Managing the data science team

Jeff Leek

@jtleek

www.jtleek.com

- Individual meetings
  - Regular
  - Updates, problems, and goals

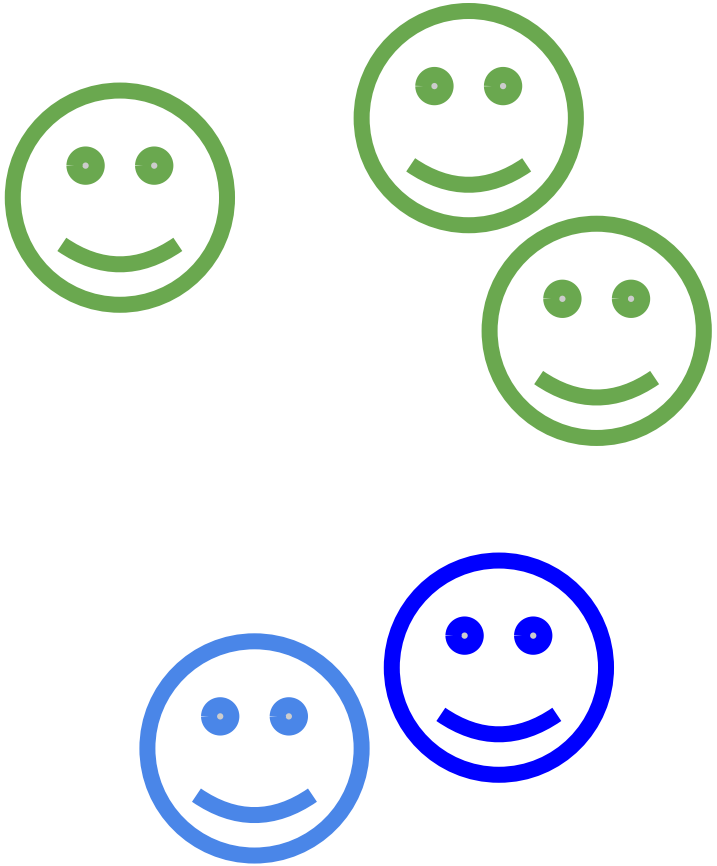
http://www.informationweek.com/big-data/big-data-analytics/5-ways-data-science-teams-can-raise-their-profiles/a/d-id/1316532

- Data science team meetings
  - Regular
  - Updates, problems, and goals
  - Peer review
  - Team priorities and motivation

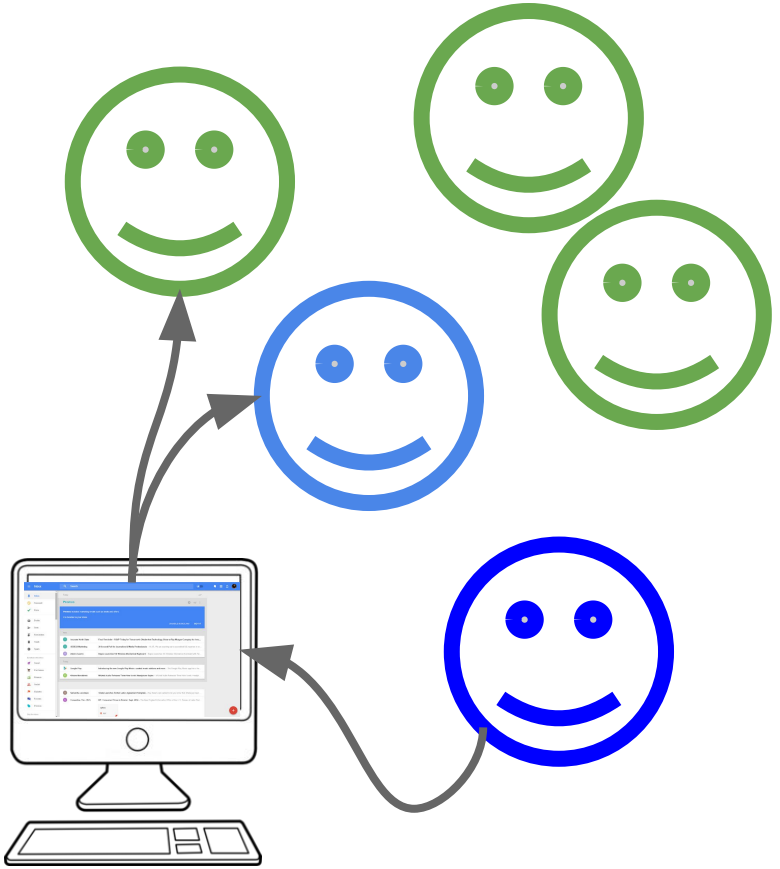

http://www.pulseheadlines.com/psychology-studies-reliable/4258/
COS
CENTER FOR SCIENCE

Monitoring interactions

- Actively
- Passively



Active management

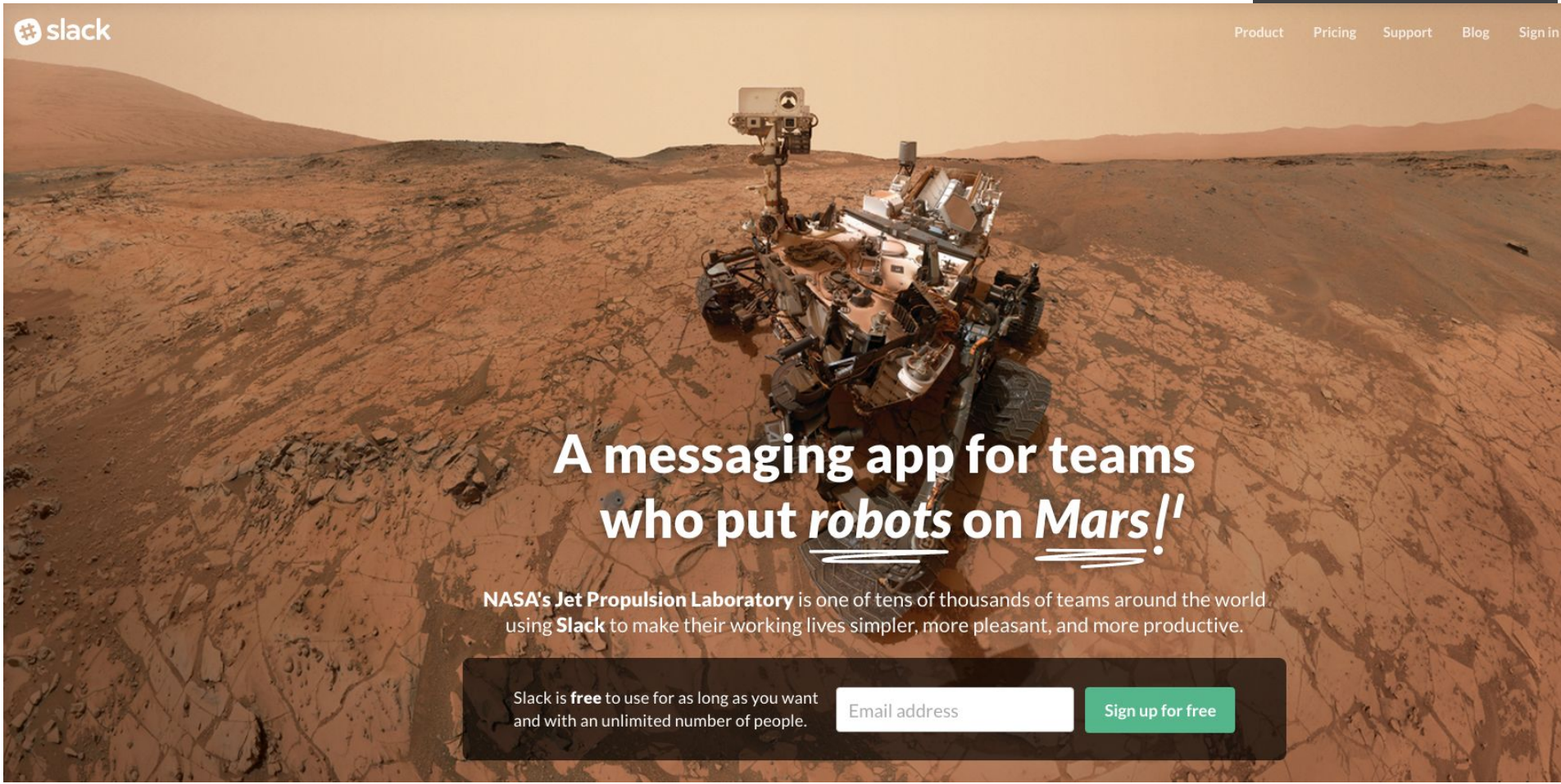


Passive management

Keeping things running

- Open door policy
- Quick questions via chat/email

https://slack.com/



Managing growth

- Introduction to new tools
- Introduction to new interactions
- Opportunities for advancement

README.md

https://github.com/apache/spark

Apache Spark

Spark is a fast and general cluster computing system for Big Data. It provides high-level APIs in Scala, Java, Python, and R, and an optimized engine that supports general computation graphs for data analysis. It also supports a rich set of higher-level tools including Spark SQL for SQL and DataFrames, MLlib for machine learning, GraphX for graph processing, and Spark Streaming for stream processing.

http://spark.apache.org/

Online Documentation

You can find the latest Spark documentation, including a programming guide, on the project web page and project wiki. This README file only contains basic setup instructions.

Building Spark

Spark is built using Apache Maven. To build Spark and its example programs, run:

```
build/mvn -DskipTests clean package
```