

Fluree Sense

High-Level Overview



About Fluree Sense

Fluree Sense is unique in the industry in how it uses AI to assist in executing some of the most complex data management processes on Data Lakes.

It was built through many years of working on engagements to either help salvage an existing failed Data Lake implementation, or to help customers modernize from on-premises Hadoop to newer Cloud technologies. It has grown with core methodologies **tested**, **tuned**, and **proven** through production in client environments of Tier-1 Financial institutions.

Founded in 2019

Headquartered in the US;
with offices in India

Data Scientists
averaging 8+ years of
experience

Fortune 500
Clients

Specialized in Large Scale Advanced Data Integrations,
Data Analytics and Data Science based Solutions

Formerly ZettaSense®, is now part of the
Fluree Database Ecosystem™

STRATEGIC PARTNERS »

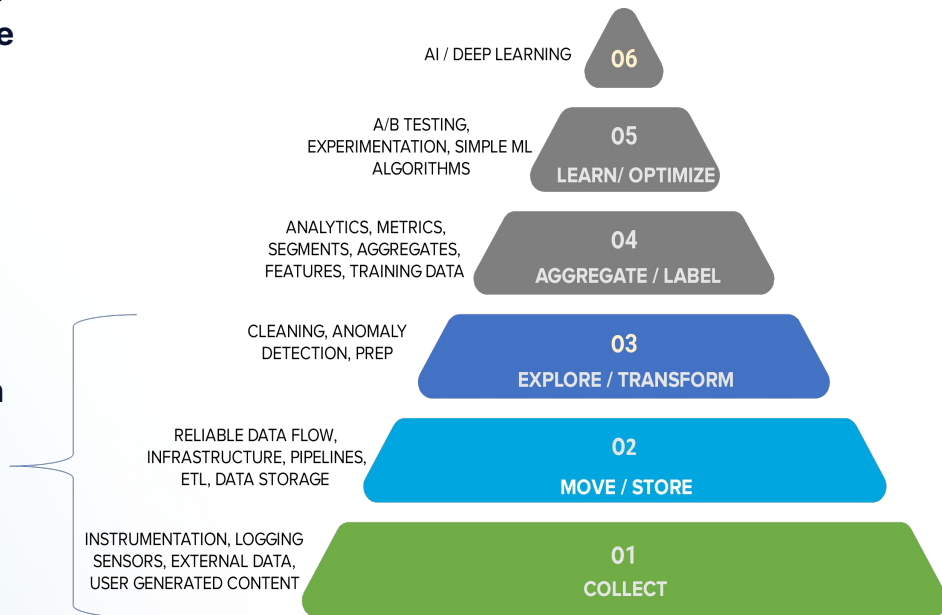


What Fluree Sense Does

Fluree Sense is an innovative software suite that uses its state-of-the-art and fine-tuned **Machine Learning** and **AI algorithms** to automate the process of

- **Data Integration,**
- **Data Quality** and
- **Master Data Management**

Fluree Sense learns and builds metadata from multiple data sources, identifies data duplication and uses machine learning to tag, clean and remediate data, with no programming required



Data is constantly changing and often leads to a massive disruption in the upstream data flow or business consumption process:

- **Data Catalog and Governance systems need to be updated**
- **New rules/SQL code have to be updated for matching to business data needs, new transformation pipeline have to be created**
- **Data Quality and Remediation systems need to be updated for additional quality and code-based rules check**

These disruptions can cause months of delay in implementing efficient business analytics and can be a resource/budget drain to an organization.

Why Fluree Sense

Fluree Sense streamlines the data management process by leveraging its self-learning AI/ML software to efficiently identify data source changes and deploy to the client's business target model with minimal human intervention, typically delivering 10x faster than traditional manual intervention models

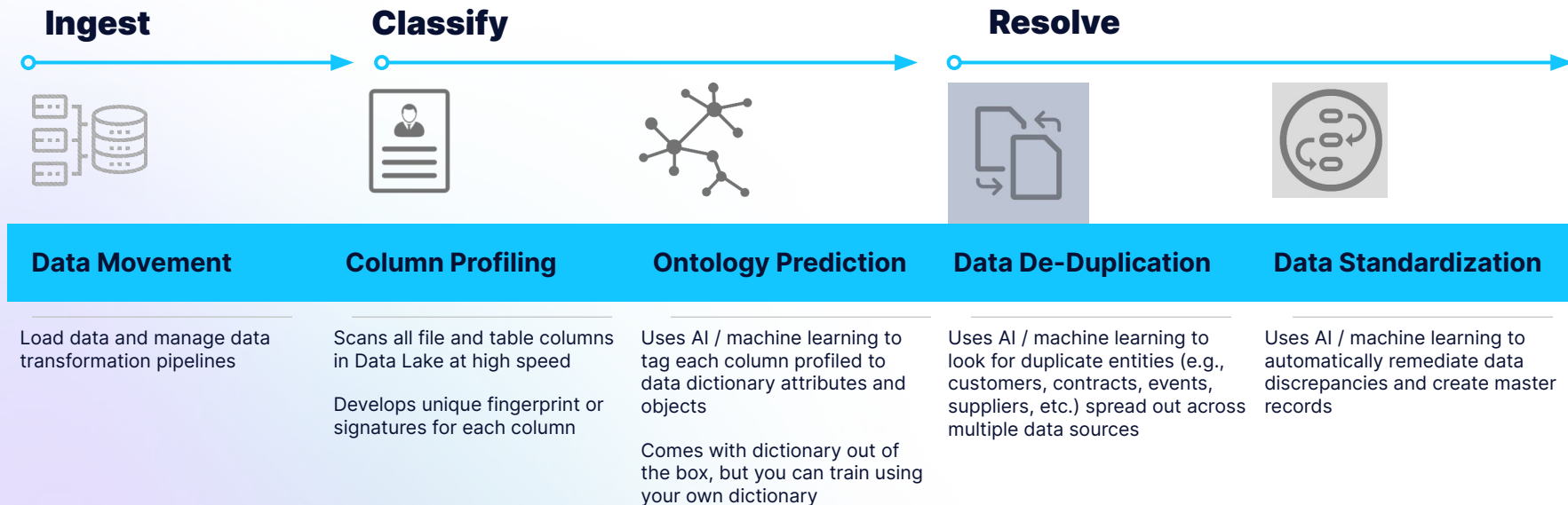
KEY FEATURES

- ★ Auto Data Discovery in data sources using ML
 - Connect and auto populate existing data catalog system or use prebuilt data catalog from Fluree Sense
- ★ Auto generated Data Pipelines exportable to Azure Data Factory
 - Fluree Sense learns and builds pipelines to map source data to client data model
- ★ Auto generated rules for Data Quality & Master Data Management with better reconciliation of data
 - System generates rules to create golden or mastered data
- ★ Maintains data consistency and accuracy due to the automated learning that maps business metadata to the physical data

BENEFITS

- ★ Using AI/ML algorithms (Automation), reduces development efforts by about 60% with about 40% lesser resources
- ★ Drastically reduces the manual coding of data transformation logic, pipelines or MDM rules reducing TCO by 60%
- ★ Built for scale – data entity/model changes or requirements can be configured in days, for quicker adoption of new analytics use cases
- ★ AI system continually improves - the system learns quickly and gets smarter by usage leading to increased productivity
- ★ Uses Modern Data replication methods that support real-time and batch processing

Fluree Sense uses state of the art AI / ML Algorithms fine tuned for better & faster learning based on user behaviors to expedite Data Integration, Data Quality and MDM reducing manual coding and resource/development needs.



Fluree Sense Components

Ingest

Classify

Resolve

Source Data : Disparate DBs

Data Ingestion

1. Raw Data Copy

2. Canonical Form of Data

3. Mastered Data Sets

4. Consumption (Data Views)

Many data tables

Many table / file partitions

May have inconsistent schemas

Onboard one or more sources at a time

Data Extraction + Load (no transformation)

Batch

Job scheduling

Data ingestion quality management

Data Tables and Files

Data Tables and Files

Semantic Copy of Data Tables & Files

Semantic Copy of Data Tables & Files

Mastered Data by Subject Area

Risk View

Finance View

Exact duplicate copies of Data Source

Files loaded into HDFS into a common Fluree Sense technical file format

Raw data archive

Data column headings changed to conform to business ontology

Physicalized on demand (when needed)

Data tables from disparate sources integrated

Data values cleansed and reconciled

Use of machine learning + crowdsourcing to clean values

Physicalized on demand

Versioned and saved in HDFS

Unique views created per business use case

Aggregations, row and column filters

Data cell-level entitlement and control

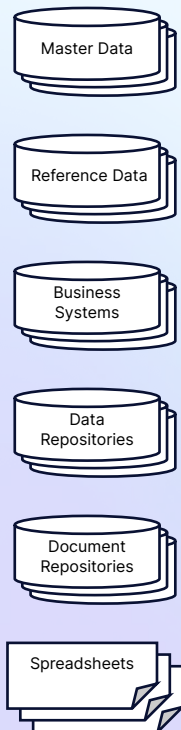
Full traceability and lineage

Data Flow

	Manage Data Ingestion pipelines from many sources	Automated Metadata Tagging models for how to classify and tag data	Match Learning Model how to cluster and find duplicate entities	Survivorship / Merge Learning Model Merge duplicate records into one golden record	Self-Service Fast implementation with minimal training; vs. requiring services	Crowdsourcing Distribute training as tasks and marker-checker or peer-review workflows
						
						
						
						
						
						
						

Modern Data Architecture

Client Data Ecosystem



Ingest

DATA INGESTION

Batch Multi-table
Data Movement

Delta & Change
Data Capture

Streaming Data
Movement

Secure File /
Document
Transfer

Classify

Resolve

ENTERPRISE DATA LAKE (HADOOP ON PREMISES OR CLOUD)

Raw Zone

Original
Copy of Data

Semantic Zone

Tagged,
conformed
Copy of Data

Mastered Zone

Cleansed,
reconciled
Copy of Data

Data Processing

YARN

Map Reduce

SPARK

Data Storage

HDFS

Cloud Data Lake

CONSUMPTION ZONE (NOSQL DATABASE)

Structured
Information
Storage (NoSQL)

In-Database
Processing

Data Consumer
Subscription
Management

Data Feed
Management



cloudera



MAPR

The Fluree Sense Advantage

Fluree Sense Approach Is Different Compared To Traditional Data Management Processes

We Used To...

- Have developers program lots of code to link and match data (ETL & SQL)
- Program thousands of business rules to measure data quality
- Hire large data management teams to manually remediate poor or missing data

But Now We Can...

- Use heuristics (instead of rules and code) to teach machines how to spot and fix data quality issues
- Have machines programmatically scan billions of rows of data and merge / fix them
- Crowdsource Human Expertise: have anyone at an organization teach the machine how to improve quality (rather than dedicated large data management teams)
- Do this with No Programming!!! (No ETL & No SQL)

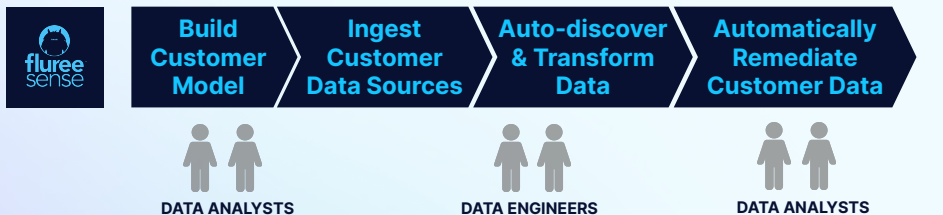
Which Leads To...

- A way to generate and link golden sources of customer, product, account and transaction master data
- A true repository of all organization activity, reconciled to official books and records (transactions, positions, balances)
- An ability to view all activities of a customer across all of their product relationships

The Fluree Sense Advantage

Improved Time to Market 6x
Reduced Overhead cost by 10%

8 weeks, 2 business analysts, 2 data engineers, ~\$60k labor cost



TIMELINE TO
DEMONSTRABLE
RESULTS

30 weeks, 3 business analysts, 6-8 data engineers (IT developers), ~\$650k labor cost

Typical approaches



Thank you

Learn More at www.flur.ee/sense