

# 目录

.....

.....  
.....

.....

.....

J

.....

.....  
.....

B

.....

.....  
.....

.....  
.....

.....  
.....

.....  
.....

.....

.....  
.....

.....  
.....

.....2

.....  
.....

.....

.....4

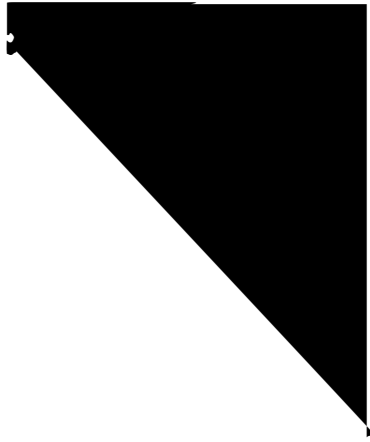
.....4

.....

.....

3143

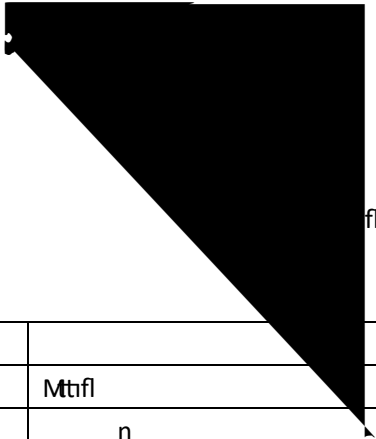
n



Mtfl

n

n



fl

n

3	Mtfl	
4	n	
4		

Mtfl

Mtfl

Python  
Python

1

2

OS

1

4

n

n

Hadoop

1 HDFS HDFS ;  
Hive  
HQL

1  
4

Python  
HTTP  
HTTP  
Cookie

urllib3 Requests HTTP  
Chrome Xpath Beautiful Soup  
JSON

1  
4

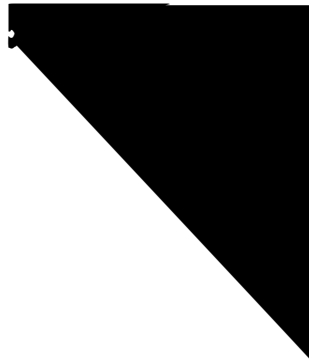
matplotlib pyplot

1

1

4

311



1

4

4

3		
4	Mtfl	Mtfl Mtfl Mtfl
4	n	n G n n n n
		4      oo      M n M fl

3

41

3 3 3

41

3	413	fl n o 413 14 4 1 fl
4	413 4 4 4	fl no 4 1 n o fl 413 134 433 3
4		fl no 4 1 n o fl 413 3443314 43
		fl no 4 1 n o fl 413 14133 4441