

New Approach to Big Data

The Snagajob Story

Robert Fehrmann

Principal Architect @ Snagajob



4.5 million
monthly job
applications



ONE million
new workers added
to the site each month

70+ million
registered hourly workers

LARGEST

marketplace for hourly work

3.5 million
positions filled in 2016



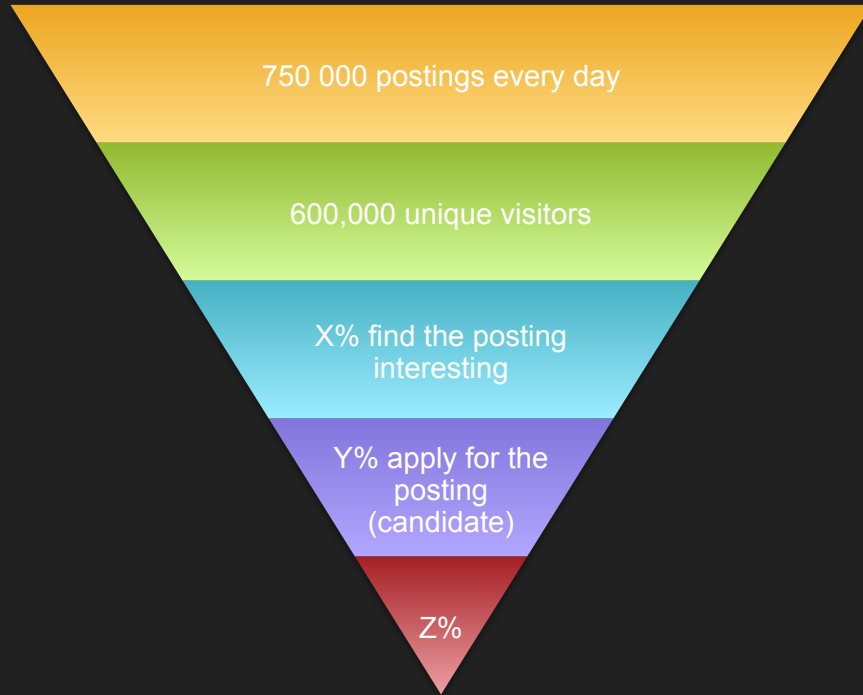
300,000
employer locations

About Me

- Master Degree in Computer Science from “Technische Universitaet Braunschweig”
- 25 years building the data tier for applications in different verticals
- Evangelist for polyglot data environments
- Community involvement (MongoDB User Groups / DevOps)



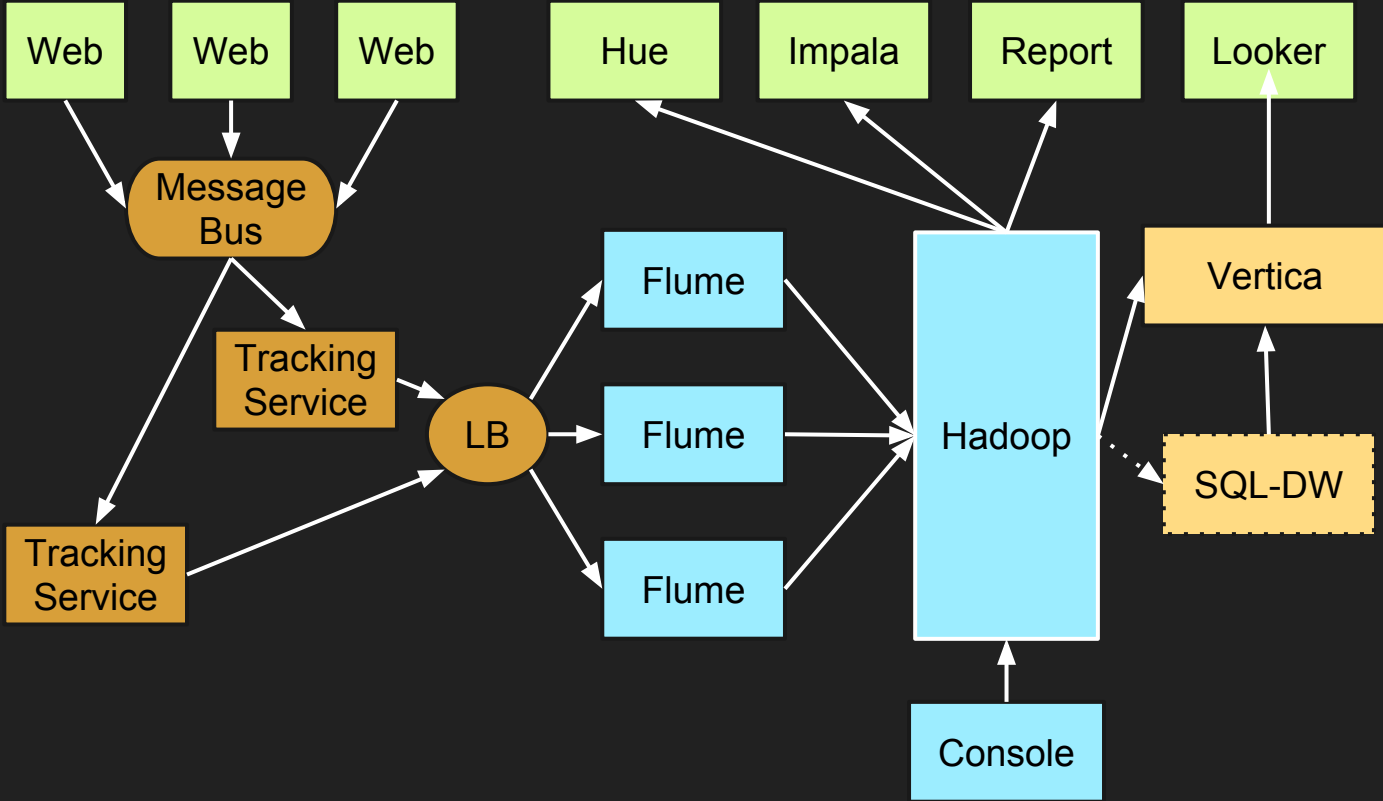
Funnel Analysis



Using Analytics to understand the funnel

- Geographical Analysis
- Customer Analysis
- Historical Analysis
- Industry Analysis
- Click through rate & abandoning the search
- What makes a Posting Interesting, ...

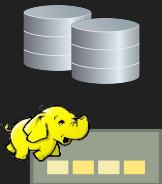
Data Collection Framework V1



Evolution

Data warehouse & platform software
(on premise)

Vertica Data Warehouse



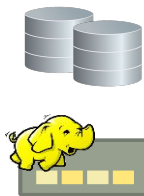
Hadoop

"We want to be a cloud based company"
Peter Harris, CEO



Move to Cloud
Doesn't solve all problems

Vertica Data Warehouse



Hadoop

Search Continues
For a true cloud solution till

...

2012

2014

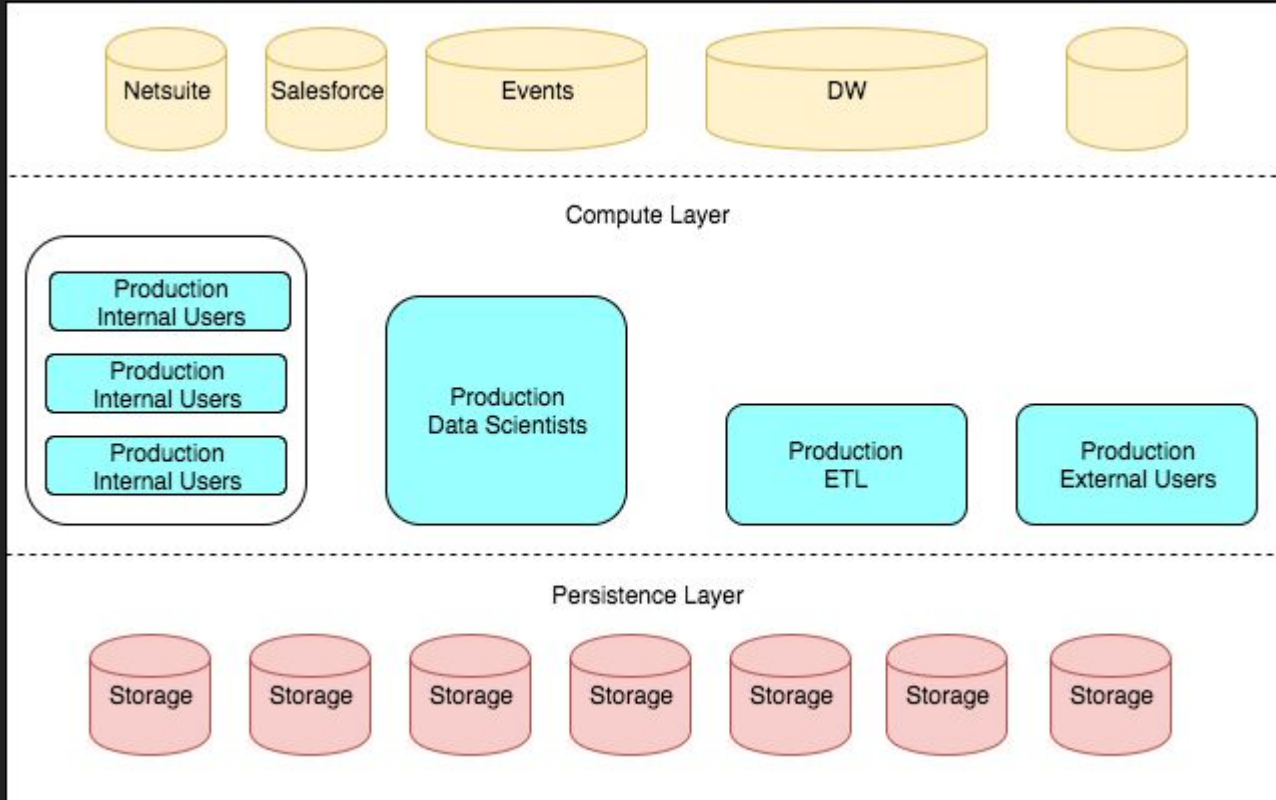
2015

2016

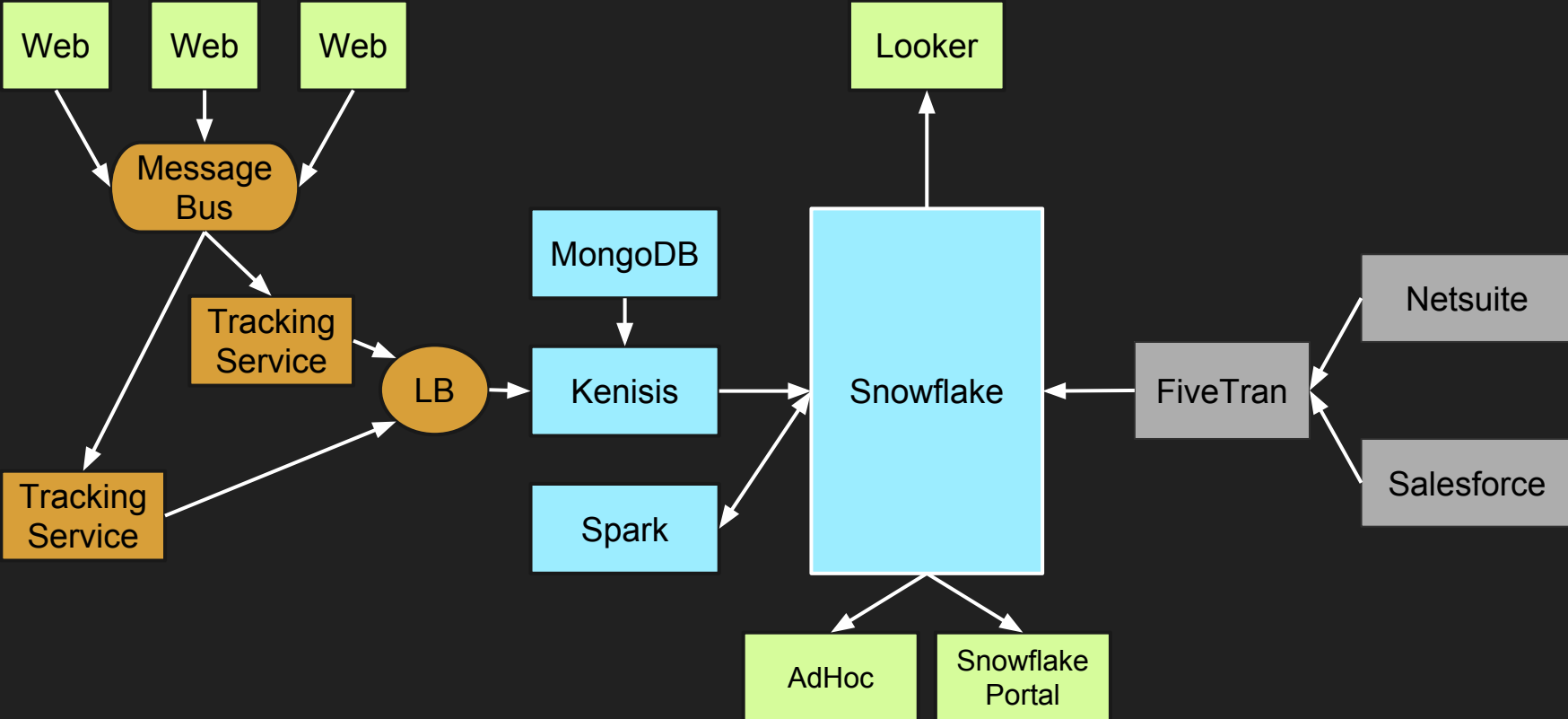
Goals for Next Generation Solution

- Horizontal Scalability
- PaaS
- Stability
- Ease of Use
- Can't be more expensive

Architecture



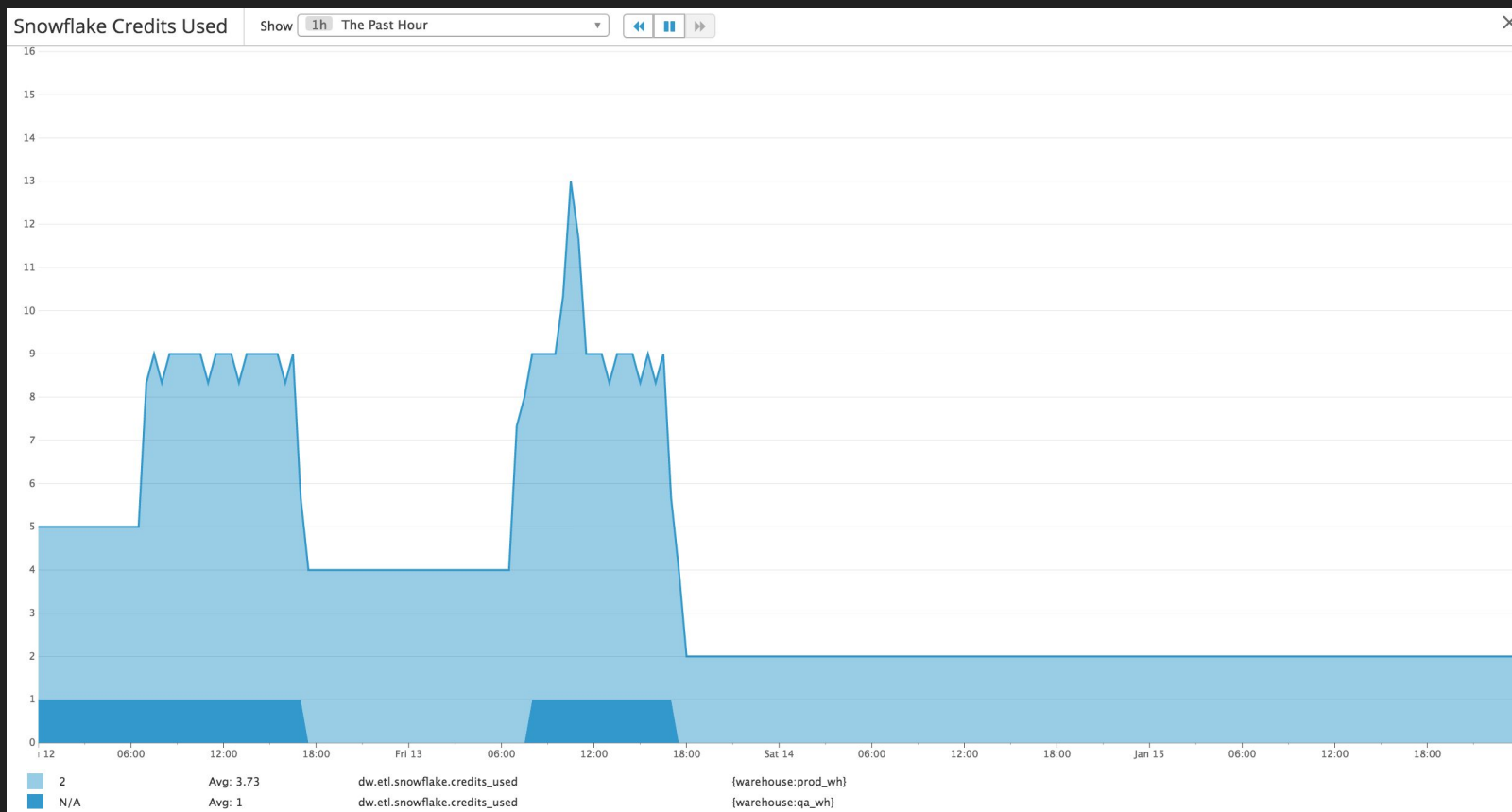
Data Collection Framework V2



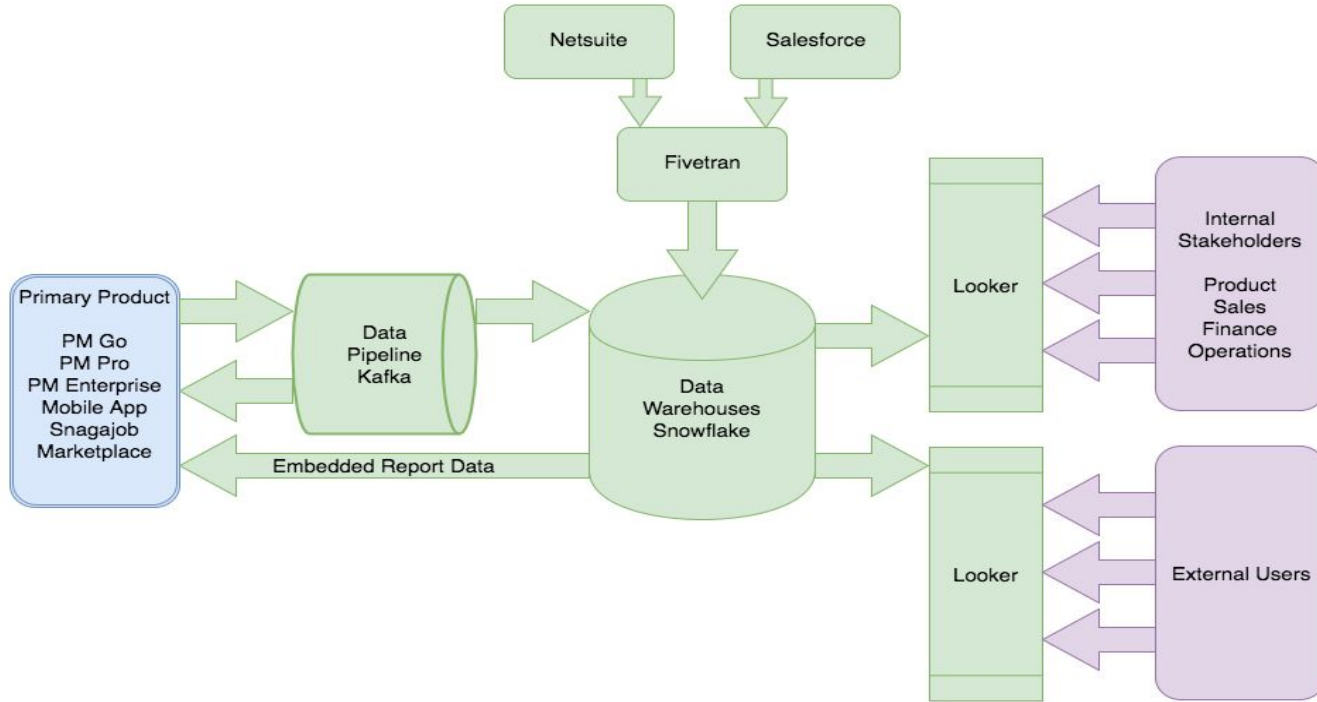
Results: Performance



Results: Better Use of Resources



Snagajob Platform



Other Features

- Undrop (DB, Table, Schema) no restore required
- Clone (DB, Table, Schema) (metadata only operation)
- Native JSON Parsing (as well as CSV, AVRO, XML, Parquet)
- Automatic Encryption of Data
- Automatic Query Optimization (no tuning)
- All Data in one place (single source of truth)

