

Distributed Time Travel for Feature Generation

DB Tsai

March 24, 2016 at SF Big Analytics Meetup

NETFLIX

Who am I?

- I am a Senior Research Engineer at Netflix
- I am an Apache Spark Committer



in die ire furoris sui. **Mem** De recessu
miste ignem in ossibus meis: et exarduit
me. **Ep** adit rege pedibus meis: con
uerit me retrorsum. Posuit me deso
latur: tota die in uero confecta. **Num**
Vigilauit iugum iniquitatis meae in
manu eius: conuoluit surum et imposi
t collo meo. Infirmata est uir: mea:
dedit me dominus in manu de qua non po
tero fugere. **Samech** Abskult omnes
magnificos meos dominus de medio meo:
uocauit aduersum me reprobos: ut con
tereret electos meos. Torcular calcauit
dominus uirginem filie iuda. **Ain** Adcirco e
go plorans et oculum meum deducens
aquas: quia longe factus est a me co
solator: conuerrens anima mea. Facti
sunt filij mei pedit: quoniam inuoluit
inimicum. **He** Ep adit syon manus
suas: non est qui consoletur eam. Man
dauit dominus aduersum iacob: in reuer
tu eius hostes eius. Facta est iherusalem pol
luta meretricis inter eas. **Sade** Iustus
est dominus: quia os eius ab iniquitatem
promouit. Audite obsecro universi
populi: uidete dolorem meum. Virgines
mee et inuenies mei abierunt in captiui
tatem. **Coph** Vocauit amicos meos: et
ipsi deprece me. Sacerdotes mei et
senes mei in uice consumpti sunt: quia
querelatur ab eis sibi ut recedant ani
mas suas. **Res** Vide dominus quoniam
rebulor: rebulor est uenit me. Sub
uersum est cor meum in mentis: qui
amantudine plena sum. Foris inter
fecit gladius: et domi mors similis est.
Sen Audierunt quia ingenuis ego: et
non est qui consoletur me. Omnes in
imici mei audierunt malum meum: letati
sunt quoniam fecisti. Adduxisti de consola
tione: et huc sicut me. **Tau** Ingredi
at omne malum cor meum: et inuenerunt

ros sicut uindemasti me propter omnes
iniquitates meas. Multa enim geni
tus mei: et cor meum merco. **Alaph**
Quomodo obrepit caligine et furo
re tuo dominus filia syon. Propter
de celis in terram iudicium israel: et non
est recordatus scabelli pedum suorum in die
furoris sui. **Seth** Precipitauit dominus nec
propter omnia speciosa iacob: destrux
it in furore suo munitiones uirginis
iuda: et dixerat in terra: polluit regnum
et principes eius. **Gymel** Contergit in
ira furoris sui omne cornu israel: a
uerit retrorsum deprece sua a facie inimi
ci: succendit in iacob ignem flammam
deuorans et giro. **Delech** Incendit ar
cum suum quasi inimicus: firmavit de
struam suam quasi hostis: et occidit o
mnem quod pulcrum erat uisui. In taber
naculo filie syon effudit ignem indi
gnationis suae. **He** Factus est dominus uelut in
imicus: precipitauit israel. Precipita
uit omnia membra eius: dissipauit mu
nitiones eius: et repleuit in filia iuda
humiliatum et humiliata. **Yau** Et
dissipauit quasi ortum retrorsum suum:
demoluit tabernaculum suum. Obli
uioni tradidit dominus in syon festiuitates
et sabbatum: et in obprobrium et in indi
gnationem furoris sui regnum et sacerdotem.
Say Repulit dominus altare suum: maledi
xit sanctificationi sue: tradidit in ma
nus inimici muros ierusalem. **Voce**
deduxit in domum domini: sicut in die so
lenni. **Yeth** Cogitauit dominus dissipare
murus filie syon: retrahit funiculum suum
et non auertit manum suam a peditone.
Iuxta autem murale: et murus pariter dis
sipatus est. **Teth** Desere sunt in terra por
te eius: peditur et contumit uerbes eius: re
gum eius et principes eius in gentibus.
Non est lex: et propheta eius non inueniunt

uisionem a domino. **Jeth** Rediit in terra:
conuerit fenestrae filie syon: conspexit
dixerat capita sua. Placuit sunt aliqui
aduerit in terra capita sua uirgines
iherusalem. **Coph** Desere sunt in lacrimis
oculi mei: conuerit sunt uiscera mea.
Fissum est in terra cor meum super con
tinuam filie populi mei: cum desereret paru
lus et lactes in placentis oppidi. **Lamech**
Mentibus suis dixerat ubi est inimicus te
uimus: cum desereret quasi uulnera in
placentis ciuitatis: cum regalaret aias
suas in sinu maris suarum. **Mem** Cui
comparabo te: uel cui assimilabo te fi
lia iherusalem: Cui requirabo te et con
solabor te uirgo filia syon? Magna
enim uelut mare contitio tua. Quis
medebitur tui? **Num** Prophete tui ui
derit tibi fallia et multa: nec aperiebant
iniquitatem tuam: ut te ab impietatem
promoueret. Viderit autem tibi assump
tiones fallas et reditiones. **Samech**
Plaustrum super te manibus omnes tras
ueres per uiam: sublaueat et moueat ca
pita sua super filiam iherusalem: hecne est uerbo
dixerat pedit decori: gaudium uniuersae ter
re. **Ain** Apparet super te os suum omnes
uisioni tu: sublaueat et tenuerit den
tibus suis: dixerit deuorabit. **Am**
Ista est dies que expectabam: inue
nimus uidimus. **He** fecit deus que
rogauit: compluit seminum suum que
propter a diebus antiquis. Desere et
non propter: leuauit super te inimicum:
et regalauit cornu hostium tuorum. **Sade**
Clamauit cor eorum ad dominum: super ini
cos filie syon. Deduxit quasi torquentem
lacrimas per diem et noctem: non des
quit tibi: neque taceat pupilla oculi
tui. **Coph** Consurge: lauda in nocte
in principio uigiliarum: et funde sicut
aqua cor tuum ante conspectum domini.

Ite ad rum manus tuas per animam
paruos tuos: qui fame preter in ca
pire omnium compios. **Res** Vide dominus
et considera: quem uindemasti iuda.
Ergo ne comedat mulieres fructum suum:
paruos ad mensuram palme: Si re
nditur in laudario domini: sacerdos et
propheta? **Sen** Aquirit in terra for
puit et fruges: uirgines mee et inuenies
mei cedebat in gladio interfecti in die
furoris tui: pulchritudo: nec inuenies eos.
Tau Vocasti quasi die solemem
qui tenebat me de ciuitate: non fuit
in die furoris domini qui effugeret et reli
queretur. Quos eduxit et equauit: in
imicus meum consumpsit eos. **Alaph**
Ego uir uidemus paupere uisus: in
uenga indignationis eius. **Alaph**
Iste minauit et adduxit in tenebris: et
non in luce. **Alaph** Tenui in me ueniam
conuerit manum suam contra die. **Seth**
Uersum fecit pelle mea: et carne mea:
continuit ossa mea. **Seth** Edificauit
in gurgitibus: et circubedit me felle et la
tore. **Seth** In tenebris collocauit
me: sicut mortuos sempiternos. **Gymel**
Circubedit me aduersum me ut non
egredior: aggregauit compedem meam.
Gymel Redi et clamauit et cum roga
ro: egredietur orationem meam. **Gymel** Co
dixit uias meas lapidibus quadris:
semitas meas subuertit. **Delech** Uerbi
inuidias factus est michi: leo in abscondito.
Delech Semitas meas subuertit: con
tergit me: posuit me desolatam. **Delech**
Incendit arcum suum: et posuit me quasi
signum ad sagittam. **He** Ipsi in tem
bus meis filias pharetrae. **He** Fa
ctus sum in desolatum quoniam populo meo: tan
ticum eorum tota die. **He** Repleuit me a
magnitudinibus: inebriauit me abin
dicio. **Num** Et fregit ad numerum datus





NETFLIX ORIGINAL

MARVEL

JESSICA JONES

Hell's Kitchen just got another hero.

She's a gifted PI hiding a past that's left her broken but still fighting. Revenge may be sweet, but she isn't.



PLAY

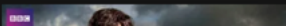
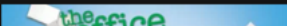
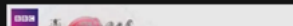


MY LIST

Popular on Netflix



Trending Now



Suspenseful TV Sci-Fi & Fantasy

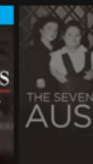
Based on your interest in:



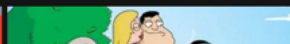
Trending Now



Recently Added



Top Picks for Aish





< THE FLASH >

2015 TV-PG 1 Season

It's showdown time in Central City as an ordinary police scientist suddenly becomes a lightning-fast superhero.

MY LIST EPISODES



TV Shows





The Twilight Zone (Original Series)

98% match 1963 TV-PG 4 Seasons

Hosted by creator Rod Serling, this groundbreaking anthology series presents tales of the supernatural that are often tinged with social commentary.

Starring: Rod Serling

Genres: TV Shows, Classic TV Shows, Cult TV Shows

This show is: Ominous, Scary, Chilling, Suspenseful



This series earned an Emmy for Best Writing, featuring stars like Robert Redford and William Shatner.



MY LIST



MY RATING

OVERVIEW

EPISODES

MORE LIKE THIS

DETAILS

< THE FLASH >

2015 TV-PG 1 Season

It's showdown time in Central City as an ordinary police scientist suddenly becomes a lightning-fast superhero.



MY LIST



EPISODES

TV Shows



Trending Now



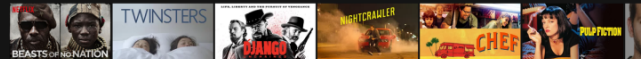
Recently Added



Top Picks for Aish



Critically-acclaimed Movies



Continue Watching for Aish



New Releases



Because you watched Jiro Dreams of Sushi



Turn on Netflix, and the **absolute best content for you** would **automatically start playing**

NETFLIX

Data Driven

- Try an **idea offline using historical data** to see if it would have made better recommendations



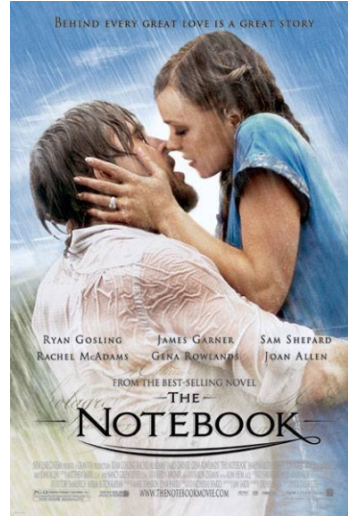
- If it did, deploy a live **A/B test** to see if it performs well in Production

Quickly try ideas on **historical data** and
transition to online A/B test

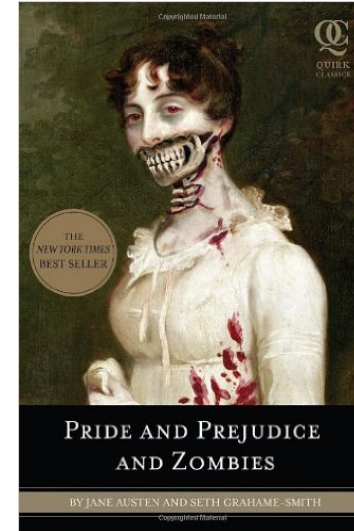
Feature Engineering



+



≠



Why build a Time Machine?

Label Leaks

- Recommendation is an analytic learning process to learn from the past to predict the future.
- $P(\text{play}_t | \text{features}_{t'})$ where $t' < t$
- This has to be very careful; otherwise the features can contain the labels such that offline metrics are very good, but online evaluation will not be performing as expected.

The Past

- Generate features based on event data logged in Hive
 - Need to reimplement features for online A/B test
 - Data discrepancies between offline and online sources
- Log features online where the model will be used
 - Need to deploy each idea into production
- Feature generation calls online services and filters data past a certain time
 - Works only when a service records a log of historical events
 - Additional load on online services



DeLorean image by JMortonPhoto.com & OtoGodfrey.com

NETFLIX

Time Travel using Snapshots

- Snapshot online services and use the snapshot data offline to generate features
- Share facts and features between experiments without calling live systems

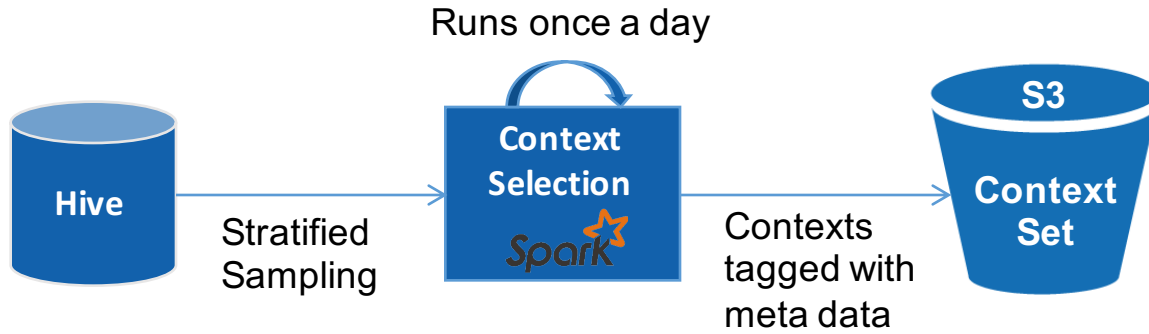
How to build a Time Machine

Context Selection

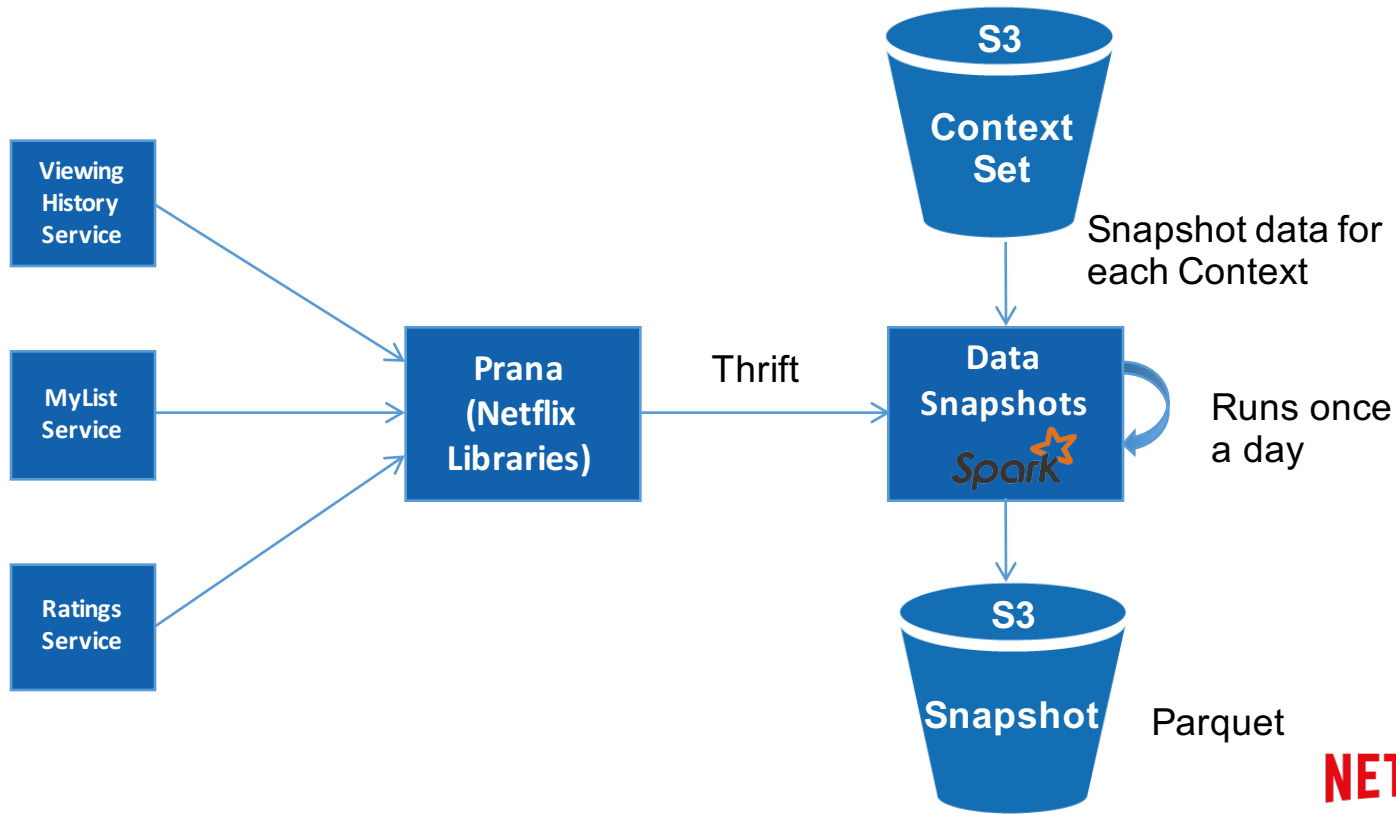
Data Snapshots

APIs for Time Travel

Context Selection



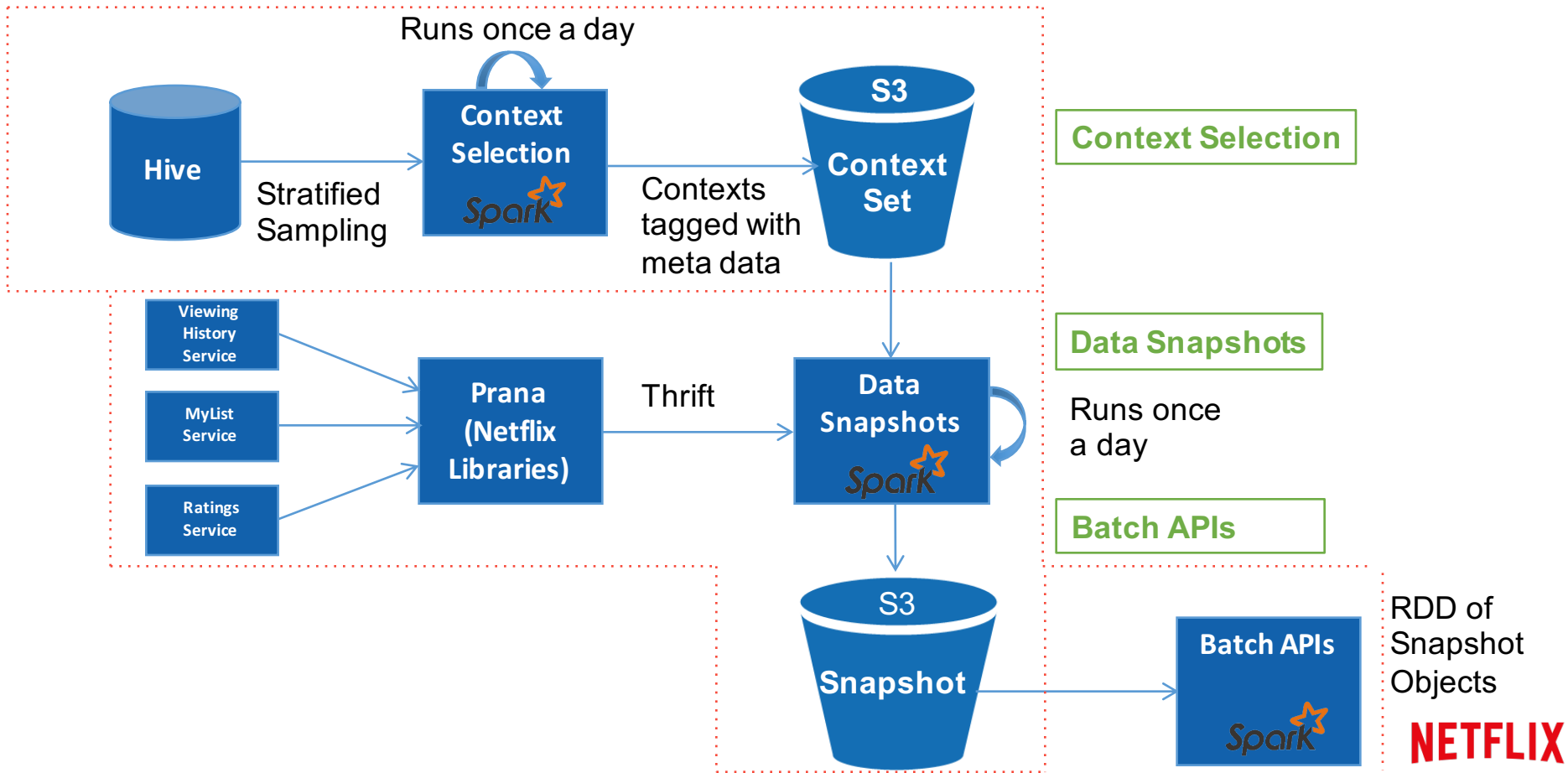
Data Snapshots



APIs for Time Travel

```
scala> val snapshot = new SnapshotDataManager(sqlContext)
      .withTimestamp(1445470140000L)
      .withContextId(OUTATIME)
      .getViewingHistory
snapshot: org.apache.spark.rdd.RDD[(Long, com.netflix.viewinghistory.ViewingHistory)]
```


Data Architecture



Generating **Features** via **Time Travel**

Great Scott! There's the DeLorean!

- DeLorean: A time-traveling vehicle
 - uses data snapshots to travel in time
 - scales with Apache Spark
 - prototypes new ideas with Zeppelin
 - requires minimal code changes from experimentation to A/B test to production

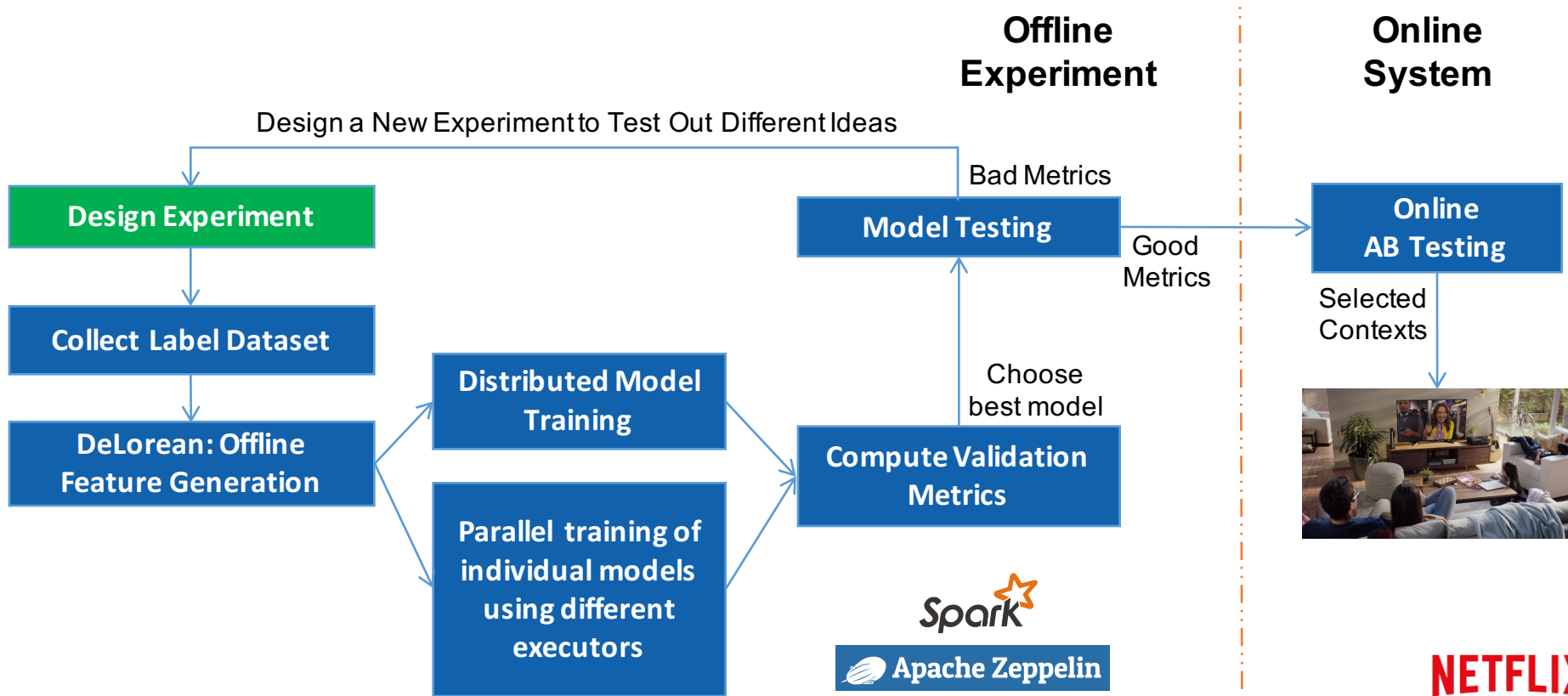


Running Time Travel Experiment

Select the **destination time**

Bring it up to **88 miles** per hour!

Running Time Travel Experiment



DeLorean Input Data

- Contexts: The setting for evaluating a set of items (e.g. tuples of member profiles, country, time, device, etc.)
- Items: The elements to be trained on, scored, and/or ranked (e.g. videos, rows, search entities).
- Labels: For supervised learning, this will be the label (target) for each item.

Feature Encoders

- Compute features for each item in a given context
- Each type of raw data element has its own data key
- Data map is a map from data keys to data objects in a given context
- Data map is consumed by feature encoder to compute features

Two type of Data Elements

- Context-dependent data elements
 - Viewing History
 - Mylist
 - ...
- Context-independent data elements
 - Video Metadata
 - Genre Metadata
 - ...



Context Independent Data Element

Video Country of Origin Matching Fraction


Video Metadata

Features



Context-Items


Context Dependent Data Element

Viewing History



Context: 

Items:

Context: 


Items:




Context: 

Items:

 = 0.5
 = 0.5
 = 0.5

Context: 

Items:


 = 1.0
 = 0.0
 = 1.0

Context: 




Items:



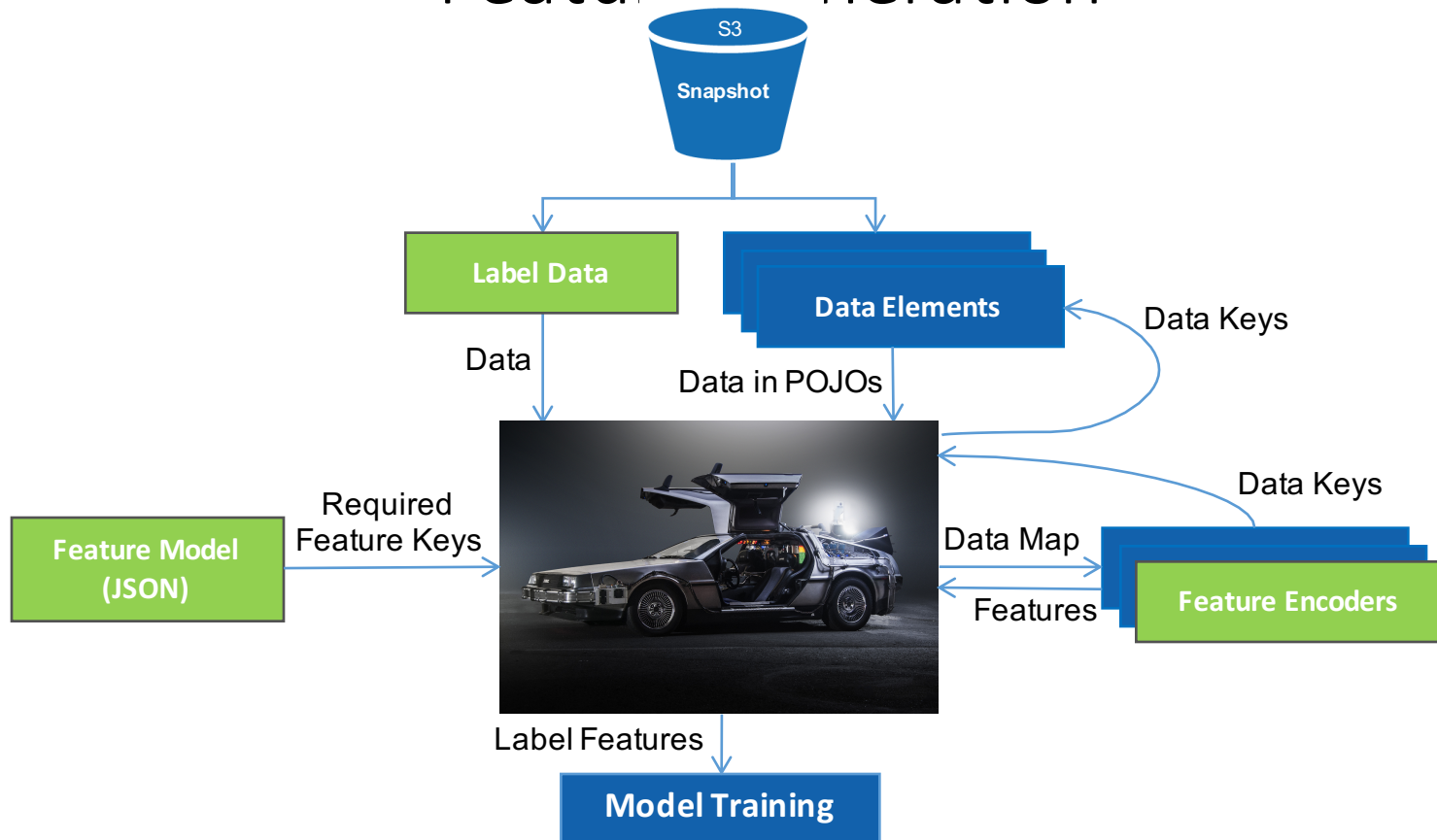


Context: 

Items:

Feature Generation



Features

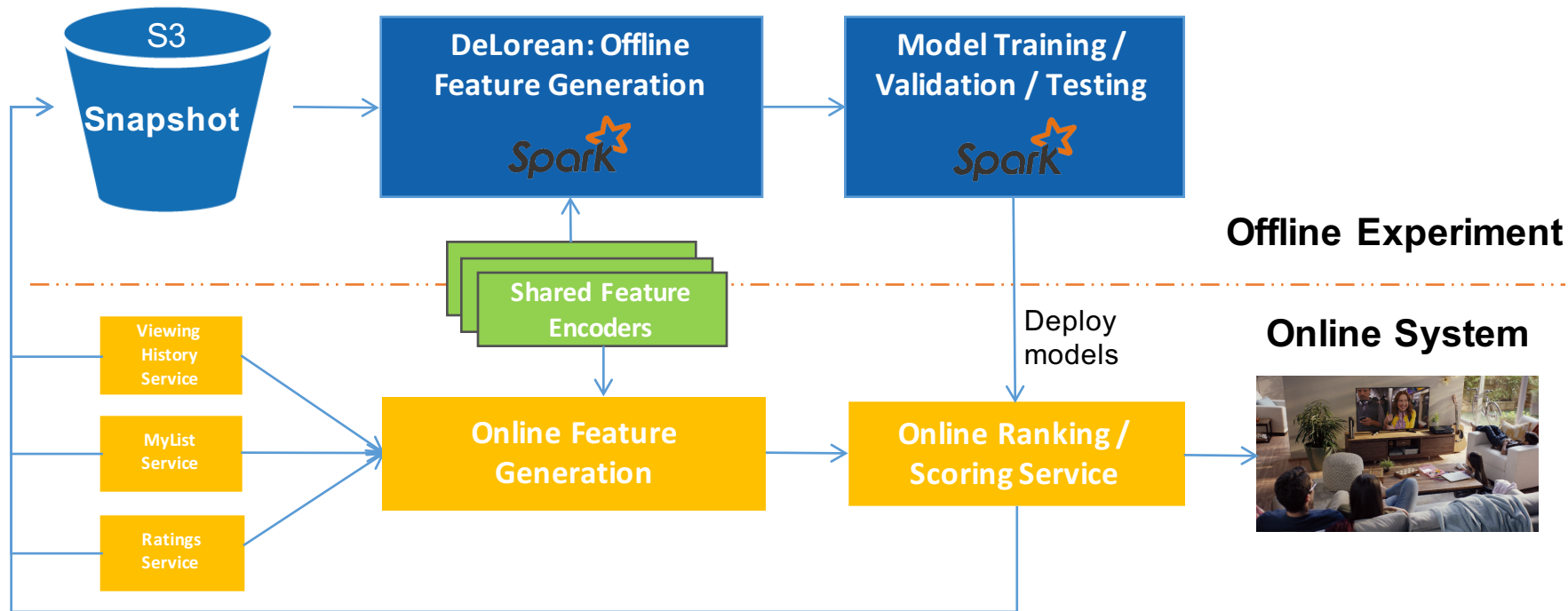
- Represented in Spark's DataFrames
- In nested structure to avoid data shuffling in ranking process
- Stored with Parquet format in S3

Features

```
root
Context |-- Visitor: long (nullable = false)
        |-- Country: string (nullable = false)
        |-- data: array (nullable = true)
        |   |-- element: struct (containsNull = true)
        |   |   |-- videoId: long (nullable = false)
        |   |   |-- weight: double (nullable = false)
        |   |   |-- label: double (nullable = false)
        |   |   |-- features: struct (nullable = false)
        |   |   |   |-- rating: double (nullable = false)
        |   |   |   |-- unpersonalizedPopularity: double (nullable = false)
        |   |   |   |-- ...
        |   |   |   |-- ...
        |   |   |   |-- ...
```

Item, label,
and features

Going Online



Conclusion

Spark helped us **significantly reduce**
the time from an idea to an AB Test

Future work

Event Driven Data Snapshots

Time Travel to the **Future!!**

We're hiring!
(come talk to us)

<https://jobs.netflix.com/>

Tech Blog: <http://bit.ly/sparktimetravel>