// Benchmarking Project Schedules

Dr. Dan Patterson, PMP
CEO & President, Acumen

March 2011

ACUMEN
Proven Project Analytics

+1 512 291 6261 // info@projectacumen.com

www.projectacumen.com

Table of Contents

Introduction

Benchmarking project schedules is a powerful means of managing and validating projects, programs and portfolios alike. However, the approach to conducting such benchmarking, whether it is determination of schedule quality/realism or execution performance, can be highly subjective.

This paper introduces concepts that help provide a framework and repeatable approach to developing not only structurally sound schedules (based on accepted Critical Path/CPM techniques) but equally importantly, that bring realism to the schedule with regards to estimated durations, required resources and forecasted rates of execution.

Benchmarking Overview

Uses of Schedule Benchmarking

A benchmark is defined as "a point of reference by which something can be measured". Schedule benchmarking can be used to determine quality and accuracy of:

1. Structural integrity of the schedule – how well built is the schedule?
2. Estimated forecasts – how accurate are the duration estimates?
3. Execution performance – how well is the project performing?

The first two categories together form the basis of a realistic forecast (planning). The third category is a measure as to how well is the project being executed relative to the given forecast (the plan).

The basis against which benchmarking is conducted gives *context* to the results: is a project that is completed three months behind schedule but also three months earlier than any other similar project conducted to date, a successful project? The answer to this question is answered by firstly establishing an agreed upon benchmark basis.

The Benchmark Basis

With regards to the basis against which benchmarking is conducted, multiple sources can be selected:

1. Previous iterations or baseline(s) of the schedule: e.g. how is the plan quality or execution performance relative to last month's update?
2. Similar projects within the portfolio: e.g. based on similar scope and operating environment, how is our project performing relative to our sister project?
3. Industry standards: often based on standard productivity rates or estimates for given types of work or industry, duration estimates can be benchmarked.
4. Historical performance of the project: a useful insight into how well the project is performing relative to the agreed upon forecast.

The Value of Schedule Benchmarking

Schedule benchmarking helps set and achieve given targets so as to improve the (planning and/or execution) quality of a project.

During the planning phase, schedule benchmarking is a valuable tool for ensuring realistic and achievable forecasts that represent the work required to achieve project completion. Basing a project on a poorly planned schedule only leads to project failure – better to lead into the execution phase with confidence about the integrity of the schedule and it's forecasts than to simply rely on the hope that the forecast is achievable.

Performance tracking of a project is in itself also a form of benchmarking. By comparing the time, cost and resources used to achieve completed milestones and deliverables against the project plan or a baseline or some other basis is a means of deducing whether performance is acceptable or not. Execution remediation (or acceleration) is only possible once we have insight into where and when in the project are the performance bottlenecks – again, a simple but effective use of benchmarking (plan V actual).

Formal stage-gate process can also be driven by project benchmarking. Setting targets for schedule quality and estimates ensures that a project can only move forward once given targets are achieved. This not only increases the chance of project success but also helps increase the level of scheduling maturity within the project or organization.

In summary, the process of schedule benchmarking is a highly effective means of both ensuring the establishment of realistic plans and the subsequent successful execution of them.

Benchmarking Schedule Soundness

Schedule Benchmarks

A quality schedule needs to be both structurally sound as well as realistic in its estimates. Validation of structural soundness is best achieved through the analysis of the schedule using schedule metrics. The topic of a separate white paper[1], these metrics are warning indicators as to where shortcomings in the schedule lie (e.g. artificially constrained activities or networks paths with negative float). Tying these metrics back to the application of benchmarks, a further beneficial technique is to apply thresholds or tolerances to the metrics. As such, rather than failing a 5,000 activity schedule simply because of a total of 5 days negative float, instead consider setting a realistic target against which the total days negative float should not be surpassed.

Various organizations have developed benchmarking standards for the quality of schedules. A recent example of this, is the development of the 14-Point Assessment by the DCMA (Defense Contracts

[1] Dr Dan Patterson, "Taming an Unruly Schedule with the 14-Point Schedule Assessment", 2010, www.projectacumen.com/resources/whitepapers

Management Agency). As well as specific schedule metrics, the 14 Point Assessment also establishes specific benchmarks for schedule quality. Examples of these include:

Schedule Metric	Benchmark Threshold
Logic (open ends)	Should not exceed 5% of the schedule
Finish Start Relationships	Should represent at least 90% of the logic links
High Duration	Should not exceed 5% of the schedule
Negative Float	Zero tolerance

Figure 1 – Example DCMA Benchmark Thresholds

In a similar manner to the DCMA benchmarks, Figure 2 shows more involved examples of benchmarks that have been defined so as to improve the quality of project schedules.

Schedule Metric	Benchmark Threshold
Logic Density[2]	Should not be less than 2 nor greater than 4
Milestone Ratio[3]	Should not be less than 1:10
Activities Riding the Data Date	Should not represent more than 10% of the schedule
Dangling (open start/finish)	Should not exceed 5% of the schedule
Lags	Should not exceed 25% of the activity duration
Logic Hotspot	Should not exceed 20% of the schedule
Wrong Status	Zero tolerance
Cost/schedule misalignment	Zero tolerance

Figure 2 – Additional Schedule Benchmark Thresholds

[2] Logic density is defined as the average number of logic links per activity and is an indicator of quality of logic within a schedule
[3] Milestone ratio is the ratio of milestone and normal activities and is an indicator of level of detail within a schedule

Establishing a Scoring System for Benchmarks

One challenge of using multiple benchmarks is the ability to summarize multiple (often competing) benchmarks into a single result. This can be overcome by combining failed benchmark tests into a single schedule quality score. There are multiple variations in scoring techniques including weightings, weighted averages, score by exception etc. – all of these serve a common purpose: to provide a scoring mechanism that can be used in the benchmarking process.

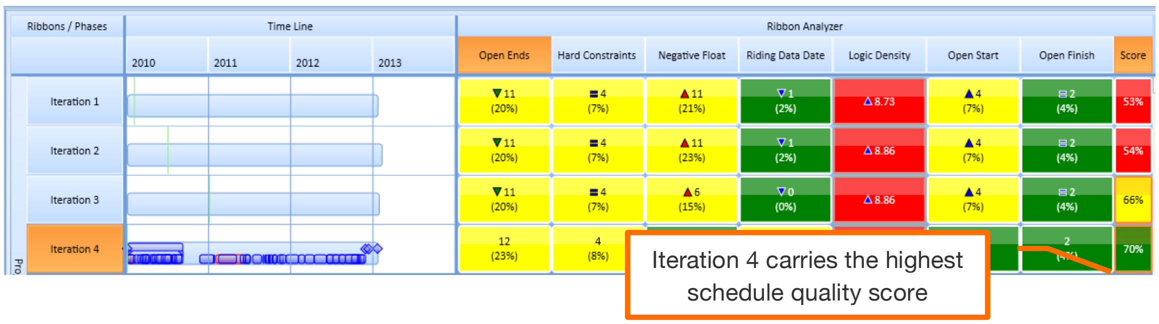

Figure 3 – Benchmark Analysis Using Weighted Scores

Figure 3 shows an example of four different iterations of a contractor schedule; each of which underwent various schedule revisions and updates. As part of the benchmarking exercise, the project owner established a benchmark pass/fail threshold of 70% based on a series of metrics. In all four iterations, the contractor consistently failed the logic density test but passed the other tests to varying degrees. As can be seen in figure 3, it took the contractor four iterations before achieving the target 70% score before passing the overall quality benchmark test.

Benchmarking Estimated Durations

Sound schedules also require realistic duration estimates – hence the need to benchmark the realism of duration estimates. As described earlier, the basis of the benchmark comparison can be from multiple sources but the approach for each is consistent.

Introducing Duration Dictionaries for Benchmarking Estimate Durations

Consider a scenario where a project schedule has been developed using duration estimates provided by multiple discipline leads. The information from these leads is accumulated by the lead scheduler and integrated into the project schedule. The next step is to establish whether the plan is a realistic plan or not. The benchmark technique used in this example is based on a "Duration Dictionary".
A duration dictionary is a look-up list of benchmark estimates categorized by a meaningful grouping that can be related back to and applied as a benchmark test against the durations in the schedule. The grouping of elements in a Duration Dictionary can be simple groupings based on Work Breakdown Structures, type of work, contractor type, location etc. within the project. Alternatively, duration dictionaries can be based on more complex bottom-up estimating approaches taking into account quantities of material, labor rates and productivity factors.

Figure 4 shows an example of a Duration Dictionary based on a corporate standard WBS Hierarchy. Three different benchmarks are referenced (Benchmark A, B and a historical reference project called Project B). These three sets of benchmarks can then be used as the basis against which to benchmark the accuracy of the duration estimates being developed.

1	WBS	Details	Benchmark A	BenchMark B	Project B
2	1	Concept	110	159	119
3	1.1	Alternate scenario development	88	127	95
4	1.1.1	In-House scenario	36	52	39
5	1.1.2	Bid review	15	22	17
6	1.2	Competitive Analysis	5	7	5
7	1.3	Requirements Definition	9	13	10
8	1.4	Technical review	23	33	25
9	1.5	Commerical review	3	4	3
10	2	Early Design	69	100	75
11	2.1	Comms design	15	22	17
12	2.2	Civil design	69	100	75
13	2.3	Mechanical design	23	33	25
14	2.4	Electrical design	2	3	2
15	2.5	Telecoms Design	27	39	29
16	3	FEED	101	146	110
17	3.1	FEED handover	23	33	25
18	3.2	FEED study	14	20	15

Figure 4 – Example WBS-Based Duration Dictionary

Once the Duration Dictionary has been established, the benchmark data can then be merged with the project schedule (in this example, based on WBS) and a subsequent metric analysis conducted.

Figure 5 shows an example of the previously referenced Duration Dictionary merged with the project schedule. The next step is to run a benchmark analysis to validate the quality of the forecasted durations.

#	Project	ID	Description	Activity Type	Baseline Duration	Original Duration	Remaining Duration	Merge.Benchmark A	Merge.BenchMark B	Merge.Project B
1	CurrentSchedule	1	Concept	Summary	0	122	0	110	159	119
2	CurrentSchedule	2	Alternate scenario development	Summary	0	98	0	88	127	95
3	CurrentSchedule	3	In-House scenario	Normal	0	40	0	36	52	39
4	CurrentSchedule	4	Bid B review	Normal	0	17	0	15	22	17
5	CurrentSchedule	5	Bid A review	Normal	0	16	0			
6	CurrentSchedule	6	Competitive Analysis	Normal	0	5	0	5	7	5
7	CurrentSchedule	7	Requirements Definition	Normal	0	10	0	9	13	10
8	CurrentSchedule	8	Technical review	Normal	0	25	0	23	33	25
9	CurrentSchedule	9	Commerical review	Normal	0	3	0	3	4	3
10	CurrentSchedule	10	Early Design	Summary	0	77	[illegible]	69	100	75
11	CurrentSchedule	11	Comms design	No[illegible]	[illegible]	[illegible]	0	15	22	17
12	CurrentSchedule	12	Civil design	No[illegible]	[illegible]	[illegible]	0	69	100	75
13	CurrentSchedule	13	Mechanical design	No[illegible]	[illegible]	[illegible]	0	23	33	25
14	CurrentSchedule	14	Electrical design	No[illegible]	[illegible]	[illegible]	0	2	3	2
15	CurrentSchedule	15	Telecoms Design	No[illegible]	[illegible]	[illegible]	0	27	39	29
16	CurrentSchedule	16	FEED	Su[illegible]	[illegible]	[illegible]	53	101	146	110
17	CurrentSchedule	17	FEED handover	No[illegible]	[illegible]	[illegible]	25	23	33	25
18	CurrentSchedule	18	FEED study	No[illegible]	[illegible]	[illegible]	15	14	20	15

Benchmark Durations from the Duration Dictionary

Figure 5 - Schedule and Duration Dictionary Merge

Using Rate Tables as the Basis for Duration and Cost Estimation Benchmarking

An alternate approach to developing benchmarking standards based on Duration Dictionaries is to adopt more of a “bottoms-up” approach using rates and quantities. By establishing rates and quantities (of resources and materials) against activities, a more detailed (and accurate) benchmark basis can be established. Further, availability of rates and quantities is generally commonplace within projects and so makes for easier benchmark basis development.

Activity duration can be determined from assigned resource rates. For example if the daily rate for laying pipe is known and the quantity of pipe to be laid is also known, then the activity duration can be calculated through a simple quantity/rate calculation. By defining quantities in a schedule and applying rates to these quantities from a rate table, accurate benchmark durations can be established and subsequently used for the basis of a benchmark analysis

Activity	Rate	Quantity	Benchmark Duration
Pipeline Lay	100 ft/day	1000 feet	10 days
Concrete Pour	3 cubic yards/day	18 cubic yards	6 days

Figure 6 – Rate-Based Benchmark Duration Calculation

Running Benchmark Metrics

Various metrics can be used for a schedule benchmark analysis including:

- *Optimistic* – schedule duration is less than the suggested benchmark value
- *Pessimistic* - schedule duration is more than the suggested benchmark value
- *Within Tolerance* – schedule duration is within a given percentage of the benchmark duration e.g. within 10%
- *Outside Tolerance* - schedule duration is outside a given percentage of the benchmark duration e.g. outside 10%.

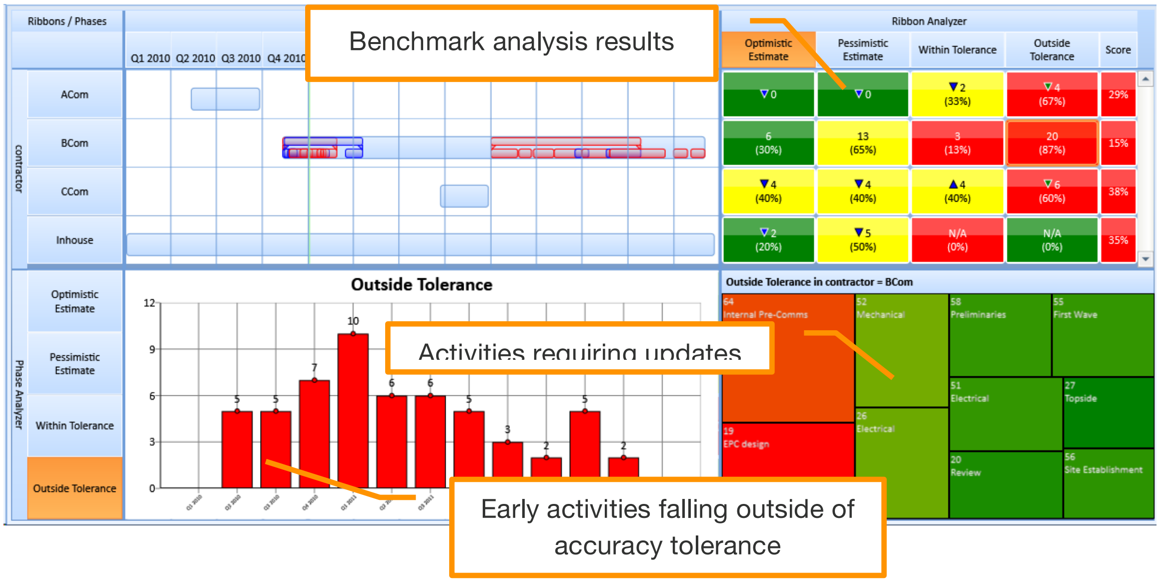

Figure 7 – Duration Benchmark Analysis

Figure 7 shows the results from applying these metrics to our example project based on groupings of activities (by contractor). Multiple conclusions from the analysis can be drawn including:

1. The majority of the “out of tolerance” estimates fall in the early/mid phase of the project.
2. The majority of these out of tolerance estimates are as a result of Contractor B’s very pessimistic schedule.
3. Specific activities requiring schedule estimate reviews can be pinpointed using the heat map chart and the metric analysis results.

From the analysis, those activities requiring duration updates can be quickly determined, reviewed and updated accordingly. Being able to conduct benchmark analysis on logical groupings of activities is a very effective means of determining trends and characteristics within a schedule, which in turn, leads to faster resolution.

Optimistic Estimate

ID	Description	Activity Type	Baseline Duration	Original Duration	Remaining Duration	Total Float	Free Float	Start	Finish	Actual Start
33	Vendor A	Normal	0.00	25.00	25.00	-7.00	0.00	6/5/2011	7/8/2011	
32	Vendor B	Normal	0.00	15.00	15.00	0.00	0.00	6/15/2011	7/5/2011	

Pessimistic Estimate

ID	Description	Activity Type	Baseline Duration	Original Duration	Remaining Duration	Total Float	Free Float	Start	Finish	Actual Start
34	Outsourced PMO	Normal	0.00	334.00	95.00	0.00	53.00	2/1/2010	5/12/2011	2/1/2010
31	Initial Long Lead items	Normal	0.00	90.00	90.00	-7.00	0.00	2/2/2011	6/7/2011	
35	Secondary Long Leads	Normal	0.00	60.00	60.00	-48.00	0.00	2/14/2011	5/6/2011	
36	Vendor C	Normal	0.00	20.00	20.00	0.00	0.00	7/6/2011	8/2/2011	
30	Bid reviews	Normal	0.00	30.00	30.00	0.00	0.00	8/3/2011	9/13/2011	

Within Tolerance

ID	Description	Activity Type	Baseline Duration	Original Duration	Remaining Duration	Total Float	Free Float	Start	Finish	Actual Start
6	Competitive Analysis	Normal	0.00	5.00	0.00	0.00	0.00	1/15/2010	1/22/2010	1/15/2010
29	Procurement	Summary	0.00	422.00	246.00	0.00	0.00	2/1/2010	9/13/2011	2/1/2010

Outside Tolerance

ID	Description	Activity Type	Baseline Duration	Original Duration	Remaining Duration	Total Float	Free Float	Start	Finish	Actual Start
1	Concept	Summary	0.00	122.00	0.00	0.00	0.00	1/1/2010	6/22/2010	1/1/2010
7	Requirements Definiti	Normal	0.00	10.00	0.00	0.00	0.00	1/1/2010	1/15/2010	1/1/2010
2	Alternate scenario dev	Summary	0.00	98.00	0.00	0.00	0.00	1/22/2010	6/9/2010	1/22/2010
3	In-House scenario	Normal	0.00	40.00	0.00	0.00	0.00	1/30/2010	3/29/2010	1/30/2010
34	Outsourced PMO	Normal	0.00	334.00	95.00	0.00	53.00	2/1/2010	5/12/2011	2/1/2010
4	Bid B review	Normal	0.00	17.00	0.00	0.00	0.00	5/15/2010	6/9/2010	5/15/2010
9	Commerical review	Normal	0.00	3.00	0.00	0.00	0.00	6/17/2010	6/22/2010	6/17/2010
31	Initial Long Lead items	Normal	0.00	90.00	90.00	-7.00	0.00	2/2/2011	6/7/2011	
35	Secondary Long Leads	Normal	0.00	60.00	60.00	-48.00	0.00	2/14/2011	5/6/2011	
33	Vendor A	Normal	0.00	25.00	25.00	-7.00	0.00	6/5/2011	7/8/2011	
32	Vendor B	Normal	0.00	15.00	15.00	0.00	0.00	6/15/2011	7/5/2011	

Figure 8 – Checklist of Activities Requiring Duration Revisions

Appropriate tolerance levels vary depending on where in the project lifecycle you are. Humphrey and Associates[4] define four discrete types of estimate (conceptual, preliminary, detailed and definitive). These estimates vary in accuracy based on the amount of information available at the time:

- Conceptual: -25% > +40%
- Preliminary: -15% > +25%
- Detailed: -10% > +20%
- Definitive -5% > +10%

These thresholds should be taken into account when defining tolerance levels assigned to benchmark metrics. i.e. metric tolerances should be updated as the project lifecycle progresses.

Using benchmarking to attain high levels of schedule quality and estimate accuracy are together two key drivers of project success. In a paper authored by Narayan Joshi, titled "Benchmarking and best Practices for Effective Turnarounds"[5], it was found that the most competitive (successful) turnaround projects were those that had the most mature scope definition, best defined execution strategy and the best planning status. All three of these factors are based upon having a sound and realistic plan in place, a very compelling argument for conducting benchmark analysis as early in the project lifecycle as possible.

Benchmarking Execution Performance

Probably one of the most common uses for benchmarking is for performance analysis. Being able to compare a project's performance against given benchmarks gives objective and quantitative insight into the true performance of a project.

[4] Humphreys and Associates, Project Management Using Earned Value, 2002, ISBN 0-9708614-0-0

[5] Narayan Joshi, "Benchmarking & Best Practices for Turnarounds", IQPC Turnaround Conference, London, UK, Sep 22 2003

There are multiple techniques available for project execution performance benchmarking ranging from rudimentary metrics such as variance from a given baseline to true performance analysis using earned value and/or earned schedule techniques.

Performance Relative to a Baseline

Tracking project execution relative to a baseline involves straightforward cost and/or schedule performance metrics comparing actual performance with that forecasted in a given baseline or original plan. Useful metrics include:

Metric	Description
Completed Ahead of Schedule	Activities that completed ahead of baseline irrespective of when they started.
Completed Behind Schedule	Activities that completed behind baseline forecast irrespective of when they started.
Accelerated	Activities that started late but finished early.
Took longer than planned	Irrespective of dates, the activity took longer to execute than originally planned
Started Ahead	Actual Start date of the activity was earlier than planned
Started Delayed	Actual Start date of the activity was later than planned
Riding the Data Date	Planned activities whose start dates get delayed along the project data date as a result of not starting on time.

Figure 9 – Example Schedule Performance Metrics

Schedule performance metrics are valuable for benchmarking execution relative to the plan. Figure 10 shows the metrics listed in figure 9 applied to a schedule grouped by contractor. As well as tracking the number of activities in a schedule that trigger a performance metric threshold, quantifying the actual **number** of days work, or effort or cost that these activities represents is a better means of actually quantifying the true impact of these performance variances. Knowing that 100 activities fell behind schedule is less valuable than knowing that this represented a total slippage of 250 labor days, for example.

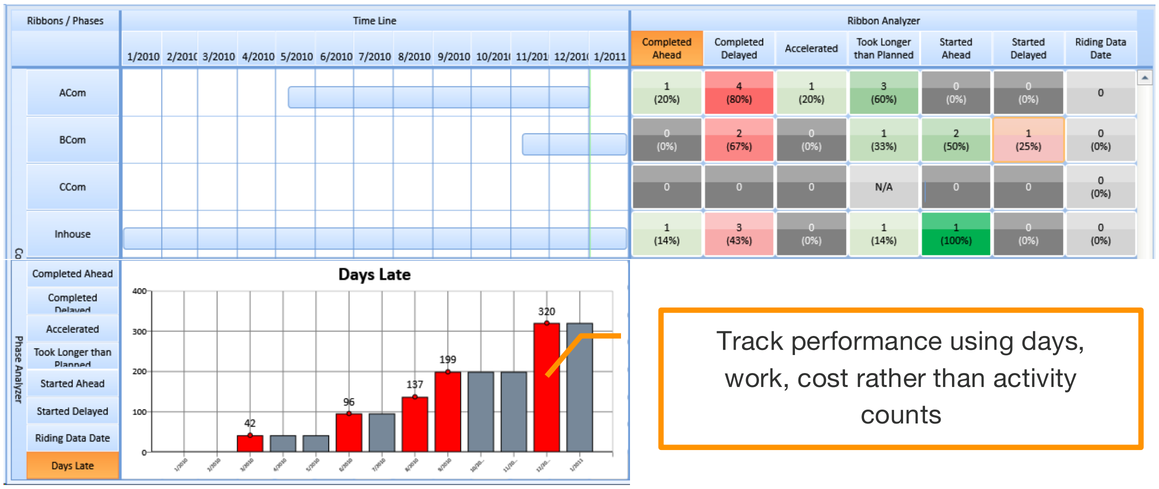

Figure 10 – Benchmarking Performance Over Time

Another useful benchmark is to differentiate slipped duration based on whether or not the activities in question are on the critical path. Figure 11 shows the same example project with execution performance benchmark metrics applied to both the critical and non-critical activities. In this example, conclusions can be drawn as to the relative performance of the activities on and off the critical path giving additional insight above and beyond that of simply benchmarking at the project level as a whole.

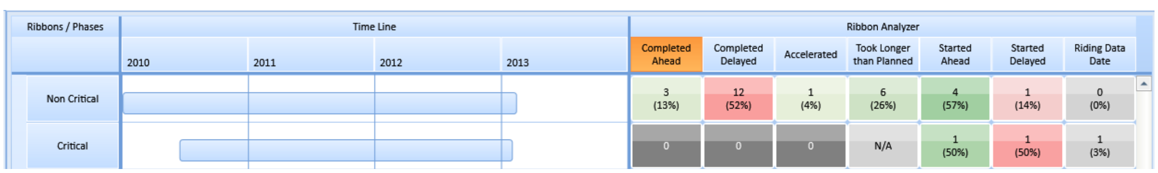

Figure 11 – Benchmark Comparison Between Critical and Non-Critical Activities

In summary, schedule performance analysis gives valuable insight into how well a project is executing relative to a given baseline. Further value can be obtained from the analysis by focusing on actual quantities of duration and work involved rather than simply looking at activity counts and/or percentages.

Earned Value

One of the drawbacks of using traditional performance metrics is that the analysis doesn't take into account the effort expended in order to achieve the current status. This shortcoming is overcome through the use of Earned Value. Earned value is essentially a comparison of three variables:

- How much progress/work should have been completed (planned cost)
- How much progress/work actually has been completed (earned value)
- How much effort has it taken to achieve the current status (actual cost)

By comparing the value of work achieved to how much it cost to achieve this progress relative to the expected progress is an excellent means of not only tracking progress but also efficiency of work. Being two weeks early on the completion of an activity but at the expense of being three times over budget is not insight that traditional performance-based benchmark analysis can provide.

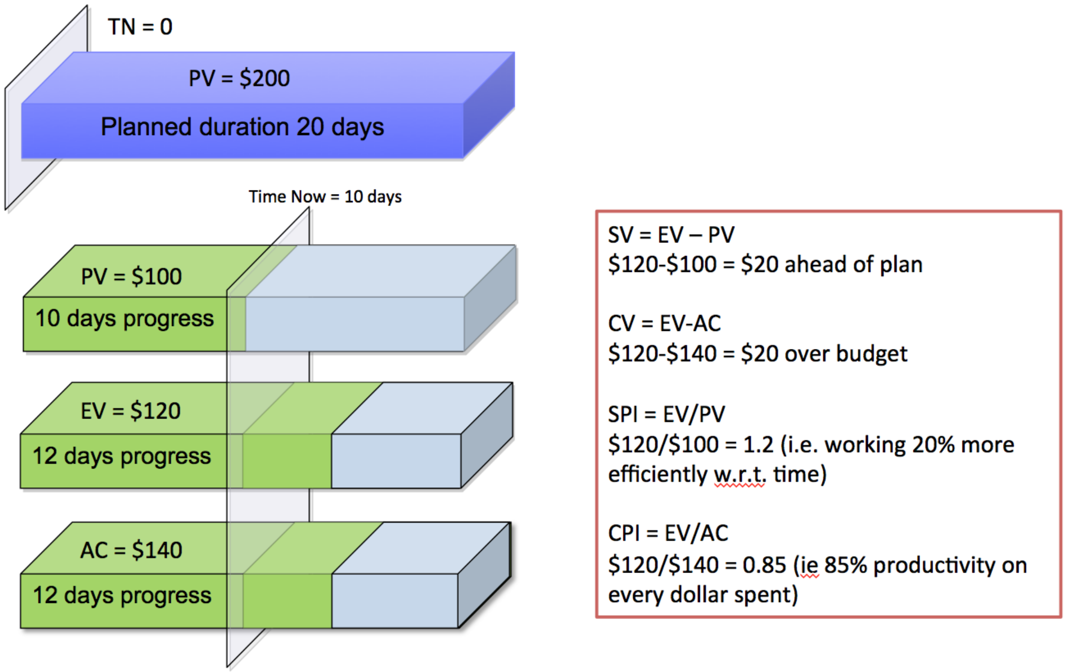

Figure 12 – Two-Way Benchmarking using Earned Value

Figure 12 shows an example activity with a planned duration of 20 days and planned cost of #200. Ten days into the project, the assumption is that 10 of the 20 days of planned duration should be completed (at a cost of $100). In actuality, 12 days of progress has been achieved (ahead of schedule). However, when the actual cost of work performed is examined, it can be seen that it has cost $140 to achieve the 12 days worth of progress. Relating this back to benchmarking, consider Earned Value as a two-way benchmark test:

Schedule Performance Index (SPI) is an indication of how much progress has been made relative to the expected progress at that point in time. In this example, the CPI is 1.2 reflecting a 20% improvement in efficiency than originally forecast.

Cost Performance Index (CPI) is an indication of how much progress has been made relative to the cost of work. In this example the CPI equates to 0.85 indicating that for every dollar spent, only 85 cents of work has actually been achieved.

Figure 13 shows our example project undergoing an earned value analysis being sliced and diced by both Contractor and phase. This two dimensional analysis gives insight into not only which contractor is causing the largest cost variance but also when in the project lifecycle is this variance occurring.

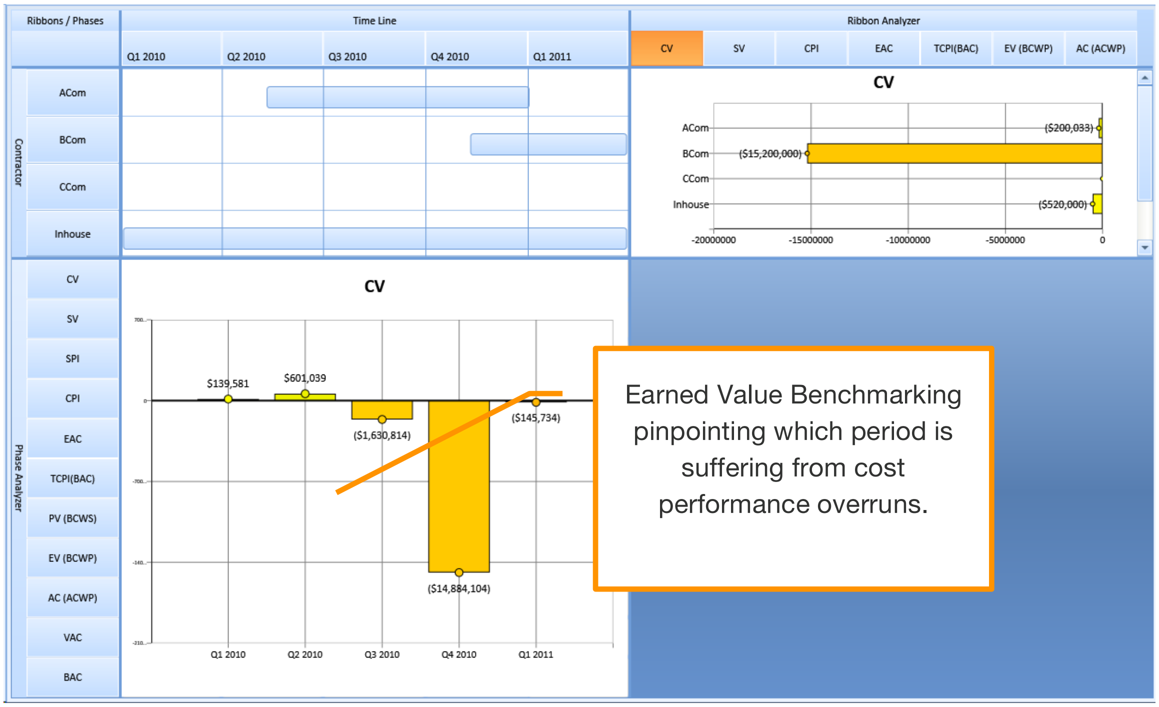

Figure 13 – Earned Value using Slice-Dice Analysis

Earned Value uses planned cost and duration as the basis for the benchmark comparison. This is all very well assuming that the planned values are accurate and realistic in the first place. This ties back to the importance highlighted in the initial two chapters of this paper discussing quality of schedule and accuracy of forecast. Earned value benchmark analysis is only as accurate as the plan/forecast basis. Further argument for spending as much effort as needed on ensuring that the basis of schedule and estimate is indeed realistic and defendable.

Earned Schedule

Earned value does however come with a price: that is, the effort involved in capturing the very detailed periodic status updates/actual costs tends to be much greater than simply capturing status in the form of percent complete. Further, not all projects directly tie their schedules with cost estimates. In such instances, it is not possible to combine schedule status with actualized costs and so earned value analysis is not possible. As an alternate solution, the concept of earned schedule has been developed. Earned schedule compares how much progress has been achieved with how much progress should have been achieved based on a given basis or baseline.

Figure 14-shows an example of Earned Schedule analysis. Twelve days into the project, the activity in question has only achieved 10 days worth of progress – therefore the Schedule Variance (t) or SV(t) is -2. In terms of a performance index, examining the Schedule Performance Index (t) SPI (t), reveals an efficiency of 0.83 (less work completed than expected).

Extrapolating this past performance out to forecast an actual completion duration is achieved using a metric known as Independent Estimate at Complete IEAC(t). In this example the IEAC(t) produces a completion forecast of 24 days (based on past performance) relative to the original forecast of 20 days.

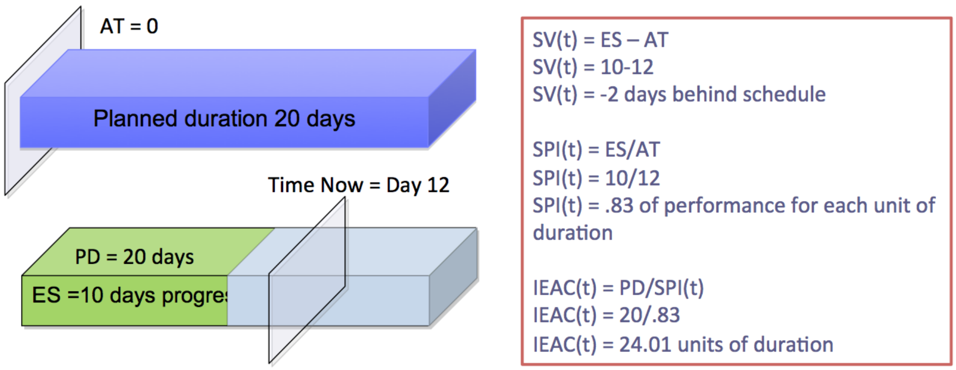

Figure 14 – Earned Schedule Benchmarking

Figure 15 shows the same sample project as previously demonstrated for the Earned Value scenario. The Benchmark in this example is the schedule baseline which enables earned schedule analysis giving insight into not only which contractors are over or under performing but also when in time are the performance bottlenecks occurring.

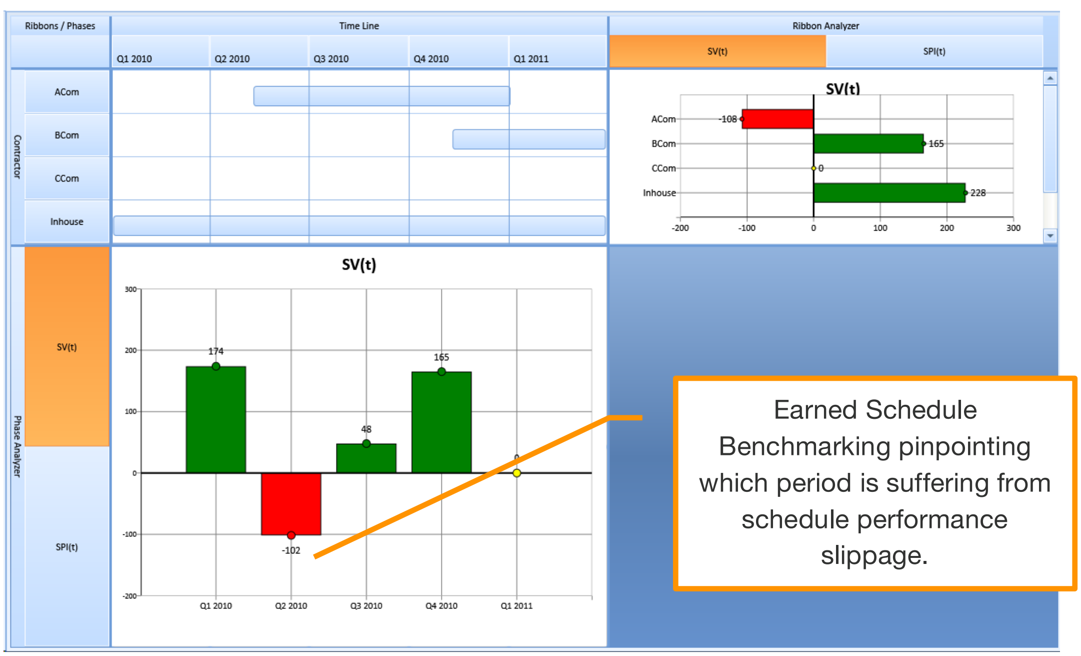

Figure 15 – Earned Schedule Benchmark Metric Analysis

Conclusions

This paper has discussed applications of schedule benchmarking under different use-cases:

- Critiquing the structural integrity of a schedule

- A highly effective means of validating durations of activities
- A means of gaining true insight into project execution performance

While these three applications of benchmarking all are very valuable, they all are based on the assumption that the underlying basis for comparison (the benchmark basis) is sound and realistic. This is yet further proof that in the world of project management, a “sound basis of schedule” is king. The use of benchmarks to develop sound, realistic and ultimately more achievable, project plans is key to achieving such schedules. If achieved, this in turn leads to a much higher chance of on-time project execution and overall project success.