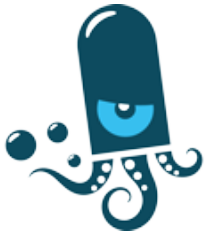


pmacct and network analytics

Paolo Lucente

pmacct

FRNOG 31, Paris – Oct 2018

whoami

Paolo Lucente

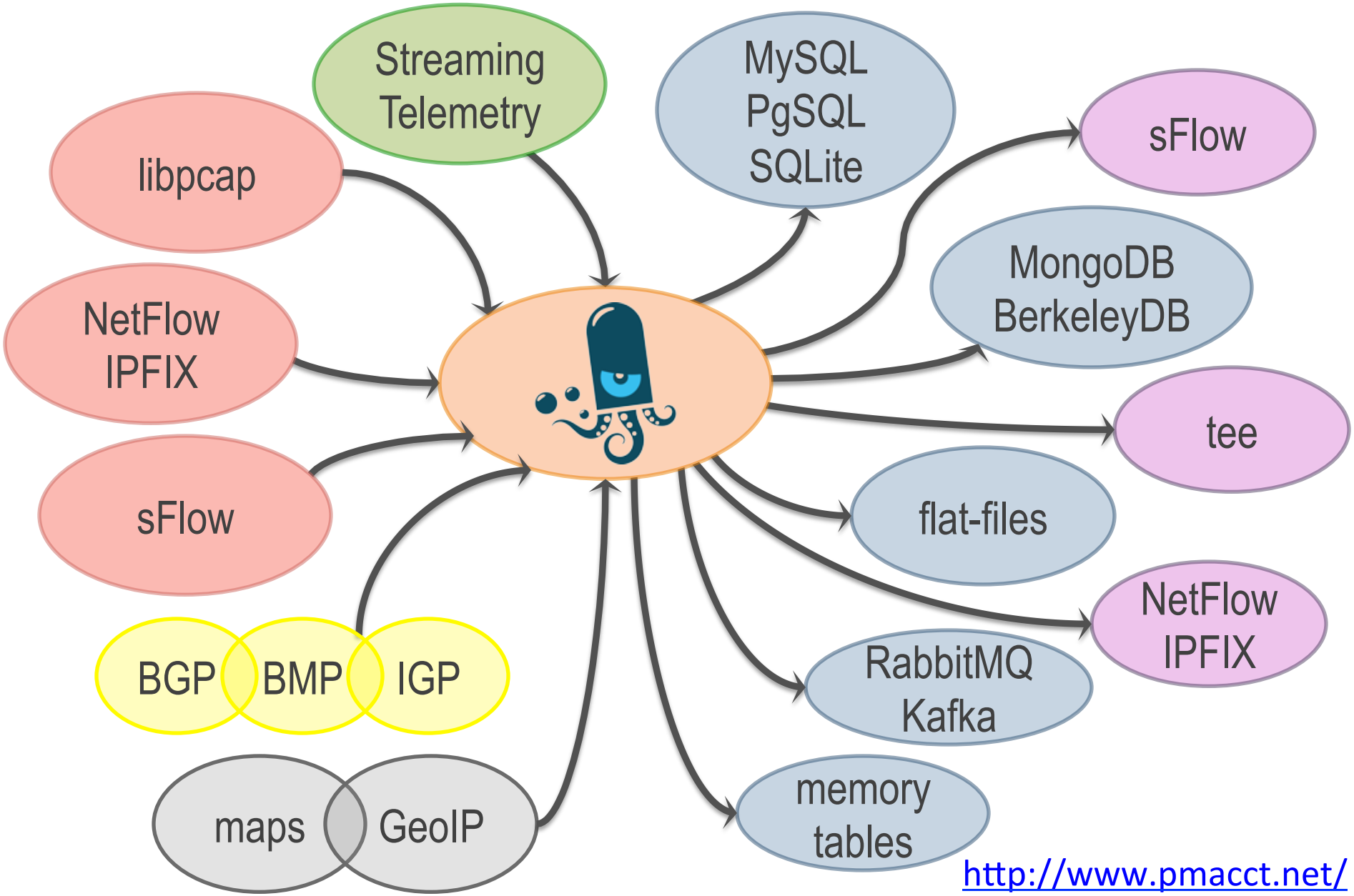
GitHub: paololucente

LinkedIn: plucente

Digging data out of networks worldwide for fun and profit for more than 10 years

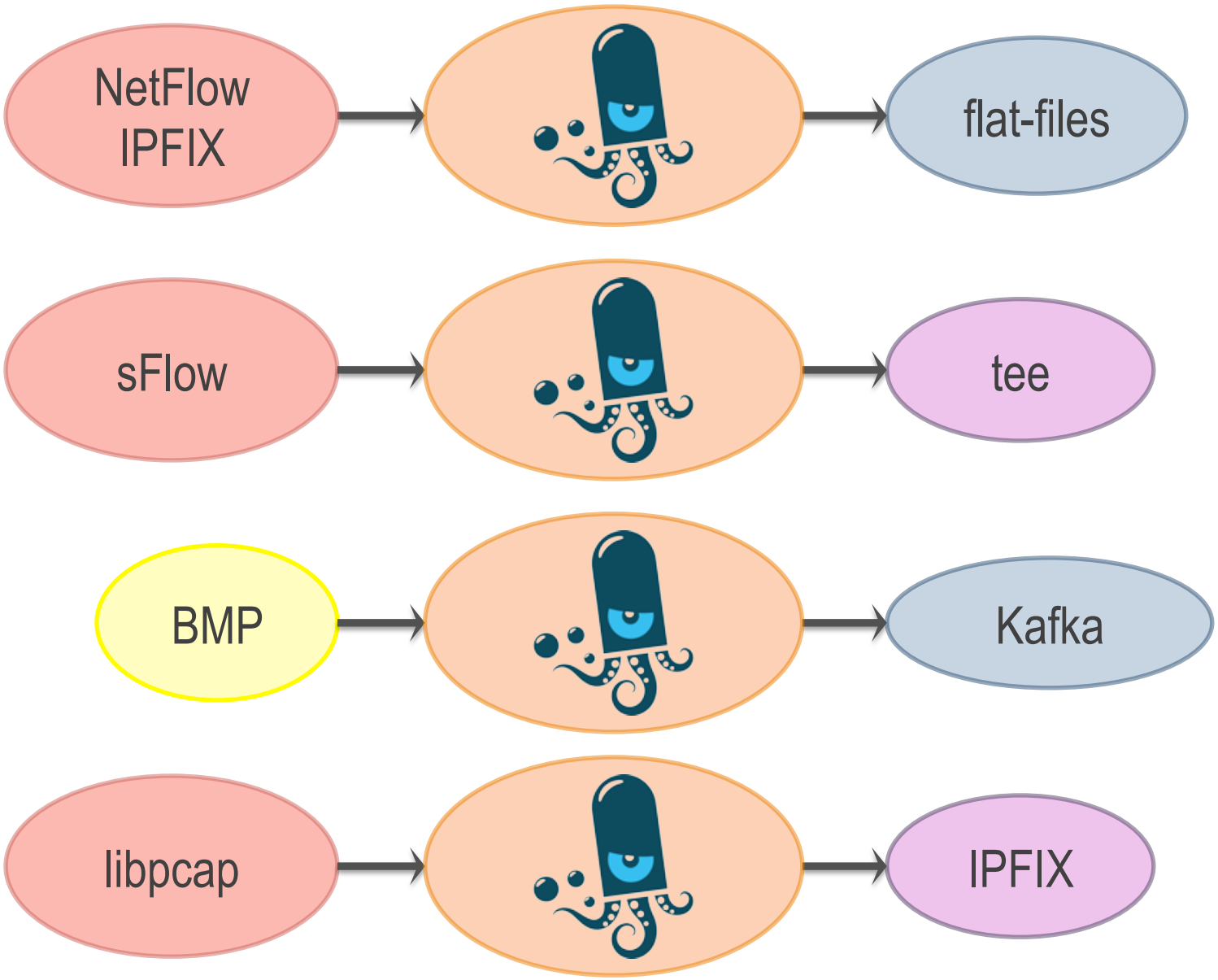
Introduction

pmacct is open-source, free, GPL'ed software

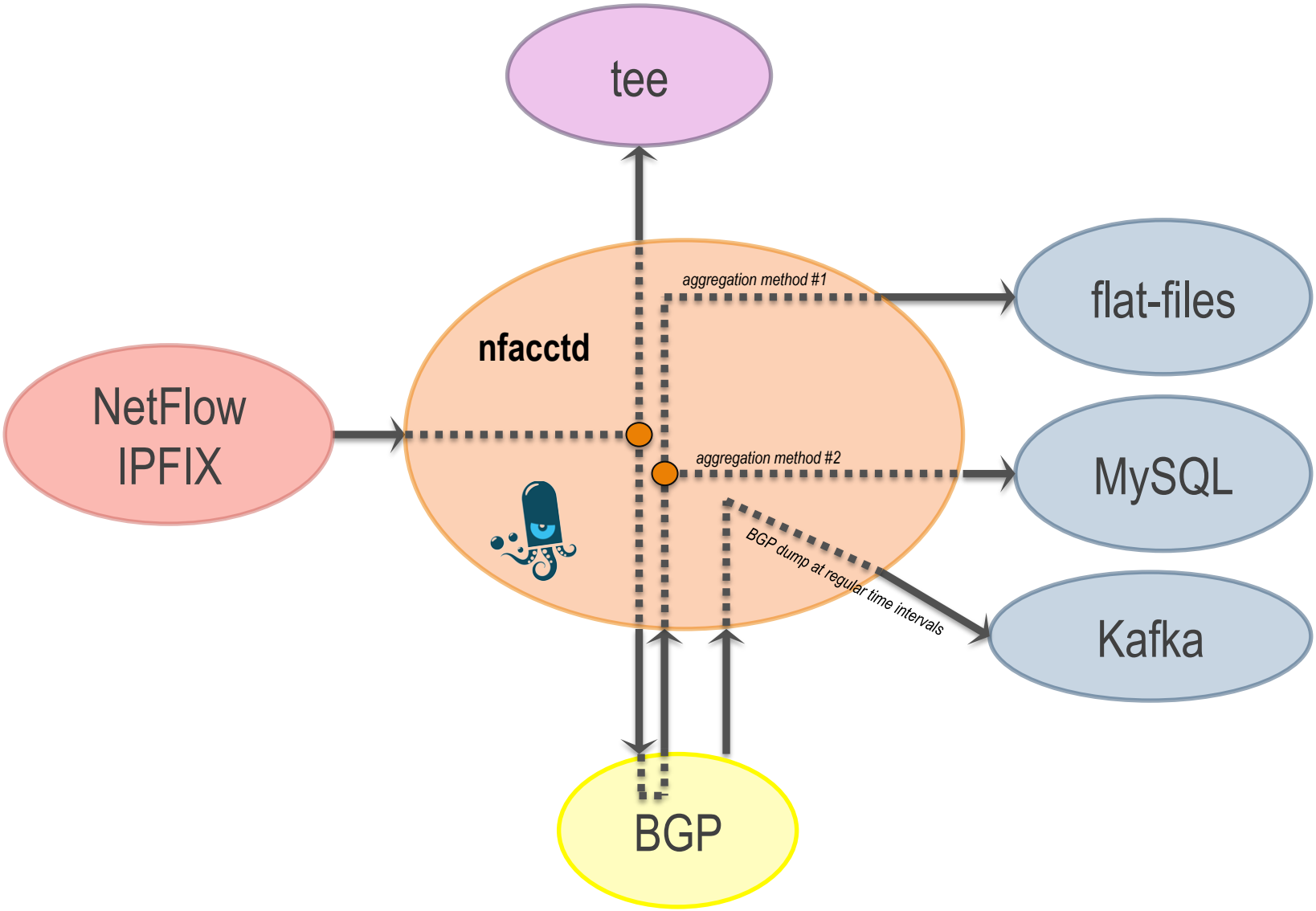


http://www.pmacct.net/

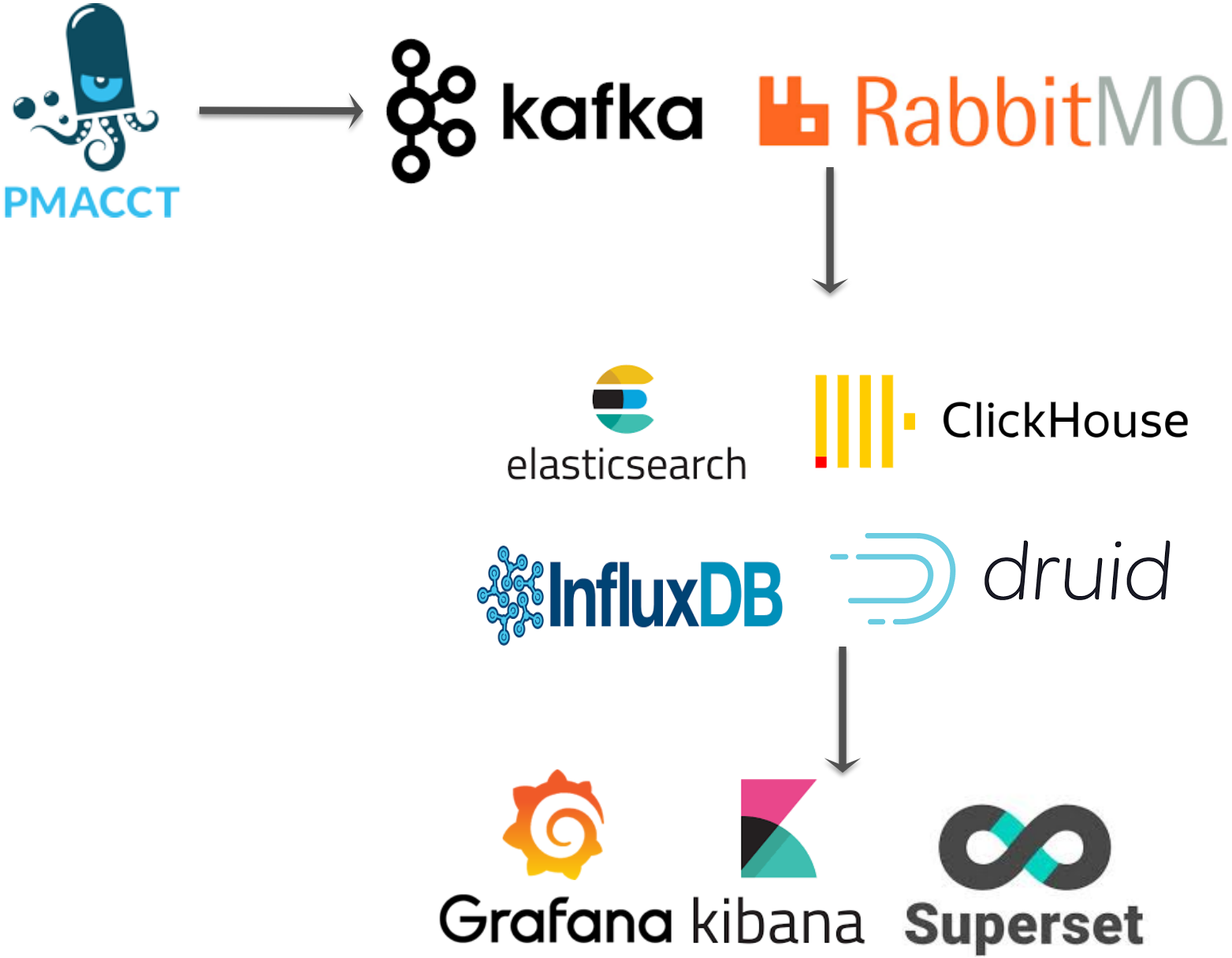
pmacct: a few simple use-cases



pmacct: a slightly more complex use-case



The use-case for message brokers



Key pmacct non-technical facts

- 15+ years old project
- Can’t spell the name after the second drink
- Free, open-source, independent
- Under active development
- Innovation being introduced
- Well deployed around, also in large SPs/IXPs
- Close to the SP/IXP community needs

Building a Network Analytics pipeline

FRNOG 31, Paris – Oct 2018

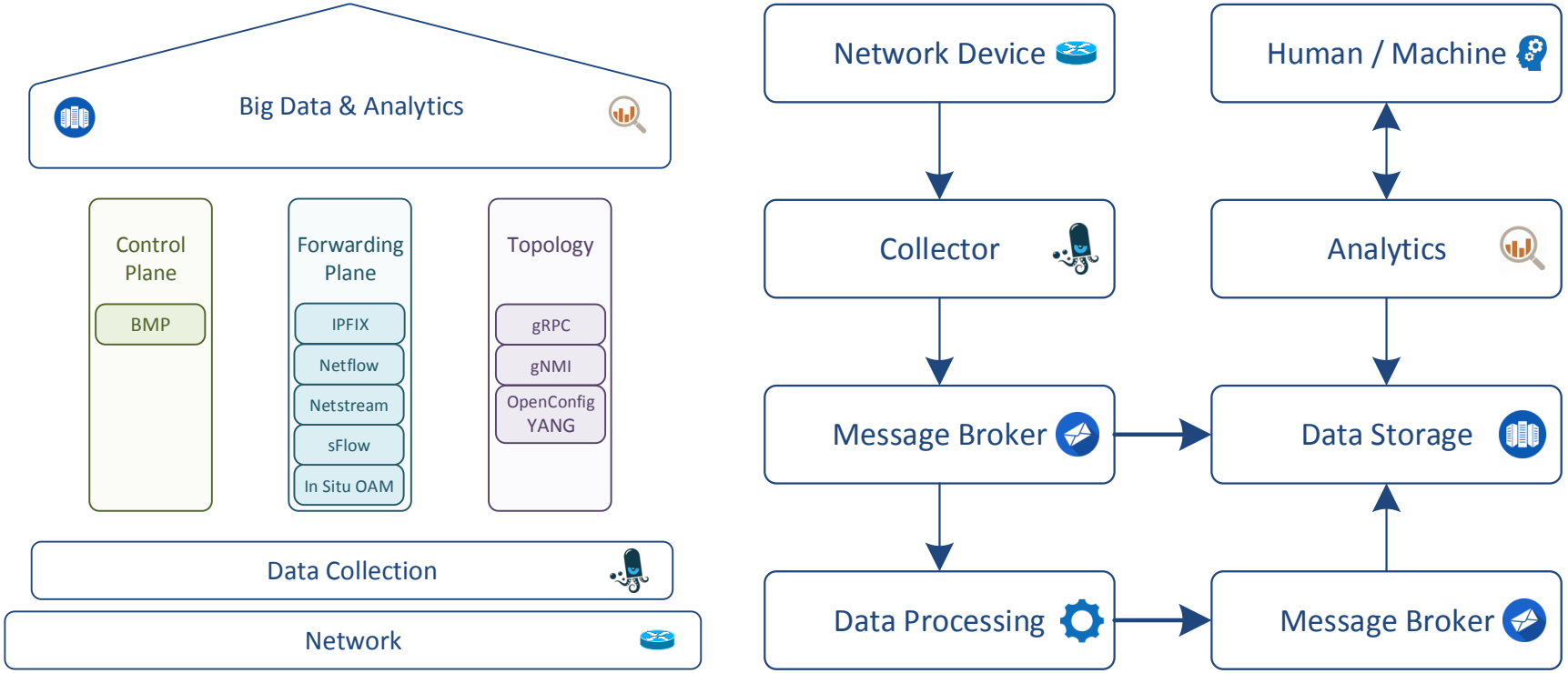
Typical goals for Network Analytics

- Business Intelligence
- Insight in traffic patterns
- Support peering decisions
- Investigation of network events
- Capacity planning
- Traffic Engineering

Sample pipeline for Network Analytics

- Input data (BGP, NetFlow, Streaming Telemetry, SNMP, …)
- Collection (pmacct, homegrown SNMP poller)
- Data encoding (JSON, Apache Avro, etc.)
- Distribution (Kafka)
- Enrichment (homegrown glue in $language)
- Ingestion (RDBMS, TSDB)
- Visualization

Sample pipeline for Network Analytics (cont.d)



Credits to: T. Graf (Swisscom) @ UBBF 2018

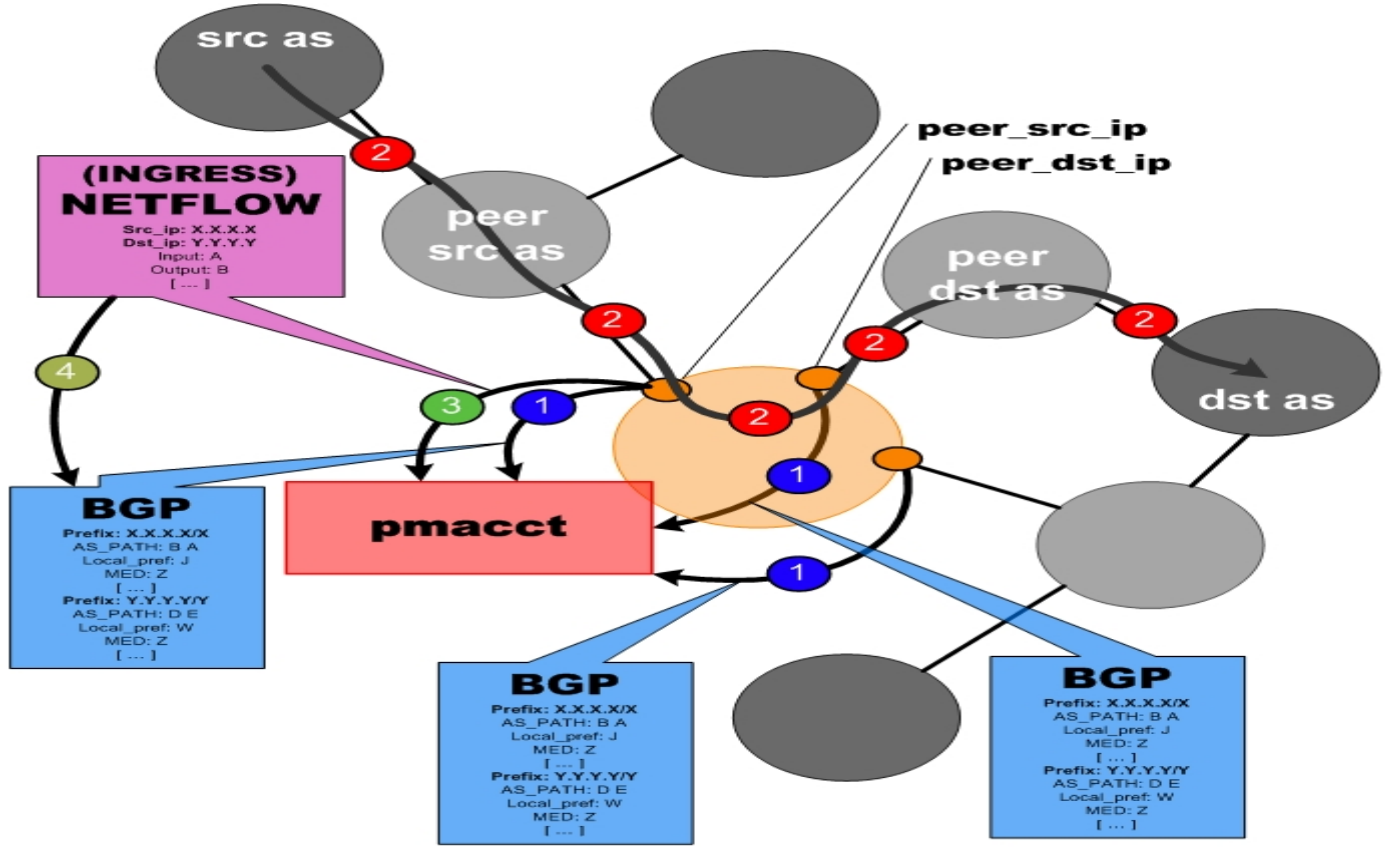
Getting BGP to the collector

- Let pmacct collector BGP peer with all PE devices: facing peers, transit and customers
  - No best-path computation at the collector: scalability preferred to optimizing memory usage
  - Count some 50MB of memory per full-routing table
- Set the collector as iBGP peer at the PE devices:
  - Configure it as a RR client
  - Collector acts as iBGP peer across (sub-)AS boundaries

Getting flow telemetry to the collector

- Export ingress-only measurements at all PE devices: facing peers, transit and customers.
  - Traffic is routed to destination, so plenty of information on where it's going to
  - It's crucial instead to get as much as possible about where traffic is coming from, ie.:
    - input interface at ingress router
    - source MAC address
- Perform data reduction at the PE (ie. sampling)

Forwarding plane/control plane correlation

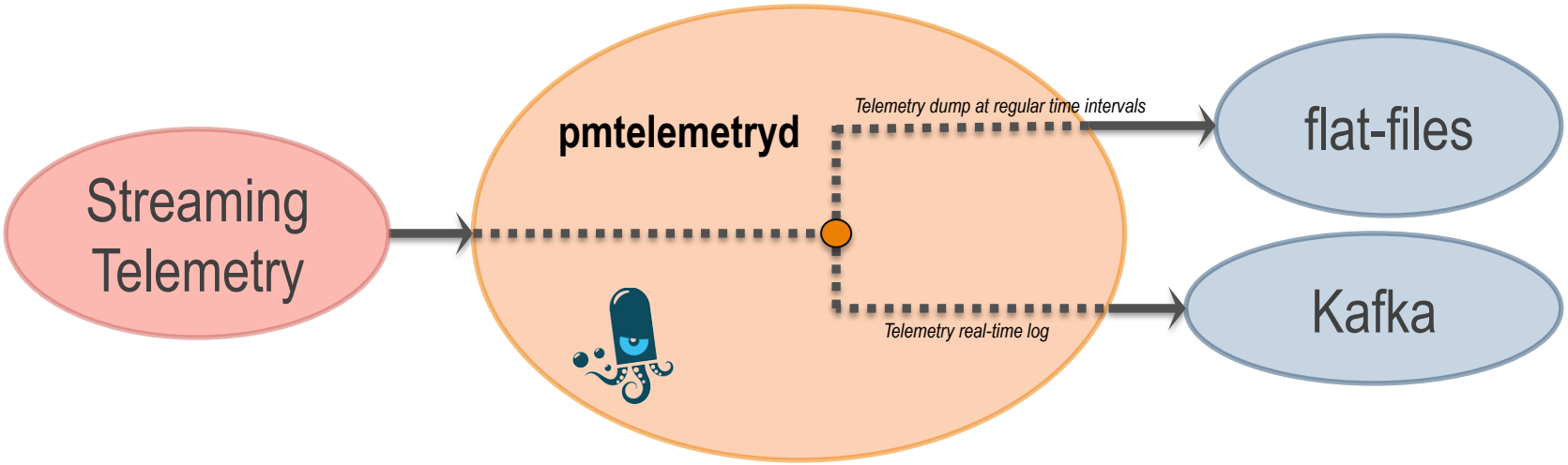


1. Edge routers send full BGP tables to pmacct
2. Traffic flows
3. NetFlow records are sent to pmacct
4. pmacct looks up BGP information: NF src addr == BGP src addr

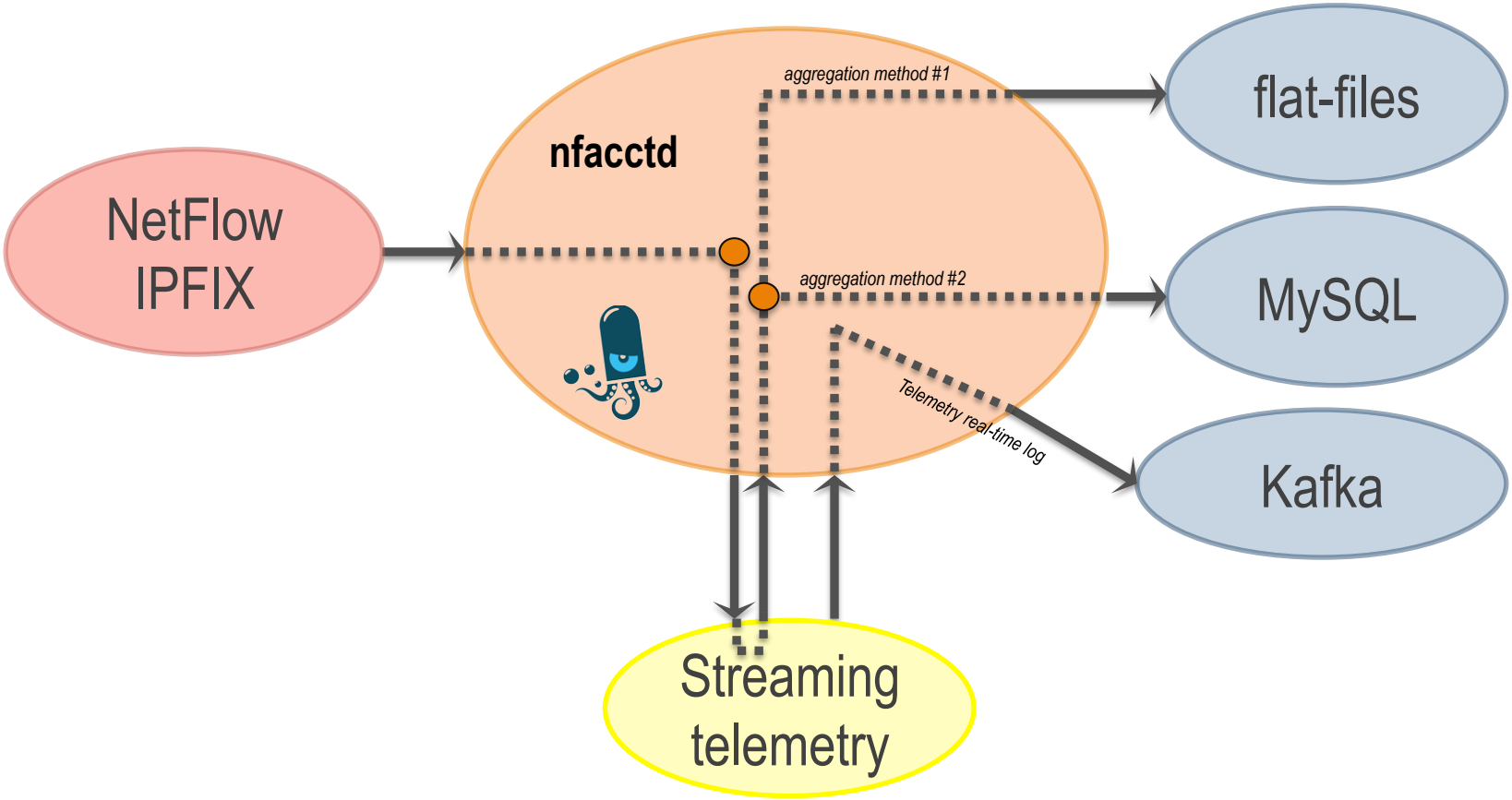
Streaming Telemetry

- A scalable replacement for SNMP:
  - Push technology
  - Subscribing to data of interest
- A long journey to standardization ahead:
  - Models: Openconfig and vendor-specific
  - Transport: traditional, Netconf and gNMI
  - RPC: Netconf (YANG Push) and gNMI
  - Encoding: JSON and GPB

pmacct & Streaming Telemetry (1/2)



pmacct & Streaming Telemetry (2/2)



Data Encoding

- JSON
  - Schemaless
  - Can be compressed successfully end-to-end
  - Simple, easy to troubleshoot and debug
  - Often that is the encoding supported at ingestion time
  - Similars: BSON, MsgPack
- Apache Avro
  - With schema
  - Binary format (when things go wrong ..)
  - Similars: Thrift, GPB, Capt'n Proto

Distribution

- Kafka: de-facto standard for data shipping
  - Easy to model different producer-consumer architectures
  - pmacct has a plugin to produce to Kafka
  - Most TSDBs can consume from Kafka
- People in the need for raw data can tap into this layer to consume directly
- Intuitive (that does not mean straightforward ..) to scale-out, balance and replicate

Storing data persistently

- If your company runs a Big Data shop, you may want to stick to one of their options
- As you may very well be trying to ingest millions of tuples per minute:
  - If in house, discuss dimensioning
  - If in cloud, think about costs and data privacy

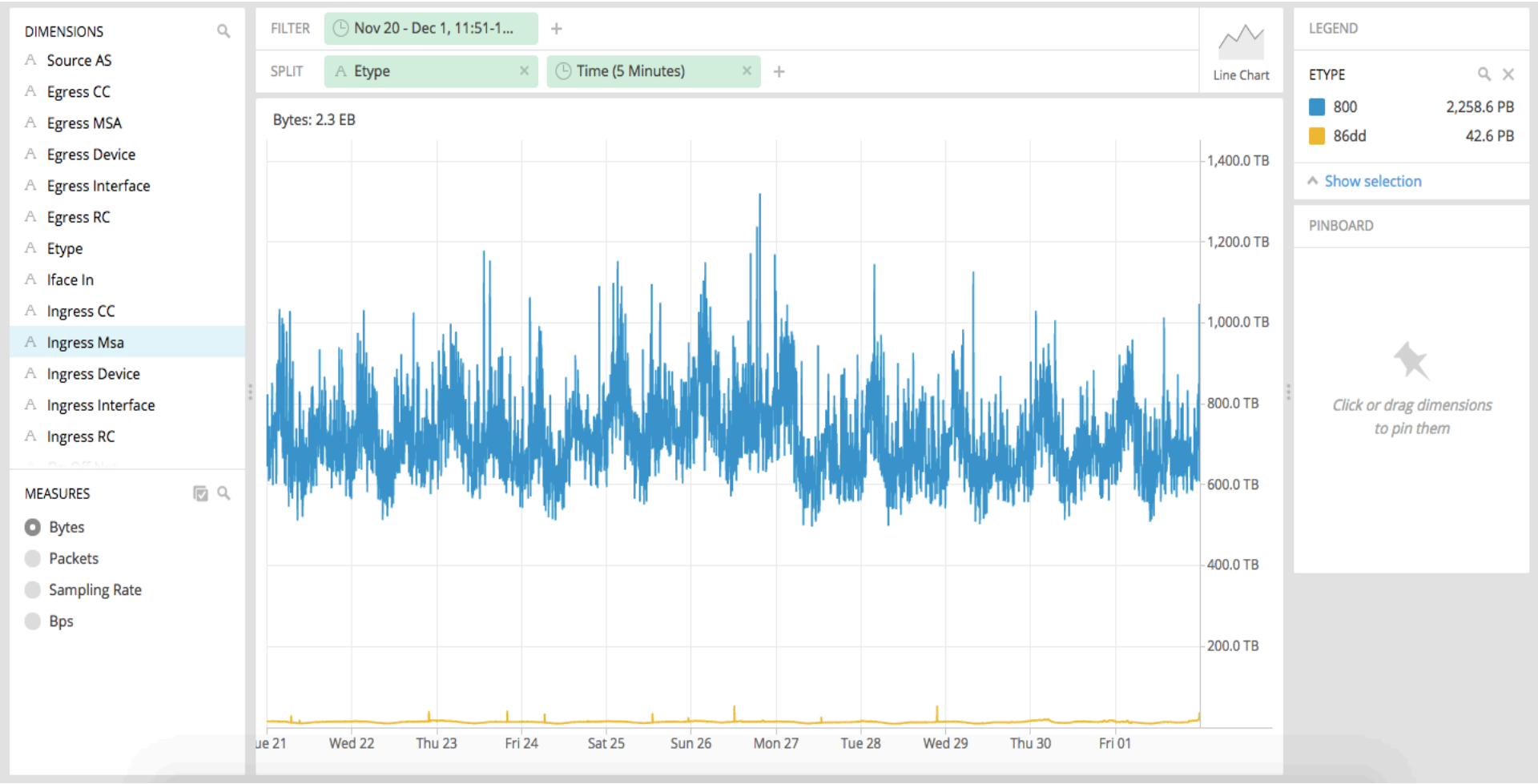
Storing data persistently (cont.d)

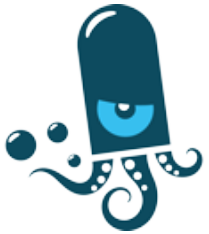
- Otherwise, select a few technologies:
  - Ingestion methods and performance
  - Query methods and performance
  - Software language
  - Support options
- Test them
- Choose one
- Congrats on becoming the Big Data shop of your company! ☺

Storing data persistently (cont.d)

- "noSQL" databases (Big Data ☺):
  - Able to handle large time-series data-sets
  - Meaningful subset of SQL query language
  - Innovative storage and indexing engines
  - Scalable: clustering, spatial and temporal partitioning
  - UI-ready: ie. ELK and TICK stacks
- Open-source RDBMS:
  - Able to handle large data-sets
  - Flexible and standardized SQL query language
  - Solid storage and indexing engines
  - Scalable: clustering, spatial and temporal partitioning

UI example


DIMENSIONS
Source AS
Egress CC
Egress MSA
Egress Device
Egress Interface
Egress RC
Etype
Iface In
Ingress CC
Ingress Msa
Ingress Device
Ingress Interface
Ingress RC
MEASURES
Bytes
Packets
Sampling Rate
Bps
FILTER
Nov 20 - Dec 1, 11:51-1...
SPLIT
Etype
Time (5 Minutes)
Line Chart
Bytes: 2.3 EB
1,400.0 TB
1,200.0 TB
1,000.0 TB
800.0 TB
600.0 TB
400.0 TB
200.0 TB
ue 21
Wed 22
Thu 23
Fri 24
Sat 25
Sun 26
Mon 27
Tue 28
Wed 29
Thu 30
Fri 01
LEGEND
ETYPE
800
2,258.6 PB
86dd
42.6 PB
Show selection
PINBOARD
Click or drag dimensions to pin them



pmacct and network analytics

Paolo Lucente paolo@pmacct.net

http://www.pmacct.net/ | https://github.com/pmacct/pmacct

FRNOG 31, Paris – Oct 2018

Bonus slides

Telemetry data correction

- Telemetry data may get imprecise (ie. due to sampling)
- Use interface stats as gold standard
- Mold telemetry data .. to match interface stats:
  - Builds on Traffic Matrix estimation methods:
    - Tutorial: Best Practices for Determining the Traffic Matrix in IP Networks, NANOG 43
  - Adds telemetry data to linear system to solve
  - Solve system such that there is strict conformance with link stat values, with other measurements matched as best possible

Briefly on scalability

- A single collector might not fit it all:
  - Memory: can't store all BGP full routing tables
  - CPU: can't cope with the pace of telemetry export
- Divide-et-impera approach is valid:
  - Assign PEs (both telemetry and BGP) to collectors
  - If natively supported DB:
    - Assign collectors to DB nodes
    - Cluster the DB
  - If not-natively supported DB:
    - Assign collectors to message brokers
    - Cluster the messaging infrastructure