Creating a Centralized Database for Interconnection Services

Eric Silverman
VP of Engineering
eric@inflect.com

What is Inflect?

Commercial product for buying & selling colo & network services

Free online tool for colo and network data

Fully featured

No license fees – monetize on transactions

But let’s talk about why we are here…

Data Driven Decision making

Independent Research

Institutional Knowledge

Current Data Sources – PeeringDB, IX Reports, On Net Lists

What’s Great About PeeringDB

Free

Community based

Service Provider Neutral

Most accurate peering resource available

Focusing on 3rd party integrations

Where PeeringDB Could be Better

Peering data great—colocation data not

No Data Standardization: Facility names, address, services etc.

Errors: No data validation, opaque feedback loop

No data annotations: No way to know source

Where PeeringDB Could be Better

Volunteer effort

Cruft build up

Second guessing data

This is time consuming and error prone

Inconsistent Data is a Solvable Problem

The data you need exists

It’s just not all in one place

Engine to ingest data from lots of sources in lots of formats

The longer it runs, the better it gets

Inflect + PeeringDB

PeeringDB Partnership

Leverage PeeringDB’s dataset as a source

Combine with other datasources in industry

Improving on What’s Available

Automate the data validation process

Integrate multiple data sources

Apply confidence scoring with cross-referencing and first-hand user validation

Interactive platform that allows users to flag and correct data

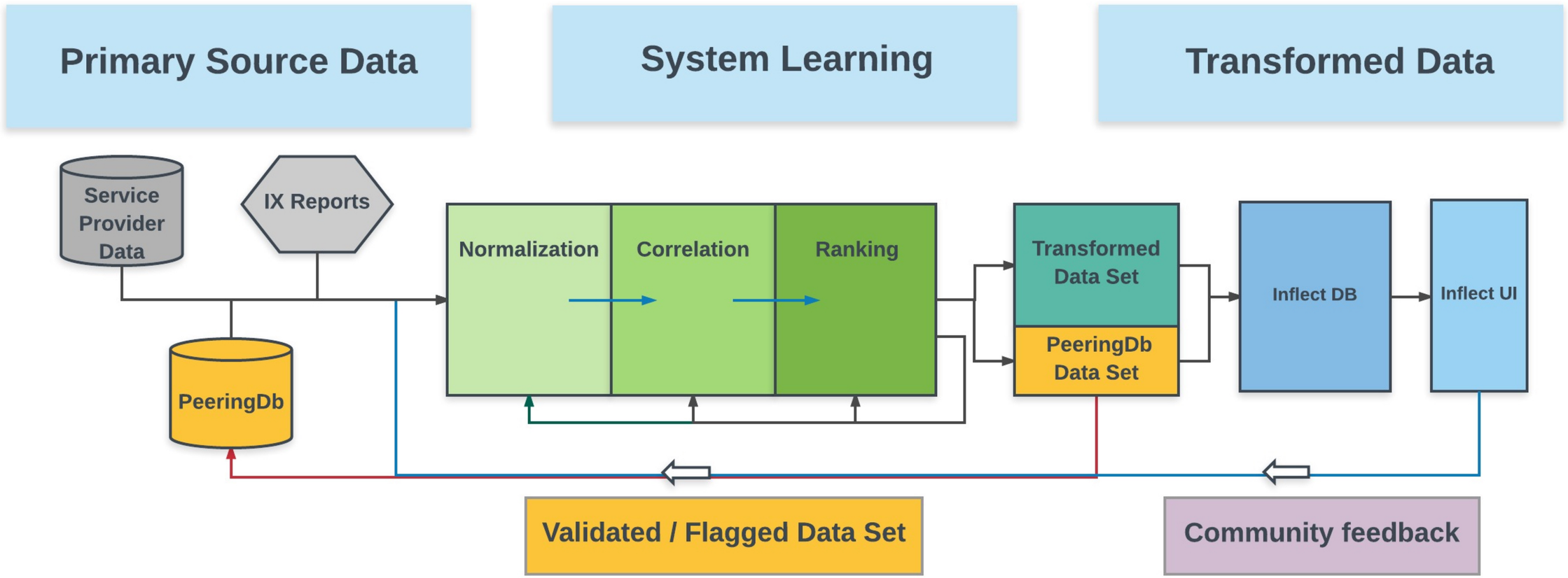
Participation

Service Providers

Data Sources

Community (that means you!)

Data Validation Flow



Data Source Scoring

Correlate and verify across multiple sources

Good peering info? ...rating score improves

Robust colo data? ...rating score improves

Data recently updated and revised? ...rating score improves

Weighting will be tuned and adjusted over time

Applying Confidence Scoring

Tracking changes from data sources

Weighting data sources based on data quality, correctness, ownership, and consistency

Software learns and adjusts weights over time

Data Types Being Normalized

Foundational information - data center names, address data, including floors and suites

Network Service Provider availability at locations

Additional products at those locations

On-boarding Data Sources

API spec

Collaborating with service providers and data sources

Standard API and spreadsheet upload

Comfortable with unstructured data

Data management portal

Integrating Data Sources

Integration layer to format data

Software ingests pre-normalized data

Data is normalized and validated with other sources

System learns and grows as more data comes in

Service Provider Participation

Free

Service Providers upload and manage their data via API and Portal

Goal is to have as much primary source data as possible

User Participation

Free

Your assets can be tagged to your account

Everyone searches the way they want (i.e., 1 Wilshire = 624 S. Grand = CoreSite LA1)

All data available can be flagged for inaccuracies

Working with PeeringDB to push validated data back to Admin team

Wrap Up

Partnering with PeeringDB

Data correctness is important, software can help solve this problem

Ingest as much data as possible, both structured and unstructured

Work with Service Providers and Peering Community

Want to provide a free resource for you to do your job better

Q & A

Eric Silverman
VP of Engineering
eric@inflect.com

Creating a Centralized Database for Interconnection Services

Eric Silverman
VP of Engineering
eric@inflect.com