

Daily Tasks

Tuesday, November 27, 2018 2:04 PM

- TNB
- ScrumMaster Exam
- Summit
- SDP
- CPF
- Badges
- Counsellor role
- Exceptional client service training
- Other revise items in mailbox to review
- Brisbane deck
- SDP deck
- James' confirmation and CCG inputs

SDP Discussion points

Thursday, November 7, 2019 12:00 PM

1. [Jeff] how much do SDP think drivers are stealing? Is it a few bricks per load, 50% of the load, entire trucks?

[Rag] The proposed PoC is for a new site. Hence, we shall be having a new perspective here. It's rollout to the pilot may or may not be on coal but this PoC shall need to circle back as to where it stops. Ideally, it's the iron ore and the outcome is solution able to show the value by someone doing a manual exercise of taking out xx % of iron ore out

2. [Benn] I think it's reasonable if we were to say we will detect and red flag a change if it's more than 10% difference?

[Jeff] We can explore the use of vision analytics for the purposes of detecting ore theft, and will help SDP evaluate the efficacy as it relates to % of theft.

[Rag] Shall we propose a tabulation at the end of PoC as the deliverable as to relate percent of theft to percent of detection. The only risk it has is a benchmark not stated as the target goal. The table can look like this

Theft proportion (% of weight)	Detection Rate(% Accuracy)
0-4	60
5-10	80
>10	95

3. [Jeff] How can we uniquely identify a truck? Reading the truck id or license plate isn't very accurate.

[Rag] Do we have a proposal in terms of the camera positioning for this? The position of plates is fairly consistent between trucks for BUMA which makes the location fixed if we segment the truck out of the image.

4. [Jeff] Further, when looking at the provided PDFs it appears they have records of the truck, its weight, weight of coal and time in/out details. To confirm, is the ask then to apply vision analytics to verify the truck and its full load has arrived at either the jetty or processing site?

[Rag] The client has asked for a basic descriptive analysis on this pdf. Potentially, once we have the CV detection of plates, we can use the same to come up with this report. That shall be of value to client as they currently have the report post the month. This report can give further insights in to driver behaviour, driving patterns around times of the day, etc.

5. **Action items on EY side:**

Proposed cost & time:

- a. How long do we think it shall take?
- b. Camera specifications (**Camera angle, position, number and quality**) - What's Jeff's view on it?
- c. Data collection methodology:

How many data points/images/videos shall be sufficient to train individual model and engine verification?

Learnings

Wednesday, December 5, 2018 11:07 AM

1. Engagement experiences data sciences:
 - a. 60-85 % failure rates
 - b. Identify the right problem through stakeholders brainstorming
- Technical career framework
- BLOB Storage

Data Science Badge learnings

Wednesday, December 5, 2018 11:54 AM

1. Clustering and unsupervised machine learnings:
 - a. Concepts:
 - i. Latent and observed variables
 - ii. Gaussian Mixture models
 - iii. Cluster quality evaluation
 - iv. Cluster convergence - no more possible assignments
 - v. Soft K Means - fuzzy membership in each class (probability of alignment); thus defining probabilistic cluster responsibilities
 - vi. Mathematical guarantee
 - b. Tools:
 - i. Visualization: Scatterplot, dendrogram, hierarchical plots
 - c. Applications:
 - i. Document search - use to search vectors (built on documents using TF-IDF or bag of words approach)
 - d. Capabilities:
 - i. Can visually show the segmentation
 - ii. Can visually show the progression to clustering
2. Dimensionality reduction

Resources

Wednesday, December 19, 2018 11:51 AM

Atlas for Analytics in audit resources
EY Canvas

Structured Problem solving

Thursday, December 20, 2018 9:13 AM

- State Organize anaLyzE Verify Express - EY Framework approach: PDW (Problem Definition Worksheet)
 - o State the problem
 - Basic question
 - Level of detail
 - Relevance
 - Solvability
 - Context and trigger
 - Decision-makers
 - Solution threshold
 - Constraints
 - Scope of work
 - o Organizing the problem
 - Logic trees
 - Big picture
 - Decomposing and structuring problems
 - Bringing rigor and structure
 - Facilitating communication
 - Making projects more efficient
 - Types of logic tree
 - What ?
 - How ?
 - ◆ Dupont tree - Profit and turnover evaluation
 - Why ?
 - MECE (Mutually Exclusive Completely Exhaustive) tree
 - Financial sector example
 - o Power map - stakeholder management: Buyer value management

SCQA framework - Proposals

Monday, July 29, 2019 9:53 AM

Situation

Complication

- Global examples to support own complications

Question

Answer

:

Question Trigger

Research

Understand

Important elements

Patterns and trends

Reading

Research - detailed

Beyond borders - DBB from EY App Store

Tasklist

Tuesday, March 12, 2019 12:11 PM

MY010030846

Funds transfer
register for the events in summit

11/11:

1. Check on communities on yammer and how to go ahead with own community - what to share and how much can be shared + Presentation to be submitted via KM
 - a. We can create our own group or participate in the few available online with high member presence
 - b. This can be a platform to share the material on assets created
 - c. This can also act like a blog share but the hosting of blogs has to be on a separate platform
 - d. Use EY Cognistreamer:
<https://find.ey.net/discover/sitempages/resource.aspx#detailPath=https://internal.ey.net/sites/rescat/Lists/Resource Catalog/DispForm.aspx?ID=36830>
2. Add badge in signature along with PSM I logo
3. Learnings in the Microsoft space
4. Participate in the GE challenge
https://ey-home.ey.net/wps/myportal/?ut/p/a0/IY_LDolwEEW_yMyUDICXRSCy0Bjf7YY0WLVRwWiDvy_d6cqYzGbmnpPcAQ170K3p3cl417XmGnad1NI8MRMZk4himmAlKdpI4hEiAwVKpx_AVsYDMKYsiTjHKoVd7-wLVqBBLx6uN97Wzfm5tPera0yVgxJxFKeiKBHTvGRs_EkeHuboZ93BgpqH-2-z6VpvWx_iyXpEjHhOVPJcFoUsKBiC5NCcShyir1b_mbtfj98v6iZ4mPryBubT2F8!/

Create a video of the Blockchain based implementation of vendor aggregation or another aspect - focused on process optimization - call it Process Optimization Framework

5. Explore yammer strategy and the discover on the PEGA platform
6. Compile learning on CV at a github location
7. Make demo video/dashboard for the Summit
8. Amend past timesheets for new project code
9. EY badge: timelines to check for James

Personal -

1. Explore YCombinator literature and finalize one of them as approach to the book
 2. Study leaves planning - TOGAF next steps and see any internal material in EY
 3. Check on Rshiny security issues for publishing, data upload, etc.
 4. Browse on the analytics badges - AI gold, etc.
 5. Explore market data model shared on Linkedin
-
1. Read material on Azure ML and stuff if needed for presentations
 2. Go through TNB - DGS + SC and Spend Analytics stuff
 3. Create a summary slides for point 2 above
 4. File expenses - EPF

5. Prepare EPF summary to discuss with Adrian
6. Data science badge learnings to complete
7. Send questionnaire to DGS guy - Madhu
8. Create a specification for DS stuff machine
9. Install PowerBI, Azure Marketplace setup and Python in the required form

Practice leadership meeting

Wednesday, October 23, 2019 8:33 AM

CCG

Plan for the skills training

Langkawi Summit

Friday, August 23, 2019 9:09 AM

1. Workbench.ey.net
2. CTPs through workbench
3. Jeff's slide on CTP - Srujan
4. Can query on sharepoint for workbench
5. Check with Blair on the slides with Microservices
6. Check with Blair on access to Project Wolf assets

Hackathon

Friday, August 23, 2019 9:59 AM

Business Problem statement

- Clear problem articulation
- Solution
- Business Value
- Clear Approach
- Innovative

Technology

Presentation

Proposed views

Monday, September 23, 2019 11:26 AM

1. **Career path for associates** - APP development vs Data Science vs AI (Automation) vs Data Engineer vs Business Analyst
- 2.

Meeting Rooms booking

Friday, October 4, 2019 11:21 AM

[Shahrin Shahrulzaman:
sho.yven.tan@my.ey.com

PowerBI

Friday, October 4, 2019 12:14 PM

Real time data options:

1. Streaming
2. Push datasets
3. Pub nub

Synaptic panel

- Bring in new images as background

Friday, October 4, 2019 2:31 PM

Local counselling excellence code to charge 1 hour on Friday 4th October

MS DevOps

Thursday, December 19, 2019 9:56 AM

Password: Awaredevops1@

Claims

Tuesday, February 11, 2020 2:15 PM

DQML

Saturday, February 15, 2020 8:39 PM

To research on algorithms that can help in:

1. Reusable algorithms can be rapidly applied across multiple data sets
2. pattern analysis across multiple data sets and variables identifies more issues
3. validation of data relationships , etc.
4. Learns to recognise data types and apply previously confirmed DQ rules
5. Reduce errors
6. Out-of-sample predictions of missing values with ML algorithms
7. Capacity to analyse a broader range of data sets to identify patterns and suggest values
8. Where root cause is understood and remediation approach approved, DQML can be used to rapidly deliver learned and consistent data fixes
9. Unsupervised learning provides improved basis for advanced analysis activities
10. Big data scalability
11. Being able to visualize
12. Being able to connect with multiple data bases and sources
13. Flexibility between univariate and multivariate cases
14. Basic profiling dashboard to evaluate completeness and accuracy of datasets
15. Flexibility to select ML Algorithm
16. Data view and stats view for each algorithm
17. Overall algorithm view to compare and make selection
18. New relationship identification
19. Limited to csv and API connectors
20. Flexibility to upload business rules
21. EDA on full or subset as per user selection
22. Export or connect via API
23. Enable tuning of confidence levels or show the confidence levels along side the results
24. DQ identification & remediation model
25. Transaction & reference data
26. Use of Active Learning to integrate feedback from experts in order to improve performance
27. Nature of datasets: timeseries, etc.

Dashboard overview (from Video):-

1. Basic data profiling
 - a. Univariate Distribution
 - b. Univariate Frequency distribution
 - c. Missing value analysis
 - d. Duplicates
2. Algorithms listed for anomaly detection:
 - a. DBScan
 - b. OneSVM
 - c. Isolation Forest
 - d. LOF
 - e. SVM
 - f. Decision Tree-Entropy
 - g. Decision Tree-Gini
 - h. Autoencoder
3. View of the dashboard:

- a. Business view
- b. Statistical view

Problem today:

- 1. Structured vs unstructured data
- 2. Sampled vs Unsampled data

	Datastore	Draw: TLHJ connection to Azure or anyother VM
Dev environment		
Architecture		
Environment agnostic		
More docker oriented		

Intern process

Thursday, March 5, 2020 9:07 AM

- Ask Naz on how the communication starts
- Charging rate
- Follow up strategy

Azure Certification

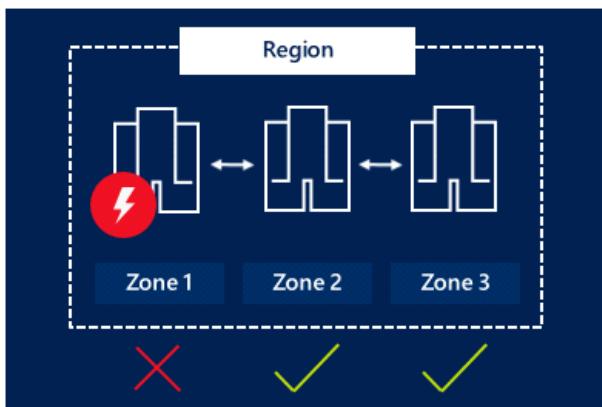
Thursday, April 2, 2020 11:14 AM

MSP3754A7B9F

Notes

1. Availability options in Virtual Machines:

- a. <https://docs.microsoft.com/en-us/azure/virtual-machine-scale-sets/availability>
- b. **Availability zones** expand the level of control you have to maintain the availability of the applications and data on your VMs. An Availability Zone is a physically separate zone, within an Azure region. There are three Availability Zones per supported Azure region. Each Availability Zone has a distinct power source, network, and cooling. By architecting your solutions to use replicated VMs in zones, you can protect your apps and data from the loss of a datacenter. If one zone is compromised, then replicated apps and data are instantly available in another zone.



- c. A **fault domain** is a logical group of underlying hardware that share a common power source and network switch, similar to a rack within an on-premises datacenter.
- d. An **update domain** is a logical group of underlying hardware that can undergo maintenance or be rebooted at the same time. This approach ensures that at least one instance of your application always remains running as the Azure platform undergoes periodic maintenance. The order of update domains being rebooted may not proceed sequentially during maintenance, but only one update domain is rebooted at a time.

e. Virtual Machines Scale Sets

Azure virtual machine scale sets let you create and manage a group of load balanced VMs. The number of VM instances can automatically increase or decrease in response to demand or a defined schedule. Scale sets provide high availability to your applications, and allow you to centrally manage, configure, and update many VMs. We recommended that two or more VMs are created within a scale set to provide for a highly available application and to meet the 99.95% Azure SLA. There is no cost for the scale set itself, you only pay for each VM instance that you create. When a single VM is using Azure premium SSDs, the Azure SLA applies for unplanned maintenance events. Virtual machines in a scale set can be deployed across multiple update domains and fault domains to maximize availability and resilience to outages due to data center outages, and planned or unplanned maintenance events. Virtual machines in a scale set can also be deployed into a single Availability zone, or regionally. Availability zone deployment options may differ based on the orchestration mode.

f. Availability sets

An availability set is a logical grouping of VMs within a datacenter that allows Azure to understand how your application is built to provide for redundancy and availability. We recommended that two or more VMs are created within an availability set to provide for a highly available application and to meet the 99.95% Azure SLA. There is no cost for the Availability Set itself, you only pay for each VM instance that you create. When a single VM is using Azure premium SSDs, the Azure SLA applies for unplanned maintenance events.

Passwords

Monday, April 13, 2020 11:36 AM

GlobalOne: Naughtyboy12@

EY Delivers: Naughtyboy15@

Acclaim: Acclaim1@

Docker: Dockerragh1@

AI DS Competency

Wednesday, May 6, 2020 3:07 PM

1. Project stuck
2. Recognition to understand the competency - involve in to internal projects
3. Skillset have to be understood and breed around
4. Insights on the survey areas: People pillar

Solution pillar:

1. x-labs
2. Use experience from the projects, past experience, come up with problems in scenario basis

Create a plan for the engagement on learning, etc.

Thursday, May 14, 2020 3:02 PM

Send feedback to Naz, Benn, Indrajit

Review James' candidature for the badge application