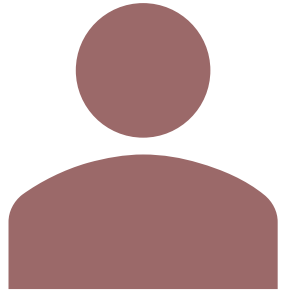
# POWERING OTT WITH DATA SCIENCE

Thulasiram

# AGENDA

Goals



Team Structure

# DATA SCIENCE USE CASES

| | | | | |
|---|---|---|---|---|
| Content | Subscription | Product Analytics | Customer Life Cycle Management | Performance marketing |
| Social Media Analytics | Customer Success | Recommendation Engines | Advertising | Instrumentation |

# CONTENT

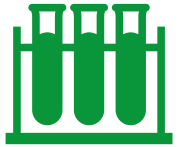
# SHORT TERM

Attributes to capture

Inputs for Front end & Database Design

Assigning ID's to video assets

Co-ordinate with content team to publish assets



QA / Testing

# MEDIUM TERM

- Type of content to create
- Casting & Renumeration
- Conversion
- Engagement
- Dubbing

# LONG TERM

- Creating Blockbusters
- Content Green lighting
- Neuro Research
- Generating Content Metadata
- Generating Subtitles
- Generating Creatives

https://www.ted.com/talks/sebastian_wernicke_how_to_use_data_to_make_a_hit_tv_show?language=en

SUBSCRIPTION

# SHORT TERM (B2B & B2C)

ID Management

Database design

Frontend design

Integration with Payment Gateways

Capturing Auto-renewals

QA / Testing of data collection

# MEDIUM TERM

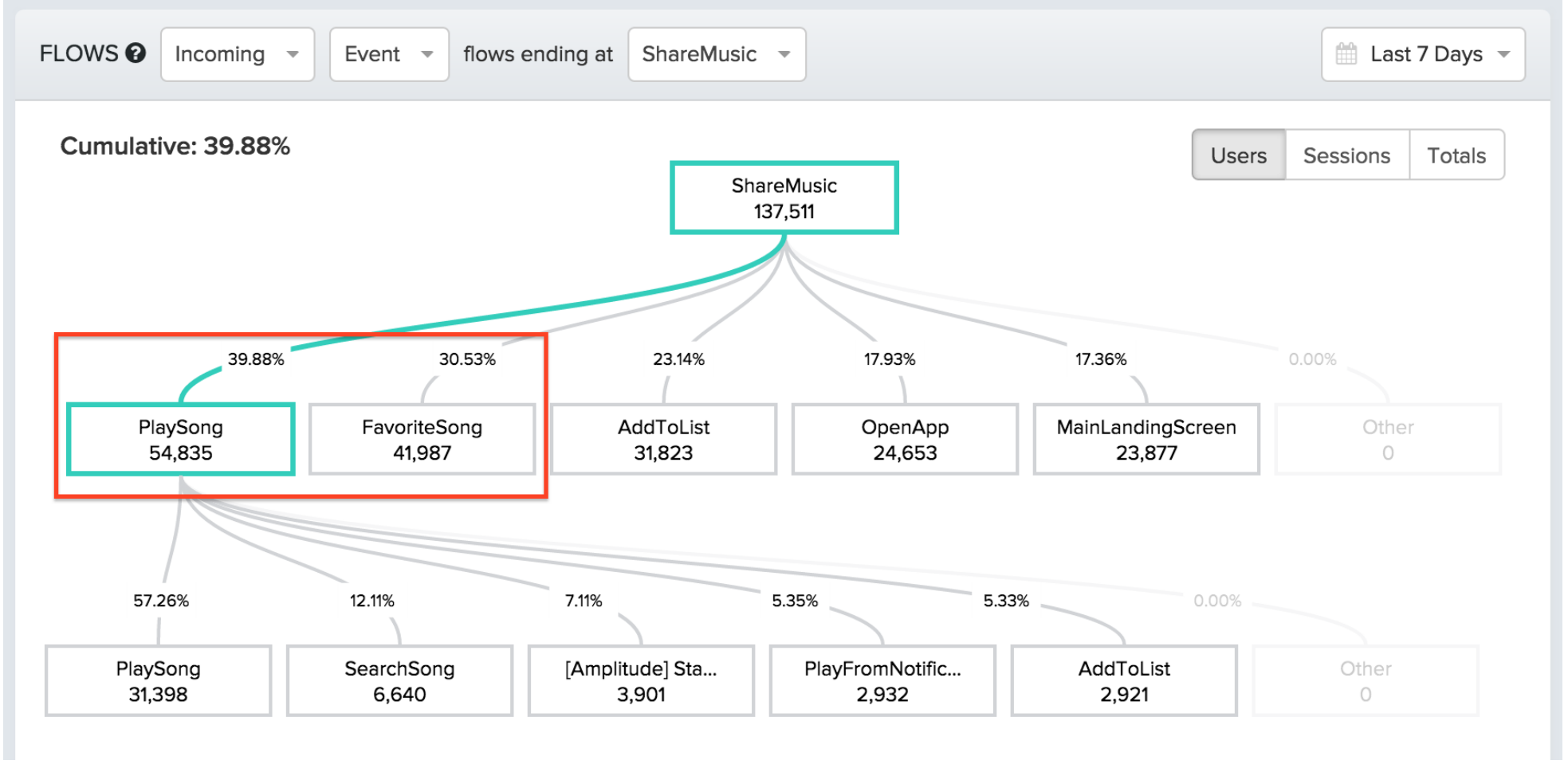
- Manual / Auto-renewal
- Payment gateway failures
- Revenue at stake on a given day
- Revenue generated
- Packs working / Price optimization
- Subscribers changing packs
- Customer LTV

# MEDIUM TERM

- Window shoppers / Hoppers
- Watch pattern of subscribers on various packs
- Content attributing to subscriptions
- Registration funnel
- Subscription funnel
- Retention analysis
- Subscription dashboard

# MEDIUM TERM

# Flow analysis



# LONG TERM

CHURN PREDICTION

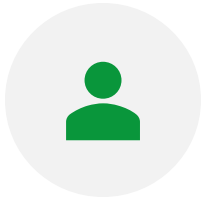
TASTE CLUSTERS

FORECASTING REVENUE / SUBSCRIBERS

MODELLING & REAL TIME NOTIFICATIONS

COUPON CODE ANALYSIS

MODELS FOR ALERTING



MODELS TO CONVERT USER TO VIEWER

# PRODUCT ANALYTICS

# SHORT TERM - INSTRUMENTATION

Deciding events to capture

Tool for instrumentation

Co-ordination for implementation

QA / Testing / Pitfalls to avoid

Data collection / Latency

Processing & automation in medium and long term

# QUALITY OF EXPERIENCE

- Requirement gathering
- Tool selection and Integration
- Data collection - EBVS, VSF, Bitrates, ISP, Errors, Playing time etc
- QA / Testing
- Plan to process the data

# SHORT TERM - RECOMMENDATION ENGINES

Capturing user ratings

UI design & Data capture

Data processing

Deriving content metadata & similarities

Recommendations based on content meta

# MEDIUM TERM - RECOMMENDATION ENGINES

COLD START CHALLENGES

COLLABORATIVE FILTERING

EVALUATING QUALITY OF RECOMMENDATIONS

OPTIMIZING THE ALGORITHM

# LONG TERM - RECOMMENDATION ENGINES

ARTWORK PERSONALIZATION

TEXT PERSONALIZATION

# EXPERIMENT DESIGN

Short term

Platform capability

Medium term

Website optimization

Utility of new features

Long term

UI for experimentation

Automation

Experimentation Dashboard

PERFORMANCE
MARKETING

# SHORT TERM

TOOL FOR MEASURING INSTALLATION

INTEGRATION

DATA COLLECTION

QA / TESTING

PROCESSING THE DATA

# MEDIUM TERM

- Campaign performance
- Creative performance
- Channel performance
- Analysis of Installs / Uninstalls
- Revenue measurement
- First touch attribution
- Dashboard

# LONG TERM

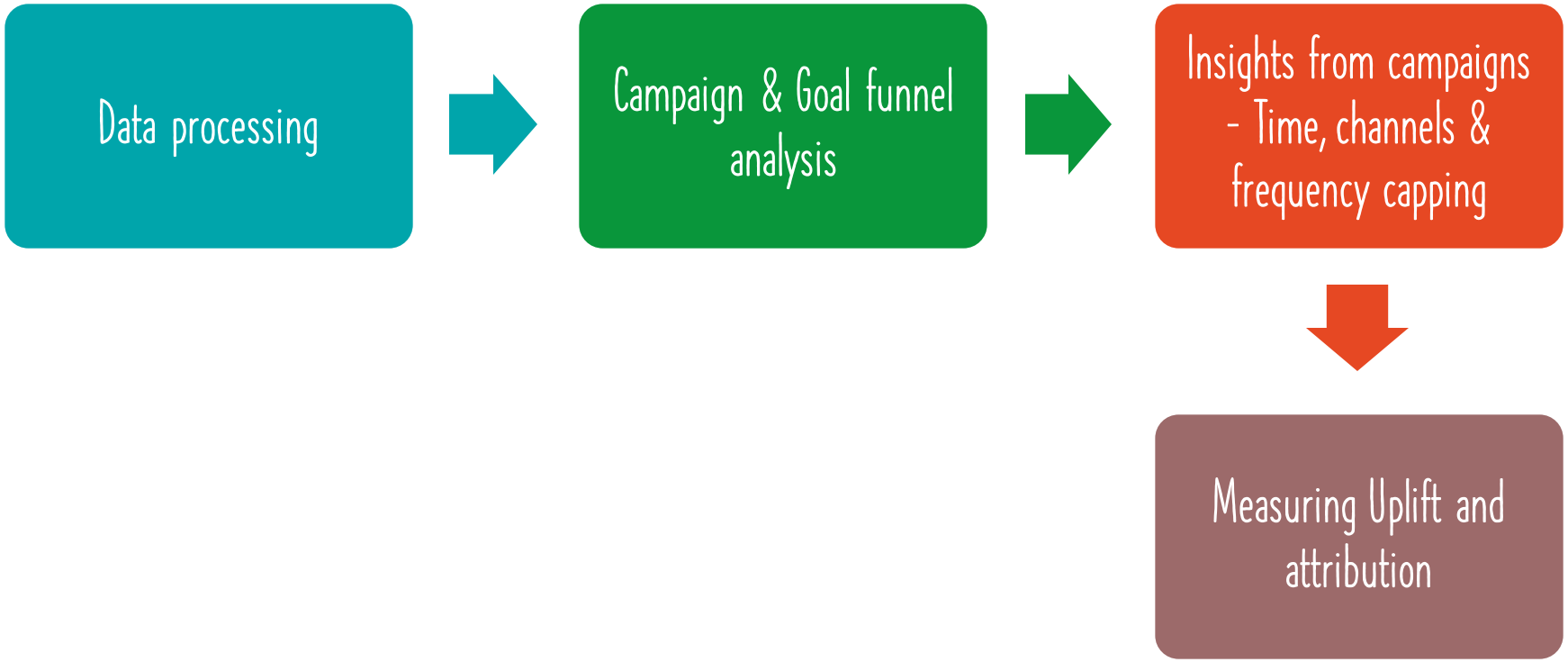
- Multi touch attribution
- Fraud detection
- Predicting Uninstalls
- Retargeting
- RFM analysis / Clustering

# CUSTOMER LIFECYCLE MANAGEMENT

# SHORT TERM

- Features required & tools selection
- Tool integration & channel activation
- Data Collection
- QA & testing
- Architecture for processing
- Attribution for video views / subscription
- Contact policy
- Goal setting

# MEDIUM TERM


Data processing
Campaign & Goal funnel analysis
Insights from campaigns - Time, channels & frequency capping
Measuring Uplift and attribution

# LONG TERM

Campaign performance dashboards

Automation workflows for B2B

Integration with churn models

Integration with taste clusters

Creative performance

Predicting CTR

# SOCIAL MEDIA ANALYTICS

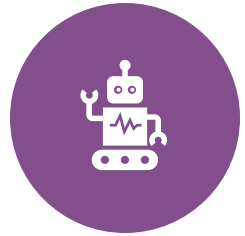
# SHORT TERM

CHANNEL SELECTION

COMPETITORS SELECTION

PLAN FOR DATA COLLECTION & TOOLS SELECTION

PIPELINE FOR DATA COLLECTION



AUTOMATION & PROCESSING

# MEDIUM TERM

- Brand performance
- User engagement
- Show / Movie page performance
- Creative performance
- Social media dashboard
- App store rating and sentiment

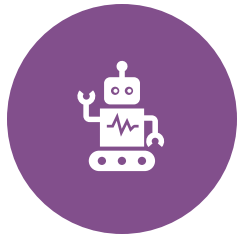
# LONG TERM

SENTIMENT MODELLING

TOPIC MODELLING

TOP POSITIVE / NEGATIVE SENTIMENTS

COMPETITION DASHBOARD



AUTOMATION

# ADSALES

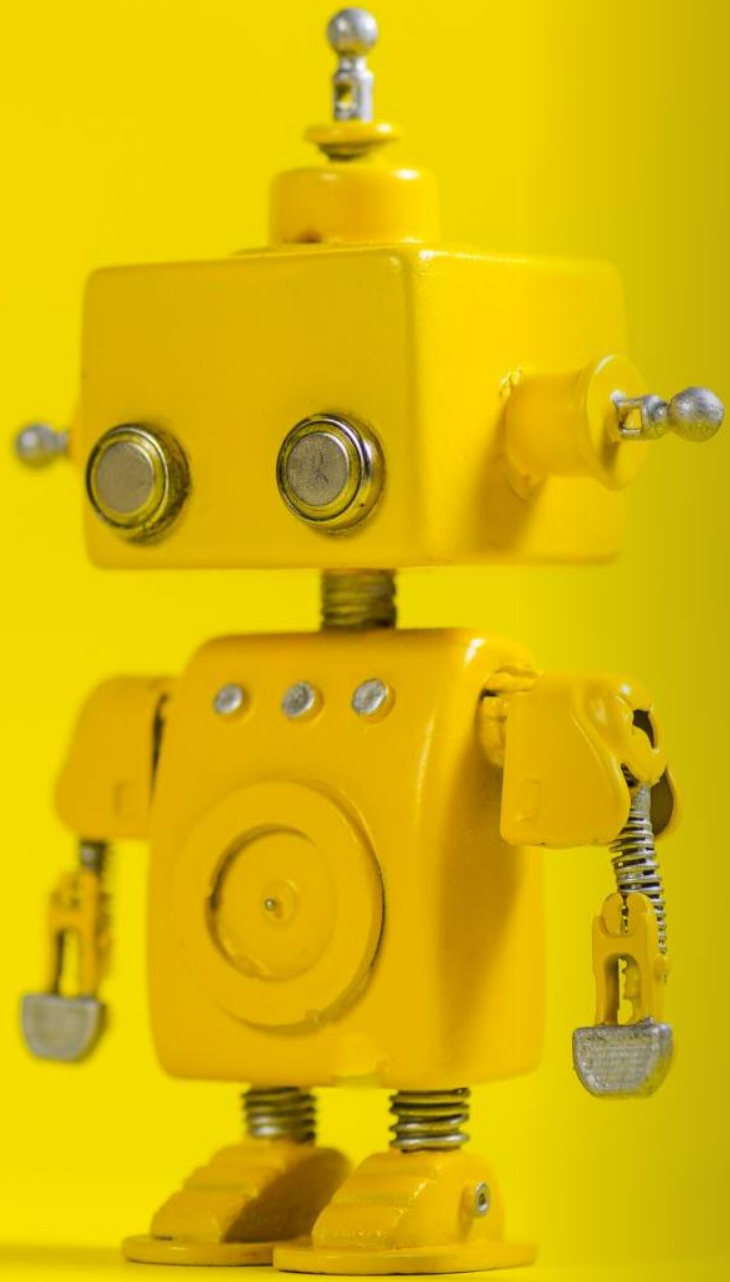
# MEDIUM TERM

- 1st , 2nd and 3rd party data
- Segmentation for targeting
- Integration with SSP & DSP
- Data capture of ad campaigns
- Data processing
- Adsales Dashboard
- Ad frequency capping / Ad fatigue

# LONG TERM

- Ad campaign performance
- Ad campaign dashboards
- DMP integration
- RTB
- Age / Gender prediction
- CTR prediction

# ARTIFICIAL INTELLIGENCE



# SHORT TERM

GENERATING SUBTITLES USING CLOUD APIS

TRANSLATING SUBTITLES

CONTENT METADATA USING CLOUD APIS

THUMBNAIL GENERATION

# MEDIUM / LONG TERM

Video summarization for creatives

Creative generation for various genres

Script generation for new content

CUSTOMER SUCCESS

# SHORT TERM

DESIGN & DATA COLLECTION - CONTACT US

DATA COLLECTION FROM CRM SYSTEMS

QA / TESTING

ARCHITECTURE TO COLLECT AND PROCESS THE DATA

# MEDIUM TERM

Insights from support tickets

Inputs to product & tech for improvements

Agent performance

Time to resolution

# LONG TERM

Model to categorize support tickets

Automatic ticket assignment / Model learning from re-assignment

Recommendations for resolution / Similar support tickets

Chatbot support

# MISCELLANEOUS

# DATA ENGINEERING

- Short term
  - Designing data pipelines
  - Planning Data warehouse and Data Marts
- Medium term
  - Processing and execution
  - Automating dashboards
- Long term
  - Creating tools for use by data scientists

# INFRASTRUCTURE TEAM

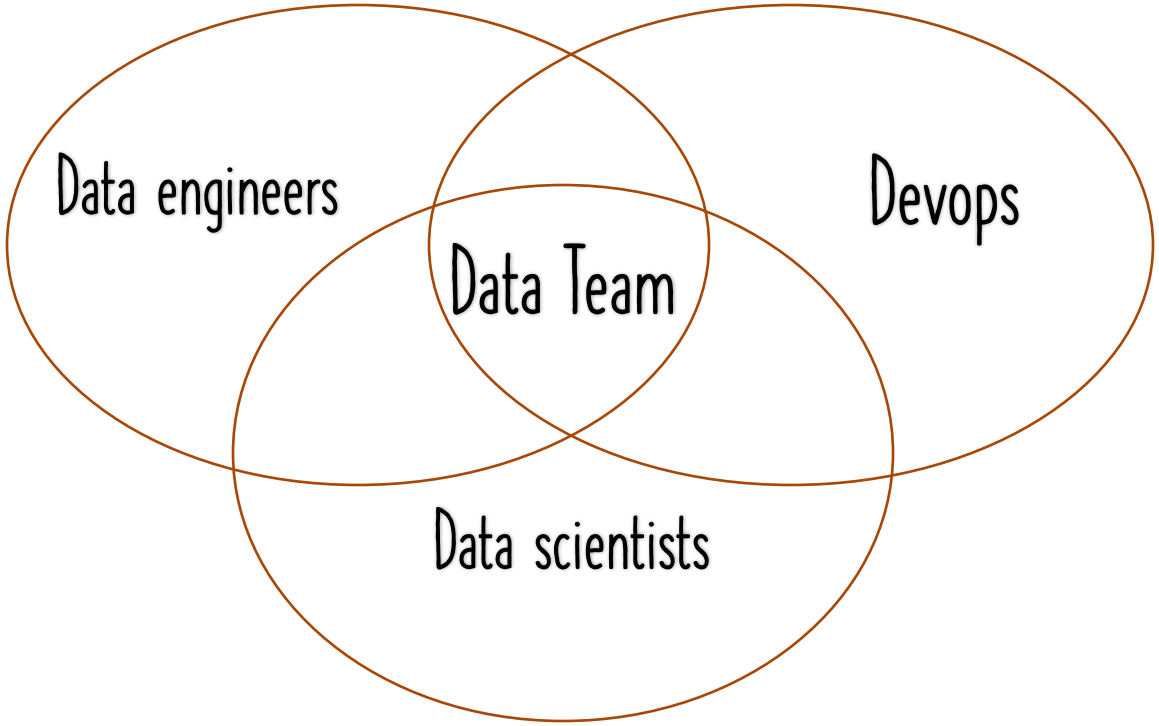
- Short term
  - Tool selection - Cloud services, Notebook environments, versioning, CI / CD, deployment, Monitoring, Business Intelligence
  - POC
  - Access control
- Medium term
  - Containerization
  - Cost optimization

# DATA GOVERNANCE - ON GOING

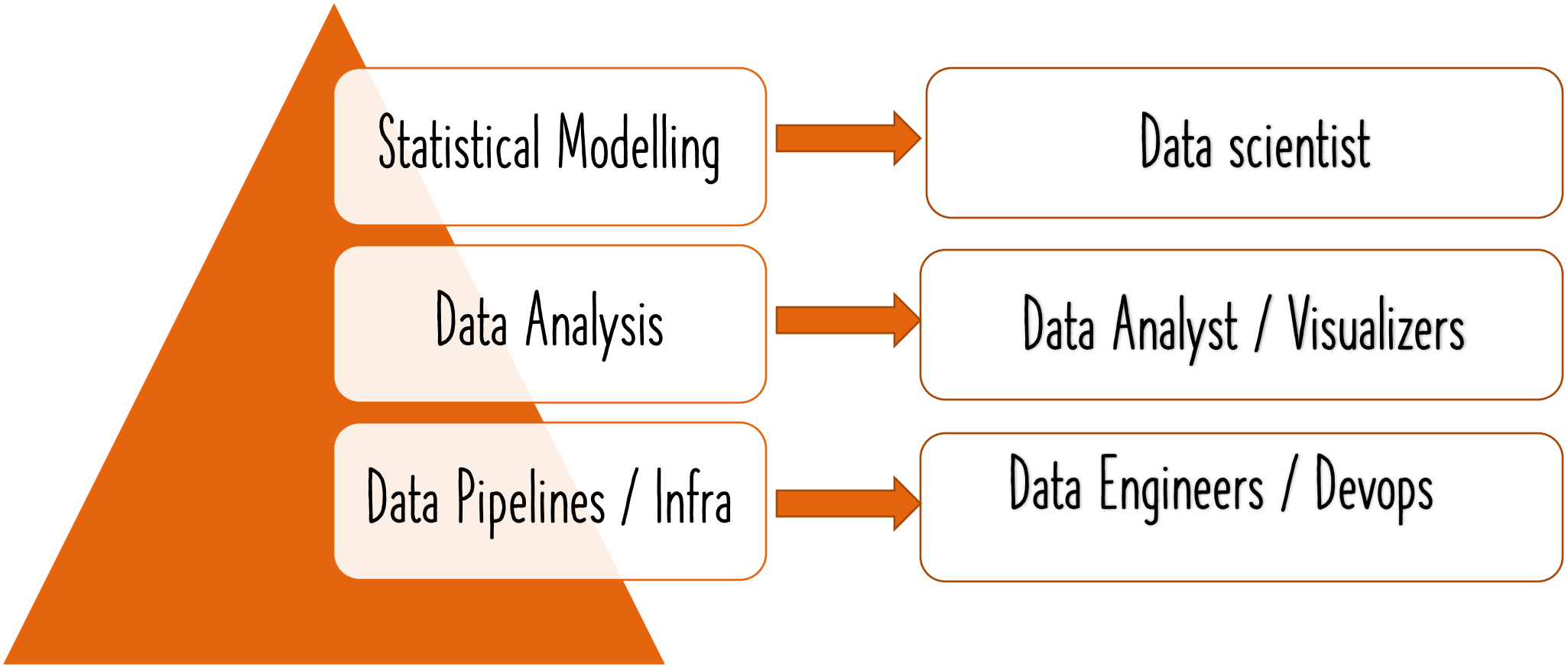
- Data cataloguing
- Data discovery platform
- Compliance to GDPR & CCPA
- Governance mechanism
- Auditing
- Documentation

# TEAM STRUCTURE

# TEAM COMPOSITION

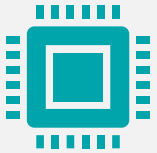


# HIRERACHY




Reference: https://www.datateams.io/

# SHORT & MEDIUM TERM CENTRALIZED



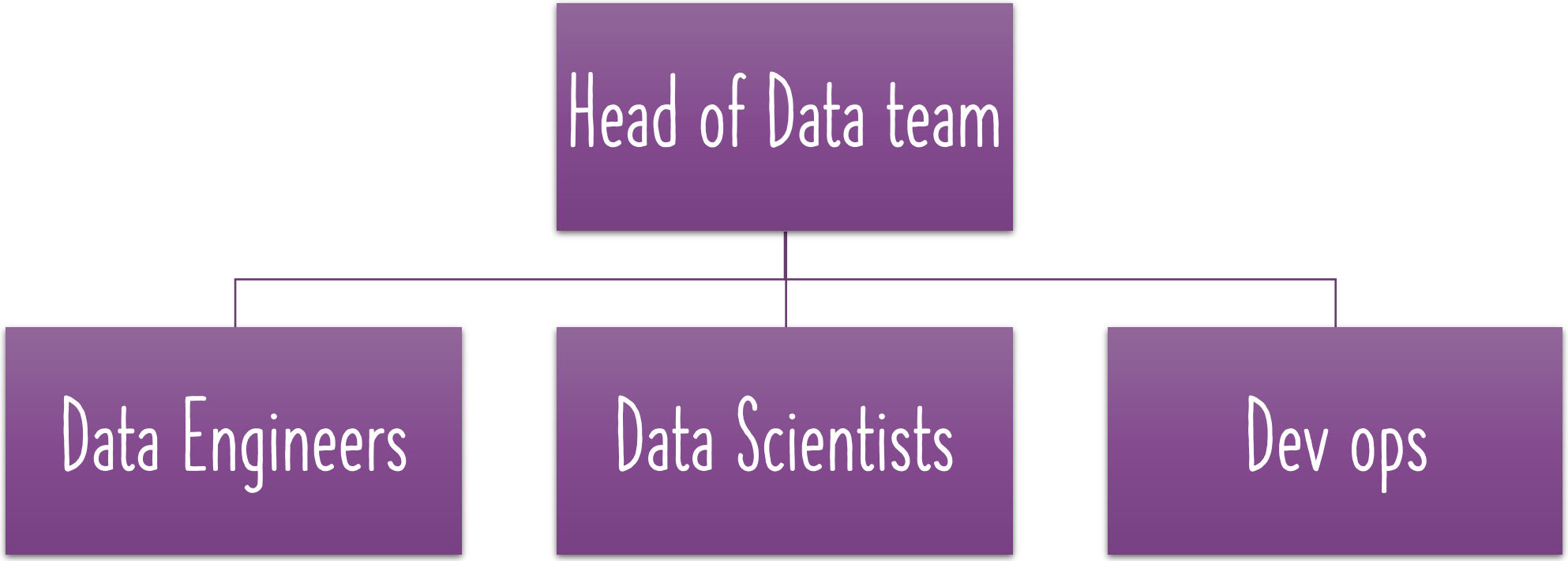
80 percent - Data engineers

15 percent - Devops

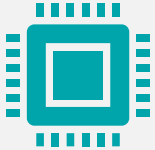
15 percent - Data Scientists

Reference: https://www.datateams.io/

# SHORT & MEDIUM TERM ORG STRUCTURE



# LONG TERM - CENTRALIZED / HUB-SPOKE MODEL

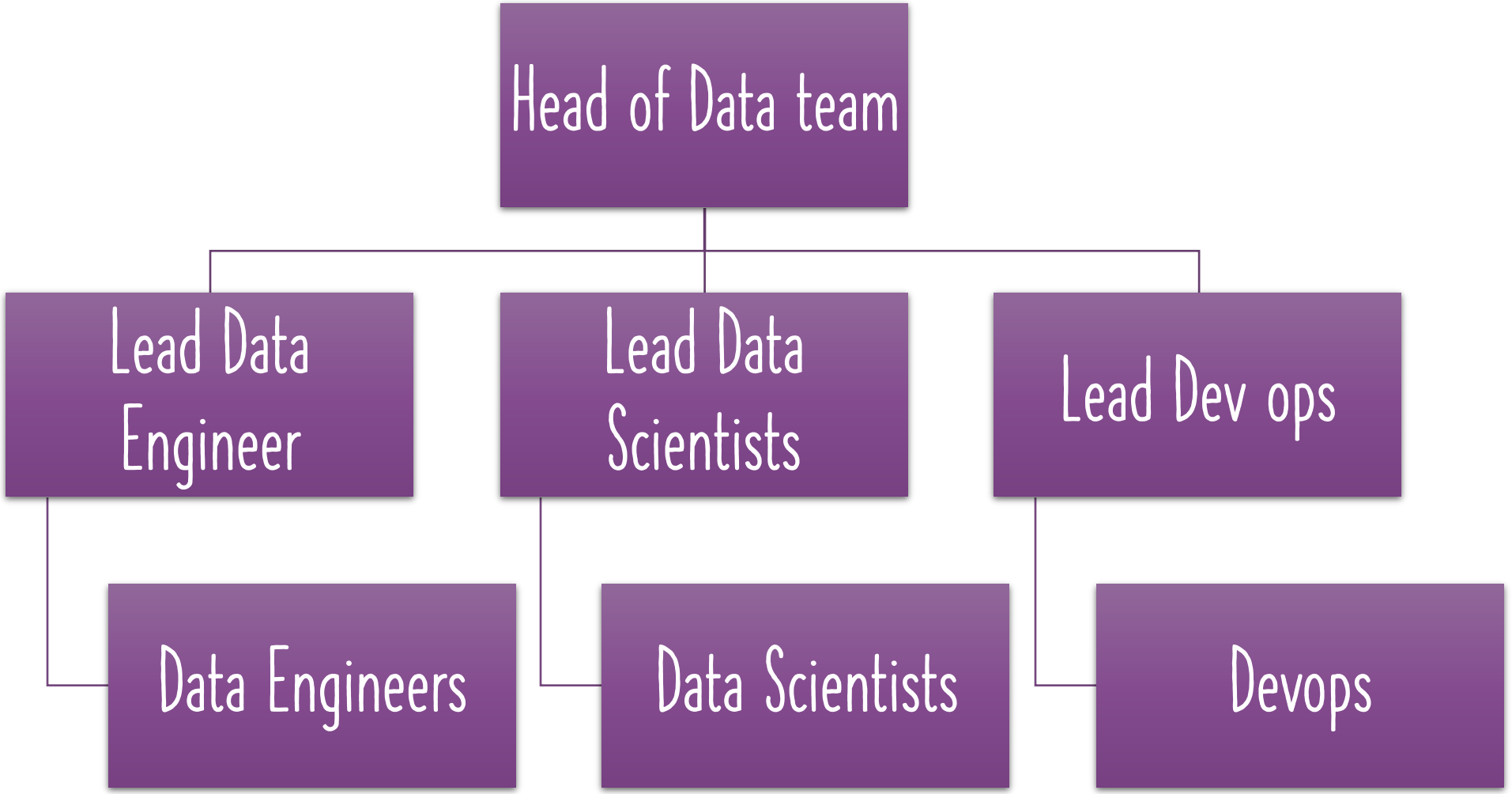


50 percent - Data engineers

15 percent - Devops

35 percent - Data Scientists

# LONG TERM ORG STRUCTURE



# THANK YOU

Thulasiram

Tulasiram.Gunipati@gmail.com

Advaitha.github.io