Americas' SAP Users' Group

L3's Innovative Approach to Adapt ML foundation & Text Analysis

Arvind Patel, L3 Technologies
Vivek RR , SAP America Inc

Session ID #84067

About the Speakers

Arvind Patel

- Senior Architect, L3 Technologies
- 7 years of SAP Experience in solution design, architecture, product evaluation and deployment projects such as SAP HCM, SuccessFactors, SAP Analytics Cloud & Digital Boardroom

Vivek RR

- Technology Architect, SAP
- 12 years of SAP Experience focusing on native HANA Developments, XSA , SCP , S/4 Analytics & ABAP Programming

Key Outcomes/Objectives

1. Understanding SAP Leonardo Machine Learning foundation (MLF)
2. Understanding SAP HANA Text Analysis
3. Building hybrid applications using MLF APIs & structuring data using Text Analysis

Agenda

- L3 Technologies – Who we are
- Evaluation Phase
- Technical Evaluations (in detail)
- Summary
- Conclusion

Who We Are – L3 Technologies

- Aerospace & Defense
- L3 develops advanced defense technologies and commercial solutions in pilot training, aviation security, night vision and EO/IR, weapons, ISR market, maritime systems and space.
- $10.2 B Revenue (2018)
- Digital Transformation
  - Becoming data-driven to Intelligent enterprise
- Enterprise Analytics - 2018 SAP Innovation Award winner

ASUG

L3 Technologies – Global Footprint

Business Challenge

Business Issues:

- ✓ L3 is looking for a solution to extract Budget intelligence from a US Defense Department budget report that is published as a PDF
- ✓ To extract key information from Defense related blogs & press releases to identify market trends and upcoming technologies involved in defense related areas

Key Benefits

- ✓ Ability to mine information & use as part of forecasting and planning process
- ✓ Simplify the Budget Intelligence reporting process
- ✓ Automating E2E process during Data Intelligence platform & capabilities
- ✓ Use key market trends to pursue new business opportunities

Our Process

1. Identify Customer Pain Points, Value and Desired Outcomes
2. Engage SAP Max Attention Innovation Channel to Explore Use Cases
3. Design Proof of Concept
4. Provide Data for Evaluation
5. Identify Solutions
6. Review Solutions and Explore Cost
7. Access Architecture Needs
8. Determine Next Steps

Evaluation Phase – Use Case 1

Use Case 1 – DoD Budget Marks

Objective:

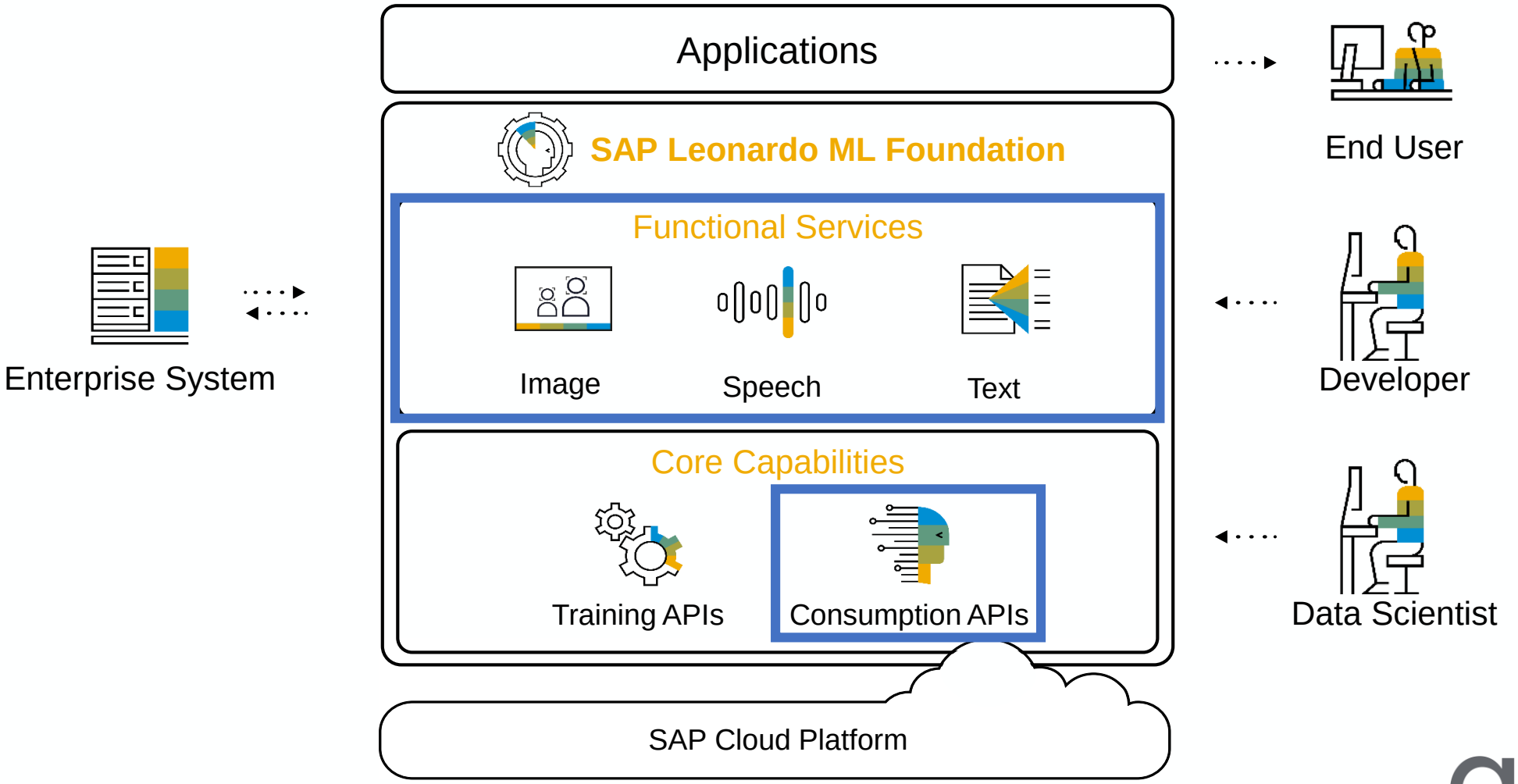
- To extract program elements(PE) and the associated information from defense budget reports
- To compare PE data against the master data stored in HANA
- Capture key differences as change in budget, and reasons

Scope of Evaluation:

- SAP Leonardo Machine Learning Foundation – OCR Functional Service

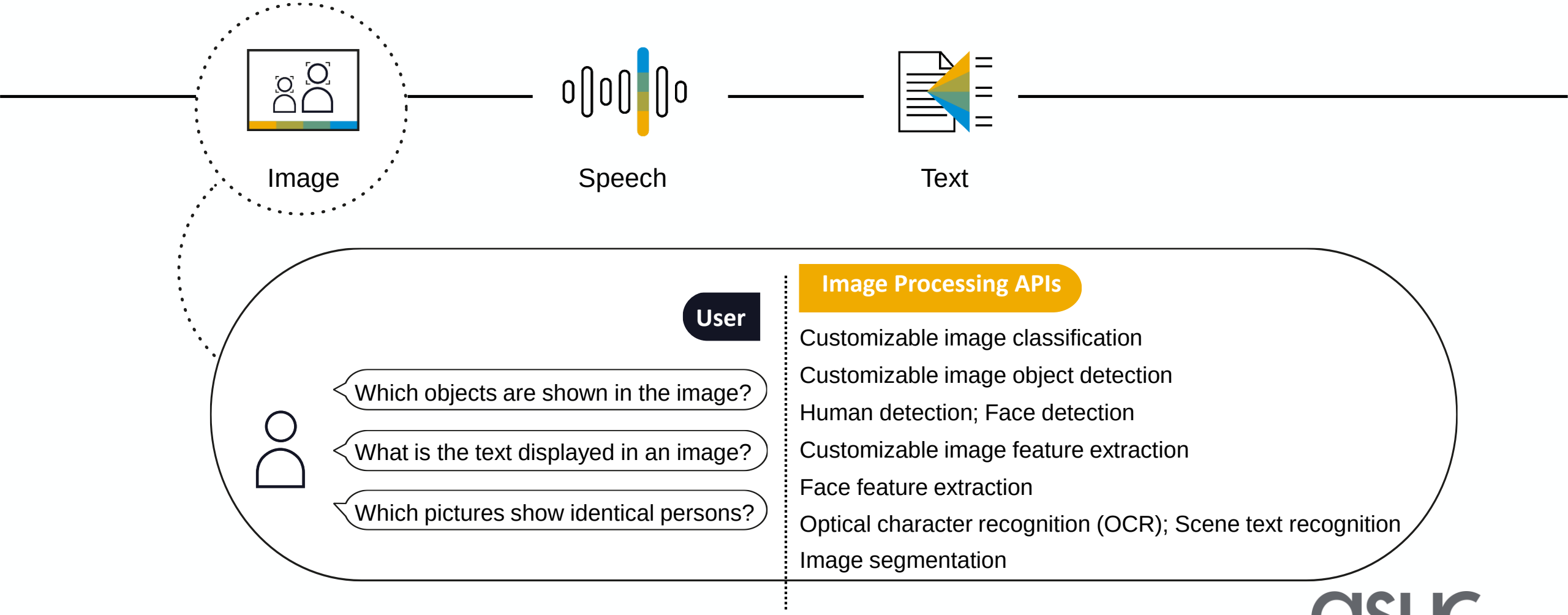
Ready-to-Use Functional Services

Enabling customers and partners to build the Intelligent Enterprise



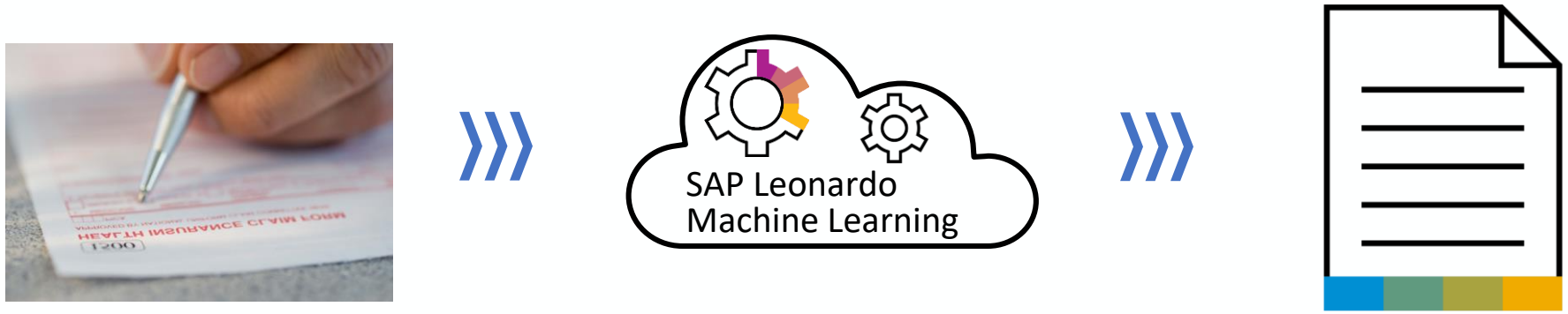
Ready-to-Use Functional Services

Machine learning services portfolio



Optical Character Recognition (OCR)
Capabilities

Optical Character Recognition service takes an uploaded file and returns the text characters detected in the input.



API Business Hub: https://api.sap.com/shell/discover/contentpackage/SAPLeonardoMLFunctionalServices/api/ocr_api

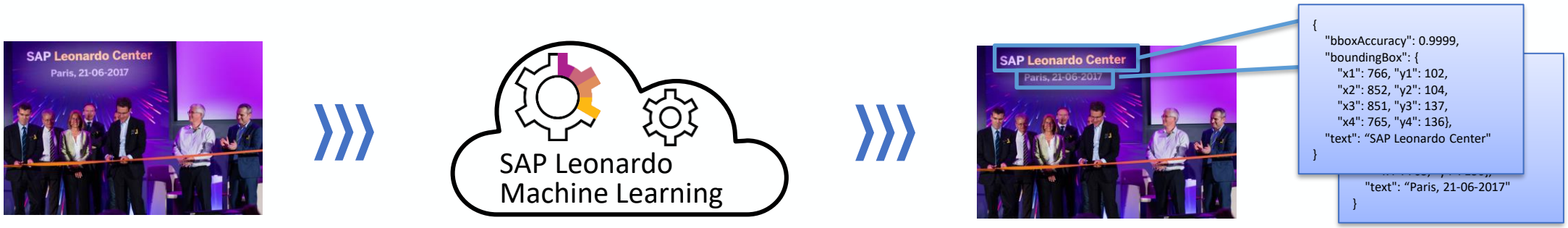
Remark: “Scene Text Recognition API” is an alpha-status API with similar capabilities:

”When the formats from which the text has to be read are documents or print media scans, the OCR service should be used whereas in case of natural images (e.g. reading the counter of a utility meter or the number-plate of an automobile from a security camera feed), the Scene Text Recognition service should be used.”
https://help.sap.com/viewer/product/SAP_LEONARDO_MACHINE_LEARNING_FOUNDATION/1.0/en-US

Scene Text Recognition (Alpha)

Capabilities

– The Scene Text Recognition service localizes and reads text from natural images and scenes.



– It detects individual texts, and outputs those as character strings along with their location within the image.

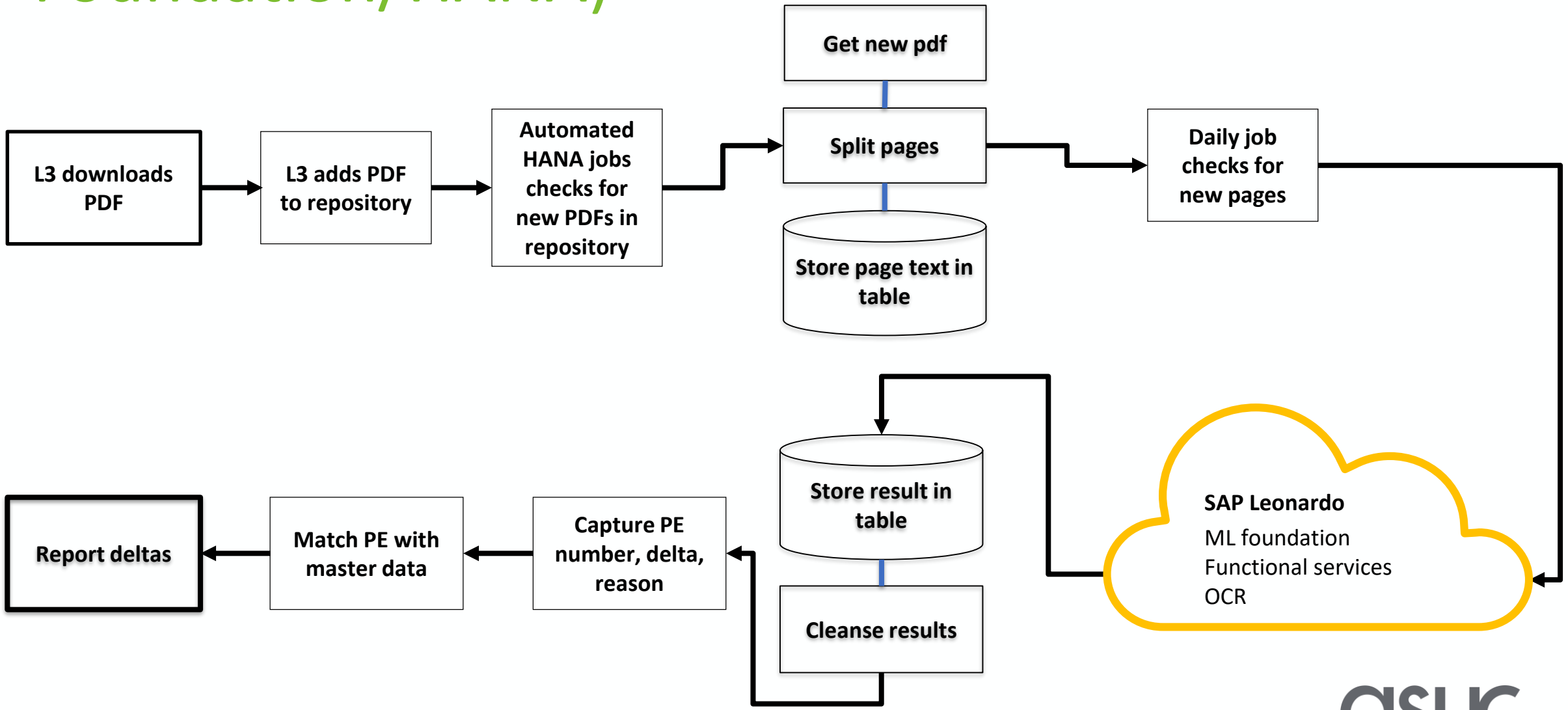
– In comparison to the Optical Character Recognition service, the Scene Text Recognition service offers

- Works with real-life color images
- Ability to work with font-less text
- Extract word-art/picturized text
- Works in different orientations of texts
- Text occurring in natural images like low-contrast, emboss/engrave

– API Business Hub: https://api.sap.com/shell/discover/contentpackage/SAPLeonardoMLFunctionalServices/api/scene_text_recognition_api

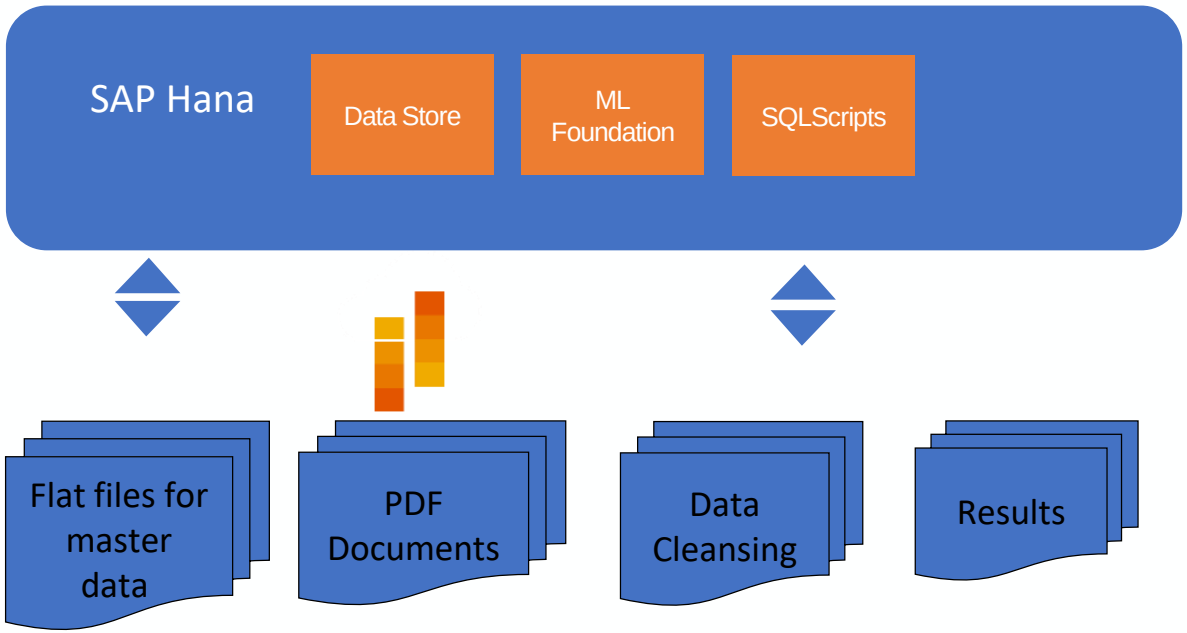
– Documentation: https://help.sap.com/viewer/b04a8fe9c04745b98ad8652ccd5d636f/1.0/en-US/d80662841b94438a968d800d4e152723.html

Use Case 1 – DoD Budgets Marks(ML Foundation/HANA)



Scope – Production Ready Model

Develop and test model based on Evaluation Project plus additional requirements to make model Production Ready



Scope Item	Description
Data Sources	Flat files(Excel), PDF documents
Data Extraction Options	ML foundation – OCR functional services
Engines	SAP Hana Libraries MLF
Models/Tools	Machine Learning, SQLScripts, UI5
Output	Structured data for dashboard reporting

Evaluation Phase – Use Case 2

Evaluation Phase Use Case 2 – DoD Contract Awards

Objective:

- To extract key information about competitors, upcoming technologies, and emerging market trends
- To extract the information from blogs/html feeds and store the text contents and keywords in HANA
- To retrieve the relevant text or structured content on request from business

Scope of Evaluation:

- SAP HANA Text Analysis

Why does SAP HANA provide text analysis functionality?

- Enterprise Challenges

- Massive amounts of unstructured data are being captured in operational, CRM, maintenance, engineering, R&D, and call center systems as well as social media, blogs, forums, e-mails, documents, etc.
- Companies are struggling to:
    - Search on unstructured text related content
    - Extract meaningful, structured information from unstructured text
    - Combine unstructured with structured data
    - Leverage data in real-time to gauge and guide their business strategy and solve critical problems

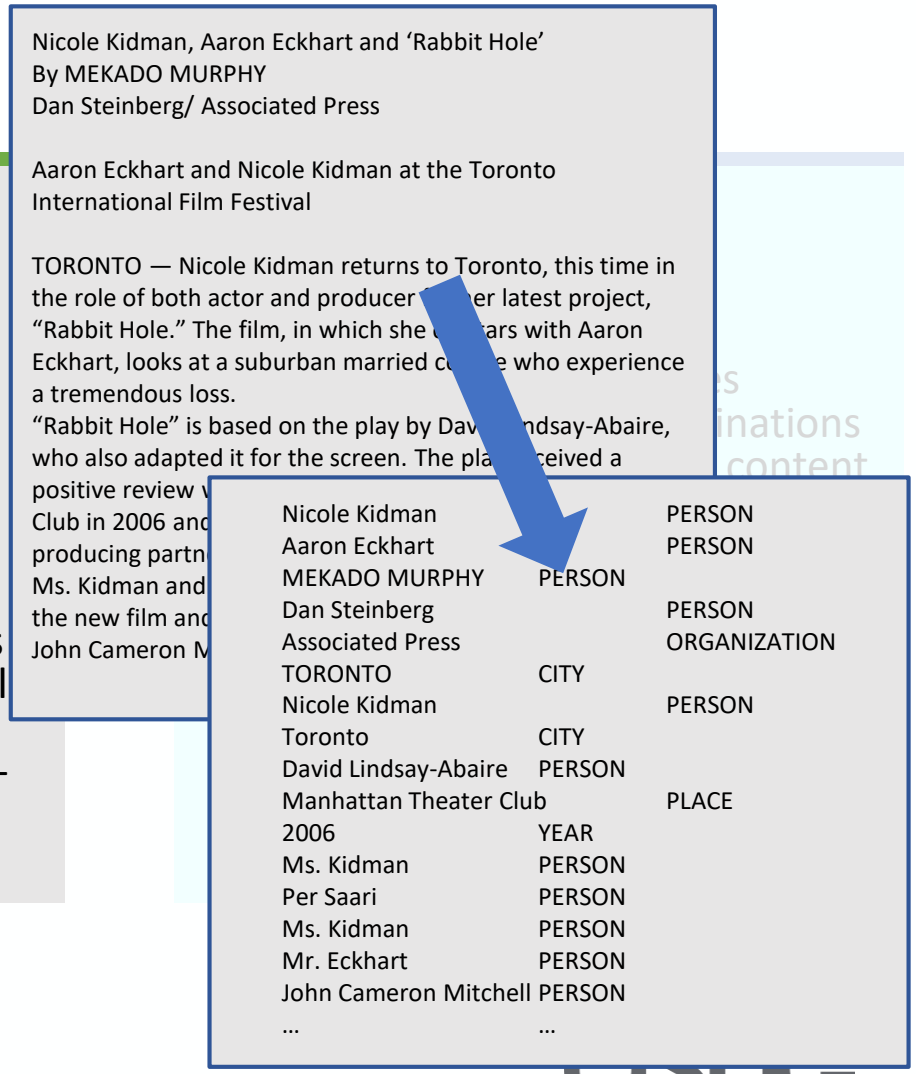
What types of text processing capabilities are supported?

- Search
  - In addition to string matching, HANA features **full-text search** which works on content stored in tables or exposed via views. Just like searching on the Internet, full-text search finds terms irrespective of the sequence of characters and words.
- Text analysis
  - Capabilities range from basic tokenization and stemming to more complex semantic analysis in the form of **entity and fact extraction**. Text analysis applies within individual documents and is the foundation for both full-text search and text mining.
- Text mining
  - Text mining makes semantic determinations about the overall content of documents relative to other documents. Capabilities include **key term identification and document categorization**. Text mining is complementary to text analysis.

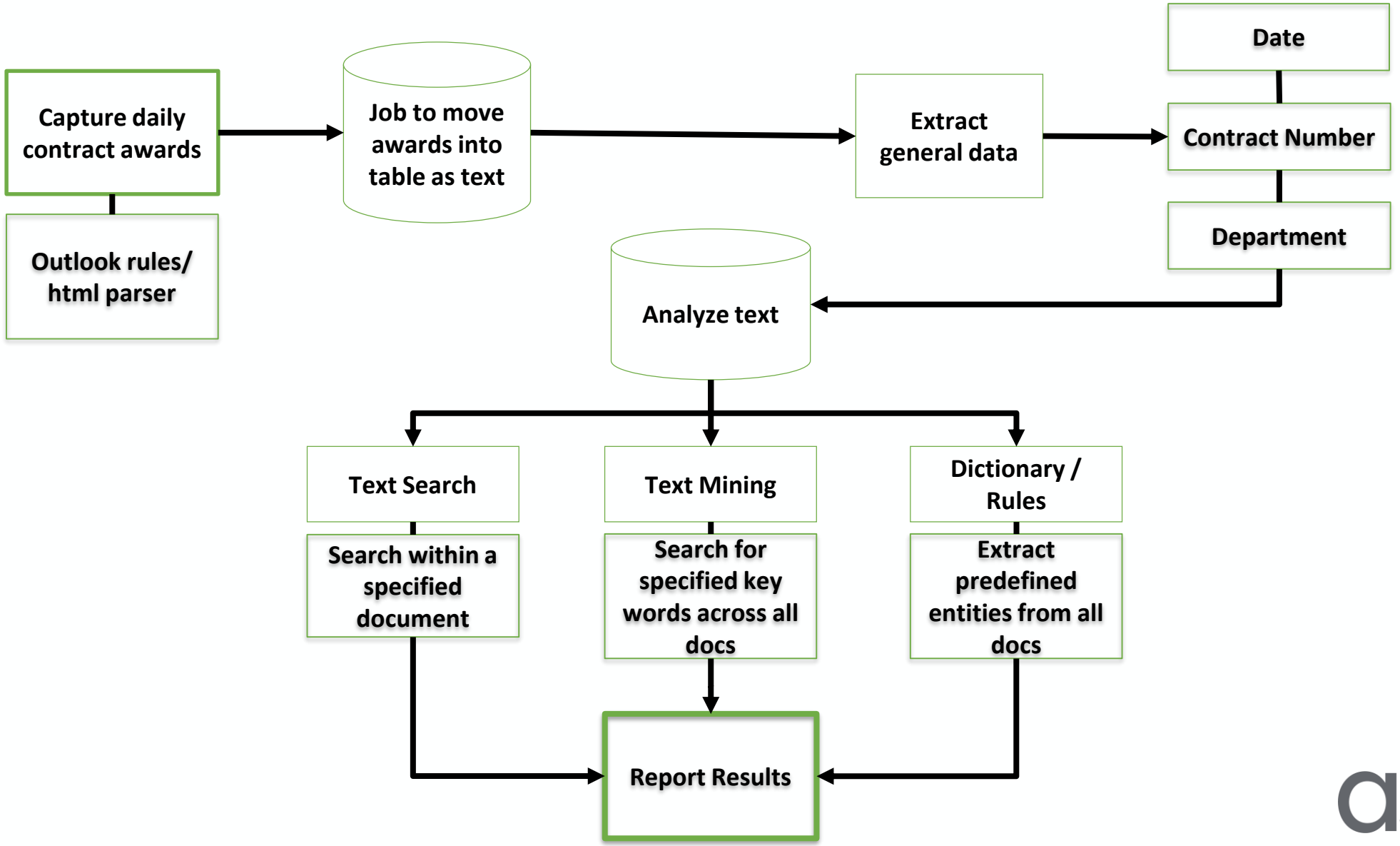
What types of text processing capabilities are supported?

- Search
  - In addition to string matching, HANA features **full-text search** which works on content stored in tables or exposed via views. Just like searching on the Internet, full-text search finds terms irrespective of the sequence of characters and words.

- Text analysis
  - Capabilities range from basic tokenization and stemming to more complex semantic analysis in the form of **entity and fact extraction**. Text analysis applies within individual documents and is the foundation for both full-text search and text mining.

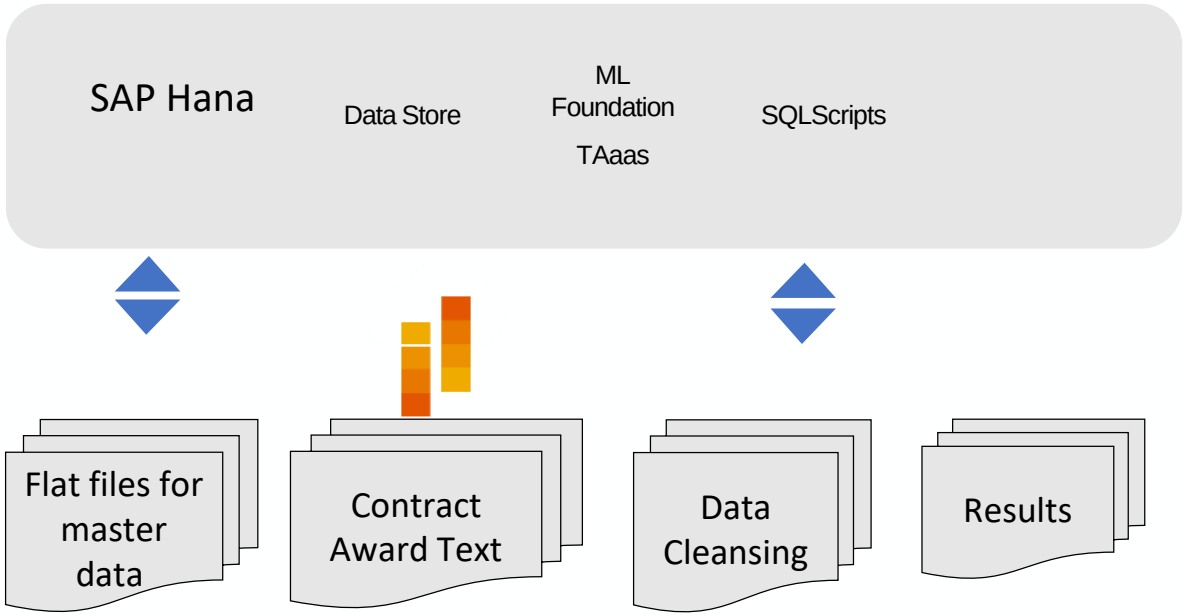


Use Case 2 – Daily Contract Awards(Phase 1)



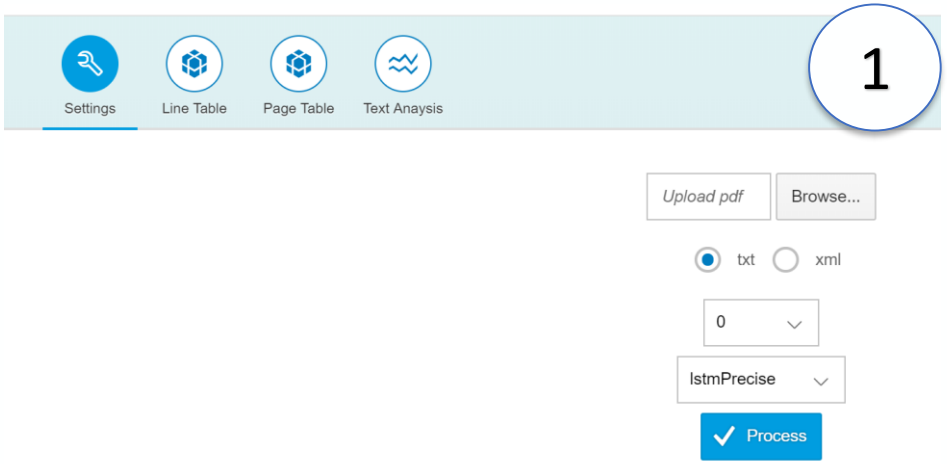
Scope – Production Ready Model

Develop and test model based on Evaluation Project plus additional requirements to make model Production Ready



Scope Item	Description
Data Sources	Flat files(email/html)
Data Extraction Options	Hana Jobs
Engines	SAP Hana Libraries MLF
Models/Tools	ML Foundation, Text Analysis as a Service 2.0(To be released), Text Mining, SQLScripts, UI5
Output	Dashboard for text searching/mining, Structured data for dashboard reporting

OCR APP – Demo/Screenshot



UNCLASSIFIED
Department of Defense
FY 2019 President's Budget
Exhibit R-1 FY 2019 President's Budget
Total Obligational Authority
(Dollars in Thousands)

2

Appropriation	FY 2017 (Base + OCO)	FY 2018 PB Request with CR Adj Base	FY 2018 Total PB Requests* with CR Adj Base	FY 2018 PB Request with CR Adj OCO	FY 2018 Total PB Requests+ with CR Adj OCO
Research, Development, Test & Eval, Army	8,852,507	8,273,447	8,273,447	342,356	342,356
Research, Development, Test & Eval, Navy	17,851,955	17,116,976	17,091,976	326,537	326,537
Research, Development, Test & Eval, AF	28,381,681	27,577,477	27,577,477	365,205	365,205
Research, Development, Test & Eval, DW	19,542,639	18,639,241	18,639,241	442,104	442,104
Operational Test & Eval, Defense	188,654	184,666	184,666	2,725	2,725
Total Research, Development, Test & Evaluation	**74,817,436**	**71,791,807**	**71,766,807**	**1,478,927**	**1,478,927**
Other RDT&E Budget Activities Not Included in the Research, Development, Test and Evaluation Title					
Office of the Inspector General	4,615	2,800	2,800		
Defense Health Program	2,101,627	673,215	673,215		
Chem Agents & Munitions Destruction	515,609	839,414	839,414		
National Defense Sealift Fund	7,237	18,622	18,622		
Total Not in Research, Development, Test & Evaluation	**2,629,088**	**1,534,051**	**1,534,051**		

3

UNCLASSIFIED
Department of Defense
FY 2019 President's Budget
Exhibit R-1 FY 2019 President's Budget
Total Obligational Authority
(Dollars in Thousands)
FY 2017
Appropriation
Research, Development, Test & Eval, Army 8,852,507
Research, Development, Test & Eval, Navy 17,851,955
Research, Development, Test & Eval, AF 28,381,681
Research, Development, Test & Eval, DW 19,542,639
Operational Test & Eval, Defense 188,654
Total Research, Development, Test & Evaluation 74,817,436
Other RDT&E Budget Activities Not Included in the Research, Development, Test and Evaluation Title
(Base + OCO)
FY 2018
PB Request
with CR Adj
Base
8,273,447
17,116,976
27,577,477
18,639,241
184,666

4

Department of Defense FY	ORGANIZATION/GOVERNMENT	2:2:13
2019	YEAR	2:2:38
President	TITLE	2:2:43
Budget Exhibit	NOUN_GROUP	2:2:55
2019	YEAR	2:2:77
President's Budget Total Obligational Authority	ORGANIZATION/GOVERNMENT	2:2:82
FY	PROP_MISC	2:2:153
2017	YEAR	2:2:156
Appropriation Research	NOUN_GROUP	2:2:161
Eval	PROP_MISC	2:2:205
Army	ORGANIZATION/GOVERNMENT	2:2:211

Summary

Technical Evaluation Using	Pros	Cons	Recommendation
OCR APIs (Use Case 1)	• Extracts text from Scanned images/pdf • Supports Multiple Languages • Provides more lstm options	• Does not extract hand written text	New inhouse solution to be released soon for handwritten text

Summary

Technical Evaluation Using	Pros	Cons	Recommendation
Text Analysis (Use Case 2)	• Provides native NLP algorithms for extracting unstructured text • Support multiple languages • Support custom dictionaries & custom extraction rules	• Does not extract meaningful information from tables or images	Use the hybrid approach of extracting information from OCR API and then structuring the data using Text Analysis capabilities

Next Steps

Future Engagement

- ✓ L3 to evaluate the technology options for the production type development model
- ✓ SAP Max Attention to be involved for additional use case involving Data Hub/deployment platform
- ✓ Additional use cases might require hybrid technology deployments(HANA XSA/Machine Learning) depending on complexities of the use case.

Take the Session Survey.

We want to hear from you! Be sure to complete the session evaluation on the SAPPHIRE NOW and ASUG Annual Conference mobile app.

ASUG

Presentation Materials

Access the slides from 2019 ASUG Annual Conference here:

http://info.asug.com/2019-ac-slides

Q&A

For questions after this session, contact us at Arvind.Patel@L3T.com and v.rr@sap.com

Let’s Be Social.

Stay connected. Share your SAP experiences anytime, anywhere.

Join the ASUG conversation on social media: **@ASUG365 #ASUG**

ASUG