Introduction

Background

• History
  - Founded in 2006
  - Launched in Sweden 2008
  - Launched in USA 2011

• Stats
  - Over 24 million active users
  - Over 6 million paying
  - Over 20 million songs
  - Available in 32 countries

Spotify quick facts
The beginning of Big Data at Spotify

Hadoop
- First cluster in the fussball room
- 10 nodes
Enter analytics infrastructure

Analytics
• Infrastructure/engineers
• Analysts
Strategy: platformization

Big bet
• Hardware
• Engineers
• $$$
Why invest in Big Data?

It’s all about feedback & learning
- Product
- User
Product feedback

- Company KPIs
- AB-testing
- Funnel analytics
- Growth analytics
- Operational analytics
- Reporting
- Partner data
- Marketing
User feedback

- Toplists
- Artist/album/track recommendations
- Notifications
- Popularity
- Radio
- Music Graph
Data Infrastructure **Architecture** at 10 000 feet
Core technology

- Hadoop
- Hive
- Luigi
- Storm: Distributed and fault-tolerant realtime computation
- Apache Kafka: A high-throughput distributed messaging system

etc...
What’s next?

Projects
• Data development environment
• Streamlined AB-testing infrastructure
• Evaluation of third-party visualization tools
• etc...

Future Challenges
• Next generation hadoop
• What do you do when hadoop no longer scales?
• Real-time processing
Next-gen hadoop?

Partnership
• Moving from Cloudera to Hortonworks
• Upgrading to HDP 2.0

Use-cases
• Storm on Yarn
• Giraph on Yarn
• Hive via Tez
Want to join the band?

Check out http://www.spotify.com/jobs or @SpotifyJobs for more information.

Or mail: fabian@spotify.com

October 1, 2013
Questions?

Mail: fabian@spotify.com
or visit one of our meetups

477 SHUGers

1,417 Music Hackers

October 1, 2013