

## GSOC 2015 R STATISTICAL PROJECT NATURAL LANGUAGE PROCESSING PROPOSAL.

HÉCTOR APOLO ROSALES PULIDO

### 1. INTRODUCTION

The following is a proposal for the Natural Language processing project, this proposal is made based on the information based on the site: <https://github.com/rstats-gsoc/gsoc2015/wiki/Natural-language-processing>

### 2. BACKGROUND

I am originally Guadalajara, México where I completed a bachelor degree in Computer Science at the Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) campus Guadalajara.

I'm now completing a Master's degree in Business Intelligence and Data Mining at the UPC(Barcelonatech), and I'm looking forward to work with machine learning and Natural language processing.

I worked for Continental corporation, in the **Core Development Software(CDS)**, developing an internal Tracing tool that was embedded in several automotive systems for testing features of other developers. We had teammates in Germany with whom had meetings twice a week to discuss and develop the tool. We also gave support to other business units in other countries, notably Singapore and Romania, whenever there was a problem with our tool, or to present our tool to them. It was developed using Python, and Qt in the PC side, and C/C++ in the embedded side. The project was managed using GIT, and Atlassian's Jira, to manage tasks.

### 3. PERSONAL MOTIVATIONS

**Natural language processing (NLP)** is becoming more relevant every day, the computational capabilities available to store and process enormous quantities of data is becoming cheaper everyday, so applications that use NLP such as measuring the mood in Twitter, or Facebook, a brand's worth etc. are becoming more popular, along with those that translate text from one language to the next.

I would like to be part of this project, because I would like to get experience with NLP and machine learning, and even though I'm still a student I believe that I can make a significant contribution to the development of the project. At the UPC(Universitat Politècnica de Catalunya) I've been studying machine learning techniques and I've been using R to do my analysis, I've had a small introduction to NLP, and by the end of the semester I'll be finishing several projects using these tools.

---

*Date:* March 2015.

I would also like to be part of the development of the R statistical tool, I think is a great tool, working in its development would help me get a better understanding of it and it would be a valuable thing to potential employers in the future.

#### 4. PARTICIPANTS

The google summer of code provides with two participants that are involved in the project

- **A mentor:** *A person or group of people responsible for monitoring the progress of each accepted student and to mentor her/him as the project progresses.*
- **Student(s):** Who will develop the approved proposal.

#### 5. PROPOSAL

The idea would be to evaluate a user query regarding a product quality or features, could be related to price, warranty, or other things. We can use a dataset such as the one found here: Stanford's snap

#### 6. IMPORTANT DATES

First we have to be aware of the important dates in which we have to deliver documents to google, and the time we might have to complete any previous training if selected to work on the project.

- **April 27th 2015:** Interim period starts
- **May 25th 2015:** Coding starts
- **June 26th 2015:** Mid term evaluation by our mentor.
- **August 21st 2015:** Final delivery of the project.

#### 7. MOTIVATION

When buying things online, it's not easy to find references of products that are not phisically within our reach, and given the growth that online shopping is having, it's important to have some way to make more informed decisions, looking at comments may be a standard practice in some sites, but there are sites in which there are no information available about either the manufacturer or the product itself.

#### 8. ITERIM PERIOD

If selected, I hope I can have a meeting with the participants of the project, the mentor(s) and if there are other students with whom I will work, we should be able to have a virtual meeting first of all to get to know team members to present ourselves and to establish the working dynamic, which software are we going to use, and if there are any external players we should know about.

After that we should be able to define the scope of the project, what things are feasible or not, also the technologies that we can use for this project, and other things we should take into account.

During this period I will take steps to kickstart my developing environment, and also work on improving my abilities needed to efficiently work in the project once it starts, if i need to learn how the R development environment works for example.

Testing should also be discussed, having every part of code unit-tested before pushing it, and documenting the code that is pushed would be better at this stage.

## 9. DEVELOPMENT METHODOLOGY

I would like to work in a scrum agile methodology, having meetings as often as possible with the team to discuss the things that we had done, and if there is anything to modify in our weekly goals. Document anything that is released before pushing it, and if possible do unit test at least in the more crucial parts of the code. Having a fluent communication through email would be a special point to take care, we could also use the emails to document the project as discussions for future reference.

## 10. START OF THE PROJECT














By now the scope of the project and a list of main task to develop the project should be specified, we should also prioritize the tasks, so that we finish what's more important first, and then add the features that would be *nice to have*. A goal to aim to for the mid term evaluations should be defined while taking into account that I will during my last weeks of class, finishing school projects and exams during this period.

## 11. MID TERMS

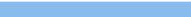














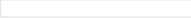



Here we would be finishing all the main tasks, and start completing the "nice to have" features, and the documentation of the code should be condensed in a vignette for future use.

## 12. FINAL DELIVERY.

It would be prudent to plan the project as if it would finish August 17th, to use the remaining days fine tuning the evaluation, and deployment of the product.

ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors	
0		<b>Software Development Plan</b>	<b>92 days?</b>	<b>Mon 4/27/15</b>	<b>Tue 9/1/15</b>		
1		<b>Important Dates</b>	<b>89 days</b>	<b>Mon 4/27/15</b>	<b>Fri 8/28/15</b>		
2		Start Interim Period	0 days	Mon 4/27/15	Mon 4/27/15		
3		Start Coding	0 days	Mon 5/25/15	Mon 5/25/15	2	
4		Mid Term Evaluations Period	6 days	Fri 6/26/15	Fri 7/3/15	3	
5		Pencils Down date	0 days	Fri 8/21/15	Fri 8/21/15	4	
6		Final Evaluation Date	0 days	Fri 8/28/15	Fri 8/28/15	5	
7		<b>HR Communications</b>	<b>5 days?</b>	<b>Tue 4/28/15</b>	<b>Tue 5/5/15</b>		
8		Presentation to the community	1 day?	Tue 4/28/15	Wed 4/29/15	14	
9		Get contact information	1 day?	Wed 4/29/15	Thu 4/30/15	8	
10		Set communication channels	1 day?	Thu 4/30/15	Fri 5/1/15	9	
11		Define development process	1 day?	Mon 5/4/15	Tue 5/5/15	12,10	
12		Define office hours	1 day?	Fri 5/1/15	Mon 5/4/15	10	

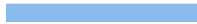


















  














Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			




















  



Page 1




















ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors
13		<b>Scope</b>	<b>6.5 days</b>	<b>Tue 4/28/15</b>	<b>Wed 5/6/15</b>	
14		Determine project scope	4 hrs	Tue 4/28/15	Tue 4/28/15	
15		Define preliminary resources	1 day	Tue 5/5/15	Wed 5/6/15	11,12,14
16		Scope complete	0 days	Wed 5/6/15	Wed 5/6/15	15
17		<b>Interim Period</b>	<b>20 days</b>	<b>Mon 4/27/15</b>	<b>Fri 5/22/15</b>	
18		Prepare development environment	1 day?	Mon 4/27/15	Mon 4/27/15	
19		Obtain access to relevant tools in the organization	1 day?	Mon 4/27/15	Mon 4/27/15	
20		Prepare the technologies in which I'm not very strong.	1 day?	Mon 4/27/15	Mon 4/27/15	
21		<b>Analysis/Software Requirements</b>	<b>52.63 days</b>	<b>Mon 6/1/15</b>	<b>Wed 8/12/15</b>	
22		List tasks to be completed	5 hrs	Tue 8/11/15	Tue 8/11/15	24,17
23		Priorityse task delivery	1 day	Tue 8/11/15	Wed 8/12/15	22
24		<b>Update tak list</b>	<b>51 days</b>	<b>Mon 6/1/15</b>	<b>Mon 8/10/15</b>	<b>16</b>
36		Analysis complete	0 days	Wed 8/12/15	Wed 8/12/15	23,24



Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			




















ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors
37		<b>Design</b>	<b>1.5 days</b>	<b>Wed 8/12/15</b>	<b>Fri 8/14/15</b>	
38		Design a web prototype	1 day	Wed 8/12/15	Thu 8/13/15	36
39		Obtain approval to proceed	4 hrs	Thu 8/13/15	Fri 8/14/15	38
40		Design complete	0 days	Fri 8/14/15	Fri 8/14/15	39
41		<b>Development</b>	<b>61.13 days</b>	<b>Mon 5/25/15</b>	<b>Tue 8/18/15</b>	
42		Review functional specifications	1 day	Fri 8/14/15	Mon 8/17/15	40
43		Identify modular/tiered design parameters	1 day	Mon 8/17/15	Tue 8/18/15	42
44		Develop code (also document code)	61 days	Mon 5/25/15	Mon 8/17/15	
45		Developer testing (primary debugging)	15 days	Tue 7/28/15	Mon 8/17/15	44FS-75%
46		Development complete	0 days	Mon 8/17/15	Mon 8/17/15	45
47		<b>Testing</b>	<b>55 days</b>	<b>Fri 6/5/15</b>	<b>Thu 8/20/15</b>	
48		<b>Develop unit test using product specifications.</b>	<b>51 days</b>	<b>Fri 6/5/15</b>	<b>Fri 8/14/15</b>	
60		<b>Integration Testing mid terms</b>	<b>4 days</b>	<b>Tue 6/30/15</b>	<b>Fri 7/3/15</b>	

Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			

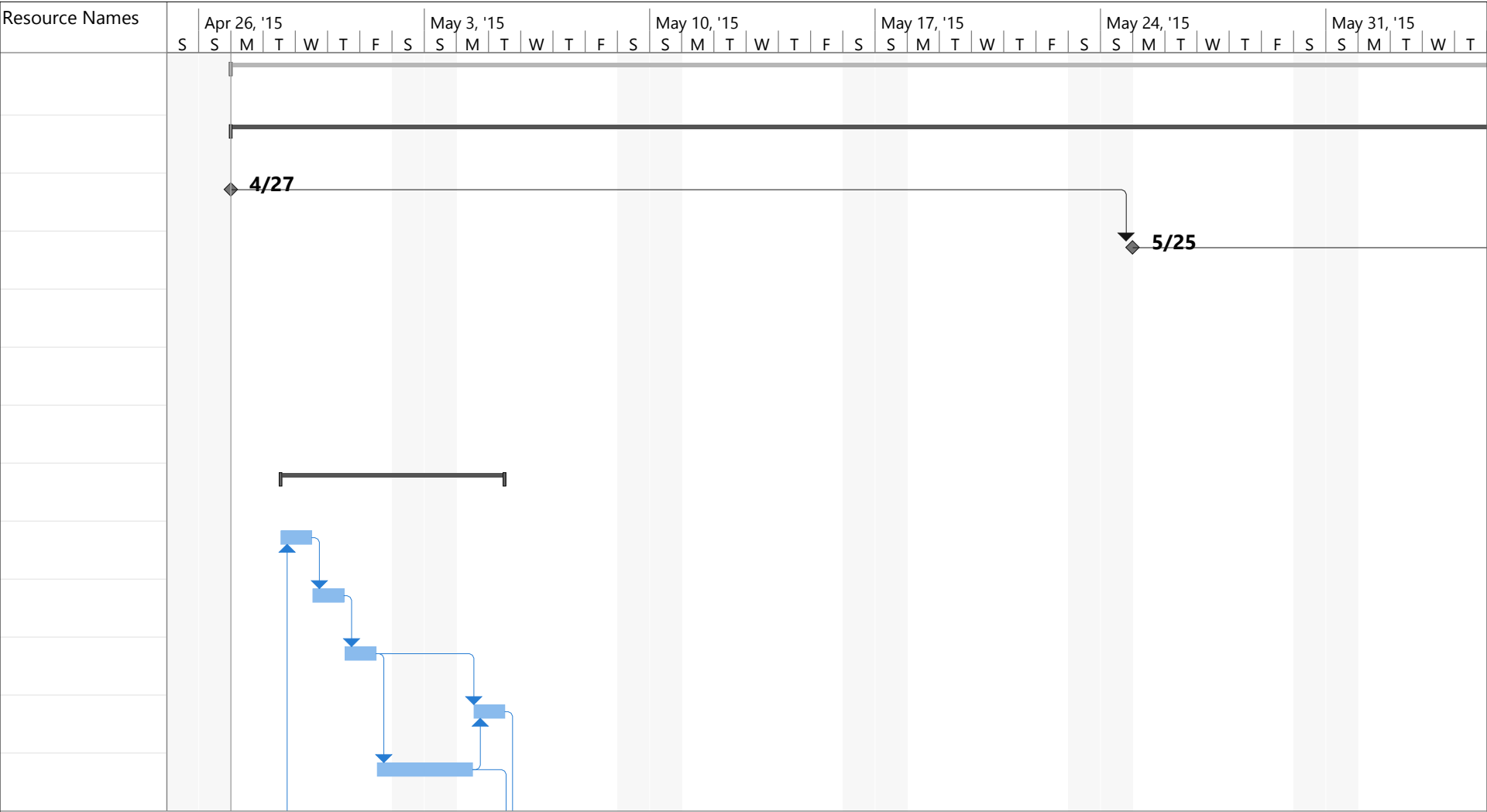
ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors
66		<b>Integration Testing final</b>	<b>4 days</b>	<b>Mon 8/17/15</b>	<b>Thu 8/20/15</b>	
72		<b>Documentation</b>	<b>55 days</b>	<b>Mon 6/8/15</b>	<b>Fri 8/21/15</b>	
73		Develop user manuals specifications	2 days	Mon 6/8/15	Tue 6/9/15	
74		First user manual delivery	0 wks	Fri 7/3/15	Fri 7/3/15	
75		Final Manual delivery	0 days	Fri 8/21/15	Fri 8/21/15	
76		Review all user documentation	2 days	Thu 8/20/15	Fri 8/21/15	
77		Documentation complete	0 days	Fri 8/21/15	Fri 8/21/15	
78		<b>Deployment</b>	<b>5 days</b>	<b>Mon 8/24/15</b>	<b>Fri 8/28/15</b>	
79		Mid term deployment	1 day	Fri 7/3/15	Fri 7/3/15	
80		Deploy software	1 day	Fri 8/21/15	Fri 8/21/15	79
81		Deployment complete	0 days	Mon 8/24/15	Mon 8/24/15	80
82		<b>Post Implementation Review</b>	<b>5 days</b>	<b>Wed 8/26/15</b>	<b>Tue 9/1/15</b>	
83		Document lessons learned	1 day	Mon 8/24/15	Mon 8/24/15	

Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			

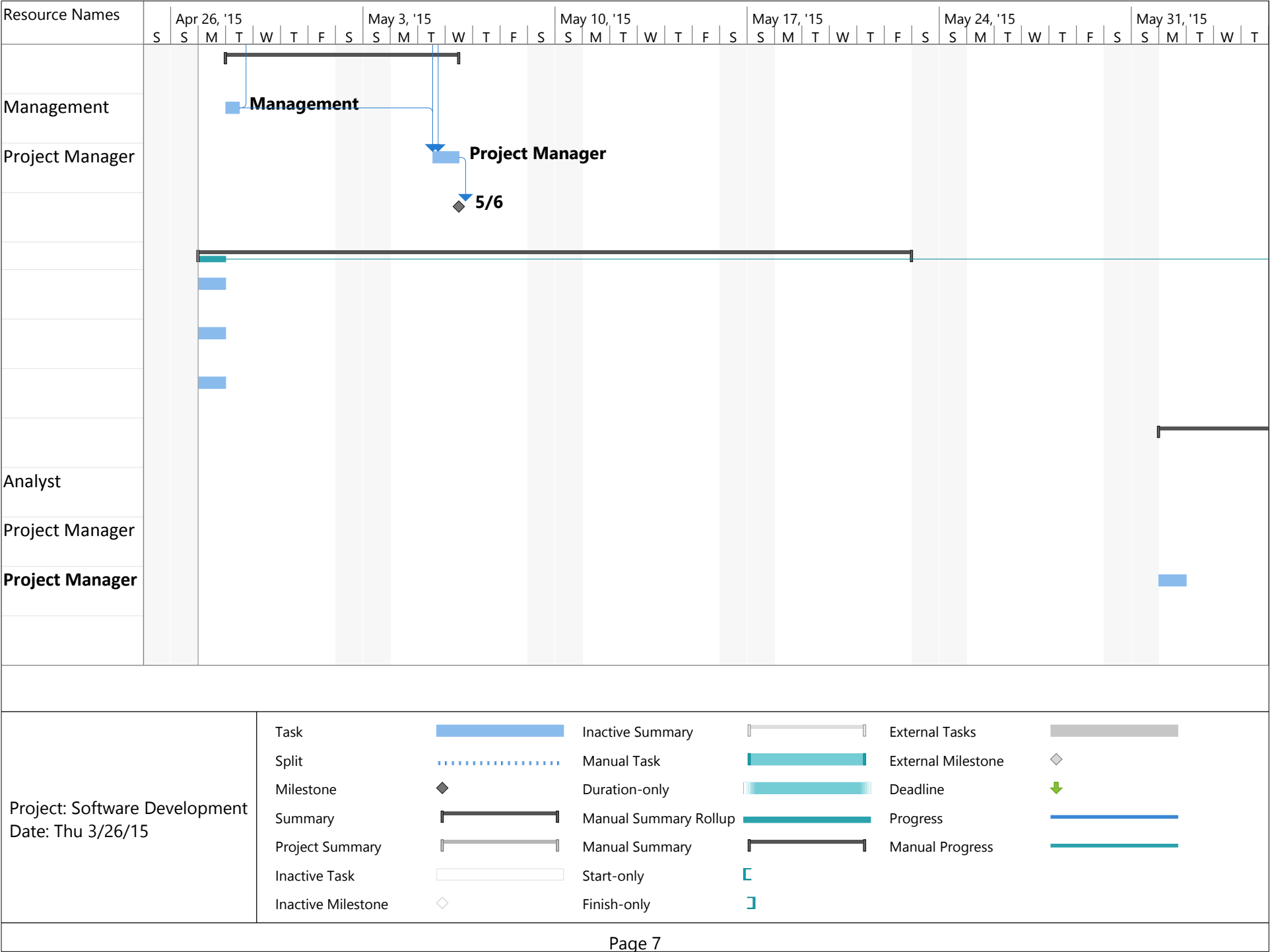
ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors
84		Distribute to team members	1 day	Fri 8/21/15	Fri 8/21/15	
85		Post implementation review complete	0 days	Fri 8/21/15	Fri 8/21/15	
86		Software development template complete	0 days	Fri 8/21/15	Fri 8/21/15	85

Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			





Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			



[illegible]

Resource Names		Apr 26, '15							May 3, '15							May 10, '15							May 17, '15							May 24, '15							May 31, '15							
		S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T		
Technical Communicators																																												
Technical Communicators																																												
Technical Commu																																												
Technical Communicators																																												
Deployment Team																																												
Deployment Team																																												
Project Manager																																												

Project: Software Development

Date: Thu 3/26/15

Task

Split

Milestone

Summary

Project Summary

Inactive Task

Inactive Milestone

Inactive Summary

Manual Task

Duration-only

Manual Summary Rollup

Manual Summary

Start-only

Finish-only

External Tasks

External Milestone

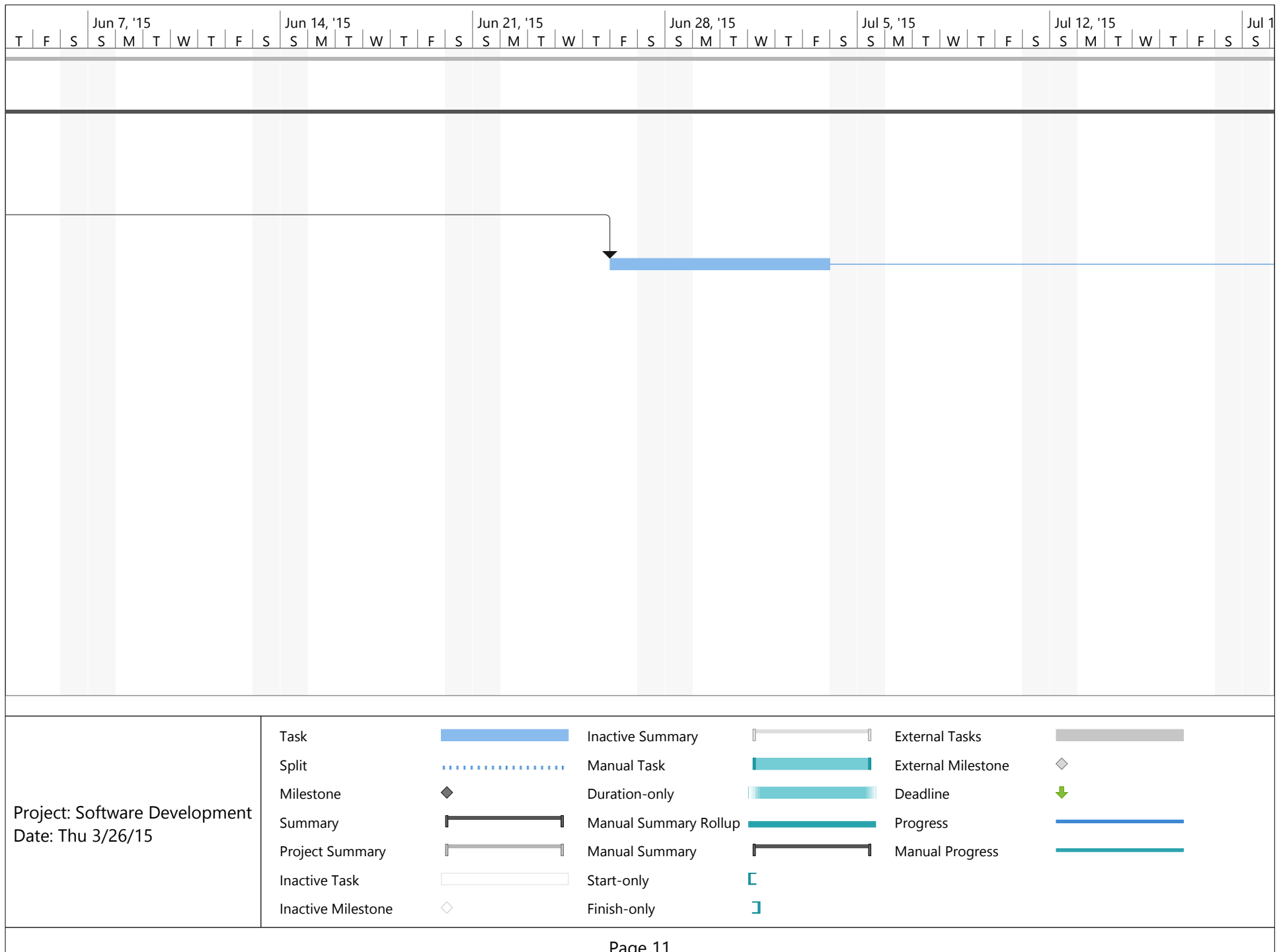
Deadline

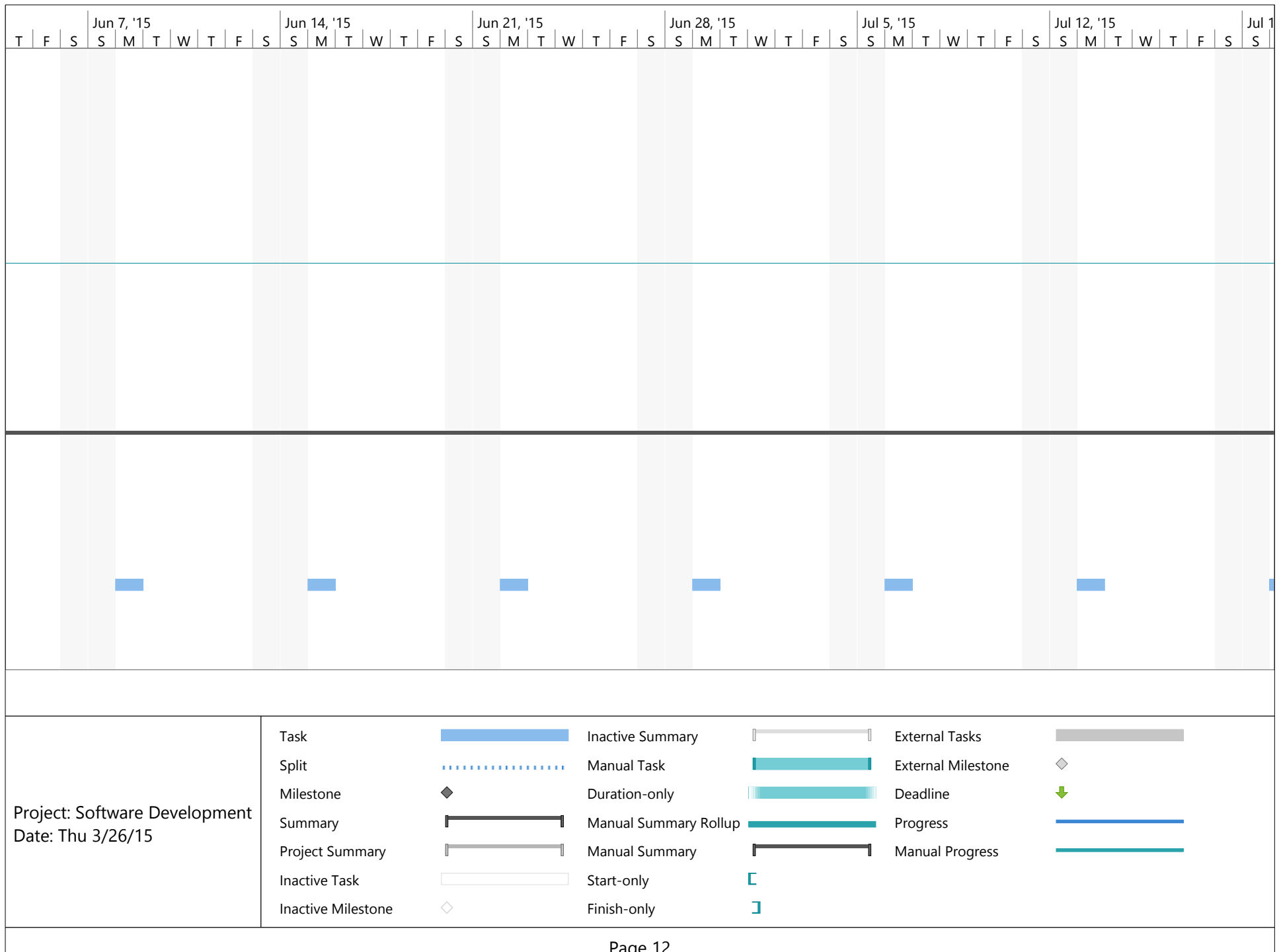
Progress

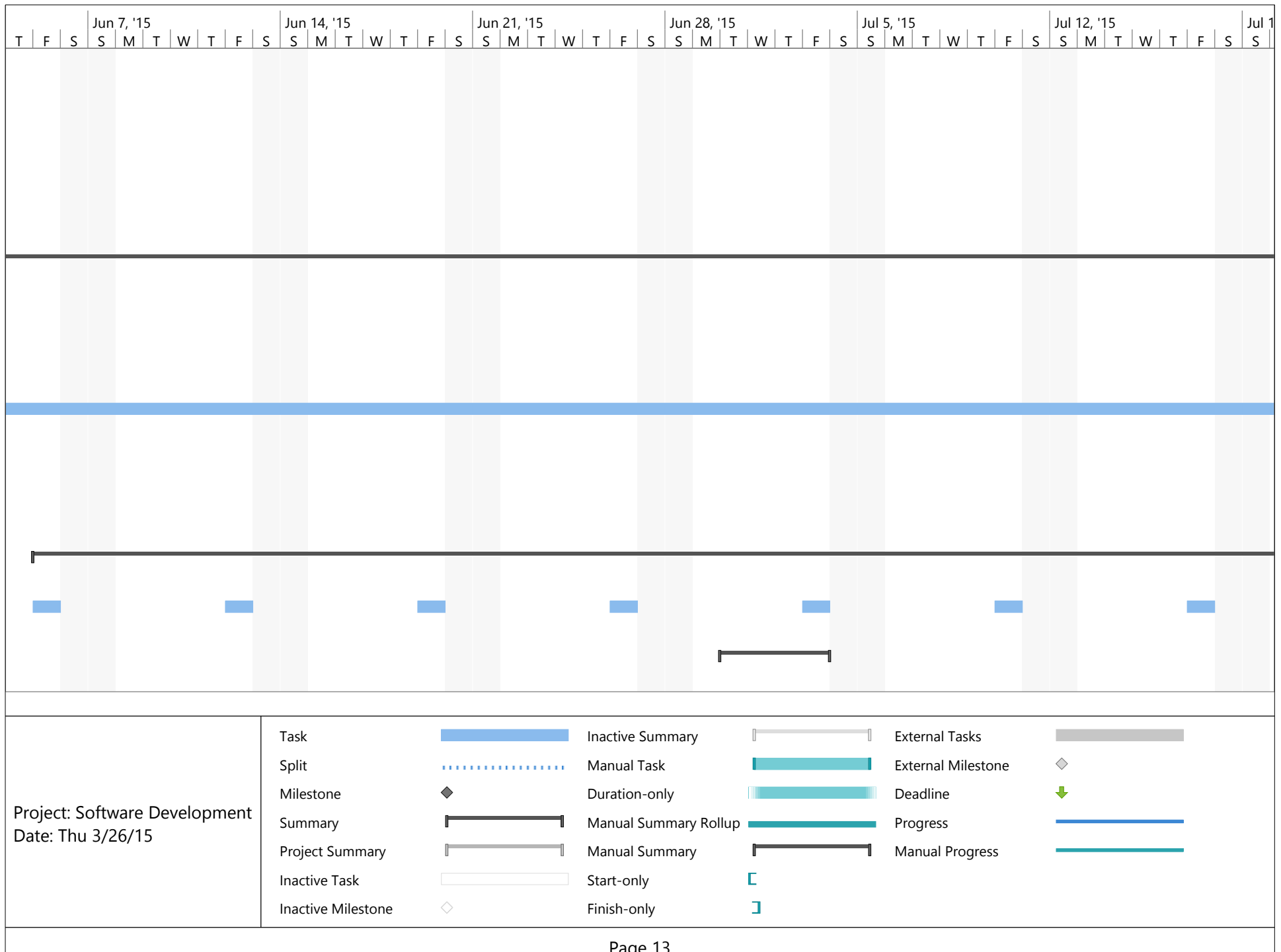
Manual Progress

Page 9

[illegible]



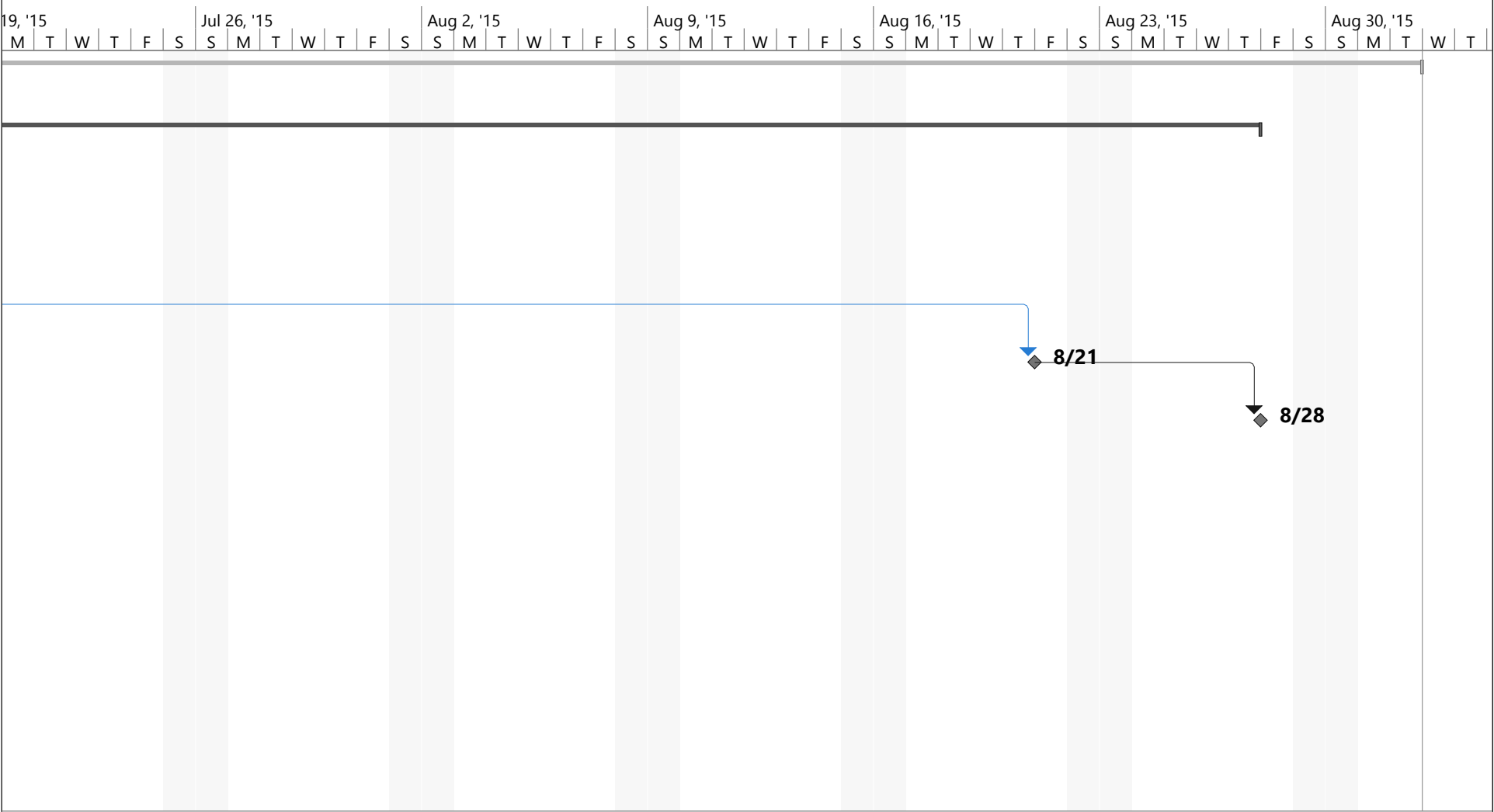






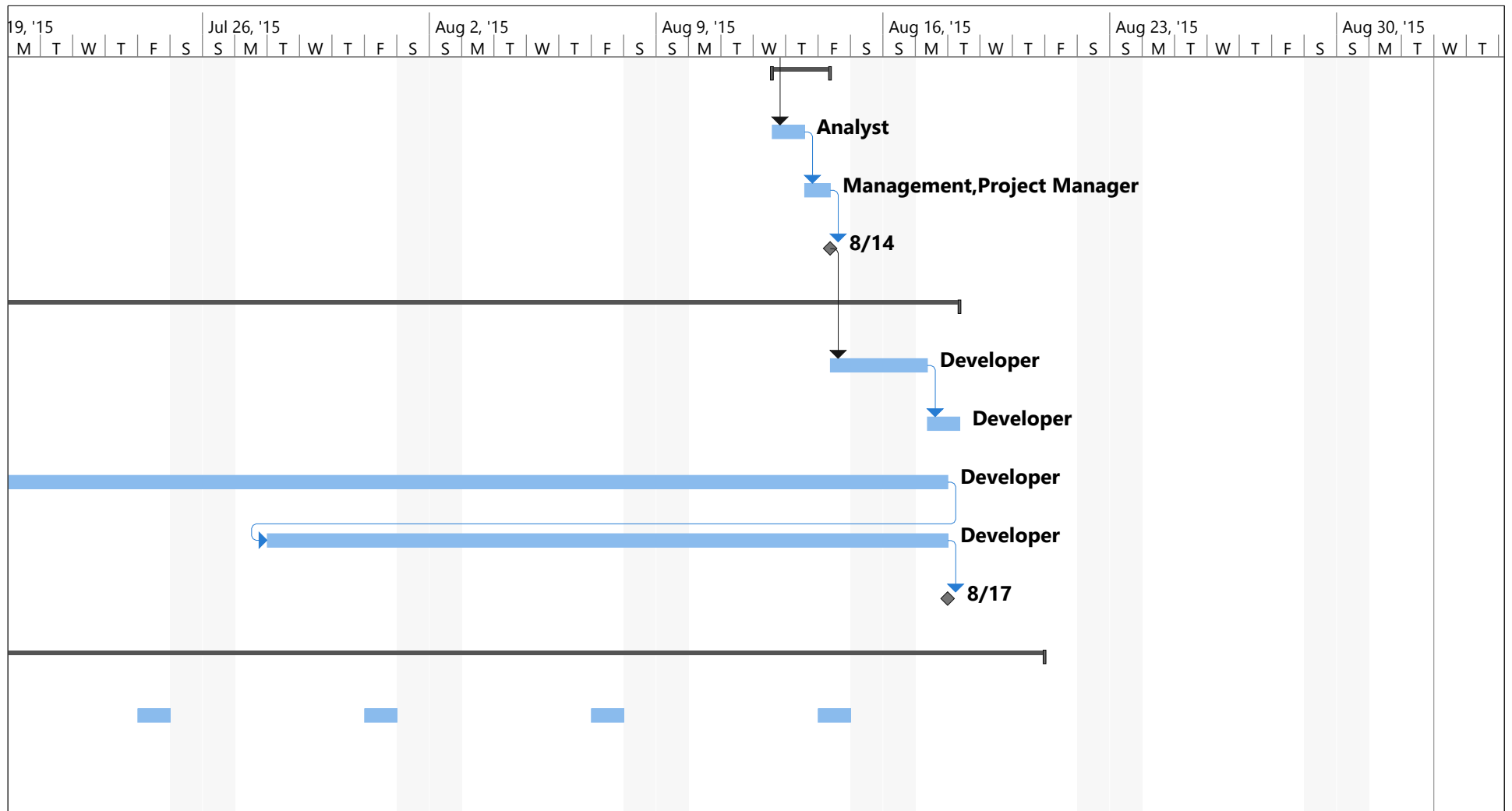





















[illegible]

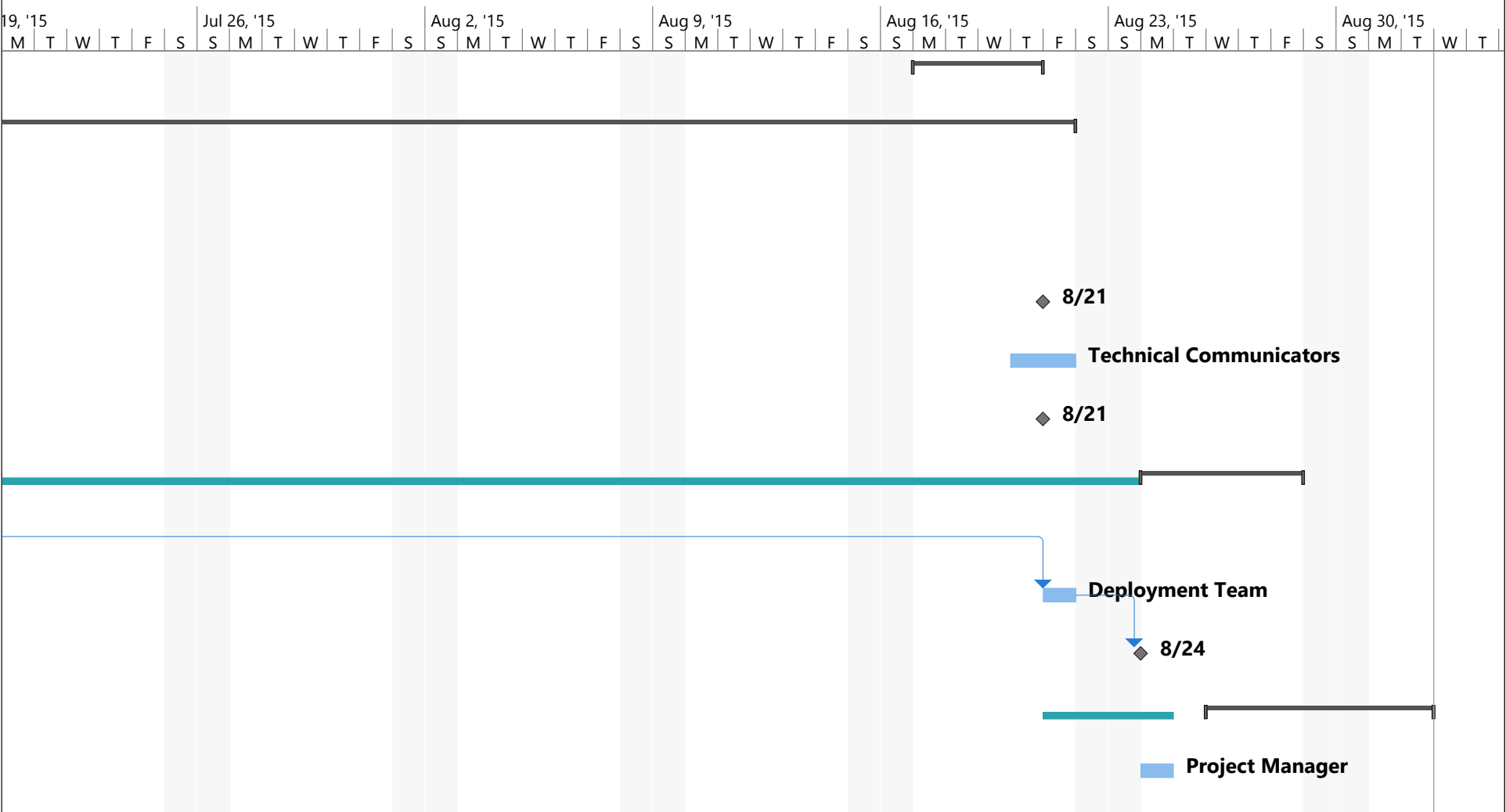


Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			

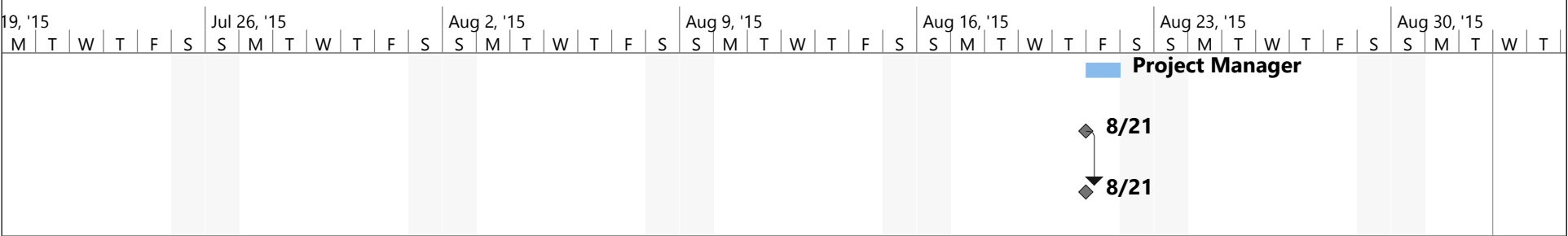




Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			



Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			



Project: Software Development Date: Thu 3/26/15	Task		Inactive Summary		External Tasks	
	Split		Manual Task		External Milestone	
	Milestone		Duration-only		Deadline	
	Summary		Manual Summary Rollup		Progress	
	Project Summary		Manual Summary		Manual Progress	
	Inactive Task		Start-only			
	Inactive Milestone		Finish-only			