Final Intern Presentation

Data and Dashboards

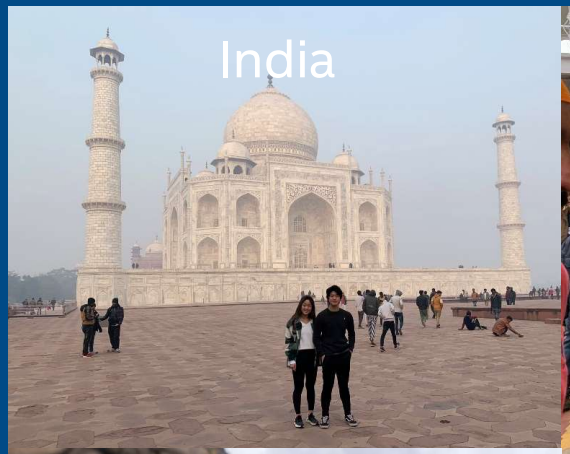
By Phoeland (Momo) Siu

intel®

Agenda

- About Me
- ASPN Dashboard
  - Why ASPN?
  - Demo
  - Impact and Improvements
- Catalyst Application Health Dashboard
  - What is Catalyst?
  - Work Completed
  - Next Steps
- Learning Points
- People at Intel
- Acknowledgements
- Conclusion

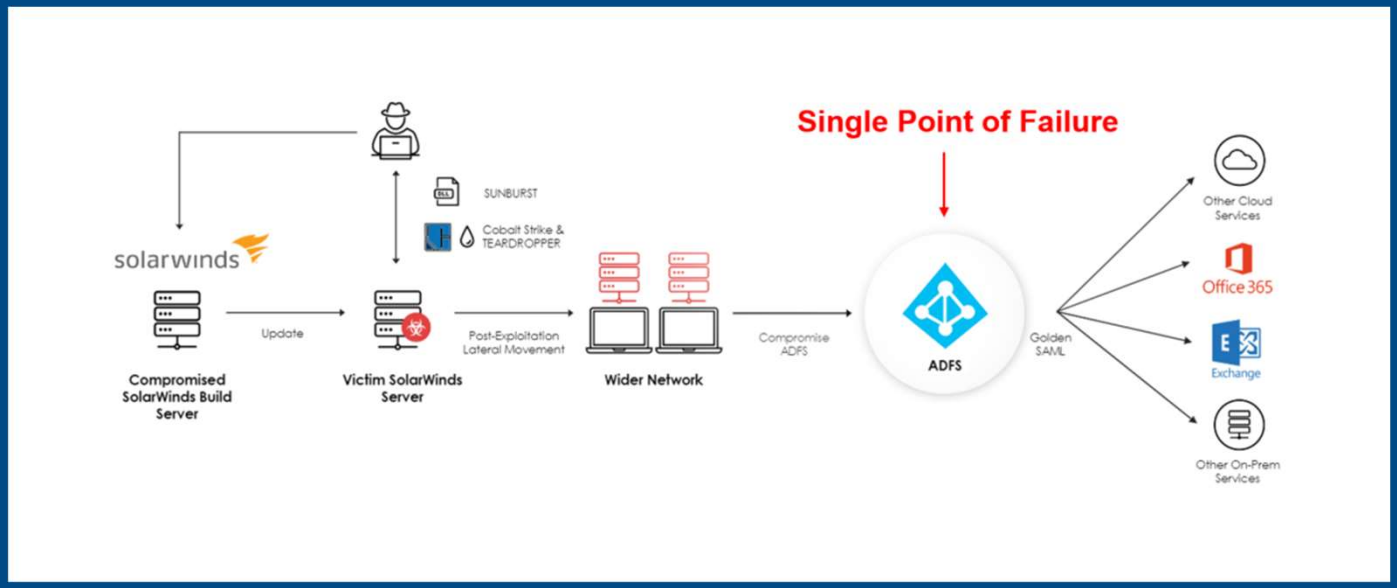
About Me



Why Application Security Posture Now?

- Cyber attacks are inevitable. Our best response is to be ready for them when they happen.

"Imagine those Reese's Peanut Butter Cups going into the package and just before the machine comes down and seals the package, some other thing comes in and slides a razor blade into your Reese's Peanut Butter Cup,"



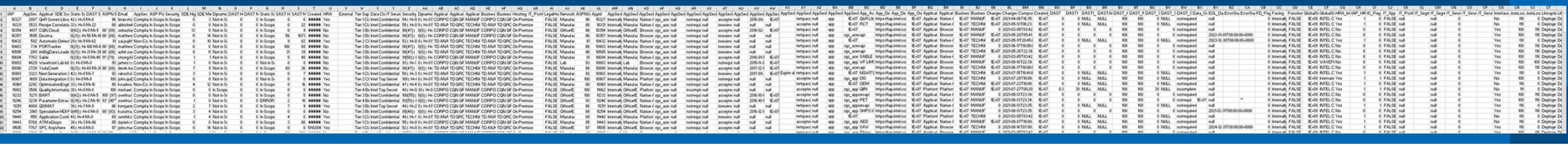
Why the ASPN Dashboard?

Tasks Completed:
Register ✓
Plan ✓

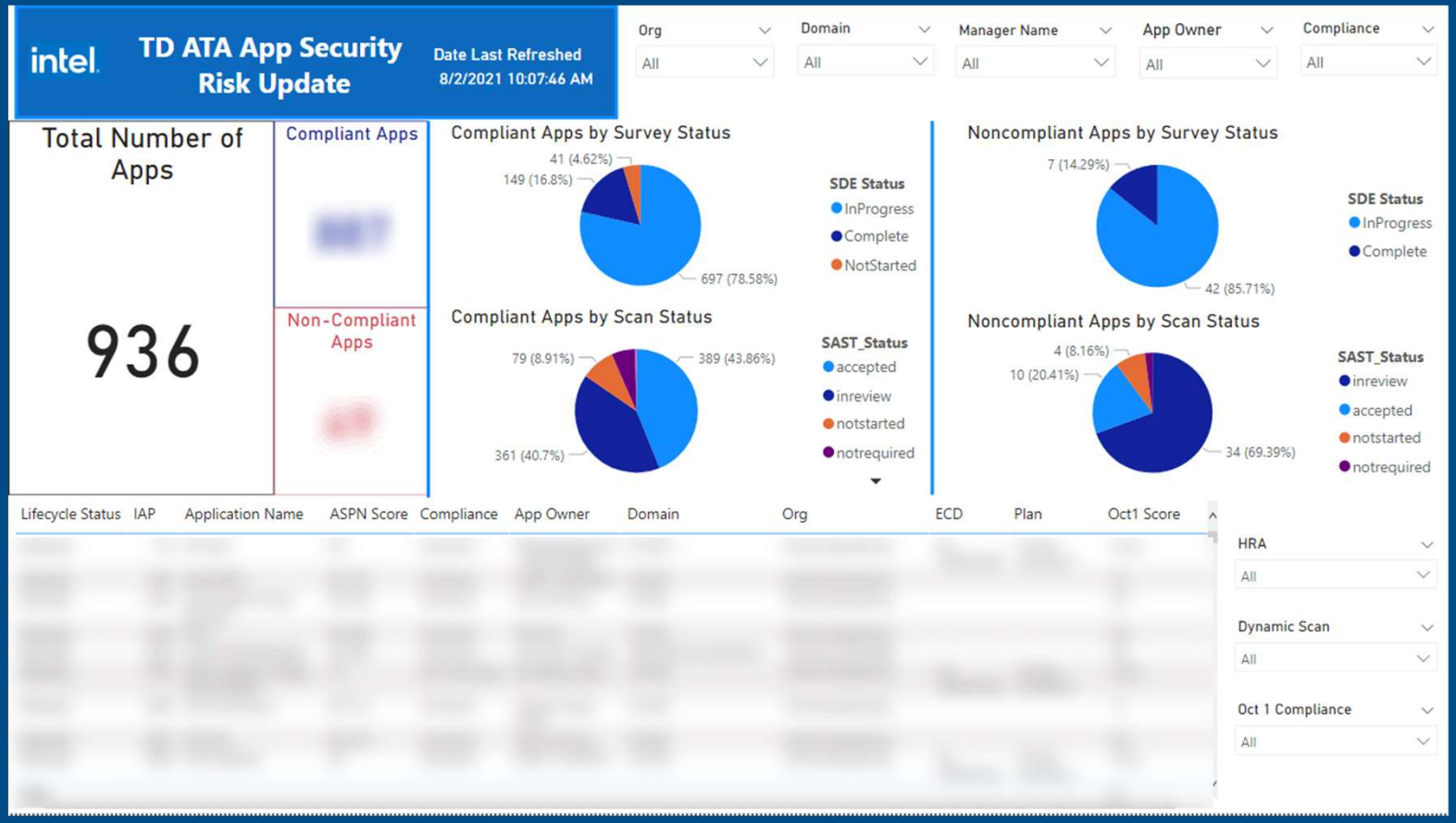
Tasks To Do:
Scan
Remediate
Maintain

New Problem:

The data from 900+ applications created across the TD ATA department created a spreadsheet that was **936 rows x 217 columns**. This was **extremely** inconvenient for managers to maintain and monitor their applications.



ASPN Dashboard



ASPN Dashboard & Query Demo

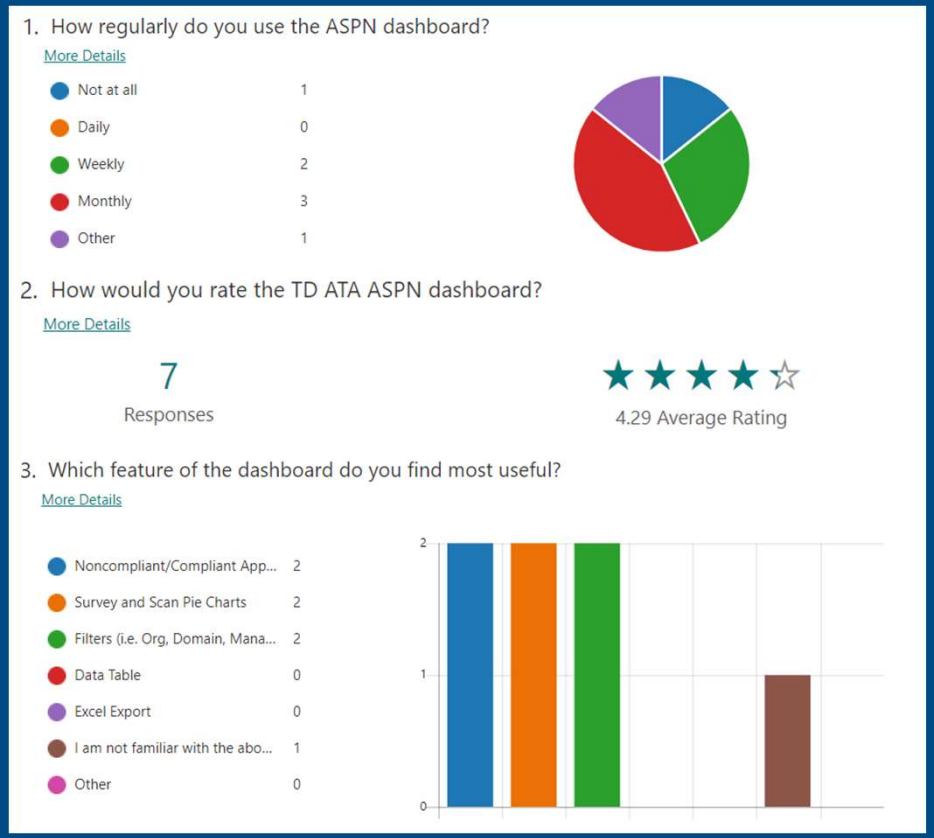
ASPN Dashboard Impact

Observations from Data:

- 7 managers responded out of ~60 accounted for in the dashboard
- Respondents use dashboard, on average once a month.
- Generally, respondents give the dashboard a 4.29/5 star rating.
- Respondents enjoy different features within the dashboard.

Questions to Consider:

- How are managers currently maintaining their applications?
- How can we make the dashboard more apparent to managers?
- What can we change about the dashboard to make it more useful to managers?



Future Improvements

Suggested Changes

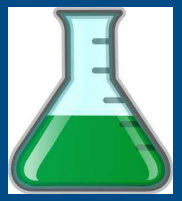
- Monthly incremental progress chart
- Change IAP column to have links to the Application Security Portal
- Show the number of open tasks from survey/scan for each app

Other Possible Improvements

- Writing a query to send custom reports to app owners
- Sending summary points of dashboard to managers (i.e. daily total compliant/non-compliant numbers)
- Drill downs for applications indicating their tasks completed and to-dos
- Automated alerts to managers if apps become noncompliant

ASPN Dashboard Q&A

What is Catalyst?



Catalyst: A **unified suite of offline reporting / analysis** tools for Yield trend & pareto, PCS, Machine Learning, Commonality, Image Retrieval, Interconnect / Shape, Customer Returns / DPM, Cycle/Process/Queue Time **C#/WPF**

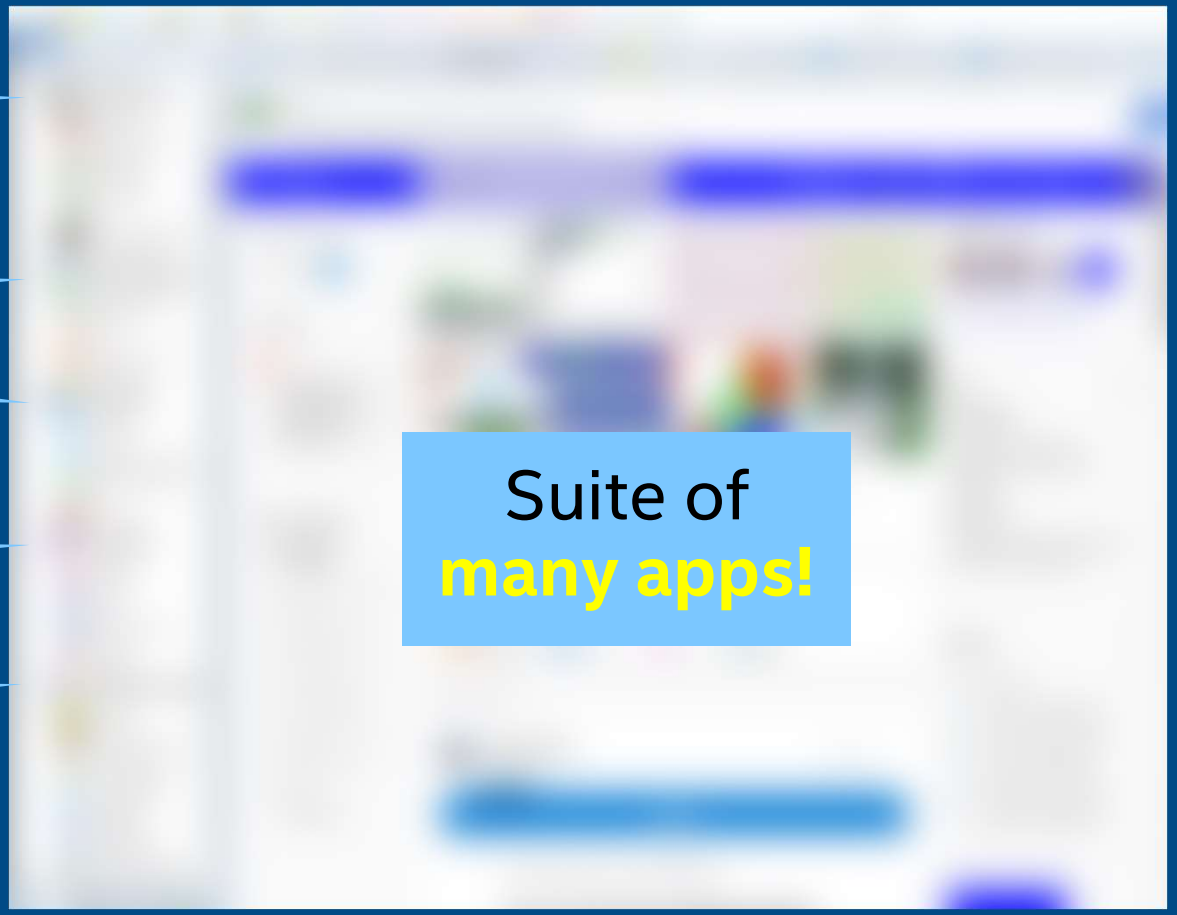
CATALYST_ZEN_AMPS_OVERVIEW_JUNE_2021.pptx by Randall Goodwin

Apps within Catalyst: Wafer Map, PCS Health –SPC++, APP.XAML.CS, PCS Trends, PCS SPC++, Yield Builder, AMPS – PDK MBIT, PCS Admin, Bintrends, Cycle Time, Ship Sort Yield, Yield Pareto, Commonality, Yield Forecast vs Actuals, Testtime, OCT, and MORE

Disclaimer: The data I will present are NOT confidential, as it has company-wide access

Catalyst Applications

- Commonality & Machine Learning
- Images and Interconnect
- Cycle Time, Routes, Equipment Perf
- Offline PCS Charts & Health, Limits
- Yield Reporting & Analysis



CATALYST_ZEN_AMPS_OVERVIEW_JUNE_2021.pptx by Randall Goodwin

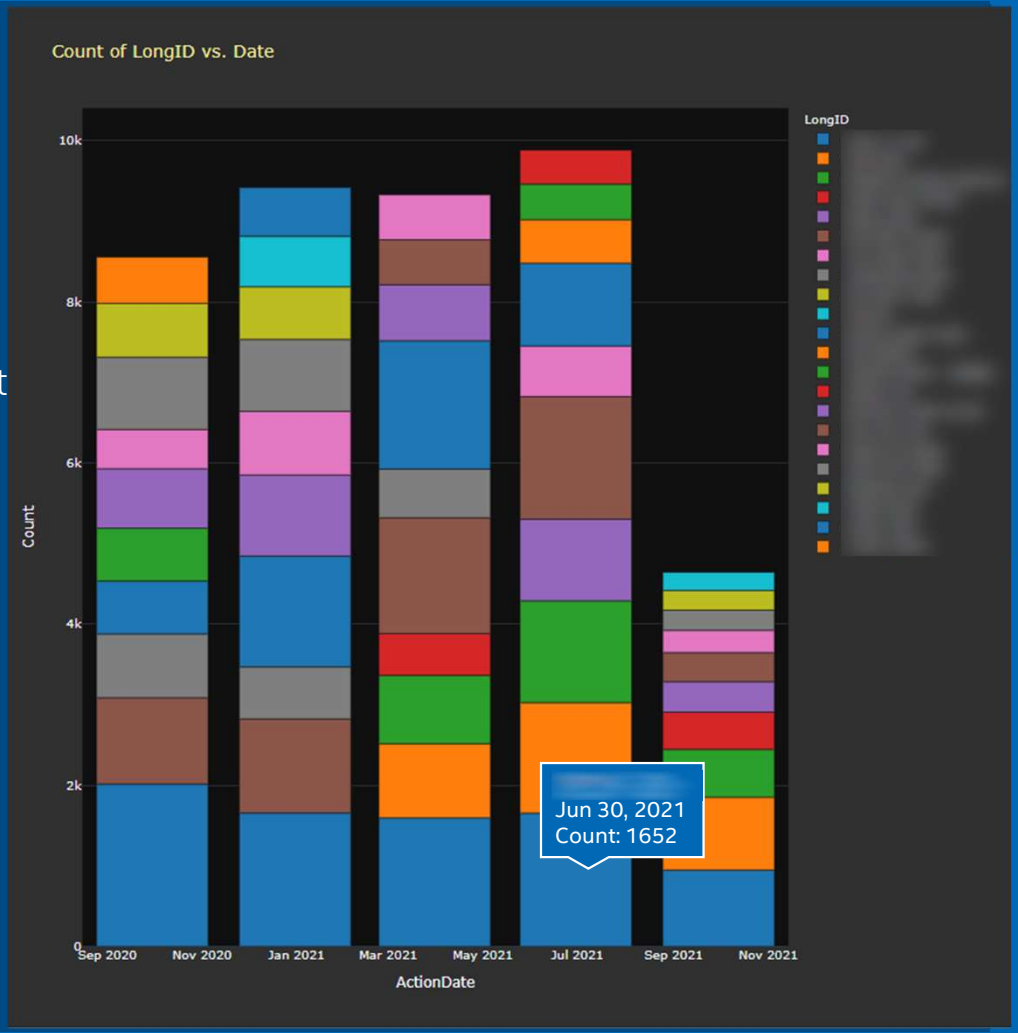
Gathering User Insights

Purpose

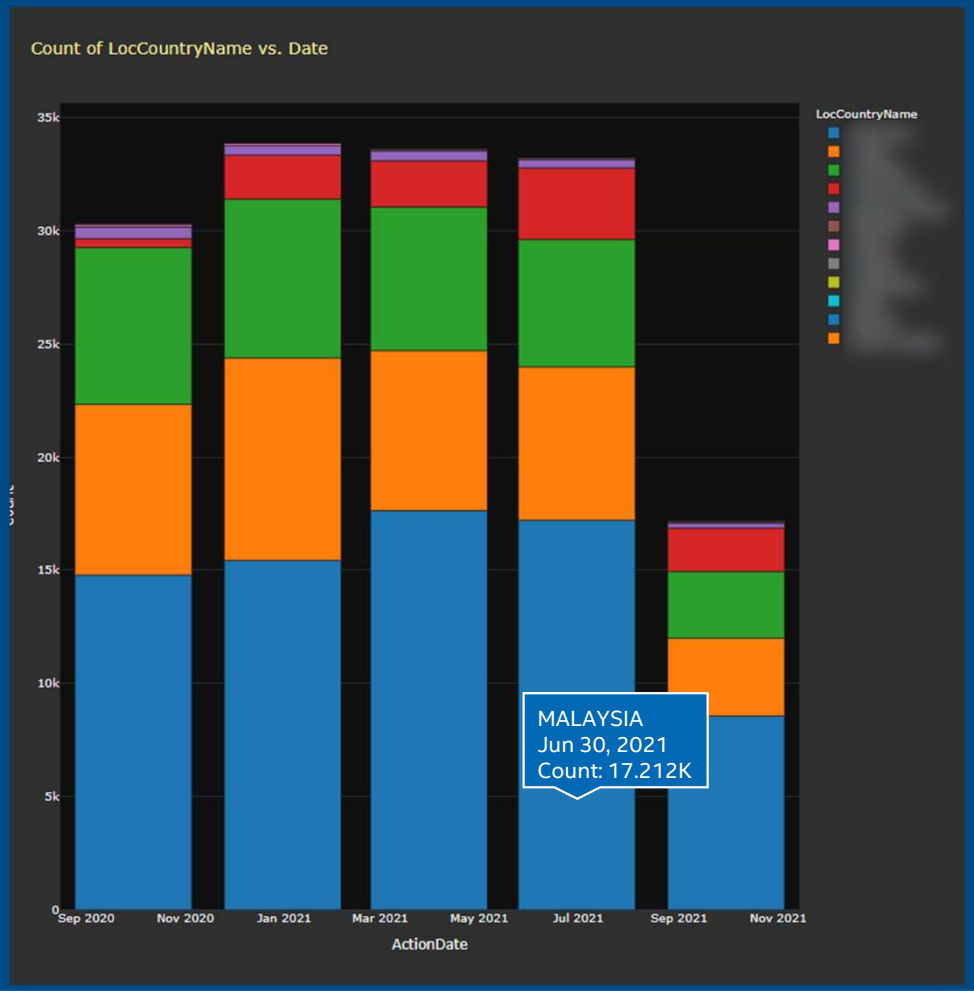
- Users of Catalyst come from all over Intel. Information on users allows us to:
  - 1. Listen to their feedback for future improvement
  - 2. Understand the reason behind their usage

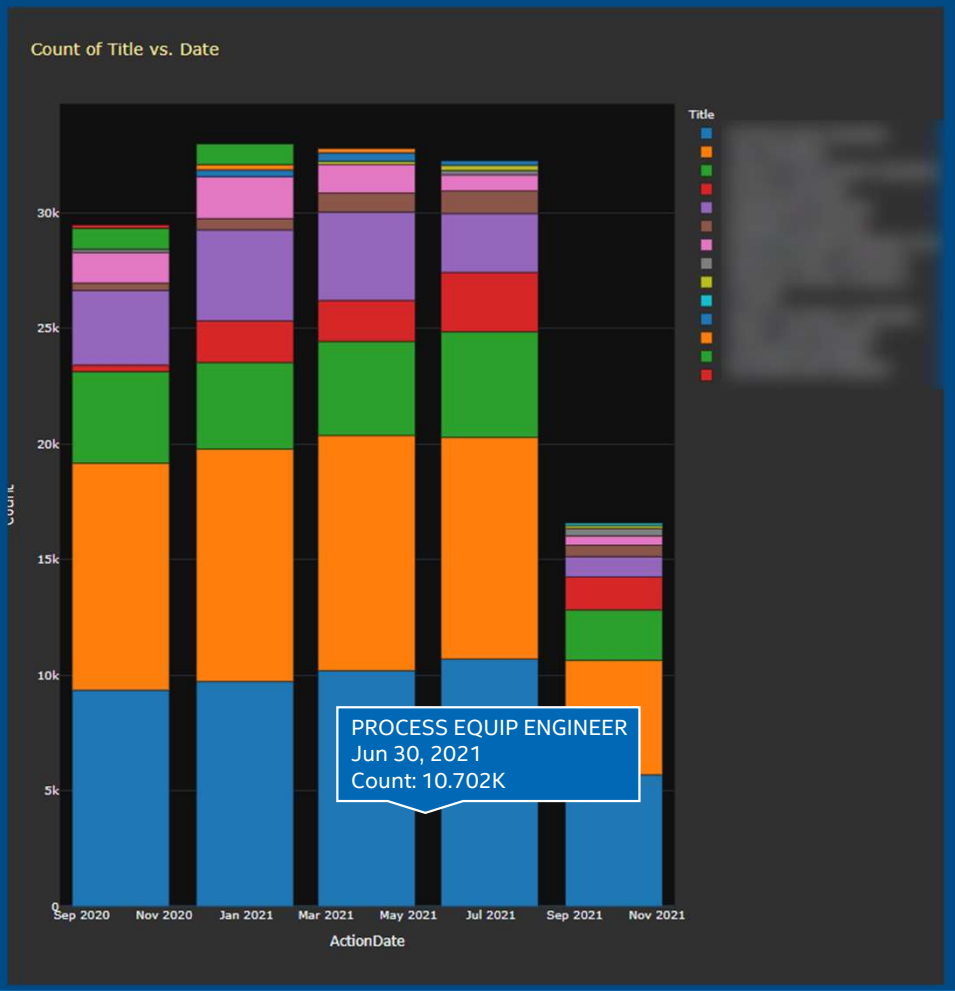
Result

- Filtered data by Module and put into dataframe
- Split data by quarter
- Aggregated the last 5 quarter's data
- Plotted LongID/Country/Title by Action Date and aggregated count in bar graph



Now, we are able to see the user data much more clearly.





What is a Service log? What is an ETL log?

Ship Sort

OCT

Service log:

Documents requests initiated by the user. Contains details about what they requested to see, how long it took to process, and if it send back a valid response.

ETL log:

Script that routinely pulls data from individual app databases. Reports metrics such as runtime and rows packed.

UI

Third type of script used, but not relevant to this project.

Extracting Service log data

Unedited log file

```
2021-06-15 13:52:49,399 - INFO - ShipSort 785 rows packed.
2021-06-15 13:52:49,399 - INFO - ShipSort Response In 0.016 Seconds.
2021-06-15 13:56:12,795 - INFO - Returned ShipSort Meta in 0.000 Seconds.
2021-06-15 13:56:15,700 - INFO - ShipSort 785 rows packed.
2021-06-15 13:56:15,700 - INFO - ShipSort Response In 0.016 Seconds.
```

Revised log file

```
***
2021-06-15 13:52:49,399 - INFO - Rows packed: 785.
2021-06-15 13:52:49,399 - INFO - Runtime: 0.016 Seconds.
2021-06-15 13:56:12,795 - INFO - Returned ShipSort Meta in 0.000 Seconds.
***
2021-06-15 13:56:15,700 - INFO - Rows Packed: 785.
2021-06-15 13:56:15,700 - INFO - Runtime: 0.016 Seconds.
***
```

Visualizing Service log Data

Dataframe:

Unnamed: 0	ActionDate	Grouping	Filter	Idsid	UUID	Rows Packed	Runtime

Plotted Data:

Extracting ETL data

Unedited log file

```
2021-06-22 11:58:42,155 - INFO - Configured and started logging system.
2021-06-22 11:58:42,155 - INFO - Starting Oct ETL in WIP Environment.
2021-06-22 11:58:42,155 - INFO - Start Oct ETL
2021-06-22 12:03:45,661 - INFO - Queried 90 days of OCT Data Loaded To MIDAS in 303.51 Seconds.
2021-06-22 12:03:46,650 - INFO - Queried 1825 days of OCT Data Loaded To EATS in 0.99 Seconds.
2021-06-22 12:03:46,760 - INFO - Queried OCT Data Loaded To EATSOPER in 0.11 Seconds.
2021-06-22 12:04:52,950 - INFO - Queried 90 days of OCT Data Loaded To EATSSTAGGING in 66.19 Seconds.
2021-06-22 12:05:02,299 - INFO - Number of rows in the merged data 7713421.
2021-06-22 12:05:20,464 - INFO - Number of rows in the xeus data 1770335.
2021-06-22 12:05:20,464 - INFO - Number of rows in the appended data 9483756.
2021-06-22 12:05:20,778 - INFO - Number of unique lots in the xeus data 162439.
2021-06-22 12:05:20.778 - INFO - Number of unique lots in the merged data 1297601.
[redacted]
2021-06-22 12:14:53,298 - INFO - Oct ETL Run Time: 971.14 Seconds
2021-06-22 12:58:42,151 - INFO - Configured and started logging system.
2021-06-22 12:58:42,151 - INFO - Starting Oct ETL in WIP Environment.
2021-06-22 12:58:42,151 - INFO - Start Oct ETL
2021-06-22 13:06:36,203 - INFO - Queried 90 days of OCT Data Loaded To MIDAS in 474.05 Seconds.
2021-06-22 13:06:37,225 - INFO - Queried 1825 days of OCT Data Loaded To EATS in 1.02 Seconds.
2021-06-22 13:06:37,336 - INFO - Queried OCT Data Loaded To EATSOPER in 0.11 Seconds.
2021-06-22 13:07:50,627 - INFO - Queried 90 days of OCT Data Loaded To EATSSTAGGING in 73.29 Seconds.
2021-06-22 13:08:00,093 - INFO - Number of rows in the merged data 7718943.
2021-06-22 13:08:18,302 - INFO - Number of rows in the xeus data 1770178.
2021-06-22 13:08:18,302 - INFO - Number of rows in the appended data 9489121.
2021-06-22 13:08:18,619 - INFO - Number of unique lots in the xeus data 162429.
2021-06-22 13:08:18,619 - INFO - Number of unique lots in the merged data 1298461.
[redacted]
2021-06-22 13:18:14,676 - INFO - Oct ETL Run Time: 1172.52 Seconds
```

Revised log file

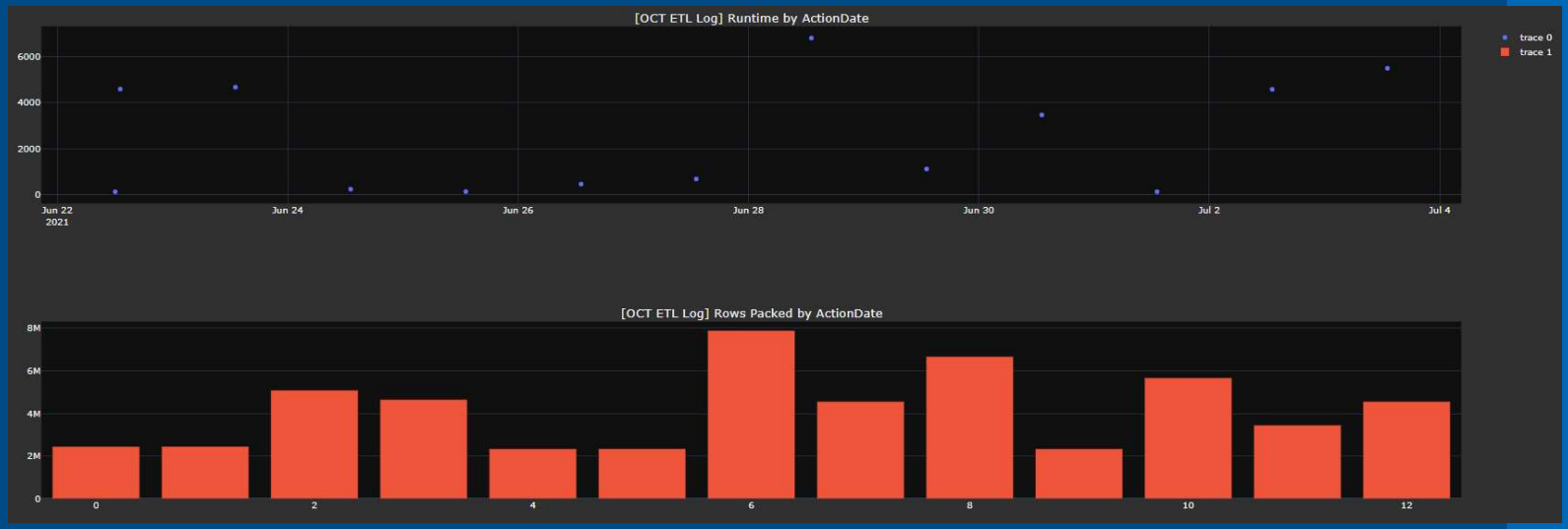
```
2021-06-22 12:03:45,661 - INFO - Queried MIDAS: 303.51 Seconds
2021-06-22 12:03:46,650 - INFO - Queried EATS: 0.99 Seconds
2021-06-22 12:03:46,760 - INFO - Queried EATSOPER: 0.11 Seconds
2021-06-22 12:04:52,950 - INFO - Queried EATSSTAGGING: 66.19 Seconds
2021-06-22 12:05:02,299 - INFO - Num merged rows: 7713421
2021-06-22 12:05:20,778 - INFO - Unique lots: 1297601
2021-06-22 12:14:53,298 - INFO - Oct ETL Run Time: 971.14 Seconds
***
2021-06-22 13:06:36,203 - INFO - Queried MIDAS: 474.05 Seconds
2021-06-22 13:06:37,225 - INFO - Queried EATS: 1.02 Seconds
2021-06-22 13:06:37,336 - INFO - Queried EATSOPER: 0.11 Seconds
2021-06-22 13:07:50,627 - INFO - Queried EATSSTAGGING: 73.29 Seconds
2021-06-22 13:08:00,093 - INFO - Num merged rows: 7718943
2021-06-22 13:08:18,619 - INFO - Unique lots: 1298461
2021-06-22 13:18:14,676 - INFO - Oct ETL Run Time: 1172.52 Seconds
```

Visualizing ETL Data

Dataframe:

	ActionDate	Rows Packed	Runtime
0	2021-06-22 12:03:45	2454345	123.14
1	2021-06-22 13:06:36	2456760	4576.52
2	2021-06-23 13:06:36	5096500	4657.52
3	2021-06-24 13:06:36	4657899	234.52
4	2021-06-25 13:06:36	2343223	131.52
5	2021-06-26 13:06:36	2346877	456.52
6	2021-06-27 13:06:36	7900778	676.52
7	2021-06-28 13:06:36	4565445	6787.52
8	2021-06-29 13:06:36	6675456	1111.52
9	2021-06-30 13:06:36	2343223	3454.52
10	2021-07-01 13:06:36	5676544	124.52
11	2021-07-02 13:06:36	3454334	4565.52
12	2021-07-03 13:06:36	4564456	5476.52

Plotted Data:



Next Steps

WIP → PROD

We have only parsed OCT and Ship Sort. ~33 other applications whose Service and ETL files must be parsed.

Once in production, these visuals can be used across Intel to track the health and performance of these applications.

Implications: Improve app performance, repair efficiency, user satisfaction, and more

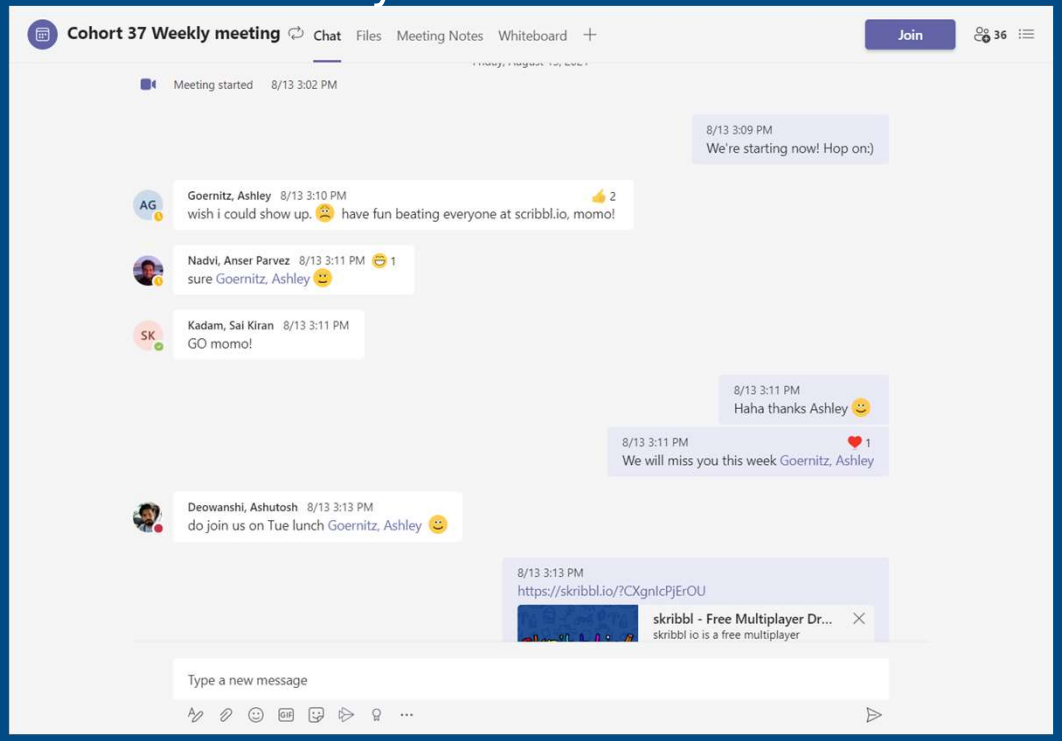
Learnings

- Power BI
- Python queries
- Excel Vlookups
- Joining/Merging datasets
- Jupyter Notebook + Excel
- Displaying data
- Python dataframes
- Plotly & Dash
- Log file parsing
- Jupyter Notebooks
- Agile software development
- Daily standups
- Sprint planning
- Quarterly planning

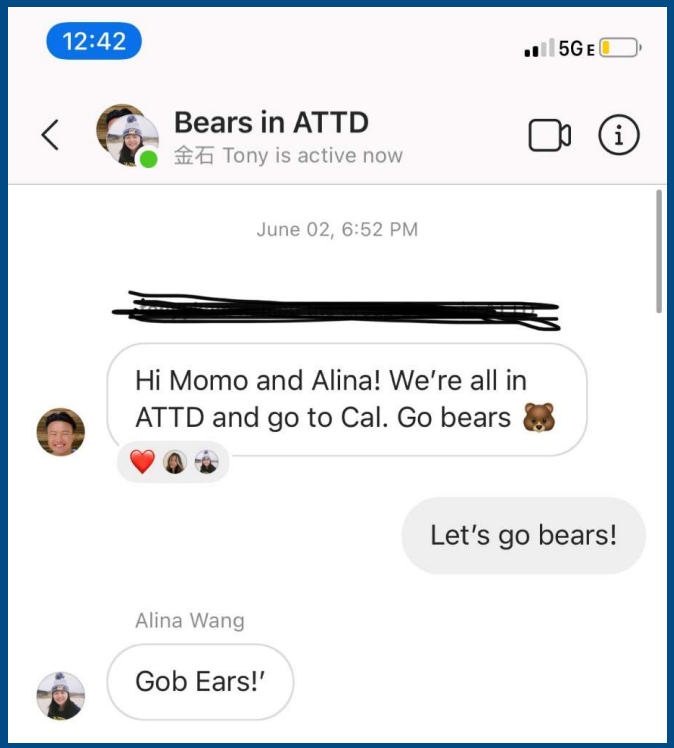
Catalyst Application Health Dashboard Q&A

People at Intel

Weekly Cohort Skribbl.io



Bears at Intel



Acknowledgements

Lisa Pivin
Somaiah Thimmaiah Balekuttira
Chris Maloney
Randall Goodwin
Marie Winnie
Yield Engineering and Analysis Team
Cohort 37 Leads/Co-leads
New Friends at Intel☺

Thank you for listening!