talend

HOW NLP AND SPARK CAN ENRICH CUSTOMER DATA

MARK BALKENENDE

MARK BALKENENDE

Director Technical Product Marketing
Talend

@markbalk

talend

Walgreens

Abbott

MOTOROLA

talend CRAFT BEER AND DATA

#CraftBeerandData

AGENDA

01 UNSTRUCTURED DATA

02 NATURAL LANGUAGE PROCESSING

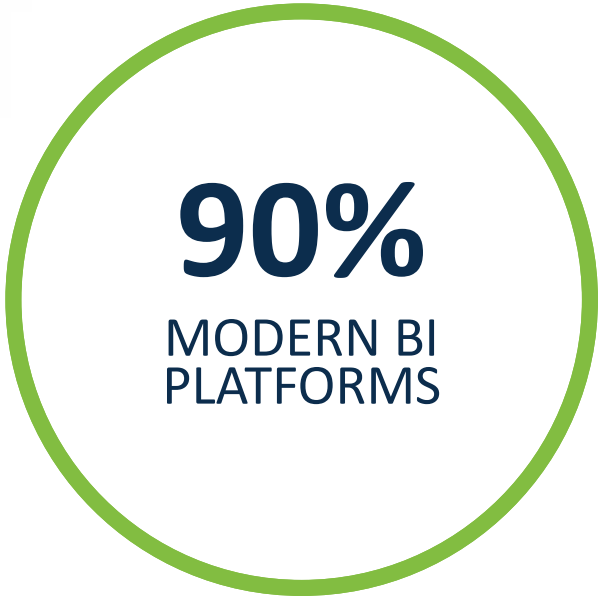
03 WAYS TO IMPLEMENT

04 STEPS TO TRAIN

UNSTRUCTURED DATA - TEXT

talend

UNSTRUCTURED DATA CHANGES NATURE OF ANALYTICS



By 2020, natural-language generation and artificial intelligence will be a standard feature of 90% of modern BI platforms

By 2020, 50% of analytic queries will be generated using search, natural-language processing or voice, or will be autogenerated

Through 2020, the number of citizen data scientists will grow five times faster than the number of data scientists

2015 2016 2017 2018 2019 2020

talend

EXAMPLE

I WORK FOR A COMPANY SELLING

ARTIFICIAL INTELLIGENCE

AN EXISTING OPPORTUNITY IN SALESFORCE

Opportunity
Tesla AI

Hide Feed

Post File Link Poll

Write something...

Share

Follow

Followers

No followers.

Show All Updates

There are no updates.

« Back to List: Contacts

Products [0] | Open Activities [0] | Activity History [0] | Notes & Attachments [0] | Contact Roles [0] | Partners [0] | Competitors [0] | Stage History [1] | Orders [0]

Opportunity Detail

Edit Delete Clone

Opportunity Owner	Mark Balkenende [Change]	**Amount**	$0.00
Private		**Expected Revenue**	$0.00
Opportunity Name	Tesla AI	**Close Date**	8/22/2018

Description In the Facebook Live, Mark Zuckerberg says he opposes the spread of fear surrounding the potential of artificial intelligence.

"I have pretty strong opinions on this. I am optimistic," says Zuckerberg. "I think you can build things and the world gets better. But with AI especially, I am really optimistic.

Created By	[illegible] PM	**Last Modified By**	Mark Balkenende, 8/23/2018 6:39 PM

Description acebook Live, Mark Zuckerberg says he opposes the spread of fear surrounding the potential of artificial intelligence.

"I have pretty strong opinions on this. I am optimistic," says Zuckerberg. "I think you can build things and the world gets better. But with AI especially, I am really optimistic.

Custom Links Delivery Status

NOTES AND DESCRIPTIONS?

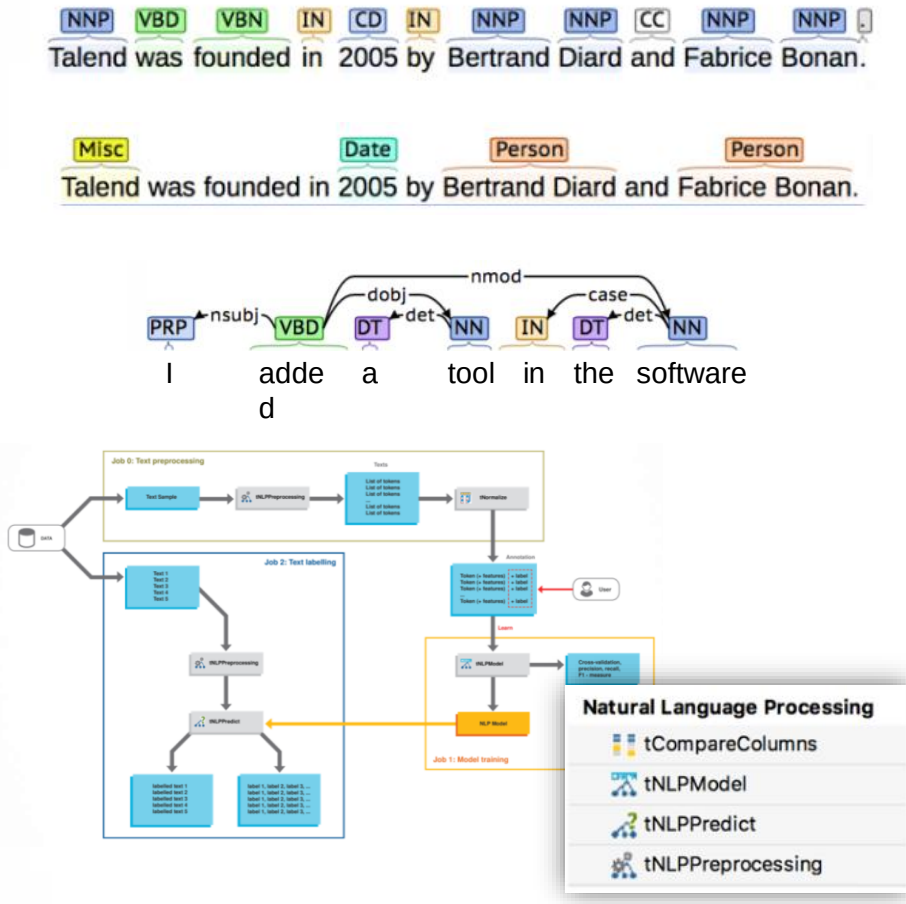
WHAT	What value may exist in open text fields?
WHERE	Is there value in Notes or Comments?
HOW	How do we know if there is value?
PARSE	Can we parse the data for what our company needs?

talend

NATURAL LANGUAGE PROCESSING (NLP)

talend

NATURAL LANGUAGE PROCESSING (NLP)



- Extract useful information
- Finds people names, companies, tools
- Classify discussions
- Entity linking (e.g. persons and organizations linking)

Benefits: Make your integration intelligent. Create new data-driven insights.

GREAT NLP USE CASES

talend

WAYS TO IMPLEMENT NLP

talend

HOW TO IMPLEMENT NLP

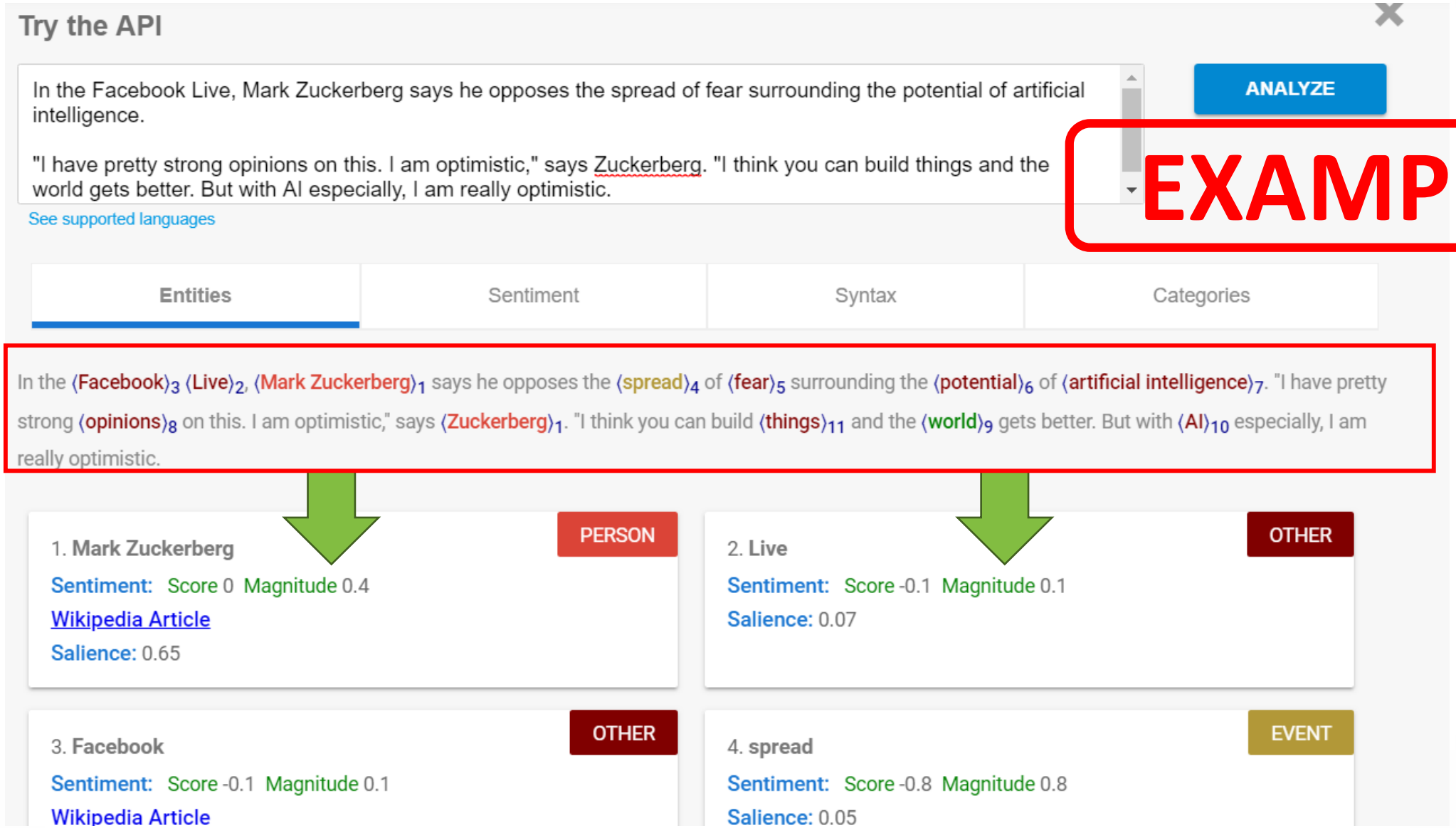
Google
Azure
AWS

Klevu - eCommerce
Insight Engines - Enterprise
MindMeld – ChatBots

Spark ML
Standford
TensorFlow

talend

GOOGLE'S NLP SERVICE - EXAMPLE



TARGET – PREPACKAGED MODELS

agolo Signup Login Developers Careers

What Agolo does

Scan

Ingests public text and internal documents.

Organize

Identifies relationships among sources.

Summarize

Uncovers the most relevant information.

EXAMPLE

See it in action

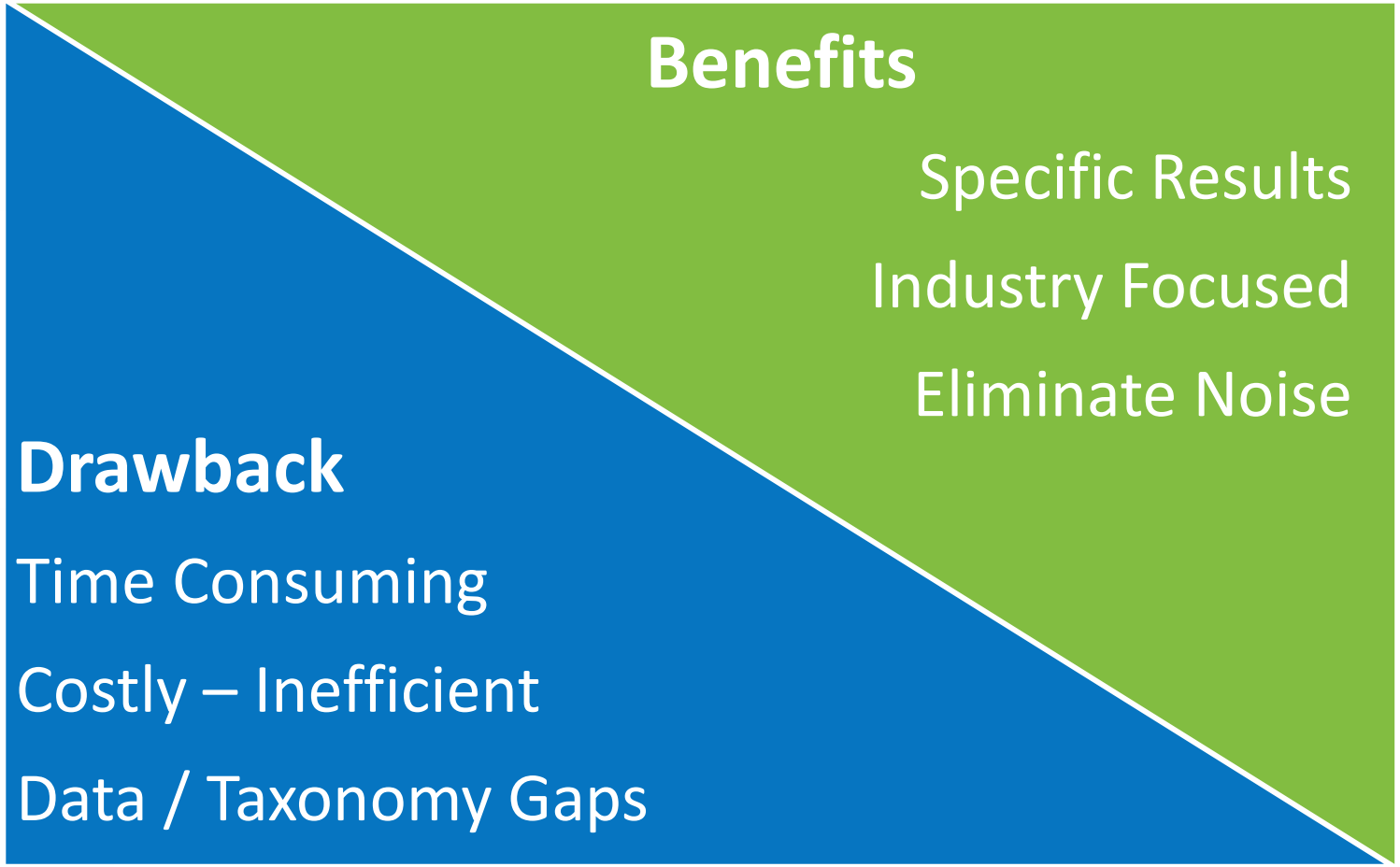
Newsfeed from Reuters	Article from the Wall Street Journal	Internal PDF
Long list of China high-yield issuers due to make May debt payments	China Bondholders Lose Their Beijing Bailout Confidence	People's Bank of China

CLICK ME TO SUMMARIZE

TRAIN AN NLP MODEL

talend

TRAIN THE MODEL – USE YOUR DATA



Frameworks

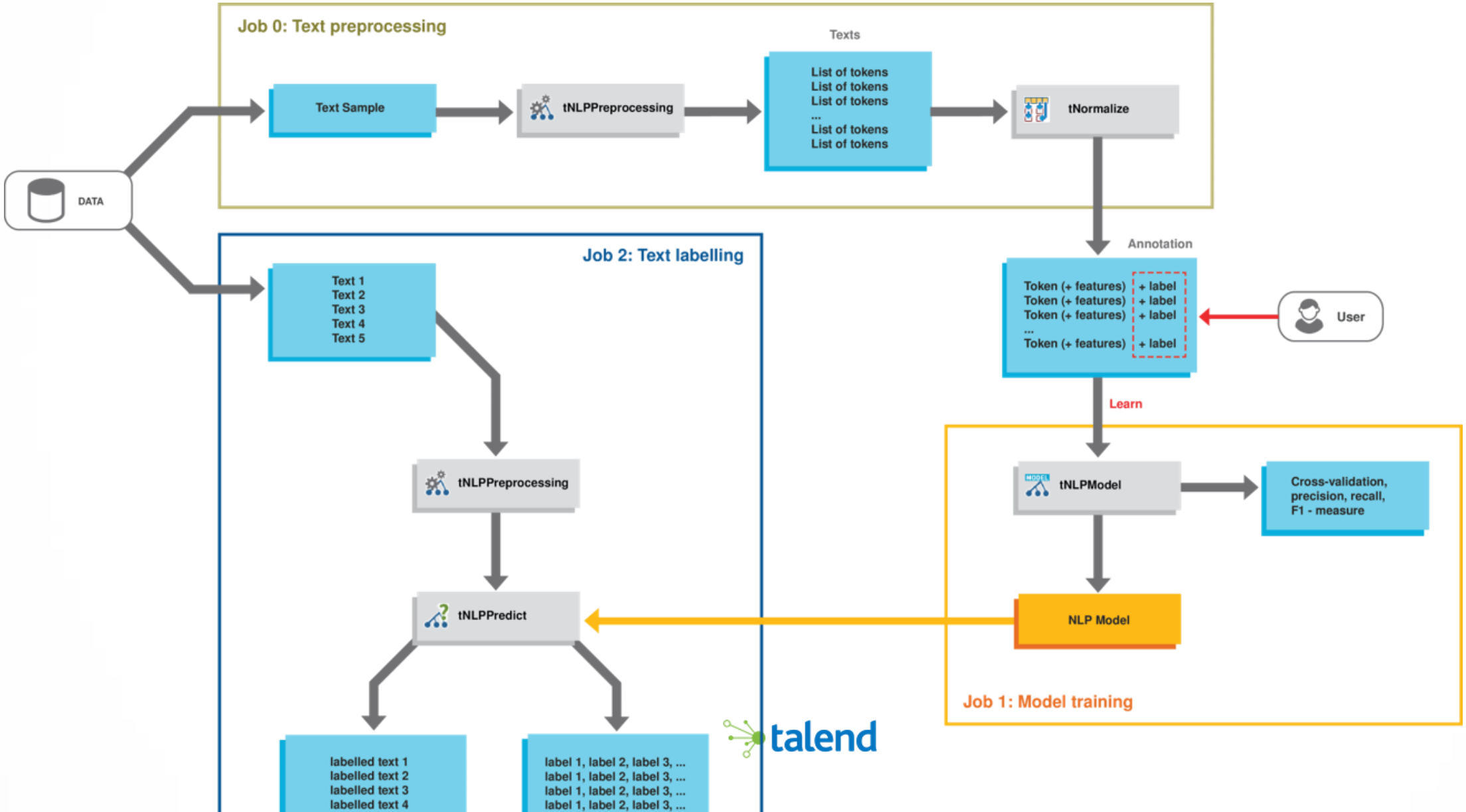
TensorFlow

Stanford's CoreNLP

OpenNMT

talend

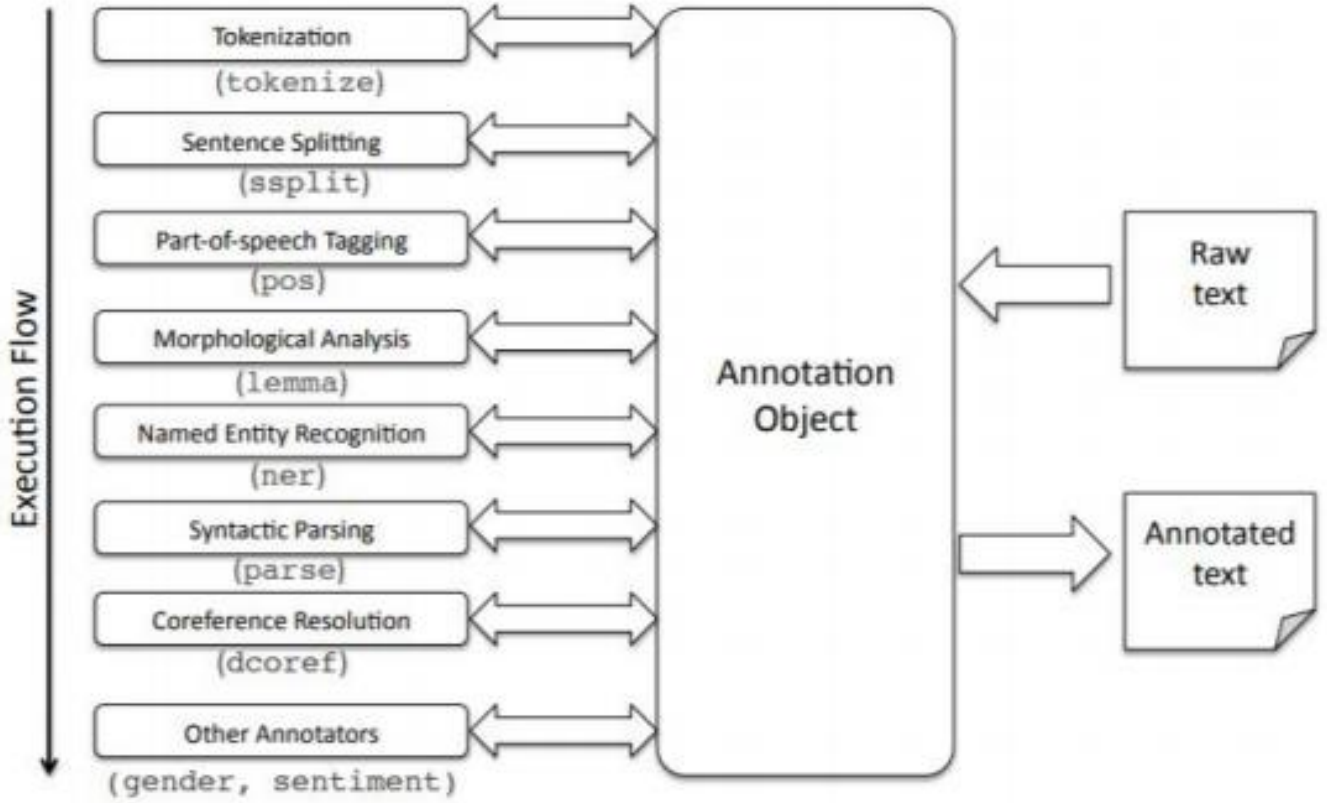
HOW TO TRAIN A NLP MODEL?



BASIC FLOW TO NLP TRAINING

Natural language processing basic tasks include:

- Text tokenization
- Label / Annotations
- Syntax/Semantics/Discourse
  - Sentence splitting
  - Named entity recognition

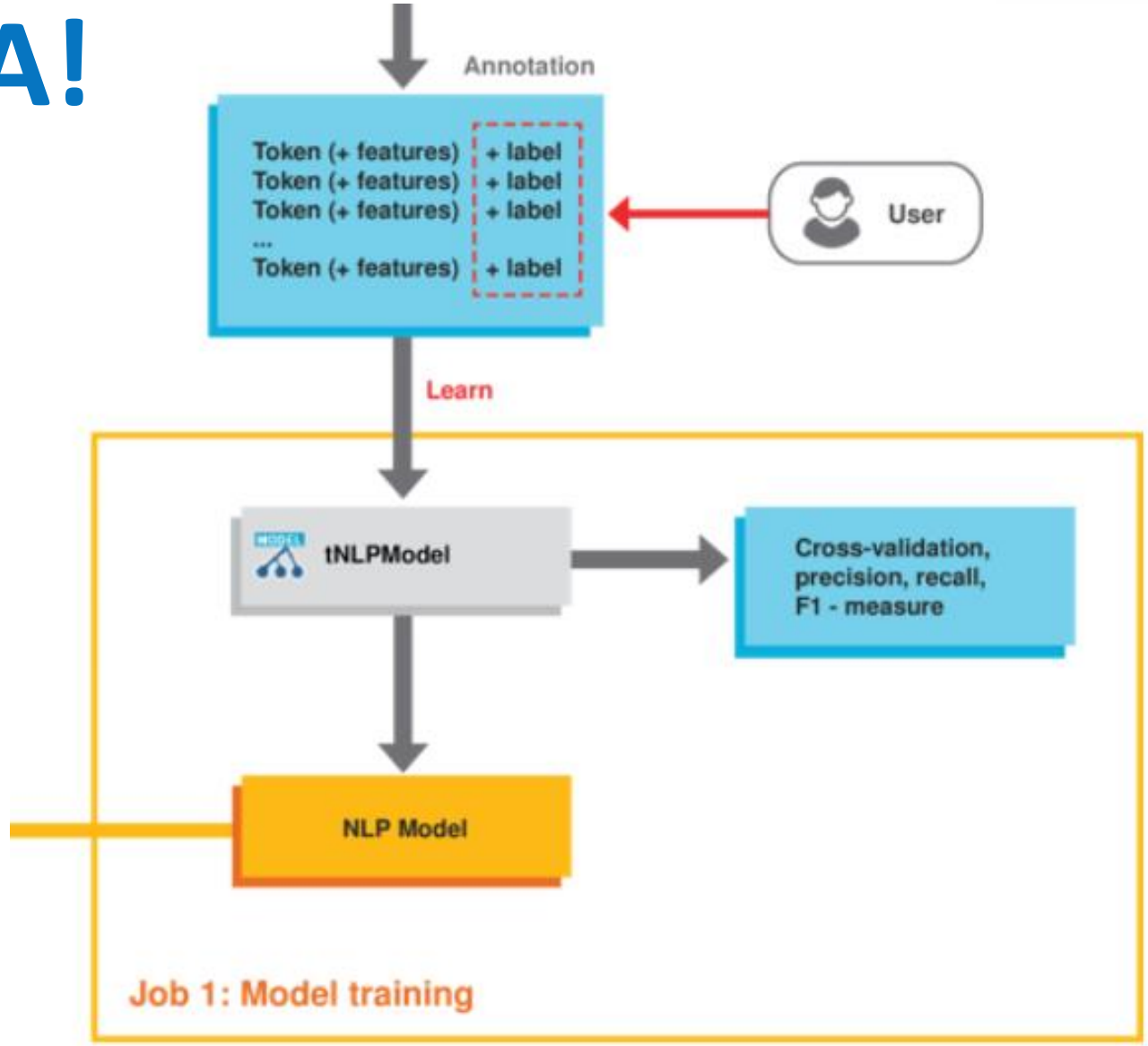


*Stanford CoreNLP

talend

LABEL YOUR DATA!

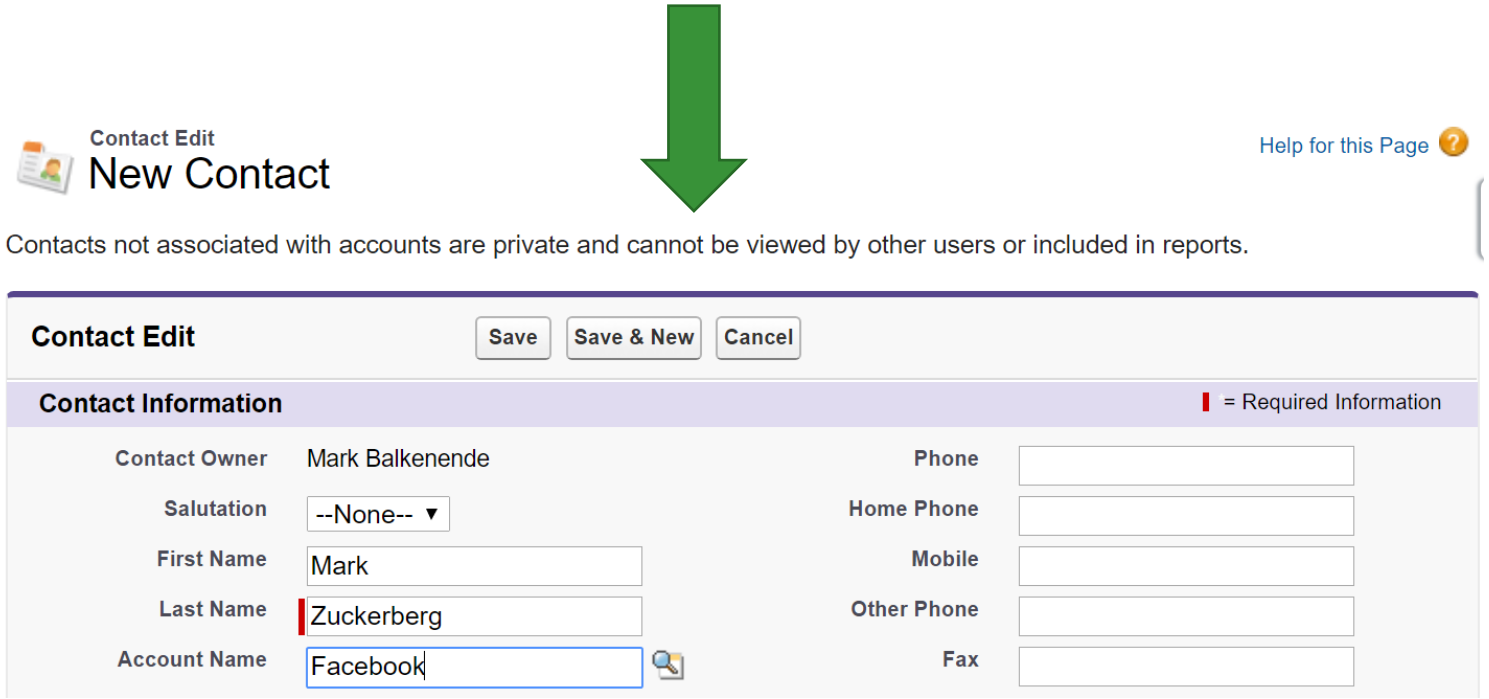
```
using    0.46     F    F    O
Postgres     0.64     F    F    O
with     0.6 F    F    O
Talend   0.56     F    F    O
?    0.0 F    F    O
Best     0.75     F    F    O
regards 0.68     F    T    O
,    0.0 F    F    O
Tim 0.83     F         PER

Hi   0.56     F    F    O
,    0.0 F    F    O
ppm 0.5 F    F    O
install 0.44     F    F    O
Text-CSV_XS 0.51     F    F    O
does     0.61     F    F    O
```



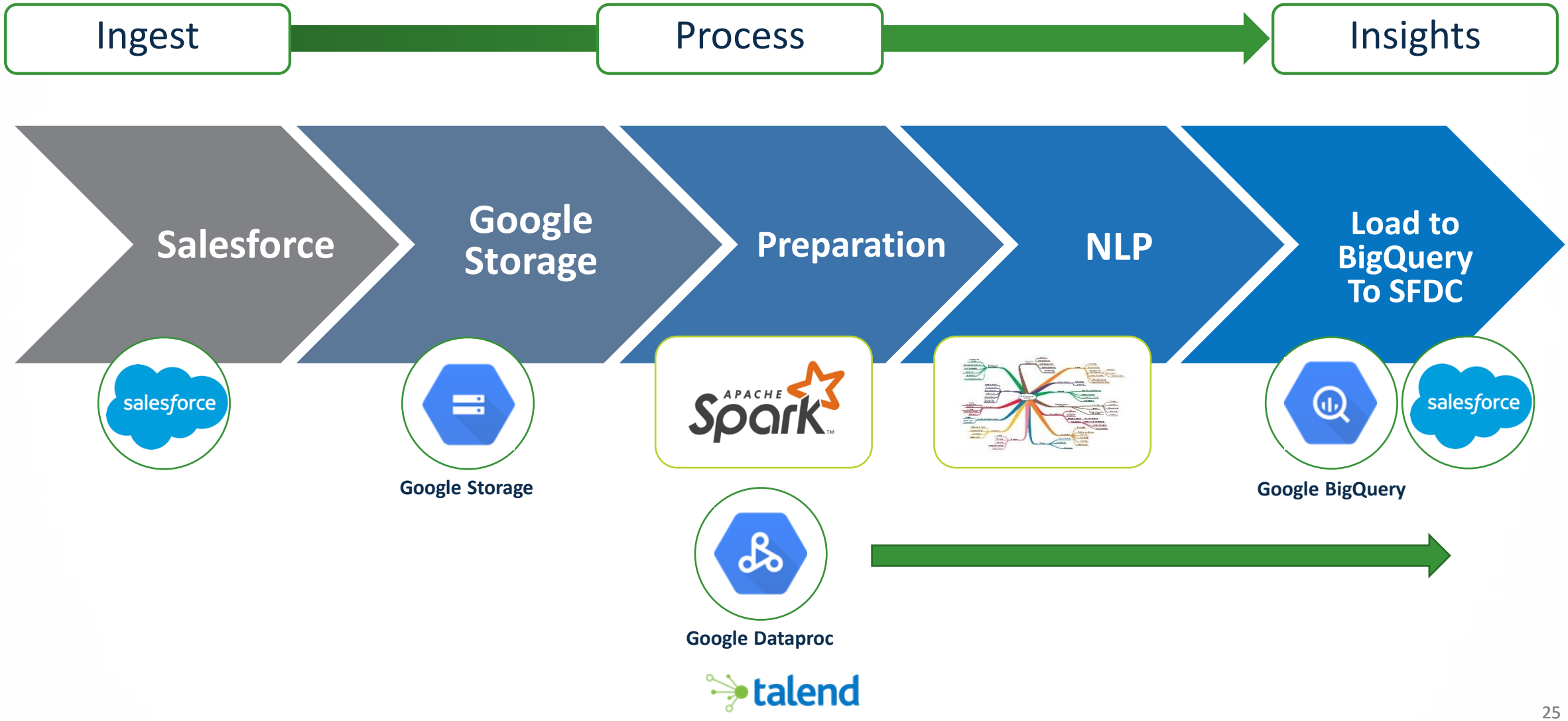
talend

PARSE RESULTS AND LOAD!

```
Grif Thanks for the reply. I did for get to mention all 3 computers are running XP. Two have the Home Ver and 1 the Pro Ver. The one
giving me the problem has the Home Ver and is my "main" desktop computer. The weird thing is the other two (a laptop & 2nd desktop) are
still networked. The "main" desktop has my printer connected to it, so until I can get it "networked" with the others, I can't print
from either of them. Can I assume that since none of them have Vista that I do not need to download and run the program you suggested?
Alan
;<PER>Grif</PER> Thanks for the reply. I did for get to mention all 3 computers are running XP. Two have the Home Ver and 1 the Pro
Ver. The one giving me the problem has the Home Ver and is my "main" desktop computer. The weird thing is the other two (a laptop & 2nd
desktop) are still networked. The "main" desktop has my printer connected to it, so until I can get it "networked" with the others, I
can't print from either of them. Can I assume that since none of them have Vista that I do not need to download and run the program you
suggested? <PER>Alan</PER>
;PER    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   PER
```



CUSTOMER FLOW



TAKE AWAYS

EXTRACT	Get value with all your data – text, voice, you name it
CHOICE	Pick your service or build your own
SPARK	Scale your NLP with existing Spark Platforms
START	Start now with Talend and Spark and NLP

talend