



# Creating a Data Engineering Culture

- **What Is a Data Engineering Culture?**
- Data Engineering Culture Stories
- How to Create a Data Engineering Culture?
- Common Reasons for Failure

**85% of Big Data projects  
fail to get into production**

Source: <http://tiny.bdi.io/gartnerfail>

**Projects that do make it  
into production provide  
incredible value**

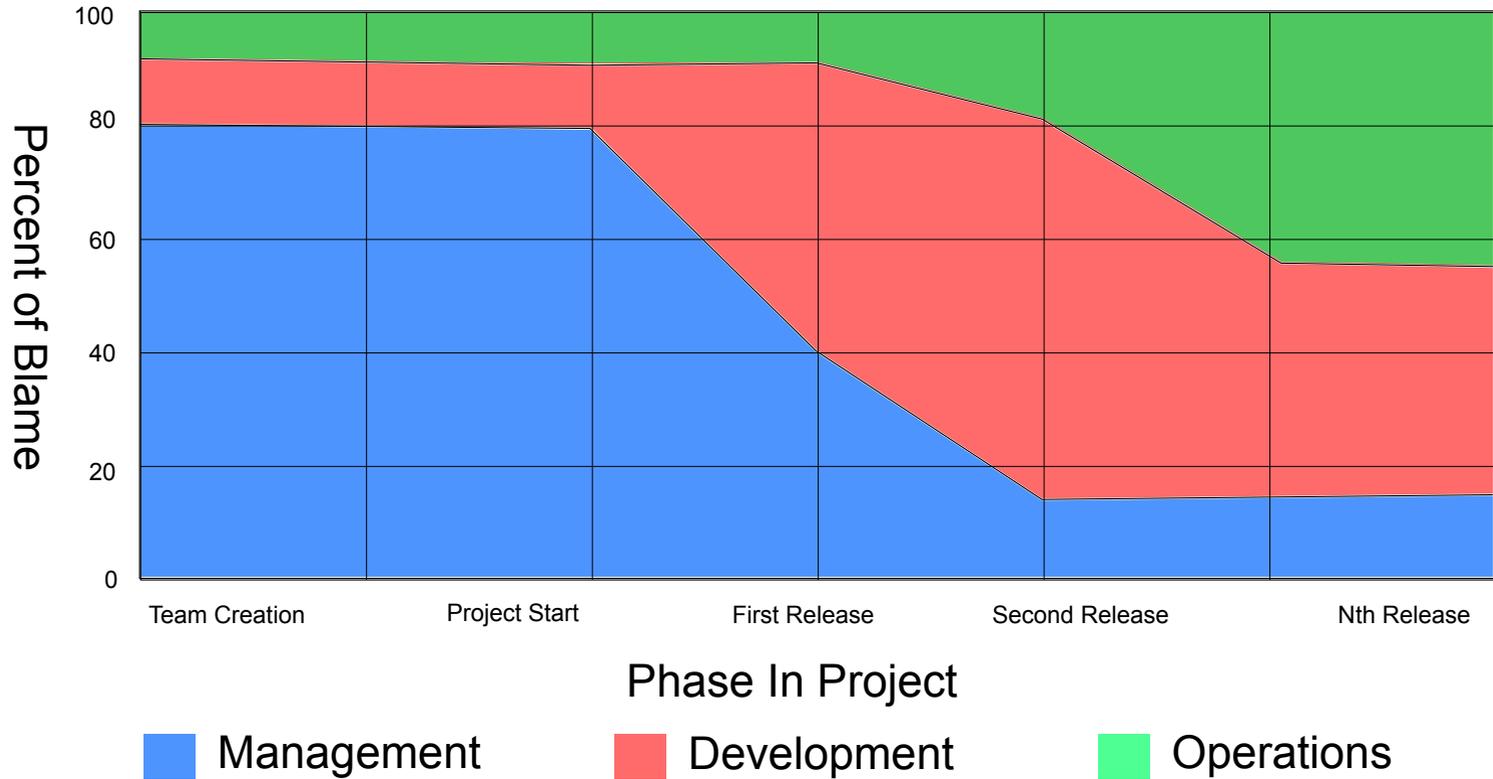
# What is a Data Engineering Culture?

A culture where the value and importance of data engineering is recognized at all levels of the organization

Organization-wide realization that data science and big data require data engineering

Manifestation: a correct ratio of data engineers to data scientists

# Early Success and Failure Is Mostly Management



# Why Create a Data Engineering Culture?

Companies doing big data without data engineers will under-perform or fail

Many companies don't even realize they're under-performing

Conferences and successful Big Data companies assume they have a data engineering team

- Sometimes mentioned as an aside that they need to improve their data engineering

# Creating a Data Engineering Culture

- What Is a Data Engineering Culture?
- **Data Engineering Culture Stories**
- How to Create a Data Engineering Culture?
- Common Reasons for Failure

Data scientists were stopping once it became too difficult on the data engineering side

Data scientists were doing data engineering work because there weren't any data engineers

- Easier for data engineer - virtually impossible for data scientists

Data engineering accelerates data science

Data engineering democratizes data for data scientists, data analysts, and others

Data infrastructure stops being the bottleneck for data projects

# Creating a Data Engineering Culture

- What Is a Data Engineering Culture?
- Data Engineering Culture Stories
- **How to Create a Data Engineering Culture?**
- Common Reasons for Failure

# Who Should Be on the Team?

Data Engineer: a data engineer is someone who has specialized their skills in creating software solutions around big data

Data Scientist: a data scientist is someone who has augmented their math and statistics background with programming to analyze data and create applied mathematical models

DBA: a SQL-focused person

See <http://tiny.bdi.io/devds>

# What Skills Are Needed?

The skills needed on a data engineering team are:

- Distributed systems
- Programming
- Analysis
- Visual communication
- Verbal communication
- Project veteran
- Schema
- Domain knowledge

# Creating a Data Engineering Culture

- What Is a Data Engineering Culture?
- Data Engineering Culture Stories
- How to Create a Data Engineering Culture?
- **Common Reasons for Failure**

DBA Definition - Someone whose only programming language is SQL

- This includes data warehouse, SQL Developers, etc

Big Data is not an extension or the logical extension of data warehousing

It's much much more complex

- <http://tiny.bdi.io/complex>

It's not just a skills gap; it's an ability gap

- <http://tiny.bdi.io/abilitygap>

Big Data is complex

- <http://tiny.bdi.io/complex>

Beginners need to be give the time and resources to learn

It takes at least 6 months for a beginner to become proficient

As you look at successful case study talks, they leave out

- Expert resources provided
- Starting proficiency of the team
- Total time used

Schema problems don't manifest immediately

- Takes 6-12 months to see

You can't lay down PBs of data and change it

Data pipelines need to change data formats

Which role typically has this skill?

- DBAs (I didn't say no DBAs-I said not just DBAs)

A project veteran is someone who has put a Big Data or distributed system in production

Beginners to distributed systems and Big Data are the sources of the worst abominations

- Average time lost is 1-2 man months

Very different to whiteboard and erase than code and rewrite

- Only a project veteran can critique a whiteboarded architecture
- Remember it's programming and operations

# Too Ambitious

You can't go from 0 to Big Data all at once

You really can't go from 0 to the holy grail

Your team needs the time to go from beginners to intermediate to advanced

You need to build momentum first

- Projects without momentum get canceled

Take an honest evaluation of the team

- Skills
- Abilities
- Use case
- Resources

Does the team have a skills gap?

Does the team have an ability gap?

- <http://tiny.bdi.io/abilitygap>

# Data engineering teams need to be multidisciplinary

<http://tiny.bdi.io/detbook>

## Some teams say they don't need help

- Technical people think it's not needed (small data mentality)
- Admission of failure
- "We tend to do everything in house" (NIH)

Very important to take an honest look at the team

Training

Consulting

- Very important to get a company with a good track record

Mentoring

- On going help for the technical and business teams

## Early

**Never too late to fix, but  
fixing will be much more  
costly**

# About Me

Current: Instructor, Thought Leader, Monkey Tamer

Previously:

- Curriculum Developer and Instructor @ Cloudera
- Senior Software Engineer @ Intuit

Covered, Conferences and Published In:

- GigaOM, ArsTechnica, Pragmatic Programmers, Strata, OSCON, Wall Street Journal, CNN, BBC, NPR

See Me On:

- <http://www.jesse-anderson.com>
- [@jessetanderson](#)
- <http://tiny.bdi.io/linkedin>
- <http://tiny.bdi.io/youtube>