

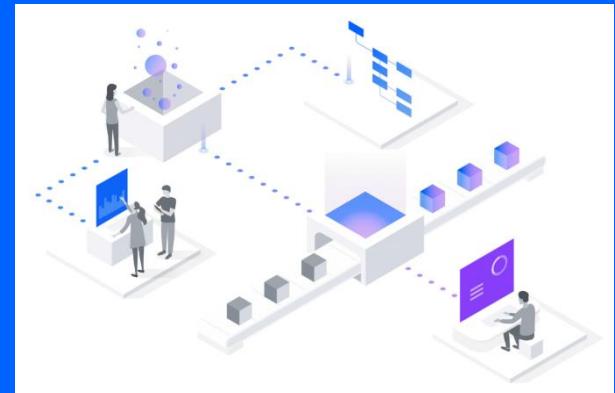
## IBM Business Automation Insights with Business Automation Workflow

Virtual Technical Workshop for EMEA Business Partners  
November/December 2020

---

Radek Šulc, radek\_sulc@cz.ibm.com  
Digital Business Automation Software Engineer  
Technical Engagement Team, IBM EMEA

---



Credits to **Paul Pacholski**, IBM SWAT Team and many other IBMers working on the labs  
Special thanks to **Reinhold Engelbrecht**, Technical Sales Leader EMEA - IBM Digital Business Automation - for logistic support

Latest version of the materials: <http://ibm.biz/bai-workshop-materials>

# Agenda of the day (CET time) with materials

- 09:30-11:00
    - Welcome and overview of the day
    - Overview IBM BAI with live demo
    - How to access your lab environment
    - Introduction to the lab
  - 11:00-12:15
    - Lab: Setup BAI events for BPMN, Case and ODM with Kibana dashboards – Part 1
  - 12:15-12:30
    - Demo of the new Performance Center
  - 12:30-13:30
    - Lunch break
  - 13:30-15:00
    - Introduction to the lab
    - Lab: Setup BAI events for BPMN, Case and ODM with Kibana dashboards – Part 2
  - 15:00-15:30
    - IBM BAI, BAW and ODM with Machine Learning – introduction, demonstration
  - 15:30-16:30
    - Lab: IBM BAI, BAW and ODM with Machine Learning
  - 16:30-17:00
  - Q&A session, Closing
- Items in RED are performed together on-line**
- Workshop materials with labs:  
<http://ibm.biz/bai-workshop-materials>
  - All attendees are on mute by default
  - The instructor will be available all the time for your questions and troubleshooting
  - Do not hesitate to speak up, chat, ask questions
  - Please be informal 😊
  - Take a minute for after-survey  
<https://ibm.biz/bai-workshop-survey>
  - Your feedback is important to us!

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# What is IBM Business Automation Insights?



## COLLECT

Scalable cloud capability to capture end to end business data from DBA platform components to operational data store and long-term store (data lake)

Data for use to gain operational insights and machine learning opportunities



## VISUALIZE

Provides real-time operational visibility to Business Managers via Dashboards: (i) pre-defined operational and IT authored and managed (Kibana); (ii) Business User authored and managed (Business Performance Center) New

Gain visibility and deep understanding of business operations running on the automation platform



## LEARN & GUIDE

Correlate and measure the data based on collected business and operational metrics.

Apply MLs to the operational data lake to make recommendations to business managers and knowledge workers

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# Hyperautomation

#1

Top Strategic Technology  
Trend For 2020

“Hyperautomation deals with the application of advanced technologies, including artificial intelligence (AI) and machine learning (ML), to increasingly **automate processes and augment humans.**”

“**RPA and Intelligent Business Process Management Suites** are key components of hyperautomation.”

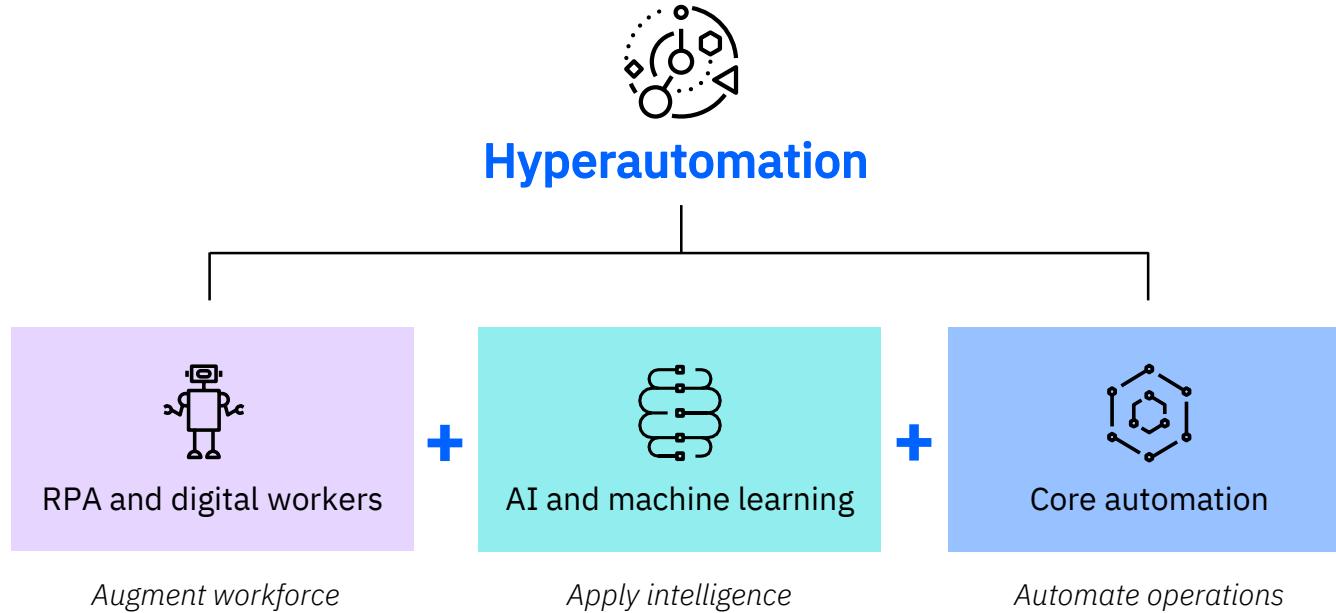
Gartner

2019, [Gartner Top 10 Strategic Technology Trends for 2020](#)

Published October 21, 2019

# Gartner®

# IBM Hyperautomation



# Components of IBM hyperautomation



## RPA and digital workers

- Perform repetitive tasks
- Act on data extracted from unstructured documents
- Converse via chatbots using text and voice response



## AI and machine learning

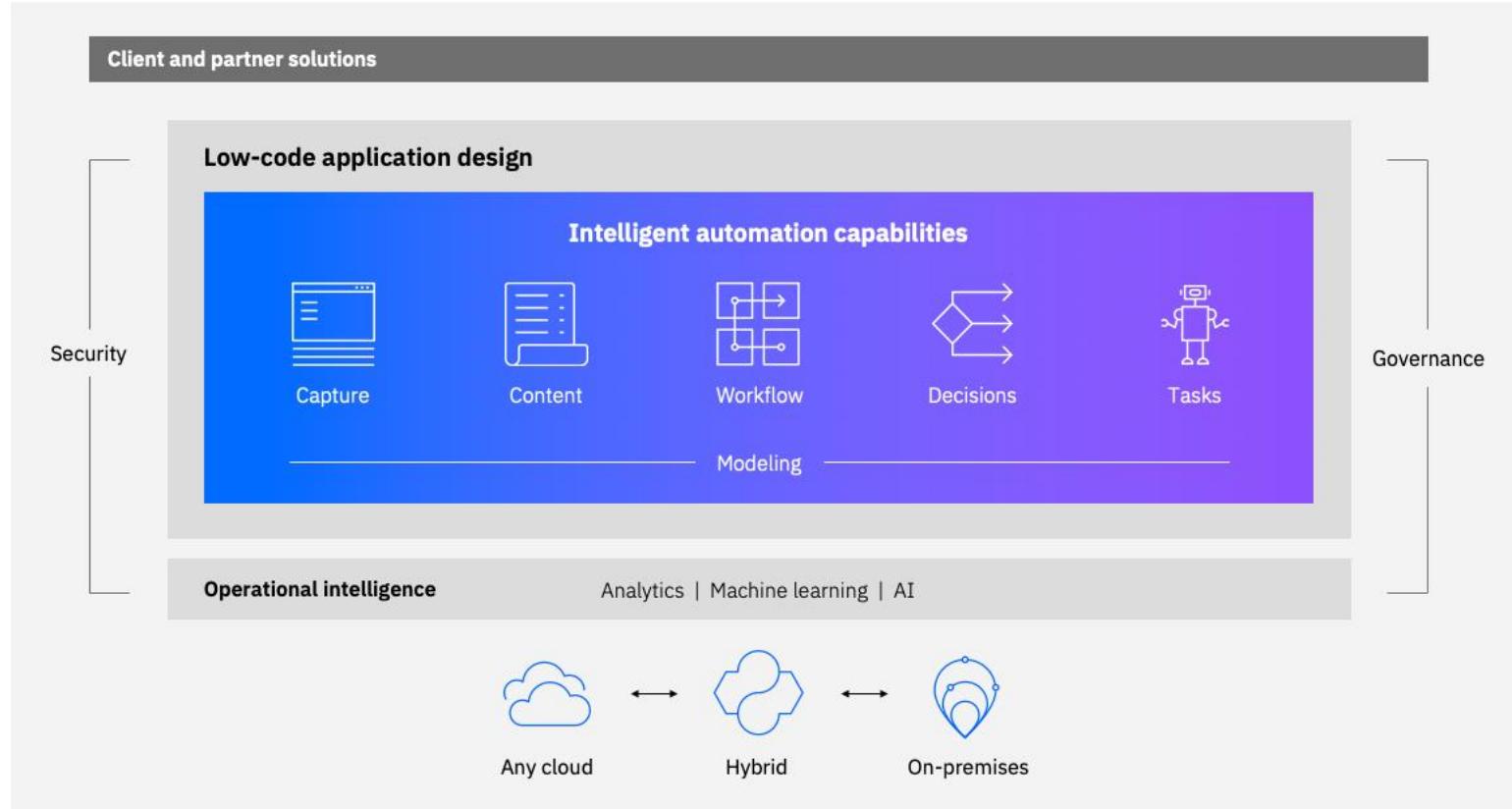
- Classify and understand content
- Prioritize tasks
- Make predictive decisions
- Build intelligent chatbots
- Apply AI to operational data



## Core automation

- Capture and digitize content
- Manage work – processes, cases
- Automate decisions
- Build low-code applications
- Monitor and report

# IBM Cloud Pak for Automation



# Cloud Paks – Enterprise-ready containerized software

*A faster, more secure way to move your core business applications to any cloud through enterprise-ready containerized software solutions*

## IBM containerized software

Packaged with Open Source components, pre-integrated with the common operational services, and secure by design



## Container platform and operational services

Logging, monitoring, security, identity access management



AWS

MS Azure

Google Cloud



## Complete yet simple

*Application, data and AI services that are modular, term licensed, and easy to consume*

## IBM certified

*Full software stack support, and ongoing security, compliance and version compatibility*

## Run anywhere

*On-premises, on private and public clouds, and in pre-integrated systems*

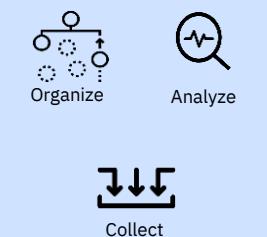
# Cloud Paks – Pre-integrated for cloud use cases

*Today, IBM offers clients the first six Cloud Paks...*

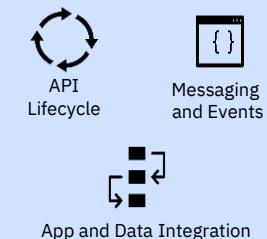
## Cloud Pak for Applications



## Cloud Pak for Data



## Cloud Pak for Integration



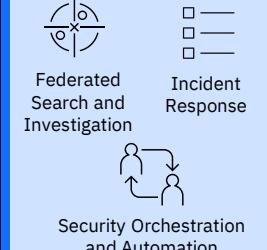
## Cloud Pak for Automation



## Cloud Pak for Multicloud Management



## Cloud Pak for Security



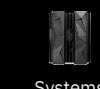
Container platform and operational services



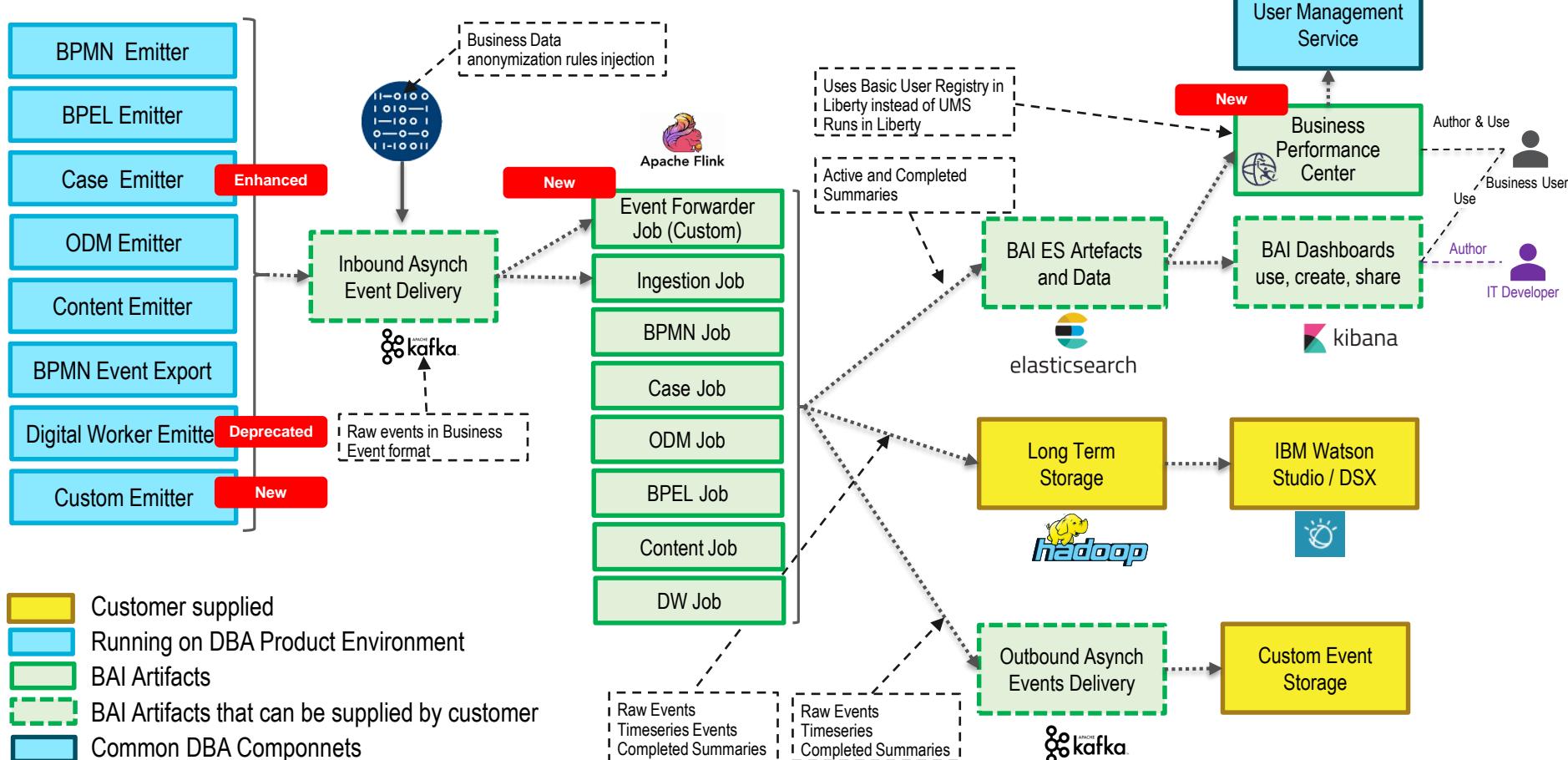
AWS

MS Azure

Google Cloud



# Business Automation Insights Architecture



# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

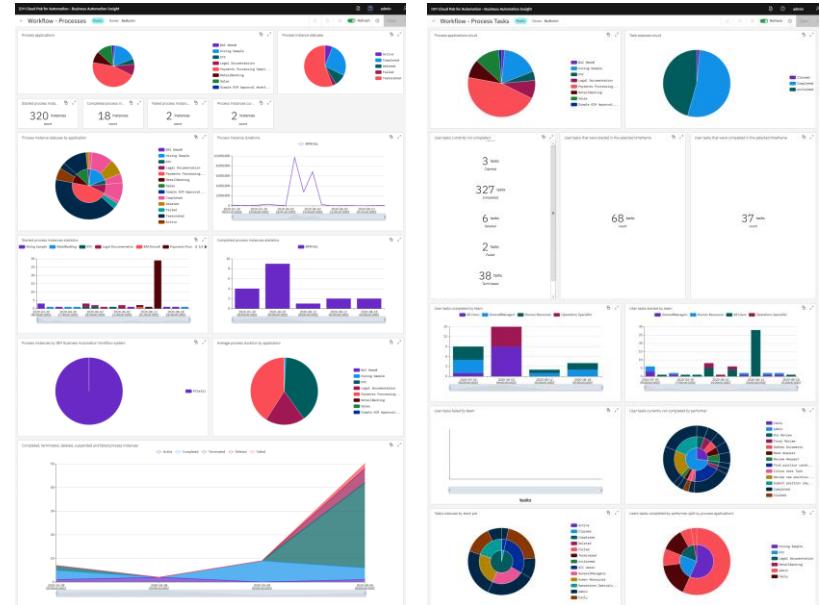
What's New

Summary

# What's in the Box?

- Predefine ELK artifacts
  - ELK Indices
  - Saved Searches
  - Visualization
  - Dashboards for most DBA Pillars

- Business Performance Center New  
<https://ibm.box.com/v/BusinessPerformanceCenter>



# Elasticsearch Indices

During install a Kubernetes job creates all Elasticsearch indices

Kubernetes Elasticsearch cluster in shards in K8s Persistent volumes

Elasticsearch shards in K8s Persistent volumes

**Elasticsearch** https://9.30.188.25/elasticsearch-ibm-dba-ek-ba Connect bai-tecsales-ibm-dba-ek-elasticsearch cluster health: green (50 of 50) Info

Overview Indices Browser Structured Query [+] Any Request [+] Cluster Overview Sort Cluster Sort Indices View Aliases Index Filter Refresh

process-summaries-completed-idx-ibm-bai-2018.08.22-000001  
size: 638ki (1.25Mi)  
docs: 48 (96)  
Info Actions

process-summaries-active-idx-ibm-bai-2018.08.22-000001  
size: 497ki (993ki)  
docs: 25 (145)  
Info Actions

case-summaries-completed-idx-ibm-bai-2018.08.22-000001  
size: 39.9ki (67.1ki)  
docs: 3 (10)  
Info Actions

case-summaries-active-idx-ibm-bai-2018.08.22-000001  
size: 81.5ki (165ki)  
docs: 18 (81)  
Info Actions

.kibana  
size: 352ki (704ki)  
docs: 105 (285)  
Info Actions

bai-tecsales-ibm-dba-ek-client-8487c4fc5d-q8545  
Info Actions

bai-tecsales-ibm-dba-ek-data-0  
Info Actions

bai-tecsales-ibm-dba-ek-data-1  
Info Actions

bai-tecsales-ibm-dba-ek-master-7f86dfbd96-76b2d  
Info Actions

undefined x

# Elasticsearch Summaries Index Mappings

## Mapping

Mapping is the process of defining how a document, and the fields it contains, are stored and indexed. For instance, use mappings to define:

- which string fields should be treated as full text fields.
- which fields contain numbers, dates, or geolocations.
- whether the values of all fields in the document should be indexed into the catch-all `_all` field.
- the **format** of date values.
- custom rules to control the mapping for **dynamically added fields**.

Includes Index Mappings  
for Completed  
Summaries

```
{ "process-summaries-completed-ibm-bai": { "state": "open", "settings": { "process-summaries": { "index": { "creation_date": "1534928293905", "number_of_shards": "5", "number_of_replicas": "1", "uuid": "WEGA0M5WRaKPs3AYE4TFwg", "version": { "created": "6030199" }, "provided_name": "<process-summaries-completed-idx-ibm-bai-{now/d}-000001>" } }, "mappings": { "process-summary": { "dynamic": "true", "dynamic_templates": [ { "data_key": { "path_match": "*_.key", "mapping": { "type": "keyword", "index_options": "docs" } } ], "data_string": { "path_match": "*_.string", "mapping": { "omit_norms": true, "fields": { "keyword": { "type": "keyword" } } } } } } } } }
```

# Elasticsearch Index Patterns

## Defining your index patterns

Index patterns tell Kibana which Elasticsearch indices you want to explore. An index pattern can match the name of a single index, or include a wildcard (\*) to match multiple indices.

The screenshot shows the Kibana Management interface with the 'Management' tab selected. The left sidebar lists various management tools: Discover, Visualize, Dashboard, Timelion, Dev Tools, and Management. The 'Management' tool is highlighted with a blue background. The main content area is titled 'Management / Kibana' and shows the 'Index Patterns' tab is active. A button '+ Create Index Pattern' is visible. A list of existing index patterns is shown, including:  
★ \*summaries\*  
\*-summaries-\*  
\*rocess-summarie\*  
\*rocess-summaries\*  
\*rocess-summaries-\*  
case-summarie\*  
case-summaries\*  
case-summaries-\*  
process-summ\*  
process-summa\*  
process-summar\*  
process-summarl\*  
process-summarie\*  
process-summaries\*  
process-summaries-\*

The page title is 'process-summa\*'. A green box highlights the 'Time Filter field name: resumedTime'. Below it, a message states: 'This page lists every field in the process-summa\* index and the field's associated core type as recorded by Elasticsearch'. There are three tabs: 'Fields (109)', 'Scripted fields (0)', and 'Source filters (0)'. A 'Filter' input field is present. A table lists the fields with their types and formats:

Name	Type	Format
_id	string	String
_index	string	
_score	number	
_source	_source	
_type	string	

# Kibana Visualizations & Searches

A screenshot of the Kibana interface. At the top, there are three tabs: 'Dashboards (5)', 'Searches (24)', and 'Visualizations (60)'. The 'Searches (24)' tab is highlighted with a blue border. Below the tabs is a search bar with the placeholder text 'Search...'. Under the search bar are several filter options represented by checkboxes:

- Title
- Active [do not change]
- Active Cases [do not change]
- Active Tasks [do not change]
- CaseSummary [do not change]
- Completed [do not change]
- Completed Cases [do not change]

## Searching Your Data

You can search the indices that match the current index pattern by entering your search criteria in the Query bar. You can use Kibana's standard query language (based on Lucene [query syntax](#)) or the full JSON-based [Elasticsearch Query DSL](#). Autocomplete and a simplified query syntax are available for the Kibana query language as experimental features which you can opt-in to under the options menu in the Query Bar.



A screenshot of the Kibana interface. At the top, there are three tabs: 'Dashboards (5)', 'Searches (24)', and 'Visualizations (60)'. The 'Visualizations (60)' tab is highlighted with a blue border. Below the tabs is a search bar with the placeholder text 'Search...'. Under the search bar are several visualization options represented by checkboxes:

- Title
- Average Age of Active Cases [do not change]
- Average Age of Active Tasks [do not change]
- Average Case Duration [do not change]
- Average Elapsed Time of Active Tasks by State [do not change]
- Average Elapsed Time of Completed Cases [do not change]
- Average Elapsed Time of Completed Tasks [do not change]

## Visualize

[Visualize](#) enables you to create visualizations of the data in your Elasticsearch indices. You can then build [dashboards](#) that display related visualizations.

Kibana visualizations are based on Elasticsearch queries. By using a series of Elasticsearch [aggregations](#) to extract and process your data, you can create charts that show you the trends, spikes, and dips you need to know about.

You can create visualizations from a search saved from [Discover](#) or start with a new search query.

# Out of the Box DBA Kibana Dashboards

## Dashboards

[Create new dashboard](#)

Search...

<input type="checkbox"/> Title ↑	Description	Actions
<input type="checkbox"/> Automation Digital Worker	} Digital Worker	Edit
<input type="checkbox"/> Content Dashboard	} ECM: Content	Edit
<input type="checkbox"/> Decisions Dashboard	} ODM: Decisions	Edit
<input type="checkbox"/> Workflow - BAW Advanced Activities		Edit
<input type="checkbox"/> Workflow - BAW Advanced Order Processing Sample		Edit
<input type="checkbox"/> Workflow - BAW Advanced Processes		Edit
<input type="checkbox"/> Workflow - BAW Advanced Tasks		Edit
<input type="checkbox"/> Workflow - Case Activities		Edit
<input type="checkbox"/> Workflow - Cases		Edit
<input type="checkbox"/> Workflow - Hiring Sample		Edit
<input type="checkbox"/> Workflow - Process Tasks		Edit
<input type="checkbox"/> Workflow - Processes		Edit

Rows per page: 20 ▾

The screenshot shows a list of dashboards in a Kibana interface. The dashboards are listed in rows, each with a checkbox, a title, a description, and an 'Edit' button. Red annotations have been added to group certain dashboards: a bracket on the left groups the first three items ('Automation Digital Worker', 'Content Dashboard', 'Decisions Dashboard') under the heading 'Digital Worker'; another bracket on the left groups the next five items ('Workflow - BAW Advanced Activities', 'Workflow - BAW Advanced Order Processing Sample', 'Workflow - BAW Advanced Processes', 'Workflow - BAW Advanced Tasks', 'Workflow - Case Activities') under the heading 'BAW: BMN & BPEL Processes and Cases'.

<input type="checkbox"/> Title ↑	Description	Actions
<input type="checkbox"/> Automation Digital Worker	} Digital Worker	Edit
<input type="checkbox"/> Content Dashboard	} ECM: Content	Edit
<input type="checkbox"/> Decisions Dashboard	} ODM: Decisions	Edit
<input type="checkbox"/> Workflow - BAW Advanced Activities		Edit
<input type="checkbox"/> Workflow - BAW Advanced Order Processing Sample		Edit
<input type="checkbox"/> Workflow - BAW Advanced Processes		Edit
<input type="checkbox"/> Workflow - BAW Advanced Tasks		Edit
<input type="checkbox"/> Workflow - Case Activities		Edit
<input type="checkbox"/> Workflow - Cases		Edit
<input type="checkbox"/> Workflow - Hiring Sample		Edit
<input type="checkbox"/> Workflow - Process Tasks		Edit
<input type="checkbox"/> Workflow - Processes		Edit

# Dashboards

Process Dashboards

Case Dashboards

Decisions Dashboards

Content Dashboards

# Kibana BPMN Process Dashboard

Dashboard / Process - IBM Business Automation Workflow

Search... (e.g. status:200 AND extension:PHP)

Full screen Share Clone Edit ⌘ Auto-refresh < ⌂ Year to date > Options

Add a filter +

Process applications (Click to apply filter)  
The container is too small to display the entire cloud. Tags might be cropped or omitted.

Hiring Sample for taiga\_accept  
Hiring Sample  
Nested Tasks  
Claim Approval Sample  
Hiring Sample clone for taiga accept  
Count - Process Application

Process instance statuses (Click to apply filter)  
Failed Terminated  
Deleted Completed Active  
Count - Status

Started process instances: 204 started process instances

Completed process instances: 175 completed process instances

Failed process instances: 3 failed process instances

Process instances currently at risk: 0 process instances currently at risk

Process instance statuses by application

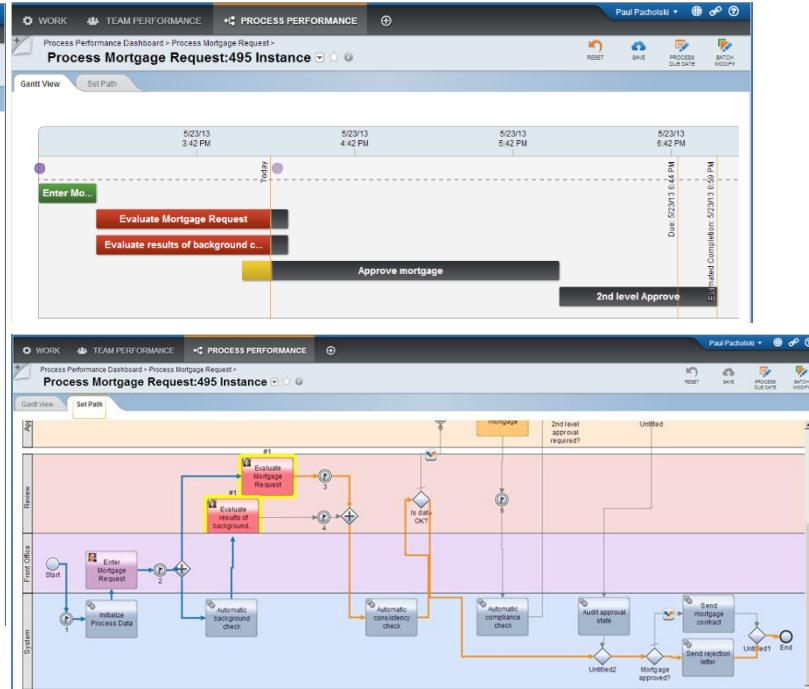
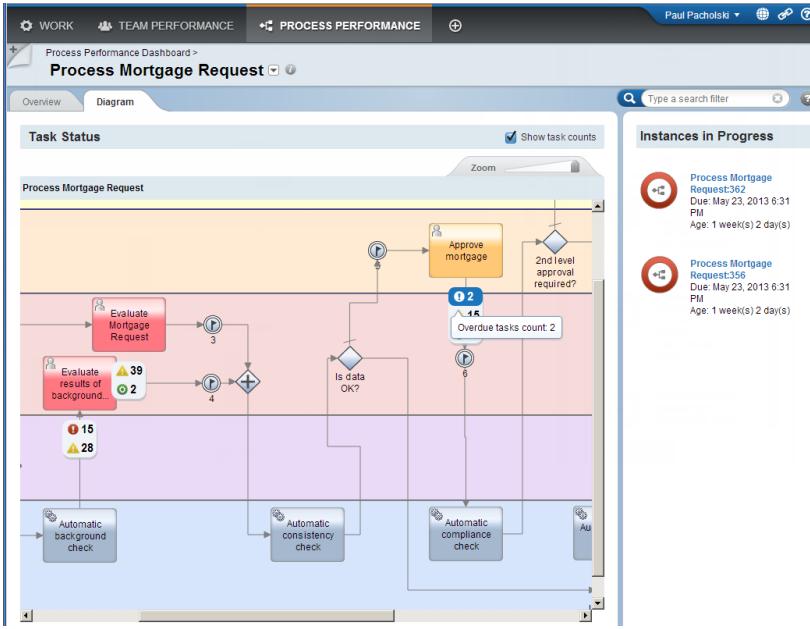
Application	Deleted	Completed	Terminated	Active
app1	~140	~10	~5	~0
app2	~60	~20	~10	~0
app3	~50	~20	~10	~0
app4	~10	~10	~0	~0
app5	~10	~0	~0	~0

Process instance durations

Average duration

# Process Performance

- Most features included in Kibana Processes Dashboard
- But does not include Gant Charts or Process Path and Aggerated Task Status views



# BPEL Process Instances

BAW Advanced - Process instance statuses cloud [do not change]

PROCESS\_COMPLETED  
PROCESS\_TERMINATED PROCESS\_STARTED PROCESS\_DELETED PROCESS\_FAILED

Count - Status

BAW Advanced - Started process instances count [do not change]

31  
Count

BAW Advanced - Completed process instances count [do not change]

27  
Count

BAW Advanced - Failed process instances count [do not change]

1  
Count

BAW Advanced - Process instance statuses by process template [do not change]



BAW Advanced - Process instance durations for completed processes [do not change]



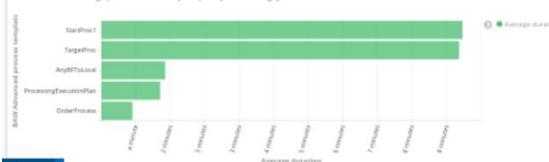
BAW Advanced - Started process instances statistics [do not change]



BAW Advanced - Completed process instances statistics [do not change]



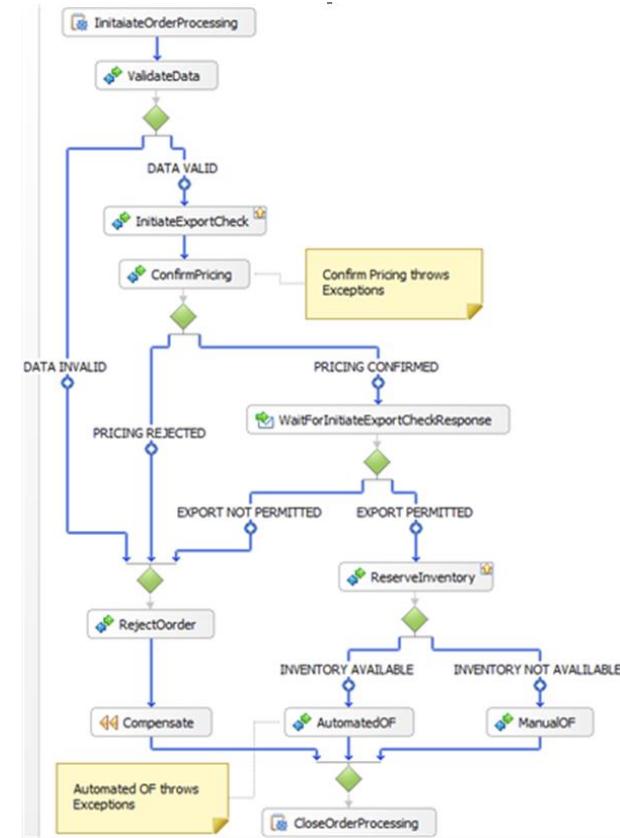
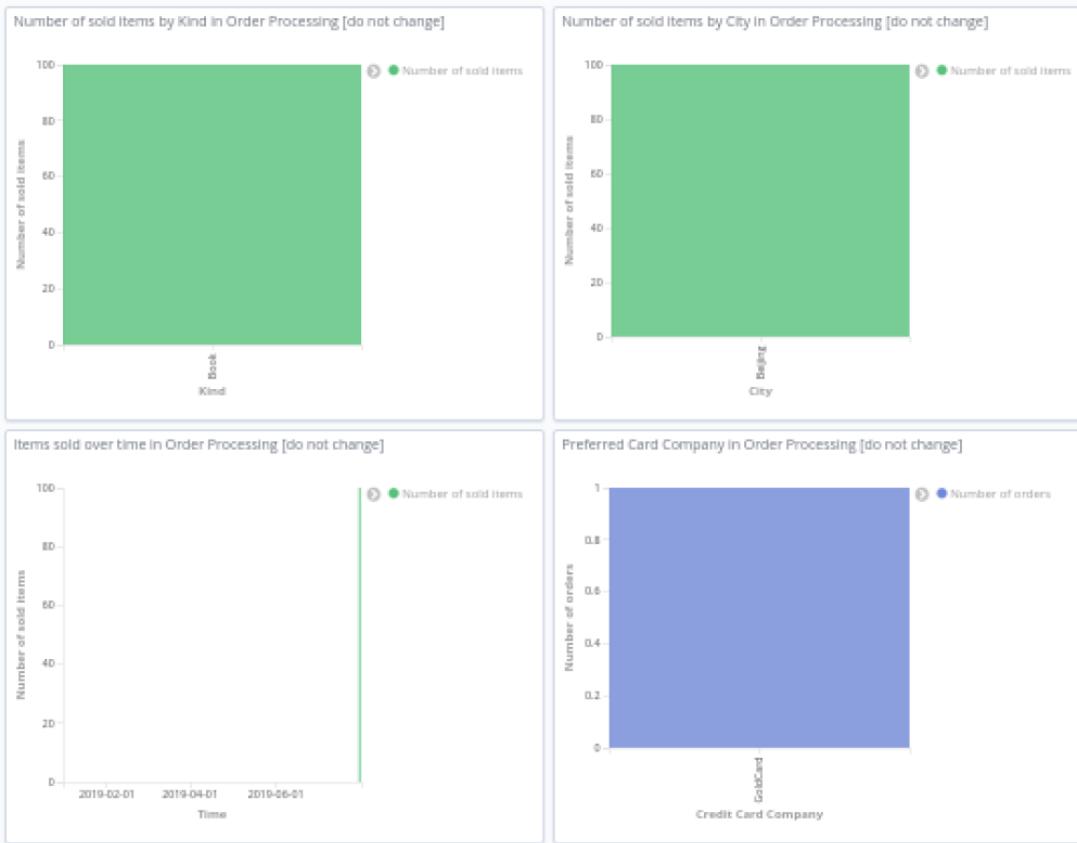
BAW Advanced - Average process duration by template [do not change]



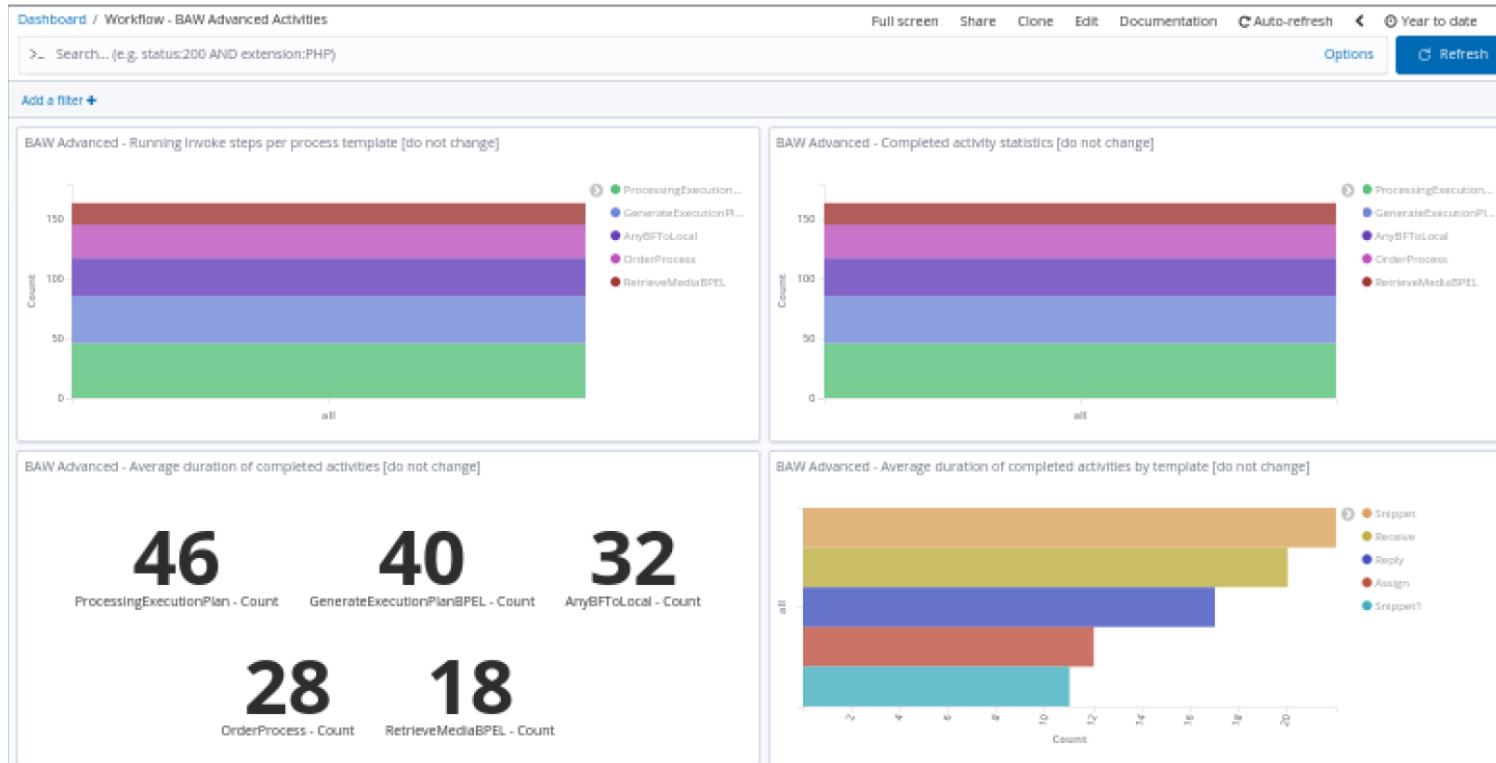
BAW Advanced - Completed, terminated, deleted, suspended and failed process instances [do not change]



# Example of a Custom BPEL Process Dashboards



# BPEL Process Activities



# Kibana BPMN Process Tasks Dashboard

Dashboard / Process Task - IBM Business Automation Workflow

Search... (e.g. status:200 AND extension:PHP)

Full screen Share Clone Edit ⚡ Auto-refresh ⏪ ⏴ Year to date ⏵ Options 🔍

Add a filter +

Process applications (click to apply filter)

Hiring Sample Advanced

# Hiring Sample

Procurement Sample

Count - Process Application

Number of user tasks that are currently not completed

13

not completed user tasks

Task statuses (click to apply filter)

Claimed

# Completed

Unclaimed

Count - state.keyword: Descending

User tasks that were started in the selected time...

45

started user tasks

User tasks that were completed in the selected ti...

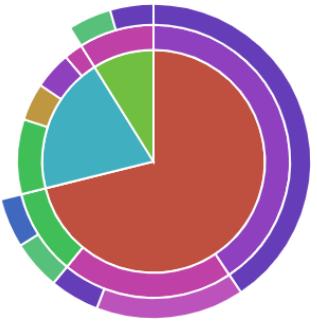
32

completed user tasks

Collapse

# Kibana BPMN Process Tasks Dashboard

**Tasks statuses by team**



- Completed
- Unclaimed
- Claimed
- GeneralManagers
- All Users
- Human Resources
- General Managers
- Laura
- deadmin
- Erick
- Sandra

**Users tasks completed by performer split by process applications**

- Laura
- deadmin
- Patti
- John
- Michael

**User tasks that are currently waiting to be processed**

startTime	state	claimedTime	processApplicationName	name	potentialPerformerName	performerName
August 28th 2018, 12:18:05.569	Unclaimed	-	Hiring Sample	Find position candidates	Human Resources	-
August 28th 2018, 12:04:23.270	Unclaimed	-	Hiring Sample	Review new position request	GeneralManagers	-
August 28th 2018, 11:15:31.469	Unclaimed	-	Hiring Sample	Review new position request	GeneralManagers	-
August 28th 2018, 04:35:18.216	Unclaimed	-	Hiring Sample Advanced	Approve / reject requisition	General Managers	-
August 28th 2018, 04:32:39.918	Unclaimed	-	Hiring Sample Advanced	Approve / reject requisition	General Managers	-
August 28th 2018, 04:31:01.149	Claimed	August 28th 2018, 04:31:15.289	Procurement Sample	ApproveReplenishmentOrder	All Users	Laura
August 27th 2018, 10:11:11.817	Claimed	August 27th 2018, 10:11:13.809	Hiring Sample	Submit position request	-	deadmin
August 27th 2018, 10:06:10.185	Claimed	August 27th 2018, 10:06:10.773	Hiring Sample	Submit position request	-	deadmin

# BPMN Process Tasks

- Equivalent to Team Performance
- But does not include any actionable capabilities such as Modify Task, View Process Diagram, Assign Task, etc...

The screenshot displays a user interface for managing BPMN process tasks. At the top, there are three tabs: WORK, TEAM PERFORMANCE, and PROCESS PERFORMANCE. The TEAM PERFORMANCE tab is active, showing a 'Team Performance Dashboard' with four pie charts:

- Mortgage Approvers:** 4 Overdue (red), 12 At Risk (yellow), 5 On Track (green).
- Senior Mortgage Approvers:** 1 Overdue (red), 1 At Risk (yellow), 1 On Track (green).
- Reviewers:** 3 Overdue (red), 33 At Risk (yellow), 2 On Track (green).
- Front Office:** 4 Overdue (red), 24 At Risk (yellow), 1 On Track (green).

Below the dashboard, the 'Mortgage Approvers' section is expanded, showing 'Open Tasks' and 'Overdue (3)' tasks:

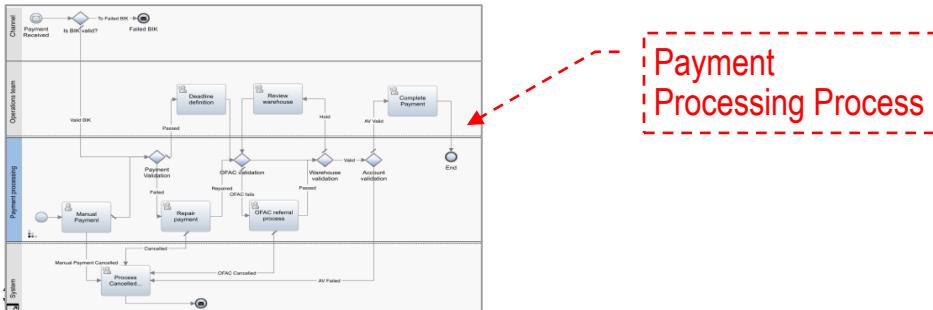
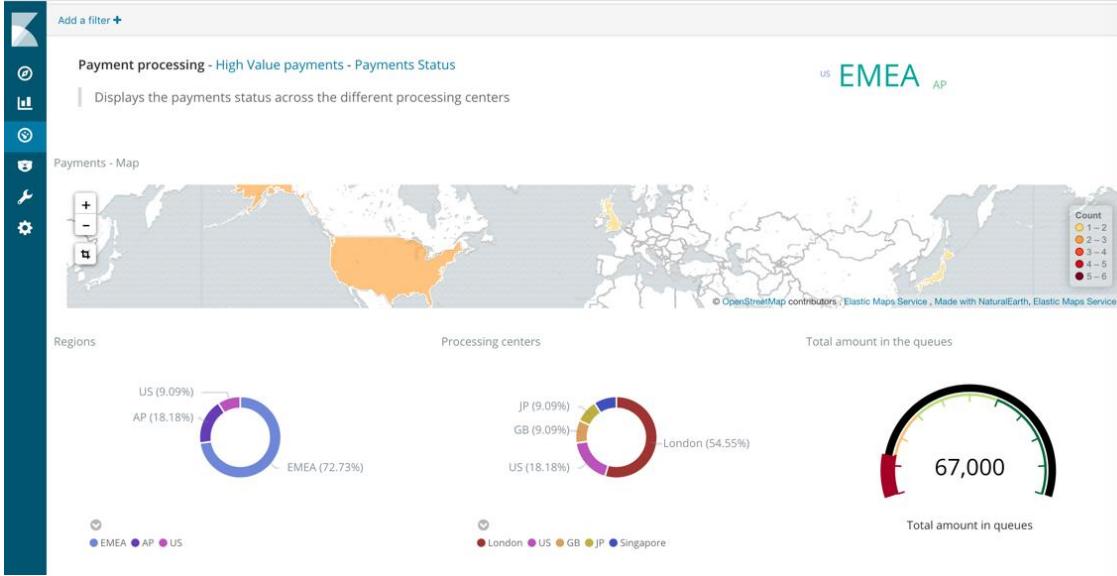
- Step: Approve mortgage** (Process Mortgage Request: 356) - Due: May 22, 2013 2:58 PM
- Step: Approve mortgage** (Process Mortgage Request: 355) - Due: May 22, 2013 2:58 PM

A context menu is open over the first overdue task, listing options: Modify Task, Reassign Back to Team, View Instance, Modify Instance, View Audit History, and View Process Diagram.

To the right, a modal window titled 'Assign: Step: Approve mortgage' shows the assignee 'Paul Pacholski' and an 'Assign' button.

At the bottom right, a 'Roster' section lists 'Allen Makowski' with 'Assigned Tasks: 1' and 'Tasks Completed Today: 5'.

# Example of a Custom Process Dashboard



# Dashboards

Process Dashboards

Case Dashboards

Decisions Dashboards

Content Dashboards

# Kibana Case Dashboard

**Cases in Progress** A bar graph in which each bar shows the number of cases that are active for a case type. When you hover on a bar in the graph, a tooltip displays the solution name, case type name and the number of cases in progress.

**Cases in Each State** A pie chart shows the proportion of the cases in each of two states: working and Complete. This is further split across case type and Solution Name. When you hover on a sector of the pie, a tooltip displays the number of cases in that state, case state, case type and solution name.

**Average Age of Active Cases** A bar graph in which each bar shows the average age of active cases for each case type. The average age is calculated by adding the elapsed time for each case and dividing by the number of cases. When you hover on a bar in the graph, a tooltip displays the case type name, solution name and the average age.

*Note: Due to the Kibana constraints, the elapsed time of an in-Progress case is taken as the difference between the time of the last processed event on that case and the start time of that case.*

**Average Elapsed Time of Completed Cases widget** A line graph that shows the average elapsed time for each case type of the cases that are completed. When you hover on a point in the line, a tooltip displays the solution name, case type name and the average elapsed time of completed cases.

**Incoming and Outgoing Cases in Past 3 Months widget** A bar graph in which two bars show the number of cases that opened and closed in the past three months for a case type. When you hover of a bar in the graph, a tooltip indicates whether the bar represents incoming or outgoing cases. The tooltip also displays the case type name and the number of incoming or outgoing cases.

**Total Cases Count** The count of total cases.

**Average Case Duration** The average duration of the cases in seconds.

**Cases Started Statistics** The cases that have been started based on the selected time frame in Kibana.

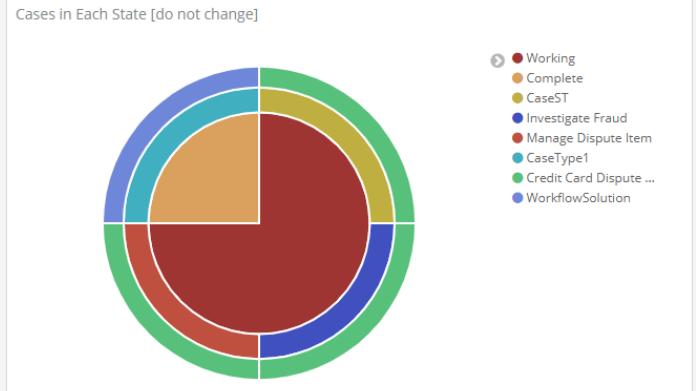
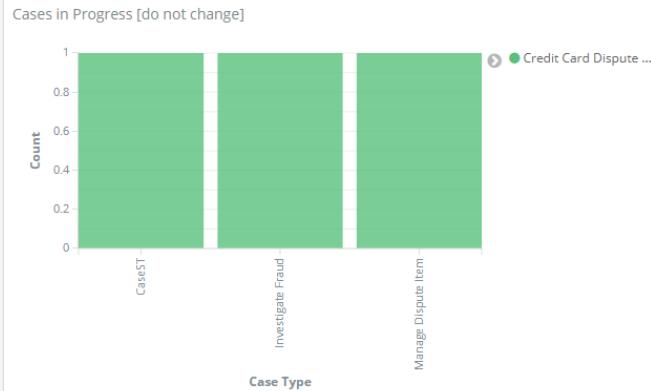
**Cases Completed Statistics** The completed tasks based on the selected time frame in Kibana.

# Kibana Case Dashboard

kibana

- Discover
- Visualize
- Dashboard**
- Timelion
- Dev Tools
- Management

Collapse



# Kibana Case Task Dashboard

**Tasks in Progress** A bar graph in which each bar shows the number of tasks that are active for a task type. When you hover on a bar in the graph, a tooltip displays the case type, task type name and the number of tasks in progress.

**Tasks in Each State** A pie chart that shows the proportion of tasks in each of the states: waiting, ready, working, complete and Failed. When you hover on a sector of the pie, a tooltip displays the number of tasks in that state. The tooltip also displays the case type and task type.

**Average Age of Active Tasks** A bar graph in which each bar shows the average age of active tasks for a task type. The average age is calculated by adding the elapsed time for each task and dividing by the number of tasks. When you hover on a bar in the graph, a tooltip displays the task type, case type and the average age. *Note: Due to the Kibana constraints, the elapsed time of an active task is taken as the difference between the time of the last processed event on that task and the start time of that task.*

**Average Elapsed Time of Active Tasks by State** A bar graph in which each bar shows for a task type the average elapsed time of active tasks in a particular state. When you hover on a bar in the graph, a tooltip displays the task type name, task state, and the average elapsed time of that task type in particular state. *Note: Due to the Kibana constraints, the elapsed time of an active task is taken as the difference between the time of the last event on that task and the start time of that task.*

**Incoming and Outgoing Tasks in Past 3 Months** A bar graph in which two bars show the number of tasks that opened and closed in the past three months for a task type. When you hover on a bar in the graph, a tooltip indicates whether the bar represents incoming or outgoing tasks. The tooltip also displays the task type name and the number of incoming or outgoing tasks.

**Average Elapsed Time of Completed Tasks** A line graph that shows the average elapsed time for each task type of the tasks that completed. When you hover on a point in the line, a tooltip displays the case type name, task type name and the average elapsed time of completed tasks.

**Total Tasks Count** The count of total case tasks.

**Average Tasks Duration** The average duration of the case tasks in seconds.

**Task Started Statistics** The tasks that have been started based on the selected time frame in Kibana.

**Task Completed Statistics** The completed tasks based on the selected time frame in Kibana.

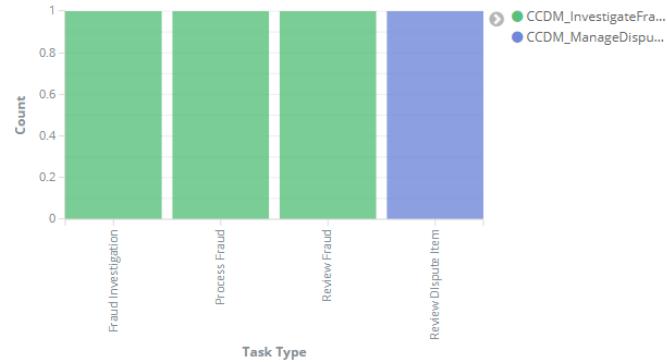
# Case Task Business Automation Workflow Dashboard

kibana

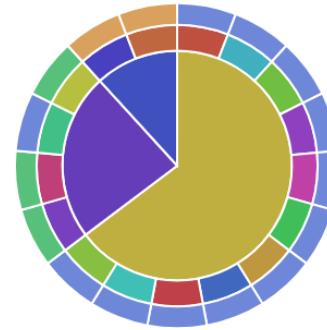
- Discover
- Visualize
- Dashboard
- Timeline
- Dev Tools
- Management

Collapse

Tasks in Progress [do not change]

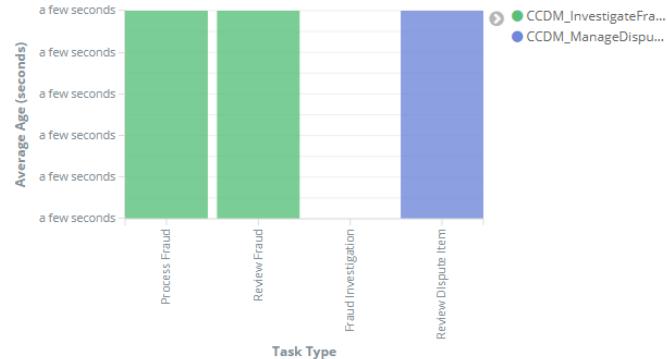


Tasks in Each State [do not change]

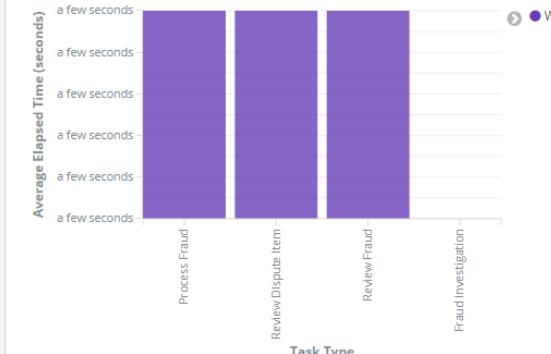


- Waiting
- Working
- Complete
- Add Document
- Arbitration
- Close Case
- Evaluate for Fraud
- Generate Close Le...
- Generate Credit L...
- Generate Fraud Le...
- Generate Provisio...
- Generate Status L...
- Process Chargeback
- Process Represent...
- Fraud Investigation
- Process Fraud
- Review Dispute I...
- Review Fraud

Average Age of Active Tasks [do not change]



Average Elapsed Time of Active Tasks by State [do not change]



# Case Stages Support in Business Automation Insights

New

In this release we added full support for Case Stages in BAI

- Modified Case Event Emitter to process stage objects and produce new ‘stage’ type events.
- Added new Flink job to
  - Process the stage events and store the timeseries in HDFS
  - Create and store the stage summaries in ElasticSearch

<input type="checkbox"/>	Workflow - Stages
<input type="checkbox"/>	In-progress stages [do not change]
<input type="checkbox"/>	Completed stages [do not change]
<input type="checkbox"/>	Stage summaries [do not change]
<input type="checkbox"/>	Incoming and outgoing stages [do not change]
<input type="checkbox"/>	Incoming and outgoing stages [do not change]
<input type="checkbox"/>	Average age of in-progress stages [do not change]
<input type="checkbox"/>	Stages in progress [do not change]
<input type="checkbox"/>	Stages in each state [do not change]
<input type="checkbox"/>	Incoming and outgoing stages in past 3 months [do not change]
<input type="checkbox"/>	Stage specification [do not change]
<input type="checkbox"/>	Total number of stages [do not change]
<input type="checkbox"/>	Average elapsed time of completed stages [do not change]
<input type="checkbox"/>	Average stage duration [do not change]
<input type="checkbox"/>	Stages completed statistics [do not change]
<input type="checkbox"/>	Stages started statistics [do not change]

Kibana artifacts related to Case Stages:

- 5 Searches
- 10 Visualizations
- 1 Dashboard

# Case Stages Dashboard and Visualizations

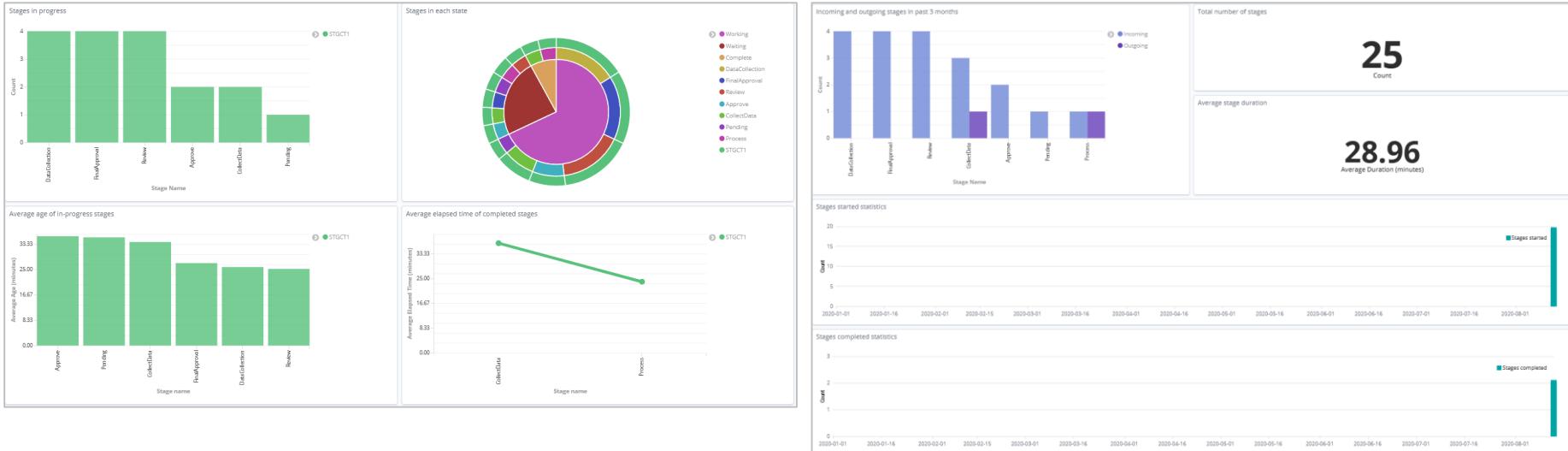
New

STG\_STGCT1\_000000180001 | Modified: 9/23/2019, 3:33 PM | STGCT1

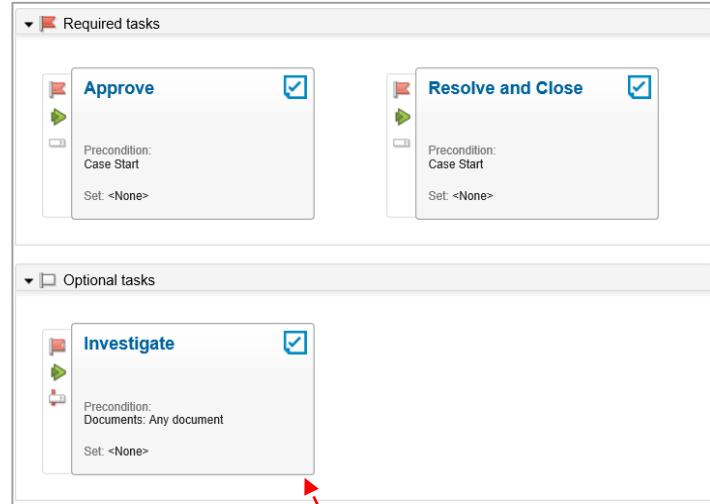
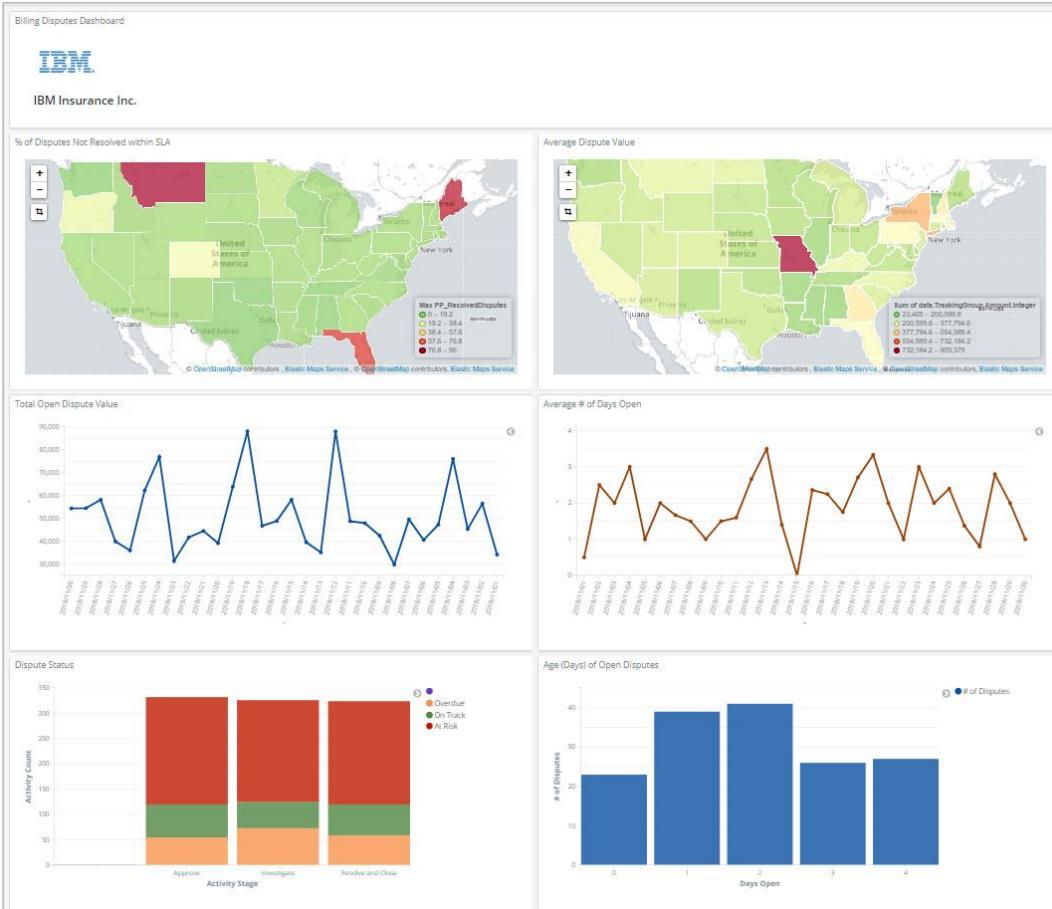
Comments Split Case ▾  
Release Hold Complete Restart

Save Close

Stage1 Stage2 Stage3 Stage4



# Example of a Custom Case Dashboards



Billing Dispute Resolution Case

Billing Disputes Case

# Using Data Lake for Data Science

IBM Watson Studio   Projects   Tools   Community   Services   Manage   Support   Docs   Paul Pacholski's Account

In [ ]:

In [1]: # The code was removed by Watson Studio for sharing.

Out[1]:

	State	Process Stage	Task Age	Task Status	Dispute Amount	Time Stamp	Open Disputes
0	OR	Approve	0	On Track	49433	2018/11/21	2114
1	DE	Resolve and Close	0	On Track	22540	2018/11/21	2641
2	NV	Investigate	4	Overdue	55039	2018/11/24	1292
3	OK	Investigate	4	Overdue	50521	2018/11/05	2962
4	ID	Approve	4	Overdue	37206	2018/11/08	1875

In [29]:  
newdf = df\_data\_1  
newdf = newdf.drop('State', axis=1)

In [33]: df = pd.concat([pd.get\_dummies(df\_data\_1[['State']]), newdf], axis=1)

In [18]: df\_data\_1['NumericState'] = le.transform(df\_data\_1['State'])

In [23]:  
import numpy as np  
np.sqrt(0.8)

Out[23]: 0.894427190999991586

In [34]: df.corr()

Out[34]:

	State_AK	State_AL	State_AR	State_AS	State_AZ	State_CA	State_CO	State_CT	State_DC	State_DE	...	State_VI	State_VL
State_AK	1.000000	-0.016545	-0.017911	-0.017604	-0.017082	-0.016654	-0.018012	-0.016869	-0.017082	-0.019656	...	-0.018112	-0.01771
State_AL	-0.016545	1.000000	-0.016932	-0.016642	-0.016149	-0.015744	-0.017028	-0.015948	-0.016149	-0.018582	...	-0.017123	-0.01671

State	Task Age	Dispute Amount	Open Disputes
State_NY	0.057523	0.568539	-0.002725
State_TX	-0.024151	0.322863	0.004377
State_MO	0.065641	0.446561	-0.008155
State_CA	0.062294	0.155869	0.007074
State_FL	0.071575	0.142554	0.002970

Significant correlation discovered between certain states and key process metrics

# Dashboards

Process Dashboards

Case Dashboards

Decisions Dashboards

Content Dashboards

# ODM - Predefined ODM Views

Select items for your own  
predefined dashboards

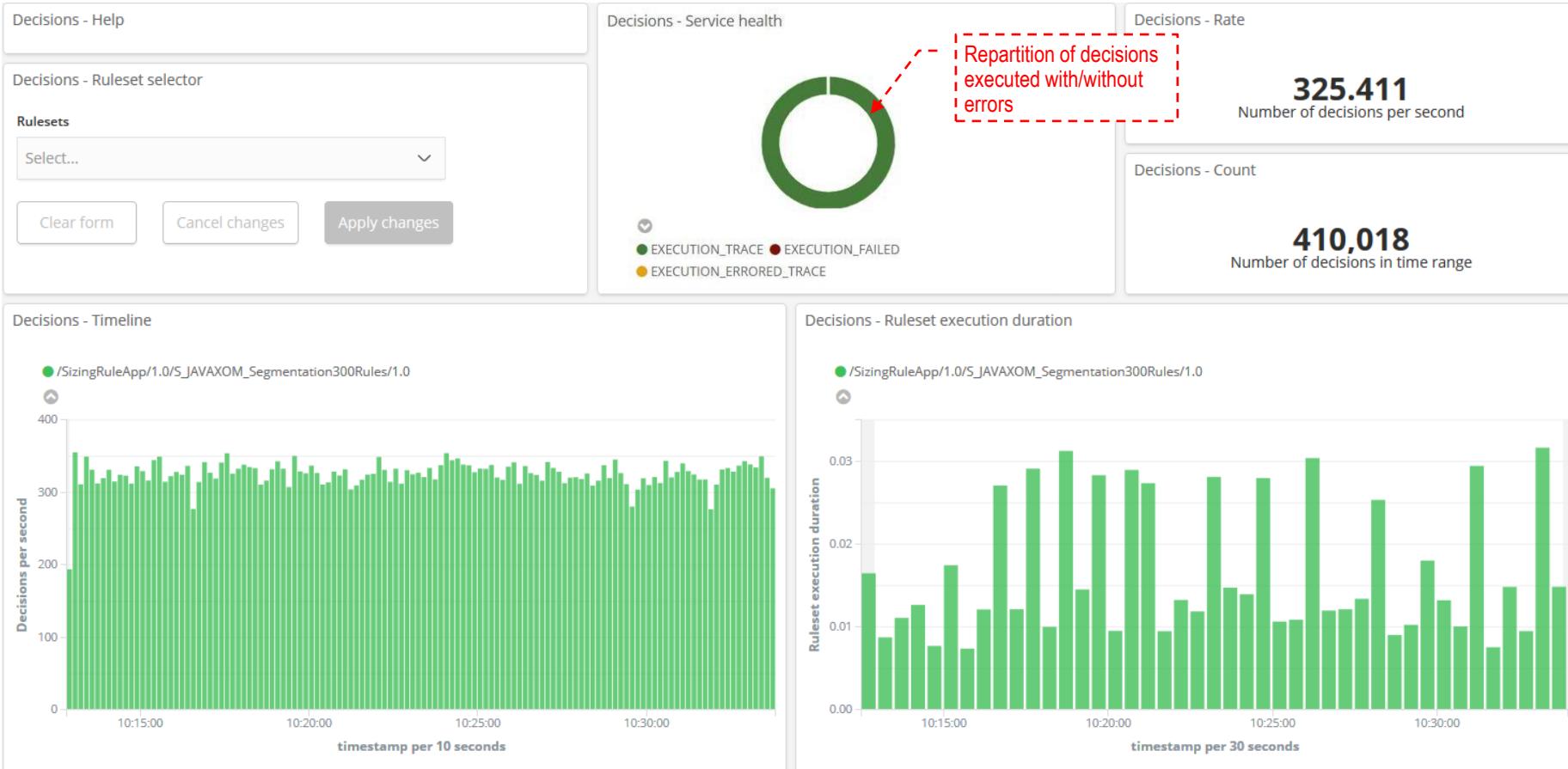
<input type="checkbox"/> Decisions average duration	 Vertical Bar	<input type="checkbox"/> Decisions rules table	 Data Table
<input type="checkbox"/> Decisions average per second	 Metric	<input type="checkbox"/> Decisions Ruleset in error	 Data Table
<input type="checkbox"/> Decisions count	 Metric	<input type="checkbox"/>  Decisions Ruleset selector	 Controls
<input type="checkbox"/> Decisions distribution	 Pie	<input type="checkbox"/> Decisions tasks table	 Data Table
<input type="checkbox"/> Decisions duration	 Vertical Bar	<input type="checkbox"/> Decisions timeline	 Vertical Bar
<input type="checkbox"/> Decisions errors cloud	 Tag Cloud	<input type="checkbox"/>  Decisions trace selector	 Controls
<input type="checkbox"/> Decisions global distributions	 Timelion		

# ODM - Detailed decision data

1-50 of 5

Time	rulesetPath	duration	type
▼ February 5th 2019, 13:37:55.481	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0	1.00	EXECUTION_TRACE
<a href="#">View surrounding documents</a>		<a href="#">View similar</a>	
<a href="#">Table</a>		<a href="#">JSON</a>	
Detailed decision data (input, output, executed rules) – Used to create visualizations			
t _id		Q Q □ * 05b92ab8-cb7c-41fb-bfa4-225dd9c512fa0	
t _index		Q Q □ * odm-timeseries-write	
# _score		Q Q □ * -	
t _type		Q Q □ * odm-timeseries	
# data.test_deployment.loan_validation_with_score_and_grade.in.borrower.birth		Q Q □ * -101,779,200,000	
# data.test_deployment.loan_validation_with_score_and_grade.in.borrower.birthDate		Q Q □ * -101,779,200,000	
# data.test_deployment.loan_validation_with_score_and_grade.in.borrower.creditScore		Q Q □ * 600	
t data.test_deployment.loan_validation_with_score_and_grade.in.borrower.firstName		Q Q □ * Paul	
t data.test_deployment.loan_validation_with_score_and_grade.in.borrower.lastName		Q Q □ * Wilson	
? data.test_deployment.loan_validation_with_score_and_grade.in.borrower.latestBankruptcy		Q Q □ * ▲ -	
# data.test_deployment.loan_validation_with_score_and_grade.in.borrower.latestBankruptcyChapter		Q Q □ * -1	
# data.test_deployment.loan_validation_with_score_and_grade.in.borrower.latestBankruptcyDate		Q Q □ * -	
t data.test_deployment.loan_validation_with_score_and_grade.in.borrower.latestBankruptcyReason		Q Q □ * -	
? data.test_deployment.loan_validation_with_score_and_grade.in.borrower.spouse		Q Q □ * ▲ -	
t data.test_deployment.loan_validation_with_score_and_grade.in.borrower.ssn.areaNumber		Q Q □ * 89	
# data.test_deployment.loan_validation_with_score_and_grade.in.borrower.ssn.digits		Q Q □ * 7	
t data.test_deployment.loan_validation_with_score_and_grade.in.borrower.ssn.fullNumber		Q Q □ * 89-45-896	
t data.test_deployment.loan_validation_with_score_and_grade.in.borrower.ssn.groupCode		Q Q □ * 45	
t data.test_deployment.loan_validation_with_score_and_grade.in.borrower.ssn.serialNumber		Q Q □ * 896	
t data.test_deployment.loan_validation_with_score_and_grade.in.borrower.ssncode		Q Q □ * 89-45-896	
# data.test_deployment.loan_validation_with_score_and_grade.in.borrower.yearlyIncome		Q Q □ * 554,000	

# ODM - Decisions performance monitoring



# ODM - Detailed decision artifact executions

Decisions - Tasks table

Tasks	Rulesets
loanvalidation	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0
loanvalidation#initResult	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0
loanvalidation#validation	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0
loanvalidation#computation	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0
loanvalidation#eligibility	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0
loanvalidation#insurance	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0

Executed tasks sequence

Decisions - Rules table

Rules	Rulesets
validation.borrower.checkAge	/test_deployment/1.0/loan_validation_with_score_a
validation.borrower.checkName	/test_deployment/1.0/loan_validation_with_score_a
validation.borrower.checkSSNareanumber	/test_deployment/1.0/loan_validation_with_score_a
validation.borrower.checkSSNdigits	/test_deployment/1.0/loan_validation_with_score_a
validation.borrower.checkZipcode	/test_deployment/1.0/loan_validation_with_score_a
computation.initialCorporateScore	/test_deployment/1.0/loan_validation_with_score_a
computation.neverBankruptcy	/test deployment/1.0/loan validation with score a

Executed rules sequence

Decisions - Rulesets table

Rulesets
/test_deployment/1.0/loan_validation_with_score_and_grade/1.0

Ruleset list

Executions      Average Execution duration (ms)

1,000      0.33

# ODM - Decision Health monitoring

The screenshot displays the ODM Decision Health monitoring interface, which includes four main sections:

- Decisions - Status selector:** A form with a dropdown menu for "Status" (with options "Select...", "All", "Error", "Warning", "Info", "Trace", "None"), and buttons for "Clear form", "Cancel changes", and "Apply changes".
- Decisions - Errors cloud:** A section showing "Execution error messages if any" (indicated by a red dashed box and arrow).
- Decisions - Rulesets in error:** A section showing "No results found" (indicated by a red dashed box and arrow).
- Decisions - Search:** A table showing search results for decisions. The columns are "Time", "rulesetPath", "value", and "type". The data is as follows:

Time	rulesetPath	value	type
February 5th 2019, 13:37:55.481	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0	1.00	EXECUTION_TRACE
February 5th 2019, 13:37:55.481	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0	3.00	EXECUTION_TRACE
February 5th 2019, 13:37:55.465	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0	3.00	EXECUTION_TRACE
February 5th 2019, 13:37:55.278	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0	1.00	EXECUTION_TRACE
February 5th 2019, 13:37:55.262	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0	2.00	EXECUTION_TRACE
February 5th 2019, 13:37:55.262	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0	1.00	EXECUTION_TRACE
February 5th 2019, 13:37:55.242	/test_deployment/1.0/loan_validation_with_score_and_grade/1.0	5.00	EXECUTION_TRACE

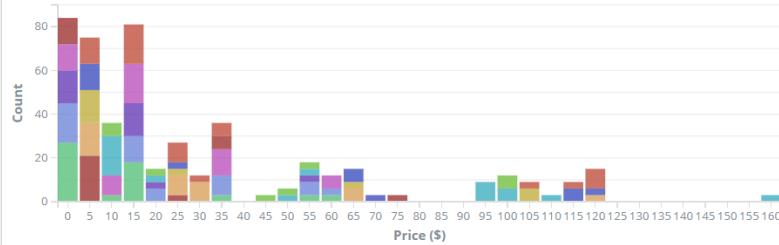
# ODM - Example of Custom Business Dashboard

Sample Decisions Shipment - Help

Sample Decisions Shipment - Count

**600**  
Shipment pricing requests      **29.225**  
Average price (\$)

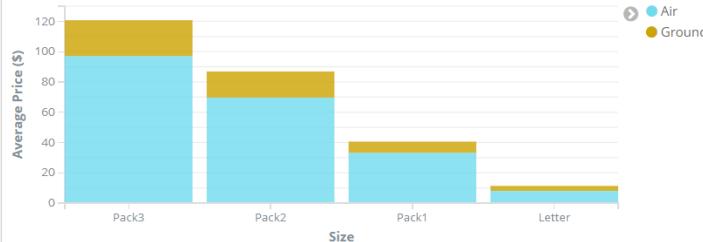
Sample Decisions Shipment - Price per distance



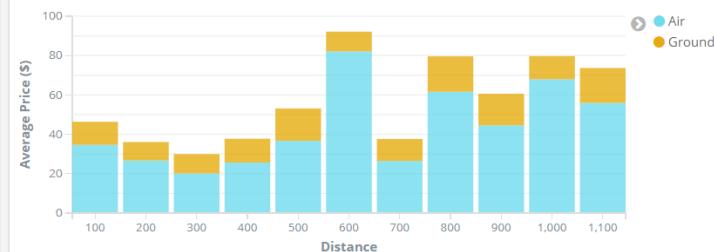
Sample Decisions Shipment - Distribution



Sample Decisions Shipment - Average Price per size



Sample Decisions Shipment - Average price per distance



# Dashboards

Process Dashboards

Case Dashboards

Decisions Dashboards

Content Dashboards

# FileNet Content Manager Dashboards

**kibana**

Content - Content Specification  
**Content dashboard**  
The Content dashboard page displays chart widgets that provide content-related statistics. The Content dashboard page contains the following widgets:

Content - Object store selector  
Object store selector  
OS1 X  
Clear form Cancel changes Apply changes

Content - Creators  
KBarlow JGarcia GHearst Dr.Smith

Content - Number of documents for each document class  

Document Class	Percentage
Resume	39.92%
BackgroundCheck	35.93%
SupportingDocument	20.16%
OfferLetter	3.99%

Content - Total number of documents  
**2,505**  
Total number of documents

Content - Document classes  
SupportingDocument  
**Resume**  
OfferLetter  
BackgroundCheck

Count - Creators  
Content - Number of documents group by creator  

Creator	Number of documents
JGarcia	~1300
Dr.Smith	~900

Count - Document classes  
Content - Number of documents group by lastModifier  

Last Modifier	Percentage
JGarcia	~75%
GHearst	~20%
KBarlow	~5%
Dr.Smith	~0%

Gain operational insights regarding the nature and volume of content being ingested

Monitor technical and business metadata attached to documents

# Machine Learning Examples

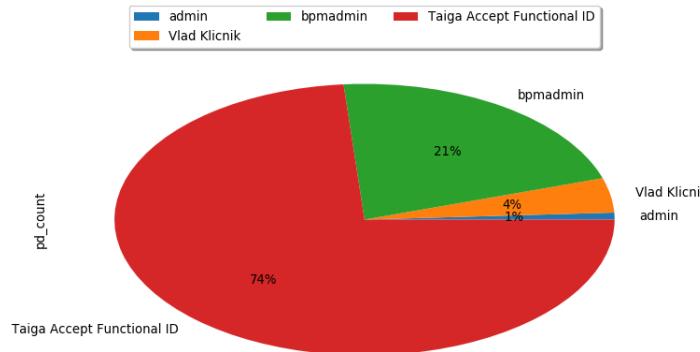


- Examples in the form of Jupyter Notebooks in IBM Watson Studio that show how to exploit business data that BAI stores in Hadoop file system (HDFS).
- Data charts
  - Uses Apache Spark to query data and create charts from Business Automation Insights data that is stored in an HDFS data lake
  - [https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/4575ea5c-bcce-4968-91ec-c0ebb7267b68/view?access\\_token=f57fba08d0084803dc050304c4c1ee98034cfbe37d6918ed738db71895c39340](https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/4575ea5c-bcce-4968-91ec-c0ebb7267b68/view?access_token=f57fba08d0084803dc050304c4c1ee98034cfbe37d6918ed738db71895c39340)
- Recommendation service
  - Uses BAI data to create a recommendation service for an IBM Business Automation Workflow application.
  - [https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/1e7bec1b-a07f-4d68-a76e-84bab4a97ee6/view?access\\_token=9c794a25947504d2dd50b056c98f8fd812deabafd6d860427aa869da2efaa309](https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/1e7bec1b-a07f-4d68-a76e-84bab4a97ee6/view?access_token=9c794a25947504d2dd50b056c98f8fd812deabafd6d860427aa869da2efaa309)
- Estimation of process duration:
  - Uses BAI data to estimate the duration of processes, based on their start time and business data
  - [https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/0b395fce-04b0-4729-86e3-e5b1b605ef8a/view?access\\_token=91d4093f33e4a70831269cb87870ad4b018e9cd78c46b7cf2f2d8b64abd02b3c](https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/0b395fce-04b0-4729-86e3-e5b1b605ef8a/view?access_token=91d4093f33e4a70831269cb87870ad4b018e9cd78c46b7cf2f2d8b64abd02b3c)

# Data Chart Sample

Query Business Automation Insights data and use charts to visualize

- This notebook shows how to retrieve and use data stored by Business Automation Insights for Business Automation Workflow.
- The notebook uses Spark SQL to retrieve process and task data and uses also pixiedust to display charts.
- The notebook is configured to retrieve data from the HDFS data lake where Business Automation Insights stores that data.

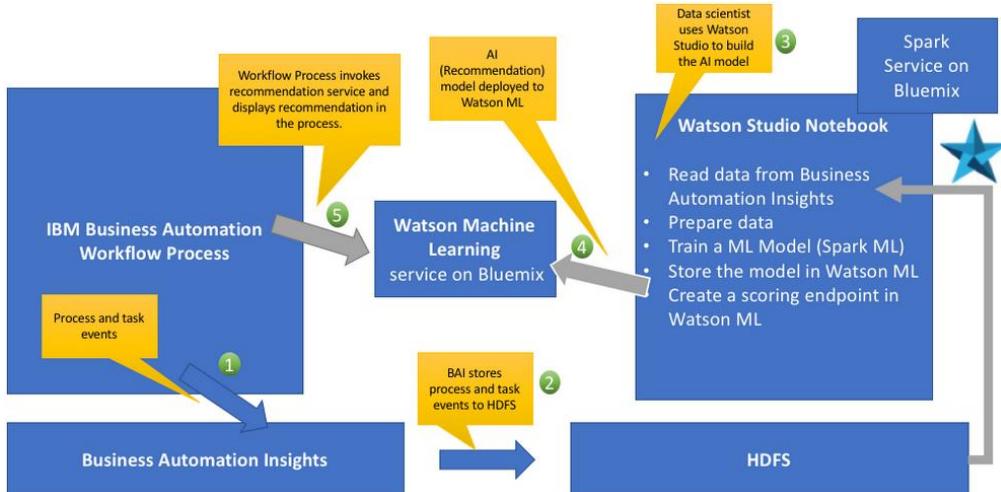


```
spark.sql("select performerName, count(*) from activitysummaries group by performerName") .show()
```

performerName	count(1)
Taiga Accept Func...	3581
null	1351
bpmadmin	1041
Vlad Klicnik	201
admin	4

# Recommendation Service

- Data from a human task “classification decision” is captured by BAI and stored in
- Machine-learning algorithm can assist in task completion by recommending the best decision based on past decisions.



To display a recommendation for a decision on a claim within the Workflow process itself we invoke the Recommendation service REST API

The screenshot shows a user interface for "Claim Approval". The form includes fields for Customer Name (John Smith), Credit Score (399), Vehicle, and Claim. Below these, there are fields for Approved Amount (854) and Estimate Amount (854). In the "Recommendation" section, there is a green smiley face icon and the text: "Based on the previous decision we recommend to APPROVE the claim with a confidence of 95". At the bottom, there is a checkbox labeled "I Approve the claim" and a "OK" button.

# ODM - Example of Watson Studio Notebooks

## 2. View the rule coverage distribution

You visualize the rule execution statistics metered when running a decision service. From these statistics you get insights on:

- the rules that are never executed. Rules can be never executed against a data set when their conditions are never met. Typically if they are triggered for specific cases that don't appear in your data set. Otherwise have a deeper look to your rules and dataset.
- the rules that are always executed. Common case is a rule that checks the compliance of input parameters. But if you see a rule that should be triggered on a subset of the requests and is reported with a 100% then check its conditions and the request set.

```
In [5]: from matplotlib import rcParams
import matplotlib.pyplot as plt; plt.rcParams()
import numpy as np
import matplotlib.pyplot as plt

rcParams.update({'figure.autolayout': True})
rcParams.update({'font.size': 10})

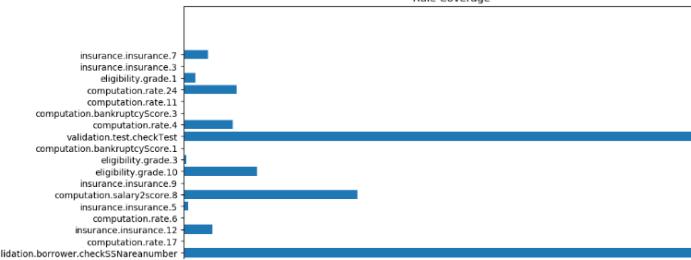
plt.figure(figsize=(12, 16))
#plt.yscale('log', nonposy='clip')

rules = df['Rule']
percentages = df['Percentage']
counts = df['Count']
y_pos = np.arange(len(rules))

plt.barh(y_pos, percentages, align='center')
plt.yticks(y_pos, rules)
plt.xlabel('percentage')
plt.title('Rule Coverage')
plt.ylabel('rule')
plt.savefig('rulecoverage.jpg')
plt.show()

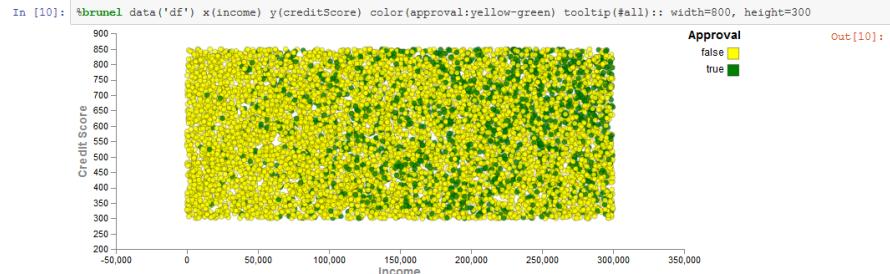
/usr/local/src/conda3_runtim/home/envs/DSX-Python35-Spark/lib/python3.5/site-packages/matplotlib/figure.py:1999:
rWarning: This figure includes Axes that are not compatible with tight_layout, so results might be incorrect.
warnings.warn("This figure includes Axes that are not compatible")
```

Rule Coverage



## 4. View income on credit score distribution.

Do we see trends or limits in credit score or income for accepted loan applications? We can observe graphically that the larger are the credit score and income values the more accepted approval we get.

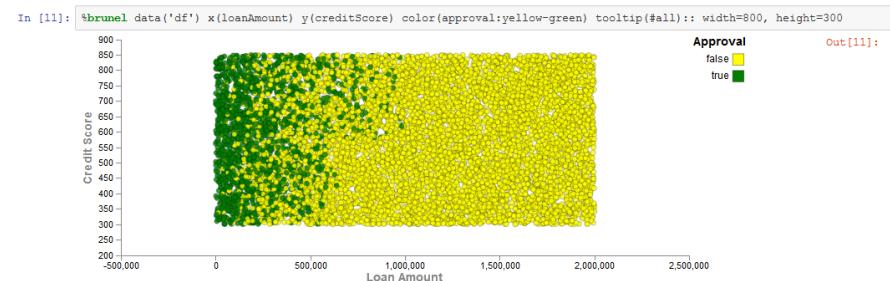


## 5. View loan amount / credit score distribution-

Do we see limits in score or amount for accepted loan applications? We observe that as expected:

- the higher the loan amount, the higher the rejection rate.
- the lower credit score, the higher the rejection rate.

We observe the absence of green points identified for loan amount greater than USD 1 000 000. It is consistent with a rule that rejects the application for amounts greater than this threshold.



# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# Install and Configure

Install BAI

Install Event Emitters

BPMN

Case

ODM

Content

# Install Options

- On Kubernetes
  - Evaluation / Demo Option
    - [https://www.ibm.com/support/knowledgecenter/en/SSYHZ8\\_20.0.x/com.ibm.dba.install/op\\_topics/tsk\\_install\\_demo.html](https://www.ibm.com/support/knowledgecenter/en/SSYHZ8_20.0.x/com.ibm.dba.install/op_topics/tsk_install_demo.html)
  - Enterprise Option
    - [https://www.ibm.com/support/knowledgecenter/en/SSYHZ8\\_20.0.x/com.ibm.dba.install/op\\_topics/tsk\\_operators.html](https://www.ibm.com/support/knowledgecenter/en/SSYHZ8_20.0.x/com.ibm.dba.install/op_topics/tsk_operators.html)
- Without Kubernetes (**BAI for a Server**)
  - [https://www.ibm.com/support/knowledgecenter/en/SSYHZ8\\_20.0.x/com.ibm.dba.install/bai\\_sn\\_topics/tsk\\_bai\\_single\\_node\\_deploy.html](https://www.ibm.com/support/knowledgecenter/en/SSYHZ8_20.0.x/com.ibm.dba.install/bai_sn_topics/tsk_bai_single_node_deploy.html)
    - **Systems - Middleware (1 Image)**  
 Select All (or use check boxes below to select image(s) to download)

---

IBM Cloud Pak for Automation v20.0.2 - Business Automation Insights for a Server Multiplatform English (CC6VTEN) - [View details](#)

**Size** 1,909MB

**Date posted** 26 Jun 2020

[License agreement](#)

[Download estimate](#)

→ [eAssembly](#)

---

# New in 19.0.3 Operator based install (CP4A)

Prepare to install containers on Kubernetes

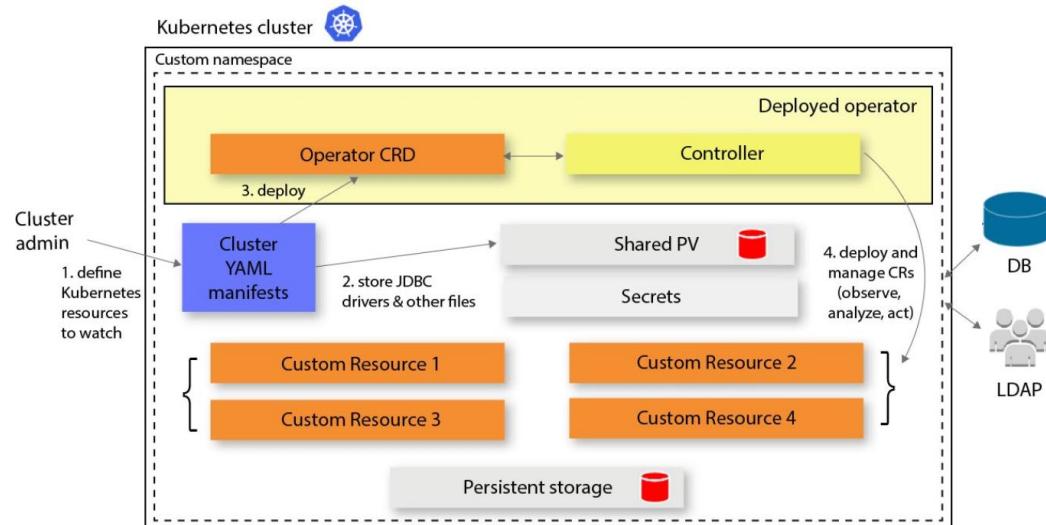
- Prepare your environment and install the necessary software before you go to the GitHub repositories to find further information on installing the IBM certified software.

Install an operator on Kubernetes

- Use the operator YAML manifest files in the GitHub repository to deploy an operator and use a custom resource to apply deployments of the automation containers on Kubernetes.

Completing post-deployment tasks

- Additional steps are required to ensure the functionality of the environment.



[https://www.ibm.com/support/knowledgecenter/en/SSYHZ8\\_19.0.x/com.ibm.dba.install/op\\_topics/tsk\\_operators.html](https://www.ibm.com/support/knowledgecenter/en/SSYHZ8_19.0.x/com.ibm.dba.install/op_topics/tsk_operators.html)

# Steps before installing BAI (CP4A)

- Install Kafka (two options)
  1. Use your own or
  2. Use IBM Event Streams (DBA for Multiload non-chargeable supporting feature)
- Create Kafka Ingress/Egress/Service topics
- Install Elasticsearch and Kibana (two options)
  1. Use your own (required for production) or
  2. Use the one that comes with BAI install (**Recommended**)
- Prepare the Persistent Volume and Persistent Volume Claim in ICP Cluster
  - PVs Optional for Elasticsearch (but required in production)
  - PVs are required by Flink



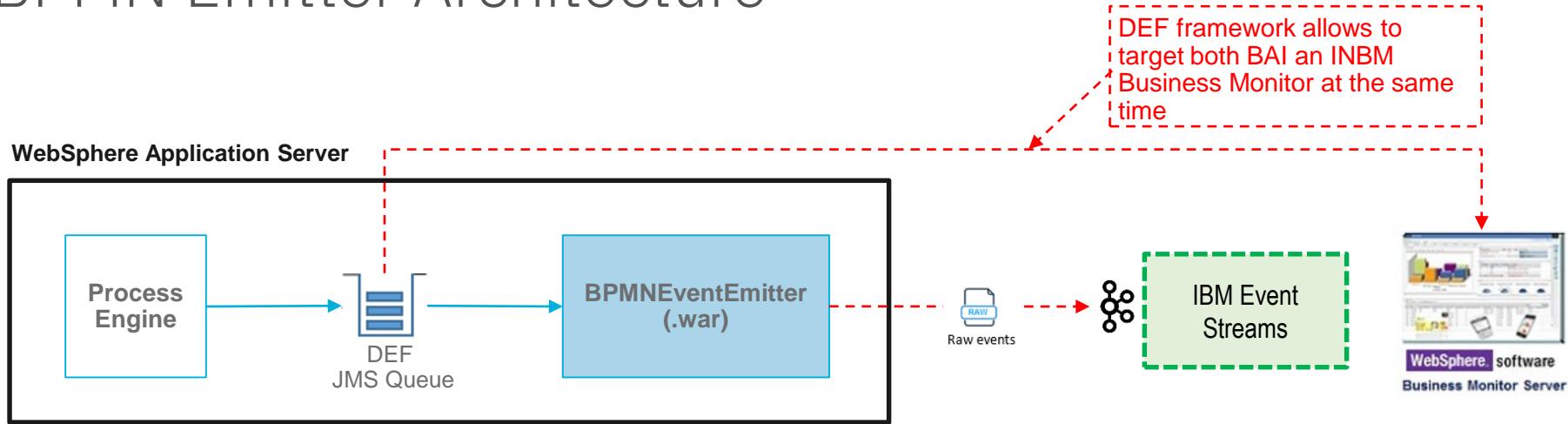
# Install and Configure

## Install BAI

### Install Event Emitters

- BPMN
- Case
- ODM
- Content

# BPMN Emitter Architecture



- The BPMN event emitter is a Java MDB running on BAW App Server
- It gets events from the DEF JMS queue
- It formats DEF events as raw events in common JSON format
- It posts the raw events to Kafka Ingress Topic securely
- Performance Impact on the Workflow Server: 5-10%
  - See detailed study in Performance section of this deck

# Web Process Designer Considerations

- Once the BAI Emitter is installed you do not have to enable anything in Web PD
  - Do not need to enable monitor or update tracking definitions



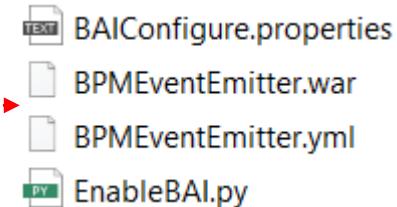
- We recommend that you turn off PDW
  - See: <https://developer.ibm.com/answers/questions/167196/disabling-tracking-data-generation-for-a-process-s.html>
- To emit business data enable tracking of process variables

A screenshot of the 'Tracking' tab in the 'Standard HR Open New Position' process. The tab is highlighted in blue. Below the tab, there are two sections: 'Tracking Groups' and 'Auto-Tracked Fields'. The 'Auto-Tracked Fields' section contains five entries, each with a field name and a corresponding value in a text input field:

Field Name	Value
requisition.requester	HiringManager
requisition.empType	EmploymentStatus
requisition.department	Department
requisition.location	Location
requisition.gmApproval	GMAccrual

# BPMN Emitter Installation

- Prior to 18.0.0.2 the following APAR must be applied on BAW  [8.6.10018001-WS-BPM-IFJR59785.zip](#)
- For information about all perquisites see tech note:  
<https://www.ibm.com/support/docview.wss?uid=swg20734037>
- Configuration:
  - Edit BAIConfigure.properties – to setup event encryption (i.e. Kerberos + SSL)
  - Edit BPMEventEmitter.yml – to set up Kafka endpoints
- Run EnableBAI.py
  - Unlike Case, you do not need to specify what Process to enable (this will enable all Processes)



# Controlling What Processes Emit Events

- You can configure the Dynamic Event Framework (DEF) and subscribe to events that you want to receive.
- Configure and run
  - wsadmin –lang jython –f c:\SampleReloadDEF.py

To listen for every event for all applications, use the wildcard character as shown in the following example:

```
subscriptions=[  
    '*/**/*/*/*/*'  
]
```

The following example shows how you might register to receive events for the **Hiring Sample**:

```
subscriptions=[  
    'HSS/*/*/*/*/*/*/*',  
    'HSS/*/*/*/*/*/*/*',  
    'HSS/*/*/*/*/*/*/*',  
    'HSS/*/*/*/*/*/*/*'  
]
```

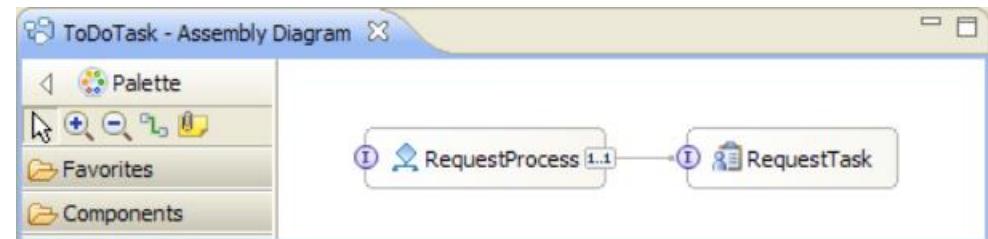
Hiring Sample  
Process App  
acronym

- Details
  - [https://www.ibm.com/support/knowledgecenter/SS8JB4\\_18.0.0/com.ibm.wbpm.admin.doc/topics/capturingevents.html](https://www.ibm.com/support/knowledgecenter/SS8JB4_18.0.0/com.ibm.wbpm.admin.doc/topics/capturingevents.html)

# BPEL Emitter

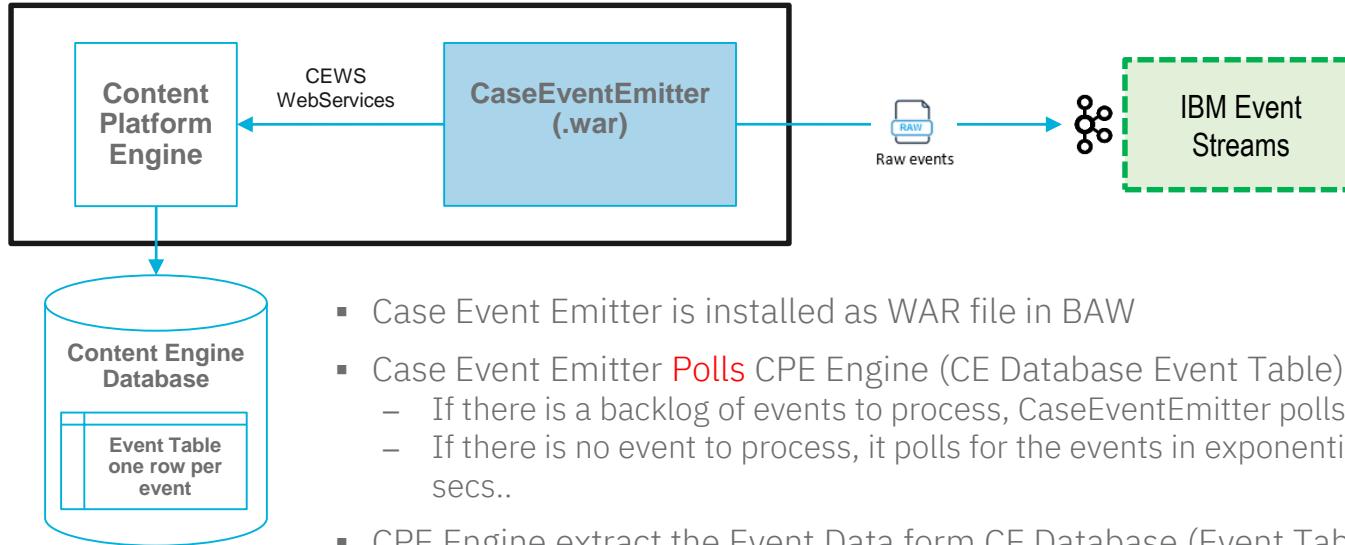
NEW

- Same EnableBAI.py script that is used to configure BPMN event emission:
  - Shared DEF listener
  - Shared event emitter
- Additional setStateObserver.py script call to enable the BPEL events
- Events are generated from both BPEL processes and Human Tasks (both are stateful)



# Case Emitter Architecture

## WebSphere Application Server



- Case Event Emitter is installed as WAR file in BAW
- Case Event Emitter **Polls** CPE Engine (CE Database Event Table)
  - If there is a backlog of events to process, CaseEventEmitter polls the events continuously..
  - If there is no event to process, it polls for the events in exponential intervals of 2, 4, 8, 32, 64 secs..
- CPE Engine extract the Event Data form CE Database (Event Table)
- CPE Engine's "Audit Processing Bookmark" object is used to ensure events are not emitted multiple times
- Case Event Emitter formats Raw Events as JSON and posts them to Kafka Ingress Topic securely
- Performance Impact on the Workflow Server: no data

# Configure Case to Emit Data

## Create Audit Configuration

The screenshot shows the 'Case Solutions' screen. A red arrow points from the 'New' button on the left to the 'Actions' dropdown menu. Another red arrow points from the 'Mortgage Ap' solution name in the list to the 'Audit Configuration' option in the dropdown menu. A third red arrow points from the 'Create an audit configuration' radio button to the right.

**Case Solutions**

New Import Search

Solution Name: Mortgage Ap (MORT)

Actions

- Import
- Deploy
- Export
- Delete
- Manage
- Audit Configuration

Solution name: Mortgage Application

Solution prefix: MORTG

Description: Solution for new mortgage applications.

\* Options:

- Create an audit configuration
- Edit an audit configuration

## Select Properties to Emit



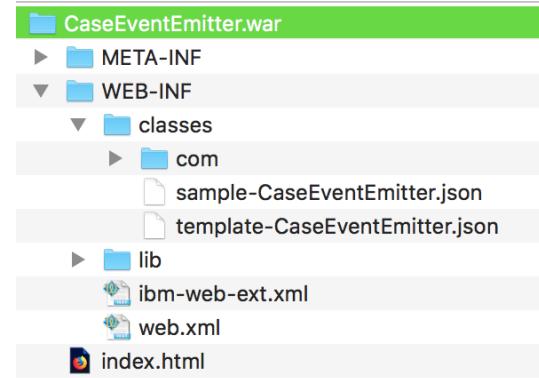
# Case Emitter Installation

Configuration script along with Case Eventer Emitter is shipped with BAW 18.0.0.2

- **CaseEventEmitter.war** – Case Event Emitter Application with out any configuration in it.
- **EnableCaseBAI.py** – Configuration script to work with Case Event Emitter application.
- **sample-CaseEventEmitter.json** – Sample configuration file for CaseEventEmitter.json.
- **template-CaseEventEmitter.json** – Explains each and every field in the json and how it is used.

# Case - Configuration script EnableCaseBAI.py

- Uses WebSphere Jython APIs to communicate with WebSphere Application Server
- Installs / upgrades / uninstalls Case Event Emitter application
  - Takes Case Event Emitter war and Json configuration file as inputs
  - Creates a new Case Event Emitter war application in a new directory and pushes the Json configuration file into the new emitter war.
  - Optionally configures the Kerberos setting in WebSphere Application Server based on the configuration.



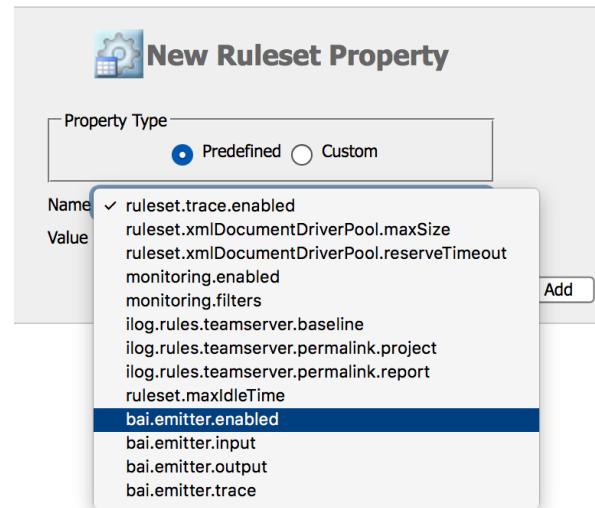
# ODM Emitters - Install and Configure

## 1. Configure Emitters

- ICP
  - Use IBACC or the ODM helm chart
- On Prem using one the following:
  - ra.xml deployment descriptor
  - plugin-configuration.properties file
  - res-setup Ant task
- Configuration instructions:
  - [https://www.ibm.com/support/knowledgecenter/en/SSQP76\\_8.10.0/com.ibm.odm.distrib.overview/topics/con\\_bai.html](https://www.ibm.com/support/knowledgecenter/en/SSQP76_8.10.0/com.ibm.odm.distrib.overview/topics/con_bai.html)

## 2. Activate for selected ruleset

- Configuration instructions
  - <https://github.com/ODMDev/decisions-bai-gettingstarted>



# ODM BAI Emitter: Requirements, Limitations, Perf.

- ODM emitter available through:
  - Either DBAMC 18.0.2,
  - or ODM 8.10.1 running on-prems with WAS Liberty (no support yet for other app servers)
  - Not available for: ODM On Cloud and ODM for z/OS
- ODM emitter compatible with :
  - Rule Execution Server only (Not Decision Server Insights or Decision Runner)
  - Decisions based on:
    - Java Object Model only (not XML binding)
    - Decision engine only (not Classic Rule Engine)
    - Regular Decision Service or Decision Model Service
- Performance
  - Impact on CPU usage (create the event, optionally encrypt it, generate decision trace)
  - Impact on Decisions per Seconds (each ruleset in the pool uses a few more ms per decision)
  - From early performance tests, the impact depends on:
    - Response time of the decision service (small response time -> high relative impact of event emission)
    - Quantity of data sent in the events (trace / input / output / number of fields in input and output)

# Content BAI Emitter

- Install and configure steps:
  - From boot master node of BAI ICP installation, retrieve the bai-content-emitter.jar file
  - In Admin Console for CPE configure BAI Emitter Event Action for each Object Store
    - You will need the bai-content-emitter.jar file
  - Configure Kafka security configuration in cpe-cfgstore/BAIConfiguration directory and create configuration file
    - Ex: Instead of using the default configuration properties to provide the keys and password, you can customize the configuration file by providing only the plain username and password combination
- Content emitter configuration info:

[https://www.ibm.com/support/knowledgecenter/SSYHZ8\\_19.0.x/com.ibm.dba.bai/topics/con\\_bai\\_cntevmtr\\_config.html](https://www.ibm.com/support/knowledgecenter/SSYHZ8_19.0.x/com.ibm.dba.bai/topics/con_bai_cntevmtr_config.html)

Product Version	Cloud Pak	On Prem
IBM FileNet Content Manager V5.5.3	✓	✗

# Historical Data BPMN Export to BAI

- BPM and BAW customers may have a vast amount of data already stored in the BPM instance tables as well as Performance Data Warehouse
- As they adopt BAI, they now have a mechanism available to transfer this historical data into BAI.
- What does Historical Data BPMN Playback for BAI do?
  - Converts BPM instance and PDW databases to build corresponding raw BPMN events
  - It emits them to a BAI system using the installed BPMN emitter.
  - This enables existing customers to leverage BAI with their legacy data rather than having to build up the data from scratch.
- How does it work? **POST** `/std/bpm/historical_data_playback`
  - A new ops REST API was introduced to move historical data from BPM/PDW to BAI
    - Asynchronous ops API executable only by an administrator
    - Allows optional filtering by date range and/or container acronym(s)
    - Updates the PDW DB to track emitted events and prevent duplicates

# New REST API

POST

/std/bpm/historical\_data\_playback Play back historical event data and send it to IBM Business Automation Insights.

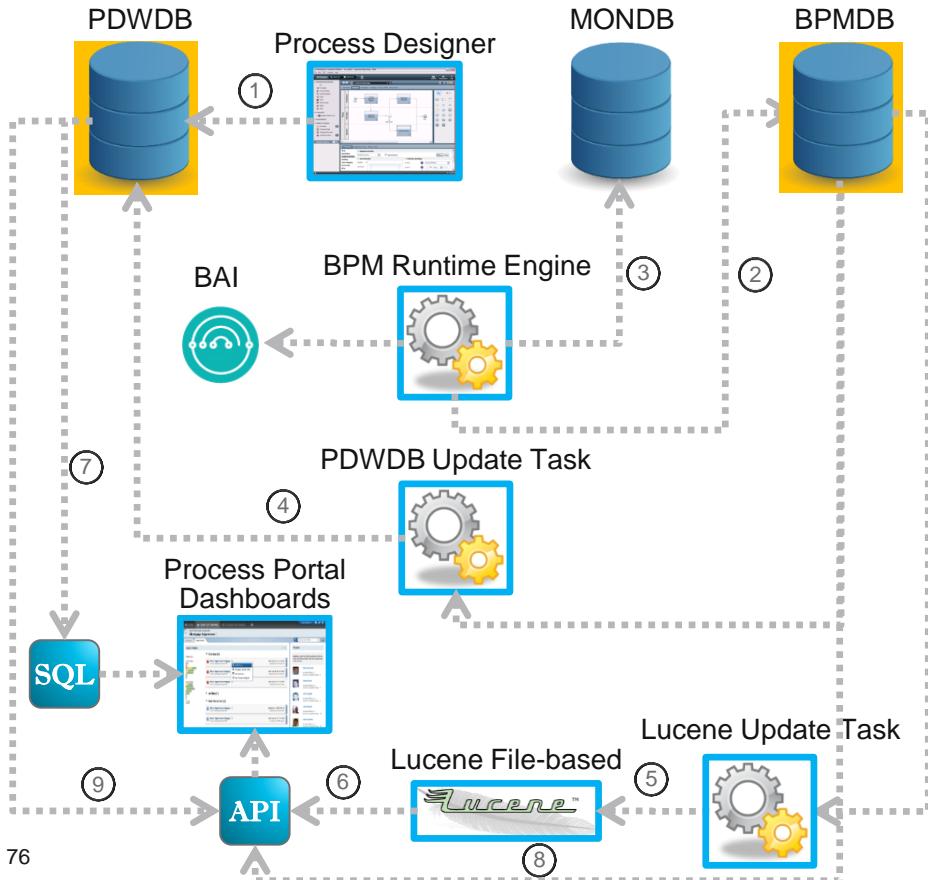


Plays back process-related events stored in the PDW and BPM databases and sends them to Business Automation Insights.  
Only Business Automation Workflow administrators are authorized to perform this call.

## Parameters

Name	Description
BPMCSRFToken <small>* required</small> <small>string (header)</small>	Cross-site request forgery prevention token for IBM Business Automation Workflow REST APIs
containers <small>array[string] (query)</small>	A comma-separated list of container acronyms. Restricts the playback of events to instances that belong to the specified process application acronyms.
start_date <small>string(\$date-time) (query)</small>	The date from which to start processing events. If not specified, process instance data is played back up to the end_date, if specified.
end_date <small>string(\$date-time) (query)</small>	The date to stop processing events. If not specified, process instance data is played back up to the current date and time.
skip_duplicates_check <small>boolean (query)</small>	Use this parameter in conjunction with tracing to troubleshoot event playback. By default, events are generated and played back only once from Performance Data Warehouse; subsequent calls to the API will not send the same events again. However, you can replay these events without sending them to Business Automation Insights by setting this parameter to true.
event_count_only <small>boolean (query)</small>	If the value is set to true, return only an estimated count of the number of events that will be sent to IBM Business Automation Insights. The estimate is based on the values of the other parameters. If a value is not specified, the default is false.

# Understanding Process Data for Dashboards



## Source of data for BAI export.

Clearly, the export will not work, if data is purged from PDWDB and completed instances are deleted in BPMDB .

1. "Update Tracking Definitions" action creates Tracking Group tables in PDWDB. It is possible to associate tracking group with task instance,
2. 'Business Data' is stored in BPMDB with the instance. Tracked variable data is stored in BPMDB temporarily, it's final destination is PDWDB.
3. DEF events sent to MONDB. Every 30 sec (this is configurable) tracked data is moved from temporary BPMDB tables to PDWDB tables.
4. Every 30 sec (this is configurable) tracked data is moved from temporary BPMDB tables to PDWDB tables.
5. Every 5 seconds (this is configurable) the Lucene index gets refreshed. It contains information about running and completed tasks/instances, as well as business data. Note: OS will likely cache the most recently utilized files in OS memory, which makes the file access generally efficient.
6. New JS API designed to support dashboards retrieve running instance data from Lucene index in order to support full text searching and filtering.
7. User authored SQL queries retrieve data from Tracking Group tables. No instance information. Simply tracked data associated with BPDs. Must have tracking enabled. Data is used for critical path analysis.
8. Traditional JS APIs go against live data in BPMDB for runtime object access. This is the 'system of record' for BPM data.

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

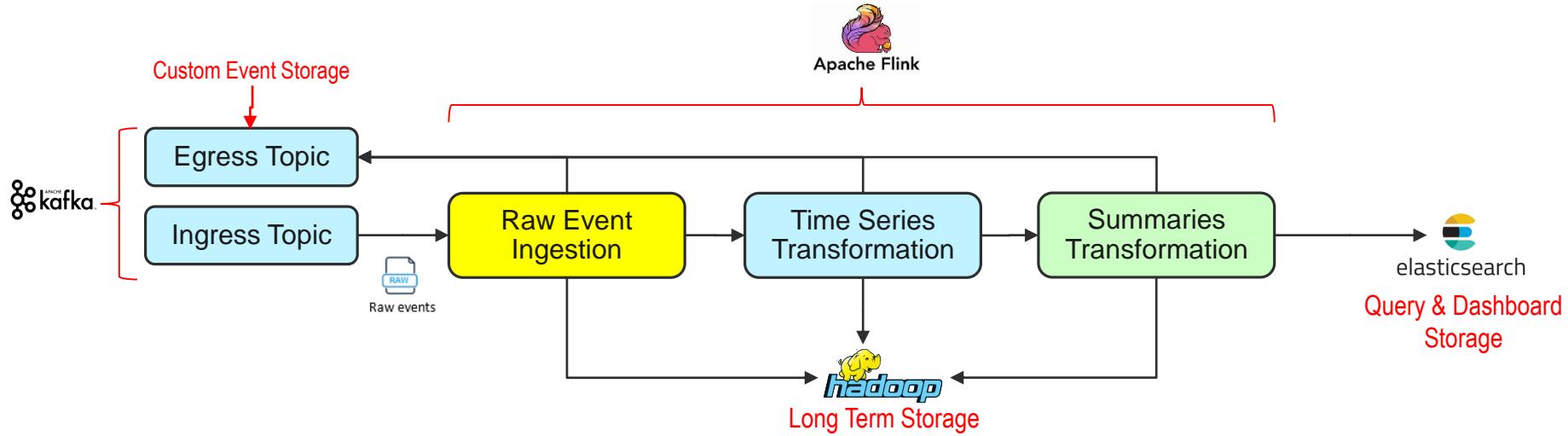
BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# BAI Event Processing



- Raw Events in common JSON format arrive from Processes and Cases
- Apache Flink jobs are responsible for event processing and staging
  - **Raw Event** Ingestion and distribution to: Egress Topic, HDFS
  - **Time Services** Transformation and distribution to: Egress Topic HDFS
  - **Summaries** Transformation and distribution to: Egress Topic, HDFS and Elasticsearch
- Flink includes persistent storage for point recovery and to store data between the three processing stages

# Raw Events

Raw Event  
Ingestion

- Raw Events are unprocessed JSONs as they come from BAI Event Emitters
  - Raw Events are deeply nested data structures, not suitable for indexing or AI processing algorithms
- Stateless operation in Flink
  - Passes Raw Events to the next stage of event processing
  - Delivers Raw Events unchanged to

▪ Egress Topic



▪ HDFS



# Time Series Events

- Time Series Events are flattened version of the Raw Events
  - Flattened data structures are suitable for indexing or AI processing
    - MI processing algorithms require flattened data – rows and columns
  - Ready to use by Data Scientist!
  - Easy to index by Elasticsearch
- Stateless operation in Flink
  - The Raw Events are parsed and transformed on the fly to Time Series Events and passed on in the pipeline for next stage of event processing
  - Delivers Time Series Events to **Egress Topic** and **HDFS**



# Time Series Events Types

- Six Case event types:
  - Case events with type case
  - Task events with type task
  - Case comment events with type case-comment)
  - Task comment events with type task-comment)
  - Document events with type document)
  - Document comment events with type document-comment)
- Three BPMN event types:
  - Process events with type PROCESS\_\* for top-level process)
  - Activity events with type ACTIVITY\_\* and TASK\_\*)
  - Tracking events with type EVENT\_THROWN and EVENT\_TRACKING\_THROWN)

# Time Series Event Examples

## Case Event

```
{
  "id": "[604C8965-0000-C46C-893B-42824BE88D48]",
  "offset": 2141,
  "partition": 1,
  "timestamp": "2018-08-30T05:26:49.082Z",
  "type": "case",
  "case-instance-id": "[604C8965-0000-C11D-B1D9-FC64F56BDC14]",
  "case-instance-name": "000000140003",
  "state": "Initializing",
  "audit-sequence": 90096,
  "case-folder-id": "[604C8965-0000-C11D-B1D9-FC64F56BDC14]",
  "source-class-id": "[030F12CA-90D6-4639-9A02-5897A3D4B2EC]",
  "source-class-display-name": "CaseST",
  "case-type-name": "CCDM_CaseST",
  "solution-name": "Credit Card Dispute Management",
  "version": "1.0.0",
  "category": "icm",
  "user": "intgpreadmin",
  "data": {
    "CCDM_AssigneeName": "IBM",
    "CCDM_Priority": 5,
    "CCDM_CaseOwner": "Administrator",
    "CCDM_CreatedTime": "2014-04-09T09:30:00.000Z",
    "CCDM_DateRequested": "2014-04-01T09:30:00.000Z",
    "CCDM_Country": "Asia Pacific",
    "CCDM_AllocatedTo": "AllocatedTo",
    "CCDM_DateSent": "2014-04-02T09:30:00.000Z",
    "CCDM_AccountingYear": "AccountYearTest",
    "CCDM_CustomerStatus": "Silver"
  },
  "trace-id": "604c8965-0000-c11d-b1d9-fc64f56bdc14",
  "span-id": "933d6a6f8bc5a078"
}
```

Business data may be present in any of Case Time Series events under 'data' field.

## Process Event

```
{
  "id": "02568186e2065461792162189",
  "offset": 0,
  "partition": 0,
  "version": "1.0.0",
  "type": "PROCESS_STARTED",
  "timestamp": "2018-07-01T06:23:08.033-07:00",
  "bpmSystemId": "5de24de3-6984-4a02-82b9-2696c6f1b7c3",
  "bpmCellName": "PCCell1",
  "processId": "27bfdcb4-dca9-44ad-b89d-f11339fdf3f4",
  "processVersionId": "2064.090b62c2-66b8-4e86-8da0-38eb409ef306",
  "processName": "Extreme Tracking Linked Process",
  "processSnapshotName": "v1.0.4",
  "processApplicationId": "085b1c62-5be1-4fc1-adf2-7ed44445e000",
  "processApplicationSnapshotName": "v1.0.4",
  "processApplicationVersionId": "2064.090b62c2-66b8-4e86-8da0-38eb409ef306",
  "processApplicationName": "Extreme Tracking",
  "processInstanceId": "b0279d9d-4286-4196-b273-0e2fb953f64b.5de24de3-6984-4a02-82b9-2696c6f1b7c3.936-34",
  "startingProcessInstanceId": "b0279d9d-4286-4196-b273-0e2fb953f64b.5de24de3-6984-4a02-82b9-2696c6f1b7c3.936",
  "sequenceId": 42,
  "processState": "Active",
  "trace-id": "d80cf208-d15f-43ce-a22c-7996cf1399fa",
  "span-id": "dda9662471079ab3"
```

- There is no Business Data in Process and Task/Process/Activity Time Series Events
- Business data is included in EVENT\_THROWN and EVENT\_TRACKING\_THROWN event types

# Summary Transformation Events

Summaries  
Transformation

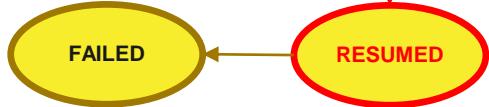
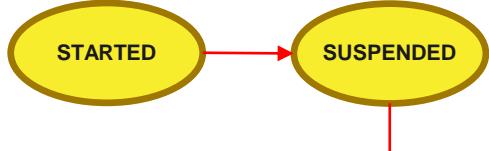
- Summary Transformation Events are “time-aggregated” version of the Time Series Events
- Two type of events
  - **ACTIVE** - reflects the current state of a Process, Task or Case (current state, current duration)
  - **COMPLETED** - records the completed state of Process, Task Case (final state, final duration)
- **Stateful** operation in Flink
  - What is persisted?
    - Since Raw events can be emitted out of order, the Time Series are likely to arrive out of order, a state is used for process summary and activity summary to store events that can not be aggregated yet (for example activities or task that arrive before a process started event)
    - The current process/task event is recorded
    - The current process/task duration is recorded
  - Delivers Summary Events to **Elasticsearch**, **Egress Topic** and **HDFS** (only COMPLETED state)



# Active and Completed Summaries

Summaries Transformation

## Current Process State



## Events in Elasticsearch Indices and HDFS



# Flink Dashboard to Monitor Jobs Execution

Apache Flink Dashboard

Overview Version: 1.5.2 Commit: 1a9b648

Overview Running Jobs Completed Jobs Task Managers Job Manager Submit new Job

 3 Task Managers

 6 Task Slots

 0 Available Task Slots

Total Jobs

Running	3
Finished	0
Canceled	6
Failed	0

Running Jobs

Start Time	End Time	Duration	Job Name	Job ID	Tasks	Status
2018-08-16, 12:43:47	2018-08-20, 20:20:32	4d 7h	dba/bai-lcm	cb7e650521d2c750a3b2161ff0e5daf8	 16 0 0 0 16 0 0 0 0 0	RUNNING
2018-08-17, 17:58:30	2018-08-20, 20:20:32	3d 2h	dba/bai-ingestion	0b8cd76ff54a67c79242abd8bc86f897	 6 0 0 0 0 0 0 0 0 0	RUNNING
2018-08-20, 10:59:56	2018-08-20, 20:20:32	9h 20m	dba/bai-bpmn	b781595ae99c81b36095cf53f097d1c	 16 0 0 0 16 0 0 0 0 0	RUNNING

Completed Jobs

Start Time	End Time	Duration	Job Name	Job ID	Tasks	Status

# Service Level Agreement for BAI

- Is it acceptable to lose events ?
- Which policy should be implemented when Kafka is not reachable or slow (drop events / or block decisions ) ?
- The Kafka client policy configuration is part the BAI Emitter configuration, using standard Kafka parameters
- It is recommended to have an HA Kafka and BAI topology (default configuration with Kubernetes)

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# Embedded Elastic Search & Kibana



Open Distro for Elasticsearch

<https://opendistro.github.io/for-elasticsearch/>

- BAI includes Amazon OpenDistro for Elasticsearch 0.9.0
  - The 0.9.0 release includes Elasticsearch 6.7.1, Kibana 6.7.1 from upstream
    - <https://aws.amazon.com/elasticsearch-service/>
  - Includes Elasticsearch security plugin:
    - Provides a security plugin allowing role-based access to the Elasticsearch cluster
    - Will be installed when upgrading from a previous BAI release

**security-kibana-plugin**  
Open Distro for Elasticsearch Security Kibana Plugin  
JavaScript Apache-2.0 29 82 13

# Open Distro for Elasticsearch Security Kibana Plugin

- This plugin for Kibana adds a configuration management UI for the Open Distro for Elasticsearch Security and Security-Advanced-Modules features, as well as authentication, session management and multi-tenancy support to your secured cluster
  - <https://github.com/opendistro-for-elasticsearch/security-kibana-plugin>
  - <https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/security.html>
- Basic features
  - Kibana authentication for Open Distro for Elasticsearch
  - Kibana session management
  - Open Distro for Elasticsearch Security configuration UI
  - Multi-tenancy support for Kibana
- Security plugin limitations
  - Users and passwords are defined in a yaml file (aka internal user database)
  - Alternative authentication methods (LDAP, ActiveDirectory, SAML, OpenID Connect ) and custom certificates are features that are provided as a technical preview, that is, without any support from IBM. (Planned for Q3)
    - LDAP: <https://aws.amazon.com/blogsopensource/ldap-integration-for-open-distro-for-elasticsearch/>
  - Multitenancy is supported in BAI 19.0.2 or higher

# Configuring ES and Kibana Security Using Open Distro

## Customizing Open Distro security configurations by using a secret

- By default, Open Distro for Elasticsearch provides a predefined configuration, which, however, is not suitable for a production environment. For better security, you can customize the security configuration by using a secret.

## Using custom certificates

- By default, embedded Elasticsearch uses the certificates from the Open Distro for Elasticsearch demonstration but you can also provide your own certificates.

## Authenticating and authorizing users by using LDAP

- After you have defined the necessary users and backend roles, you can customize your Open Distro security configuration by using either the Kibana internal user database and LDAP, or LDAP only.

## Changing which user is associated with the admin security role

- You can assign a security role to a different user than the IBM Business Automation Insights default admin user.

## Using multitenancy to manage access to Kibana resources

- You can use the Kibana multitenancy capability to manage Kibana dashboards and their dependencies in separate named containers called tenants.

## More information:

- [https://www.ibm.com/support/knowledgecenter/SSYHZ8\\_19.0.x/com.ibm.dba.bai/topics/tsk\\_bai\\_esk\\_sec\\_rolebased-access.html](https://www.ibm.com/support/knowledgecenter/SSYHZ8_19.0.x/com.ibm.dba.bai/topics/tsk_bai_esk_sec_rolebased-access.html)

# Elasticsearch Security configuration UI

- Access
  - [https://<k8s\\_master\\_ip>:31501/app/security-configuration](https://<k8s_master_ip>:31501/app/security-configuration)
- Security plugin documentation
  - <https://github.com/opendistro-for-elasticsearch/security>
- Authentication Backends
  - Support for Active Directory, LDAP, Kerberos, SAML, and OpenID Connect
- Permissions and Roles
  - Roles define the actions that users can perform: the data they can read, the cluster settings they can modify, the indices to which they can write, etc.
  - Roles are reusable across users, and users can have multiple roles.

The screenshot shows the Kibana interface with a dark blue sidebar on the left containing the Kibana logo and links to various features: Discover, Visualize, Dashboard, Timelion, Alerting, Dev Tools, Management, and Security. The 'Security' link is highlighted with a blue background. To the right, the main area is titled 'Security' and contains three sections: 'Permissions and Roles' (with 'Role Mappings', 'Roles', and 'Action Groups' sub-sections), and 'Authentication Backends' (with 'Internal User Database' sub-section).

Security

Permissions and Roles

Role Mappings

Roles

Action Groups

Authentication Backends

Internal User Database

# Elasticsearch Security configuration UI

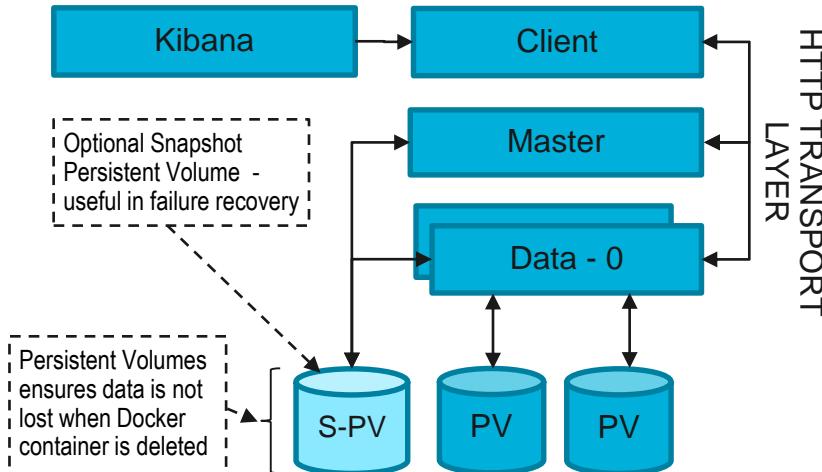
Term	Description
Permission	An individual action, such as creating an index (e.g. indices:admin/create). For a complete list, see <a href="#">Permissions</a> .
Action group	A set of permissions. For example, the predefined SEARCH action group authorizes roles to use the _search and _msearch APIs.
Role	Security roles define the scope of a permission or action group: cluster, index, document, or field. For example, a role named delivery_analyst might have no cluster permissions, the READ action group for all indices that match the delivery-data-* pattern, access to all document types within those indices, and access to all fields except delivery_driver_name.
Backend role	(Optional) Additional, external roles that come from an authorization backend (e.g. LDAP/Active Directory).
User	Users make requests to Elasticsearch clusters. A user has credentials (e.g. a username and password), zero or more backend roles, and zero or more custom attributes.
Role mapping	Users assume roles after they successfully authenticate. Role mappings, well, map roles to users (or backend roles). For example, a mapping of kibana_user (role) to jdoe (user) means that John Doe gains all the permissions of kibana_user after authenticating. Likewise, a mapping of all_access (role) to admin (backend role) means that any user with the backend role of admin (from an LDAP/Active Directory server) gains all the permissions of all_access after authenticating. You can map each role to many users and/or backend roles.

# Elasticsearch-Kibana Pods

```
$ kubectl get pods  
NAME  
master-dev-bai-ibm-dba-ek-client-65f5676886-nbtsw  
master-dev-bai-ibm-dba-ek-data-0  
master-dev-bai-ibm-dba-ek-data-1  
master-dev-bai-ibm-dba-ek-kibana-668fc497cf-cr7nw  
master-dev-bai-ibm-dba-ek-master-75f4f5d4c7-dpxm9
```

	READY	STATUS	RESTARTS	AGE
master-dev-bai-ibm-dba-ek-client-65f5676886-nbtsw	2/2	Running	0	43m
master-dev-bai-ibm-dba-ek-data-0	1/1	Running	0	43m
master-dev-bai-ibm-dba-ek-data-1	1/1	Running	0	43m
master-dev-bai-ibm-dba-ek-kibana-668fc497cf-cr7nw	2/2	Running	0	43m
master-dev-bai-ibm-dba-ek-master-75f4f5d4c7-dpxm9	1/1	Running	0	43m

- **Client Pods** - Serve the Elasticsearch REST API
- **Master Pods** - Master-eligible pods are responsible of ensuring the Elasticsearch cluster is up and running. Among a quorum of running master-eligible pods, one pod gets elected as the master. If at some point the master fails/shut down, a new master will be elected among the remaining master-eligible pods running. If a quorum of running master pods is not satisfied then the ES Cluster will not serve any request
- **Data Pods** are responsible of storing and retrieving the Elasticsearch data from the Persistent Volumes.



# Elasticsearch Head Tool

The screenshot shows the Elasticsearch Head Tool interface. At the top, it displays "cluster health: green (50 of 50)". Below this, the "Cluster Overview" section lists several indices:

- process-summaries-completed-idx-ibm-bai-2018.08.22-000001
- process-summaries-active-idx-ibm-bai-2018.08.22-000001
- case-summaries-completed-idx-ibm-bai-2018.08.22-000001
- case-summaries-active-idx-ibm-bai-2018.08.22-000001
- .kibana

For the first index, a context menu is open, showing options like "Info", "Actions", "Flush", "Optimize...", "Gateway Snapshot", "Test Analyser", "Close", and "Delete...".

On the left, there's a sidebar with cluster status and node details:

- bai-tecsales-ibm-dba-ek-master-7f86dfbd96-76b2d
- bai-tecsales-ibm-dba-ek-client-8487c4fc5d-q8545
- bai-tecsales-ibm-dba-ek-data-0

At the bottom, five rows of green boxes represent shards across multiple nodes, with some boxes containing numbers like 0, 1, 2, 3, or 4.

- *Elasticsearch-Head* is a web front end for browsing and interacting with an [Elastic Search](#) cluster.
- [\*elasticsearch-head\*](#) is hosted and can be downloaded or forked at [github](#)

# Data Retention - Elasticsearch

- Purging old events in Elasticsearch is a customer's responsibility
- In Elasticsearch customers are encouraged to regularly roll over their indices so they can then delete the old ones (this is an ES feature):

<https://www.elastic.co/guide/en/elasticsearch/reference/master/indices-rollover-index.html>

<https://github.ibm.com/dba/taiga-project/blob/master/architecture/decisions/0006-indices-architecture.md>

<https://github.ibm.com/dba/taiga-project/blob/master/architecture/decisions/0010-data-storage-and-rollover.md>

- BAI enables the ES Rollover features by
  - Writing to an index alias
  - Linking the write alias to one and only one index
- Example for Completed Summaries
  - Flink jobs write the completed summaries to index Alias
    - alias process-summaries-completed-write-ibm-bai
  - Which is linked to
    - index process-summaries-completed-idx-ibm-bai-{date}-000001

**process-summaries-completed-  
idx-ibm-bai-2018.08.22-000001**

size: 638ki (1.25Mi)

docs: 48 (96)

# Data Retention - HDFS

- Purging unwanted events in HDFS is a customer's responsibility
- For HDFS the data path contains a date so it is easy to find "old" data and can be used for deleting events:
  - Process time series

/ibm-bai/bpmn-timeseries/[processApplicationId]/[processApplicationVersionId]/process/[processId]/[**date**]/[part]

- Activity time series

/ibm-bai/bpmn-timeseries/[processAppId]/[processAppVersionId]/activity/[processId]/[activityId]/[**date**] /[part]

- Tracking time series

/ibm-bai/bpmn-timeseries/[processAppId]/[processAppVersionId]/tracking/[trackingGroupId]/[**date**]/[part]

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

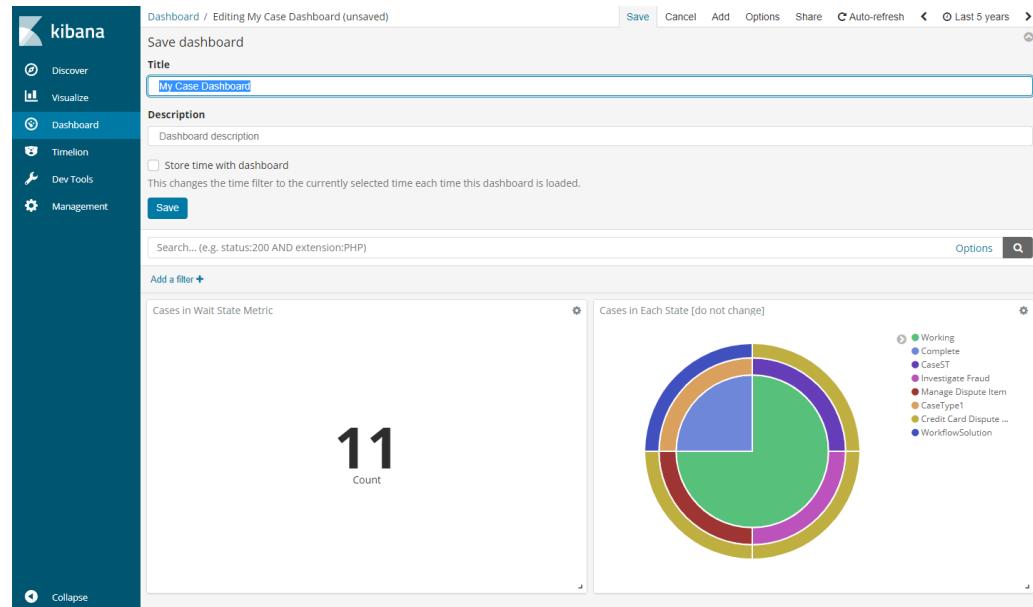
Labs and Demos

What's New

Summary

# Building Kibana Dashboards

1. Create (reuse) Kibana Index Patterns
2. Create a Search
3. Create Visualization
4. Specify Source of Data for Visualization
5. Create a Dashboard
6. Add Visualizations to Dashboard



# 1. Create Index Patterns

The screenshot shows the Kibana Management interface. The left sidebar has a dark blue background with icons for Discover, Visualize, Dashboard, Timelion, Dev Tools, and Management. The Management icon is highlighted with a teal bar. The main area is titled "Management / Kibana" and shows the "Index Patterns" tab selected. A red dashed box highlights the "+Create Index Pattern" button. Below it is a list of index patterns starting with "\*summaries\*". On the right, there's a sidebar with a star icon and the text "This page lists every index pattern in your cluster". It also includes a "Fields (112)" section with a "Filter" input field.

## Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

### Step 1 of 2: Define index pattern

#### Index pattern

You can use a \* as a wildcard in your index pattern.  
You can't use spaces or the characters \, /, ?, ", <, >, |.

Your index pattern can match any of your 5 indices, below.

.kibana

case-summaries-active-idx-ibm-bai-2018.08.22-000001

case-summaries-completed-idx-ibm-bai-2018.08.21

process-summaries-active-idx-ibm-bai-2018.08.22-

process-summaries-completed-idx-ibm-bai-2018.0

activityId

activityId.keyword

activityVersionId

activityVersionId.keyword

### Step 1 of 2: Define index pattern

#### Index pattern

You can use a \* as a wildcard in your index pattern.  
You can't use spaces or the characters \, /, ?, ", <, >, |.

✓ Success! Your index pattern matches 2 indices.

case-summaries-active-idx-ibm-bai-2018.08.22-000001

case-summaries-completed-idx-ibm-bai-2018.08.22-000001

# Index Patterns

In the Kibana dev console we can query against Elasticsearch indexes using wildcards\*. But for Kibana dashboard objects we can't hit the indexes directly, we must use index patterns.

The screenshot shows the Kibana Dev Console interface. On the left, there is a code editor window displaying a Elasticsearch search query:1 #Find all vehicles
2 GET /vehicles\*/\_search
3 {
4 "query": {
5 "match\_all": {}
6 }
7 }The URL path "/vehicles\*/\_search" is highlighted with a red box. To the right of the code editor is the Kibana interface. At the top, it says "2 hits". Below that is a search bar containing ">\_ Search... (e.g. status:200 AND extension:PHP)". To the right of the search bar are buttons for "New", "Save", "Open", "Share", "Inspect", and "Auto-refresh" (with "Auto-refresh" currently selected). Below these buttons are "Options" and "Refr" buttons. The main area shows the search results under the heading "\_source". There are two results listed:> make: Honda model: CRV color: Silver HP: 180 mileage: 24,500 year: 2,015
 price: 12,000 date-added: February 15th 2020, 00:00:00.000 \_id: 1001 \_type: car
 \_index: vehicles-current \_score: 1

> make: Volkswagen model: Polo color: Yellow HP: 95 mileage: 2,500 year: 2,018
 price: 8,995 date-sold: April 1st 2020, 01:00:00.000 date-added: January 5th 2020,
 00:00:00.000 \_id: 1002 \_type: car \_index: vehicles-previous \_score: 1At the bottom of the results area, there is a dropdown menu labeled "Selected fields" which has "vehicles\*" selected. A red box highlights this dropdown menu. The Kibana sidebar on the left includes links for Discover, Visualize, Dashboard, Timelion, Alerting, Dev Tools, and Management.

For vehicles\* to be available in Kibana we must've created an index pattern called vehicles\* first.

# Index Patterns

The name of the index pattern must be a substring of the index or indices you are targeting:  
:

## Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

Include system indices

### Step 1 of 2: Define index pattern

#### Index pattern

vehicles\*

You can use a \* as a wildcard in your index pattern.  
You can't use spaces or the characters \, /, ?, ", <, >, |.

> Next step

✓ Success! Your index pattern matches 2 indices.

vehicles-current

vehicles-previous

If Kibana finds a date fields in your index it will ask if you want to use one of the date fields as a time filter

## Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

Include system indices

### Step 2 of 2: Configure settings

You've defined vehicles-\* as your index pattern. Now you can specify some settings before we create it.

#### Time Filter field name

Refresh



date-added  
date-sold

I don't want to use the Time Filter

> Show advanced options

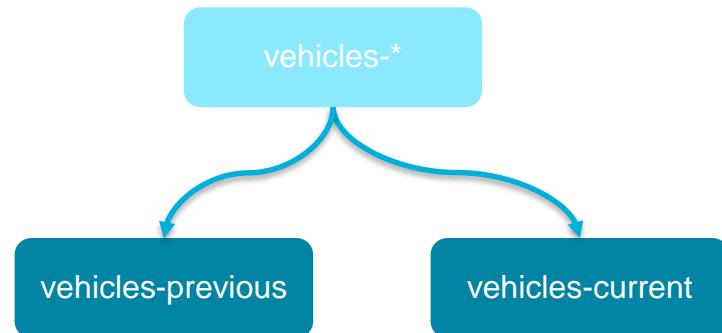
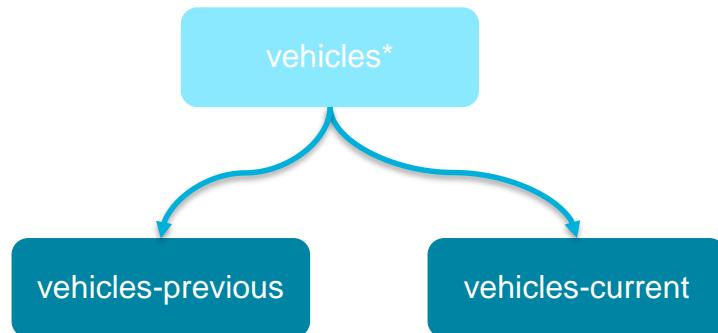
< Back

Create index pattern

I'll do this twice. First to create vehicles-\* and again to create vehicles\*

# Index Patterns

We've got two index patterns called `vehicles*` and `vehicles-*`. One of them has a time filter, one of them doesn't. You can't tell this just from the name of the index pattern.



**vehicles\***

**vehicles-\***

Time Filter field name: date-added

# BPMN Documents

So what does that have to do with BAI ?

```
GET /process-su*/_search
{
  "query": {
    "bool": {
      "must": [
        { "match": {"processApplicationName" :
          "Cash Allocation"}}
      ]
    }
  },
  "size": 20
}
```

BAI is maintaining two indexes for process and task data. One for active stuff & one for completed stuff.

When stuff happens to your process instance in BAW, BAI creates/updates/deletes the doc in ElasticSearch.

```
_index" : "process-summaries-active-idx-ibm-bai-2020.01.10-000001",
_type : "process-summary",
_id" : "7e641241-a3e7-4feb-b121-a55ff5ce6ab3.39d25db3-d087-4941-8ac3-14c038b5b180.1013",
"score" : 3.4465055,
"_source" : {
  "version" : "1.0.4",
  "timestamp" : "2020-04-17T08:41:17.924-05:00"
}
"type" : "process",
```

```
_index" : "process-summaries-completed-idx-ibm-bai-2020.01.10-000001",
_type : "process-summary",
_id" : "7e641241-a3e7-4feb-b121-a55ff5ce6ab3.39d25db3-d087-4941-8ac3-14c038b5b180.1012.5",
"score" : 3.492166,
"_source" : {
  "version" : "1.0.4",
  "timestamp" : "2020-04-17T08:41:14.244-05:00",
  "bpmCellName" : "BAWCloudServer1903",
  "bpmSystemId" : "39d25db3-d087-4941-8ac3-14c038b5b180",
  "potentialPerformerId" : "All Users_T_da7e4d23-78cb-4483-98ed-b9c238308a03.d6fbddfa-f329-4
  "potentialPerformerName" : "All Users",
  "performerId" : "cp4aAdmin",
  "performerName" : "cp4aAdmin",
  "type" : "userTask",
  "name" : "Receive Payment",
  "id" : "7e641241-a3e7-4feb-b121-a55ff5ce6ab3.39d25db3-d087-4941-8ac3-14c038b5b180.1012.5"
}
```

# BAW Index Patterns

We have multiple index patterns that contain BAW BPMN data, but why ?

## BPMN index patterns

For all the index patterns that use a time field, the matching documents that are returned by Kibana queries have that time field and you can use the Kibana time range picker .

### process-su\*

This index pattern provides a way to parse BPMN summaries stored in Elasticsearch. It does not provide any time field. Therefore, all searches and visualizations based on this index pattern span the entire data set, and the time range picker is not available for them.

### process-sum\*

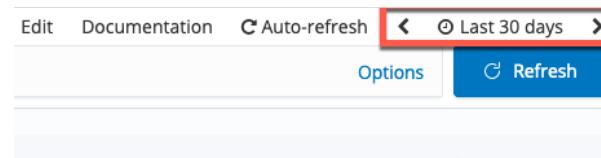
This index pattern enables you to parse BPMN summaries that are stored in Elasticsearch, based on the **timestamp** field, which allows you to filter out results by using the time range picker. The **timestamp** field is updated each time the BPMN processing job updates a summary document.

Table 1. Index patterns for BPMN process summaries

Index pattern	Time field	Process instance summary	Activity summary
New in 18.0.2 process-su*	Not applicable	Yes	Yes
New in 18.0.2 process-sum*	<b>timestamp</b>	Yes	Yes
process-summaries-*	<b>startTime</b>	Yes	Yes
process-summaries*	<b>completedTime</b>	Yes	Yes
process-summarie*	<b>terminatedTime</b>	Yes	Yes
process-summarie*	<b>suspendedTime</b>	Yes	No
process-summar*	<b>migratedTime</b>	Yes	No
process-summa*	<b>resumedTime</b>	Yes	No
process-summ*	<b>atRiskDateAssignedTime</b>	Yes	No
*rocess-summaries-*	<b>failedTime</b>	Yes	Yes
*rocess-summaries*	<b>deletedTime</b>	Yes	No
*rocess-summarie*	<b>claimedTime</b>	No	Yes
New in 18.0.2 *rocess-summar*	<b>dueDateAssignedTime</b>	Yes	No

BAW index patterns contain the same data (\*) but, they're just indexed on a different **time field**.

This is important in Kibana as all dashboards have a time filter in the top-right corner.



This time filter is passed to each visualization, each visualization is built on a search and an index pattern. Different index patterns use different times for their filter.

To start, keep it simple, use **process-summaries-\*** unless process start time is unsuitable as a time filter.

(\*) Process-su\* doesn't have a time filter so it contains the entire dataset.

# Why are so many index pattern aliases?

- There are specific indices for case, bpel, bpmn etc plus active versus complete
- Searches
  - The index patterns that are pre defined to match specific combinations of these so they can be included within single searches.
- Specification of time filter fields
  - When an index pattern is defined it is possible to specify which field will be used for the time filter and as there will be different business time fields (create time, complete time etc) you will need different index patterns for each.
  - Therefore there are a number of predefine index patterns to provided for the different combinations of source indices and time fields
- For more information on aliases and their relationship with indexes see:
  - [https://www.ibm.com/support/knowledgecenter/SSYHZ8\\_19.0.x/com.ibm.dba.bai/topics/ref\\_bai\\_es\\_aliases\\_indexes.html](https://www.ibm.com/support/knowledgecenter/SSYHZ8_19.0.x/com.ibm.dba.bai/topics/ref_bai_es_aliases_indexes.html)

## 2. Create a Search

The search displays the documents that match the search

3. Name the Search and Save

2. Use Lucene syntax to create a search.

1. Specify Index Pattern

state search column results

Kibana Discover interface showing a search for 'state:waiting'. The search bar contains 'Cases in Wait State'. The index pattern is set to '\*-summaries-\*'. The search results show four documents from August 26th, 2018, all with the state 'Waiting'. A histogram on the right shows a single bar for August 26th, 2018, with a count of 8.

Time	state	_id	solution-name	start-time
August 26th 2018, 10:08:17.310	Waiting	{50907665-0200-c2b5-b4d2-07393f860ba8}	Credit Card Dispute Management	August 26th 2018, 10:08:17.310
August 26th 2018, 10:08:17.215	Waiting	{50907665-0100-cdf1-b74c-68e1518ee027}	Credit Card Dispute Management	August 26th 2018, 10:08:17.215
August 26th 2018, 10:08:17.356	Waiting	{50907665-0300-ca43-a2d5-8f3678e5c212}	Credit Card Dispute Management	August 26th 2018, 10:08:17.356
August 26th 2018, 10:08:17.373	Waiting	{50907665-0300-cda9-934f-c5a43df66cbc}	Credit Card Dispute Management	August 26th 2018, 10:08:17.373

# 3. Create a New Visualization

Visualize / New

Select visualization type

Basic Charts

- Area
- Heat Map
- Horizontal Bar
- Line
- Pie
- Vertical Bar

Data

- Data Table
- Gauge
- Goal
- Metric

Maps

- Coordinate Map
- Region Map

Time Series

- Timelion
- Visual Builder

Other

- Controls
- Markdown
- Tag Cloud
- Vega

Click the visualization type you want to instantiate

# 4. Specify Source of Data for Visualization

Visualize / Cases in Wait State Metric (unsaved)

Save Visualization

Cases in Wait State Metric

Save

Linked to Saved Search: Cases in Wait State

Search... (e.g. status:200 AND extension:PHP)

Add a filter +

\*-summaries-\*

Data Options

Metrics

Metric Count

Add metrics

Buckets

Select buckets type

Split Group

Cancel

Save Share Refresh Auto-refresh Year to date

Specify Search

11 Count

The screenshot displays the Kibana Visualize interface for creating a new visualization. The left sidebar shows navigation options: Discover, Visualize (selected), Dashboard, Timelion, Dev Tools, and Management. The main area is titled 'Visualize / Cases in Wait State Metric (unsaved)'. It includes a 'Save Visualization' section with a 'Save' button, a search bar ('Search... (e.g. status:200 AND extension:PHP)'), and an 'Options' button. Below this is a 'Linked to Saved Search' section with a link to 'Cases in Wait State'. A red dashed box highlights this link, and a red arrow points from the text 'Specify Search' to it. The central part of the interface is a '\*-summaries-\*' panel with tabs for 'Data' and 'Options'. Under 'Metrics', there is a 'Metric' section with a 'Count' field and a 'Add metrics' button. Under 'Buckets', there is a 'Select buckets type' section with a 'Split Group' button and a 'Cancel' button. At the top right, there are buttons for 'Save', 'Share', 'Refresh', 'Auto-refresh', and a time range selector ('Year to date'). On the far right, there is a large '11' with the word 'Count' below it, indicating the result of the specified search.

# 5. Create Dashboard

Screenshot of the Kibana Dashboard interface.

The left sidebar shows navigation links: Discover, Visualize, **Dashboard**, Timelion, Dev Tools, and Management. The **Dashboard** link is highlighted.

The main area is titled "Dashboard". It features a search bar labeled "Search..." and a "Create new dashboard" button with a plus icon.

A table lists five dashboards:

Name ↑	Description	Actions
Case - IBM Business Automation Workflow [do not change]		Edit
Case Task - IBM Business Automation Workflow [do not c...]		Edit
Process - IBM Business Automation Workflow		Edit
Process Task - IBM Business Automation Workflow		Edit
Workflow - Hiring Sample [18.0.0.1]		Edit

Pagination at the bottom indicates "1-5 of 5" with previous and next buttons.

Page URL: <https://9.30.188.25/kibana-ibm-dba-ek-bai-techsales/app/kibana#/dashboard>

# 6. Add Visualizations to Dashboard

Dashboard / Editing My Case Dashboard (unsaved)

Save Cancel Add Options Share Auto-refresh Last 5 years

Save dashboard

Title: My Case Dashboard

Description: Dashboard description

Store time with dashboard

This changes the time filter to the currently selected time each time this dashboard is loaded.

Save

Search... (e.g. status:200 AND extension:PHP) Options

Add a filter +

Cases in Wait State Metric

11 Count

Cases in Each State [do not change]

A donut chart titled "Cases in Each State [do not change]". The chart is divided into several segments of different colors: green, blue, purple, yellow, red, and pink. A legend on the right side lists the categories corresponding to these colors: Working, Complete, CaseST, Investigate Fraud, Manage Dispute Item, CaseType1, Credit Card Dispute ..., and WorkflowSolution. The segments represent the relative count of cases for each state.

Working  
Complete  
CaseST  
Investigate Fraud  
Manage Dispute Item  
CaseType1  
Credit Card Dispute ...  
WorkflowSolution

Collapse

# Sharing Kibana Dashboards

- Kibana dashboards can be “shared”
- BPM Process Portal
  - Exposed as iFrame served from the Kibana server
  - Dashboard Client Side Human Service can embed the iFrame in a Coach View
- IBM Content Navigator
  - ICN widgets available to display iFrame content

The screenshot illustrates the sharing mechanism for Kibana dashboards. On the left, the Kibana sidebar shows 'Discover', 'Visualize', 'Dashboard' (selected), 'Timelion', 'Dev Tools', and 'Management'. The main area displays a dashboard titled 'Process - IBM Business Automation Workflow' with sections for 'Share saved dashboard', 'Embedded iframe' (with a copy link to the full URL), and 'Link' (with a copy link to a shortened URL). A red box highlights the 'Share' button, and a red arrow points down to the 'BPM Process Portal' interface below. The BPM portal shows a 'Kibana Based Operational Dashboards' section with tabs for 'Process Performance', 'Task', 'Team Performance' (selected), 'KPI', and 'Business At a Glance'. It features a pie chart titled 'Task By Team' and a bar chart titled 'Average Task Duration by Team'. The 'Team Performance' tab lists team members: Phillip, Jones, Harmon, Geraldine, George, and Dan. The 'KPI' tab shows a single bar for 'Managers'.

# Export and Import

## Dashboard.json

The screenshot shows the Kibana Management interface with the 'Saved Objects' tab selected. The left sidebar includes links for Discover, Visualize, Dashboard, Timelion, Dev Tools, and Management. The main area displays a list of saved objects under 'Edit Saved Objects'. A red box highlights the 'Import' button in the top right corner. Another red box highlights the 'Export' button in the bottom right corner. Red dashed boxes with arrows point from the text annotations to these specific buttons.

Management / Kibana

Index Patterns [Saved Objects](#) Advanced Settings

Edit Saved Objects

From here you can delete saved objects, such as saved searches. You can also edit the raw data of saved objects. Typically objects are only modified via their associated application, which is probably what you should use instead of this screen. Each tab is limited to 100 results. You can use the filter to find objects not in the default list.

Dashboards (6) Searches (25) Visualizations (62)

Search...

<input type="checkbox"/> Title
<input checked="" type="checkbox"/> Case - IBM Business Automation Workflow [do not change]
<input type="checkbox"/> Case Task - IBM Business Automation Workflow [do not change]
<input type="checkbox"/> My Case Dashboard

Import any Saved Object (including Dashboards).

Export Everything

Export any Saved Object (including Dashboards).

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# Source of Business Data in Processes and Cases

- Auto-tracking enabled

▼ **Tracking Groups**

Enable autotracking:

Autotracking name: aEmpRequisition121381434563922

- Tracking groups



- Case properties

Properties Roles In-baskets Documents Business Objects

Add Property OK All

Name	Type	Attributes	Description
Address	String	...	
Application Form Verified	Boolean	...	

- Tracked BO elements

▼ **Details**

Visible in Process Portal:

Alias:

Track this variable:

Short name: reqNum

- Some BPMN events fields contain user names or metadata



General Implementation

Data Mapping Pre & Post Tracking

▼ **Document Property Mapping** +

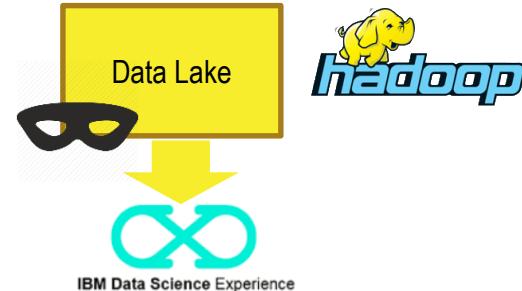
Cc (List of String) ➔	<input type="text"/>	
From (String) ➔	<input type="text"/>	
Name (String) ➔	<input type="text"/>	
Received On (D...) ➔	<input type="text"/>	
Sent On (Date) ➔	<input type="text"/>	
Subject (String) ➔	<input type="text"/>	
To (List of String) ➔	<input type="text"/>	

# Need for Business Data Protection

**GDPR** EU General Data Protection Regulation

General Data Protection Regulation (GDPR) enables EU subjects to have the right to find out what personal data an organization has stored about them and have the right to be forgotten. All organizations active across 28 EU member countries activating initiatives related to GDPR is required to continue conduction business.

GDPR will come into effect **May 25, 2019**



- Why do we need data protection?
  - Business data contained in raw events, time series or summaries are stored in **Hadoop HDFS data** lake without obfuscation.
  - Business data may contain personal information that should not be publicly exposed to data scientists
  - Data privacy regulation such as GDPR may result in non compliance
- BAI's approach to data protection
  - Use various obfuscation techniques to make the data unreadable **before it enters Data Lake**
  - Only data destined to **Long Term Storage** has data protection feature enabled

# Data Protection Rules

- JSON rules define what data is to be protected and how to protect it

```
{  
  "type": <protectionMode>,           } The protection mode to apply to business data fields  
  "params": {  
    <protectionModeParameters>      } Parameters that depend of the chosen protection mode  
  },  
  "fields": [  
    <fieldPath1>,  
    <fieldPath2>,  
    ...  
    <fieldPathN>                   ] Uses JSONPath syntax to business data fields in events  
  ]  
}
```

- Protection modes

Protection Mode	Protection Type	Description
hash	Anonymization	Irreversibly destroys any way of re-identifying the business data.
encrypt	Pseudonymization	Substitutes business data with a reversible and consistent value in such a way that additional information is required to re-identify the data.
remove	Data removal	Business data is removed from the event if it contains it

# Example of Pseudonymization Rule for Tracking Group

```
{  
  "type": "encrypt",  
  "params": {  
    "key": "MTIzNDU2NzgxMjM0NTY3OA==",  
    "iv": "MTIzNDU2Nzg4NzY1NDMyMQ=="  
  },  
  "fields": [  
    "$['trackedFields']['HiringManager.string']",  
    "$['data']['aEmpRequisition121381434563922']['HiringManager.string']",  
    "$['data']['bpm']['tracking-point']['tracked-field'][?(@.name=='HiringManager')]['content']"  
  ]  
}
```

JSON Paths to three fields that have the name of he Hiring Manger

## Pre Tracking Point

Tracking group:

 [TrackingGroup](#)

Select...

New...

X

Performance warehouse ID:

323edbc2-fd27-44bf-8239-8feb456a33e3

Sort tracked fields by:

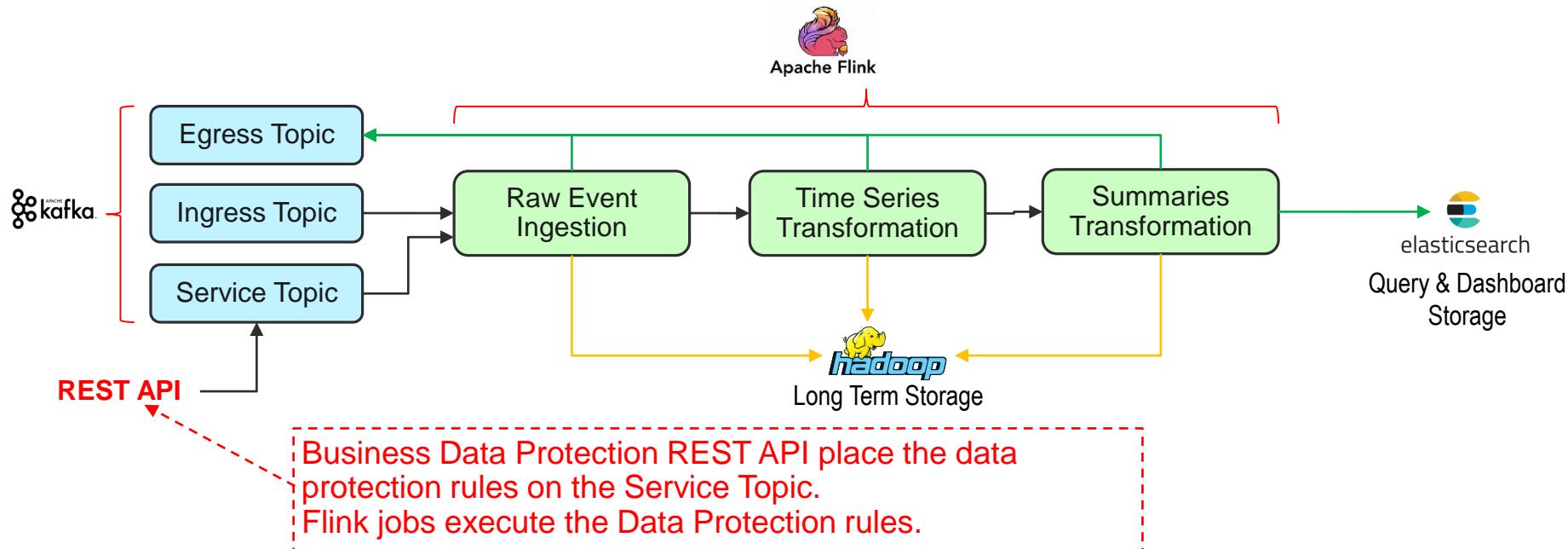
Default Order

tw.local.requisition.reqNum



HiringManager string)

# Enabling Business Data Protection



- Data Protection Rules are injected using Admin REST API which post the rules to the Admin Service Topic
- Only data destined to Long Term Storage has data protection feature enabled

# Enabling Business Data Protection REST API

- **Enabling** business data protection
  - First, the rules must be uploaded as JSON files using REST API.
  - The REST API endpoint for uploading the rules can be Ingress or NodePort
  - Ingress Example:

```
curl -X POST https://mycluster.icp/admin/api/actions
-H 'content-type: multipart/form-data'
-F file=@/users/bai/validRules.json
```
- **Disabling** business data protection
  - Restart the Flink processing job (ICM/BPMN/Ingestion)
  - Or upload an empty JSON object {} using above method

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

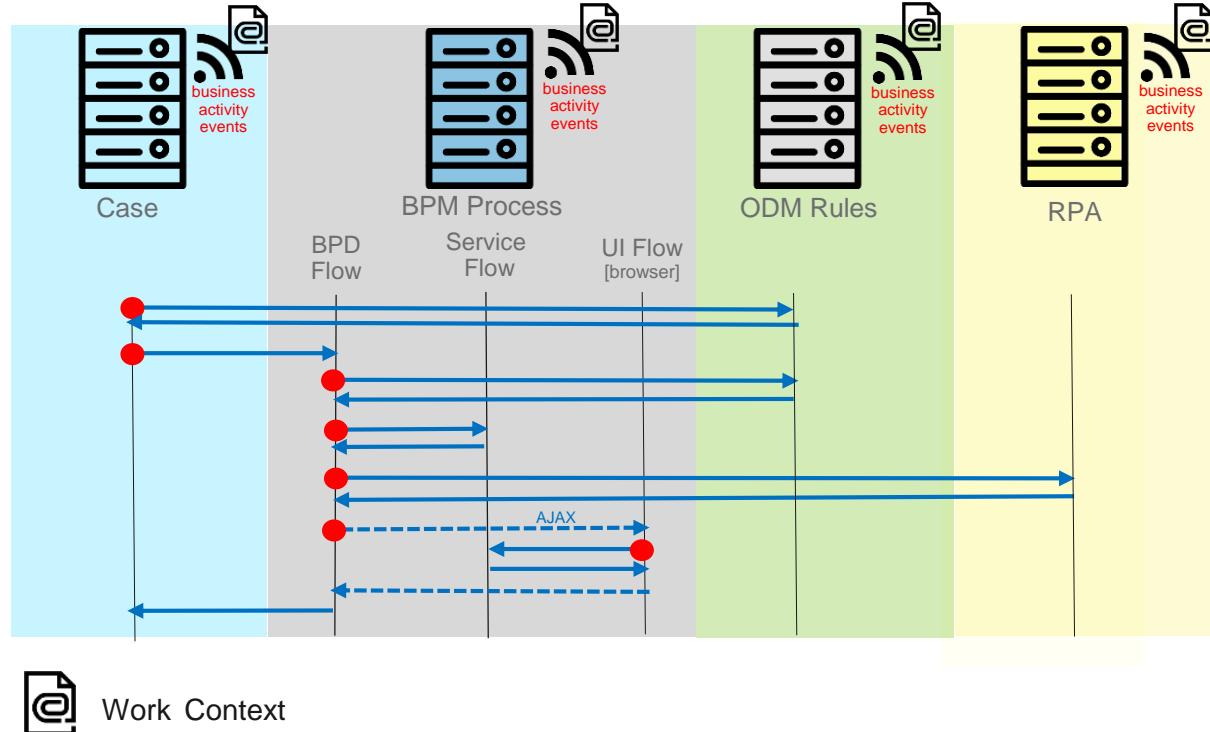
Labs and Demos

What's New

Summary

# What is Cross DBA Event Correlation?

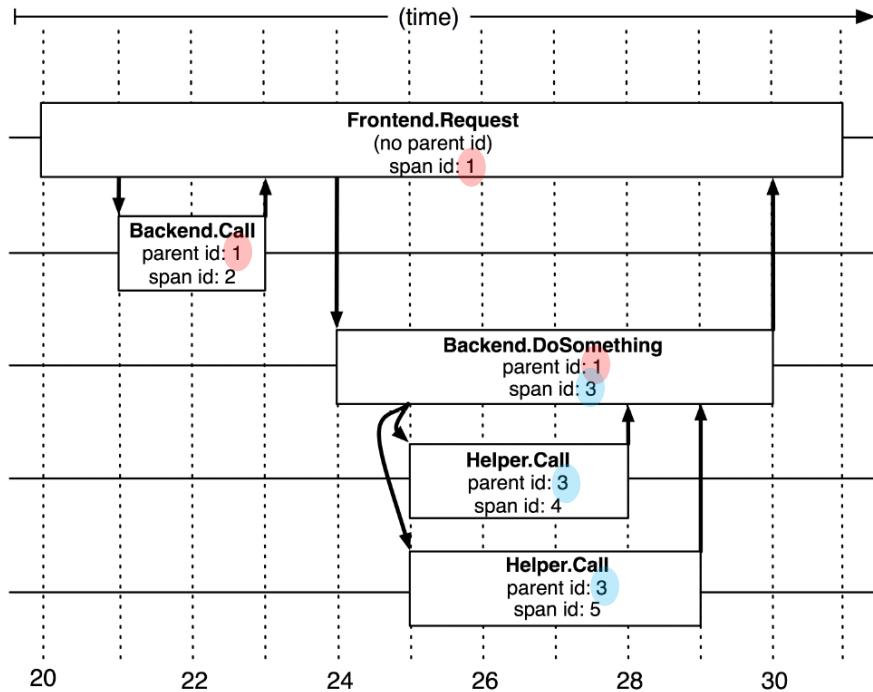
- Ability to trace the chain of calls across the separate DBA components (Case, BPM, ODM, RPA, Content etc.) as they interact with each other as part of user work.
- To be able to identify source of events and trace determine how they propagate to each DBA component, we added BAI Event “Work Context”



[https://www.ibm.com/support/knowledgecenter/SSYHZ8\\_20.0.x/com.ibm.dba.bai/topics/con\\_bai\\_events\\_workcontext\\_correlation.html](https://www.ibm.com/support/knowledgecenter/SSYHZ8_20.0.x/com.ibm.dba.bai/topics/con_bai_events_workcontext_correlation.html)

# Work Context Propagation

- Parent component manages the Work Context
  - Adds to BAI event
  - Passes it to a child (which is another DBA component i.e. Process calls ODM Decision)
- Work Context 
  - Parent id
    - Null (if top parent)
    - Span Id of parent
  - Span Id
    - Parent Id + n (where n is the number of calls made by a component, first call n=1)
- Kibana or MI tools can use Work Context to reconstruct the call sequence



# Passing Work Context to Child in API Headers

- Work Context for nested interactions are passed as header fields (transparent to downstream component)
  - http/ REST interactions can use header field names
  - WS/ SOAP interactions can use same as SOAP header fields
- This strategy allows us to **keep unchanged the exiting interfaces** for REST/WS API to start Case/ Process /ODM Decision / etc...
- BAW 18.0.0.1 3Q updates have changed the runtime
  - To expect Work Context information in the header
  - To add to Work Context Information automatically when launching components

# Enabling Correlation Context Propagation

- Included in BAW 18.0.0.2 and higher
- Not available for ODM

# Agenda

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# BAI Topologies and Sizing

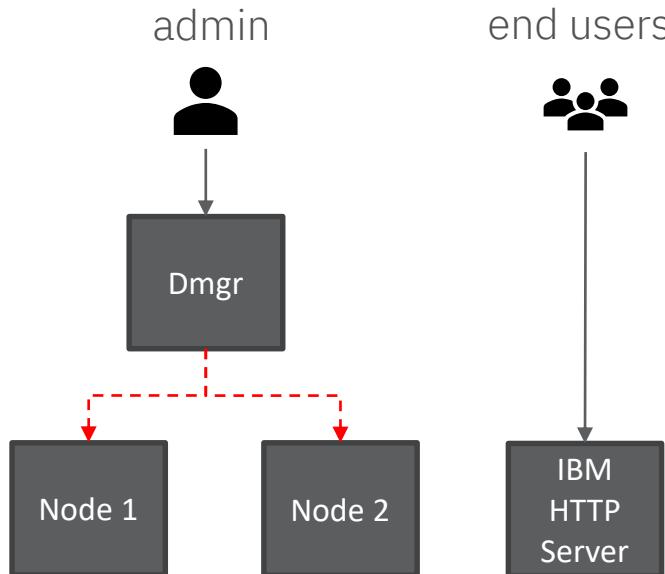
BAI on Kubernetes

BAI for a Server

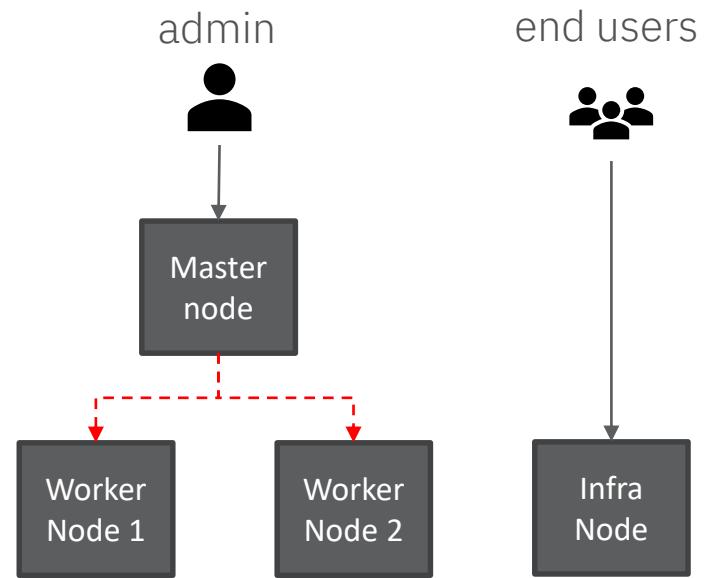
BAI Sizing

# Kubernetes Cluster versus WebSphere Cell

WebSphere Cell

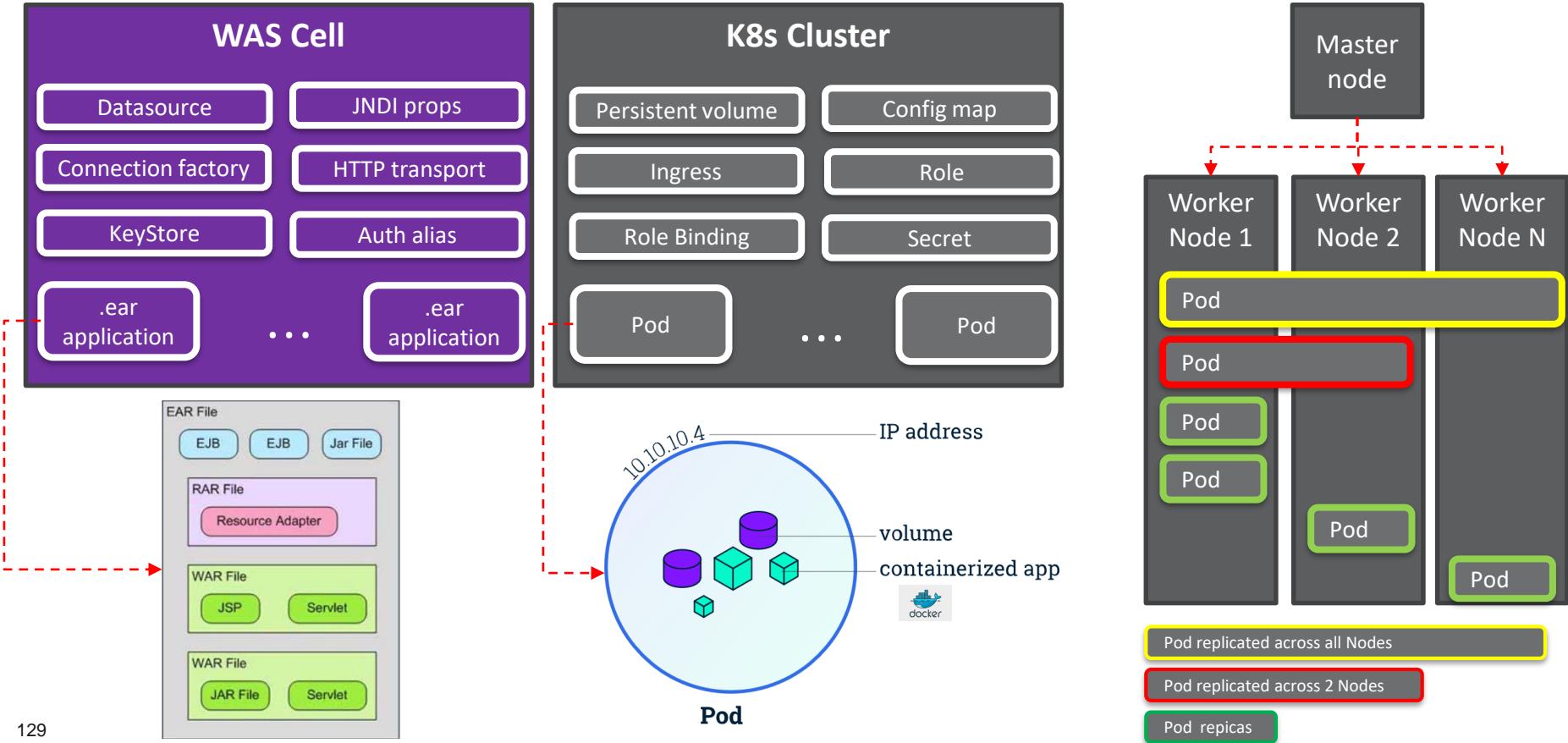


Kubernetes Cluster

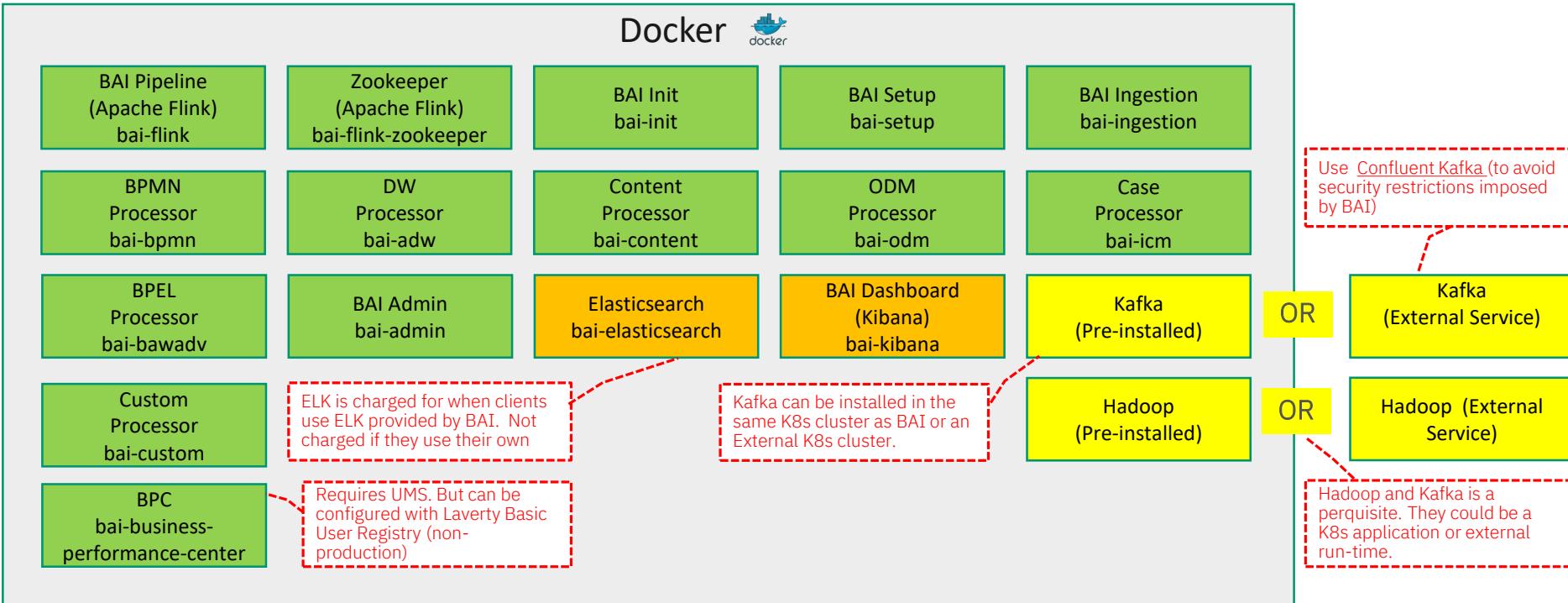


Exposes master node

# Kubernetes Cluster versus WebSphere Cell



# Business Automation Insights Docker Containers



Red Hat OpenShift Container Platform or CNCF Certified Kubernetes

# Pods – Initialization Only

Pods used only to install the event processors and Kibana dashboards for each pillar (Case, BPMN, ODM etc...)

These pods are transitory. They run only on startup and mostly initialize the environment, workload for startup.



Name	Status	
bai-poc-ibm-dba-ek-security-config-nm4kd	✓ Completed	Elasticsearch bai-elasticsearch
bai-poc-bai-bawadv-5hxb9	✓ Completed	BPEL Processor bai-bawadv
bai-poc-bai-bpmn-p292b	✓ Completed	BPMN Processor bai-bpmn
bai-poc-bai-content-z8zv8	✓ Completed	Content Processor bai-content
bai-poc-bai-icm-vn5xg	✓ Completed	Case Processor bai-icm
bai-poc-bai-odm-mzbxm	✓ Completed	ODM Processor bai-odm
bai-poc-bai-setup-k5vc2	✓ Completed	BAI Init bai-init
bai-poc-ibm-dba-ek-security-config-wkq7n	✗ Error	

# Runtime Pods

In HA mode multiple Job Manager instances are running and one is elected as a leader. If the leader fails ZooKeeper, transfers leadership to one of the other running Job Managers.

BAI Admin. Very small workflow to update anonymization

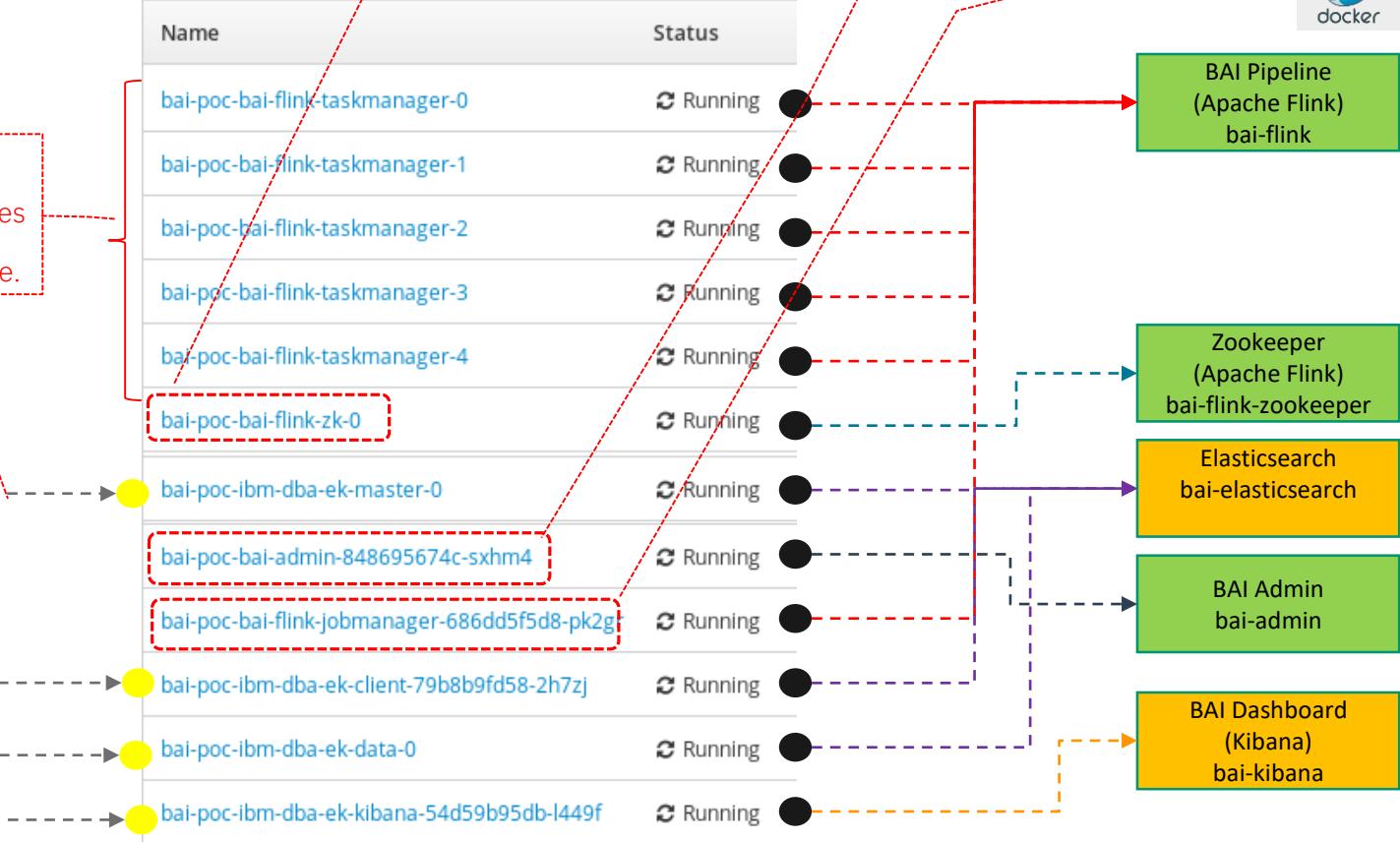
Job Manager . distributes jobs to task manager. Only 1 is needed

flink-taskmanager pods do the real work.  
They will typically need 2-3 Nodes and ample memory.  
There is one pod per emitter type.

Pods created by BAI Init.  
This Pod runs only at install time!



BAI Init  
bai-init



# Bew BPC Pod in BAI 20.0.0.2

NAME	READY	STATUS	RESTARTS	AGE
ibm-cp4a-operator-7fc6577cc-k7tdq	1/1	Running	0	42m
master-dev-bai-admin-56b6c95f8f-gq4rb	1/1	Running	0	36m
master-dev-bai-business-performance-center-65ff9b5664-bnz5s	1/1	Running	0	36m
master-dev-bai-flink-jobmanager-58976db488-mrtmc	1/1	Running	0	36m
master-dev-bai-flink-taskmanager-0	1/1	Running	0	36m
master-dev-bai-flink-taskmanager-1	1/1	Running	0	36m
master-dev-bai-flink-taskmanager-2	1/1	Running	0	36m
master-dev-bai-flink-taskmanager-3	1/1	Running	0	36m
master-dev-bai-flink-taskmanager-4	1/1	Running	0	36m
master-dev-bai-flink-taskmanager-5	1/1	Running	0	36m
master-dev-bai-flink-taskmanager-6	1/1	Running	0	36m
master-dev-bai-flink-taskmanager-7	1/1	Running	0	36m
master-dev-bai-flink-zk-0	1/1	Running	0	36m
master-dev-ibm-dba-ek-client-57b586b8df-8794g	1/1	Running	0	36m
master-dev-ibm-dba-ek-data-0	1/1	Running	0	36m
master-dev-ibm-dba-ek-kibana-697f8d469c-2w4wt	1/1	Running	0	36m
master-dev-ibm-dba-ek-master-0	1/1	Running	0	36m



# Storage

The screenshot shows the Storage interface within a project named "bai-poc". The left sidebar lists categories: Overview, Applications, Builds, Resources, Storage (selected), Monitoring, and Catalog. The main area is titled "Storage" with a "Create Storage" button. A search bar and a "Filter by label" input field are also present. The table displays three entries:

Name	Status	Capacity	Access Modes	Age
data-bai-poc-ibm-dba-ek-data-0	✓ Bound to volume bai-poc-ek-pv0	10 GiB	RWO (Read-Write-Once)	2 months
data-bai-poc-ibm-dba-ek-master-0	✓ Bound to volume bai-poc-ek-pv1	10 GiB	RWO (Read-Write-Once)	2 months
bai-poc-pvc	✓ Bound to volume bai-poc-pv	20 GiB	RWX (Read-Write-Many)	2 months

Red dashed arrows point from the application names in the table to red boxes labeled "Kibana", "Elastic Search", and "Flink jobs".

Storage

Search Catalog

Add to Project

Overview

Applications

Builds

Resources

Storage

Monitoring

Catalog

Learn More

Create Storage

Filter by label

Add

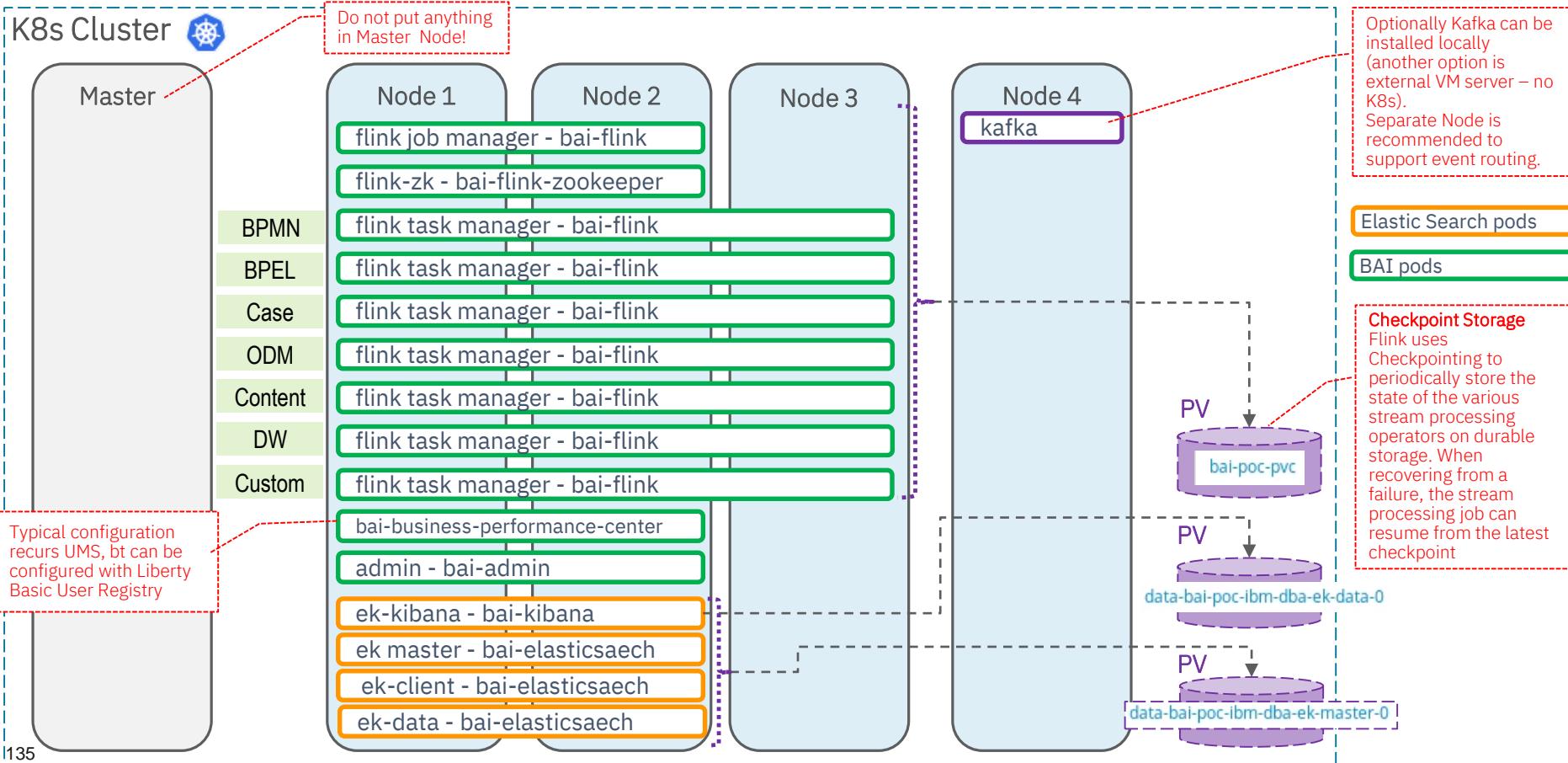
Name	Status	Capacity	Access Modes	Age
data-bai-poc-ibm-dba-ek-data-0	✓ Bound to volume bai-poc-ek-pv0	10 GiB	RWO (Read-Write-Once)	2 months
data-bai-poc-ibm-dba-ek-master-0	✓ Bound to volume bai-poc-ek-pv1	10 GiB	RWO (Read-Write-Once)	2 months
bai-poc-pvc	✓ Bound to volume bai-poc-pv	20 GiB	RWX (Read-Write-Many)	2 months

Kibana

Elastic Search

Flink jobs

# BAI Runtime Topology



# BAI Topologies and Sizing

BAI on Kubernetes

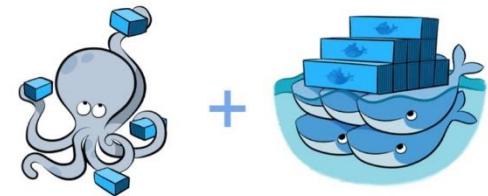
BAI for a Server

BAI Sizing

# Business Automation Insights ~~Single Node K8s Free~~

- Prerequisites

- Production environment - Red Hat 7.7 / linux
  - Development and test environment - MacOS
  - Docker runtime (Docker Compose)



- Restrictions

- For output to visualization dashboards and data science applications, it can ingest BPMN, BPEL, Case Manager, and ODM events only
  - HDFS is not supported
  - No out of the box support for HA, DR, Autoscaling
  - External installations of Elasticsearch and Kibana are not supported

# BAI for a Server Key Value

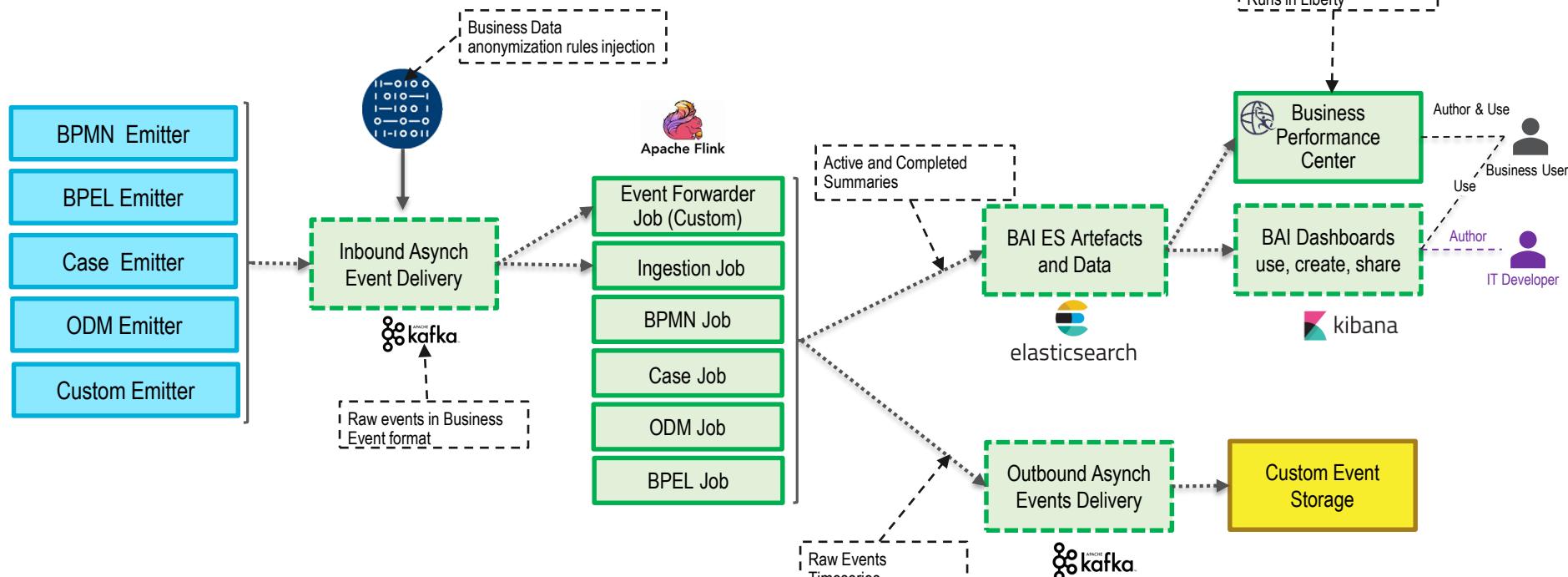
Name	Command	State	Ports
data_admin_1	/usr/bin/scl enable rh-nod ...	Up (healthy)	0.0.0.0:6892->6892/tcp
data_elasticsearch_1	/docker-entrypoint.sh /bin ...	Up (healthy)	9200/tcp, 9300/tcp
data_jobmanager_1	/bin/bash -c /data/contain ...	Up (healthy)	6123/tcp, 8081/tcp
data_kafka_1	/etc/confluent/docker/run	Up	0.0.0.0:29092->29092/tcp, 9092/tcp
data_kibana_1	bin/kibana	Up (healthy)	0.0.0.0:5601->5601/tcp
data_taskmanager_1	/bin/bash -c /data/contain ...	Up (healthy)	6123/tcp, 8081/tcp
data_zookeeper_1	/etc/confluent/docker/run	Up (healthy)	0.0.0.0:2121->2121/tcp, 2181/tcp, 2888/tcp, 3888/tcp

- Does not require the overhead of Kubernetes
  - Runs as Docker containers and operates on one computer (**hence the name Single Node**) without the need to configure a Kubernetes cluster
- Significantly less expensive
  - Less H/W needed
- Unlike K8s, Kibana is bundled and automatically installed and started
- Simpler install that installs: BAI, Kafka, Elasticsearch with Kibana and HDFS
  - Configure ([.env file](#))
  - Unzip
  - Set up security

# BAI for a Server Limitations

Limitation or issue	Description
Elasticsearch and Kafka	You can use only embedded Elasticsearch and Confluent Kafka.
Security of communications to Kafka	Salted Challenge Response Authentication Mechanism (SCRAM), plain SSL, SSL with Kerberos authentication, and Kerberos authentication are not supported. Only SSL with user name and password is supported.
Apache ZooKeeper	The version of ZooKeeper that is bundled with Confluent Kafka does not support SSL. For more information, see the <a href="#">ZooKeeper</a> page of the Confluent Kafka documentation.
Apache Flink	The Flink web interface is not accessible from Chrome on MacOS Catalina. The Flink job manager and the task manager are not automatically restarted after a machine reboot. Therefore, you must restart IBM Business Automation Insights.
Event emitters	The event emitters for FileNet® Content Manager (content) and for IBM Automation Digital Worker (adw) are not supported.
HDFS	Injection of events to an HDFS data lake is supported but with no authentication mechanisms.

# BAI for Server Architecture



  Customer supplied

  Running on DBA Product Environment

  BAI Artifacts

  BAI Artifacts that can be supplied by customer

# BAI for a Server is free for non-prod use

- Components Not Used for Establishing Required Entitlements.
- When determining the number of entitlements required for Licensee's installation or use of the Program, the installation or use of the following Program components are not taken into consideration. In other words, Licensee may install and use the following Program components, under the license terms, but these components are not used to determine the number of entitlements required for the Program.- IBM Business Automation Studio
  - IBM Business Automation Navigator
  - IBM Business Automation Application Designer
  - IBM Business Automation Application Engine when used in Non-Production
  - IBM Automation Digital Worker when used in Non-Production
  - IBM Business Automation Insights when used in Non-Production**

# BAI Topologies and Sizing

BAI on Kubernetes

BAI for a Server

BAI Sizing

# BAI Sizing Factors

Will determine:

- number of Nodes, Pod Replicas, Nodes a Pod spans
- amount of CPU, Storage (PVs)

- Number of DBA Pillars being monitored
  - The number becomes a multiplier for the factors below
- Pillar type
  - Workflow workloads are stateful (correlations, summaries) hence need more PVs/CPU/Memory
- Throughput
  - Events per day
- Event size
  - Large business data payload
- Retention duration of Kafka
  - Kafka act as a buffer between source and Flank processing jobs
- Retention duration in Elastic Search
  - How much data you want to keep in ES Indices
- Number of Dashboards and users making ES requests
  - Adequate memory needed by Kibana and ES for optimal dashboard performance (min 2GB)
- Use of HDFS
  - Storage size related to retention of data
- Providing HA/DR capabilities
  - Needs additional CPU capacity to absorb a backlog of events if a queue has been built during an outage in the chain (for example ES, or HDFS).
  - Need a minimum of 3 (odd) entities to be HA
  - Odd number to have a quorum to determine who is wrong (for example if a node has an issue with data consistency - the two others will agree to eliminate it and replace its data with a fresh replica)

[ICP4A Sizing Link](#)

# BAI Sizing Detailed Information

<https://ibm.box.com/v/BAI---Sizing>

# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary



# Labs and Demos - Visualize



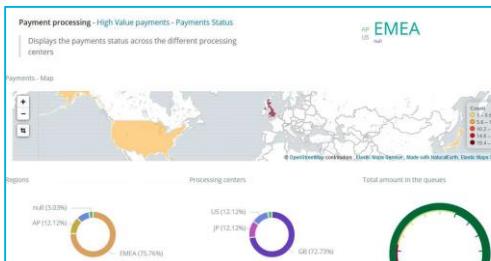
- Blue Demos: Blue Demos: <https://bluedemos.com/show/3398> (contact pacholsk@ca.ibm.com)
- Labs: (i) Learn how to demo BAI (ii) Create events either using a process or from data events generated synthetically

## Lab 1: Billing Disputes



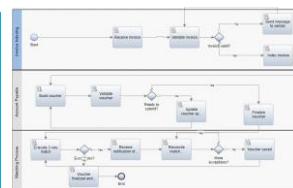
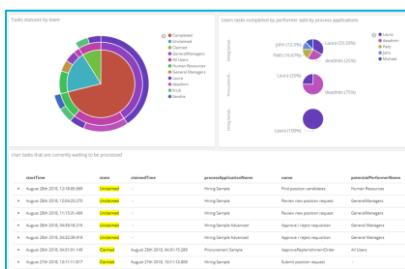
- Learn how to use Paul's Random Event Generator Technique to create BAI events.
- Create new Visualizations.
- Edit Kibana Dashboards.

## Lab 2: Payment Processing



- Learn how to read process instance input data from CVS files and start new processes instances to create BAI events.
- Create new Visualizations.
- Edit Kibana Dashboards.

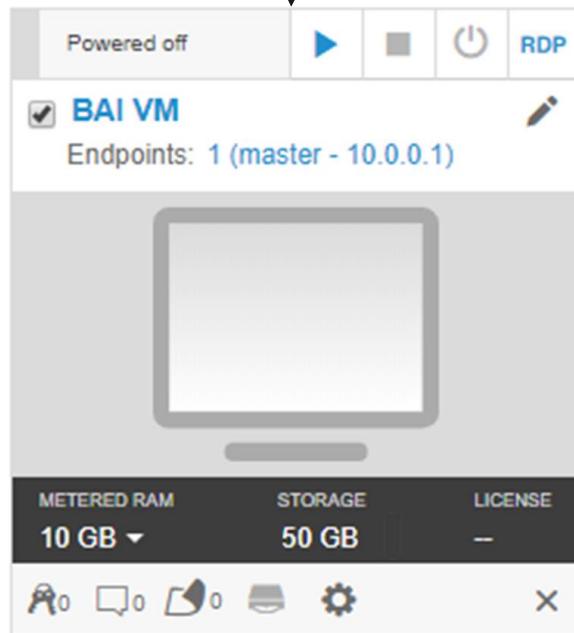
## Lab 3: Invoice Pressing



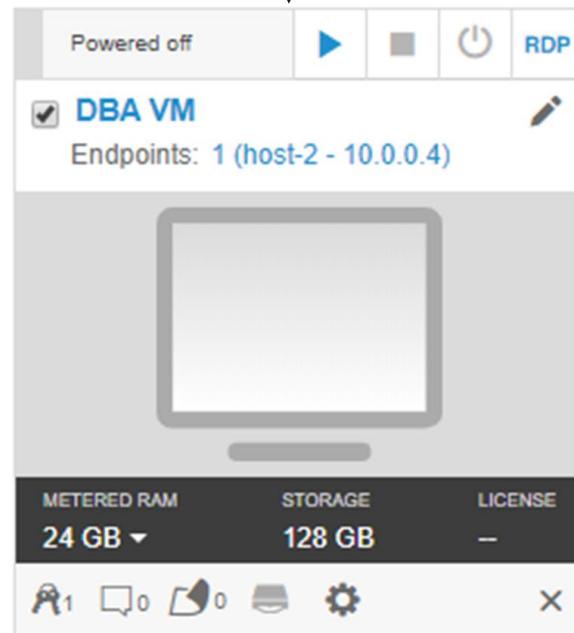
- Demo BPMN and Task OTB operational dashboard populated with real task and process events!
- Learn how to use a technique developed by Swapnil to travel in time backwards to generate past events

# Blue Demos Environment

Businesses automation Insight  
Single Node 19.0.3 Server



DBA V2.4+ VM

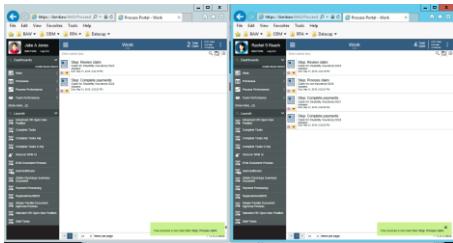


# Labs and Demos – Enable AI (Tech Preview)



- Blue Demos: <https://dte2.us1a.cirrus.ibm.com/show/3365> (contact pacholsk@ca.ibm.com)
- Labs: (i) Learn how to demo Next Best Task (lab 1 and 2); (ii) Show the value of ML assisted Process Performance dashboards

## Lab 1: Train ML Model



- Generate process instances and run complete tasks using automated procedure
- Demo how to use the ML runtime to train Next Best Task ML model

## Lab 2: Next Best Task Business Value



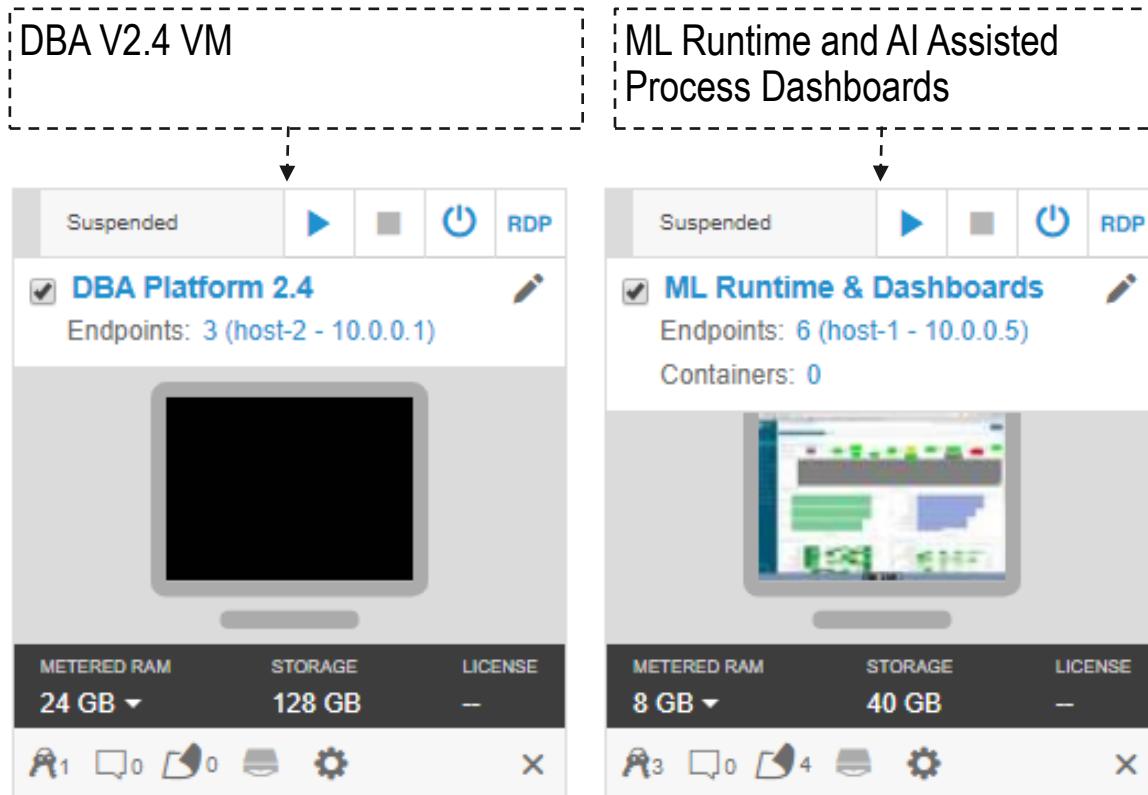
- Compare task and process completion statistics of random completion versus NBT assisted completion.
- Demonstrate the business value of using Next Best Task

## Lab 3: Process Performance Analytics



- Demo ML assisted process performance dashboards
- Learn how to use the dashboards to identify process bottlenecks and determine what actions to take to improve process performance

# Blue Demos Environment



# DBA Technical Sales Accelerators – BAI Folder

<https://ibm.box.com/s/d3nditj2ab5cd2fvmh0x3vgaogqznxr1>

All Files > ★ Digital Business Automation Technical Sales Accelerators

Folder Description  
Questions/Comments - contact: Paul Pacholski - pacholsk@ca.ibm.com  
For content description open Digital Process Automation Technical Sales Collateral.docx

Name	Updated	Size
Business Automation Insights	Yesterday by PAUL PACHOLSKI	40 Files
IBM Business Automation Workflow	Yesterday by PAUL PACHOLSKI	15 Files
IBM Cloud Pak for Automation	Feb 19, 2020 by PAUL PACHOLSKI	5 Files
Business Automation Application Designer	Feb 13, 2020 by PAUL PACHOLSKI	5 Files
DOWNLOAD	Feb 11, 2020 by PAUL PACHOLSKI	6 Files
IBM Digital Business Automation VMware Images	Feb 11, 2020 by PAUL PACHOLSKI	20 Files
DBA for SAP	Feb 10, 2020 by PAUL PACHOLSKI	2 Files
RPA - IBM RPA with Automation Anywhere	Feb 5, 2020 by PAUL PACHOLSKI	88 Files
DBA Platform Automation Pattern Demos	Jan 30, 2020 by Lars Grespan	20 Files
RPA - UiPath	Dec 11, 2019 by PAUL PACHOLSKI	2 Files
BPM Advanced Technical Overview and Introduction	Nov 21, 2019 by PAUL PACHOLSKI	10 Files



# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# Agenda

19.0.2

19.0.3

20.0.1

20.0.2

# Updates – BAI 19.0.2

- Elasticsearch and Kibana support LDAP authentication and authorization When using **embedded Elasticsearch and Kibana** security of the Elasticsearch cluster is managed through the Open Distro for Elasticsearch security plug-in. You can customize security configuration and define user authentication and authorization by using either an **internal user database and LDAP, or LDAP only**
- **Elasticsearch data snapshots** Enable snapshot storage for embedded Elasticsearch and Kibana installation. **Take snapshots from Elasticsearch data**, and also restore IBM Business Automation Insights indexes and Open Distro security indexes.
- **Packaging changes for the developer edition** Images are available on IBM Entitled Registry (on DockerHub ). Charts are available on GitHub
- **Kibana multitenancy** Restrict access to Kibana resources by defining **multiple tenants**. When multitenancy is enabled, execute scripts to retrieve Kibana global-tenant resources and import them into a new tenant
- **Confluent Kafka** Configure for **secure data communications** by using Confluent Platform Helm charts (tested/supported). Red Hat provides AMQ Stream based on Kafka 2.2.x that is supported in unsecure mode.
- **New parameter for Flink jobs** The new flink.initStorageDirectory parameter indicates whether the Flink storage directory needs to be initialized.
- **Case activity summaries** A new user parameter is available

# New Kibana Authentication Backends - LDAP

The screenshot shows the Kibana interface with the sidebar open. The 'Security' tab is selected. In the main area, there are several cards:

- Permissions and Roles**: Contains icons for **Role Mappings** (blue circles connected by lines) and **Roles** (three blue people). A tooltip says: **Available with Open Distro ES (shipped in 19.0.1)**.
- Authentication Backends**: Contains an icon for **Internal User Database** (blue person icon inside a box).
- System**: Contains icons for **Authentication & Authorization** (blue circular maze) and **Purge Cache** (blue circular arrow).

A red dashed box highlights the **Internal User Database** card. A red dashed arrow points from the text "The Open Distro security plugin can authenticate users" to this card.

**Kibana uses the Open Distro security plugin to**

- manage authentication of users in the Kibana UI
- manage authorization of authenticated users (with roles bound to users or roles bound to group of users using roles mapping)

**The Open Distro security plugin can authenticate users**

- Using an **Internal User Database**
- Using an **LDAP backend**
- Using an **Internal User Database chained to an LDAP backend**

# Data Snapshots for Backup and Restore

- Open Distro ES (shipped in 19.0.1) enables data Snapshot capability which is tested and is fully supported in 19.0.2
- API support only <https://www.elastic.co/guide/en/elasticsearch/reference/6.7/modules-snapshots.html>
- Backup. You can take a snapshot of individual indices or of the entire cluster

```
PUT /_snapshot/my_backup
{
  "type": "fs",
  "settings": {
    "location": "my_backup_location"
  }
}
```

- Restore. Snapshots can be restored into a running cluster

```
POST /_snapshot/my_backup/snapshot_1/_restore
```

- IBM Event Stream is no longer bundled with CloudPak for Automation 19.0.2 (still supported)
- Customers have to provide their own Kafka
  - On OpenShift, RedHat provides AMQ Stream based on Kafka 2.2.x that is supported in **unsecure mode**
  - In 19.0.0.2 **Confluent Kafka** was tested successfully in unsecure and **secure mode**
  - Helm charts on Github:
    - <https://github.com/confluentinc/cp-helm-charts>

# Kibana Multitenancy

- A tenant is a container for holding Kibana Saved Objects
- By default multi-tenancy is disabled (all dashboards are stored in a single global tenant)
  - When enabled, **each tenant has their own private set of Kibana objects**
- Tenants are initially empty and must be populated (using Export / Import) with: dashboards, visualizations, index-patterns and searches

The screenshot shows the Kibana interface with a sidebar on the left containing icons for Discover, Visualize, Dashboard, Timelion, Alerting, Dev Tools, Management, Security, and Tenants. The 'Tenants' icon is highlighted with a blue background. The main area is titled 'Select Tenant' and shows the message 'Active tenant: baw-tenant'. It displays a table with one row:

Name	Permissions
baw-tenant	read/write

Buttons for 'Show Dashboards', 'Show Visualizations', and 'Select' are visible at the bottom of the table.

# Agenda

19.0.2

19.0.3

20.0.1

20.0.2

# Updates – BAI 19.0.3

## Single node deployment

- To use development or production capabilities of IBM Business Automation Insights you can now deploy the component on a single computer. Learn more...

## Event emitter for IBM Automation Digital Worker

- When you install Digital Worker, you can turn on the option for sending data to Business Automation Insights. Digital Worker dashboards.

## Apache Flink 1.9

- IBM Business Automation Insights 19.0.3 now supports version 1.9 of Apache Flink.

# Agenda

19.0.2

19.0.3

20.0.1

20.0.2

# Updates – BAI 20.0.1

- Support of HDFS authentication
  - Use Kerberos 5 authentication for communications between Flink processors and HDFS
- Improved assignment of cores to Flink processors
  - The logic of the processors deployer was changed to start in non-optimal mode without changing the TASKMANAGER\_LOGICAL\_CPUS value in that case.
- Improved confidentiality of passwords stored in the .env file
  - The passwords assigned to the following variables are base64-encoded: ADMIN\_PASSWORD, KAFKA\_PASSWORD, ELASTICSEARCH\_PASSWORD, KIBANA\_PASSWORD
- Improve failure detection for Flink savepoints before starting BAI
  - When starting BAI, the bai-start script ensures that:
    - Flink savepoints or checkpoint exist in the volumes/flink directory
    - Flink savepoints are not corrupted
  - If above conditions exist
    - BAI Server will not start
    - An explanation is displayed
    - Recovery steps are provided

```
Found invalid checkpoint(s) or savepoint(s):
bawadv:/mnt/pv/savepoints/dba/bai-bawadv/savepoint-3f8f64-8ff9f1de65e1 (not found)
```

Exiting.

# Agenda

19.0.2

19.0.3

20.0.1

20.0.2

# Updates – BAI 20.0.2

- BAI Business Performance Center (BPC)
  - New dashboard application for BAW and ODM designed for business users
  - Integration in BAI for a Server (authentication based on the user local registry of WLP)
  - Integration in BAI-K8S (authentication based on UMS)
  - BPC started automatically when BAI is started
  - Technical introduction deck:  
<https://ibm.box.com/v/BusinessPerformanceCenter>
- Support of custom events
  - Integration in BAI for a Server
  - Integration in BAI-K8S
  - Emitter Samples for Confluent and IBM Event Streams
  - Dashboarding with Kibana
- BAI for a Server
  - Support of Confluent Kafka 5.4.1
  - Support of External Kafka
  - Upgrade from 20.0.1
- BAI on Kubernetes
  - New Support: OCP43, OCP44, ROCKS 4.3 and Event Streams 10.0+ (including support of SASL SCRAM)
  - Changes in the BAI CR and operator role to simplify the installation and configuration of BAI
  - Support of demo pattern for OCP43
  - Upgrade from 20.0.1

# Support for Custom Event

- Prior to v20.0.2 BAI could only receive events from ICP4A Pillar Applications such as ODM or BAW
- In this release BAI can now ingest events from any source
  - Providing the events emitted are JSON documents and comply with the Avro specification
  - BAI can process such **custom events** as raw events and send them to dashboards and HFDS data lakes
  - Custom events are consumed by a new BAI Flink job called the **Event Forwarder**
- The Avro based custom events can be produced by a **Custom Event Emitter**
  - The emitter code is custom, it is not provided by BAI
  - Sample code is available: <https://github.com/icp4a/bai-emitter-samples>

README.md

### Samples for IBM Business Automation Insights custom event emitter

This directory provides samples related to using custom events with Business Automation Insights. The samples demonstrate how to:

- Register an Avro schema.
- Send to Business Automation Insights events that conform to the registered schema, by using either Confluent or IBM Event Streams Avro APIs.

Before sending events, you must configure Business Automation Insights to process custom events. For more information about such configuration, see the [IBM Knowledge Center](#).



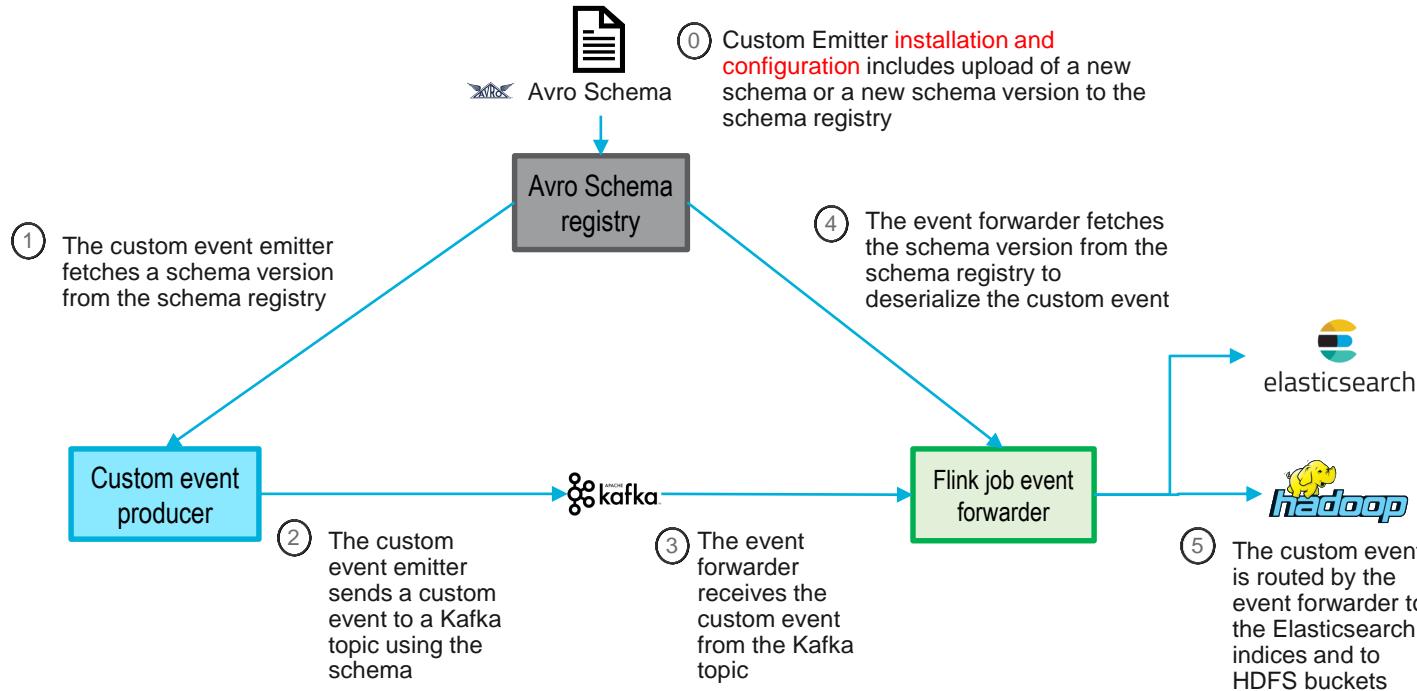
### What is Avro?

Apache Avro is a **language-neutral data serialization system**. It was developed by Doug Cutting, the father of Hadoop. Since Hadoop writable classes lack language portability, Avro becomes quite helpful, as it deals with data formats that can be processed by multiple languages. Avro is a preferred tool to serialize data in Hadoop.

Avro has a schema-based system. A language-independent schema is associated with its read and write operations. Avro serializes the data which has a built-in schema. Avro serializes the data into a compact binary format, which can be deserialized by any application. Avro uses JSON format to declare the data structures.

# Custom Emitter Architecture

**Schema Registry** provides a serving layer for your metadata. It provides a RESTful interface for storing and retrieving **Avro schemas**. It stores a versioned history of all **schemas**, provides multiple compatibility settings and allows evolution of **schemas** according to the configured compatibility setting.



# Visualizations using Custom Emitter Events

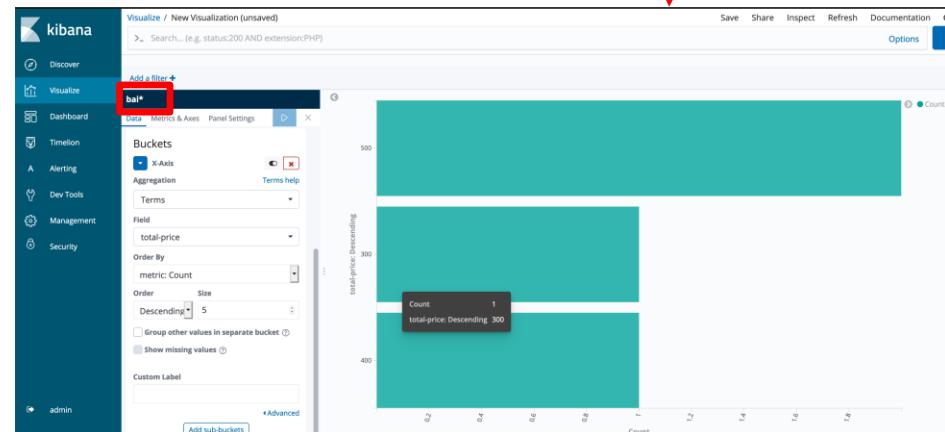
The screenshot shows the Kibana interface. On the left, the navigation bar has 'Management' selected. Under 'Index Patterns', there is a 'Create index pattern' button. A red arrow points from this button to the 'Create index pattern' step in the central panel. The central panel shows a list of index patterns starting with 'bai\*' and a form where 'bai\*' is entered into the 'Index pattern' field. A success message says 'Success! Your index pattern matches 1 index.' To the right, two numbered steps are outlined in red dashed boxes:

1. Create an index pattern matching the Elasticsearch index of the routing mappings
2. Use the index pattern to create a Visualization

## Note:

In BAI 20.0.2, Visualizations and Dashboard for custom events can be done using Kibana.

Custom Event processing in the **Business Performance Center** is NOT supported.



# Support for External Kafka in BAI for Server 20.0.2

Name	Command	State	Ports
data_admin_1	/usr/bin/scl enable rh-nod ...	Up (healthy)	0.0.0.0:6892->6892/tcp
data_elasticsearch_1	/docker-entrypoint.sh /bin ...	Up (healthy)	9200/tcp, 9300/tcp
data_jdbcmanager_1	/bin/bash -c /data/contain ...	Up (healthy)	6123/tcp, 8081/tcp
data_kafka_1	/etc/confluent/docker/run	Up	0.0.0.0:29092->29092/tcp, 9092/tcp
data_kibana_1	bin/kibana	Up (healthy)	0.0.0.0:5601->5601/tcp
data_taskmanager_1	/bin/bash -c /data/contain ...	Up (healthy)	6123/tcp, 8081/tcp
data_zookeeper_1	/etc/confluent/docker/run	Up (healthy)	0.0.0.0:2121->2121/tcp, 2181/tcp, 2888/tcp, 3888/tcp

- In BAI for a Server 20.0.2, it is possible to use:
  - An embedded Confluent Kafka server v5.4.1 (including ZooKeeper)
  - An **external Kafka server supplied customer**
  - New variables in the .env file to support external Kafka
    - KAFKA\_EMBEDDED\_SUPPORT\_ENABLED, KAFKA\_EXTERNAL\_HOSTNAME, KAFKA\_EXTERNAL\_IP, KAFKA\_EXTERNAL\_CA\_FILE
  - Limitations of using external Kafka
    - The only supported authentication scheme for external Kafka is SASL\_SSL/PLAIN
    - Event Streams 2019.4.x can be used but not Event Streams 10.0 (it uses SASL\_SSL/SCRAM)
    - There is no support for external Avro schema registry
    - There is no support for custom events
- When starting the Server (bin/bai-start --acceptLicense –init)
  - Answer NO to the question
  - Do you want to start the embedded Confluent Kafka? (yes/NO)

# Automated Configuration of Event Emitters for BAI Clients

Used by operator roles of BAI and BAI pillars (BAW, ODM...) to automatically configure their Kafka clients

- Fully automatic for **demo/evaluation**
- Script for automatic config in **enterprise**
  - <https://github.com/icp4a/cert-kubernetes/blob/20.0.2/scripts/pull-eventstreams-connection-info.sh>
  - Augments the CR with the `shared_configuration.kafka_configuration` section and deploys the connection secret
  - If Event Streams is in a different cluster or namespace, the secret must be moved to the ICP4A namespace

```
#####
# Kafka client configuration for IBM Business Automation Insights and other ICP4A products.
#
# Only used when BAI is selected as optional component.
#####
kafka_configuration:
    # Comma-separated list of hosts:port for connection to the Kafka cluster.
    # This field is mandatory for any Kafka configuration.
    bootstrap_servers:
        # Value for the Kafka security.protocol property
        # Valid values: PLAINTEXT, SSL, SASL_PLAINTEXT, SASL_SSL. Default: PLAINTEXT.
        security_protocol:
            # Value for the Kafka sasl.mechanism property
            # Valid values: PLAIN, SCRAM-SHA-512. Default: PLAIN.
            sasl_mechanism:
                # If the Kafka server requires authentication or uses SSL communications, the value of this field
                # must provide the name of a secret that holds the following keys as base64-encoded strings:
                # kafka-username: Kafka username; leave empty if no authentication
                # kafka-password: Kafka password; leave empty if no authentication
                # kafka-server-certificate: server certificate for SSL communications; leave empty if SSL protocol
```

# Case Stages Support in Business Automation Insights

New

In this release we added full support for Case Stages in BAI

- Modified Case Event Emitter to process stage objects and produce new ‘stage’ type events.
- Added new Flink job to
  - Process the stage events and store the timeseries in HDFS
  - Create and store the stage summaries in ElasticSearch

<input type="checkbox"/>	Workflow - Stages
<input type="checkbox"/>	In-progress stages [do not change]
<input type="checkbox"/>	Completed stages [do not change]
<input type="checkbox"/>	Stage summaries [do not change]
<input type="checkbox"/>	Incoming and outgoing stages [do not change]
<input type="checkbox"/>	Incoming and outgoing stages [do not change]
<input type="checkbox"/>	Average age of in-progress stages [do not change]
<input type="checkbox"/>	Stages in progress [do not change]
<input type="checkbox"/>	Stages in each state [do not change]
<input type="checkbox"/>	Incoming and outgoing stages in past 3 months [do not change]
<input type="checkbox"/>	Stage specification [do not change]
<input type="checkbox"/>	Total number of stages [do not change]
<input type="checkbox"/>	Average elapsed time of completed stages [do not change]
<input type="checkbox"/>	Average stage duration [do not change]
<input type="checkbox"/>	Stages completed statistics [do not change]
<input type="checkbox"/>	Stages started statistics [do not change]

Kibana artifacts related to Case Stages:

- 5 Searches
- 10 Visualizations
- 1 Dashboard

# Case Stages Dashboard and Visualizations

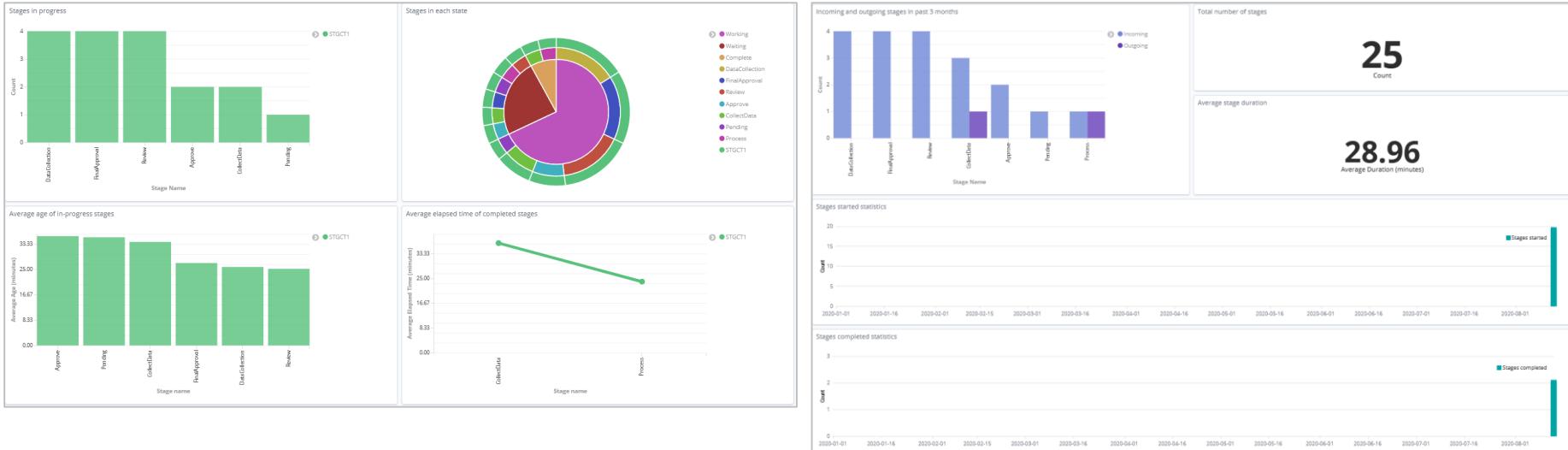
New

STG\_STGCT1\_000000180001 | Modified: 9/23/2019, 3:33 PM | STGCT1

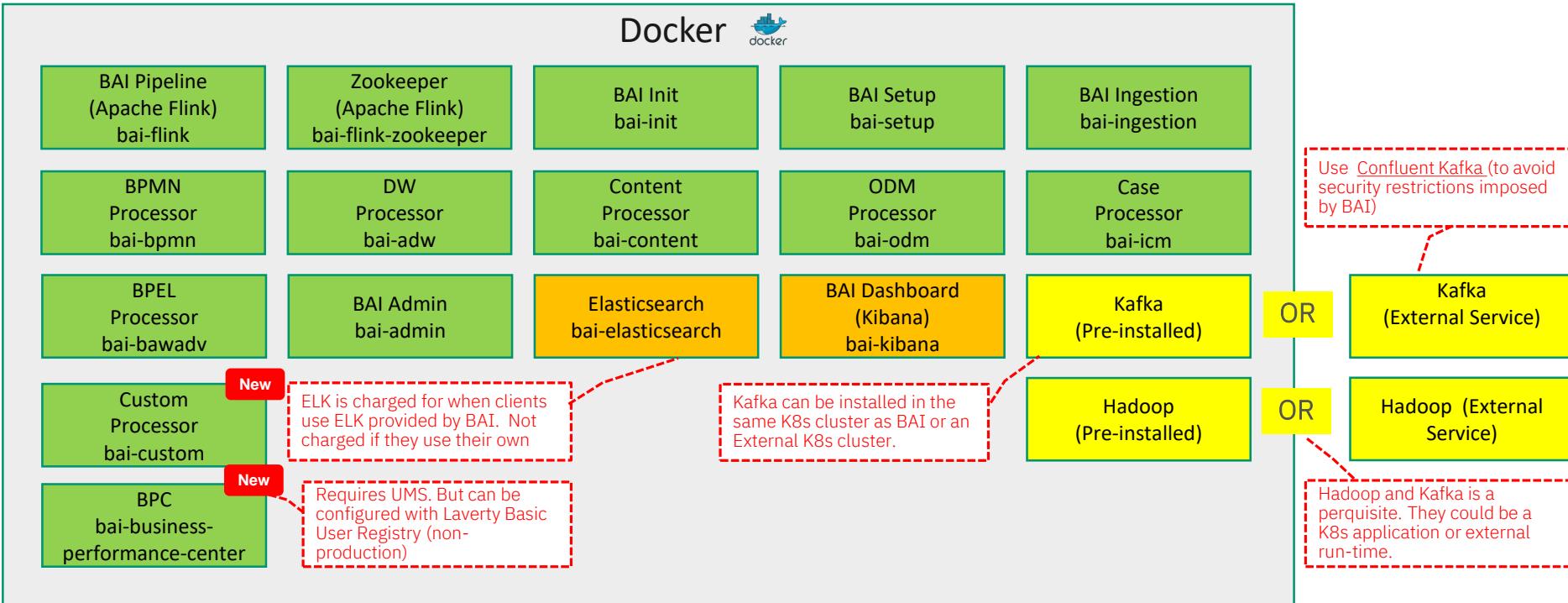
Comments Split Case ▾  
Release Hold Complete Restart

Save Close

Stage1 Stage2 Stage3 Stage4

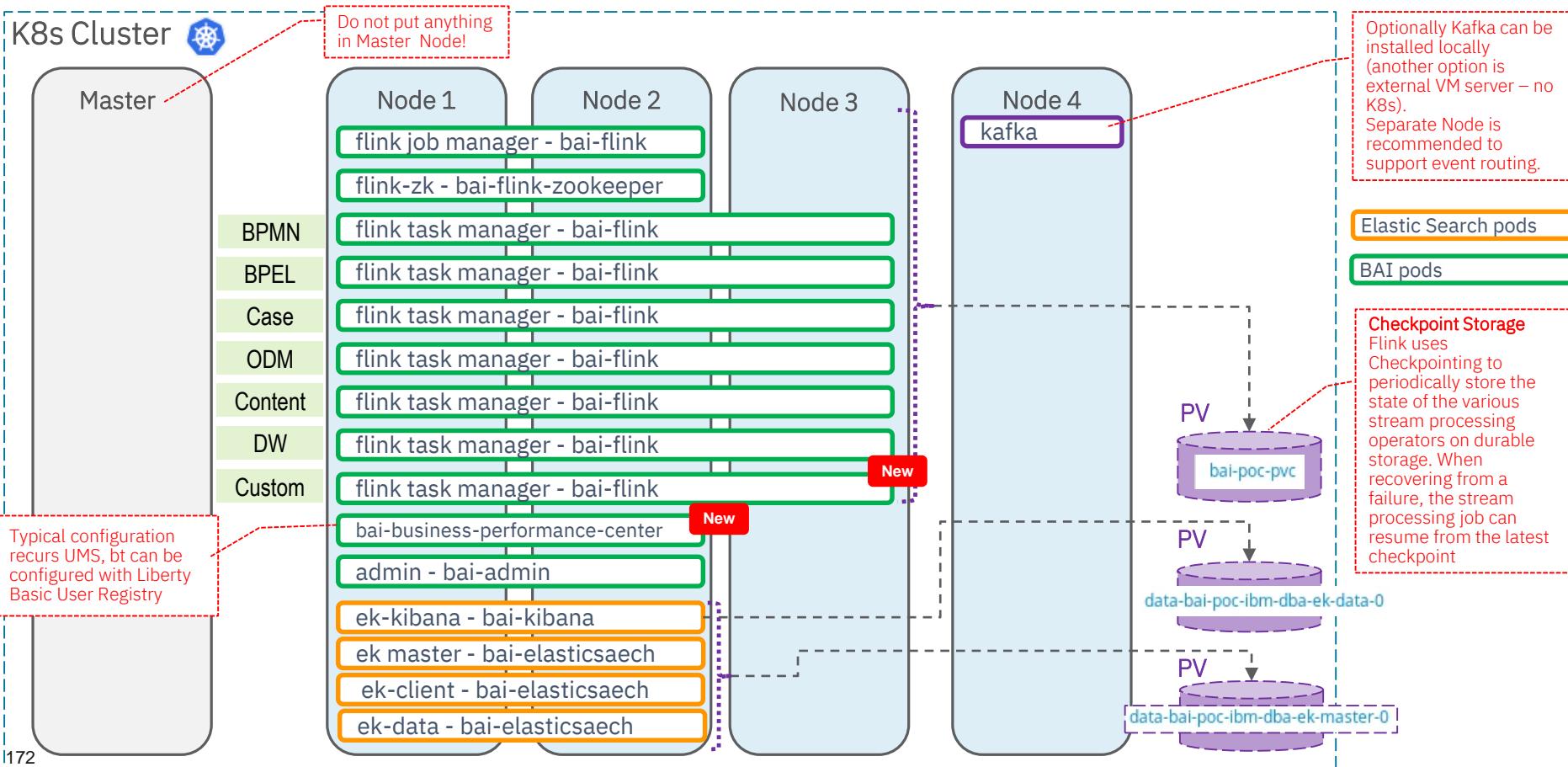


# Business Automation Insights Docker Containers



Red Hat OpenShift Container Platform or CNCF Certified Kubernetes

# BAI Runtime Topology



# Agenda

What is Business Automation Insights?

High Level Overview and Architecture

What's in the Box?

Install and Configure

Event Processing

Elasticsearch

Kibana

Business Data Protection

Cross DBA Event Correlation

BAI Topologies and Sizing

Labs and Demos

What's New

Summary

# Technical Value of IBM BAI – Build versus Buy

- BAI transforms and simplifies the raw events into **flattened format** (time series events)
  - To enable efficient usage to build Kibana dashboards and to enable consumption by AI modelling tools
- **Aggregates time series events** into summaries entities.
  - There is one summary per activity and process instance always reflecting the latest state of the activity or the process.
  - The aggregation also computes and stores activities and processes durations.
  - For BPMN and BPEL (which are stateful), considering the complexity of the BPMN raw events that is not an easy task.
- Sends raw and time series events as well a completed summaries to **HDFS** for data science scenarios
- Sends summary (both active and completed) entities to **Elasticsearch** for dashboarding / business users' scenarios
- Dynamically applies anonymization rules (ability to delete, hash or encrypt fields) before sending data into HDFS
- **Parallelizes** the ingestion and processing to provide high **throughput** scalability
- Manages the **HA and DR** capabilities such that in case of failure no data is lost, and processing can continue
- Adjusts event processing rate automatically as arrival rate increases
  - In case HDFS and/or Elasticsearch is slow at ingesting events, the **processing rate is automatically** adjusted, and no overloading happens
- BAI is at the **beginning of its journey** to provide the operational intelligence capability of the DBA platform and IBM has the intention to add additional features to BAI.
- BAI provides simple yet powerful dashboard authoring and management capabilities Business Users (**Business Performance Center**)

New



# What is a Virtual Processor Core ?

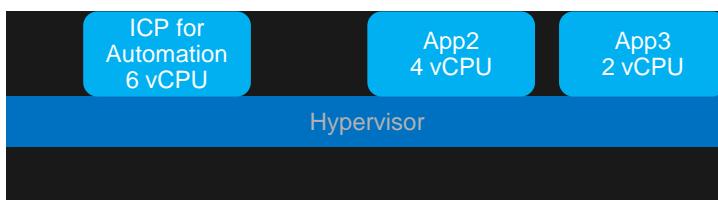
Virtual Processor Core is a unit of measure by which the Program can be licensed.

A Virtual Processor Core can be considered as either of the following:

- a physical processor core, provided that the server is not partitioned for virtual machines
- a virtual core assigned to a virtual machine

**Licensee must obtain entitlement for each Virtual Processor Core made available to the Program.**

Scenario 1: Single virtual machine on a physical server

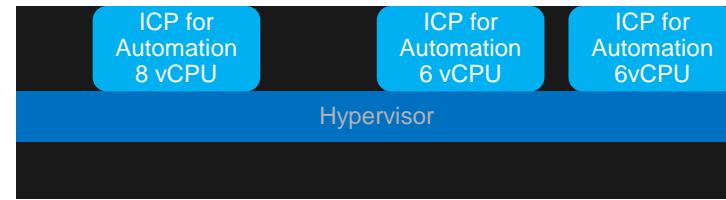


The physical server has 2 sockets, 8 cores per each socket, which gives 16 physical cores in total.

License Metric Tool reports the usage of 6 vCPU for ICP for Automation deployed on a virtual machine.

**VPC licenses required:** You must license only 6 VPCs, because that's the number of cores your application uses.

Scenario 2: Multiple virtual machines on a physical server

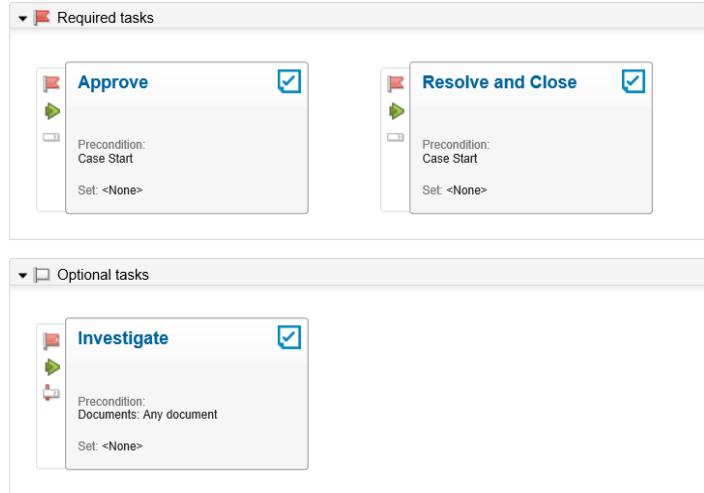
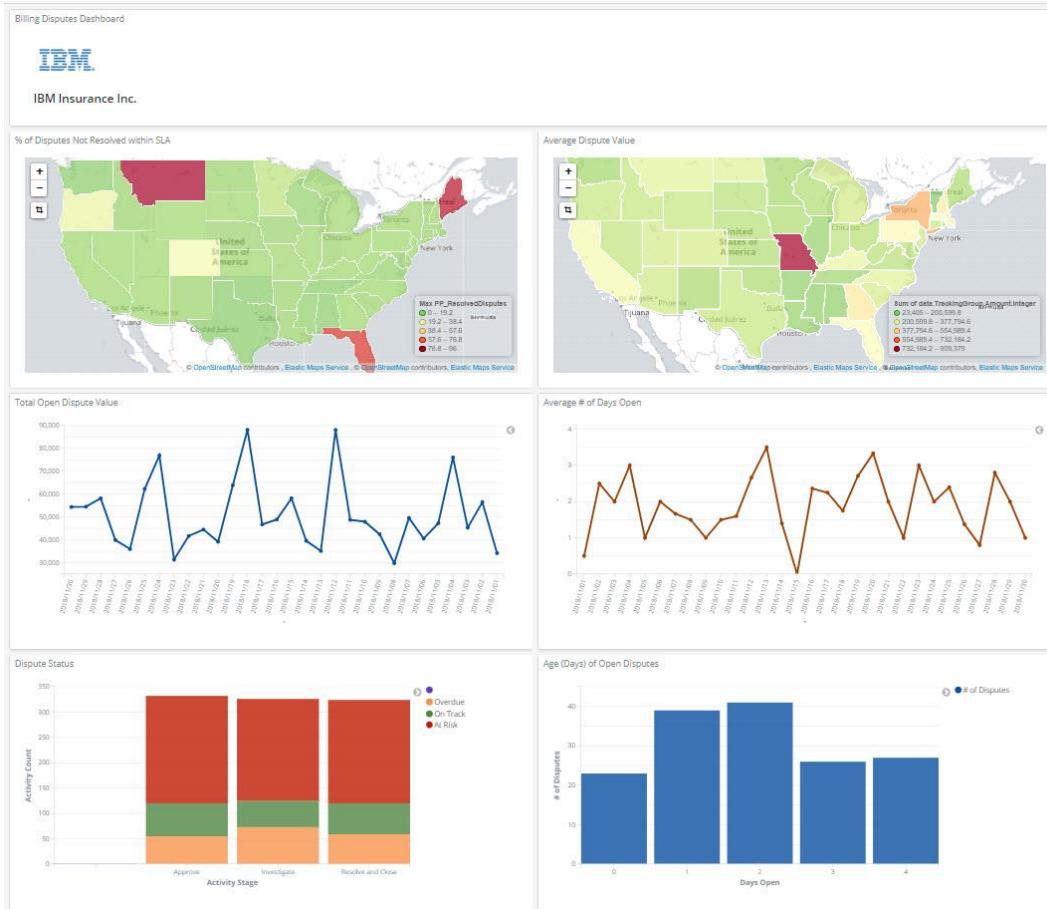


The physical server has 2 sockets, 8 cores per each socket, which gives 16 physical cores in total.

License Metric Tool reports the usage of 20 vCPU for ICP for Automation deployed on a virtual machine.

**VPC licenses required:** You must only license 16 VPCs, because that's the number of available cores.

# Lab 1: Billing Dispute Case Solution Dashboard



# Lab 2: Payment Processing Process Dashboard

Payment processing - High Value payments - Payments Status

Displays the details about the payments and their status

Payments - Activity and Status - vertical

Number of payments

27 1 4

1 to 100,000 - Count 100,000 to 1,000,000 - Count 1,000,000 to + - Count

Payments - instances

Table	Request	Response	Statistics		
Payment ID	Amount	Activity	Region	Branch	Count
24134	4,800	Manual Payment	EMEA	London	9
	2,000				1
2321	2,000	Manual Payment	AP	CBD	1
234344	2,000	Manual Payment	EMEA	London	1
234345	6,000	Manual Payment	EMEA	London	1

At Risk (red) On Track (blue) Overdue (orange) At Risk (red) Overdue (orange) null (yellow)

Payment processing - High Value payments - Payments Status

Displays the status of high value payments and if there are any delays

Currency and status

Number of payments

YEN GBP USD

data.Payment.Currency.string:keyword:Descending

Payments

data.Payment.Activity.string	data.Payment.Amount.decimal	data.Payment.Branch.string
Repair Payment	232,332	Tokyo
Manual Payment	6,000	London
Manual Payment	2,000	London
Manual Payment	2,000	Tokyo
Manual Payment	2,000	New York
Manual Payment	2,000	London

1-10 of 10

At Risk (red) On Track (blue) Overdue (orange) At Risk (red) Overdue (orange) null (yellow)

Payment processing - High Value payments - Payments Status

Displays the details about the payments and their status

Payments - Activity and Status - vertical

Number of payments

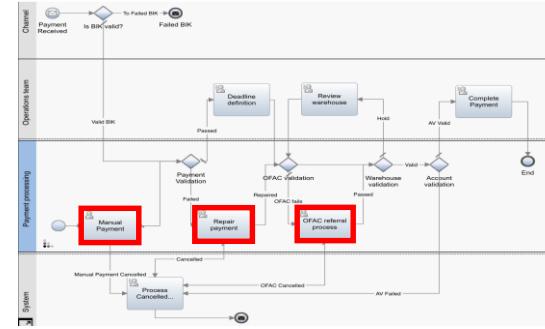
27 1 4

1 to 100,000 - Count 100,000 to 1,000,000 - Count 1,000,000 to + - Count

Payments - instances

Table	Request	Response	Statistics		
Payment ID	Amount	Activity	Region	Branch	Count
24134	4,800	Manual Payment	EMEA	London	9
	2,000				1
2321	2,000	Manual Payment	AP	CBD	1
234344	2,000	Manual Payment	EMEA	London	1
234345	6,000	Manual Payment	EMEA	London	1

At Risk (red) On Track (blue) Overdue (orange) At Risk (red) Overdue (orange) null (yellow)



# Fix for BPMN Regression Issue

- BAW 19.0.0.3 BPMN emitter introduced a regression which resulted in tasks not showing up in Kibana except for completed summaries.
- The fix will be delivered in BAW code change next quarter.
- However, there is a simple fix!
  - Add the following JVM argument to the server.xml:
    - -DuseBAIBridgeEvents=true
  - server.xml path:
    - <BAW install>\AppServer\profiles\<profile name>\config\cells\<cell name>\nodes\<node name>\servers\<server name>

```
<processDefinitions xmi:type="processexec:JavaProcessDef" xmi:id="JavaProcessDef_1576214456860" workingDirectory="${USER
<execution xmi:id="ProcessExecution_1576214456860" processPriority="20" runAsUser="" runAsGroup=""/>
<ioRedirect xmi:id="OutputRedirect_1576214456860" stdoutFilename="${SERVER_LOG_ROOT}/native_stdout.log" stderrFilename
<monitoringPolicy xmi:id="MonitoringPolicy_1576214456860" maximumStartupAttempts="3" pingInterval="60" pingTimeout="30
<jvmEntries xmi:id="JavaVirtualMachine_1576214456860" verboseModeClass="false" verboseModeGarbageCollection="true" ver
debugMode="false" debugArgs="-agentlib:jdwp=transport=dt_socket,server=y,suspend=n,address=7777" genericJvmArguments="
-Xgcpolicy:gencon -Xjit:iprofilerMemoryConsumptionLimit=67108864 -Xdump:stack:events=allocation,filter=#300m -Xgc:pref
-Declipse.bundle.setTCCL=false -DuseBAIBridgeEvents=true">
```

- Note, if you just want to use BAI with these four emitters (BPMN, BPEL, ODM, Case) already configured working and bug free just use my Blue Demos image: <https://bluedemos.com/show/3398>