



PNDa.io: when big data and OSS collide

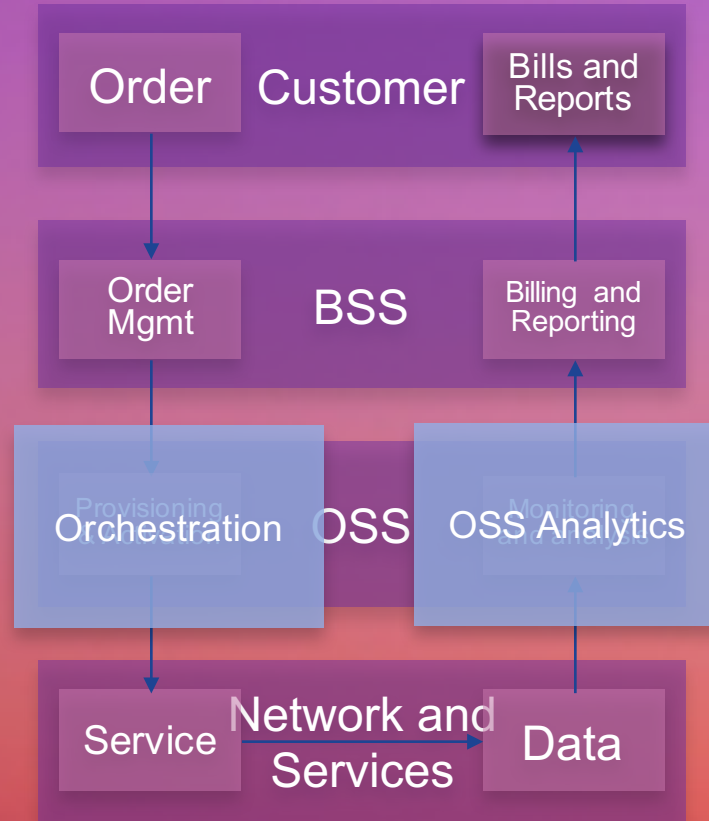
© 2016 PNDa a Linux Foundation Collaborative Project. All Rights Reserved.
Linux Foundation is a registered trademark of The Linux Foundation. Linux is a registered
trademark of Linus Torvalds.
Please see our [privacy policy](#) and [terms of use](#).



Simplified OSS / BSS Stack

Orchestration is responsible for service provisioning and pushes state to the infrastructure

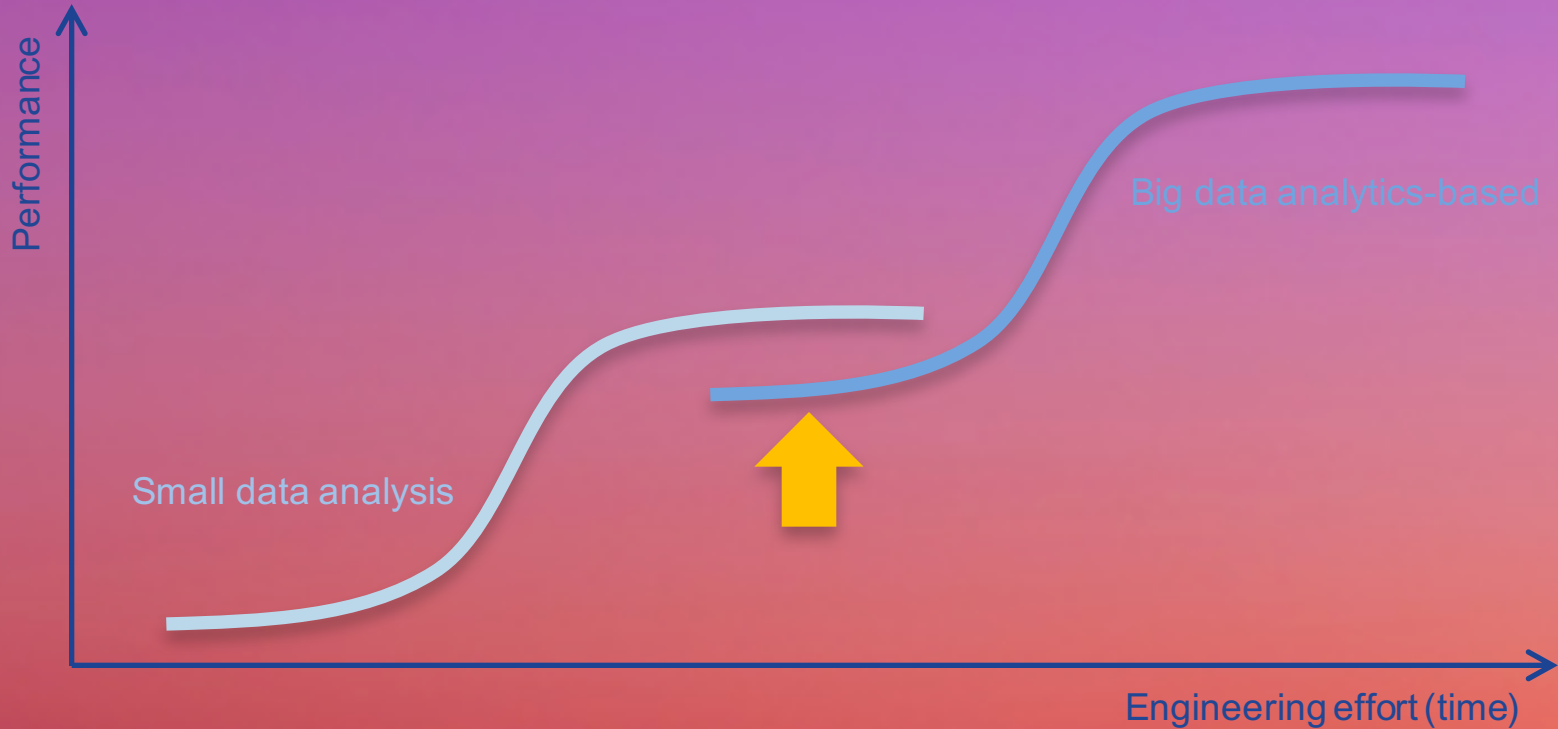
The “C” in FCAPS



OSS analytics is responsible for collecting data from the infrastructure, monitoring and analysis

The “F_APS” in FCAPS

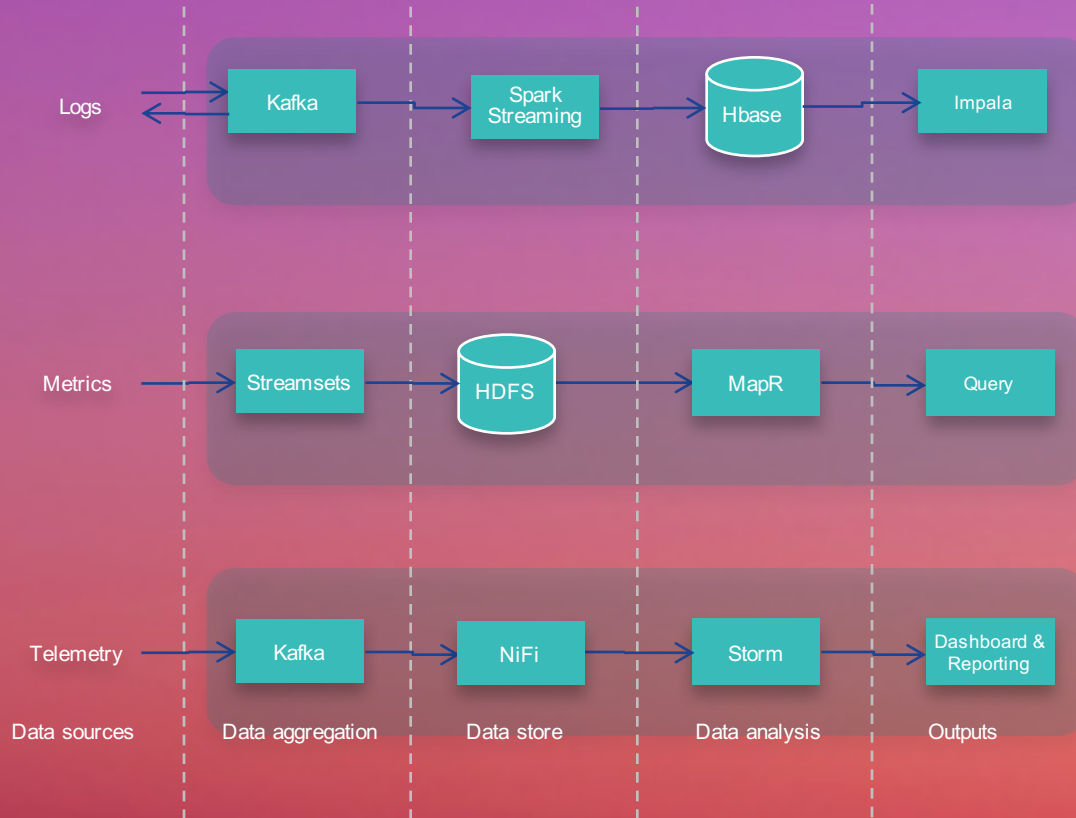
OSS Analytics is becoming a big data problem!



What changes?

	PMO	FMO
Orientation	Single domain	Cross domain
Realisation	Small data, tool driven	Big data, data driven
Data aggregation and analysis	Coupled	Decoupled
Domain Data Schema	Scheme-on-write	Schema-on-read
Analysis	Prescriptive	Prescriptive + Stochastic + ML
Customisation	Design time	Run time

Today's siloed analytics pipelines



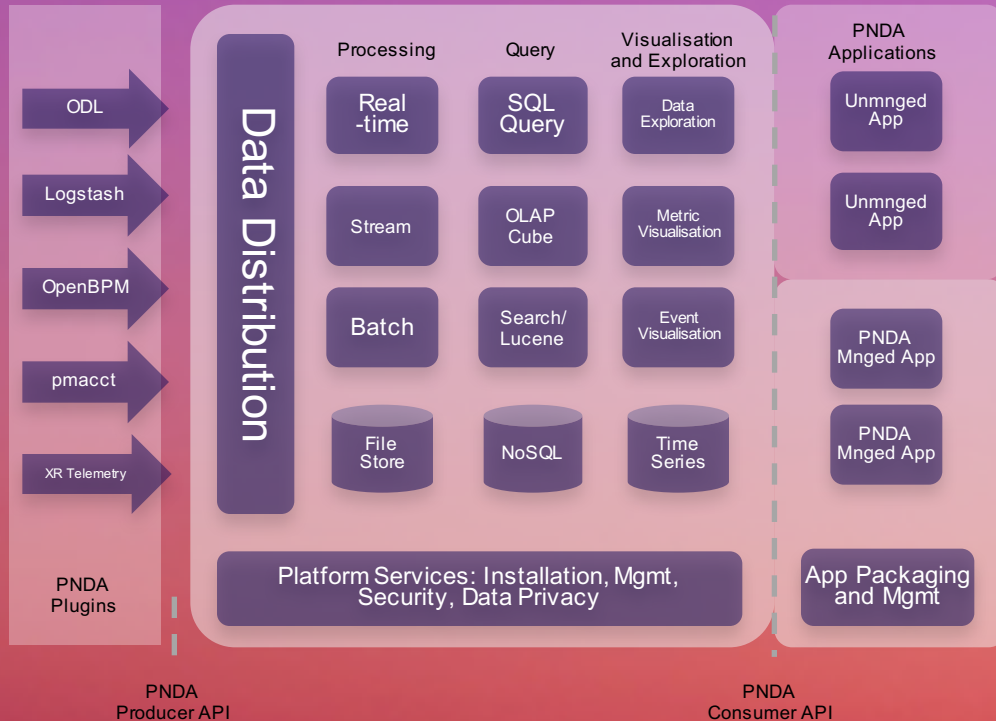
- Tight coupling of data aggregation/store/analysis
- Multiple analytics pipelines implemented from open source components
- Common design patterns ~75% of effort wasted / duplicated
- Siloes limit the potential of big data analytics and lead to industry divergence

What is PNDa?

PNDa brings together a number of open source technologies to provide a simple, scalable open big data analytics Platform for Network Data Analytics

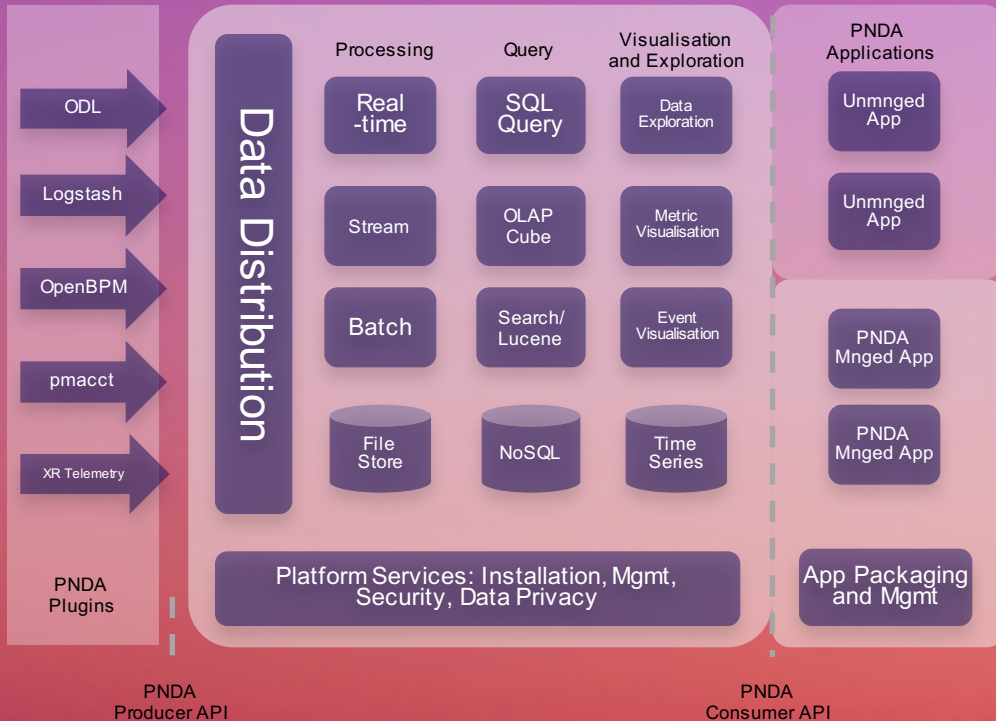
Linux Foundation Collaborative Project based on the Apache ecosystem

PNDA



- Simple, scalable open data platform
- Provides a common set of services for developing analytics applications
- Accelerates the process of developing big data analytics applications whilst significantly reducing the TCO
- PNDA provides a platform for convergence of network data analytics

PNDA



- Horizontally scalable platform for analytics and data processing applications
- Support for near-real-time stream processing and in-depth batch analysis on massive datasets
- PNDA decouples data aggregation from data analysis
- Consuming applications can be either platform apps developed for PNDA or client apps integrated with PNDA
- Client apps can use one of several structured query interfaces or consume streams directly.
- Leverages best current practise in big data analytics

Why PNDa?

There are a bewildering number of big data technologies out there, so how do you decide what to use?

We've evaluated and chosen the best tools, based on technical capability and community support.

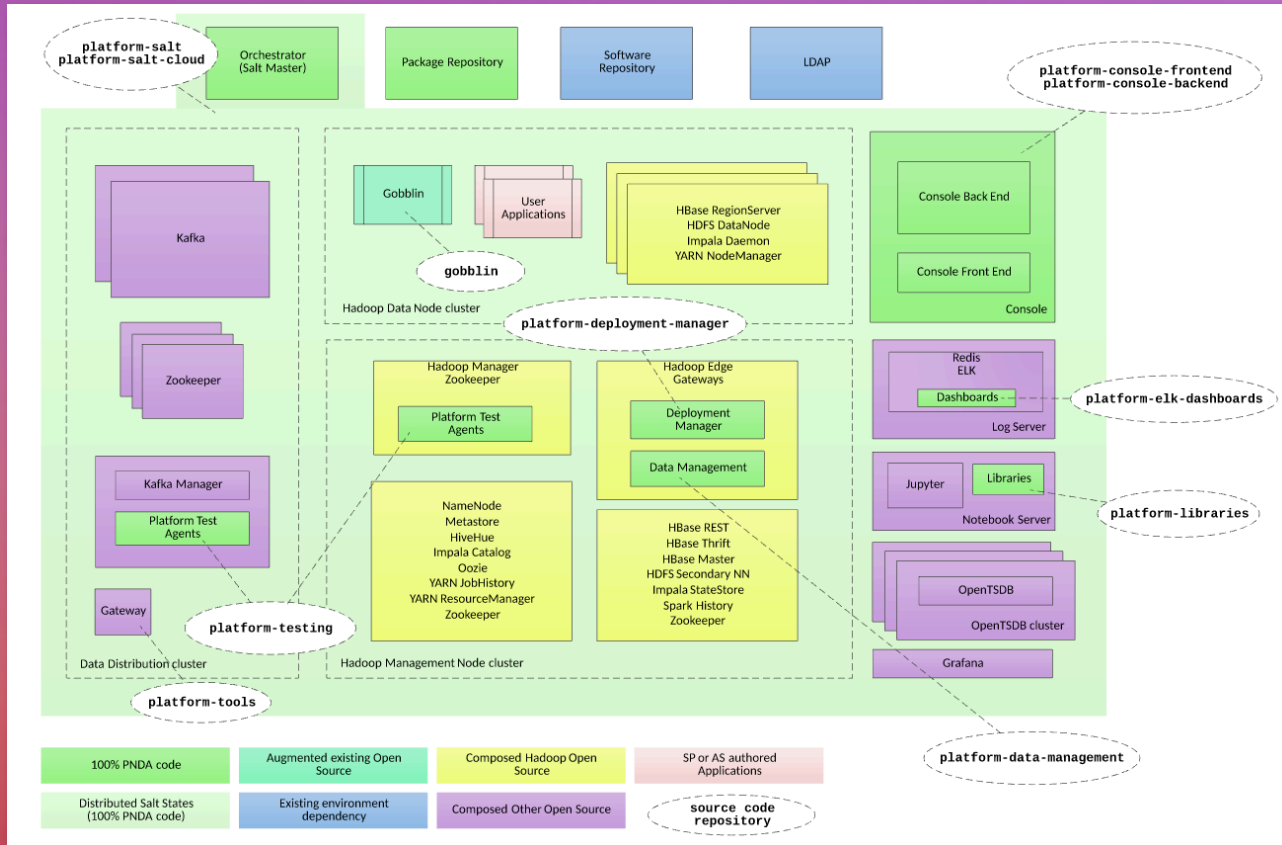
PNDa combines them to streamline the process of developing data processing applications.

Why PNDa?

Innovation in the big data space is extremely rapid, but combining multiple technologies into an end-to-end solution can be extremely complex and time-consuming

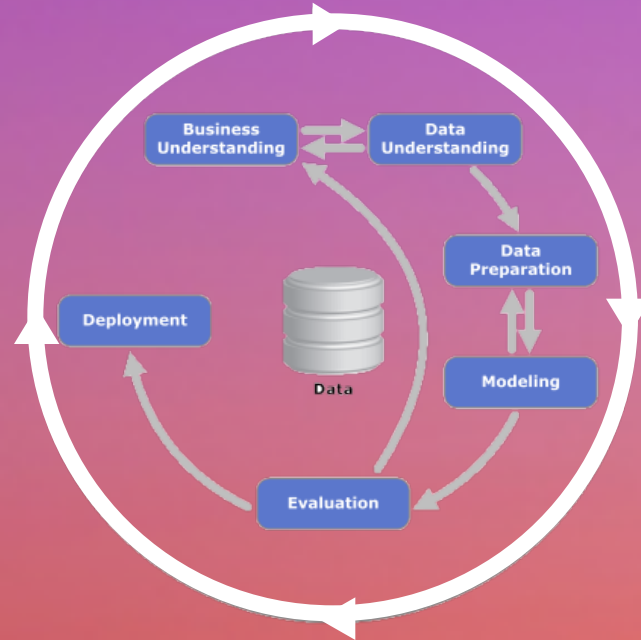
PNDa removes this complexity and allows you to focus on developing the analytics applications, not on developing the pipeline – significantly reducing the effort required and time-to-value


PNDA Software Components



Design time vs. runtime

 pnda standard



 jupyter

 pnda pico

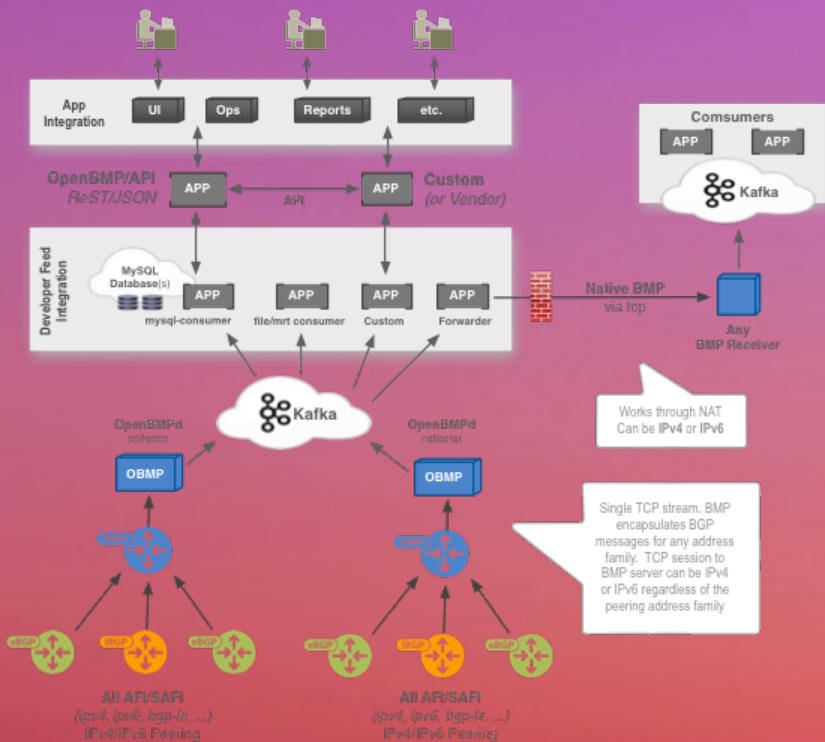


BGP meets 'Big-data'

Building the application

- Domain expertise
- Data science specialist
- Web 'full stack' specialist

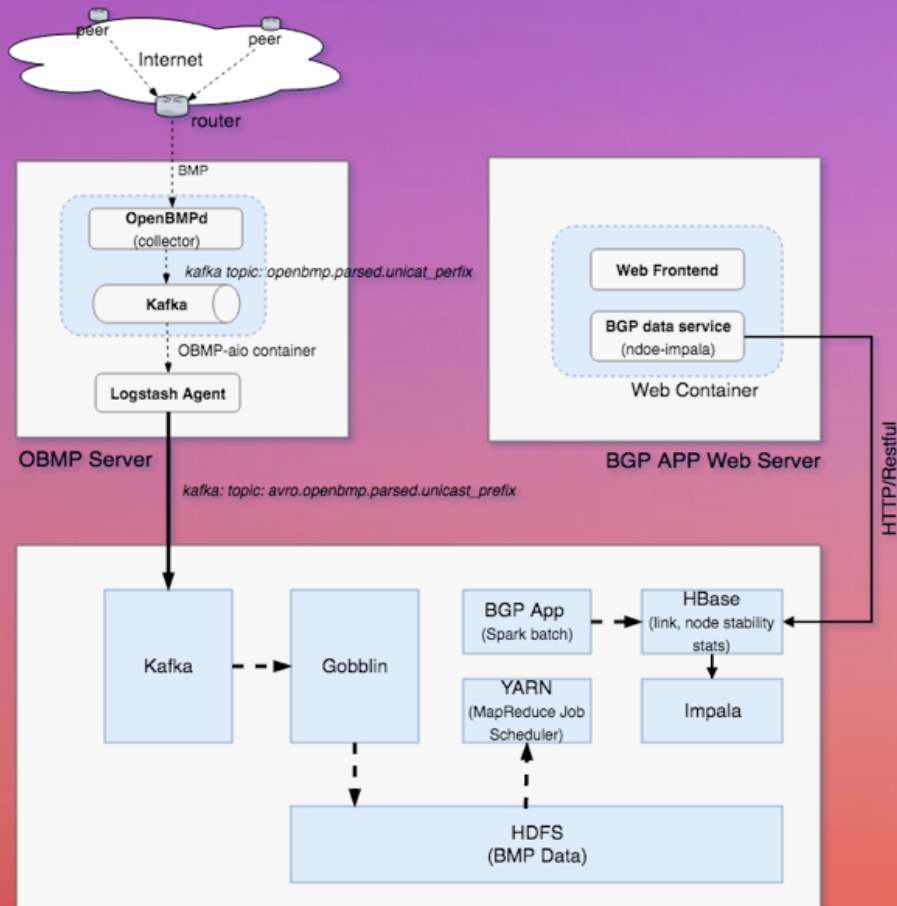
Architecture



- Needs BGP Speakers with BMP protocol support
- BMP session established between BGP Speakers and OBMP

Architecture

- Logstash required to perform 'avro' encoding of BMP data
- BGP App runs as Batch job, running periodically



What does this give us?

- OBMP gives us the ability to record the dynamics of the Internet
- PNDA platform enables
 - ‘Raw’ event recording capability, with horizontal scaling (HDFS)
 - Run analysis over large data-sets with parallelism
 - Ask questions of the aggregate data about the Internet
 - Drill down analysis
 - Per-prefix
 - Per-AS
 - Per AS-Path

Potential

- What can we do with large-scale collection of historical event information?
- Event impact analysis –
 - Stability
 - Security
 - Misconfiguration
- Application of ML/DL to data-set
- Pattern-detection and network ‘weather forecasting’



© 2016 PNDA a Linux Foundation Collaborative Project. All Rights Reserved.
Linux Foundation is a registered trademark of The Linux Foundation. Linux is a registered
trademark of Linus Torvalds.
Please see our [privacy policy](#) and [terms of use](#).

