

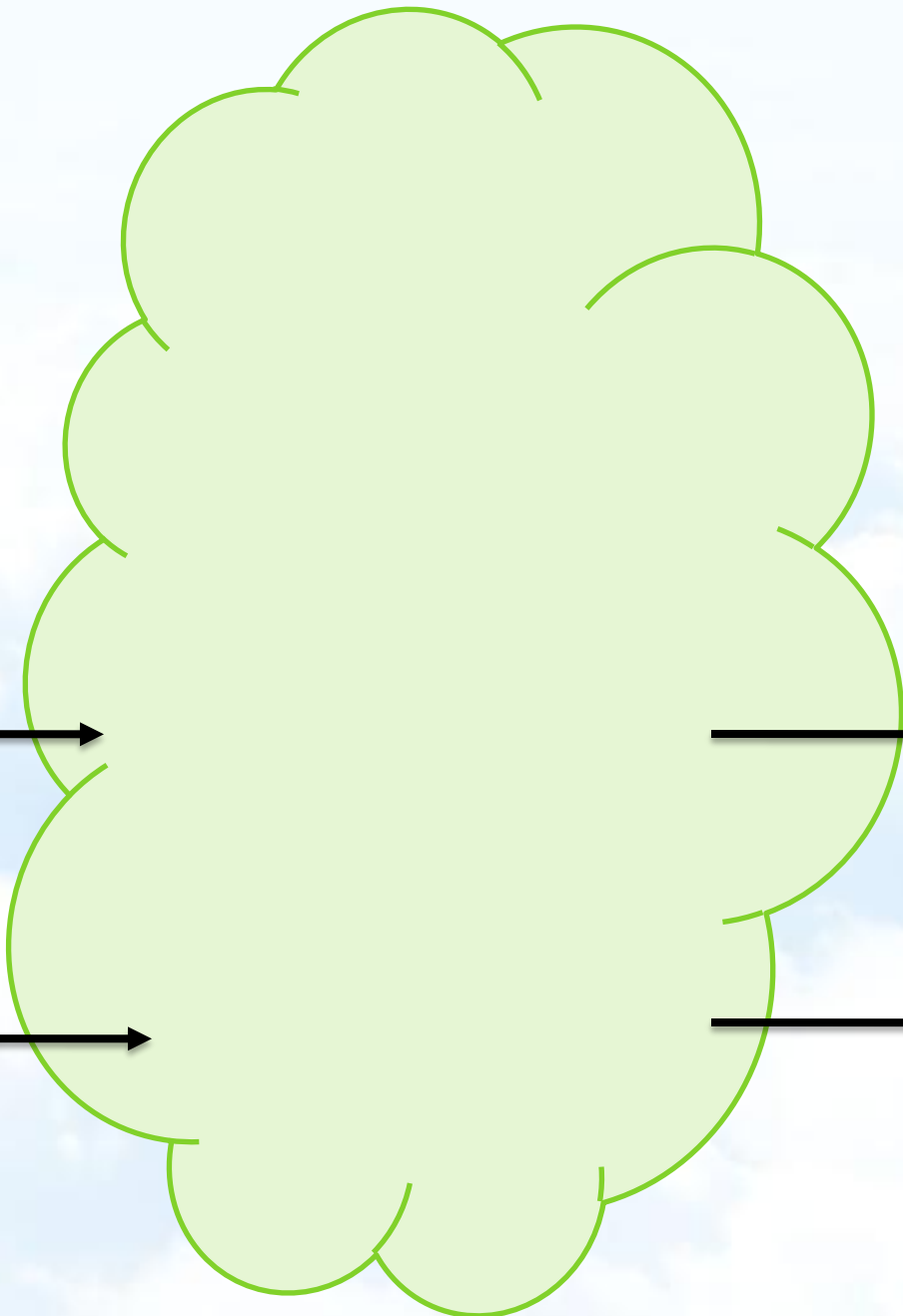
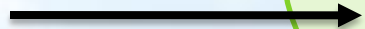
Amazon Elastic MapReduce

Jinesh Varia
Peter Sirota
Richard Cole

Start



From IDE
Command line
Web Console



End



Notify

Input Data



Get Results



Start



From IDE
Command line
Web Console



AWS EC2 Instance



Notify

End

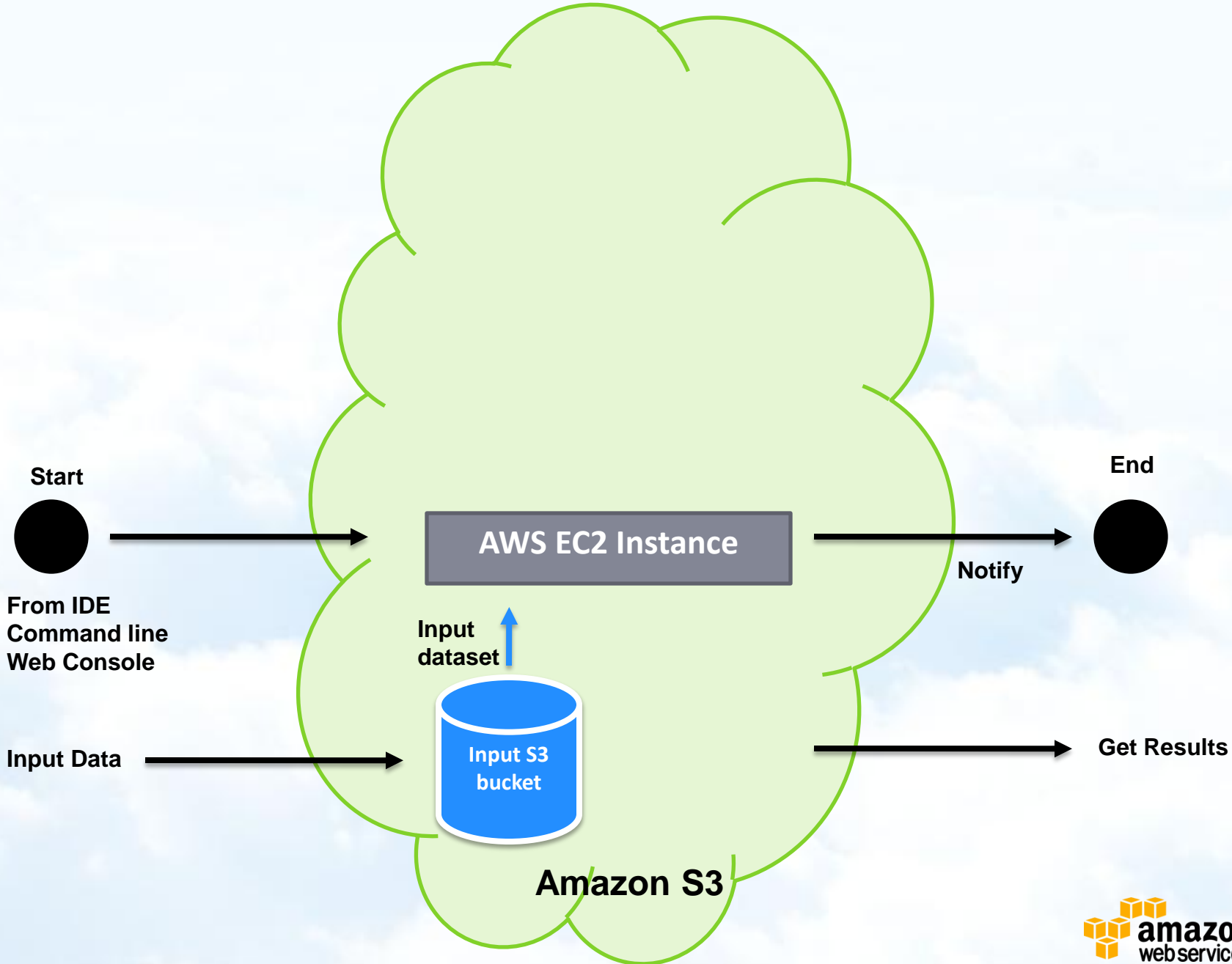


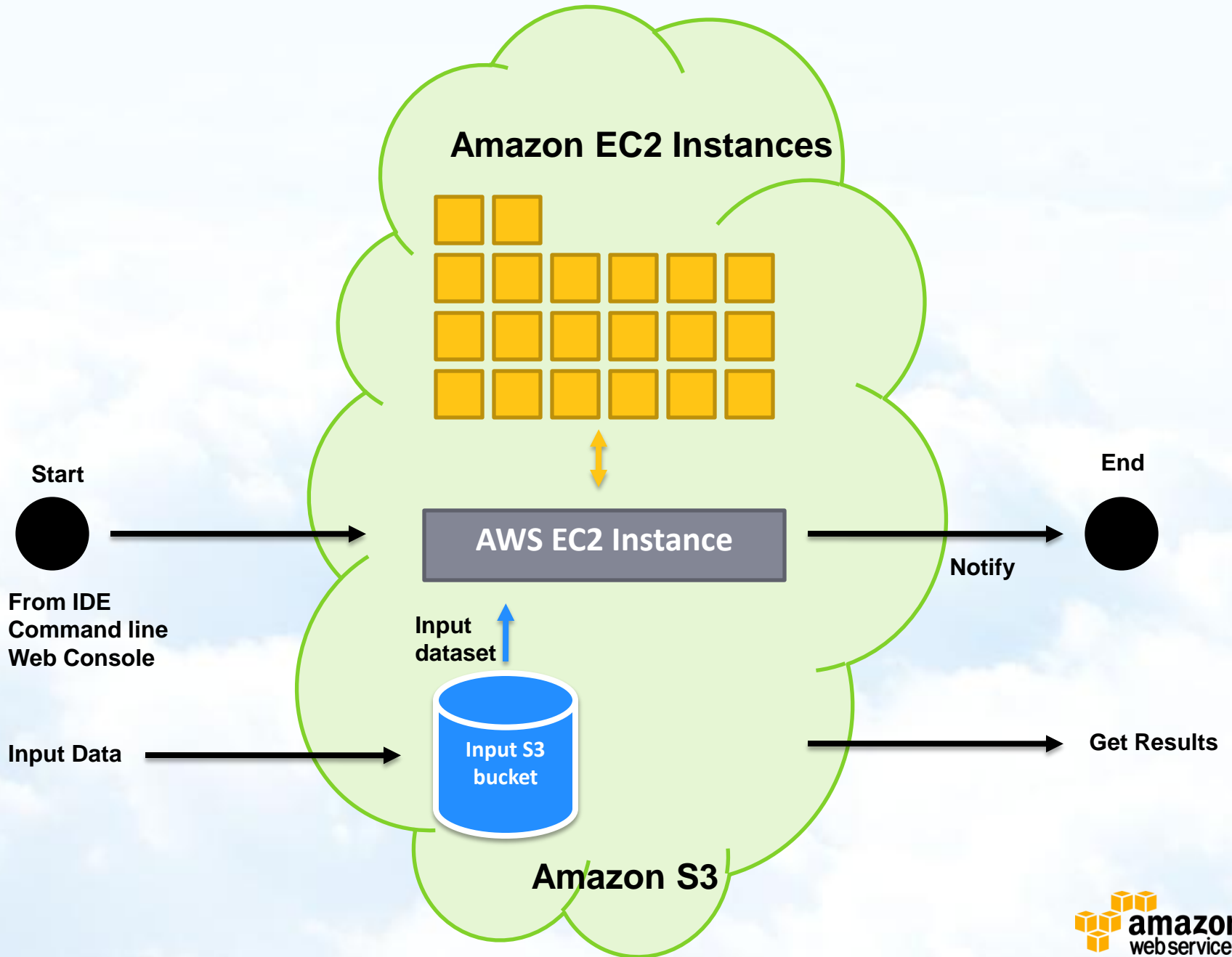
Input Data

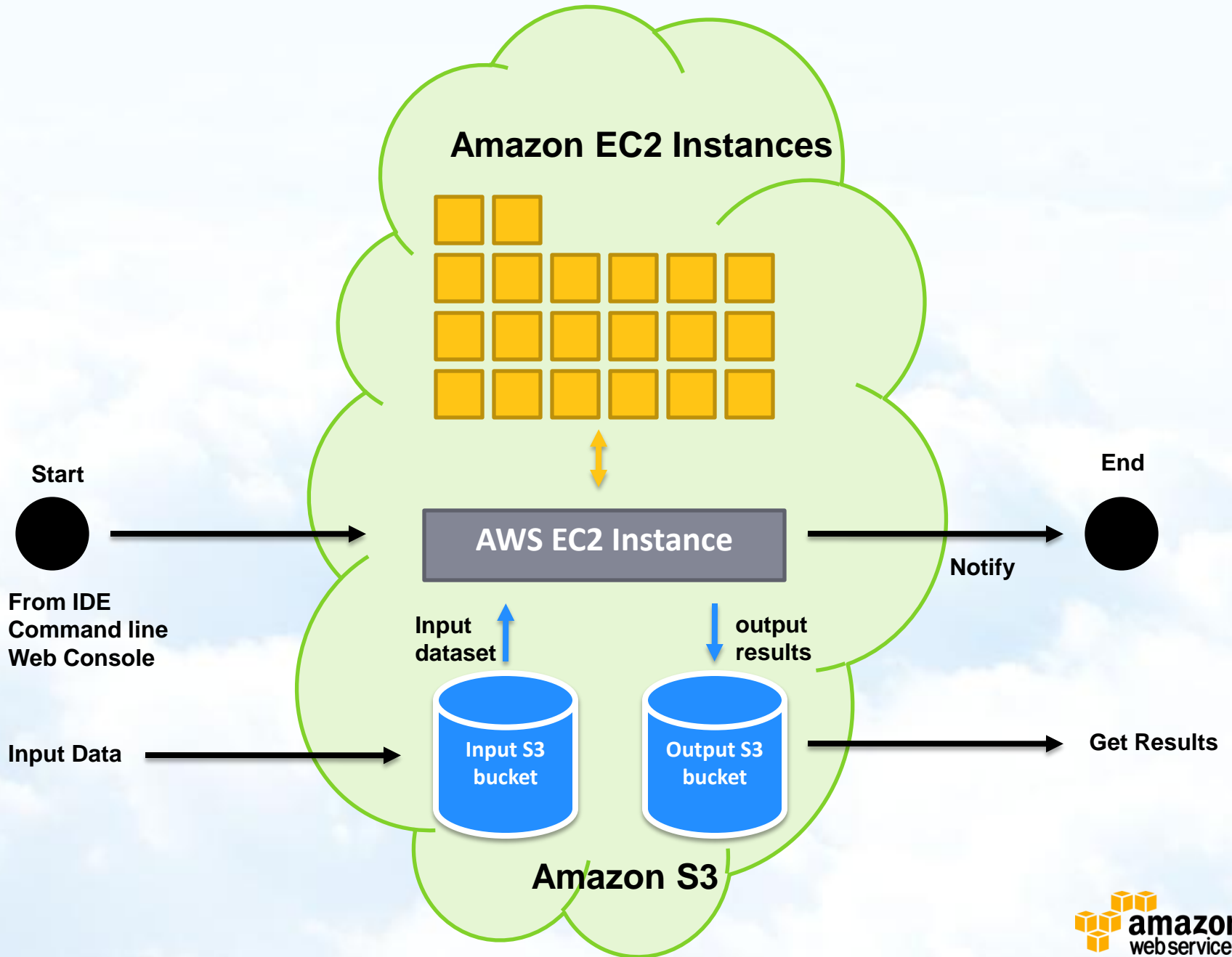


Get Results





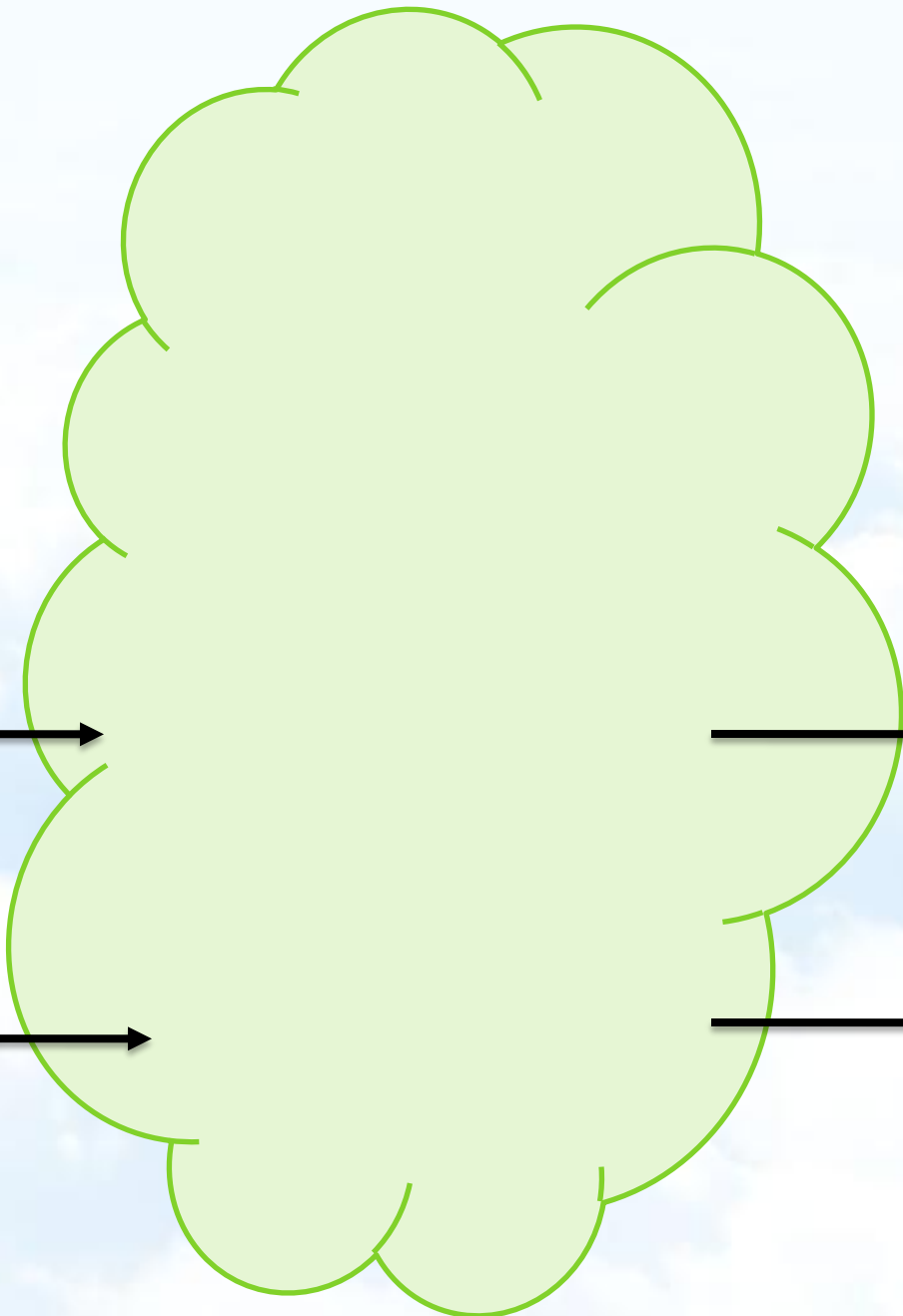




Start



**From IDE
Command line
Web Console**



End



Notify

Input Data



Get Results



Use Cases

- Bio-informatics (Genome analysis)
- Data mining (Log processing, click stream analysis, similarity algorithms, etc.)
- Financial simulation (Monte Carlo simulation)
- File processing (resize jpegs)
- Web indexing

Why Elastic MapReduce?

- Large scale data processing has a lot of MUCK and we want to remove it for our customers
 - **Hard to manage compute clusters**
 - Cluster start-up and shutdown
 - Cluster monitoring and resource management
 - Security groups management
 - **Hard to tune performance of running clusters**
 - Dozens of difficult to understand settings can affect performance 2-3x

Amazon Elastic MapReduce Benefits

Elastic

Uses as many or as few EC2 instances as needed. Spin up large or small job flows in minutes.

Easy to use

Get up and running quickly with easy-to-use web console and sample jobs. No configuration necessary.

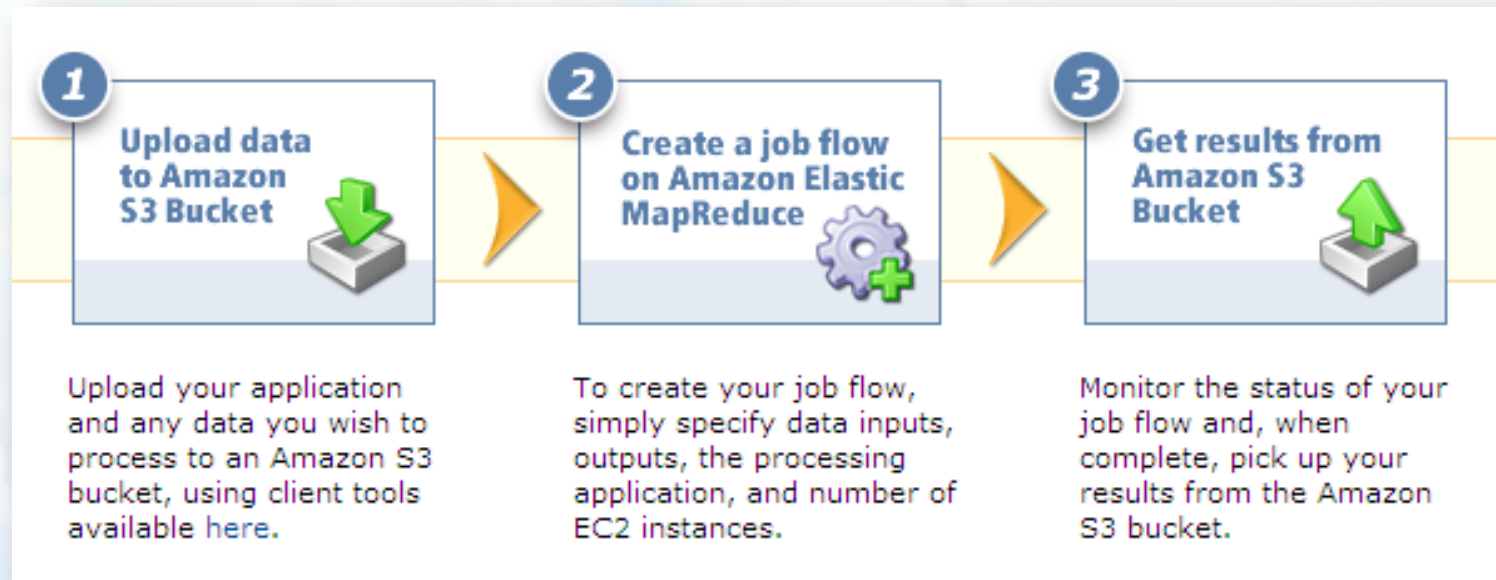
Reliable

Fault tolerant service built on top of battle-tested AWS infrastructure. Automatically retries failed tasks.

Cost Effective

We monitor progress of your jobs and turn off resources when job flow is done.

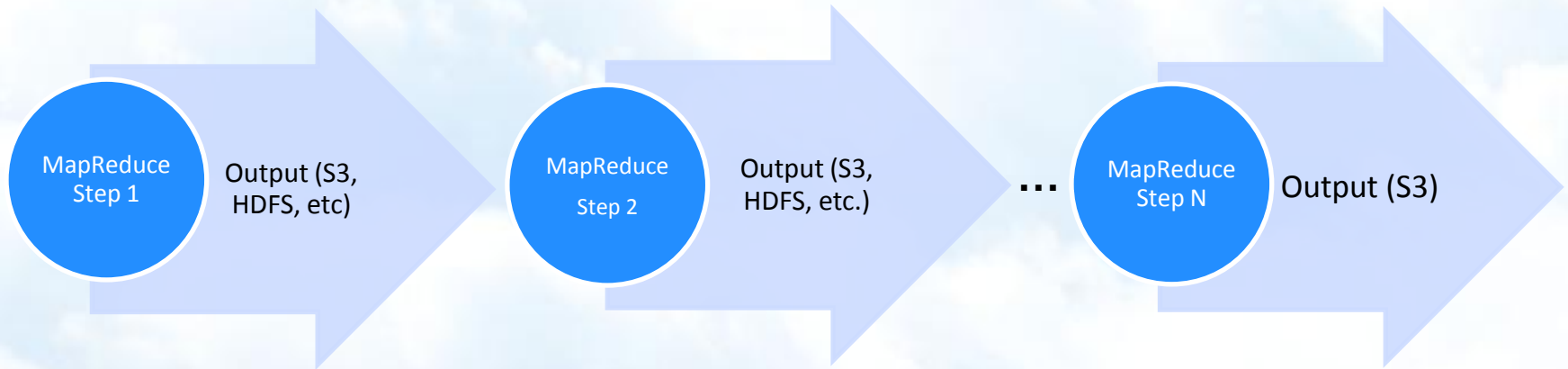
Hadoop made simple and easy



Job Flow

Job Flow inputs:

- Number and Type of EC2 instances
- Sequence of MapReduce Steps
 - Input Location
 - Output Location
 - MapReduce Algorithm Location
- Optional: log location, ssh keys



[Register](#)[Join Now! >>](#)

- Free membership
- Up to 30% cash back at over 800 stores
- Exclusive coupons and special offers

1 sign up.

IT'S FREE! Extrabux will not charge you for anything. All you need is an email address to get started. [Click here to register for free.](#)

2 shop online.

Get cash back rebates by [shopping](#) online through Extrabux - up to 30% cash back at over 800 stores. You can also use our [coupons](#) to find discounts, deals, and bargains.

3 get paid!

We will send you a check or deposit your cash back earnings directly into your [PayPal](#) account. [Learn More...](#)

Featured Stores

8%
cash back

3%
cash back

15%
cash back

3%
cash back

6%
cash back

1%
cash back

Featured Coupons

20% off any size order!*

Save 10% on any purchase, 15% on orders \$80+, and 20% on orders \$125+

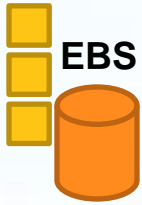
Get \$10 Off orders \$100 or more!



**Merchants
Data Feeds**



EC2



Ingestor

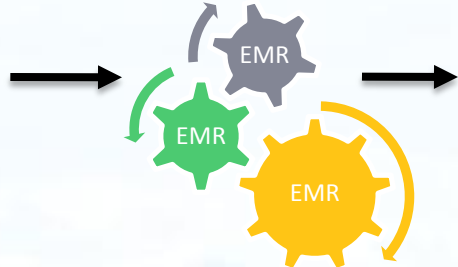


**Merchants
Data Feeds**



Ingestor

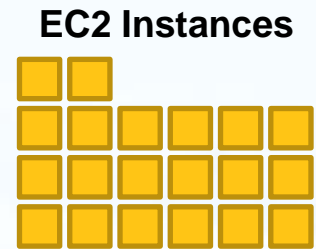
**MapReduce JobStep
(Parser)**



Converts raw product feeds
into a common format (Map-only)




HDFS

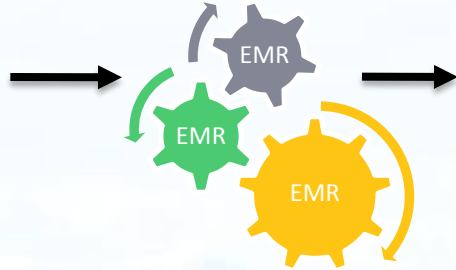


**Solr Servers
(for grouping)**


**Merchants
Data Feeds**

EC2

EBS
Ingestor

**MapReduce JobStep
(Parser)**



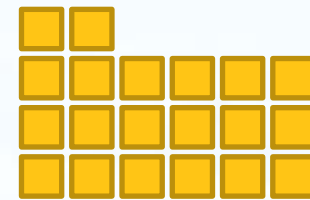
Converts raw product feeds
into a common format (Map-only)

EC2

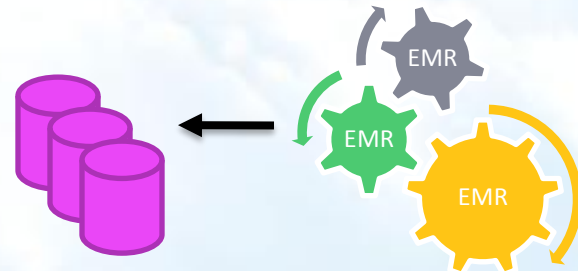


HDFS

EC2 Instances



**Solr Servers
(for grouping)**



**Same HDFS
on EC2**

**MapReduce JobStep
(Grouper)**

Uses Solr to group identical products together (Map)
and consolidates them into the correct groups (Reduce)

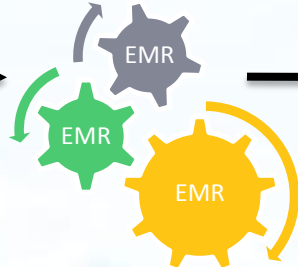


Merchants Data Feeds



Ingestor

MapReduce JobStep (Parser)



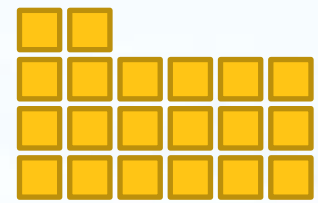
Converts raw product feeds into a common format (Map-only)

EC2



HDFS

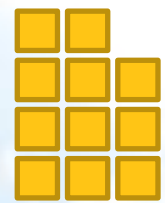
EC2 Instances



Solr Servers (for grouping)



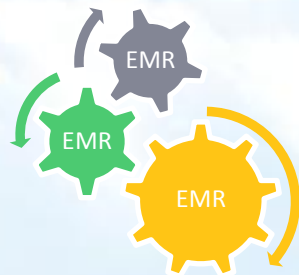
Web Servers



Solr Servers (for users)



Same HDFS on EC2



MapReduce JobStep (Grouper)

Uses Solr to group identical products together (Map) and consolidates them into the correct groups (Reduce)

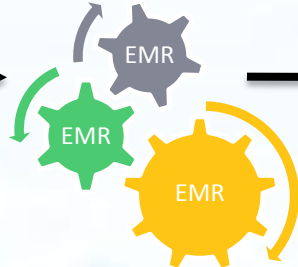



Merchants Data Feeds



Ingestor

MapReduce JobStep (Parser)



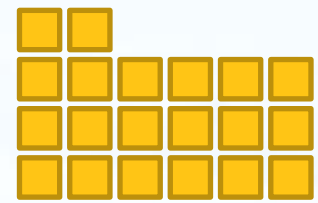
Converts raw product feeds into a common format (Map-only)

EC2



HDFS

EC2 Instances

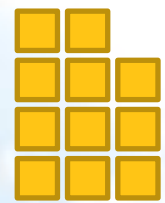


Solr Servers (for grouping)


Shopper



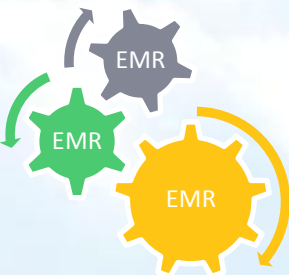
Web Servers



Solr Servers (for users)



Same HDFS on EC2



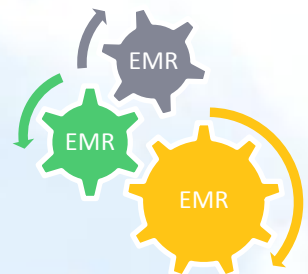
MapReduce JobStep (Grouper)

Uses Solr to group identical products together (Map) and consolidates them into the correct groups (Reduce)

S3



Product Images



MapReduce Job (Image Processor)

Ways to use Amazon EMR

- Web Management Console
 - console.aws.amazon.com
 - Point and click
- Web Services API
 - Command Line Tools by Amazon
 - scripts
 - Web Service Clients
 - Integrate from within your app

Demo

Customers Panel @ 2:00 PM

Applications Track

Expert	Company	Industry
John Barr	YieldEx	Ad inventory mgmt
Ben Hardy	eHarmony.com	MatchMaking
Paco Nathan	ShareThis	Social Web
Elias Torres	Lookery	Targeted Advertising
Ted Dunning	DeepDyve	Search

Thank You!

Jinesh Varia (jvaria@amazon.com)

Peter Sirota (sirota@amazon.com)

aws.amazon.com/elasticmapreduce

Same Pay-as-you-go pricing

Standard Amazon EC2 Instances	Amazon EC2 Price per hour (On-Demand Instances)	Amazon Elastic MapReduce Price per hour
Small (Default)	\$0.10 per hour	\$0.015 per hour
Large	\$0.40 per hour	\$0.06 per hour
Extra Large	\$0.80 per hour	\$0.12 per hour
High CPU Instances	Amazon EC2 Price per hour (On-Demand Instances)	Amazon Elastic MapReduce Price per hour
Medium	\$0.20 per hour	\$0.03 per hour
Extra Large	\$0.80 per hour	\$0.12 per hour