Open Source Evaluation

The table below covers the features required for the portal. Also, adding the open-source possible for these features.

Business Case Calculation - Portal

Portal Features	Required for Portal?	Open Source	Feature Coverage (%)	Description
Log-in/IAM, Authentication	YES	NO		
Search	YES	YES	100%	Details
Search GARD	YES	YES	100%	Details
Dashboard/status	YES	YES	95%	Details
Dashboard/status GARD	YES	YES	95%	Details
Workflow Engine	YES	YES	95%	Details
Portal BE	YES	NO		
Portal BE / GARD	YES	NO		
API Gateway	YES	YES	100%	Details
Notifications	YES	NO		
Service Requests	NO			
Self Services	YES	NO		
Change Management	NO			
Problem Management	NO			
Incident Management	NO			
Test automation	NO			
Capacity management	NO			
Inventory management	NO			
Discovery	NO			
Asset management	NO			
Analytics	YES	YES	100%	Details
Analytics Input	YES	YES	100%	Details
Process automation	NO			
Monitoring	NO			
Reservation Management	NO			
Content Management System	YES	NO		

For Final recommendation of open source tools, click here.

Search

READY

	Apache Solr	Elasticsearch	Open Semantic Search

Key features	• Full-text search • Highlighting • Faceted search • Real-time indexing • Dynamic clustering • Database integration • NoSQL features and rich document handling (Word and PDF files, for example)	• Distributed search • Multi-tenancy • An analyzer chain • Analytical search • Grouping & aggregation	• Easy full text search • Multilingual semantic search engine • Monitoring: Alerts & Watchlists • Interactive filters • Monitoring: Alerts & Watchlists • Supports multiple data sources • Privacy and data protection
Supported Data Sources	Accepts data from different sources including XML files, CSV files, PDF and data extracted from tables in a database.	Accepts data from different sources such as ActiveMQ, AWS SQS, DynamoDB (Amazon NoSQL), FileSystem, Git, JDBC, JMS, Kafka, LDAP, MongoDB, neo4j, RabbitMQ, Redis, Solr, and Twitter.	Accepts data from different data sources like files and folders, file server, file shares, databases, websites, Content Management Systems, RSS-Feeds, etc. and connects using connectors and importers.
Open Source Community	Anyone can contribute to Solr, and new Solr developers or code committers are elected based on merit only.	• Driven by a single company rather than a whole community. • Anybody can contributor but final changes are confirmed and done by employees of Elastic. • Apache 2.0	• Everyone is allowed to read, to change and to extend the source code to the developers. • GNU license
Documentation	Very well documented	Highly documented	Very well documented
Scalability and clustering	SolrCloud available	Better scalability. Designed for cloud as well	Highly scalable up to Big data
API available	Yes	Yes	Yes
Additional Features available in the paid subscription	NA	• Single sign-on • Field- and document-level security • Graph exploration	NA
Limitations	• Complicated load balance and management. • Reconfiguration if the master is lost. • Increased latency.	• Doesn't support transactions and rollbacks. • Doesn't support SQL like joins. • Not real-time – eventual consistency.	• Initialization time. • Passed file system changes during system initialization.
Links	• https://lucene.apache.org/solr/ • https://lucene.apache.org/solr/resources.html#documentation	• https://www.elastic.co/ • https://www.elastic.co/guide/index.html	• https://www.opensemanticsearch.org/ • https://www.opensemanticsearch.org/doc

API Gateway

READY

Tool Name	**Kong**	**Express Gateway**	**TYK**	**Istio (Already in use by monitoring Squad)**
Key Features	• Deep Logging enabled • Support authentication • Add, remove or manipulate HTTP params and headers on-the-fly • Live monitoring for load and performance • Plugins available for extension • OAuth2.0 authentication available • IP-restriction	• Authentication and Authorization • API Management • Microservices • Supports various client authentication such as key auth,basic auth,jwt • Custom Plugins available • Consumer Management • Built-in Policies • Hot Reload and Restart	• API Management Dashboard • API Analytics • API Developer Portal • Flexible access protocols • Metered Access • Global or granular access control • Header Injections • Developer onboarding and enrollment	• Authentication • Authorization • API Management • Connect Microservices • Monitoring for load and performance • Ambiguous network failures handling and self-healing infrastructure.

Deployment Complexity	Single node	Flexible	Single node	Flexible
Data Stores Required	Cassandra or Postgres	Redis	Redis	-
Open Source Community	Yes, Apache 2.0	Yes, Apache 2.0	Yes, MPL	Yes, Apache 2.0
Core Technology	NGINX/Lua	Node.js Express	GoLang	Java
On Premise	Yes	Yes	Yes	Yes
Community/Extensions	Large	Small	Medium	Large
Authorization/API Keys	Yes	Yes	Yes	Yes
Rate Limiting	Yes	Yes	Yes	Yes
Data Transformation	HTTP	No	HTTP	HTTP
Integrated Billing	No	No	No	Yes
Links	• https://konghq.com/ • https://docs.konghq.com/	• https://www.express-gateway.io/ • https://www.express-gateway.io/docs/	• https://tyk.io/ • https://tyk.io/docs/	• https://istio.io/ • https://istio.io/docs/

Workflow Engine

READY

Tool Name	bonitasoft	JBPM	Camunda BPM
Features	• Built-in process engine • External Data supported • Graphical modeling • Management Dashboard • Data Manipulation • Data export	• Built-in process engine • External Data supported • Graphical modeling • Management Dashboard • Data Manipulation • Data reporting • Data export	• Built-in process engine • Dashboard for overview of running and historic operations (using Camunda cockpit) • Email alert system • External Data import supported • Business model import/export supported • Supports cloud based application
OS supported	• Windows • Linux • Unix • Mac OS • Mobile application	• Windows • Linux • Unix • Mac OS • Multi-platform application	• Windows • Linux • Unix • Mac OS • Multi-platform application
Plug-ins	YES	YES	YES
Integration capabilities	• MS SharePoint • API/Web Service • Database objects	• API/Web Service • Database objects	• API/Web Service • Database objects
Open Source Community	Low	Yes, Apache 2.0	Yes, Apache 2.0
Language	Java	Java	Java
Document repository	YES	NO	YES
Process flow animation	YES	YES	NO
Time/Cost estimation	YES	YES	NO
Process templates	YES	NO	NO
Monitoring Available	Yes	Yes	No
UI Availability	Yes, Support multiple browsers	Yes, Support multiple browsers	Yes, Support multiple browsers
Desktop Availability	YES	NO	YES

Limitation	• No alert system
	• No cloud availability
	• No reporting
		• No alert system
		• No cloud availability
			• No cloud availability
			• Paid edition features
			  • Reporting
			  • Graphical modeling
			  • Data export

Links	• https://www.bonitasoft.com/ • https://documentation.bonitasoft.com/bonita/7.9/	• https://www.jbpm.org/	• https://camunda.com/ • https://docs.camunda.org/manual/7.12/

Dashboard

READY

Tool Name	Grafana	Kibana	Dashing
Features	• Multiple visualization options, including heatmaps, histograms, and geomaps • Collaboration capabilities allows sharing of data and dashboards across teams • Built-in user control and authentication mechanism • An official library with hundreds of dashboards and plugins to use	• Flexible analytics and visualization platform • Intuitive interface for a variety of users • Real-time summary and charting of streaming data • Instant sharing and embedding of dashboards	• Huge library of widgets available • Can feed data to widgets directly through Ruby • Drag & drop interface for re-arranging widgets • Provides an HTTP API for pushing data to dashboard
Technology	GO, C, Javascript	Javascript	Ruby, Sinatra
Data sources and integrations	Graphite, Elasticsearch, Prometheus, InfluxDB, OpenTSDB, KairosDB, Cloudwatch	Work only with Elasticsearch	Can connect to multiple data source using open source plugins
Community	Large, Apache 2.0	Yes, Apache 2.0	Large, MIT license
Setup, installation and configuration	• Easy to install. • Support Linux, Mac, Windows, Docker	• Easy to install. • Support Linux, Mac, Windows, Docker	• Easy to install. • Ubuntu, Mac, Windows, Docker
Data manipulation	Data filtering, data sorting	Data manipulation possible	none
Access control and authentication	Built-in user control and authentication mechanism	• Accessible to public. • Access control using add-in e.g. X-pack	Yes, Built-in
Alerts	Built in alerting engine, alerts in form of email, Slack, PagerDuty, custom webhooks	• No Built in alert engine. • Can integrate with paid addins	Built in alert engine
Usage	Generic purpose dashboard, requires developers	To perform analytics and visualization of data	Simple Dashboards
Container technology	Available	Available	Available
Limitations	• No Personal storage for source data	• Cannot export a data table or saved search from a dashboard • Cannot perform aggregations across fields that contain nested objects.	• Does not work on IE • Data manipulation not possible
Documentation	Very well documented	Well documented	Very well documented
Links	• https://grafana.com/ • https://grafana.com/docs/	• https://www.elastic.co/products/kibana • https://www.elastic.co/guide/en/kibana/current/index.html	• http://dashing.io/ • https://github.com/Shopify/dashing/wiki

Analytics

READY

Tool Name	Apache Spark	H2O	KNIME	Kibana (Paid version)
Features	• Capable for large-scale data processing. • Real-Time Stream Processing • Support Multiple Languages • Integrated with Hadoop • In-Memory Computation in Spark	• enables quick development and sharing of the analytical model. • Readily available algorithms. • Ability to scale up horizontally by provisioning dynamic clusters.	• Offers excellent integration with a wide range of other open source software such as Python, R, Spark, and even ImageJ for image analysis. • Raw data is easily ingested in the tool, processed, can be performed statistics	• In built Alerting and Reporting • Advanced security features • Machine learning • Endpoint APIs available • Endpoint detection and response mapped to MITRE ATT&CK • (For data security) • Access to ECE & ECK orchestration features (Cloud and kubernetes)
Technology	Scala, Java, Python, R, SQL	Java	Java	Javascript
Data sources and integrations	Support multiple data sources such as Hadoop Distributed File System, NoSQL databases, or relational data stores like Apache Hive	Supports data from CSV files, ORC, SVMLight, ARFF, XLS, XLSX, Avro version 1.8.0, Parquet	Support multiple data sources like files, Databases, SQL, noSQL, Big Data, REST, simulation	Work only with Elasticsearch.
Community	Yes, Apache 2.0	Yes, Strong- Apache 2.0	Low, support forum	24/7/365 support, Apache 2.0
OS supported	Microsoft Windows, macOS, Linux	MAC, Linux, OS X, Windows, Ubuntu	Linux, OS X, Windows	Linux, Mac, Windows, Docker
Limitations	• No File Management System • Memory consumption is very high • The latency of Apache Spark is higher • lesser number of Algorithms	• This Drivereless AI is not a replacement for a data scientist. • Visual presentations can be improved further.	• Visualization can be improved further • Knowledge of R/Python is required to fully use the statistical analysis • Memory usage is a problematic issue sometimes	• Cannot export a data table or saved search from a dashboard • Cannot perform aggregations across fields that contain nested objects. • Limited support for multiple indices
Documentation	Well documented	Well documented	Well documented	Well documented
Links	• https://spark.apache.org/ • https://spark.apache.org/docs/latest/	• https://www.h2o.ai/ • http://docs.h2o.ai/	• https://www.knime.com/ • https://docs.knime.com/	• https://www.elastic.co/products/kibana • https://www.elastic.co/products/kibana/features • https://www.elastic.co/guide/en/kibana/current/index.html

Analytics Input

READY

Tool Name	Logstash	Graylog	StreamSets data Collector

Features	• Logstash allows user to create customized log processing pipeline • Can collect data from different sources and send to multiple destinations. • Create and configure own pipeline • Seamless integration with Elasticsearch, Beats and Kibana	• Built-in alerting • Logs Monitoring • LDAP Integration • Scalable Log Collection • Alerts & Triggers • REST API can be used to generate reports on the history data and streaming data.	• Design and execute data pipelines that are resilient to data drift • Early warning and actionable detection of outliers and anomalies • Timely delivery of consumption-ready data • Large number of built-in integrations with source and destination systems • User Authentication
Technology	Java	Java	Java, Javascript/Angular JS, D3.js, HTML and CSS
OS supported	All OS-x, Window, Red hat	Linux ,Ubuntu, Centos, Windows	Linux and Mac OS X
Data Base	Elastic search – document oriented DB	• Elasticsearch – to store the logs • MongoDB – storing metadata and dead letter queue messages	NA- It is connector between source data and destination.
Ease of Configuration	Need to set up Elasticsearch, Logstash, Kibana, beats with cluster configuration and medium complex to setup.	Need to set up Graylog – Graylog web server, Elasticsearch, MongoDB and medium complex to set up.	Need to install MapR libraries along with Data collector due to licensing restrictions
Protocol Used For Read / Write operations	REST API, HTTP API	REST API, HTTP API	Data transferred from source to destination in XML format,
Data Collection Intervals	Real time Stream and in Batches	Real time Stream and in Batches	We can create triggers to initiate the data collection pipeline.
Data Formats	Common Log files formats ex: nginx, tomcat logs common formats.	Common log files formats. Ex: nginx [error_logs, access_logs], Syslogs, rsyslogs, GLEF format	Supports multiple files format including compressed zip files, excel, CSV, etc.
Additional Data Inputs	HTTP,TCP, Scripted inputs, Syslog input	GLEF Kafka, GLEF HTTP, Beats, Message Inputs, rsyslog and syslog-ng	Hadoop systems, NoSQL databases, Search systems
Centralized Logging Support	Available with log shippers in JSON format	Available with Logs collector Graylog sidecar, Rsyslog in GLEF-Graylog Extended Log Format	Data can be shipped using the user defined pipelines
Customization / Extensibility	• Hosted Elastic Stack provides out-of-box user management features, via X-Pack. • Can extend / integrate with other tools via webhooks / plugins and programmatic access via REST HTTP/JSON API	• Flexible to add Custom dashboards based on stream using widgets, graphs, CLI stream dashboards. • Plugins are available in Graylog, to integrate with other tools • Provides user-management, LDAP and other authentication mechanisms (single-sign on/two-factor) integrations.	• Highly scalable • Built-in processor stages let us to write custom code in either python or JavaScript. • Destination stages can be updated into existing pipelines
Community	Yes, Apache 2.0	Yes, GNU	Yes, Apache 2.0
Limitations	• Reporting can be done using add-ins • Alerting Supported through X-pack, configurations can be created • Requires intensive management • Authentication and Alerting are paid features	• Daily limitation of 5 GB data • Integrations available to send alerts to AIOps tools • Paid version has below features: • Reporting • User Audit Logs • Views / Workflow	• Need managing and monitoring • No confirmation for reliability, delivery semantics and fault-tolerance. • No reporting
Documentation	Well Documented	Well Documented	Highly documented
Links	• https://www.elastic.co/products/logstash • https://www.elastic.co/guide/en/logstash/current/getting-started-with-logstash.html • https://github.com/elastic/logstash	• https://www.graylog.org/ • https://docs.graylog.org/en/3.1/ • https://github.com/Graylog2/graylog2-server	• https://streamsets.com/products/sdc • https://streamsets.com/documentation/datacollector/latest/help/index.html • https://github.com/streamsets/datacollector

Final Recommendation

Portal Feature	Feature Coverage	Tool recommended
Search	100%	Elasticsearch
Dashboard/status	100%	Grafana
Workflow Engine	95%	Camunda BPM
API Gateway	100%	Kong
Analytics (Open Source)	95%	Apache Spark
Analytics (Paid)	100%	Kibana
Analytics Input	100%	Logstash