helm Introduction – A CDP at Comcast
- Design Goals and Architecture Overview
- Self-Service for Managing Helm
- Applications Aiding Customer Experiences
- Conclusions: Helm Journey and Evolution
Why Customer Data Platforms?

Customer Experiences

• Critical for Businesses
  • Increasingly Data driven – serve millions of customers
  • Need to be Holistic – empathize customers

• Common Data Driven Challenges
  • Disparate sources (esp. large organizations)
  • Scaling: variety, volume, and churn
    • One-off ETL (extract, transform, load) data pipelines
    • Governance and domain expertise

Customer Data Platform (CDP)

Unify, Enrich & Store

Data Collection from disparate sources

Data Activation

Customer Interactions

© 2021 SCTE®, CableLabs & NCTA. All rights reserved. | expo.scte.org
Background

• Built to understand customer journeys across all products and channels
  • Started as a Proof of Concept/Lab week project in 2012
• Common Use-Cases & Applications
  • Customer Support - Care Agents
  • Product Rollouts - Data Analysts, Business owners
  • NPS callback Program – Employees
• Self-Serve: Decentralized Data Management & Governance (Publish & Access)

Current Scale

• Billions of Monthly Interactions
  • Millions of customers
• Hundreds of Dataset types
  • Multiple domains
• Thousands of Employees (users)
  • Billions of monthly API calls

Comcast collects, stores, and uses all data in accordance with our privacy disclosures to users and applicable laws
Helm CDP

Building Blocks

- Helm CDP
  - Ingest Foundations
  - Data Activation Applications
  - Self-Service

- Integrations
  - Success
  - Operations
  - Data
Design Goals and Architecture Overview

Design Goals

- Metadata Driven
  - Domain Agnostic & Multi-Tenant Capable
  - Self-service data onboarding

- Democratized Observability
  - Quality, Availability, Notifications

- Near Real-time Streaming with High-Fidelity

- Cloud Infrastructure & Performance Isolation
  - Privacy and security by design

Example Helm Data Model
Design Goals and Architecture Overview

**Ingestion**
- Batch Sources
- Real-time Sources
- Streaming Sources
- Ingestion Adapters: Batch (AWS Lambda, EMR), HTTP's Gateway (AWS Beanstalk), Kinesis/ Kafka Bridge (Flink)
- Staging area (AWS S3), Staging area (AWS S3)
- Ingestion Library

**Routing**
- Event Bus (AWS Kinesis)
- Routing (AWS Lambda)
- Batch, Aggregate, Sessionize, Evaluate Rules

**Transform/Persist**
- Enrich, Store
- Helm Data Stores: Profile Search DB (ElasticSearch), Document Store (DynamoDB), Profile Graph DB (AWS Neptune), Historical Store/Data Lake (AWS S3)

**Services/APIs**
- Data Activation (API/UI): Profile Search & Navigation, 360° Visualization, Journeys & Actions, Watchlist, Talk, External Services
- BI Analysts, Employees & Care-Agents

**Self-Service**
- Self-service UI: HTTPs API & UI (AWS Beanstalk)
- Orchestration Workflows (AWS Step Functions): Ingestion, Infrastructure, Application Specific orch., Access Control, Operations & Observability
- High-performance Metadata store: Metadata Store for Data Plane (AWS DynamoDB)
- Metadata Integrations: collibra, Amazon Glue
Metadata Configurations

- Data Source Metadata
  - Credentials, Schema, Documentation
  - Preparation: Normalization, Encryption
- Routing, Persistence & Data Access
  - Retention and data lifecycle
  - Access controls: ABAC/RBAC
- Application Specific
  - Visualization, Sessionize, Batch
  - Rules, Aggregate, Enrich
Data Source & Transformation Metadata - Tags

- Incoming feed schema
  - Field types (JSON Schema)
- Profile type
  - Profile id and timestamp fields
  - JSON Paths
- Encryption Security
  - JSON Paths
- Field Descriptions
- Labels and Icons
- UI/Visualization
Helm Applications

- Complementary & Data Re-use
- 360° Visualization
  - Lifetime view of customer events
  - Flagship Application
Applications Aiding Customer Experiences

360° Visualization

- Chronological, “Lifetime” customer view
- Range Selector (black)
- Swim lanes per category
- Filters to hide irrelevant events
- Backed by Helm Events API
Other Helm Applications

- Talk
  - NPS callbacks
  - Customer Conversations
- Journeys & Actions
  - State machines & notifications
- Analytics
  - Metrics & distributions
Platform Evolution

- Past: Bespoke Pipelines for 360° Visualizations
  - Hard to scale and manage
  - Self-Service capabilities

- Current/Near-Future: Self-Service for Data Onboarding/Application Metadata
  - Increased complexity of Self-Service
  - Decouple metadata orchestrations

- Future: Multi-Tenancy Support
  - Current approach: New Helm instance per Tenant
Conclusions

• Helm CDP and its role at Comcast
  • Design principles, architecture overview and Journey

• Self-Serve Governance & Data Management
  • Democratized, decentralized
  • Trade-off: Freq. Configuration changes Vs. Effort to build self-serve capabilities

• Modular Applications (e.g., 360° Visualization, Analytics)
  • Complementary and re-use data
Thank You!

Sriharsha Gangam
Principal Architect,
Comcast Cable,
1800 Arch St, Philadelphia, PA 19103