Data and the Software Stack: simple data architecture steps that prep for machine learning

Miriah Peterson

weave

Bio

- Engineer at Weave in Lehi, Utah
- Board Member with Forge Foundation
- Proud Dog Mom
- Twitter captiainnobody1
- GitHub soypete
- LinkedIn

weave

Types of Data

Data: Data (treated as singular, plural, or as a mass noun) is any sequence of one or more symbols given meaning by specific act(s) of interpretation.

Types of Data

Data: Data (treated as singular, plural, or as a mass noun) is any sequence of one or more symbols given meaning by specific act(s) of interpretation.

Types:

- Machine-generated data
- User generated data

machine-generated data

What are types of machine generated data does your software produce?

Machine-generated data

What are types of machine generated data does your software produce?

- Event Data
- Error Data
- Network Usage Data
- App Actions (click-stream) data
- Change log data

User generated

What are types of user generated data does your software produce?

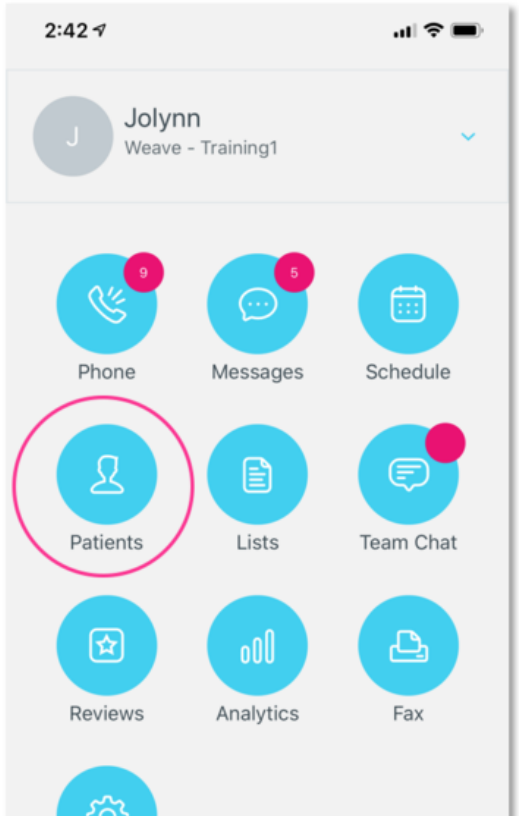
- Account information
- Personal information
- Transaction (payment/currency)
- Product Data

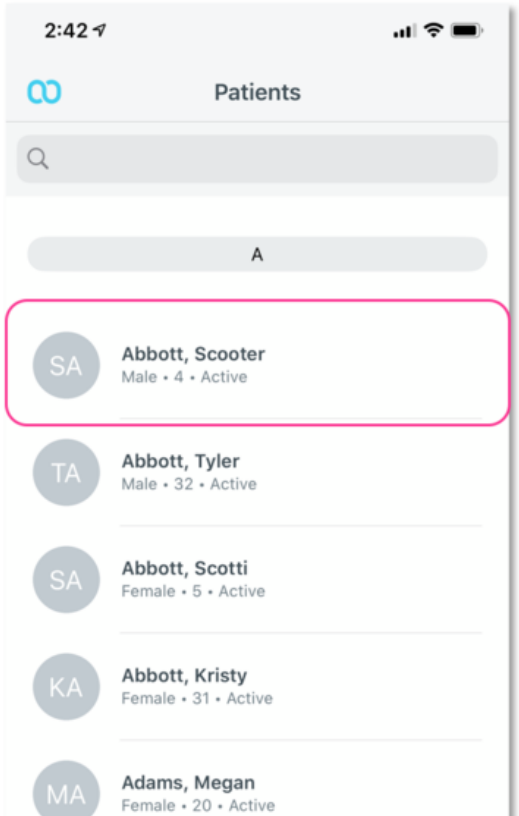
Types of Data

Which is more important?

Ways to use your data - Direct in Product

Select the Patients/Customer Icon





Tap on the name you want to add a photo to

Ways to use your data - Dashboard/Reports

weave

Ways to use your data - Dashboard/Reports

In what ways are you using machine-generated or user generated data?

Machine Learning

Machine Learning

Monitoring usage to trigger an action.

Machine Learning

Forcasting outages/high network usage

Machine Learning

- Product enhancement (weave sentiment analysis)

Data Architecture as part of planning process

Data Architecture as part of planning process

Questions to ask when planning or designing

Data Architecture as part of planning process

Questions to ask when planning

- What are creating, consuming, or communicating the data?

Data Architecture as part of planning process

Questions to ask when planning

- Am I creating, consuming, or communicating the data?
- Where is the final resting place of my data?

Data Architecture as part of planning process

Questions to ask when planning

- Am I creating, consuming, or communicating the data?
- Where is the final resting place of my data?
- How much does is cost to store the data?

Data Architecture as part of planning process

Questions to ask when planning

- Am I creating, consuming, or communicating the data?
- Where is the final resting place of my data?
- How much does is cost to store the data?
- What are availability needs?

Data Architecture as part of planning process

Questions to ask when planning

- What are availability needs?
  - Does the product need to consume this data?

Data Architecture as part of planning process

Questions to ask when planning

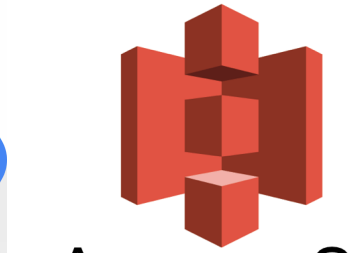
- What are availability needs?
  - Does the product need to consume this data?
  - Do other services need to consume this data?

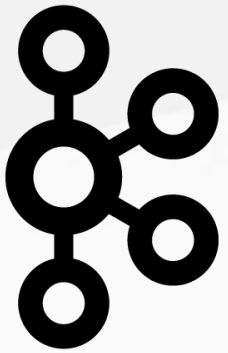
Data Architecture as part of planning process

Questions to ask when planning

- What are availability needs?
  - Does the product need to consume this data?
  - Do other services need to consume this data?
  - How often is the data consumed?

Data Architecture as part pf planning process





Data Architecture as part of planning process

Methods for adding Machine Learning Models

- API calls
- Static Model
- Part of Event Stream

Data Architecture as part of planning process

Where do we start?

Where do we start?

- Pain points

Where do we start?

- Pain points
- Central Features

Where do we start?

- Pain points
- Central Features
- Product needs

Questions?

Resources

*wikipedia data-computing
*Software Engineering Daily May 28, 2020
*machine-generated vs human-generated
*cloudera autoscaling