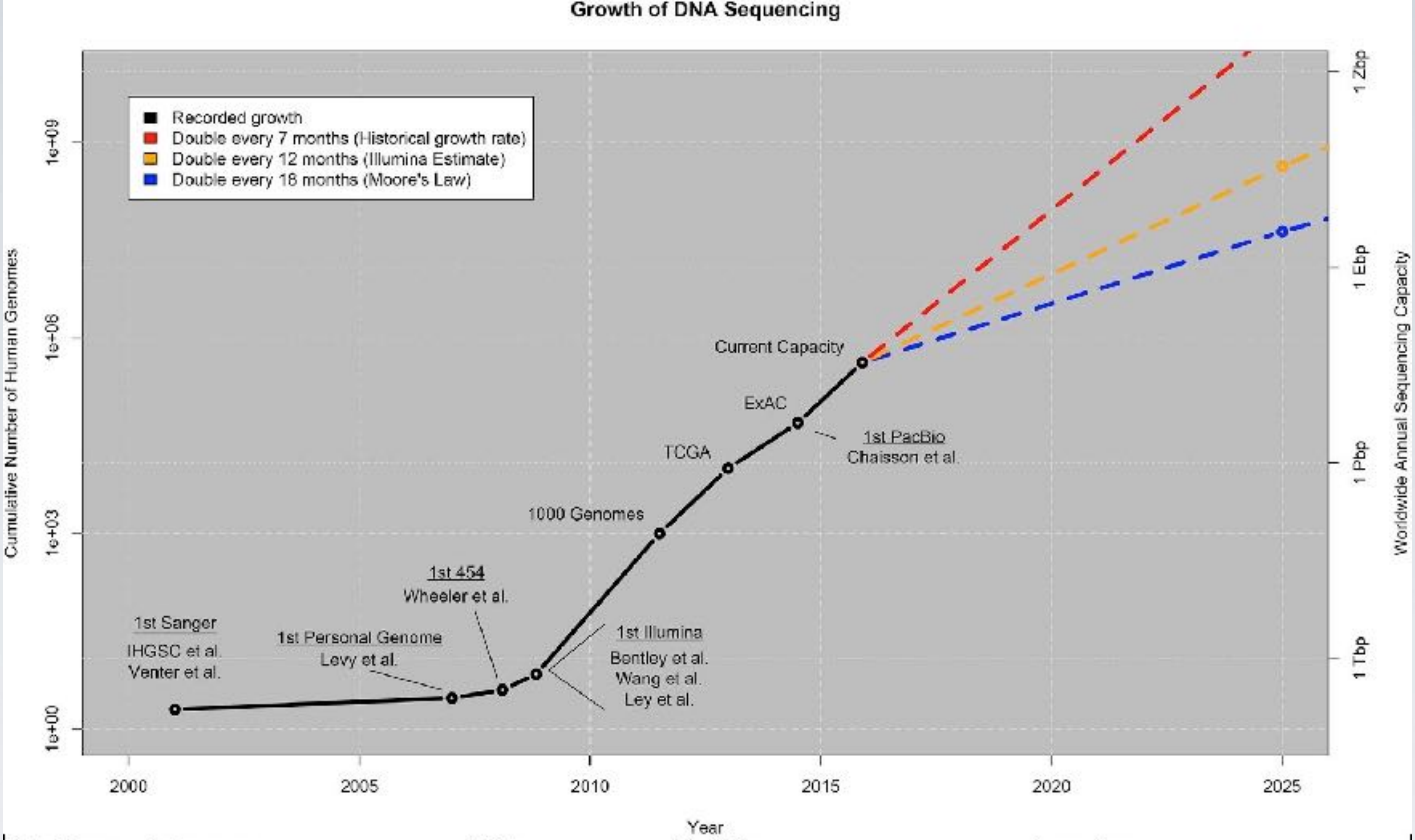
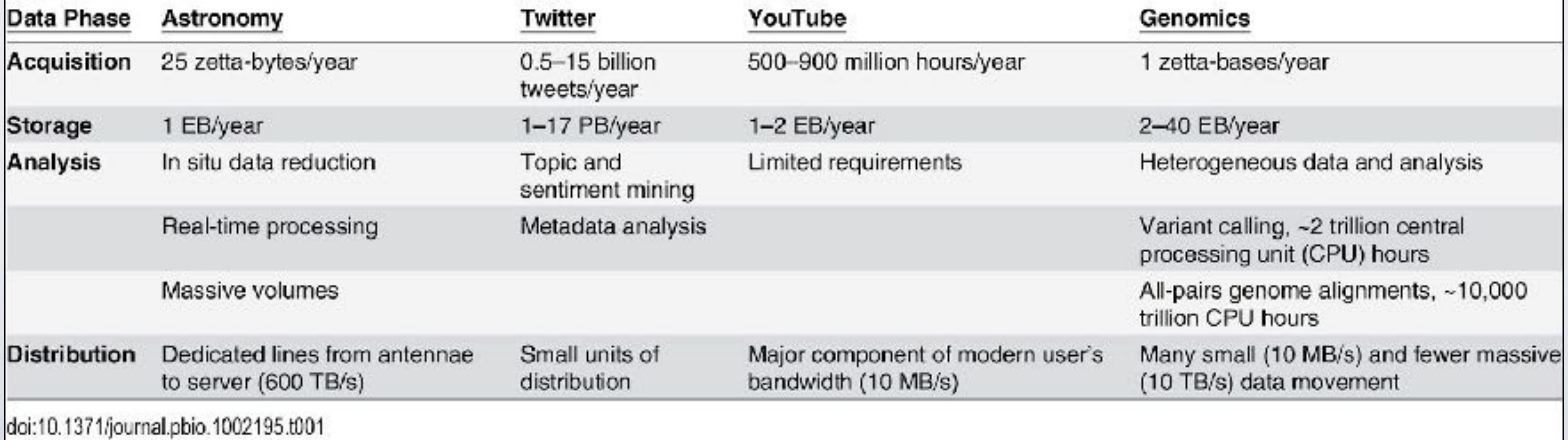
BIONODE.IO

#MozFest 2016

github.com/bionode

gitter.im/bionode/bionode





Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

WHAT

Modular and universal bioinformatics

WHAT

Modular and universal bioinformatics

- Each tool tries to do one thing well

WHAT

Modular and universal bioinformatics

☝ Each tool tries to do one thing well

♻ Provides highly reusable code and tools

WHAT

Modular and universal bioinformatics

- Each tool tries to do one thing well
- ♻ Provides highly reusable code and tools
- Scales by using Streams

WHAT

Modular and universal bioinformatics

- Each tool tries to do one thing well
- Provides highly reusable code and tools
- Scales by using Streams
- Runs everywhere

WHAT

Modular and universal bioinformatics

- Each tool tries to do one thing well
- Provides highly reusable code and tools
- Scales by using Streams
- Runs everywhere

WHAT

Using Node.js

- Each tool tries to do one thing well
- Provides highly reusable code and tools
- Scales by using Streams
- Runs everywhere

WHAT

Using Node.js

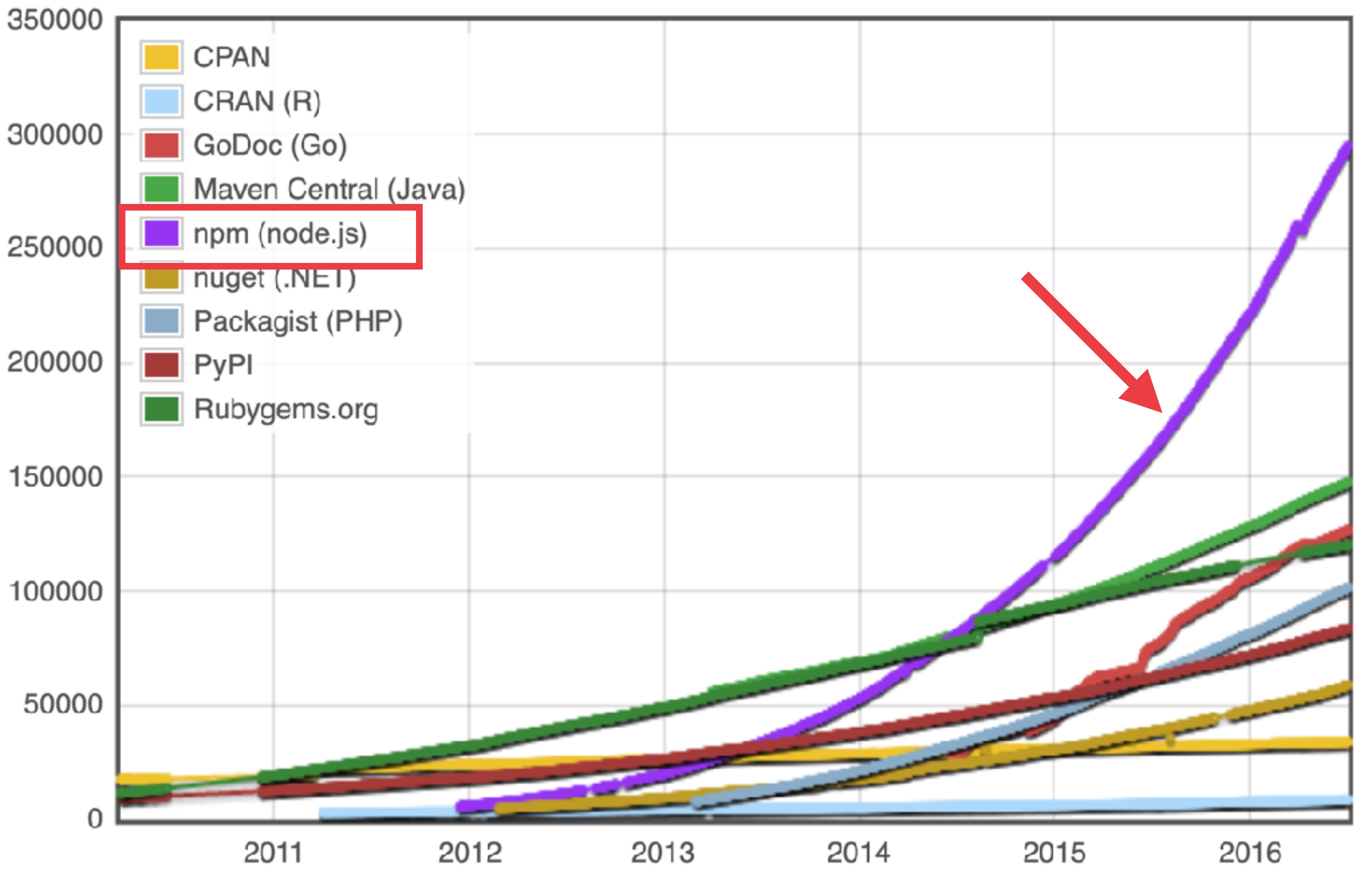
- Highly modular
- Provides highly reusable code and tools
- Scales by using Streams
- Runs everywhere

WHAT

Using Node.js

- Highly modular
- Very open community on GitHub
- Scales by using Streams
- Runs everywhere

Modules count



HOW

Using Node.js

- Highly modular
- Very open community on GitHub
- Scales by using Streams
- Runs everywhere

HOW

Using Node.js

- Highly modular
- Very open community on GitHub
- Provides native implementation of Streams
- Runs everywhere

HOW

Using Node.js

- Highly modular
- ♻ Very open community on GitHub
- Provides native implementation of Streams
- Run same JavaScript code on browser or CLI

ORIGIN

ORIGIN

During my PhD at

ORIGIN

During my PhD at

Compare genetic diversity of social vs solitary species

ORIGIN

During my PhD at

Compare genetic diversity of social vs solitary species

- Involved in biological web projects that need JS

ORIGIN

During my PhD at

Compare genetic diversity of social vs solitary species

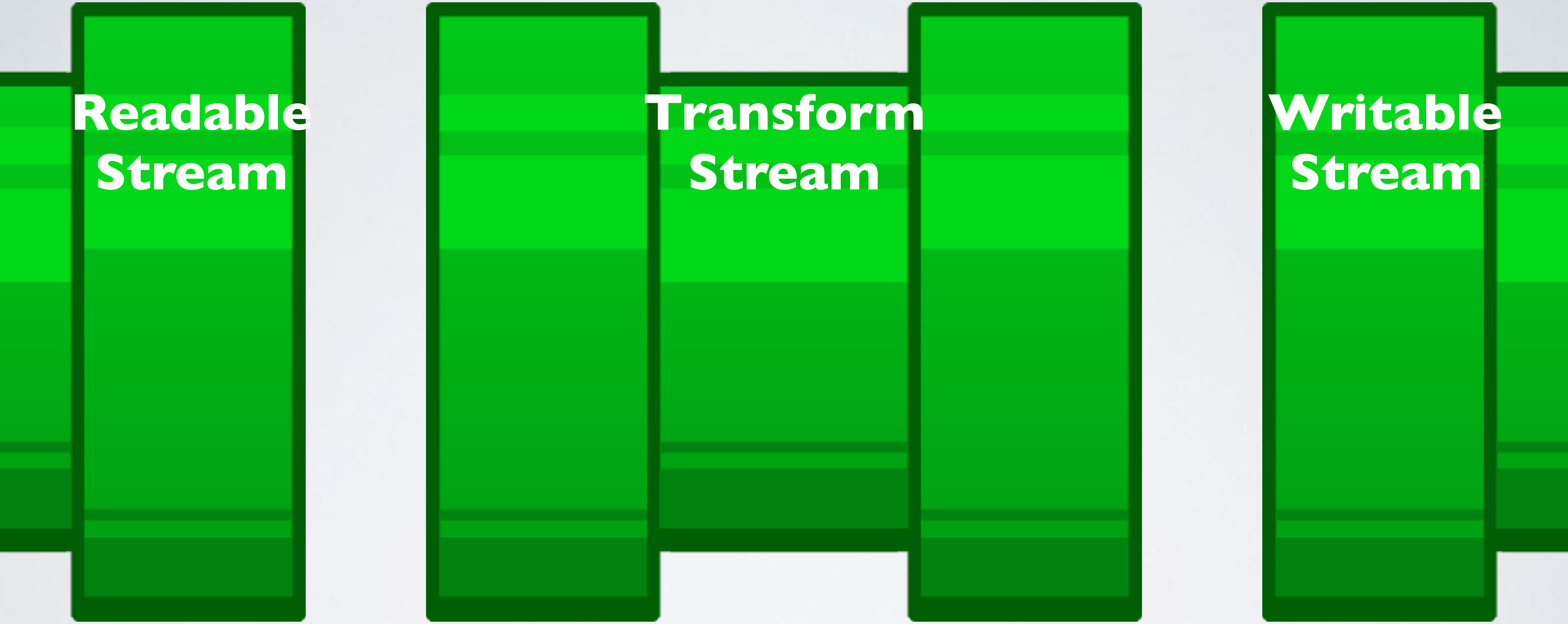
- Involved in biological web projects that need JS
- Had to find and get TB of data online

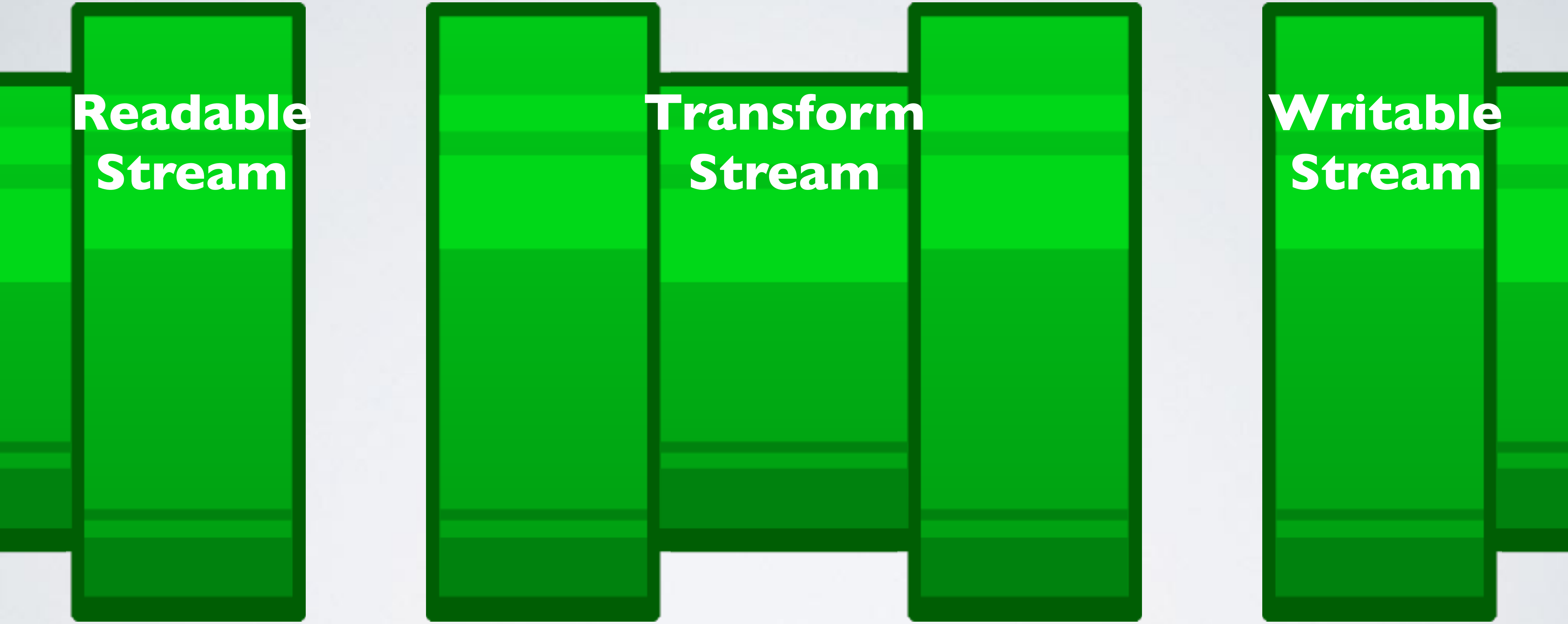
ORIGIN

During my PhD at

Compare genetic diversity of social vs solitary species

- Involved in biological web projects that need JS
- Had to find and get TB of data online
- Had to build complicated bioinformatic pipelines


Readable Stream
Transform Stream
Writable Stream



Readable Stream

Transform Stream

Writable Stream

PROCESS DATA IN CHUNKS

PROCESS DATA IN CHUNKS

fs.createReadStream(file)

request(url)

process.stdin()

PROCESS DATA IN CHUNKS

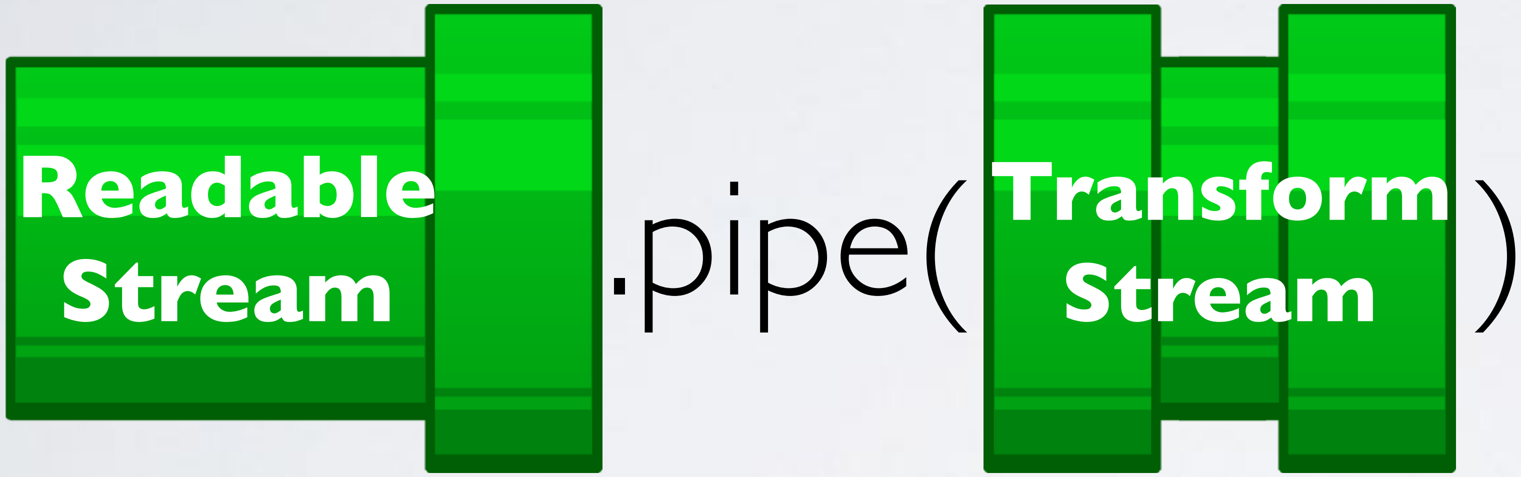
.pipe( )

fs.createReadStream(file)

request(url)

process.stdin()

PROCESS DATA IN CHUNKS



fs.createReadStream(file)

request(url)

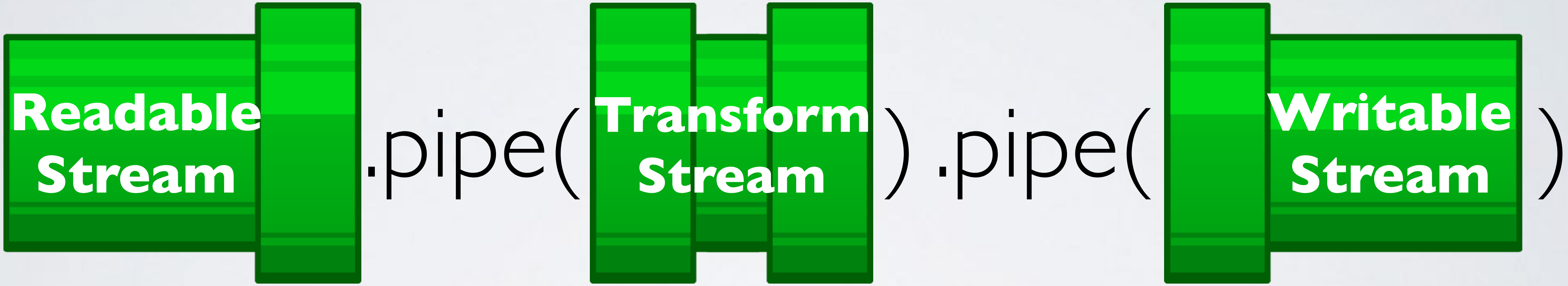
process.stdin()

JSONStream.parse()

filterFunction()

multithreadAnalysis()

PROCESS DATA IN CHUNKS



Readable Stream .pipe(Transform Stream) .pipe(Writable Stream)

fs.createReadStream(file)

request(url)

process.stdin()

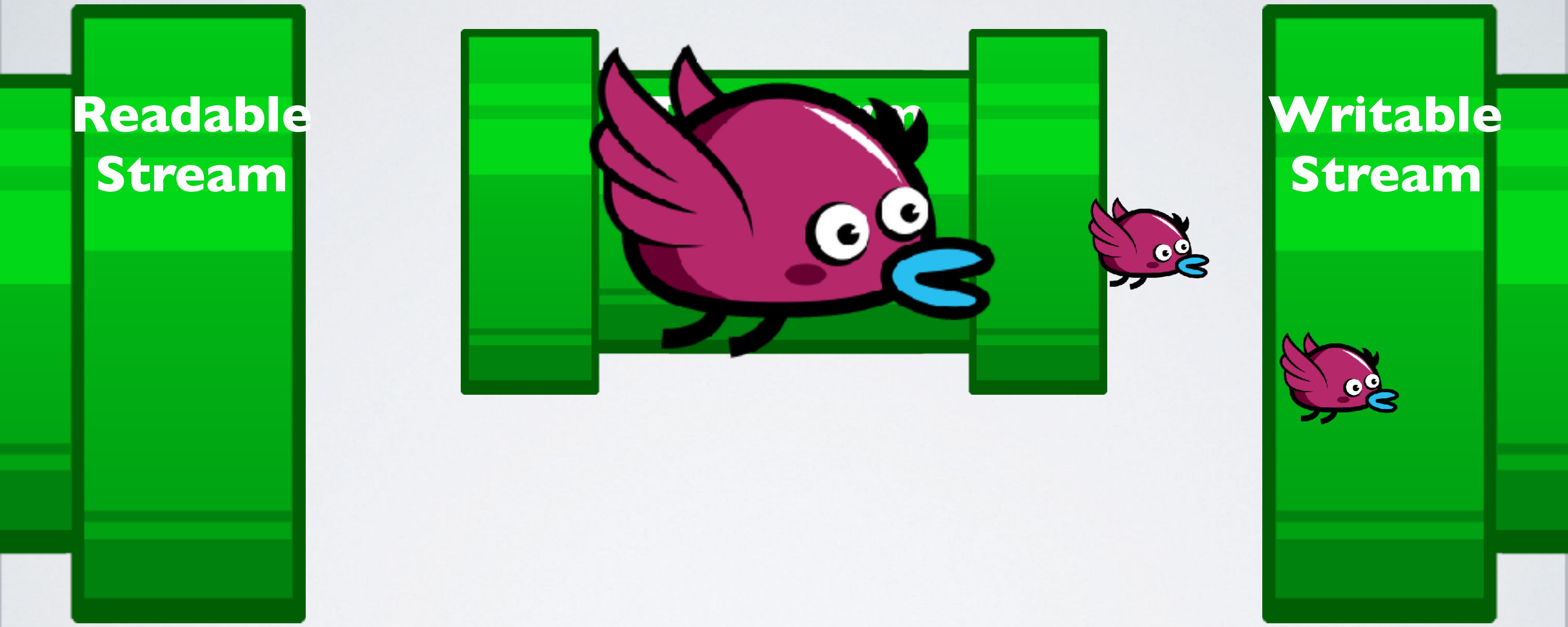
JSONStream.parse()

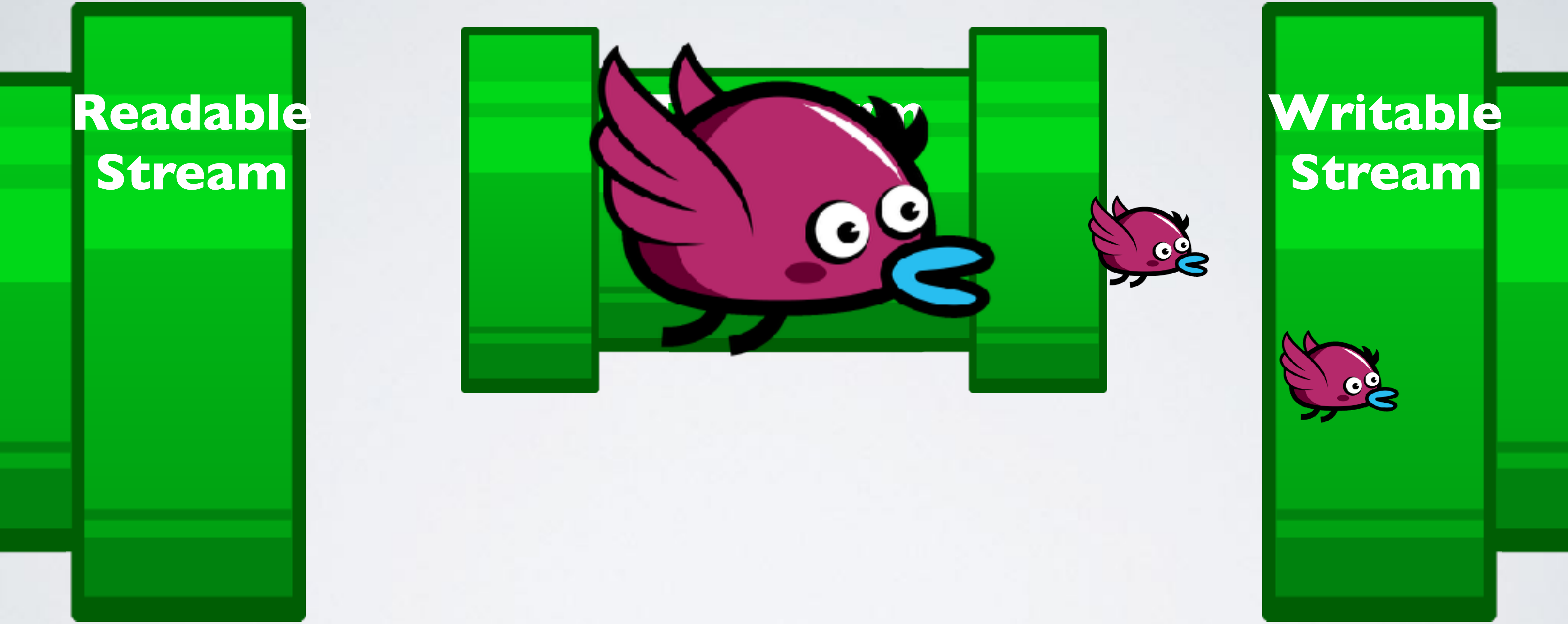
filterFunction()

multithreadAnalysis()

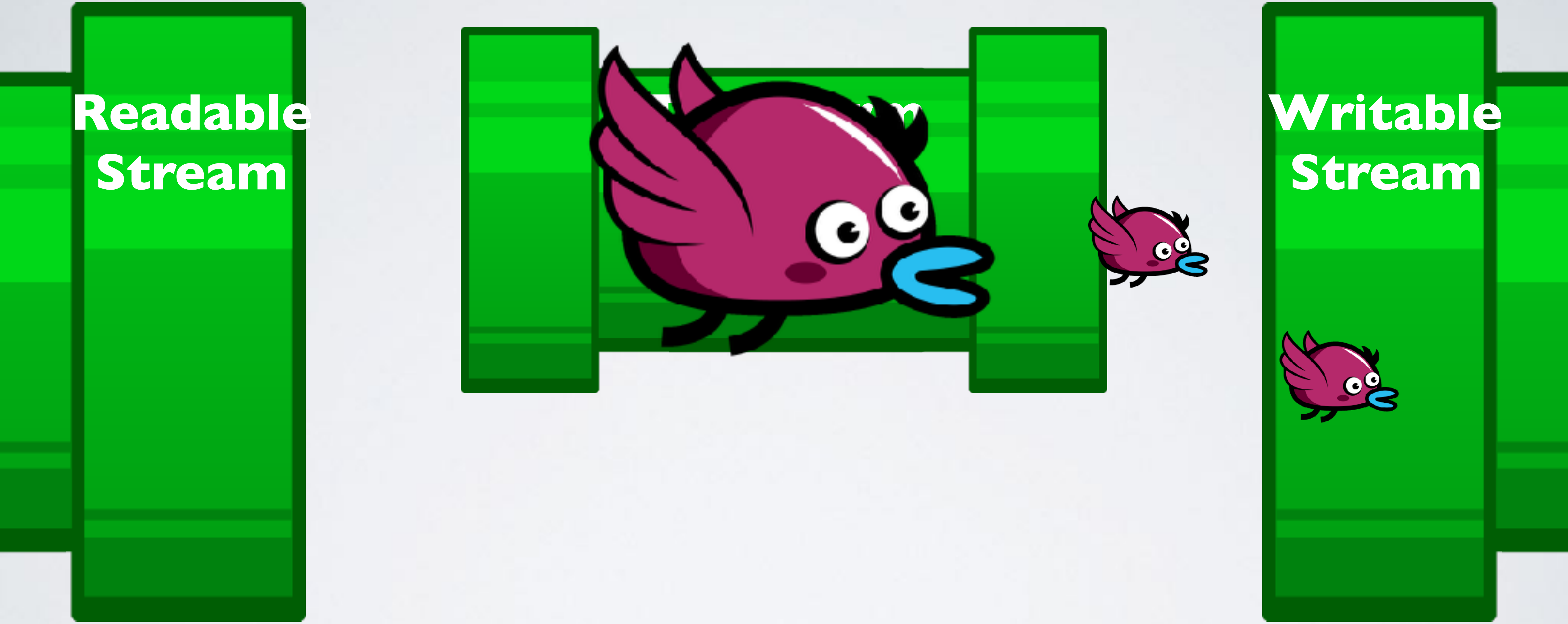
fs.createWriteStream(file)

process.stdout()


Readable Stream
Writable Stream

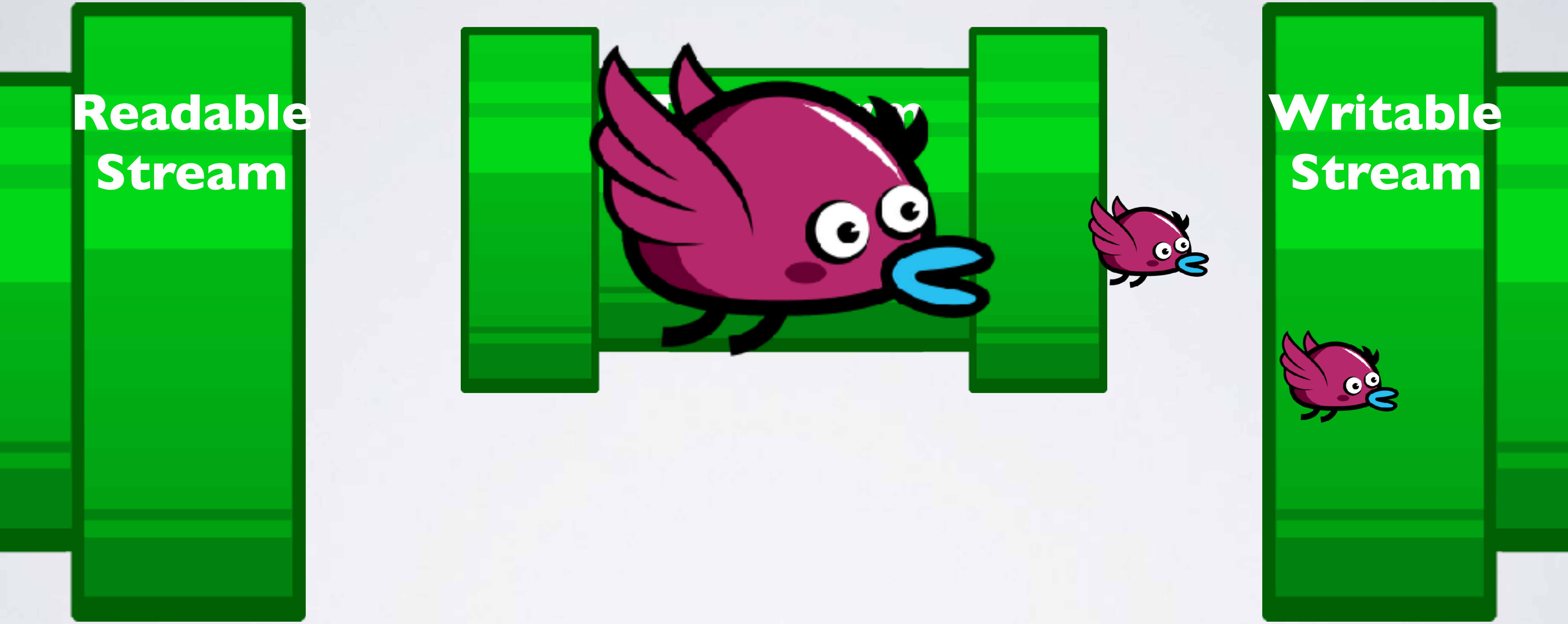


Readable Stream
Writable Stream



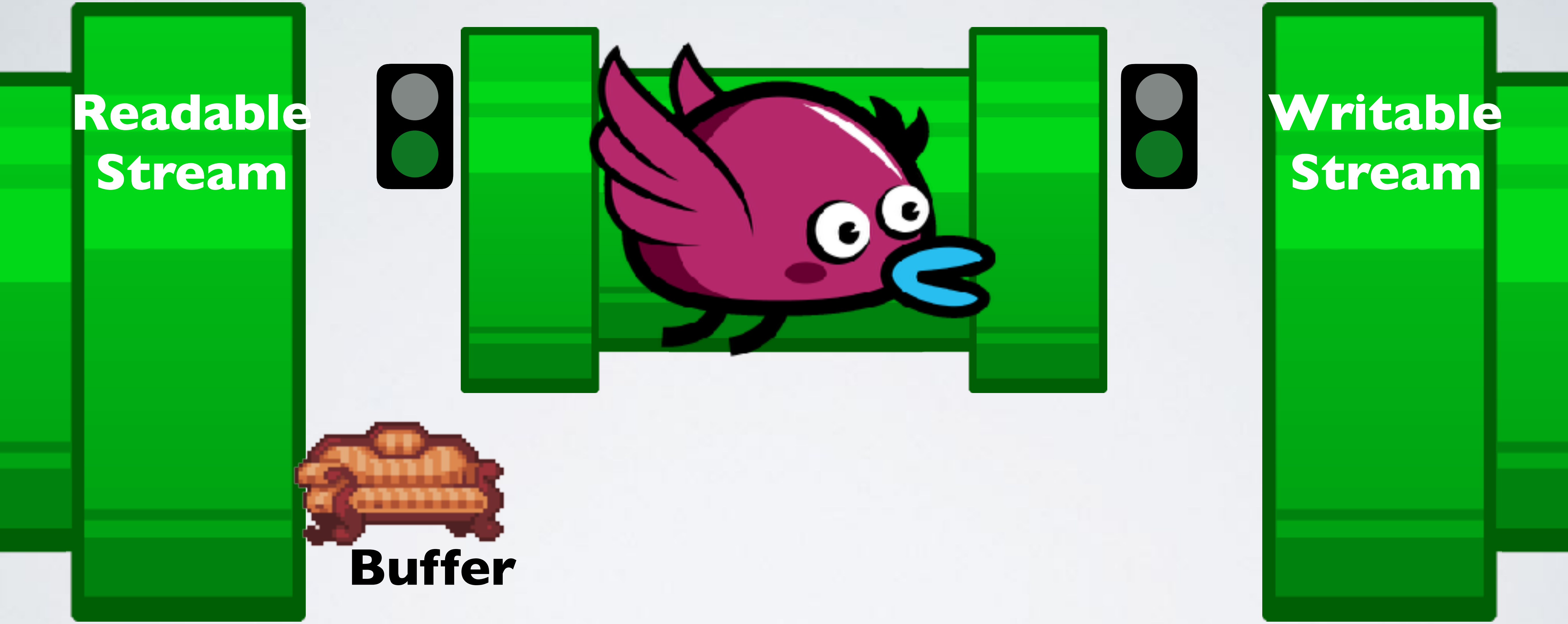
Readable Stream

Writable Stream



Readable Stream

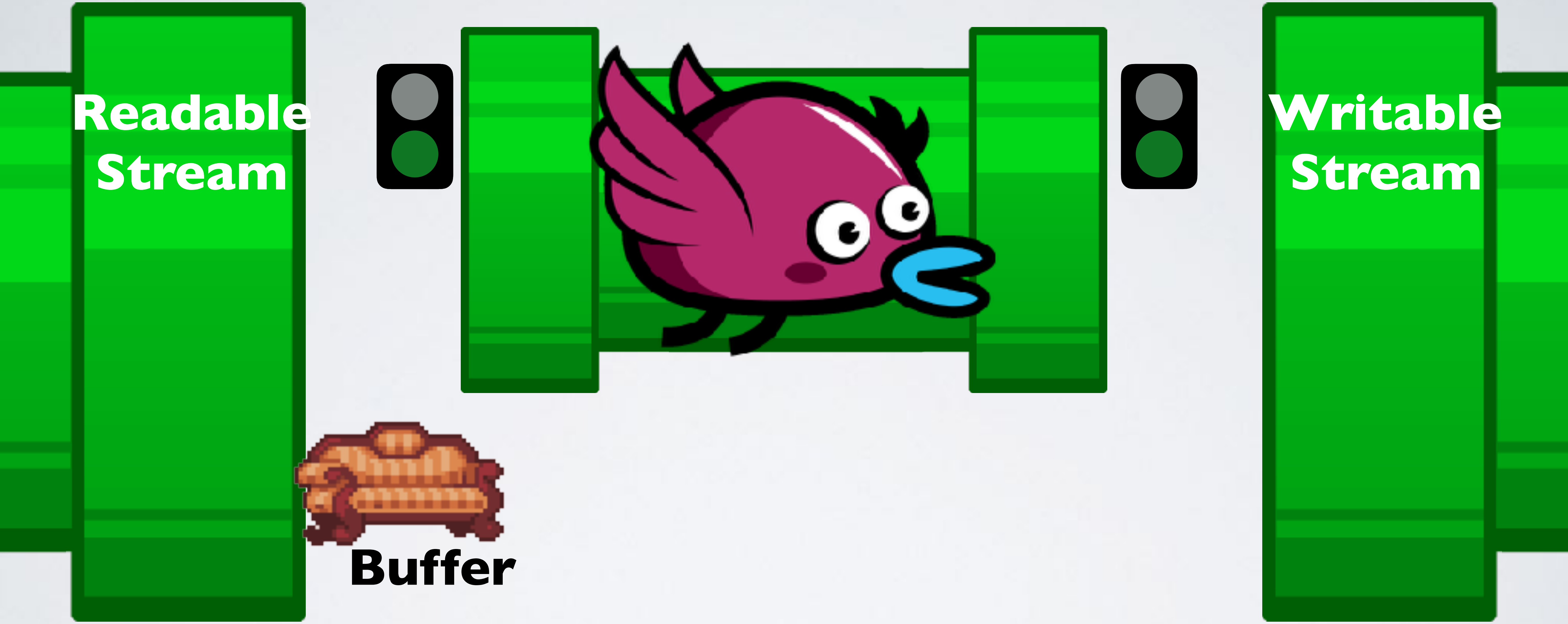
Writable Stream



Readable Stream

Writable Stream

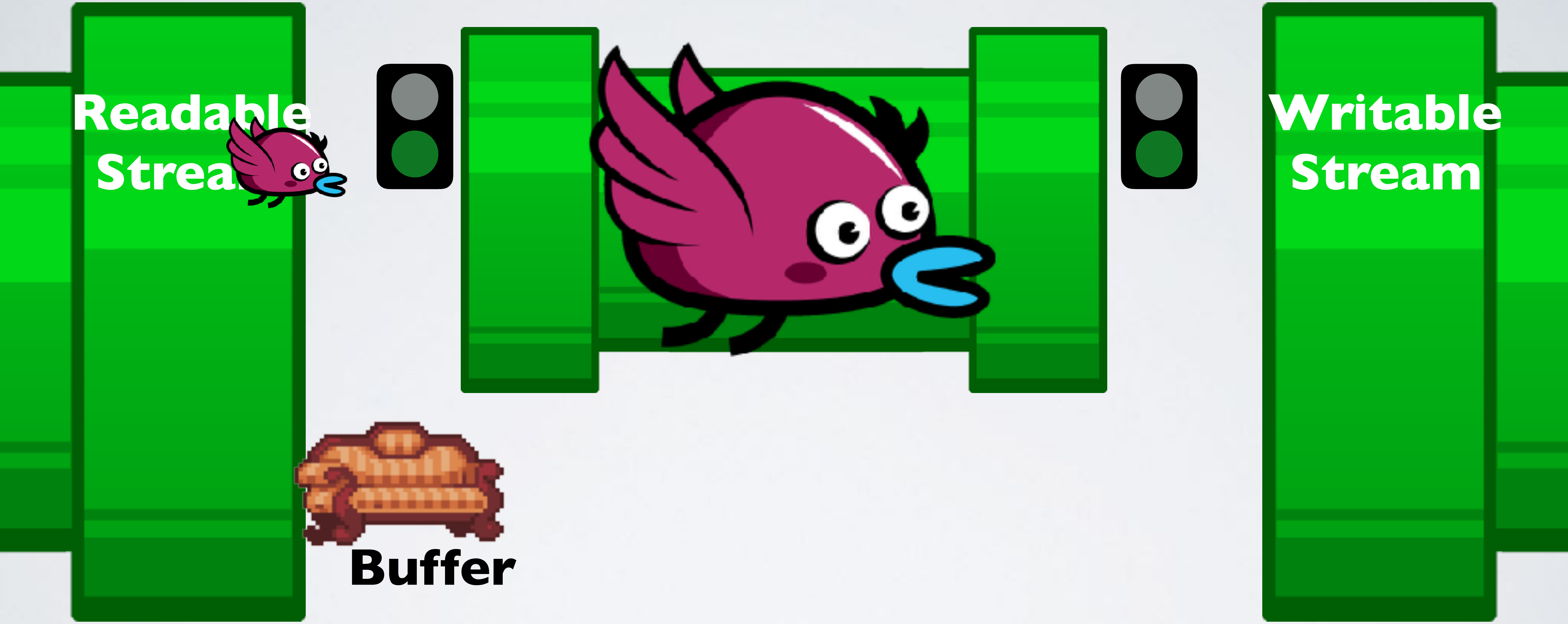
Buffer



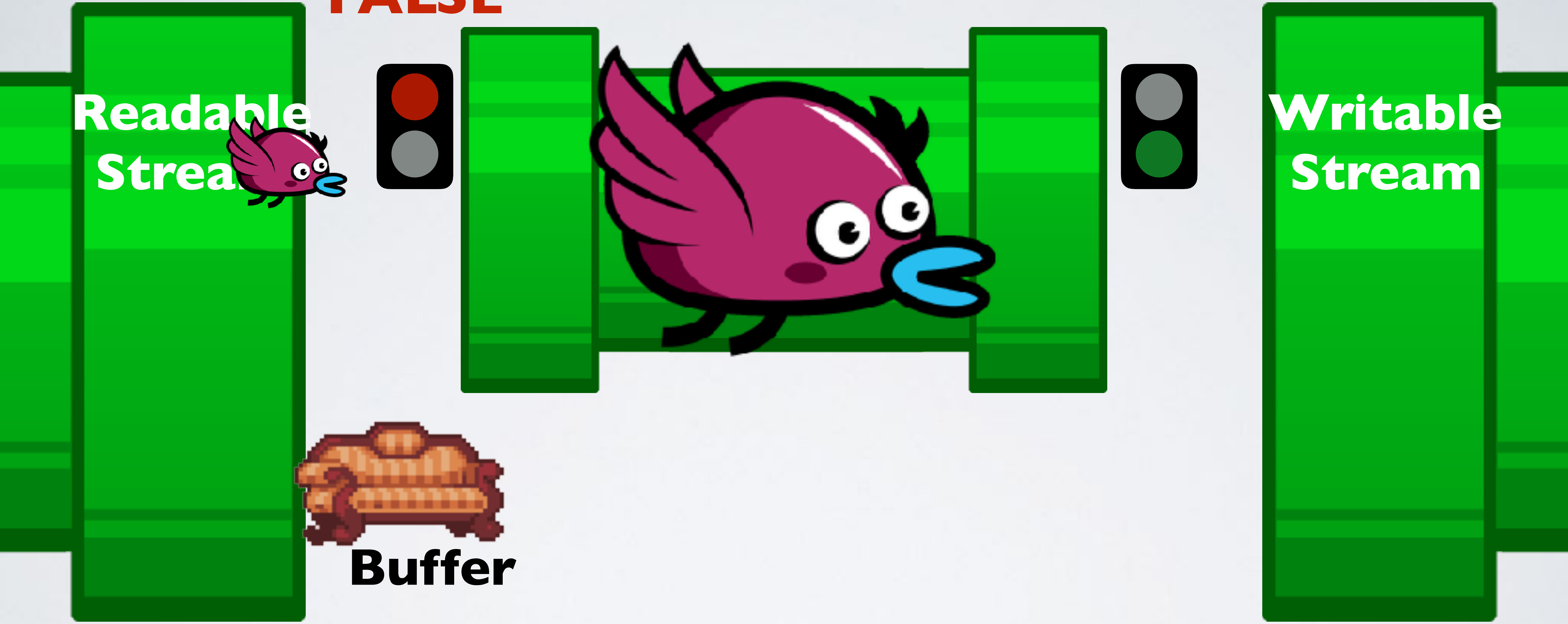
Readable Stream

Writable Stream

Buffer


Readable
Strea
Buffer
Writable
Stream

this.push(data)
FALSE

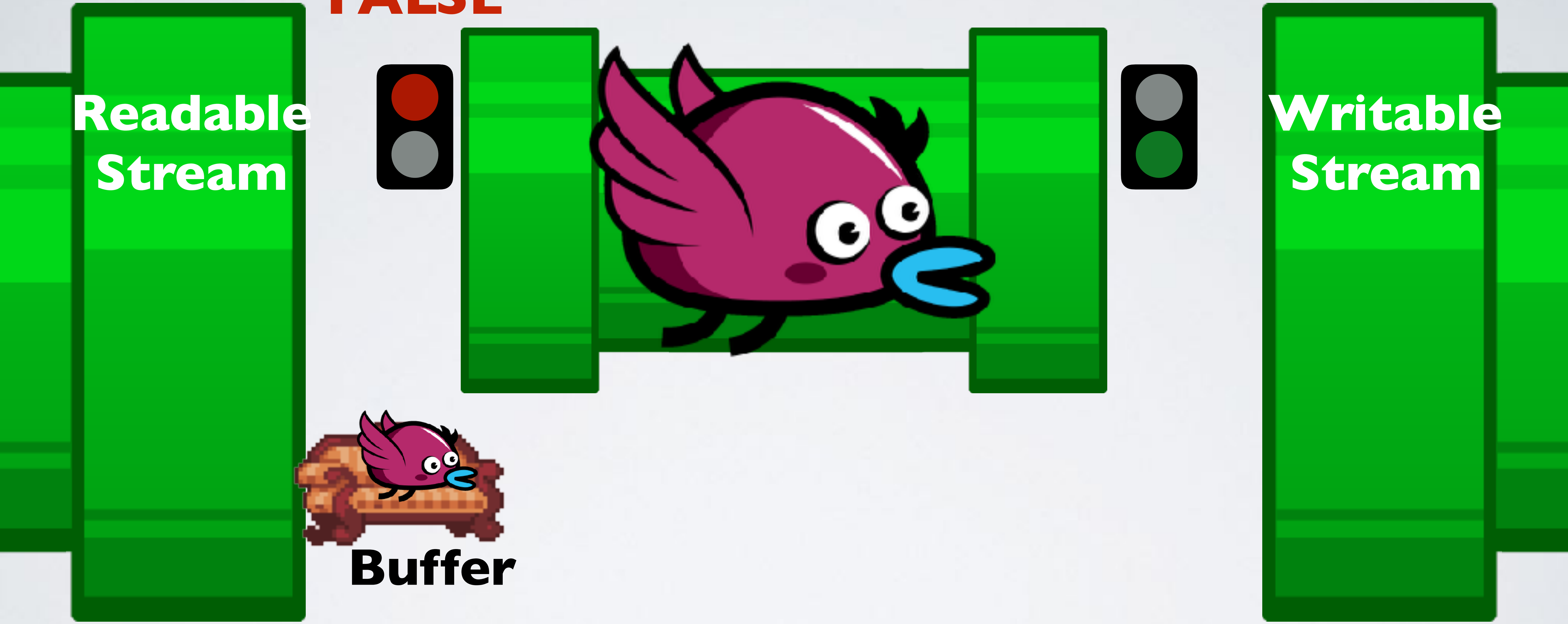


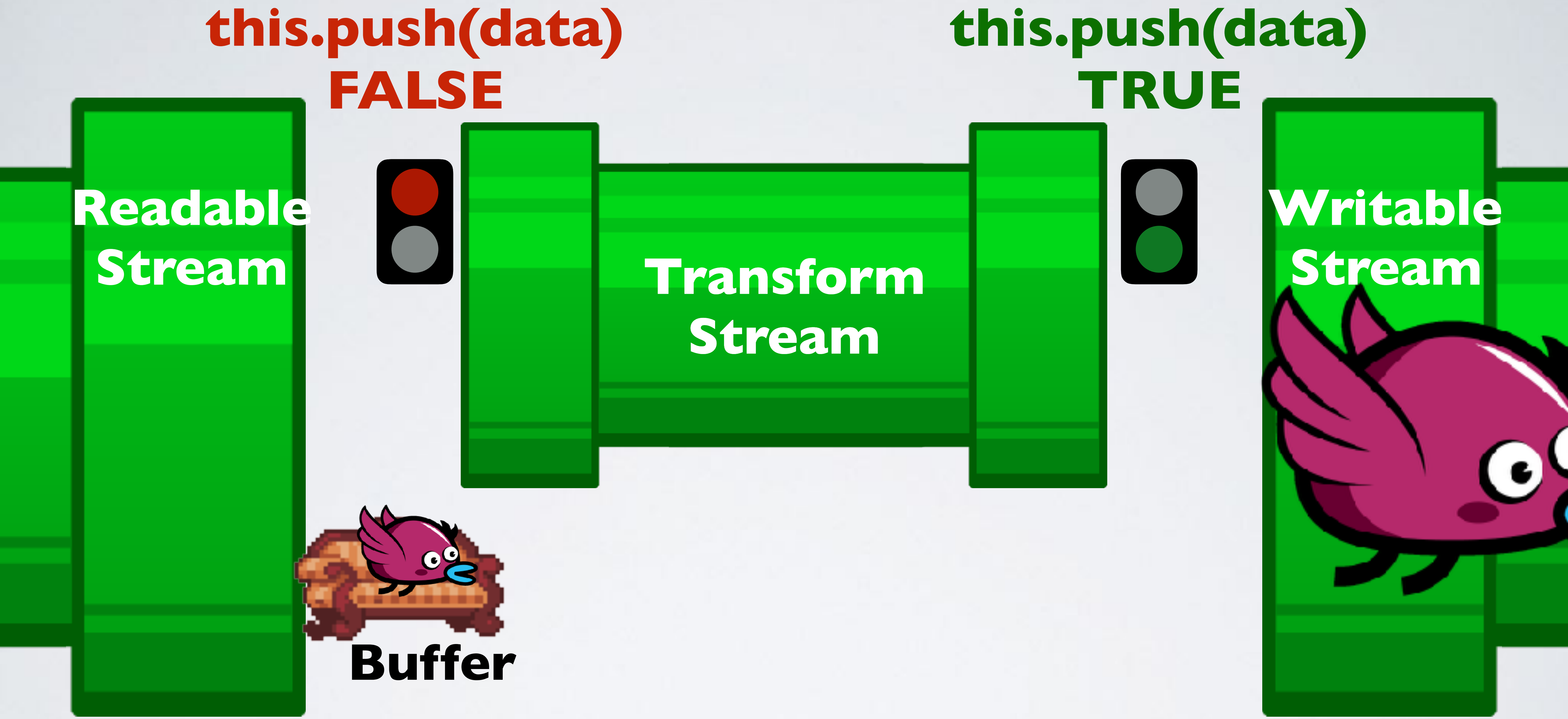
Readable Strea

Writable Stream

Buffer

this.push(data)
FALSE





this.push(data)
FALSE

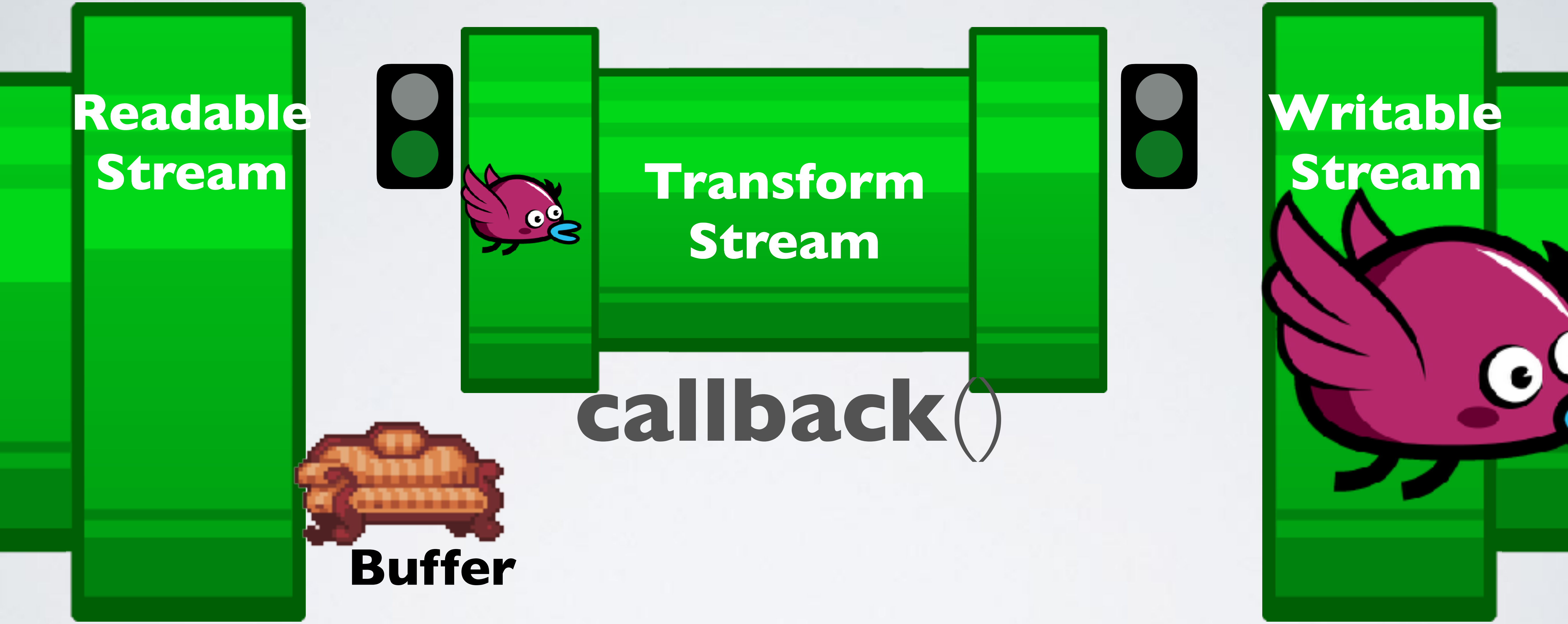
this.push(data)
TRUE

Readable Stream

Transform Stream

Writable Stream

Buffer



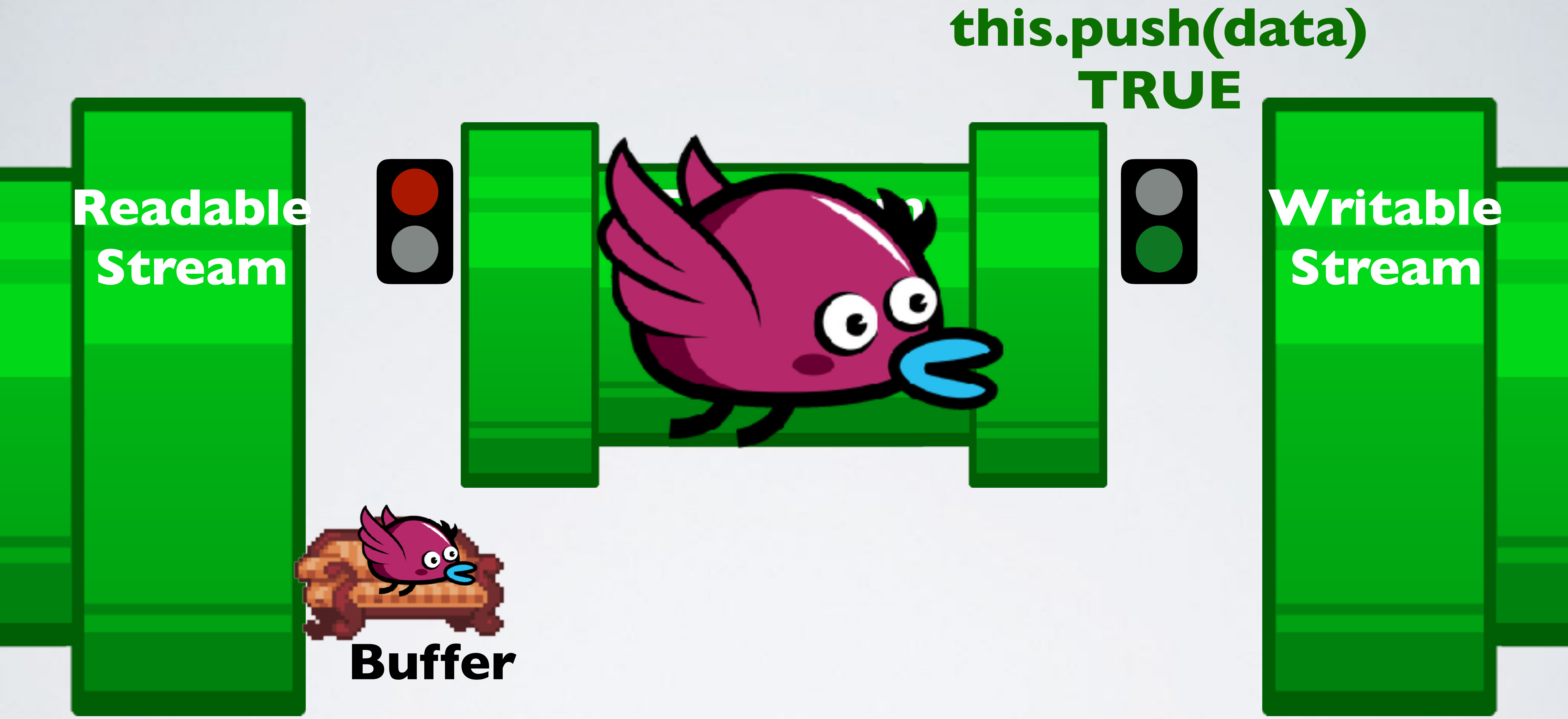
Readable Stream

Transform Stream

Writable Stream

callback()

Buffer



this.push(data)
TRUE
Readable Stream
Writable Stream
Buffer

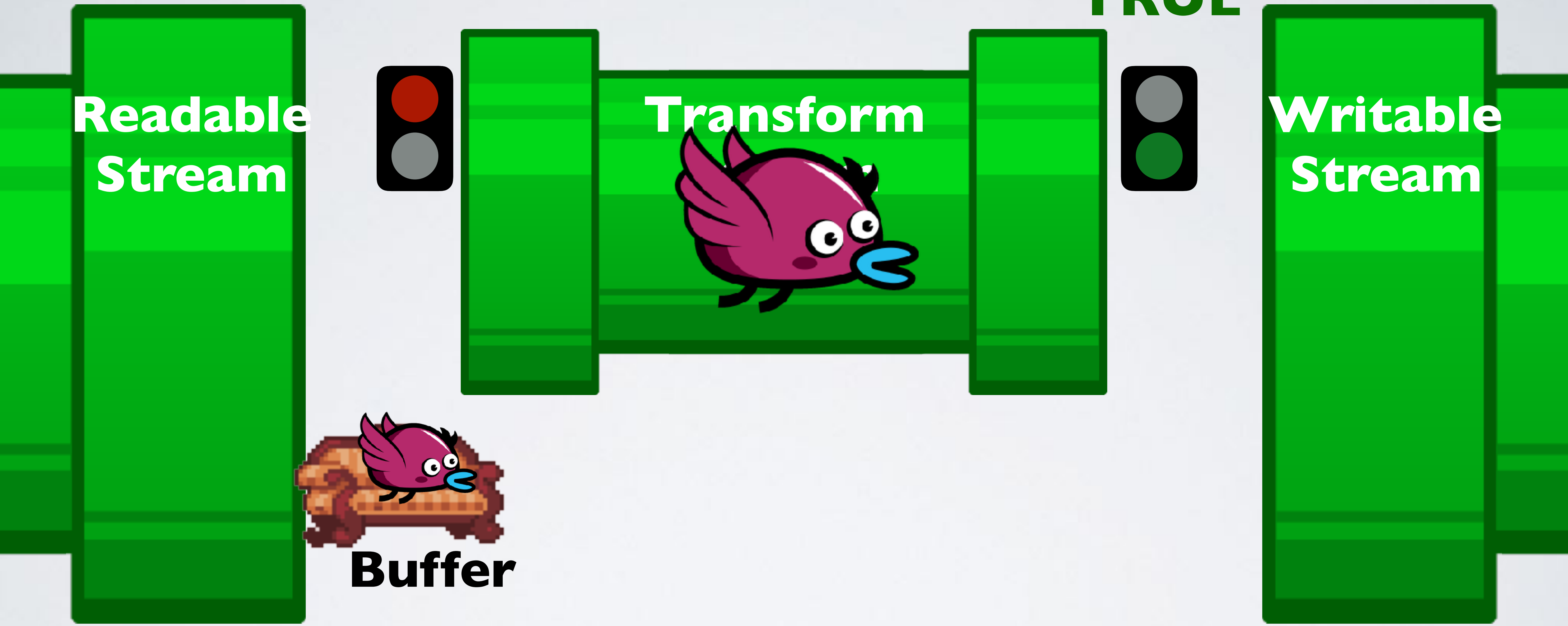
this.push(data)
TRUE

Readable Stream

Writable Stream

Buffer

this.push(data)
TRUE



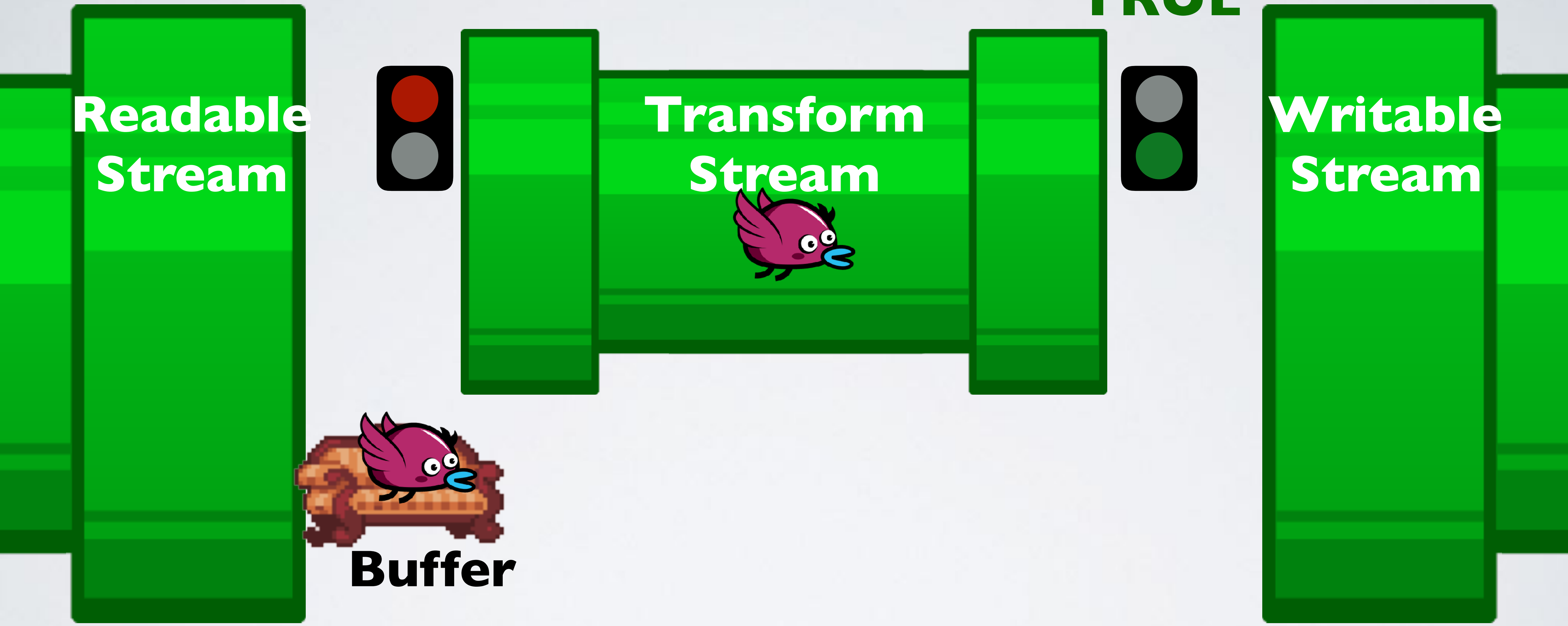
Readable Stream

Transform

Writable Stream

Buffer

this.push(data)
TRUE

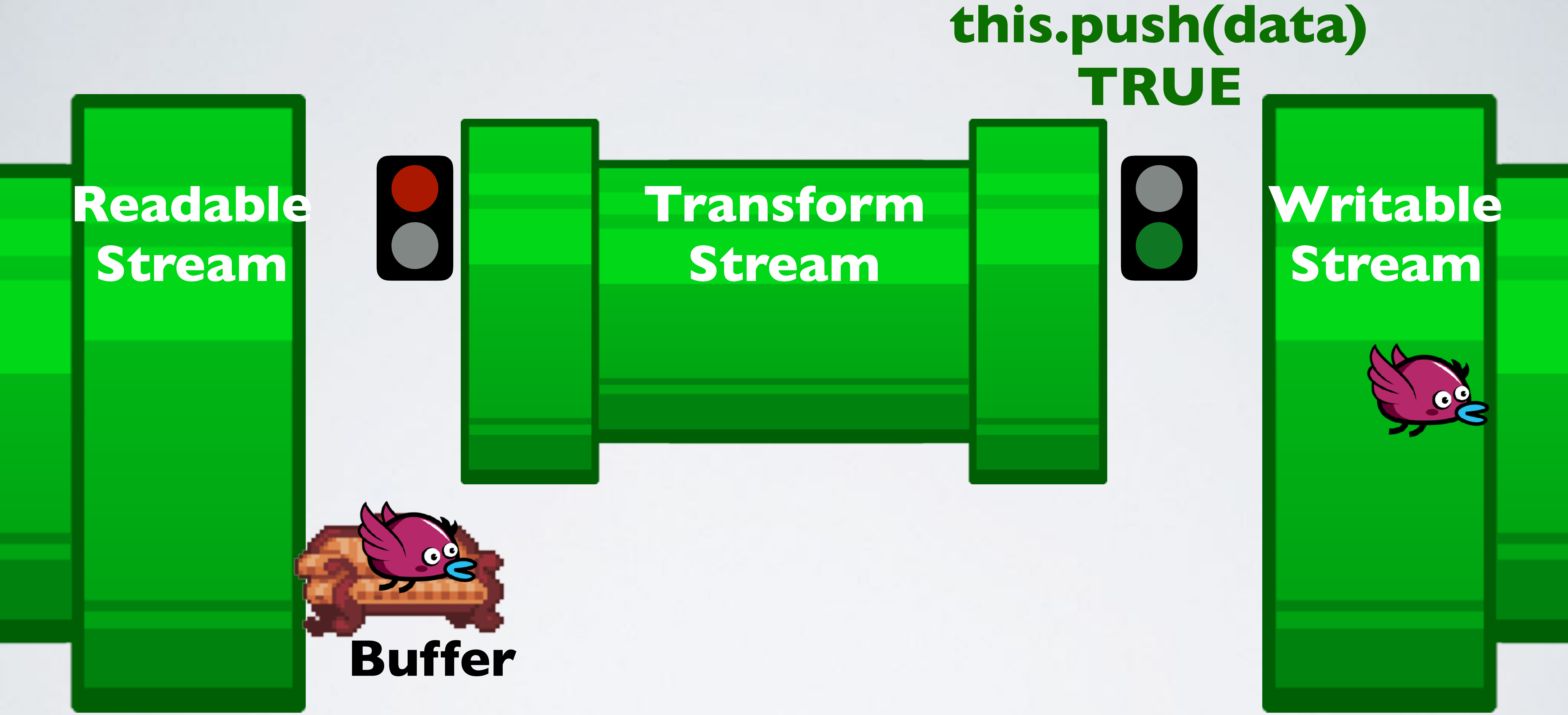


Readable Stream

Transform Stream

Writable Stream

Buffer



this.push(data)
TRUE

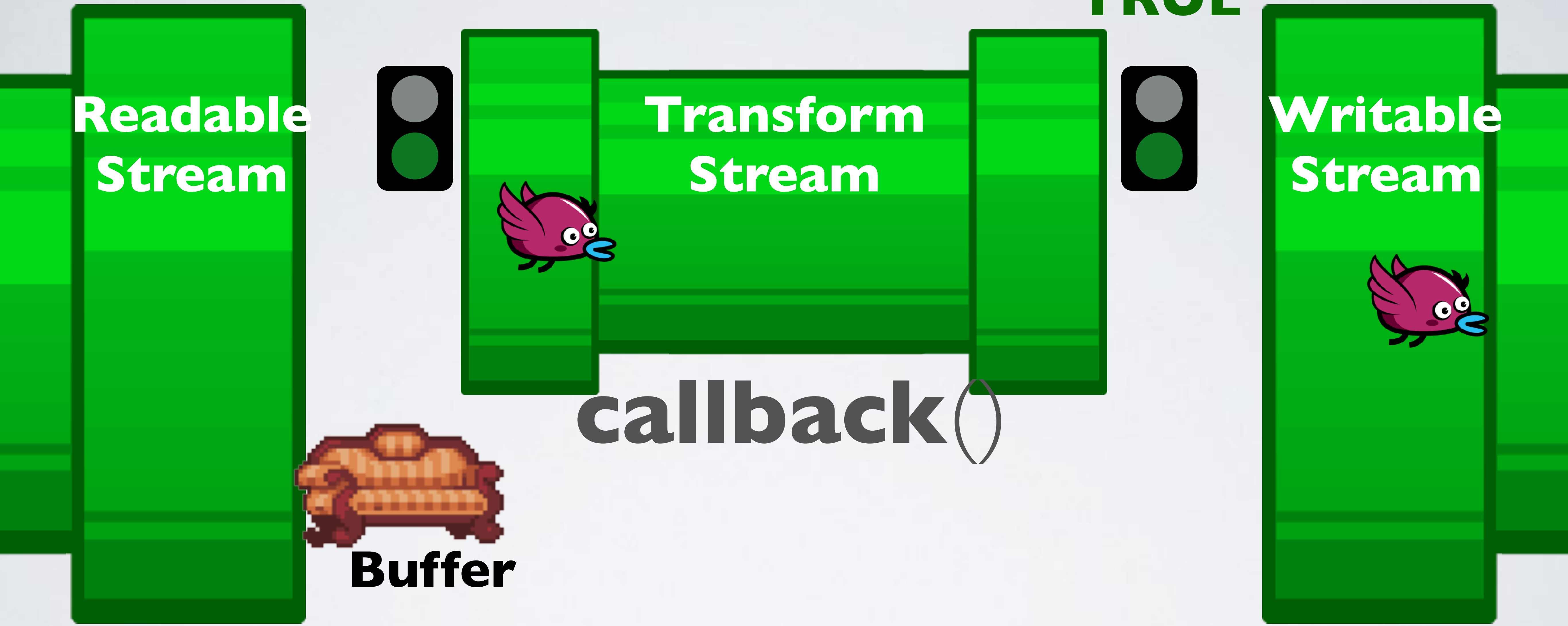
Readable Stream

Transform Stream

Writable Stream

Buffer

this.push(data)
TRUE



Readable Stream

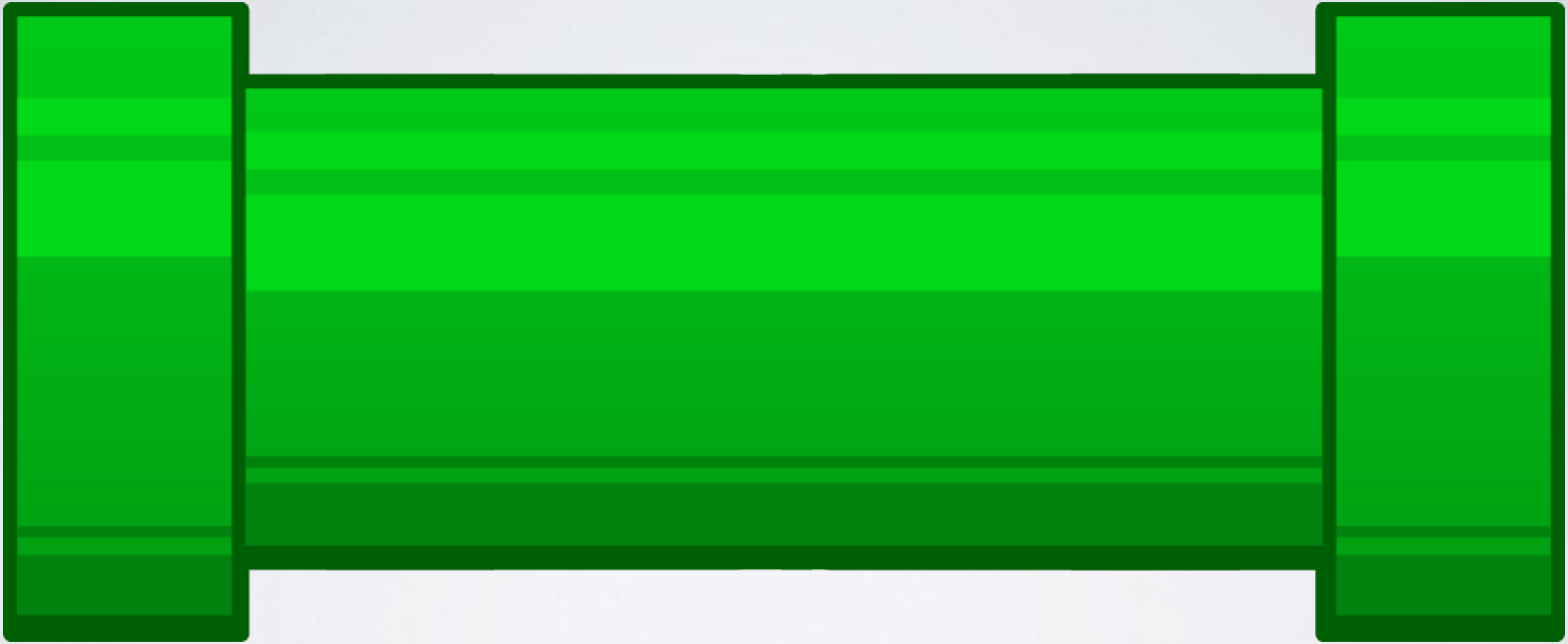
Transform Stream

Writable Stream

callback()

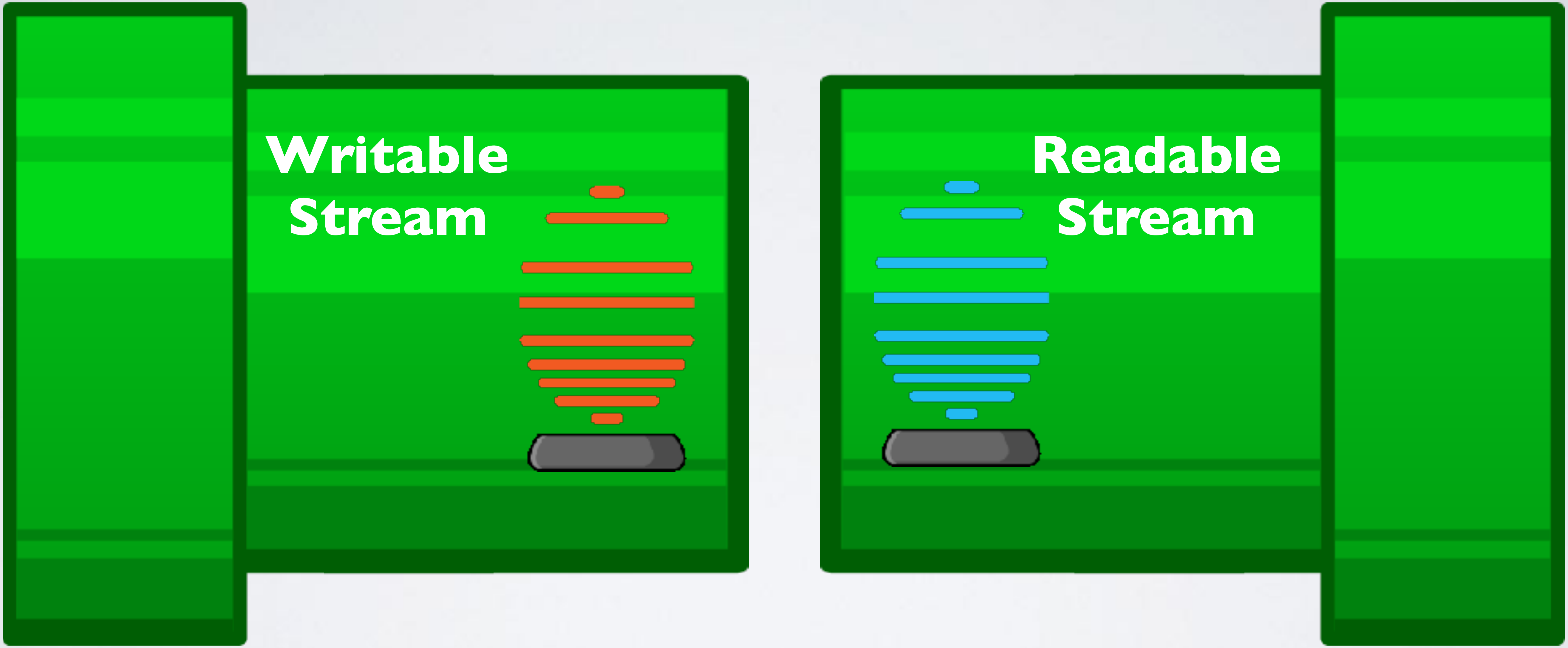
Buffer

DUPLEX STREAM



DUPLEX STREAM

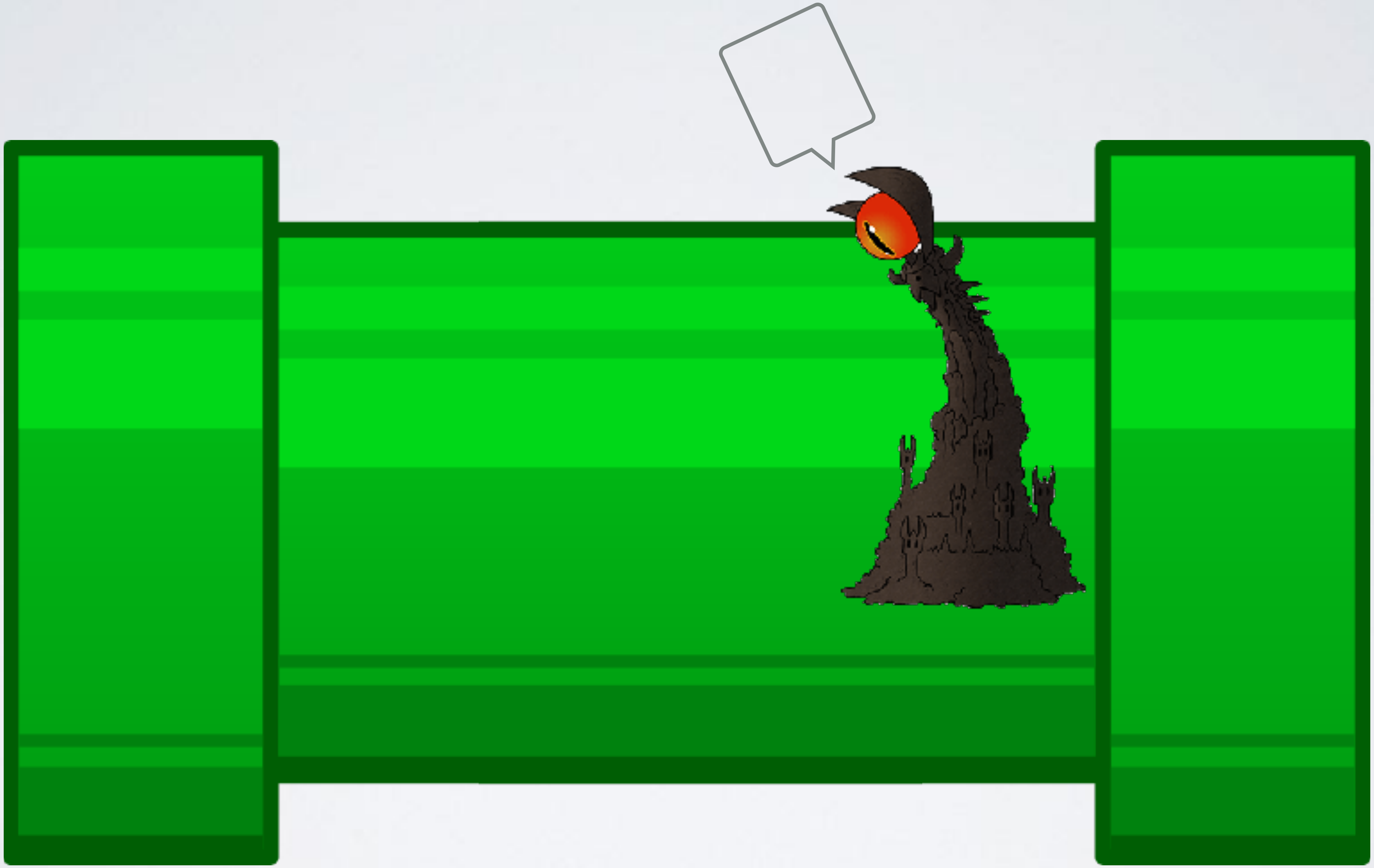
DUPLEX STREAM



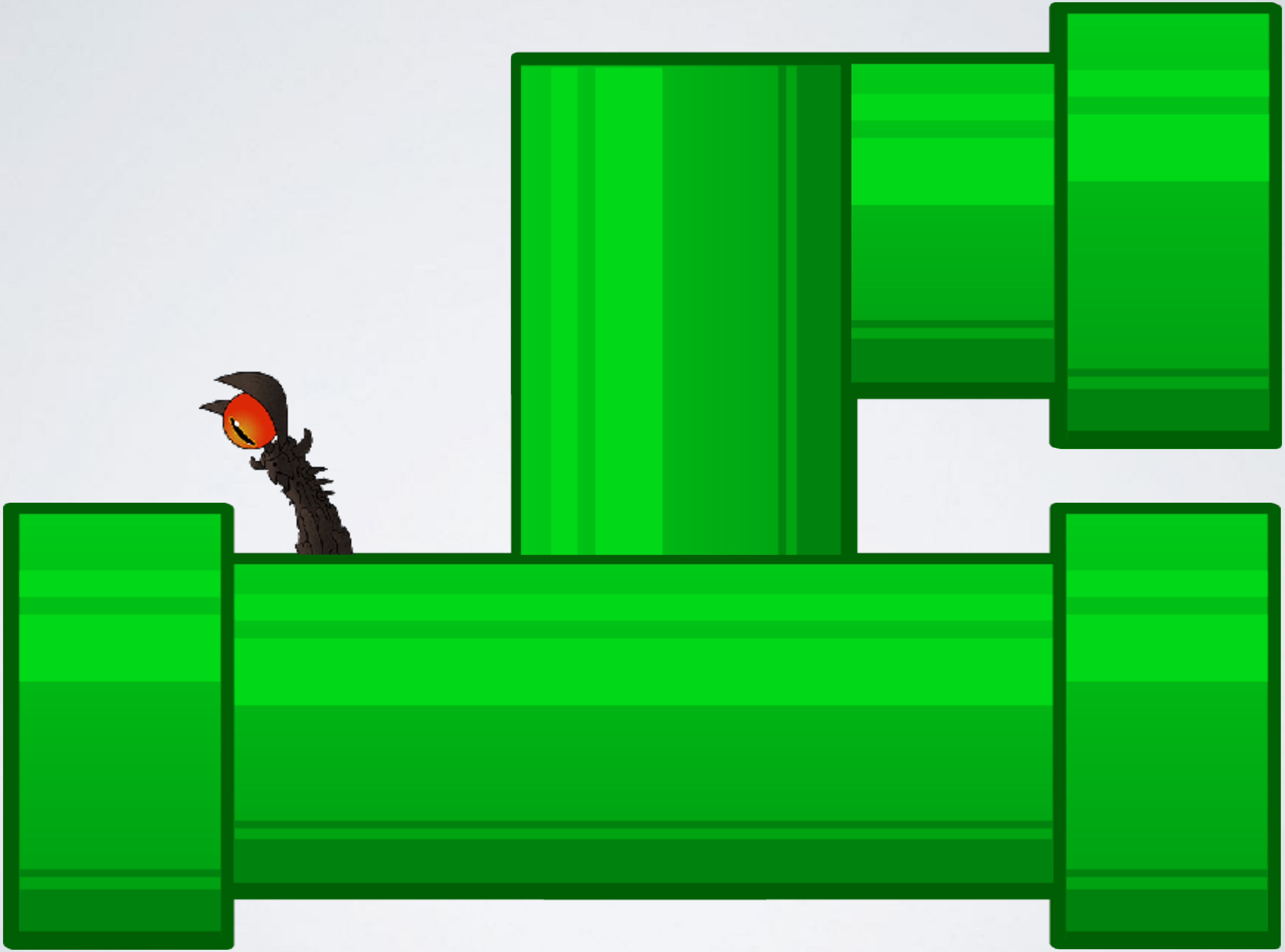
PASSTHROUGH STREAM

PASSTHROUGH STREAM

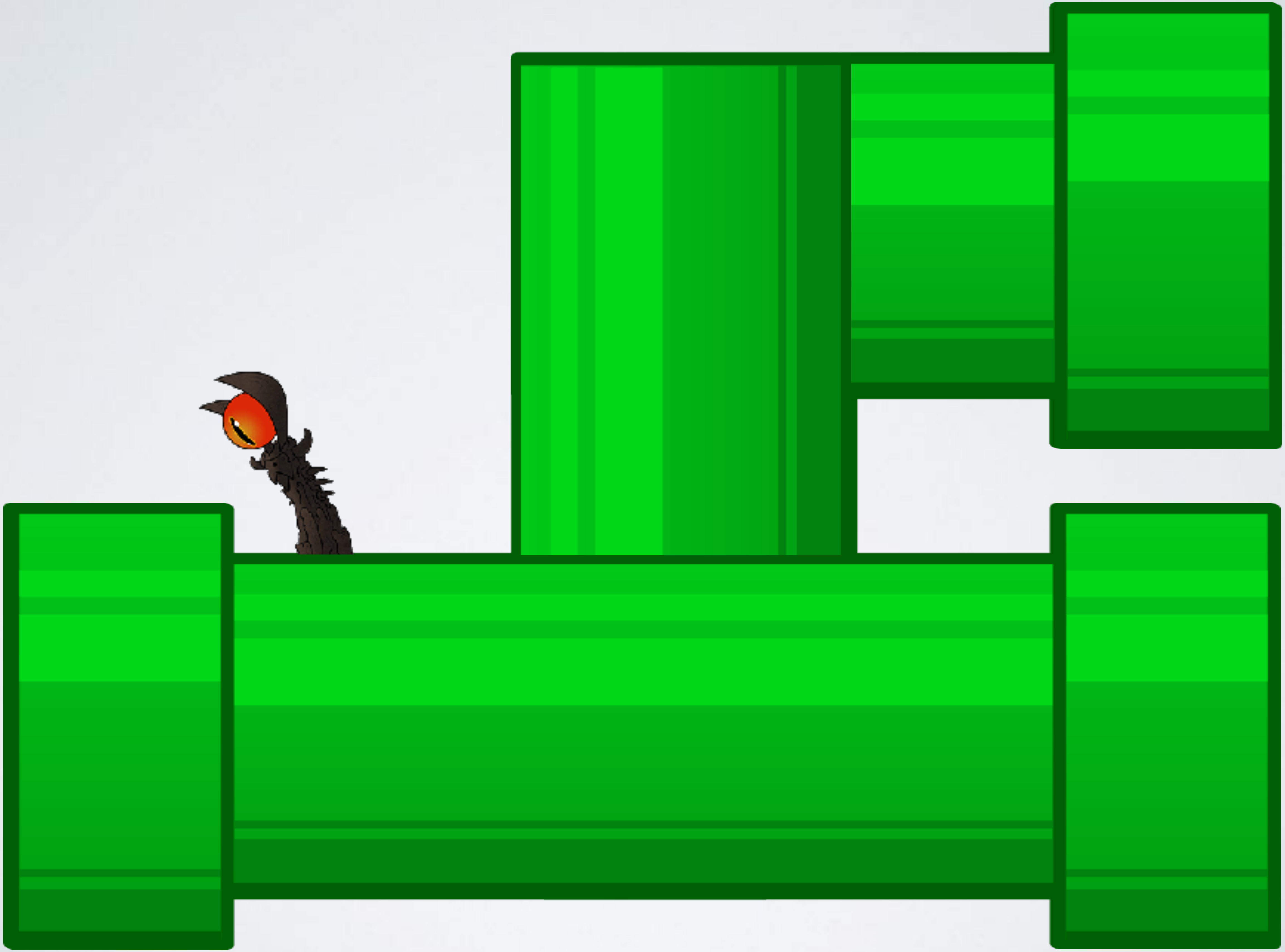
PASSTHROUGH STREAM



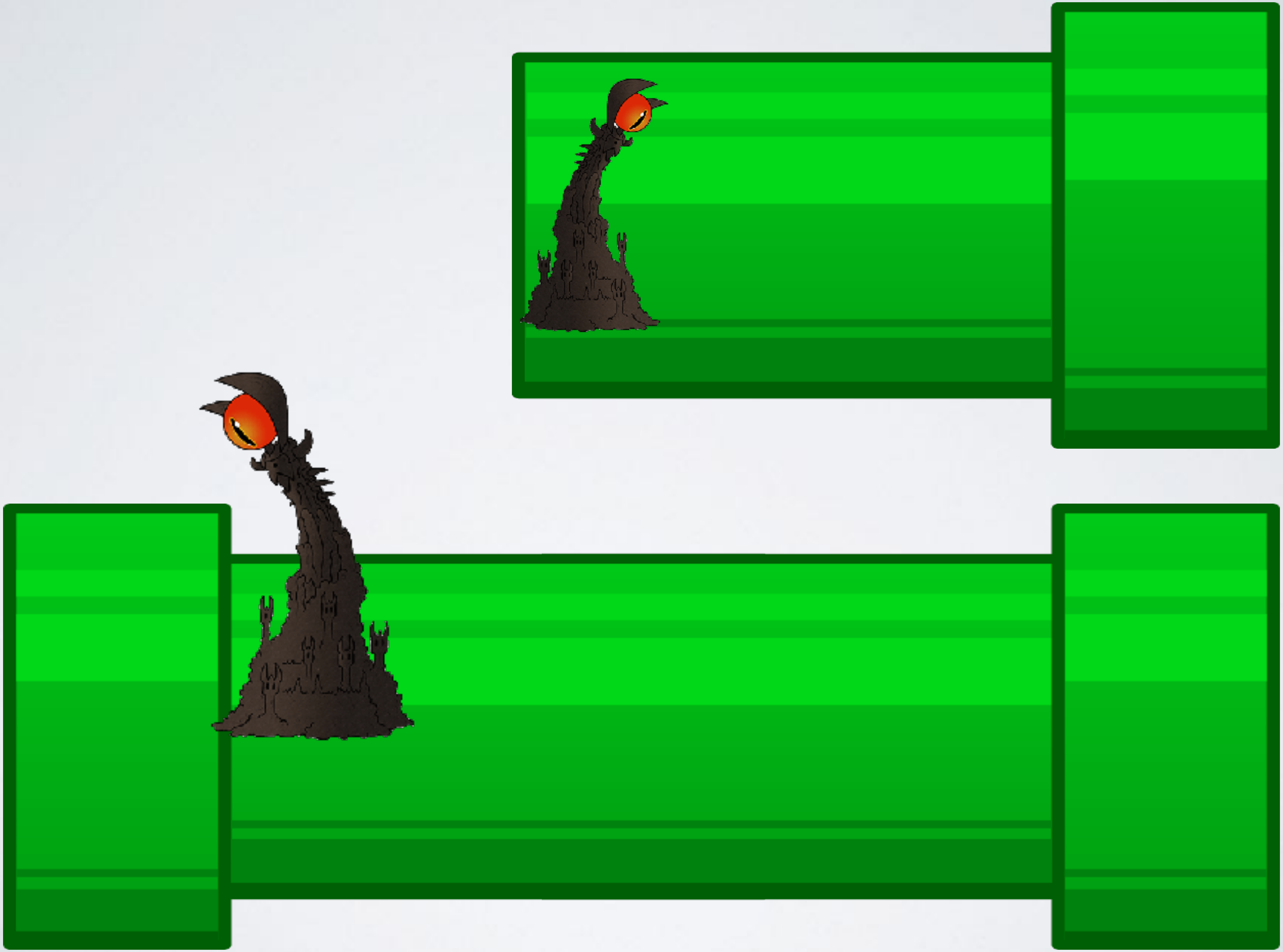
PASSTHROUGH STREAM



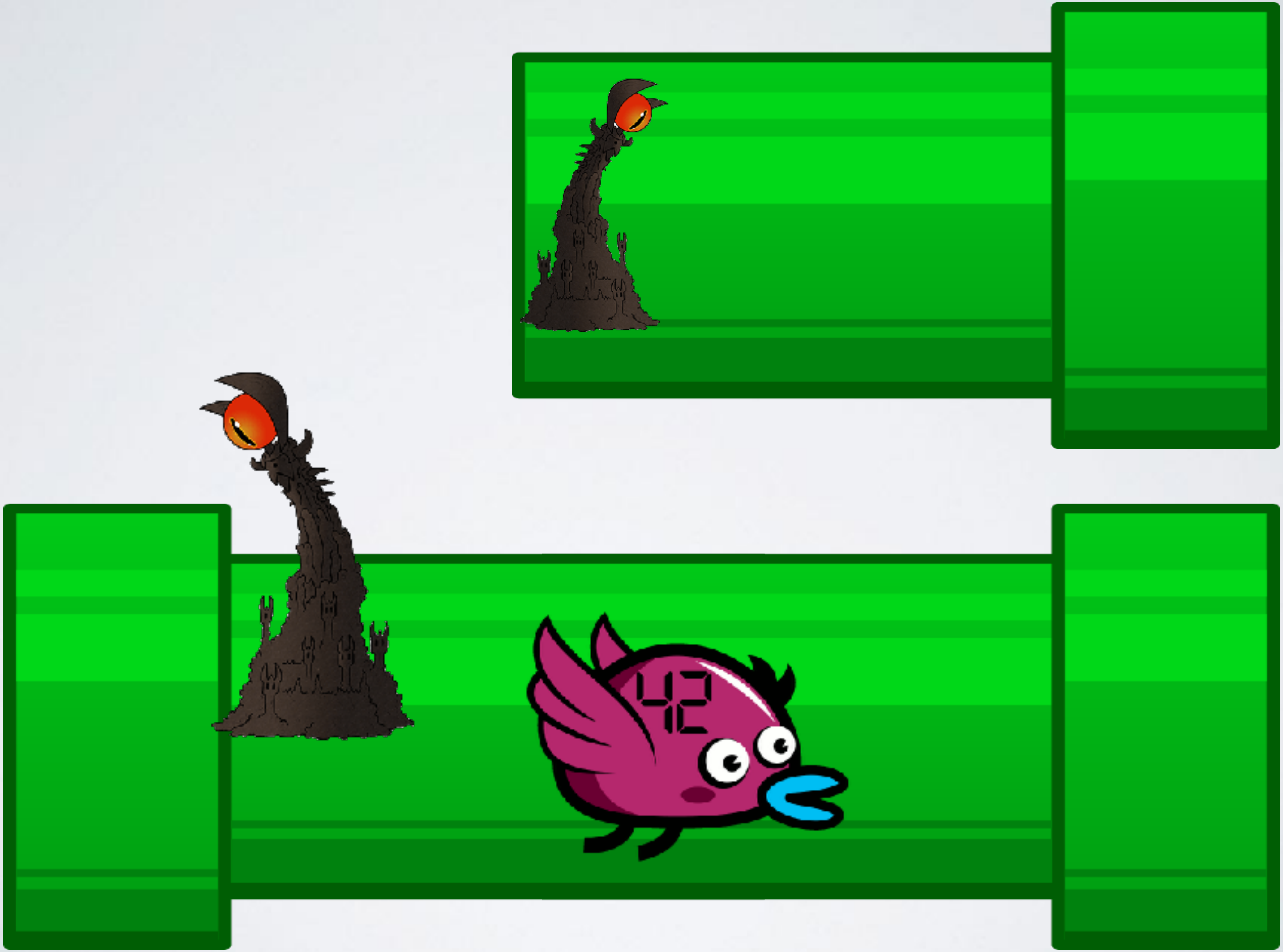
PASSTHROUGH STREAM



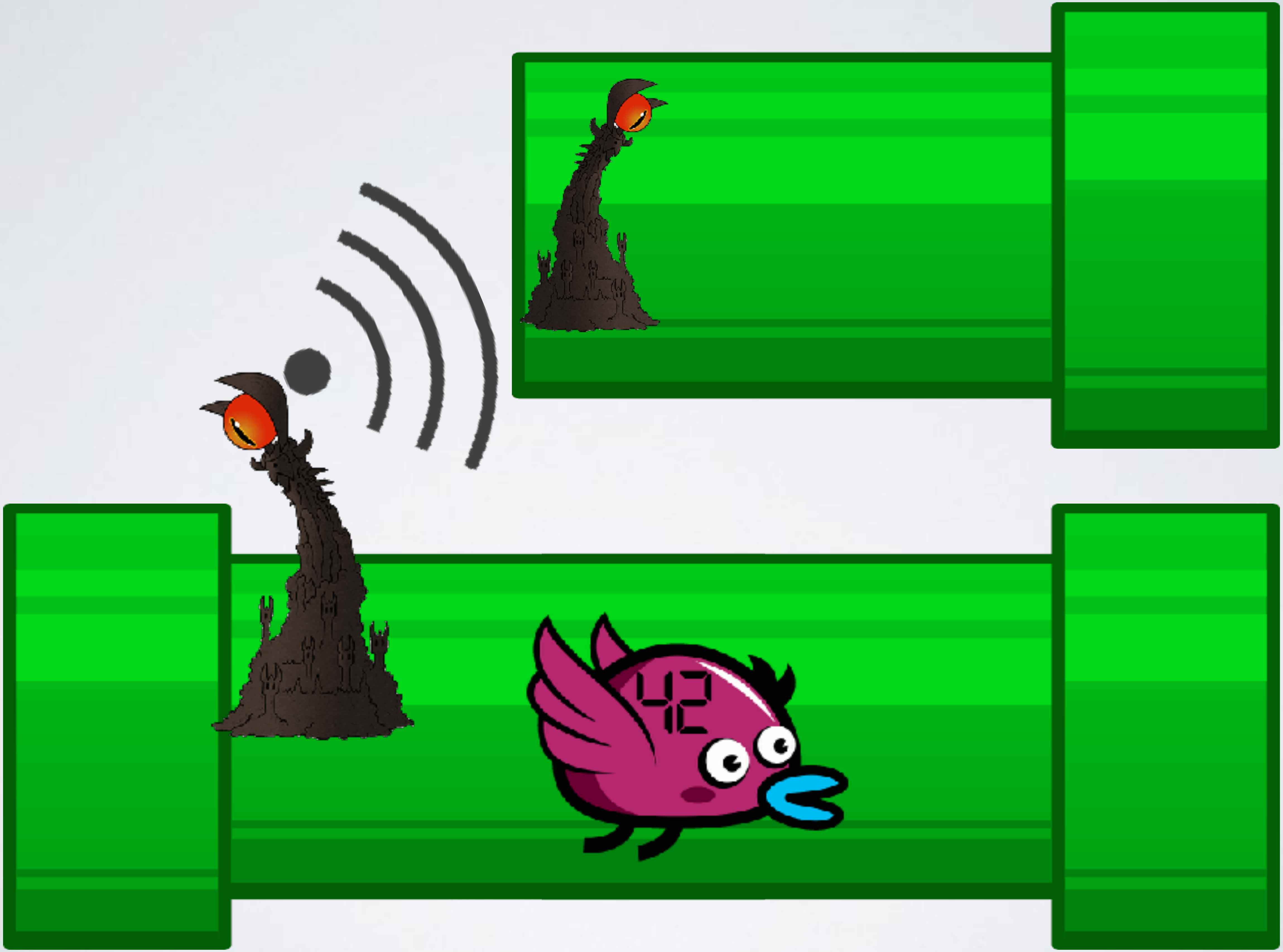
PASSTHROUGH STREAM



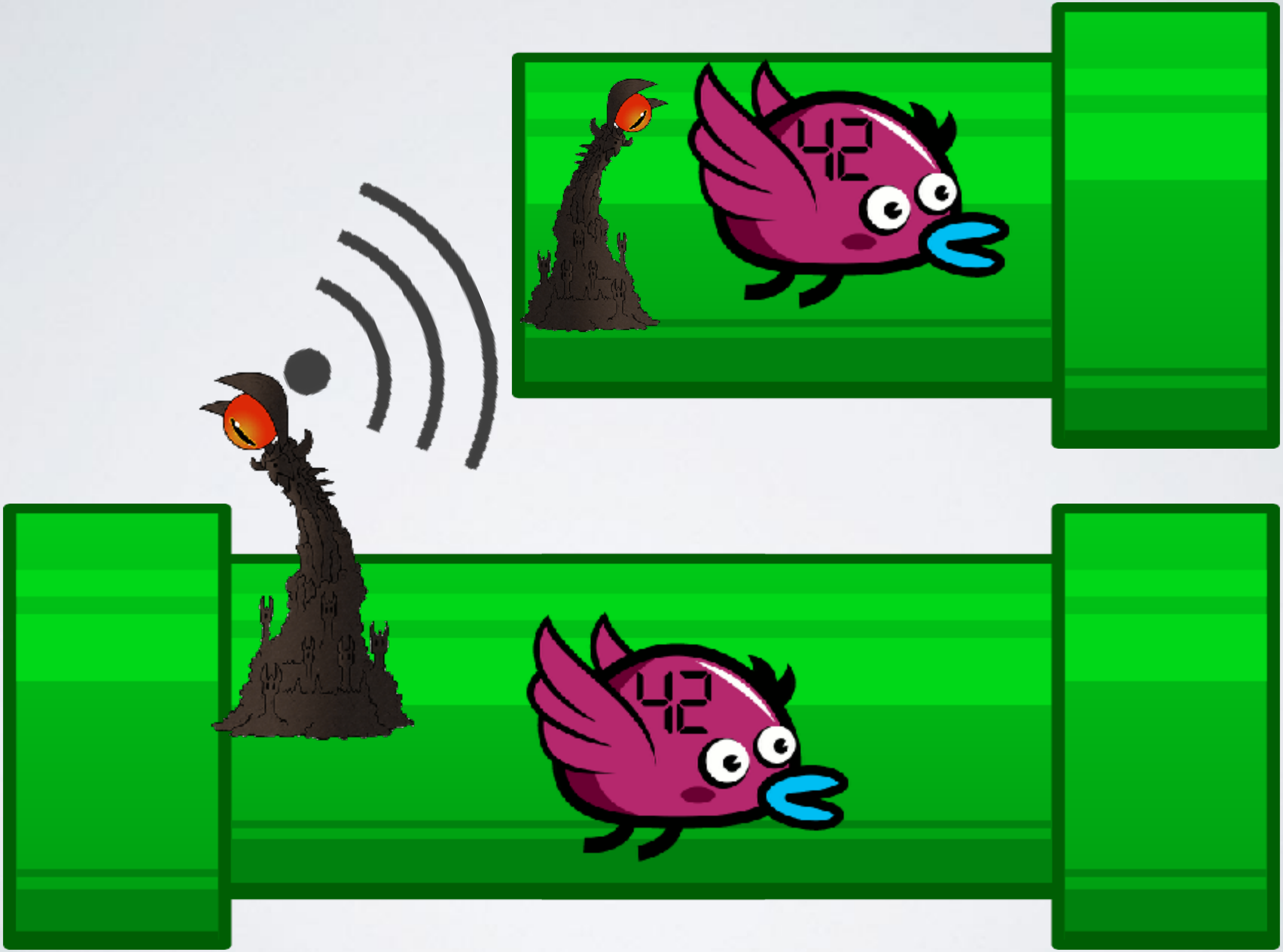
PASSTHROUGH STREAM



PASSTHROUGH STREAM

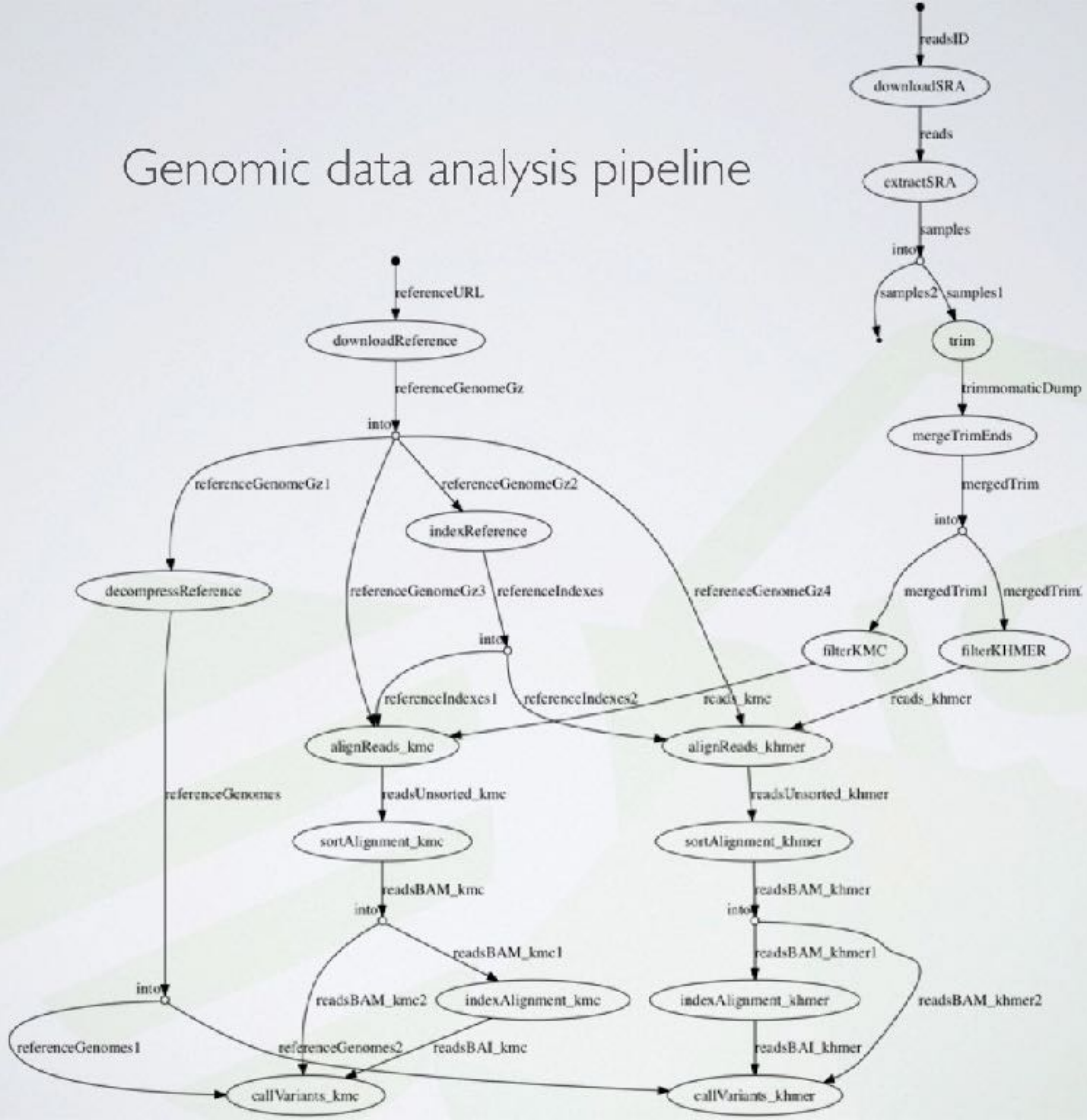


PASSTHROUGH STREAM



PASSTHROUGH STREAM

Genomic data analysis pipeline



bionode-watermill

A Streaming Workflow Engine for bioinformatics and other big data explorations

Google Summer of Code 2016

OBF

```
const samples = task({
  input: {
    db: 'sra',
    accession: config.sraAccession
  },
  output: '**/*.sra'
}, ({ input }) => ncbi.download(input.db, input.accession) )

const fastqDump = task({
  input: new File('**/*.sra'),
  output: [1, 2].map(n => new File(`*_${n}.fastq.gz`))
}, ({ input }) => shell(`fastq-dump --split-files --skip-technical --gzip ${input}`) )
```

BIONODE COMMUNITY

WurmLab.github.io

Queen Mary University of London

DNA digest

repositive.io

dat DATA

BioJS

mozilla Science Lab

- **More self-sustainable**

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background
  - Improve documentation and online tutorials to attract more users

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background
  - Improve documentation and online tutorials to attract more users
  - Apply for grants to hire a few full-time developers to improve and maintain the project

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background
  - Improve documentation and online tutorials to attract more users
  - Apply for grants to hire a few full-time developers to improve and maintain the project
- **More useful**

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background
  - Improve documentation and online tutorials to attract more users
  - Apply for grants to hire a few full-time developers to improve and maintain the project
- **More useful**
  - Reach out to specific labs/institutes to gather their needs and how this project can help them, and what features are missing

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background
  - Improve documentation and online tutorials to attract more users
  - Apply for grants to hire a few full-time developers to improve and maintain the project
- **More useful**
  - Reach out to specific labs/institutes to gather their needs and how this project can help them, and what features are missing
  - Keep improving and working on existing code

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background
  - Improve documentation and online tutorials to attract more users
  - Apply for grants to hire a few full-time developers to improve and maintain the project
- **More useful**
  - Reach out to specific labs/institutes to gather their needs and how this project can help them, and what features are missing
  - Keep improving and working on existing code
- **More collaborative**

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background
  - Improve documentation and online tutorials to attract more users
  - Apply for grants to hire a few full-time developers to improve and maintain the project
- **More useful**
  - Reach out to specific labs/institutes to gather their needs and how this project can help them, and what features are missing
  - Keep improving and working on existing code
- **More collaborative**
  - Use our existing connections to other projects like Dat (data versioning and distribution), BioJS (biological visualisation), CWL (common workflow language) to see if together we can build something amazing for Open Genomics

mozilla Science Lab

- **More self-sustainable**
  - Organise more hackathons to attract contributors, especially with more technical Node/JS background
  - Improve documentation and online tutorials to attract more users
  - Apply for grants to hire a few full-time developers to improve and maintain the project
- **More useful**
  - Reach out to specific labs/institutes to gather their needs and how this project car help them, and what features are missing
  - Keep improving and working on existing code
- **More collaborative**
  - Use our existing connections to other projects like Dat (data versioning and distribution), BioJS (biological visualisation), CWL (common workflow language) to see if together we can build something amazing for Open Genomics
  - Use the Mozilla fellowship to make new collaborations

SUNDAY 11:00

bionode.io

@bmpvieira

ACKNOWLEDGMENTS

Community

bionode.io

Research group

WurmLab
.github.io

Queen Mary
University of London

Funding

THE LEONA M. AND HARRY B.
HELMSLEY
CHARITABLE TRUST

mozilla
Science Lab

Friends

DNA
digest

repositive.io

dat
DATA

BioJS

COMMON
WORKFLOW
LANGUAGE

IMAGES SOURCES

http://opengameart.org/content/bevouliin-green-flappy-bird-sprite-sheets

http://neoriceisgood.deviantart.com/art/100-furniture-sprites-405058884

http://android272.deviantart.com/art/Teeworlds-Teleport-570298308

http://www.how-to-draw-cartoons-online.com/eye-of-sauron.html