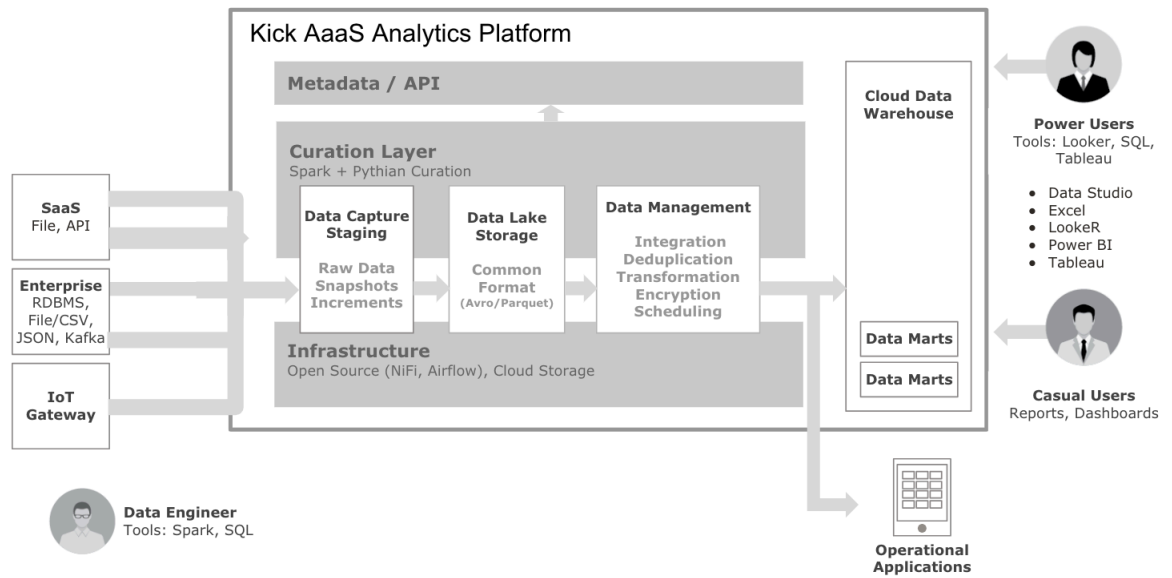


# KICK AaaS TECHNICAL PRIMER

The Pythian Kick Analytics as a Service (Kick AaaS) platform builds on available cloud services from Google Cloud Platform, Microsoft Azure and AWS, along with open source components and Pythian value-added software to automate processes for verifying, transforming and unifying complex, multi-source data on a single platform. It makes data analytics-ready and accessible to analytics users, data scientists, or other systems.



## KICK AaaS DATA HUB

The Kick AaaS Data Hub enables you to easily build large-scale and robust data processing pipelines.

- Flexible architecture with separate data ingestion, processing and warehouse layers
- A mix of best of breed open source and powerful cloud PaaS solutions: Apache Spark/Apache Beam, Apache NiFi, Airflow, Google BigQuery / Microsoft Azure SQL Data Warehouse/Amazon Redshift, Dataflow/Dataproc
- Fully automated secure infrastructure-as-code approach to deployment and updates using Terraform and Ansible
- Multi-cloud support
- Built-in monitoring

## KICK AaaS CURATION SOFTWARE

Kick AaaS Curation Software is a framework that allows you to quickly create and manage data pipelines running on the Kick AaaS Data Hub. Using a configuration file, you simply describe the data source(s) you want to add and Kick AaaS Curation Software will assemble, deploy and monitor the pipeline automatically. Kick AaaS Curation also comes with convenient tools to start, stop, remove and get information about existing pipelines. It features:

- A declarative data pipeline configuration language that lets you describe the source and the destination for your data pipeline. Kick AaaS will create and deploy an end-to-end data ingestion pipeline using Apache NiFi, Airflow (Cloud Composer support coming soon) and Dataflow or Spark on Dataproc
- Automated conversion of source schema into a destination schema with automatic data types conversion and validation.
- Conversion of various file formats (CSV, JSON, XML, etc) into a unified binary format and provides a convenient landing/staging/production data organization structure.
- Custom column-level transformations that you can apply on the fly using Spark SQL functions or custom UDFs
- Automatic scheduling of incremental data loads
- Automatic creation of pipeline SLA monitoring with a built-in alerting mechanism
- Support for both batch and real-time (coming soon) sources
- A Metadata REST API to track pipeline-related data such as which data sources are deployed, what schema they have (with versioning) as well as information about all ingestion and transformation activities. This supports use cases that answer questions like: When was the last time data was ingested for a given source? How much data arrived? Were there any issues with it? And more.
- Schema evolution support (coming soon)

## ABOUT PYTHIAN

Pythian is a global IT company that helps businesses leverage disruptive data technologies to better compete. Our services and software solutions unleash the power of cloud, data and analytics to drive better business outcomes. Our 20 years in data, commitment to hiring the best talent, and our deep technical and business expertise allow us to meet our client promise of using technology to deliver the best outcomes faster.

---

This document contains information and other content that is confidential and proprietary to The Pythian Group Inc. or its affiliates and may be protected under the law of copyright in Canada, the United States, and other jurisdictions. No part of this document may be modified, reproduced, distributed or disclosed in any form to any third party without written permission of The Pythian Group Inc.

PYTHIAN and LOVE YOUR DATA are registered trademarks of The Pythian Group Inc. All references herein to the corporate names, trade names, trademarks and service marks of third parties are intended to accurately identify such parties as the sources of specific products and services. No claim of association or license is intended or should be inferred by such use. This document is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.

© The Pythian Group Inc., 2018  
All Rights Reserved

## WORLDWIDE OFFICES

Ottawa, Canada  
New York City, USA

London, England  
Sydney, Australia

San Francisco, USA  
Hyderabad, India

**Pythian**  
love your data®