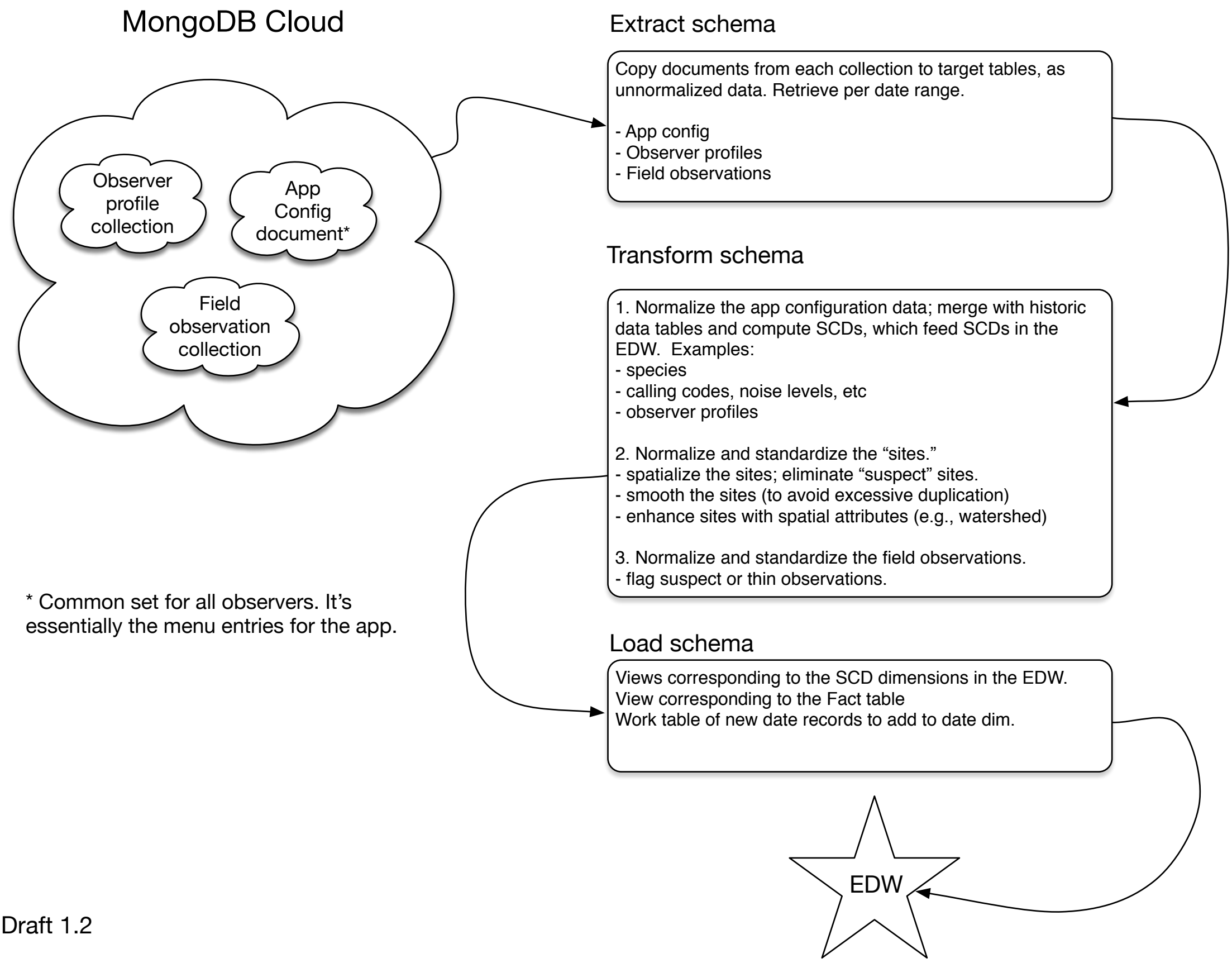
ETL Overview



* Common set for all observers. It’s essentially the menu entries for the app.

Draft 1.2

Data transformations, in PostgreSQL

1. App Data Synchronization

The Cloud collection of App Data contains one document per value picker (e.g., calling codes, wind scale, land use type, etc.). It's the current, official dataset, used by the application. For the ETL and EDW, it's been "downloaded" to the Extract schema in PostgreSQL as an unnormalized table. The "previous copy" is in the Transform schema. Sync it with the Cloud copy just downloaded to the Extract schema.

Extract Schema

Unnormalized Dataset
Collection name
Key
Value Description

Transform Schema

Species
Key
Family
Genus
Species
Common Name
Date Added

(EDW is Type 3 SCD)

Calling Codes
Code
Description
Date Added

(Type 2 SCD)

Land Use
Code
Description
Date Added

(Type 2 SCD)

Wind Scale
Code
Description
Date Added

(Type 2 SCD)

etc.

2. Observer Profile Synchronization

The cloud collection contains one document per unique observer (named or anonymous). Each observer document also contains a sub-document of the geographic sites that have been observed by this observer.

Extract Schema

Observer Documents
Observer Unique ID
Affiliation
Contact data if provided
Date Added / Updated

Site Sub-Documents
Observer Unique ID
Site Unique ID Map coodinates Land use Date Added / Updated
Habitat Sub-document Water level Date/Time Added/Updated

Transform Schema

Observer Documents
Observer Unique ID
Affiliation
Date Added / Updated

(EDW is a Type 3 SCD.
Affiliation only.)

Site per observer
Observer Unique ID (foreign key)
Site Unique ID
Map Coordinates
Land Use code
Water shed ID
Geo Admin ID (s)
Nearest Weather Station ID
Date Added / Updated

Habitats within Observer Site
Site Unique ID
Habitat label
Water Level
Date Added / Updated

(In EDW these would be a Type 4 SCD that combines the sites and habitats)

Elements in red require spatial processing

3. Observations Synchronization

Extract Schema

Observation Documents
Observer…
Site…
Species
Calling Code
Weather conditions…
tba

This section obviously is incomplete. Several upsert/ merge operations need to be spec'd.

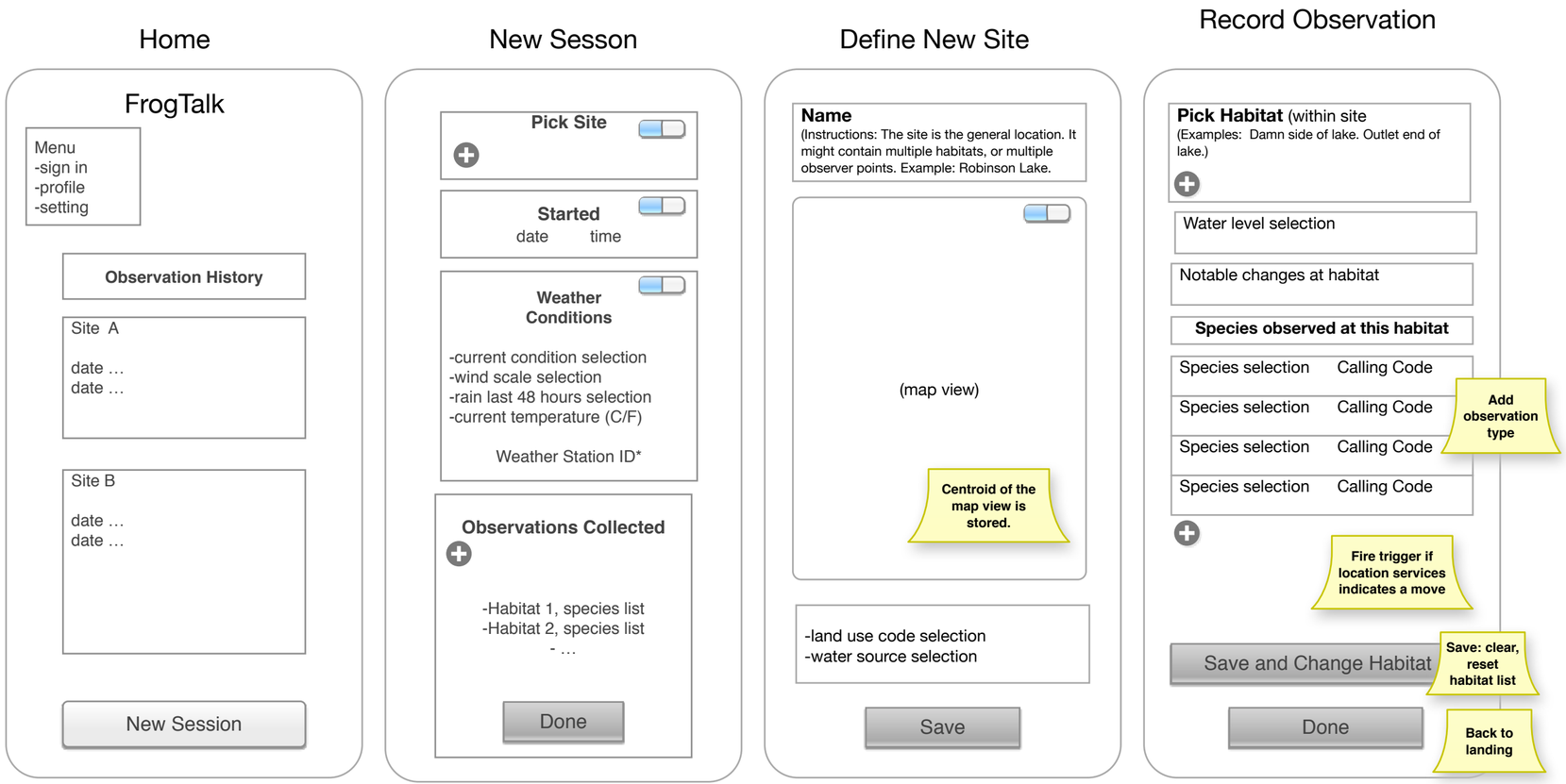
Transform Schema

Observation Records
Observer…
tba
Species
Calling Code
tba
tba
Date Added / Updated

(This table becomes the basis of the fact table.)

App Overview

(Next step. Enhance to standard iOS 9 symbology.)



* might move this to ETL processing.