

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

Miriah Peterson

Bio

- Engineer at Weave in Lehi, Utah
- Board Member with Forge Foundation
- Proud Dog Mom
- [Twitter @captainnobody1](#)
- [GitHub soypete](#)
- [LinkedIn](#)



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
- Database change capture (anomaly detection)

User generated

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

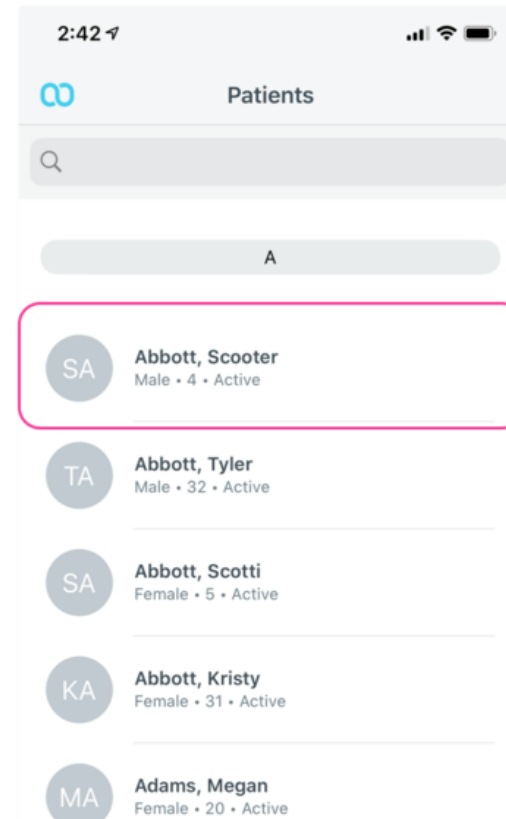
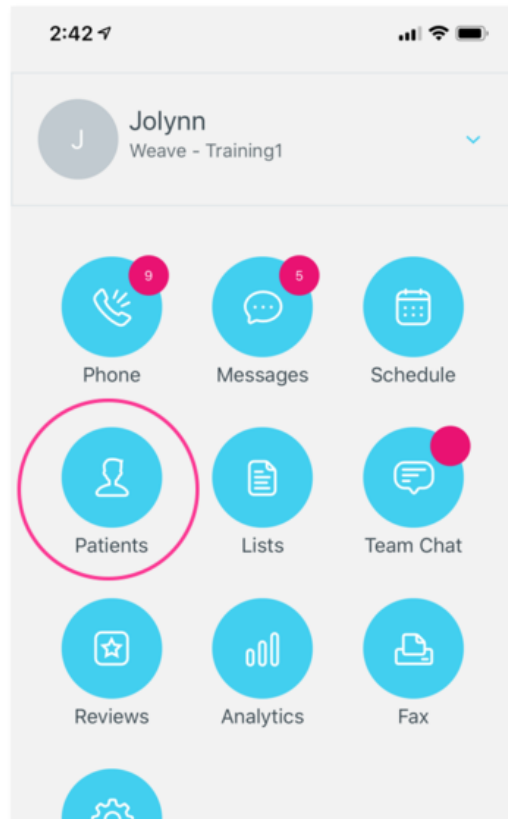
- Account information
- Personal information
- Transaction (payment/currency)
- Product Data
- *Metadata

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



Ways to use your data

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

Ways to use your data

Machine Learning from stored data.

Machine Learning

Machine Learning

Monitoring usage to trigger an action.

Machine Learning

Forecasting 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

Data Architecture as part of planning process

Questions to ask when planning

- Who is creating, consuming, or communicating the data?

Data Architecture as part of planning process

Questions to ask when planning

- Who is creating, consuming, or communicating the data?
 - Where is the final resting consumption of my data?

Data Architecture as part of planning process

Questions to ask when planning

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

Data Architecture as part of planning process

Questions to ask when planning

- Who is creating, consuming, or communicating the data?
 - Where is the final resting place of my data?
- How much does it 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 of planning process

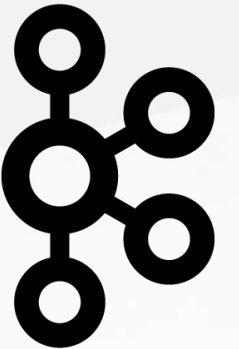


Amazon S3



ViteSS

VERTICA



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](#)