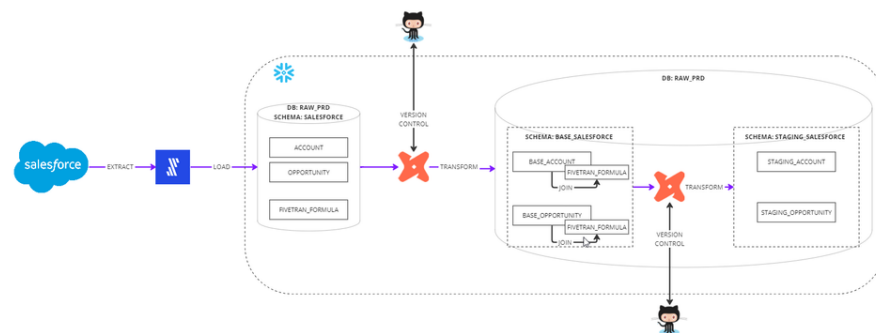# Salesforce

Salesforce Data Source

Description

Salesforce CRM (Customer Relationship Management) is a cloud-based platform designed to help Dynatrace manage our interactions with customers and prospects. Salesforce CRM can be particularly beneficial in various aspects such as lead, opportunity, quote, subscription, usage, consumption and much more.

Data Pipeline Architecture

Integration Overview

Multiple Salesforce objects are extracted and loaded every hour to our Data Hub. To provide complete and accurate data, Salesforce data transits through multiple layers, with some transformations. The following diagram highlights the transformation steps from Salesforce to the different layers in Snowflake:

Link to Miro

Components

- **Fivetran**: Extract and Load data from salesforce to snowflake using the native connector.
- **DBT Cloud**: Orchestrate the transformation from the raw layers up to the mart layer, leveraging Fivetran Salesforce Formula Utils DBT Package.
- **Git / GitHub:** DBT leverage Git and GitHub directly to version control our code.

Workflow

1. Salesforce to Snowflake Raw Database: replicates every 60 minutes with distinct table for formula.
2. Transformation from Snowflake Raw Database to Prod Database, Base Schema: Join raw table with formulas.
3. Transformation from Snowflake Prod Database, Base Schema to Prod Database, Staging Schema: Rename API column name with Label Column name.

Data Processing

Salesforce to Raw

Every hour, data is extracted from Salesforce and loaded into the Snowflake RAW database using Fivetran. Notably, the raw tables exclude formulas. Instead, Fivetran maintains a distinct table, named FIVETRAN_FORMULA. This table has one column for the original Salesforce formula and another for its corresponding SQL equivalent. To guarantee that the BASE and STG views provide the most recent data - even when querying historical information - Fivetran continuously updates the formula values in the FIVETRAN_FORMULA table with the latest versions.

> ℹ️ Only the columns selected in Fivetran will be replicated in Snowflake. If a column is missing, reach out the the data engineering team.

Transform from Raw to Base

Every day, a DBT job is run to create or update views that join the Salesforce raw metadata with the formulas metadata. This results in a base view that includes all Salesforce columns, which is then stored in the STAGING_BASE_SALESFORCE schema. The column name will match the API name, not the label name, of the field in Salesforce

If a formula is updated, all historical data will be recalculated using the new formula. This ensures that the data is always consistent with Salesforce.

Transform from Base to Staging

Every day, a DBT job is run to create or update views with light transformations. Light transformations can be replacing the API name of a column with the label name, replacing a timestamp column name with a name suffixed with _at for naming consistency or transforming a Boolean value into a text value with 'yes' or 'no'. This results in staging views which are stored in the STAGING_SALESFORCE schema.

Data Governance and Security

Data Privacy

- Salesforce data is handled in compliance with data privacy regulations, with any sensitive information anonymized or encrypted.
- **Personally Identifiable Information (PII) Hashing:** To further enhance the security and privacy of PII data sourced from Salesforce, we utilize Snowflake tagging and masking policy to hash PII data fields. Hashing irreversibly transforms PII data into cryptographic representations, preserving data privacy while permitting analysis. Access to the hashed PII data is tightly controlled and restricted to authorized personnel only.

Access Control

- Access to Salesforce data and related integration components is restricted to authorized personnel.
- The contacts provided below pertain to the data ingested into the Business Systems Data Hub.

Type	Scope	Primary Contact	Secondary Contact
Data Owner	Makes decisions about the data's permissible use, its classification, and its criticality.		
Data Steward	Ensures that data governance policies are implemented and followed. Understands the business context and use of the data and bridge the gap between IT and business units.		
Data Custodian	Handles the technical aspects of storing, securing, and maintaining data, ensuring that the data is available, reliable, and secure.	@Rutvij Sharma	@Thierry Krumeich

Compliance

Our Salesforce data integration complies with relevant data protection regulations and industry standards.

Deployment and Scaling

Deployment Strategy

- Salesforce is deployed within Dynatrace existing data pipeline infrastructure.

Scalability

- The integration is designed to scale horizontally and vertically to accommodate increased data volumes as needed.

Documentation and Maintenance

Versioning

- Version Control Mechanism: Our Salesforce integration captures version changes using DBT / GitHub as the version control system. GitHub allows us to track changes and maintain a history of modifications made to the integration scripts and configurations.

Change Log

- Change Log Documentation: The change log for the Salesforce integration is maintained using Jira tickets. Each change, enhancement, or bug fix is associated with a Jira ticket, which serves as a reference point for tracking and documenting changes. This ensures transparency and accountability in our change management process.

Runbook

- A runbook is available for troubleshooting common issues and performing routine maintenance tasks related to Salesforce data integration.