Agile Infrastructure at CERN - selected topics

Maarten Litmaath
CERN
IT/SDC

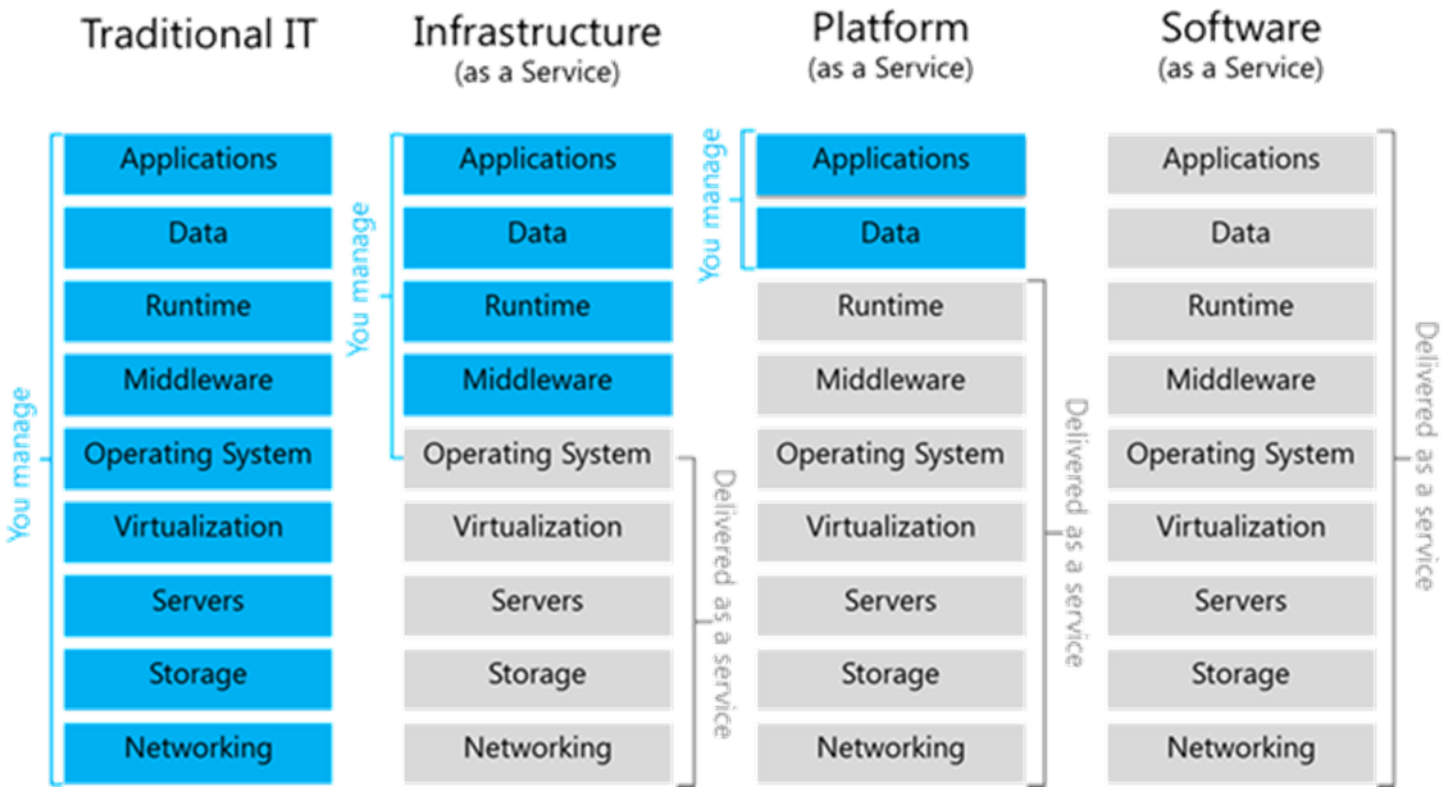
ALICE T1/T2 Workshop, CC-IN2P3
June 4-6, 2013

Sources: HEPiX Spring 2013

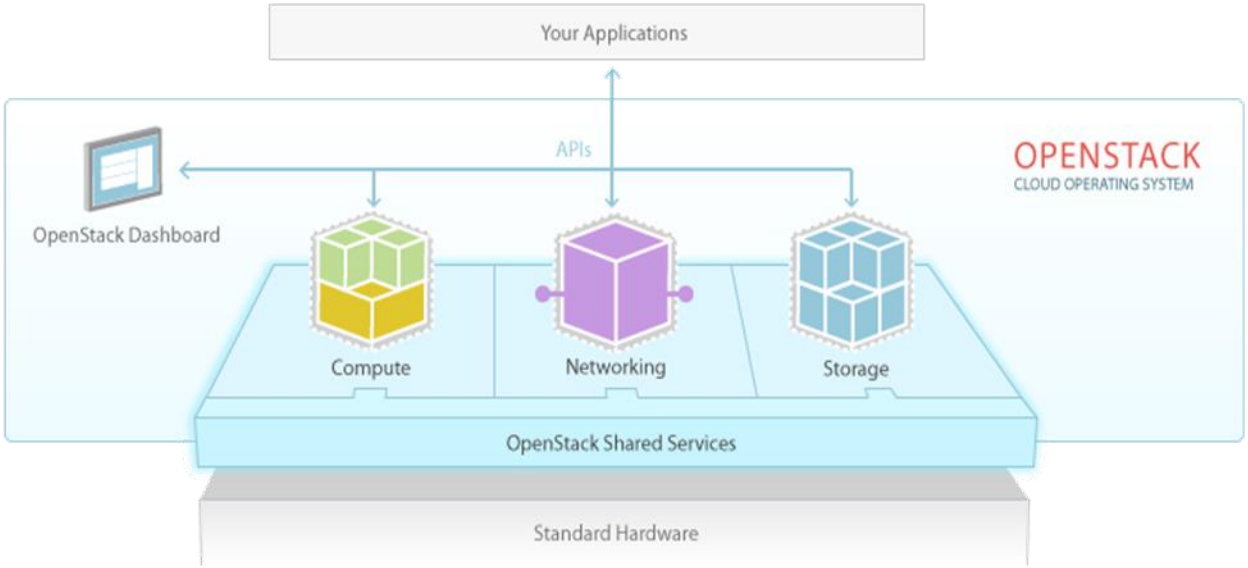
http://indico.cern.ch/conferenceTimeTable.py?confId=220443

- "Agile Infrastructure: an updated overview of IaaS at CERN"
  - Luis Fernandez Alvarez
- "Ceph: A scalable, organic option for Storage-as-a-Service at CERN"
  - Arne Wiebalck, Dan van der Ster
- "Experiences running a production Puppet infrastructure @ CERN"
  - Ben Jones
- "Agile Infrastructure Monitoring"
  - Pedro Andrade

IaaS


Traditional IT
Infrastructure (as a Service)
Platform (as a Service)
Software (as a Service)
Applications
Data
Runtime
Middleware
Operating System
Virtualization
Servers
Storage
Networking
You manage
Delivered as a service

OpenStack


Your Applications
APIs
OPENSTACK
CLOUD OPERATING SYSTEM
OpenStack Dashboard
Compute
Networking
Storage
OpenStack Shared Services
Standard Hardware

Goals

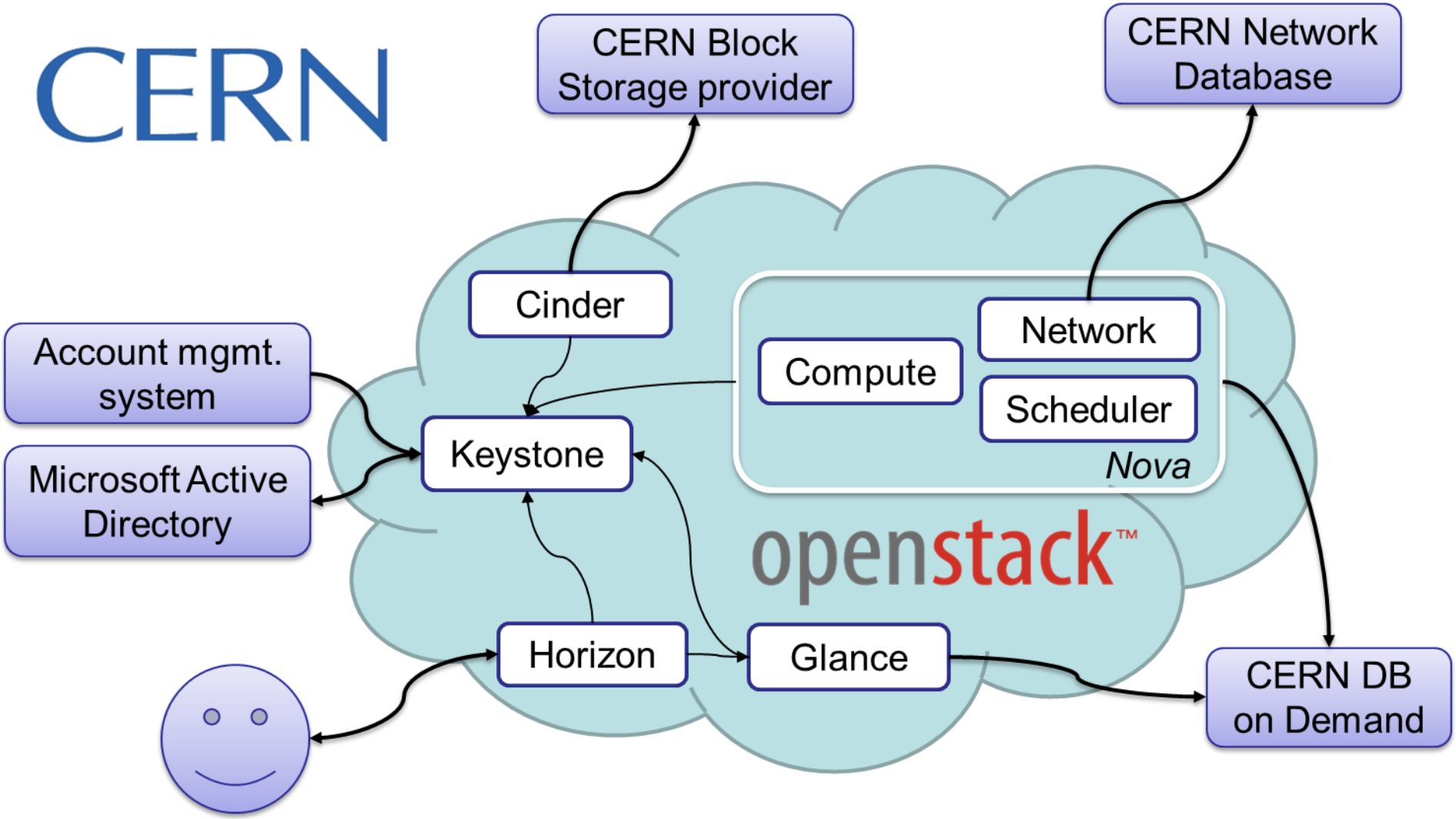
- 90% of hardware virtualized
- ~15k hypervisors needed over next two years
- Follow industry reference structure
  - Infrastructure, Platform and Applications
- Deploy multi-site
  - Extend to 2nd data center in Hungary
- More flexible → efficient use of our hardware
- Better tracking of usage
- Consolidate support
- New use cases (PaaS)

Collaboration & Contribution

- OpenStack candidates testing, bug fixing
- PuppetLabs, Foreman, Aeolus Oz, ...

- Exploit the collaboration tools benefits
  - Mailing lists, IRC, bugzilla, Launchpad, GitHub,…
- Collaborations already starting around
  - BNL, IN2P3, NeCTAR (Au), IHEP (Cn), …

Infrastructure Overview



Components

- Nova: cloud computing fabric controller
- Glance: services for discovering, registering, and retrieving VM images
- Keystone: authentication, authorization and service catalog
- Horizon: GUI for managing OpenStack services

Status

- Used in production for certain services
  - E.g. SLC6 WN
- ~1k VMs, ~500 hypervisors
- ~1900 nodes managed through Puppet
  - ~750 VM
- Steadily ramping up

Current storage services and wishes

- AFS as a distributed file system
- CASTOR for tape storage / archive
- EOS for analysis
- NetApp filers for special cases
- Block storage for Agile Infrastructure VMs
- Hadoop
- "dropbox"

Ceph for Storage as a Service

- Ceph is a distributed, open-source storage system.
  - Scalability
    - TB → EB, 10k machines
  - Reliability
    - no SPOF, configurable replication
  - Manageability
    - self-organizing, self-healing
  - Unification
    - Object/block/file store

How might it fit what we need?

- Provide block storage for AI VMs
  - OpenStack/Cinder & Glance compatible
- Further simplify AFS (and NFS)
  - Rely on safe backend
- Future options
  - Object backend for "dropbox"-like service?

Plan

- Build a serious (1 PB) prototype

- Test it for our use cases
  - OpenStack/Cinder
  - AFS backend
  - OwnCloud
  - CephFS

Puppet infrastructure

- Foreman
  - kickstart, External Node Classifiers, report visualisation
- Multiple puppet masters
  - load balanced (simple for now)
- PuppetDB
  - stored configs, complex manifests, data mining
- Hiera data separation
- git managed manifests
  - split into “modules”, “hostgroups”, “hieradata”

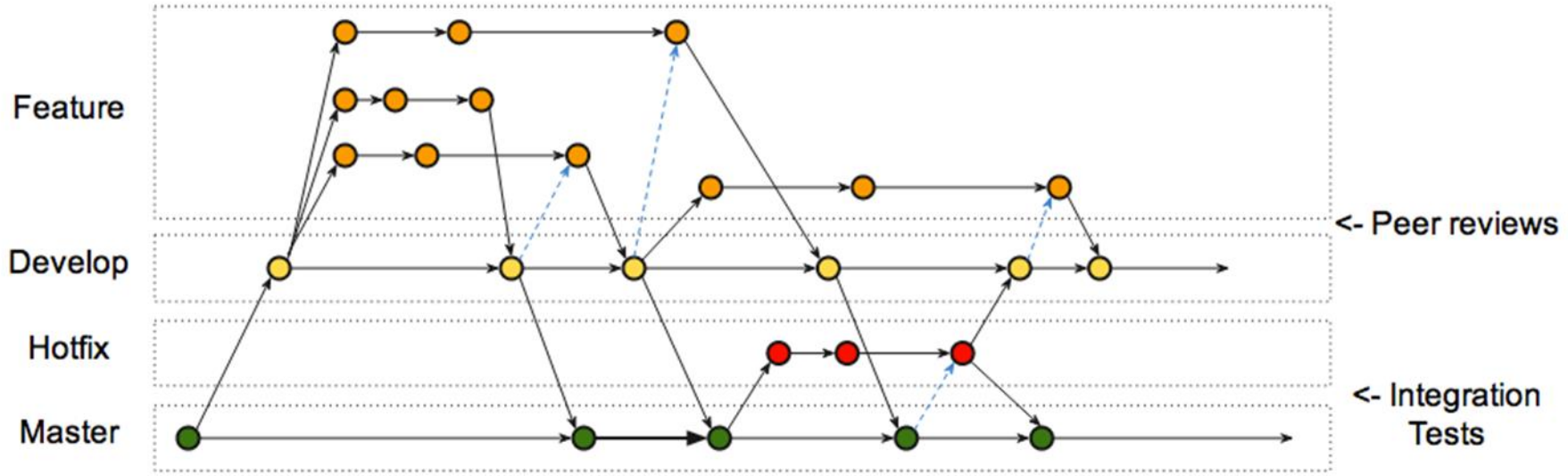
Who are using it?

- OpenStack, CVMFS, lxplus
- Modules appearing for grid services on SLC6
  - BDII, MyProxy, glexec WN, Argus, VOMS, CASTOR, …
- ~1900 nodes already
  - ~750 VM

Facter

- Unlike what we were used to in Quattor, machine facts can inform configuration through the Facter tool
- Facts + ENC + config → catalog
- Cache ENC results to avoid overloading third parties (here: LAN DB)
  - But beware of stale caches, as usual…

git workflow is easy to get wrong



- Branches → puppet “dynamic environments”

Community

- “We’re not special”
- https://github.com/cernops
- Fork where necessary but always try to return to mainstream
- Would welcome more involvement from HEPiX sites on common problems

AI monitoring context

- Monitoring in IT covers a wide range of resources
  - Hardware, OS, services, applications, files, jobs, etc.
  - Resources have several dependencies between them

- Many application-specific solutions at CERN IT
  - Similar needs and architecture
    - Publish metric results, aggregate results, alarms, etc.
  - Different technologies and tool-chains
    - Some based on commercial solutions
  - Similar limitations and problems
    - Limited sharing of monitoring data

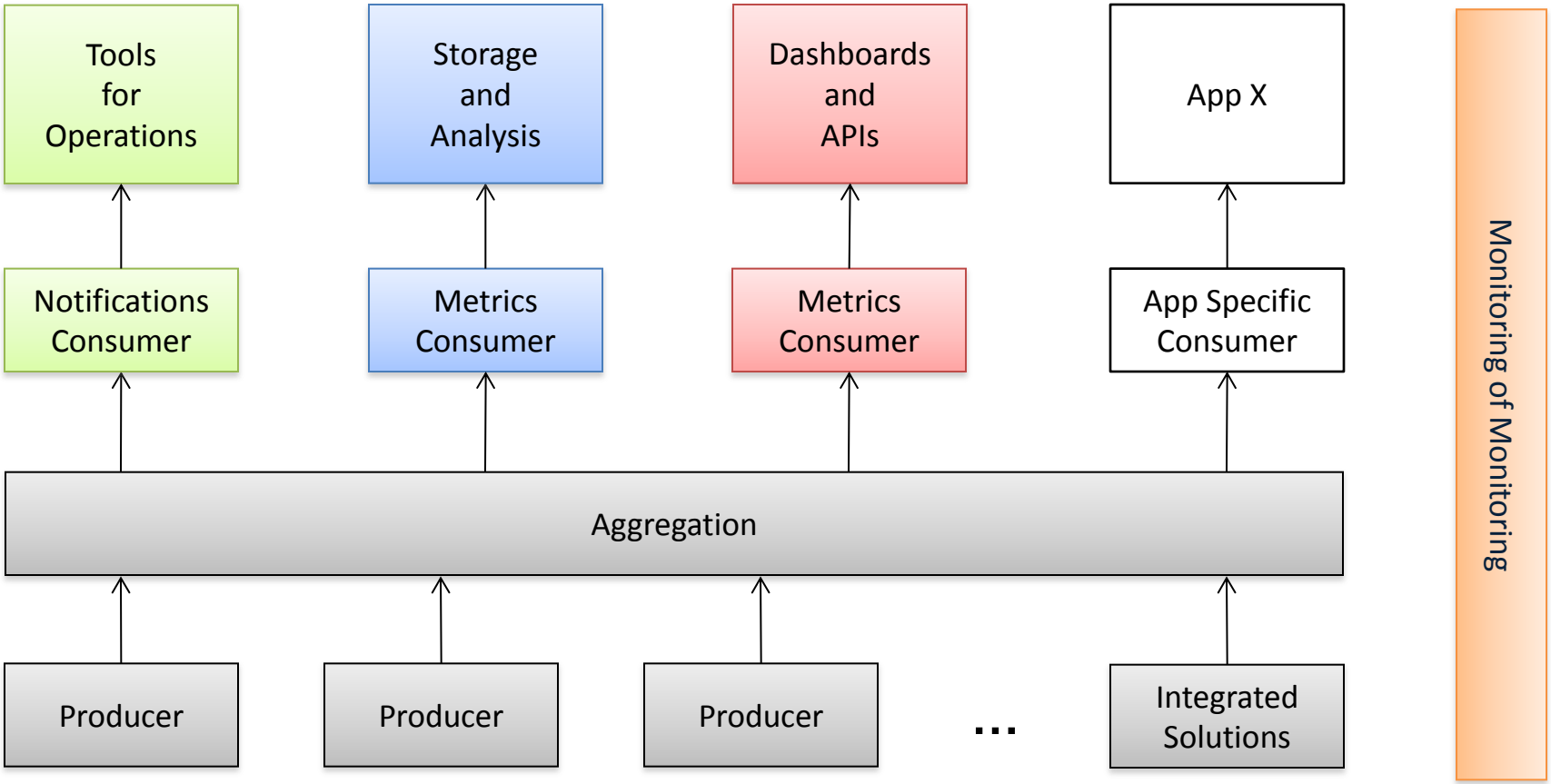
Common monitoring strategy goals

- Make monitoring data easy to access
- Combine and correlate monitoring data
- Better understand infrastructure/services performance
- Quick and easy deployment/configuration of dashboards
- Move to a virtualized and dynamic infrastructure
- Optimize effort allocated to monitoring tasks

Strategy and Architecture

- Architecture based on a toolchain approach
  - All components can be easily replaced
- Adopt whenever possible existing technologies
  - mongoDB, logstash, hadoop, kibana, etc.
- Scalability always taken in consideration
  - Horizontally scaling or by adding additional layers
- Messaging infrastructure as key transport layer
  - Messages based on a common format (JSON)
  - Messages based on a minimal specification

Components as SaaS and/or PaaS



Technologies

- General Notification Infrastructure (GNI)
- MongoDB, ActiveMQ, DB on Demand, Django
- Hadoop
- Splunk
- Logstash + Elasticsearch + Kibana