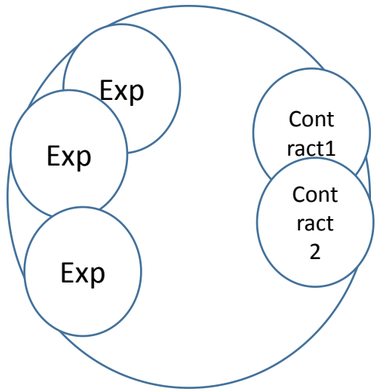
Building Content Recommendation System based on Content modeling

Tencent Zixunsun
腾讯 孙子荀

- Architecture Overview
  - *Content Side*
  - *User Side*
- User Modeling
- Content Modeling
- Content Processing Architecture

Arch Overview User Side

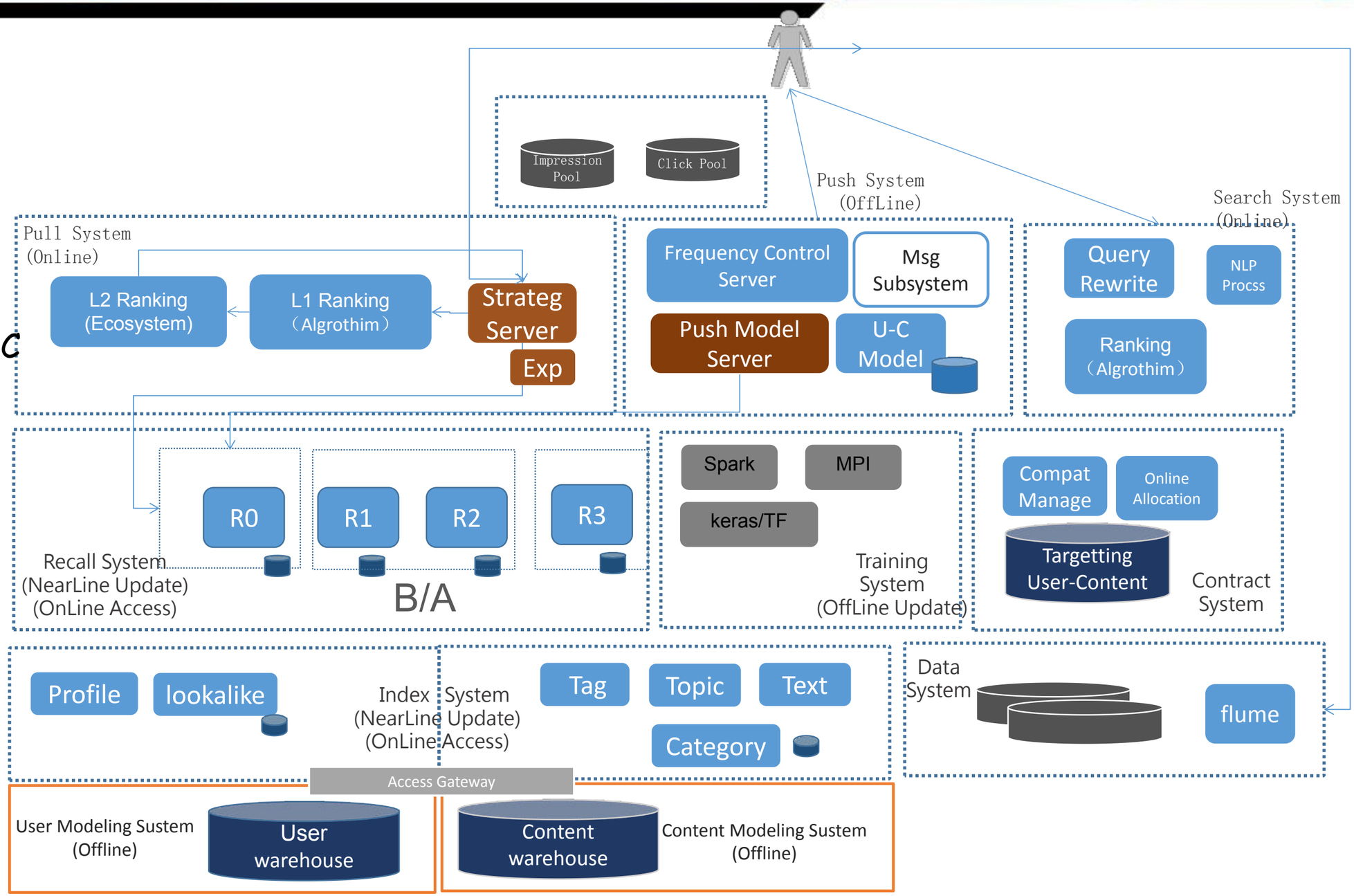


1.U>Model(U)?Model(C)->C
Simility

2.U->C->Model(C)->C
Cluster

3.U=>U->C
UserCF
U->C=>C
ItemCF

4.U->Model(U-C)->R(C)
:FM



Content Description

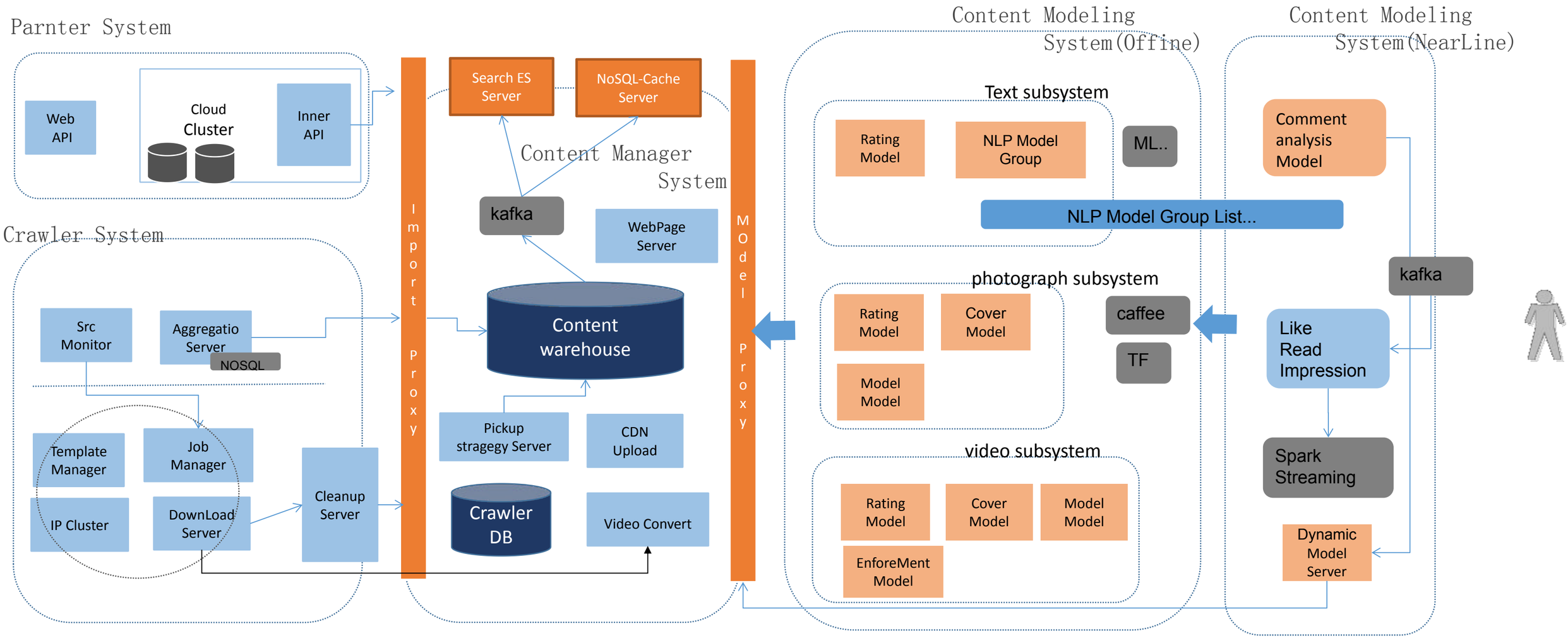
Data type: Text , URL, Video Refer

Control Type : Life Cycle Manager

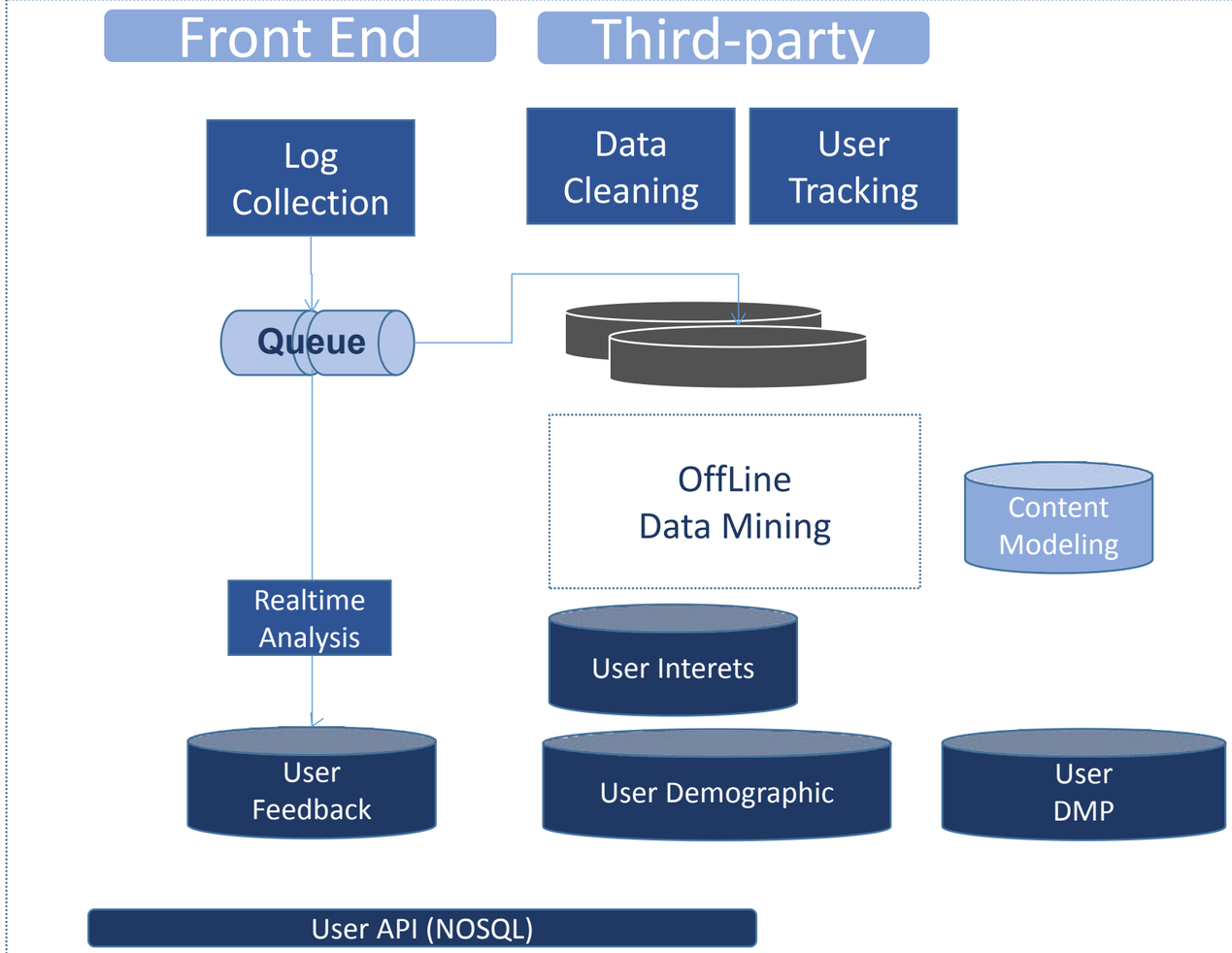
Model Type: Semantic analytics ,Mutile-Classification

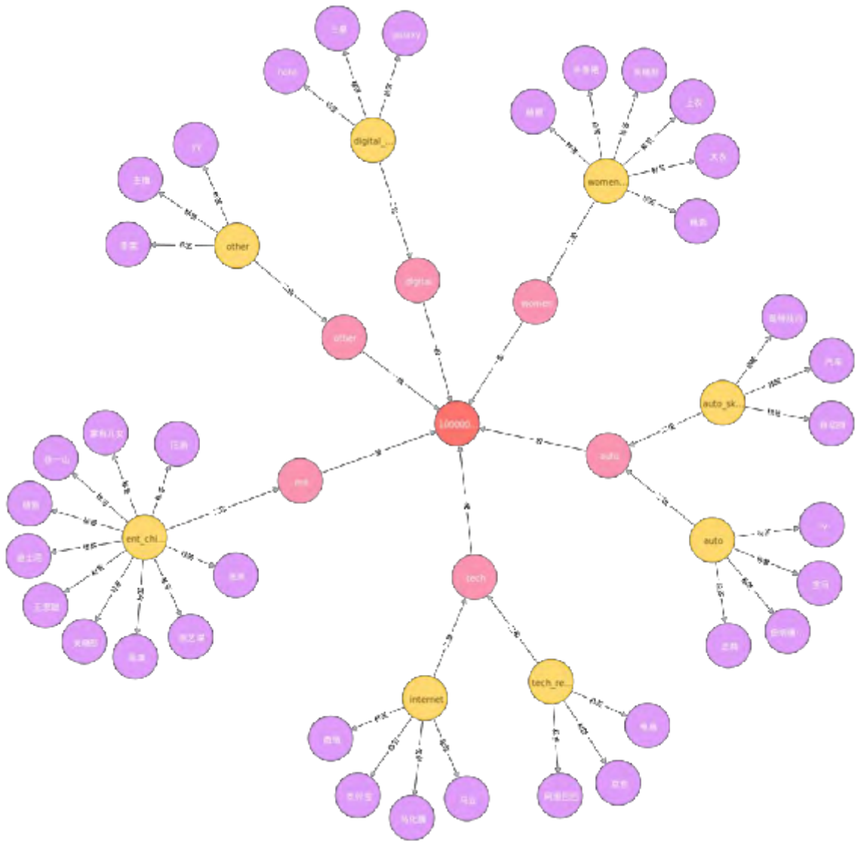
Photo Type : OCR , Enhance, Obj Tag , Rating, Cut...

Evaluation Type : Content Rating ,Media Rating


Parnter System
Web API
Cloud Cluster
Inner API
Crawler System
Src Monitor
Aggregatio Server
NOSQL
Template Manager
Job Manager
IP Cluster
DownLoad Server
Cleanup Server
Import proxy
Search ES Server
NoSQL-Cache Server
Content Manager System
kafka
WebPage Server
Content warehouse
Pickup stragegy Server
CDN Upload
Crawler DB
Video Convert
Model proxy
Content Modeling System(Offine)
Text subsystem
Rating Model
NLP Model Group
ML..
NLP Model Group List...
photograph subsystem
Rating Model
Cover Model
Model Model
caffee
TF
video subsystem
Rating Model
Cover Model
Model Model
EnforeMent Model
Content Modeling System(NearLine)
Comment analysis Model
kafka
Like Read Impression
Spark Streaming
Dynamic Model Server

	TimeLiness	Storage	latency	Data Src
Feedback	Real-Time	rockdb	local (2ms)	behavier
Interets	Near Real-Time	redis	remote (10ms)	behavier mining
Demography	Non-Real Time	redis	remote (10ms)	mutil src mining





Traditional user Interets profile building from content tag

"As is the Content, As is the User"

Tag is never Enough .

User TAG - > User Knowledge

User KnowLedge Building From

1 Content Modeling

2 User Behavier

"How deeply know the Content,

How strongly know the User"

Doc1

Content Feature：

Entity：Singer Wangxxx
Category：Entertainment
Desc ：Divorce
Emotion Analyses：Negitive

Reading completion rate
20%
Comment : Negitive

UserA
<Girl，16>

Content Feature：

Entity：Singer Wangxxx
Category：Entertainment
Desc：Marriage
Emotion Analyses：Neutral

Reading completion rate
80%

Doc2

Content Feature：

Entity：NBA Player
Category：Sports
Desc：Game Report
Emotion Analyses：Neutral

Nerver Click

UserB
<Man，45>

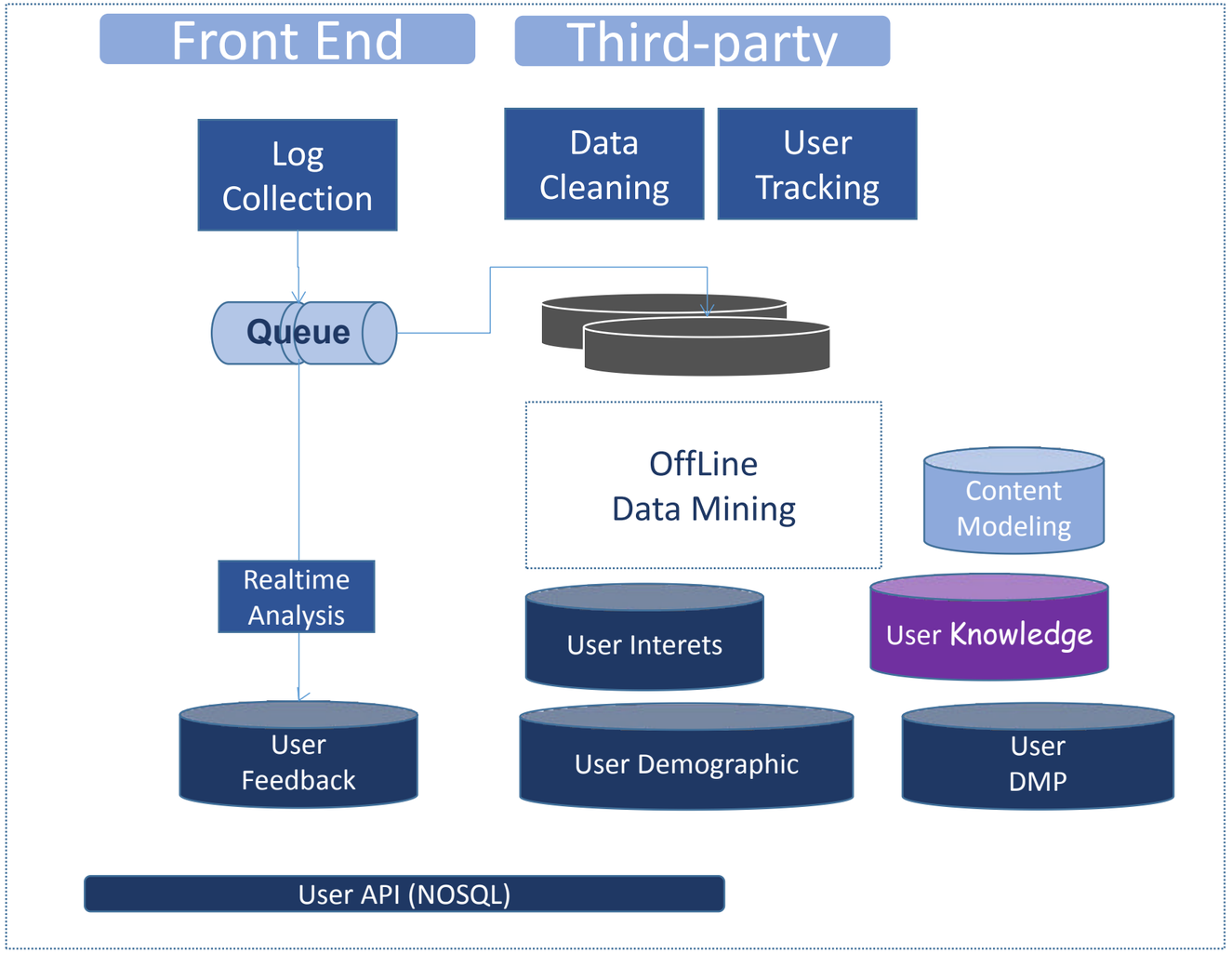
Content Feature：

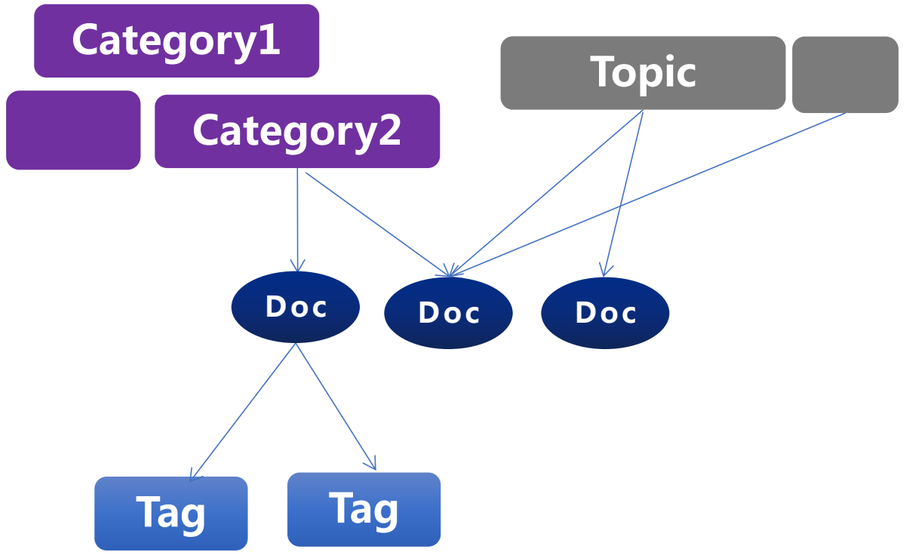
Entity：NBA Player
Category：Sports
Desc：Gossip
Emotion Analyses：Neutral

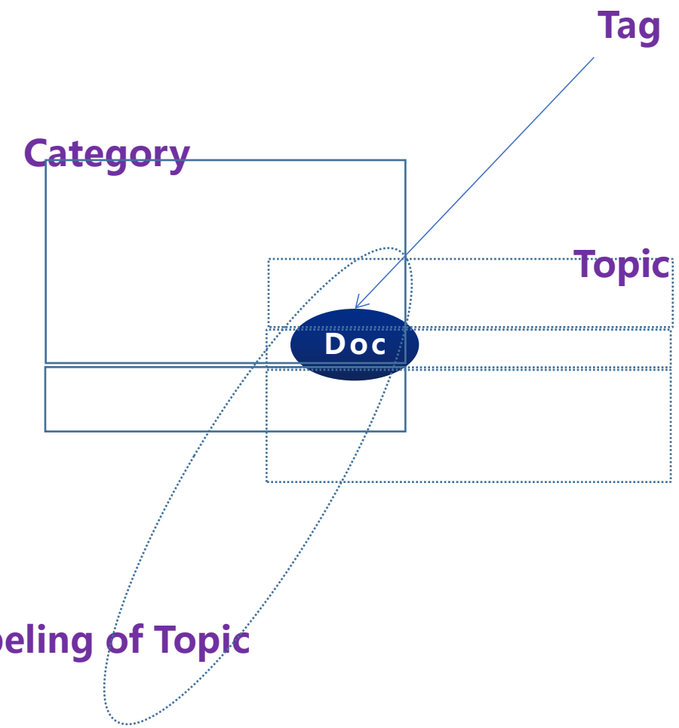
Reading completion rate
90%

```
{
{"_id":"xxx"},
{"M":"StarGossip","S":"1"},
{"Cond":[
{"tag": "Wangxxx|",
"category": "ent|sports",
"topic":"123|221"}
]}
}
```

	TimeLiness	Storage	latency	Data Src
Feedback	Real-Time	rockdb	local (2ms)	behavier
Interets	Near Real-Time	redis	remote (10ms)	behavier mining
Demography	Non-Real Time	redis	remote (10ms)	mutil src mining
Knowledge	Near Real-Time	Mongodb	remote (20ms)	Content + User Modeling



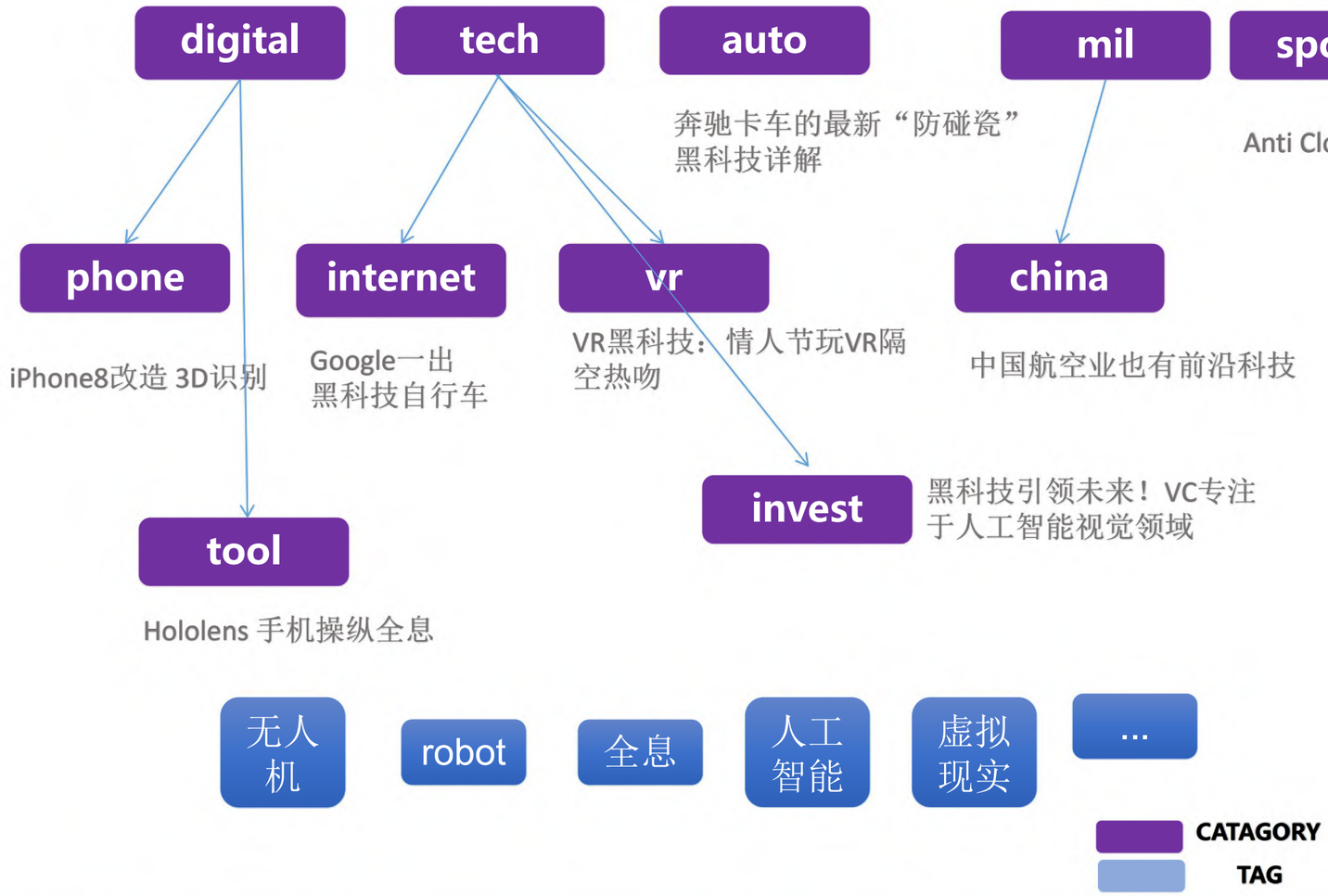




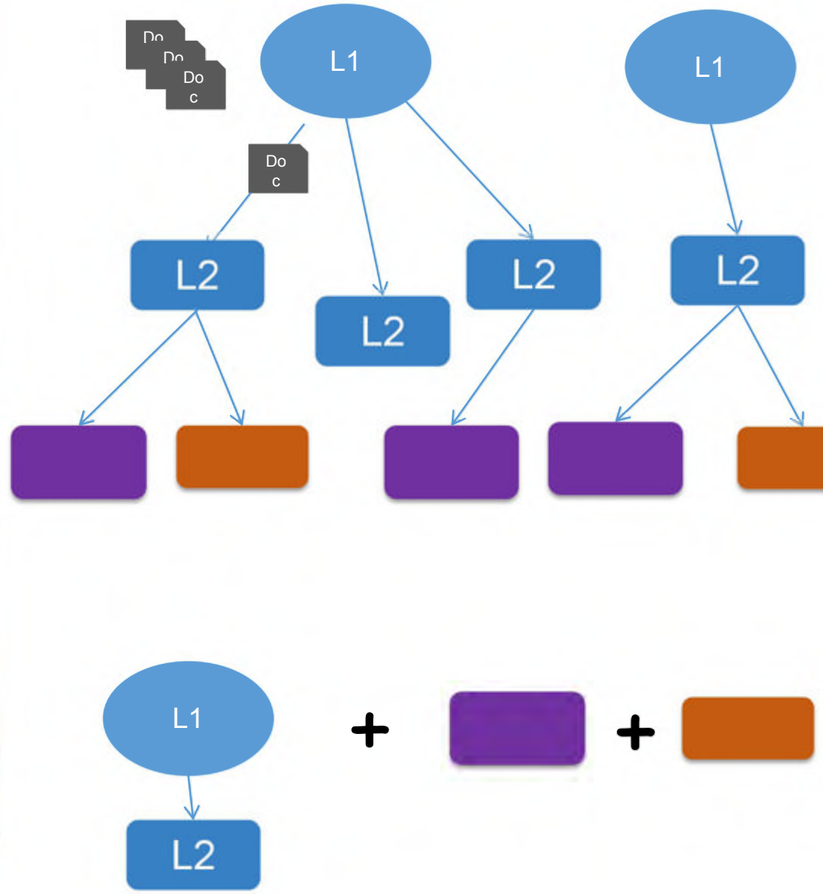
Category : Client eg“navigation menu” , Feature
Topic : Feature , User Profile， Retrieval
Tag : Retrieval , User Profile

Labeling of Topic : Client , Feature , User Profile, Retrieval
eg:Future Technologies

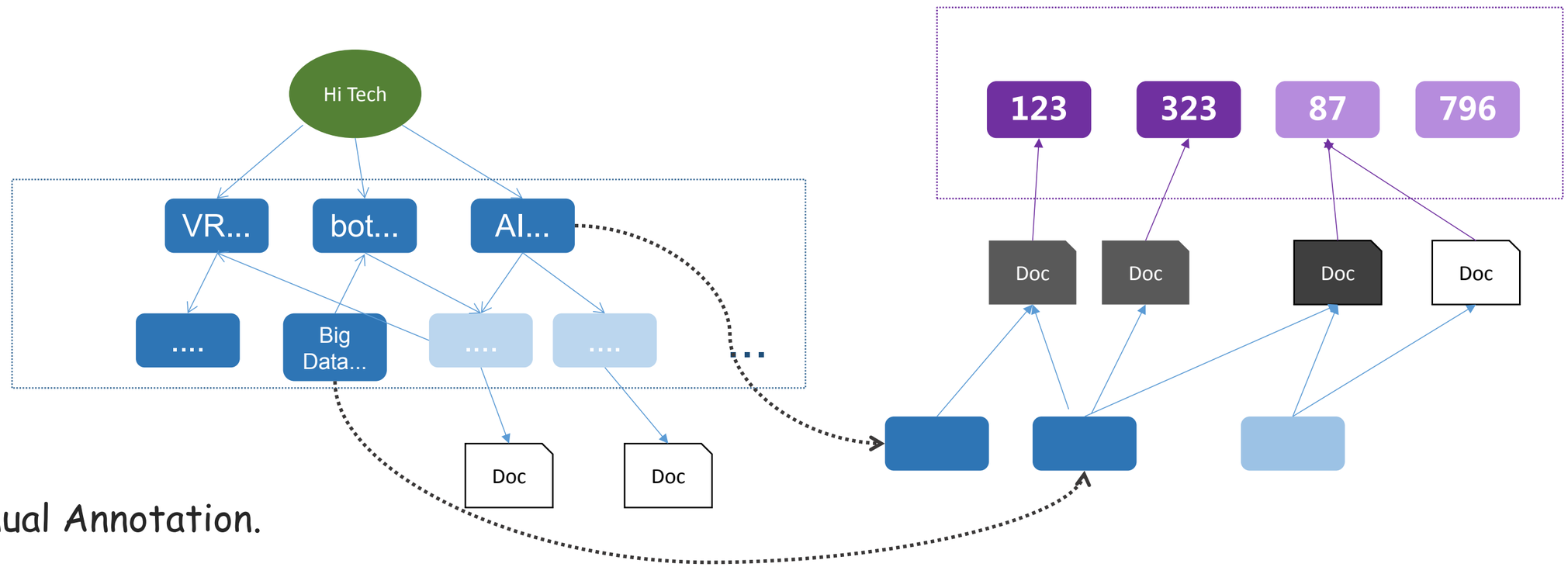
Labeling of Topic : high tech



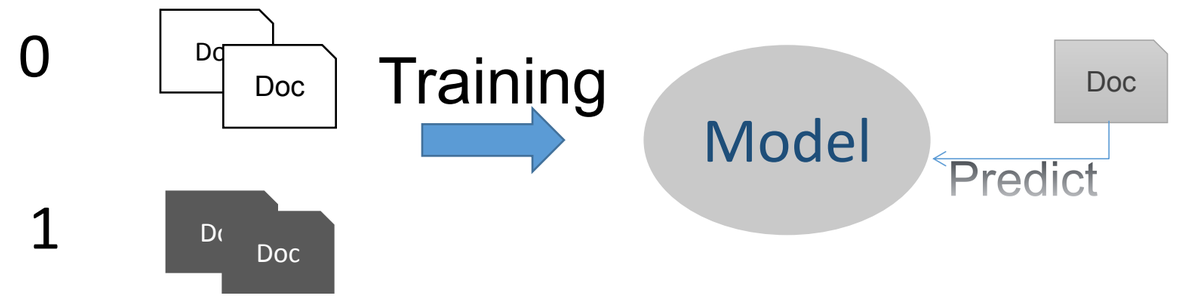
Labeling of Topic: Another dimension to describe content



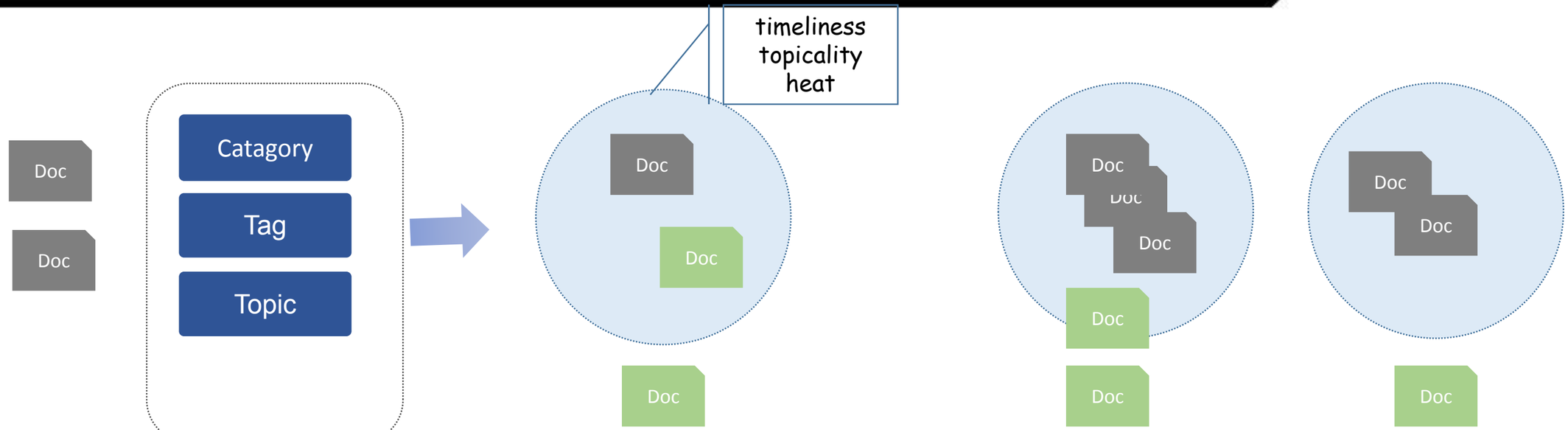
Content Modeling

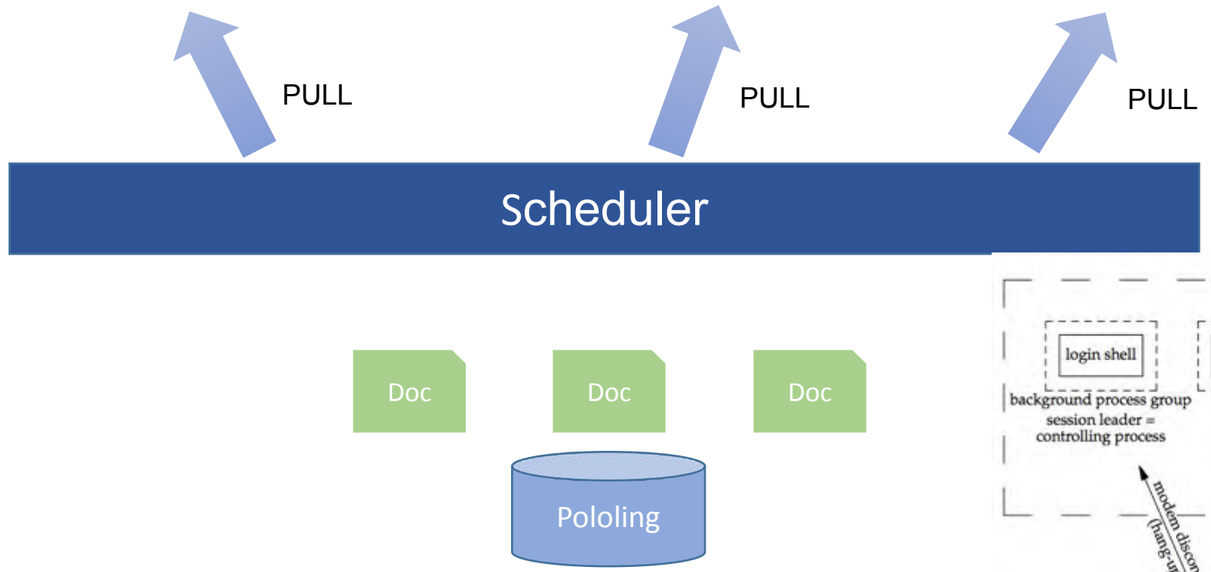


Auto Training.
No Needs Manual Annotation.

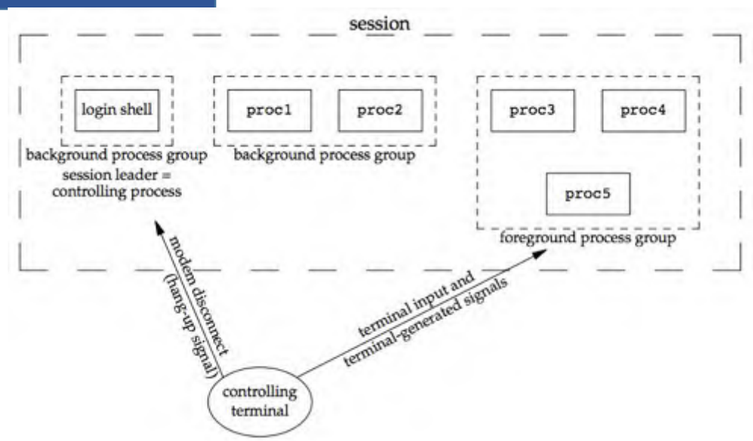


Content Manager System

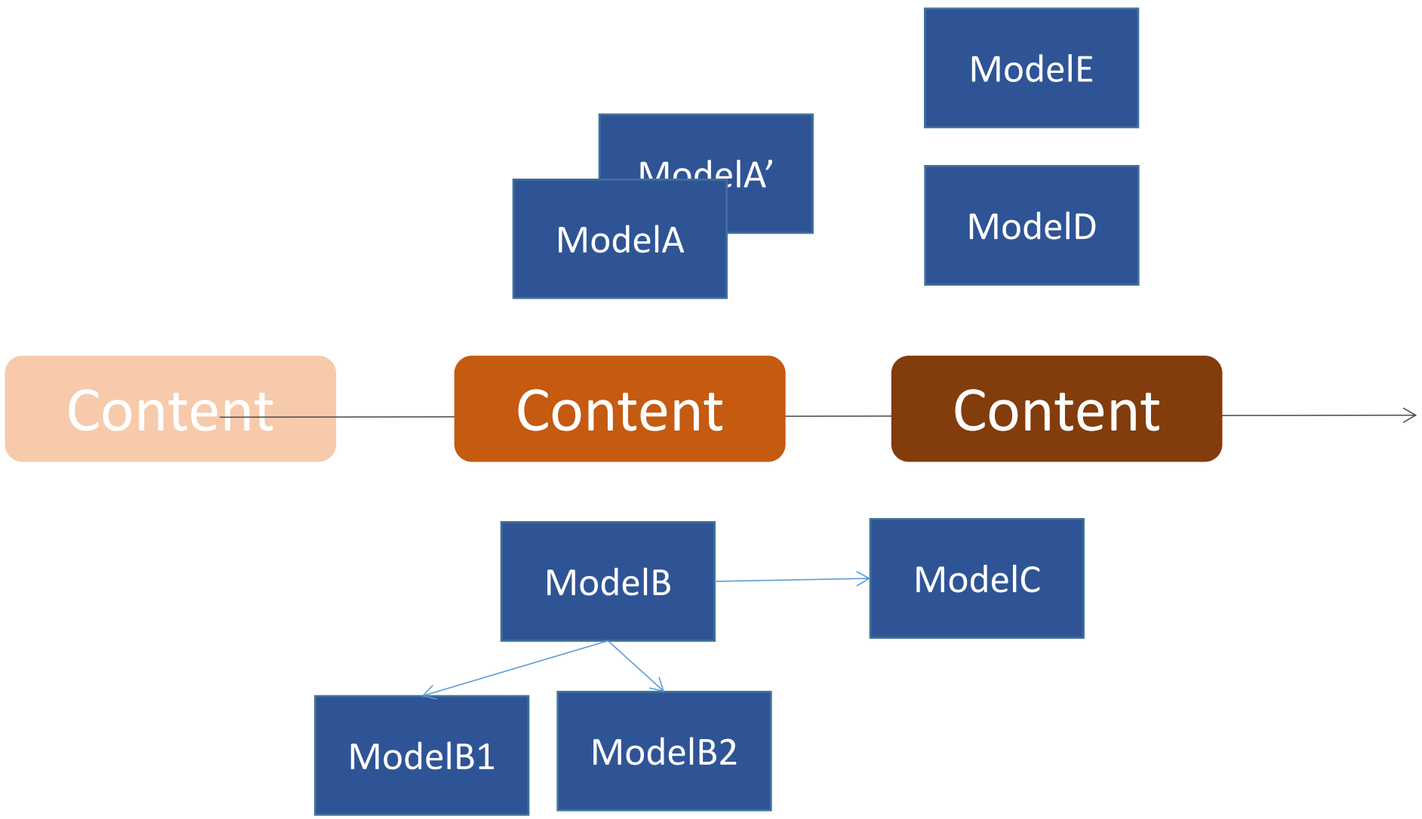


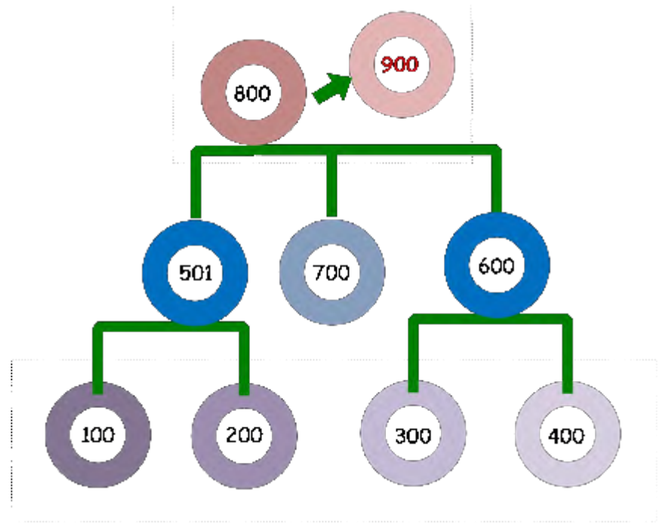


Group Modeling.
Attrs inherit from Group.
Ranking by ContentQuality.
Scheduler Contorl Content taste.



ContentQuality = Document Score + Media Score + PCTR + ...





Switch Model

Configure

```
<switch name  "st_xxx", scan_time = "600" >
                    <dependence switch = "st_name">
                    <dependence switch = "st_name2">
</switch>
<switch name  "st_vvv", scan_time = "200" >
                    <dependence switch = "st_xxx">
                    <dependence switch = "st_name2">
</switch>
```

Welcome Join Us

NLP, ML , DL
Text,Photo,Video
Search,Recomm,Ads