



Globberry



Version 1.0 - March 2019

# Globberry Data Integration and Governance

SOLUTION FOR TELECOMS AND LARGE ENTERPRISE



## Introduction

---

While it is hard to find someone arguing that the data is the fuel, powering digital economy, the transition from traditional fragmented ETL towards more sophisticated data integration and governance across the enterprise is the real challenge.

The growing number of transactions and the sheer amount of data enterprises are dealing with is common knowledge. With the rapid adoption of IoT, every telco operator is currently processing or will be soon processing the amount of data, quite recently, only the largest banks and CSPs had to deal with.

While new challenges are mounting, the old ones which have, been with the industry for ages, are still not completely resolved. With your new shining microservice architecture, you still occasionally need to parse Excel or PDF, in order to combine disparate sets of data into meaningful information at the end. Plain old data sources inconsistency and data mapping still represent a major problem for many businesses.

CSPs have the problem of data flow management for a variety of structured and unstructured data sources covered with the traditional mediation solutions, based on expensive proprietary products. However, it is not clear how such platforms can adapt to the evolution of the service provider's BSS ecosystems in the foreseeable future, including not yet clear effects of IoT adoptions as well as slow but steady migration of BSS to the public cloud.



# Our Approach

---

The Globberry approach to data integration and governance solution is inspired by our belief that soon enough the cornerstone technologies and products used by the large organizations will be based on the open source foundations. Telecommunications are lagging behind in adoption of open source tech, compared to many other industries. However, there are simply less and less reasons to pay millions in capital expenditures upfront, before seeing any benefit, be vendor locked, and pay significant license fees each year, instead of the installing and using the equally functionally and performancewise capable open-source software for free. Although, the available open source solutions may lack certain components or interfaces specific for a given industry or business vertical. In the case of telecommunication, there is the number of formats which are industry standard and used ubiquitously, for example, ASN.1.

To put it simply, the Globberry approach is to supplement best-in-class open source products with the adapters, templates and specialized telecom expertise in order to receive functionally capable solutions with no CAPEX and an order of magnitude to lower cost of ownership.

## Solutions Overview

In order to create a robust enterprise data management solution to handle routing, transformation, and mediation of the large amounts of data, you need to consider three levels of data processing:

- Data Streams Management (from producers to consumers)
- Data Flows Management (between applications)
- Data Flows Orchestration (across the organization)

There are a unique sets of challenges to address on each level. At the same time, all the levels must be able to function within same real-time cadence and be able to scale up in the same pace and at the predictable cost, so that none of these levels becomes the bottleneck.

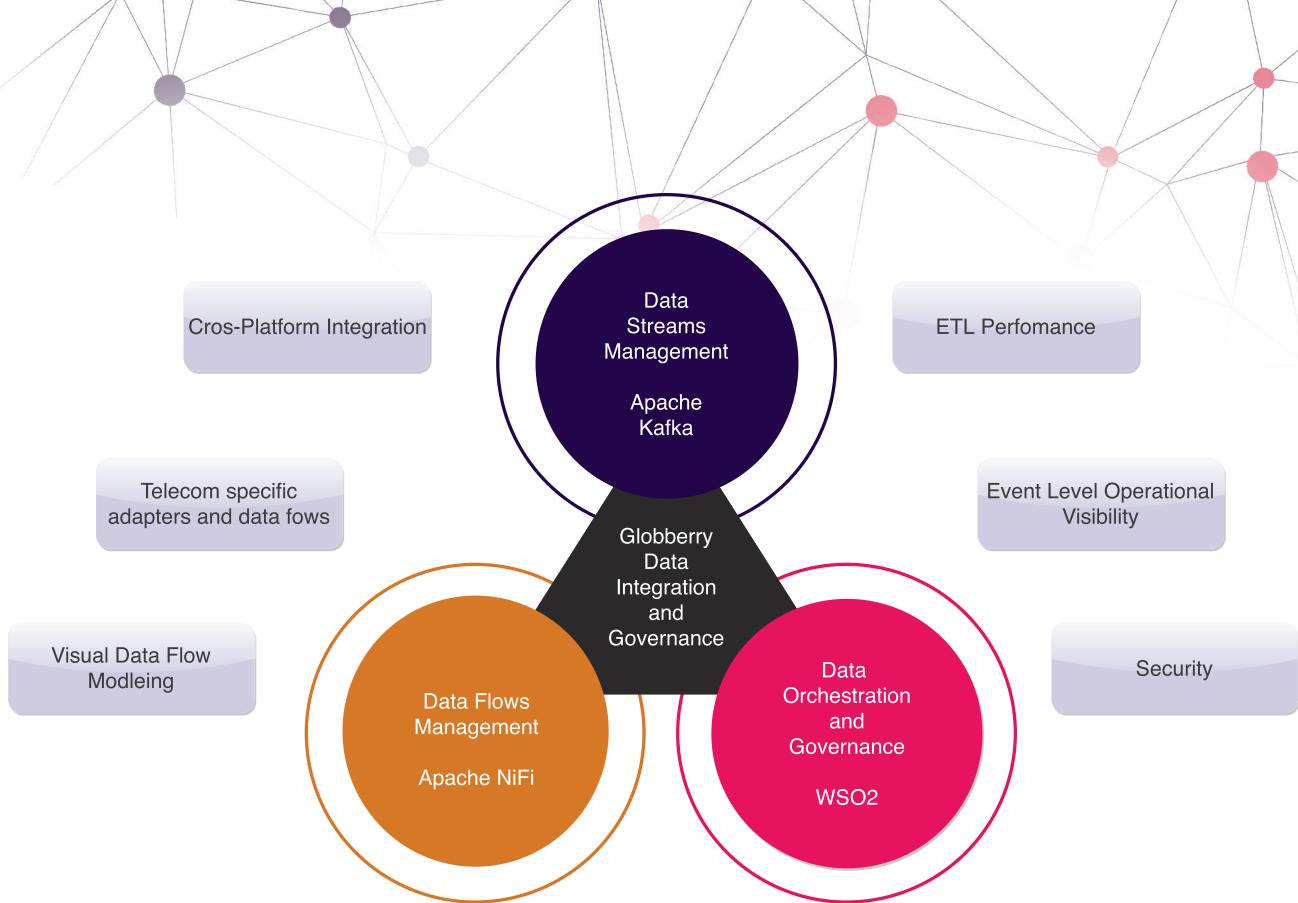


Figure1. Globberry Data Integration and Governance

In the core of our Data Management offering are three major open-source components covering each level:

- Apache Kafka for Data Streams Management
- Apache NiFi for Data Flows Management
- WSO2 Enterprise Integrator and API Manager for Data Orchestration

All the components selected are open-source projects frequently used in the telecommunications industry, but even more importantly, they are capable to scale and perform up to the level expected by Telecom Service Providers. Another criteria is that the selected products are pre-integrated and architecturally compliant. They don't functionally overlap so that they can encapsulate and isolate their level tasks from other system (e.g Stream processing from Integrations Management etc.)

In addition Globberry offers ASN.1 adapter in order to support processing of traditional CDR file formats across different components of the solution.



# Components

---

## Data Streams Management

Product for Data Streams Management shall be able to address following challenges:

- Real-time. Product must support very low latency in processing huge volumes of data in order to ensure real-time
- Reliability. It must support the management of distributed components
- Persistence. The tool must ensure data consistency and durability across decentralized endpoints

Apache Kafka is a high-performance open-source stream processing platform for collecting and processing large numbers of messages in real-time. It enables you to accept streaming data such as website click streams, events, transactions or other telemetry in real-time and at scale, and serve it downstream to stream processing applications, such as billing systems, reporting and analytics etc. Kafka is capable to process million of records per second with throughput measured in terabytes per second. The solution is the perfect fit to address the challenges in the area of data streams management the telco operators are facing. The following functions are making Kafka a perfect fit:

- Publish and subscribe to streams of records, similar to a message queue or enterprise messaging system
- Store streams of records in a fault-tolerant durable way
- Process streams of records as they occur in real-time

Project webpage: <https://kafka.apache.org/>



## Data Flows Management

Well defined data architecture heavily relies on good tools for data flow management in order to support a variety of architectural patterns (data federation, consolidation, separation in the context of the hybrid cloud).

Communication Service Providers' Data Flow Management solutions shall be able to confront its unique and relatively recent set of challenges:

- Time-to-availability. The system shall be able to synchronize all the sources producing data at different rates while consumers expect synchronous and immediate data availability
- Time-to-market. Solution is expected to provide the tools to efficiently adjust to the changing tech landscape and quickly incorporate new data producing and consuming systems
- Data variety. The system shall be able to deal with the number of data formats from many sources, including unstructured ones

Apache NiFi is a “post ETL” system for data processing and distribution. It supports powerful and scalable directed graphs of data routing, transformation, and system mediation logic. There are several features helping the system to stand out in comparison to traditional mediation platforms, making it viable to perform much wider set of task, than is traditionally expected from mediation platform. These Apache NiFi features are:

- Visual representation of global dataflow. Including visual configurability, traceability and real-time operational visibility down to individual events
- Scale out to handle gigabytes per second or scale down to handle tens of thousands of events per second
- While having all the traditional functionality such as data enrichment, splitting, aggregation, format conversion, schema transition etc.

Project webpage: <https://nifi.apache.org/>

## Enterprise Integrations, API Management and Data Orchestration

Data Streams Management and Data Flows Management is not enough to build the organization-wide comprehensive Data Integration infrastructure. You need an overarching tool to connect numerous data silos across organizations and make the company's informational resources accessible to partners and customers. Such system shall be capable to deal with:

- Uncertainty. In the complex technical environment, the formats and structures of API- and services-based data sources are subject to unannounced changes
- Complexity, which manifest itself in Variety of IT architecture patterns in siloed data islands. Starting from legacy systems, to public and hybrid clouds
- Data monetization. System shall support the transformation of company's data into valuable resource, generating revenue
- Security. System shall ensure the holistic approach to the IT security, starting from support of the necessary authentication methods, down to centralized security policies **monitoring, auditing and access control**

WSO2 Enterprise Integrator is built to address the listed above Integration challenges. It provides a centralized integration middleware/ESB with data integration, process integration and B2B integration capabilities. In addition to short-running, stateless integration flows, it can be used to manage long-running, stateful business processes. It also comes with analytics for comprehensive monitoring, message brokering capabilities that can be used for reliable messaging, and capabilities to run microservices for your integration flows.

WSO2 API Manager is an open source approach that addresses full API lifecycle management, monetization, and policy enforcement. It allows extensibility and customization, and ensures freedom from lock-in.

Both products have large footprints in particular in the telecom industry, due to its flexibility and low cost of ownership.

Platform webpage: <https://wso2.com/>

### ASN.1 Adapter

ASN.1 is the de-facto standard for CDR files. In order to enable usage of the open-source stack within BSS ecosystem we supplemented our Data Flow Management with ASN.1 converter processor. It not just makes it possible to use our open-source based solution for traditional mediation, but widens the capabilities of the mediation solution in terms of performance, variety of data sources and complexity of transformations.

# Use Cases

Offered solution can cover the following use cases in telecom and large enterprise:

- **Traditional off-line mediation.** Although every telecom operator has its mediation solution in place, they are rigid and slow to configure. Offered stack of products can complement solutions already in place adding another dimension of flexibility and configurability to them
- **Emerging IoT data orchestration.** Connecting multitude of changing data sources and formats into comprehensive system under the centralized management
- **Traditional ETL solutions alternative at lower cost**
- **Data preparation and transfer** into real-time analytics, reporting, consolidation platforms. Large volumes are delivered at the low latency to enable real-time data availability

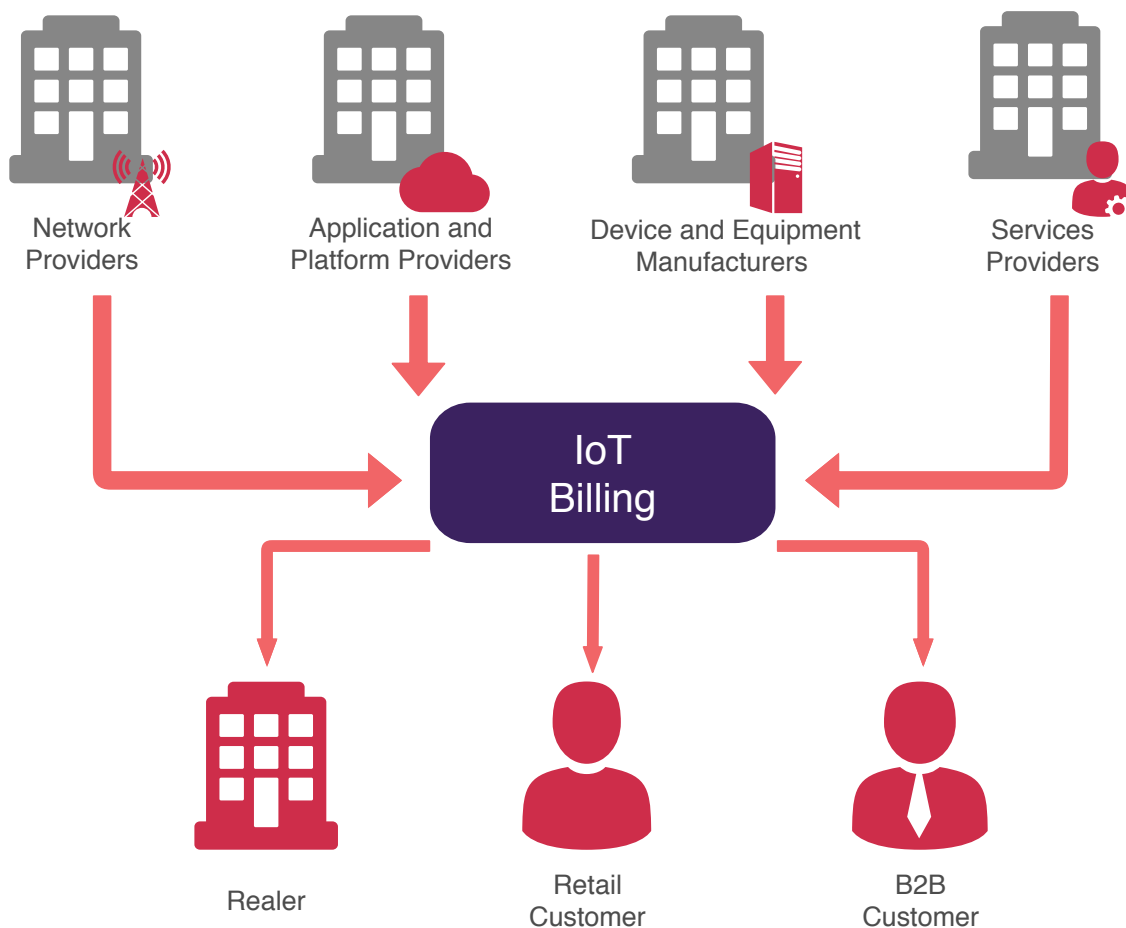


Figure2. IoT Billing Ecosystem





## Solution Benefits

- **CAPEX reduction.** All the components are built on top of open-source frameworks to reduce CAPEX by avoiding expensive licenses, meanwhile leveraging the power of the most innovative solutions produced by the industry. No upfront capital investment required
- **Cloud Native.** The solution consists of cloud-native products providing flexibility in deployment options from bare-metal to public cloud
- **Horizontal scalability.** The solution can support both huge volumes and small trial deployments. The hardware requirements grow linearly with the transaction volume increase
- **Proven:** All the products in the solution are leaders in their areas with large footprints across different industries including telecommunications
- **Multi-purpose.** Solution can be used for number of use cases in Enterprise data integration and governance, far exceeding traditional mediation
- **Short monetization cycle.** Simple configuration for charging, policy, customers, and data management components to achieve rapid services launch and monetization

## Our Offering

Globberry offers specialized telecom expertise in building and supporting mission critical BSS applications, including services in the domain of Data Integration and Governance with the products we mentioned as part of our solution.

### Consulting

Evaluation of existing solutions and target architecture of Data Integration and Governance stack definition.

### Implementation

Full cycle implementation from solution design to user acceptance testing, integration and data migration services.

### Support

Operations and maintenance, 24/7 support, second level and product support.

In addition to configuration of flows, components and integrations, we offer custom processors for Data Streaming (such as ASN.1) in order to cover major use cases.



[globberry.com](http://globberry.com)

[info@globberry.com](mailto:info@globberry.com)