

# Search engine software management

by

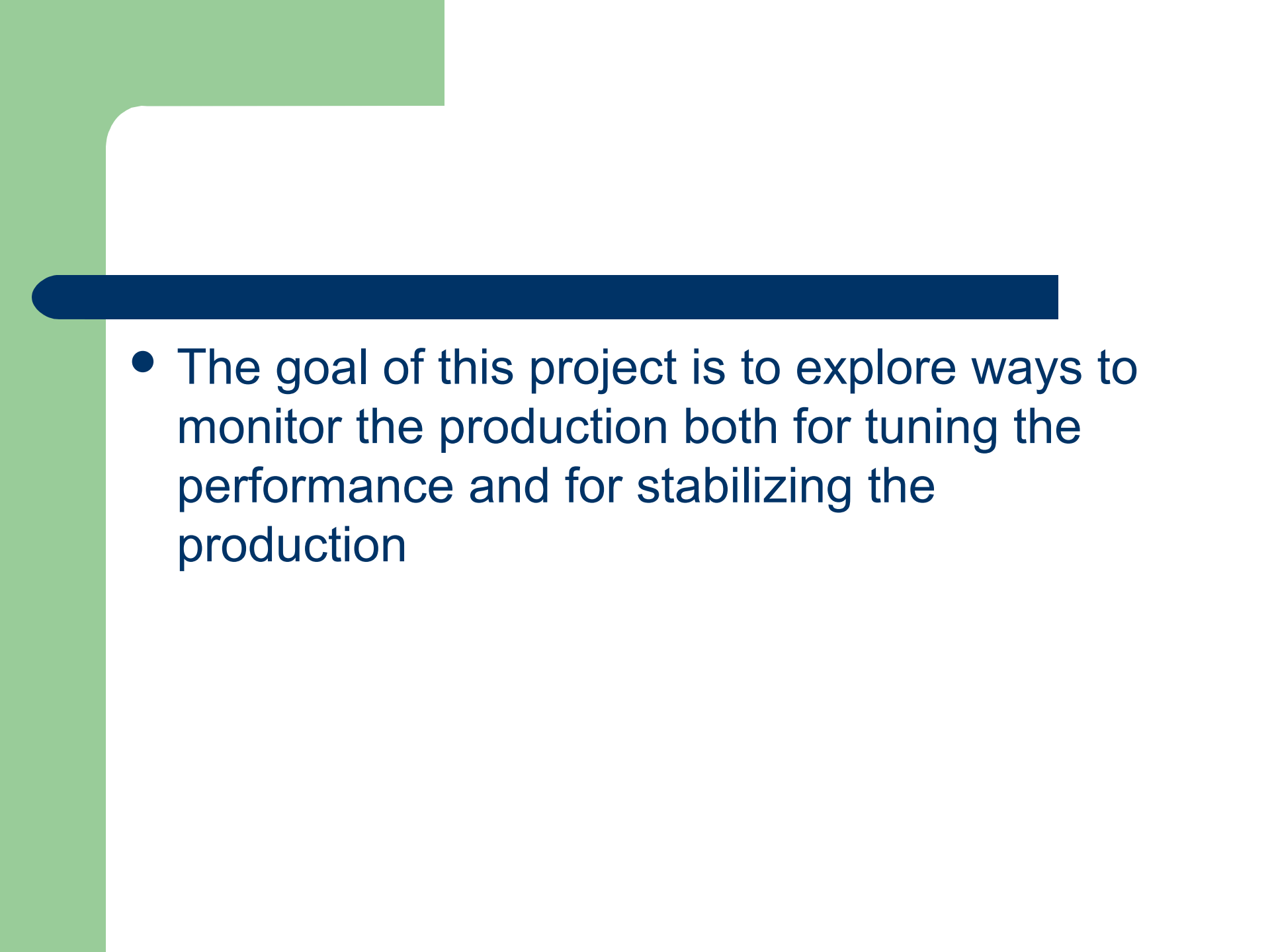
*Yang Guang, Zhu Lida*

Supervisor: Morten Goodwin Olsen



# Introduction

- In order to make sure the quality of software, we want to monitor and evaluate the security and stability of production
- Mission critical software systems need to be monitored in different ways to secure availability and stable production

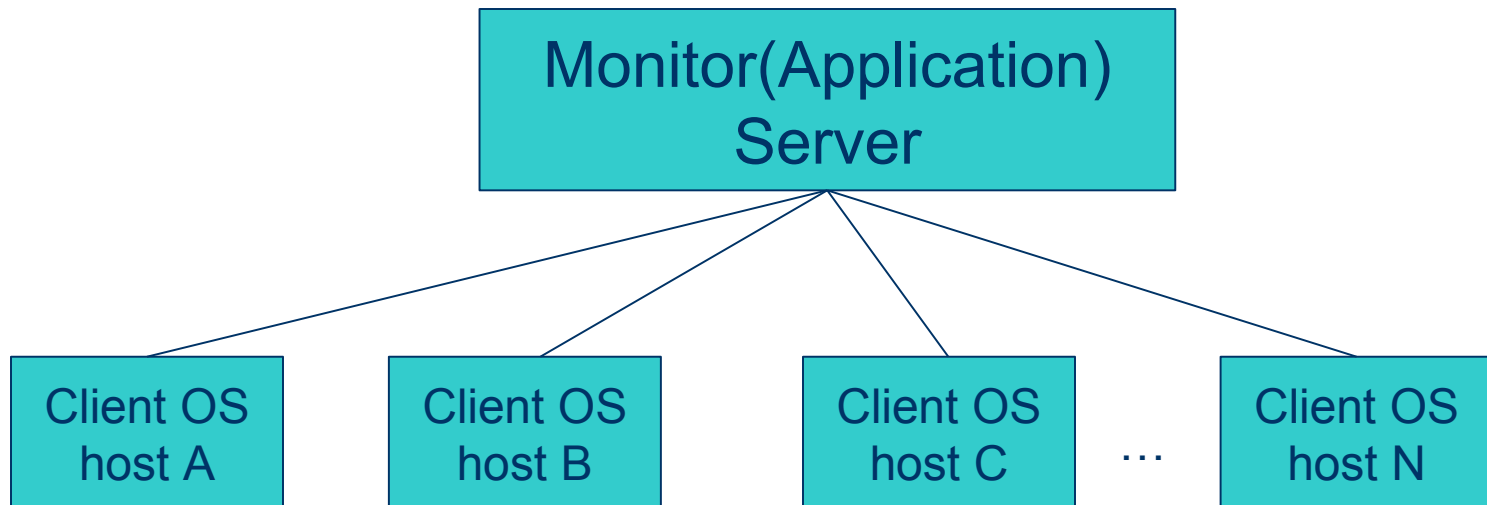
- 
- The goal of this project is to explore ways to monitor the production both for tuning the performance and for stabilizing the production

- Crawler
- ETLs (Extract, transform, and load)
- Apache server

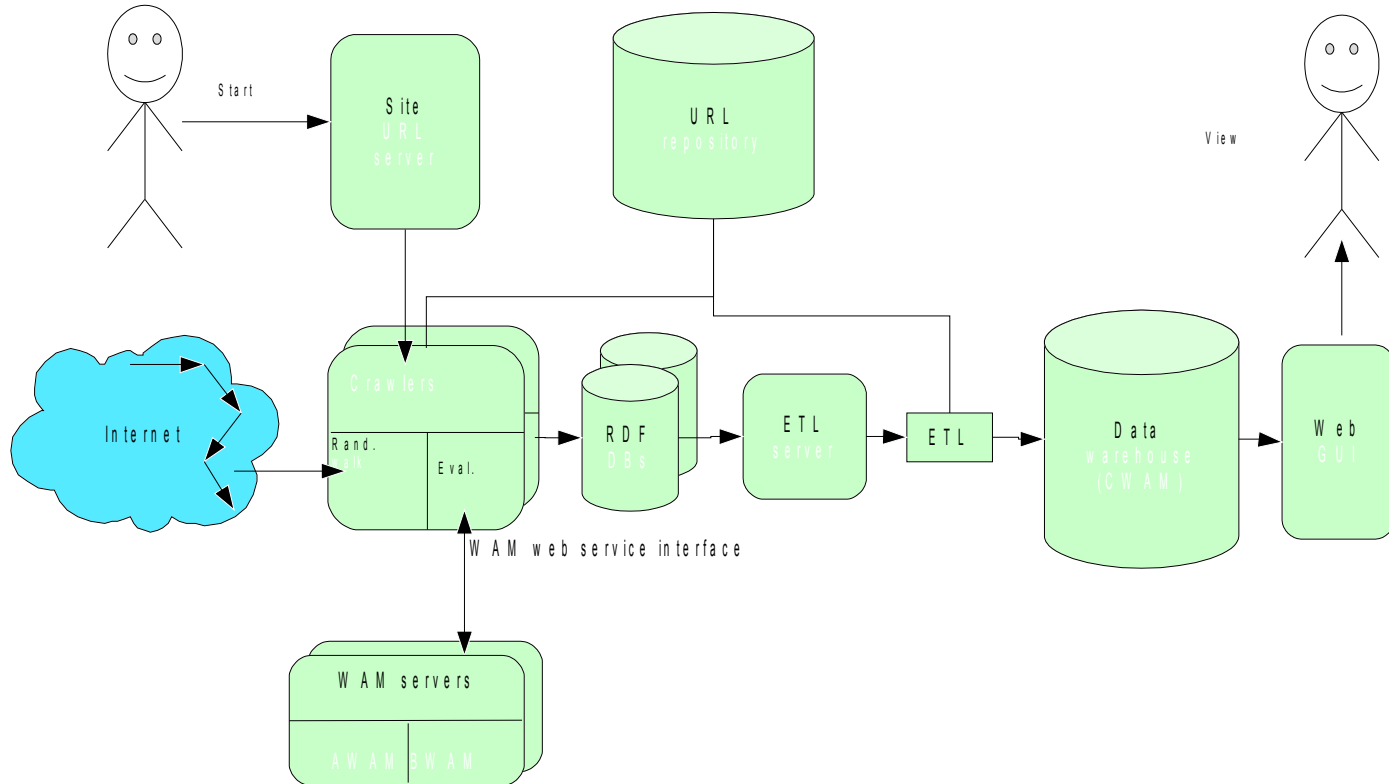
A decorative graphic on the left side of the slide, consisting of a light green vertical bar and a dark blue horizontal bar with rounded ends.

**In this project, websites monitor could be used under Linux Operation System. Considered that use Unix shell commends in Python programming is practicable and efficient.**

# Remote websites monitor



# Structure



# useful commands in Linux operator system

- Number of Crawlers running:

```
ps aux | grep -v grep | grep -v tail |grep -c crawler
```

- Get information of apache server(local):

```
ps aux | grep -v grep | grep -v tail |grep -c apache
```

- Number of ETLs;

```
ps aux | grep -v grep | grep -c dw20
```

- Number of sites finished;

```
grep -c Finished crawler.log
```

- Number of ETLs run;

# XML

```
<crawler>
```

```
<query> ps aux | grep -c crawler </query>
```

```
<text> Number of Crawler: </text>
```

```
<minimum>5</minimum>
```

```
<maximum>20</maximum>
```

```
</crawler>
```

# This is XML parser:

```
def handle(self,node):
for child in node.childNodes: #search every nodes
if child.childNodes:
for c in child.childNodes:
if str(c.nodeName)=='query':
queryres = self.query(c)
if str(c.nodeName)=='text':
self.text(c)
if str(c.nodeName)=='minimum':
mini = int(self.mini(c))
if str(c.nodeName)=='maxmum':
maxi = int(self.maxi(c))
```

```
self.test(child,maxi,mini,queryres)
def query(self,node):          #this method is used to parse 'query'
commend=node.childNodes[0].nodeValue
result=os.system(commend) #execute command which is gotten from XML
'query'
return result
def text(self,node):          #this method is used to parse 'text'
print node.childNodes[0].nodeValue #print out text in 'text'
def maxi(self,node):         #this method is used to parse 'max'
maxi=node.childNodes[0].nodeValue
return maxi                  #get value of max for further compare
def mini(self,node):         #this method is used to parse 'mini'
mini=node.childNodes[0].nodeValue #get value of mini for further compare
return mini
```

```
def test(self,node,maxi,mini,queryres):           #test part
if mini<queryres<queryres:
print "OK"
else: print "NO"
etls=os.listdir('.')                             #print out all ETLs
doc=minidom.parse('input1.xml')
monitor(doc)
```

# Testing

result show out as follows:

Content-type: text/html

2

Number of Crawlers:

0

OK

```
root 4587 0.0 0.4 4644 2112 ? S 11:18 0:00 /usr/sbin/apache
```

```
www-data 4592 0.0 0.1 4644 948 ? S 11:18 0:00 /usr/sbin/apache
```

```
www-data 4593 0.0 0.1 4644 948 ? S 11:18 0:00 /usr/sbin/apache
```

```
www-data 4594 0.0 0.1 4644 948 ? S 11:18 0:00 /usr/sbin/apache
```

```
www-data 4595 0.0 0.1 4644 948 ? S 11:18 0:00 /usr/sbin/apache
```

```
www-data 4596 0.0 0.1 4644 948 ? S 11:18 0:00 /usr/sbin/apache
```

```
root 4610 0.0 0.8 10080 4216 ? Ss 11:18 0:00 /usr/sbin/apache2 -k start -DSSL
```

```
www-data 4617 0.0 0.3 9064 1880 ? S 11:18 0:00 /usr/sbin/apache2 -k start -DSSL
```

```
www-data 4685 0.0 0.6 231416 3208 ? SI 11:18 0:00 /usr/sbin/apache2 -k start -DSSL
www-data 4687 0.0 0.6 231416 3212 ? SI 11:18 0:00 /usr/sbin/apache2 -k start -DSSL
shane 8978 0.0 0.0 1656 464 pts/0 S+ 13:36 0:00 sh -c ps aux | grep apache
shane 8980 0.0 0.1 2800 768 pts/0 S+ 13:36 0:00 grep apache
```

web server:

256

NO

Number of ETLs:

1

OK

last time wrote to the crawler log:

2007-05-06 02:06:10.000000000 +0200

# Evaluation

- **Advantage**

Flexible: one of most important advantage in this programming is its expandable

- **Disadvantage**

Application runs without graphics windows

**Thank you**

