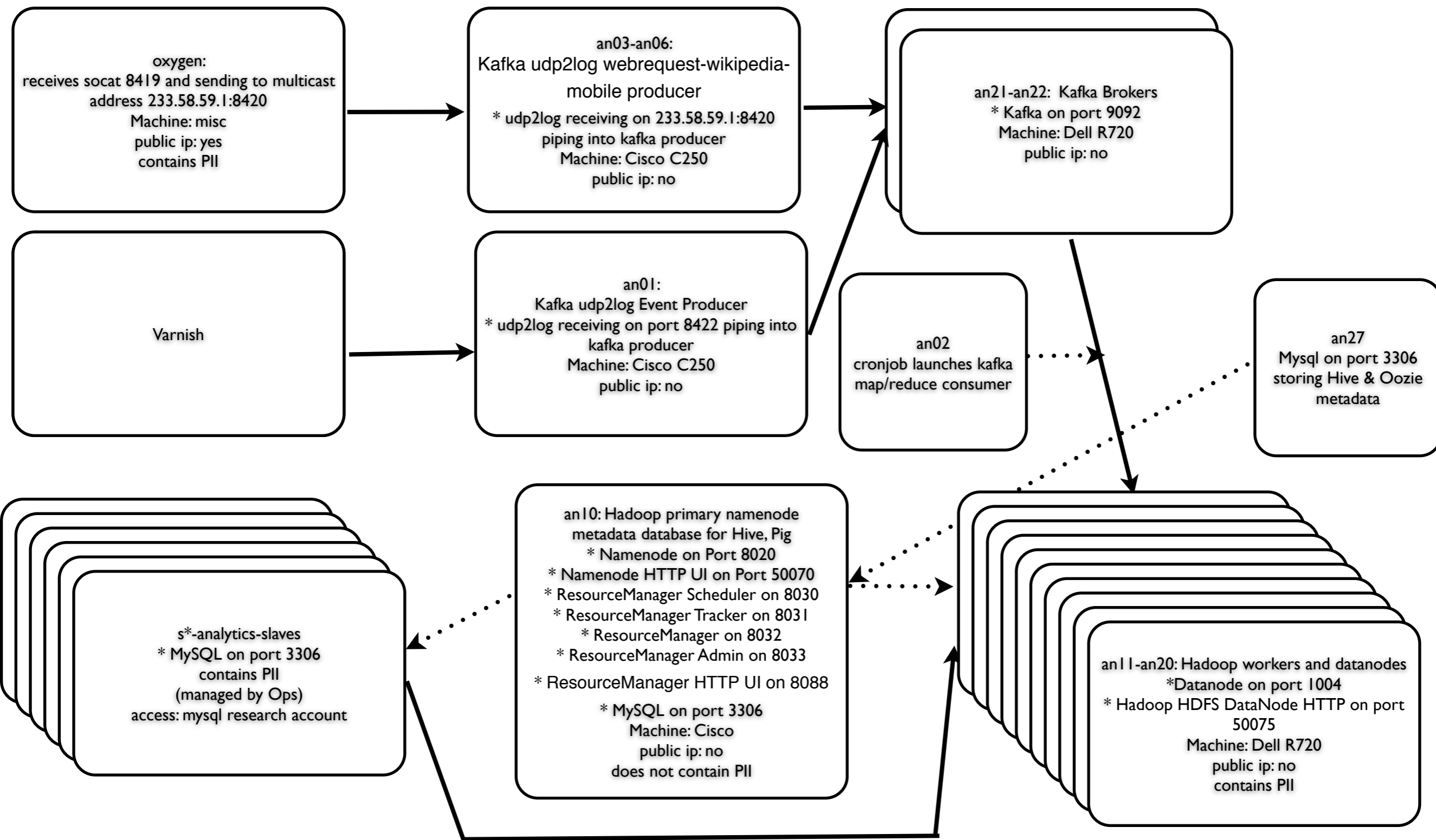


# Kraken - Minimum Viable Cluster

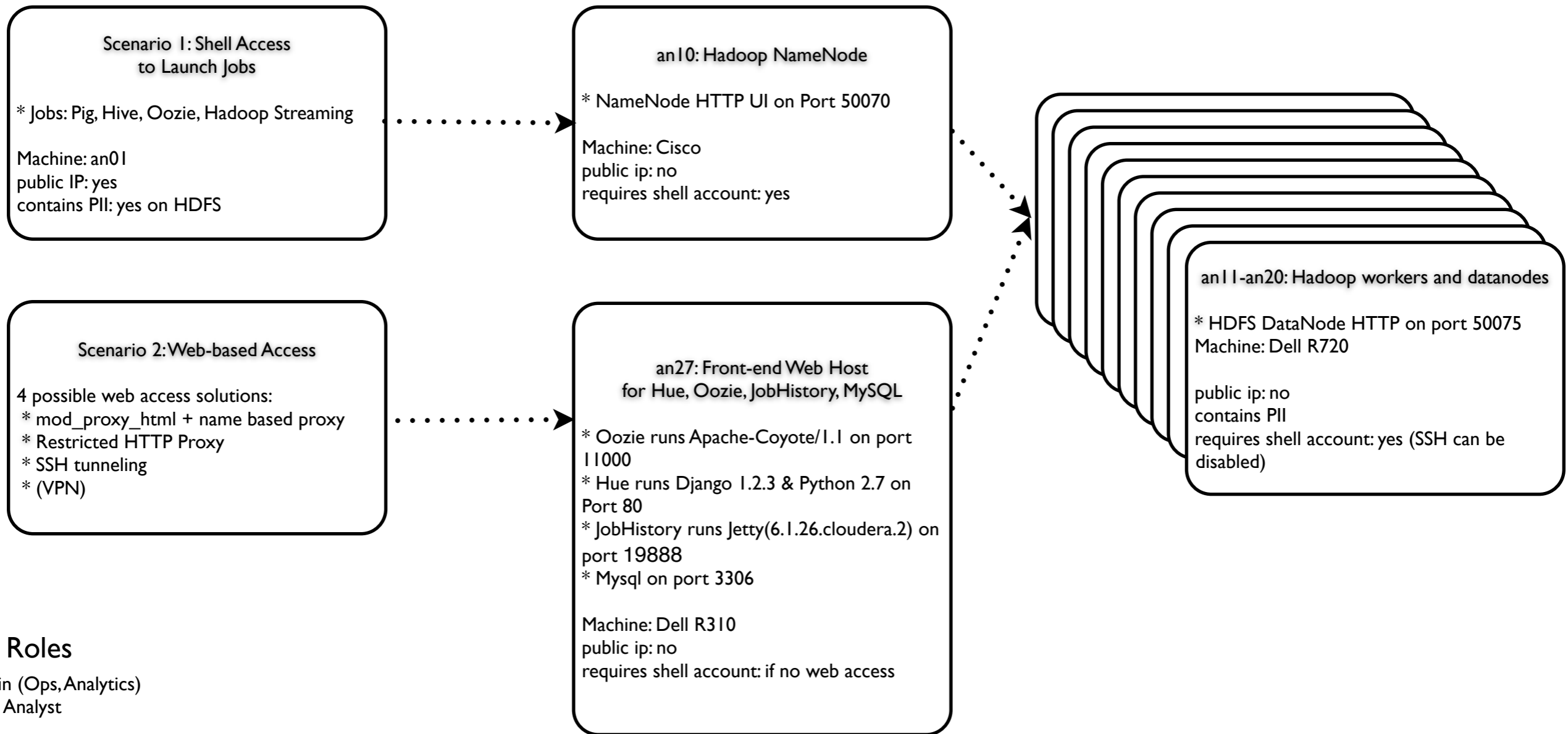
- Unsampled storage of Mobile pageview requests as txt files
- Unsampled storage of EventLogging data as JSON or txt files
- Importing MediaWiki databases
- Ability to access raw datafiles by WMF employees
- Ability to query data using Pig, Hive, MapReduce, and Hadoop streaming
- Ability to schedule jobs and workflows (Oozie)
- Automated recovery of data importers (Zookeeper)
- Not part of MVC – deferring for later discussion:
  - Realtime processing layer (Storm)
  - Guaranteed delivery of web-request data (Kafka producers and future of udp2log)

# Kraken Minimum Viable Cluster **Data Flow**



—————> dataflow  
 .....> coordination flow

# Kraken Minimum Viable Cluster Data Analyst Access



## User Roles

- 1) Admin (Ops, Analytics)
- 2) Data Analyst

**\* All are assumed to be WMF employees for now**