Improving Verification Predictability and Efficiency Using Big Data

Darron May, Mentor Graphics Inc, Darron_May@mentor.com

Abstract

Big Data is a term that has been around for many years. The list of applications for Big Data are endless, but the process stays the same: capture, process and analyze. So why shouldn't this technology help improve your verification process efficiency and predict your next chip sign-off? By providing a Big Data infrastructure, with state of the art technologies, within the verification environment, the combination of all verification metrics allows all resources to be used as efficiently as possible and enables process improvements using predictive analysis. This paper covers the technology, the metrics, and the process, and it will explore a number of techniques enabled by such an infrastructure.

What is Big Data?

The definition of Big Data was always targeted towards sets of data so large that traditional processing software is inadequate to deal with them in a useable amount of time. This was defined by the use of the 3 V's terminology.

- Increasing VOLUME, the amount of data.
- VELOCITY, the speed of data movement.
- VARIETY, the range of data types and sources.

This tends to leave the opinion that the well understood verification process with its well-understood data generation is not really Big Data but the definition has been updated as it is not only the data but the technology and methods used to transform its value that matter. These now include:

- Machine learning to detect patterns and make predictions.
- Digital footprint, which is a cost-free by-product of digital interaction.

Many industries are being transformed by Big Data analytics and machine learning. These include:

- Commerce and advertising.
- Entertainment.
- Energy industry.
- Healthcare & Medical science.
- Law and Order ….

Many of the concepts used within these industries can be applied to the data within the verification environment.

The Typical Verification Environment

Verification is typically driven by the testplan which is derived from the design specification and defines how to test or verify design features. These features are typically tested by either a directed test or by some form of coverage or combination of coverage that records when a particular feature has been tested.

A number of separate systems are required in a flow to manage the verification process. A Source Code Management (SCM) system to store HDL/HVL. A build and regression management system is used to run various verification tools to verify functionality within the specification has been covered. Bugs are tracked within a bug tracking system, and coverage closure is achieved by measuring both code and functional coverage and cross-referencing the testplan.

All of the systems and stages produce data. The compute resources are typically controlled by a load-sharing system, and the licenses by a licensing system. Again, typically data from these sources is not gathered over the period of the verification process. A complete picture of a typical verification process and the systems that produce data can be seen in (Figure 1). – The Typical Verification Environment.

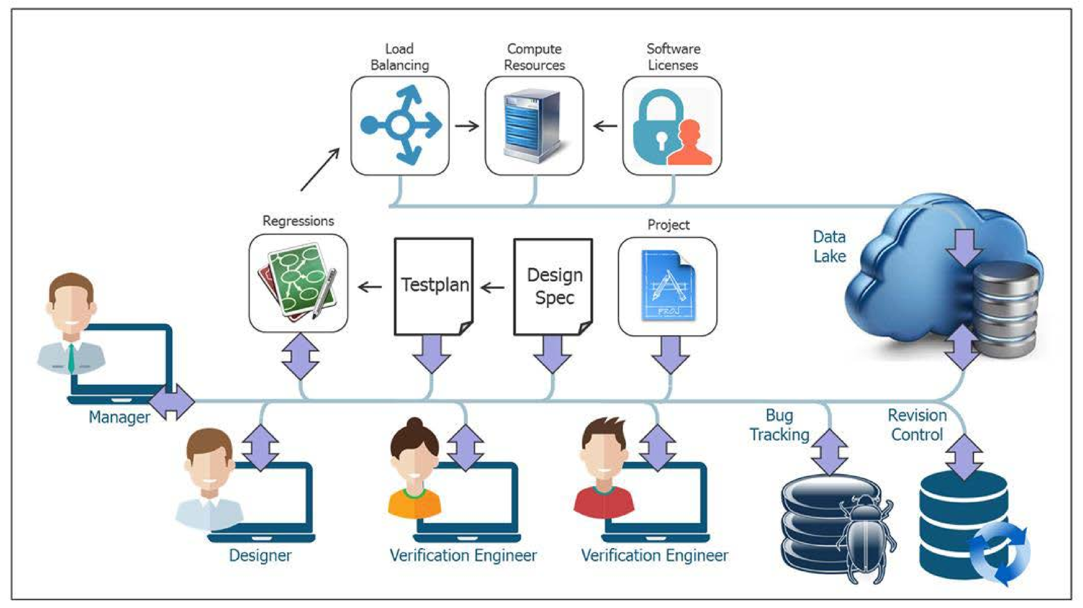


Figure 1 – The Typical Verification Environment

There are many other metrics that could be gathered during the duration of a project if there were an infrastructure to store them over time. Some of the systems described above have facilities to store information over time and maybe even analyze the data over time, but this only allows the metrics to be seen in isolation. With projects and teams spread geographically, what is required is a system that allows all of this data to be saved in a central repository that is visible in all locations.

By gathering data across many projects over a period of time, the data value will improve just like any of the Big Data applications. This will allow predictive analysis techniques to be used both within a project and across projects to help improve the process.

New Enabling Technologies

One of the major requirements when harvesting vast amounts of data is to have somewhere to store it. Enterprise databases require heavy initial investment in resources, time and money. This means that a generic backend infrastructure must give the flexibility to choose the database technology but still provide security, authentication and maintenance. Once the backend infrastructure is in place, the major investment involves building the business logic to achieve the required analysis and views. The major blocks of the system can be viewed in (Figure 2).

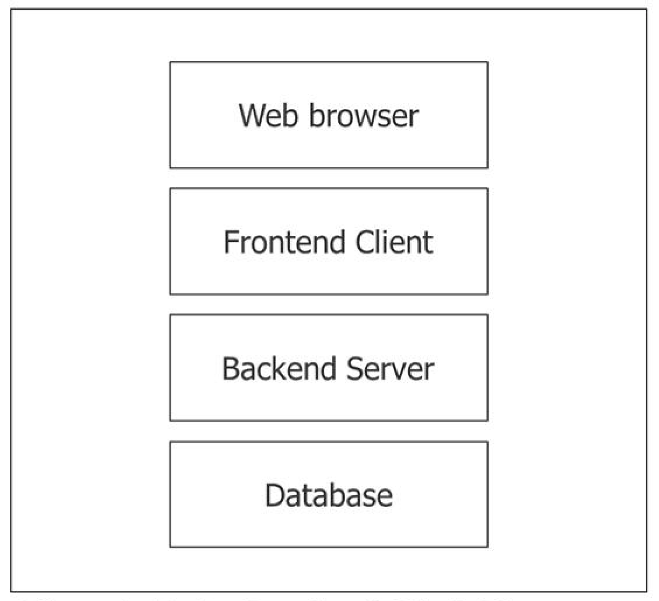


Figure 2 -Technology Stack Block Diagram

There are many technologies that can be used in the blocks shown above. These include:

- Database
  - Structured – Microsoft SQL Server, SQLite, Oracle, PostgreSQL, and IBM DB2
  - Unstructured - MongoDB, Hadoop, Cassandra, CouchDB, HBase, and Redis
- Backend
  - JAVA, Ruby on Rails, Django
- Frontend
  - HTML5, Advanced JavaScript, AngularJS, Bootstrap, NodeJS

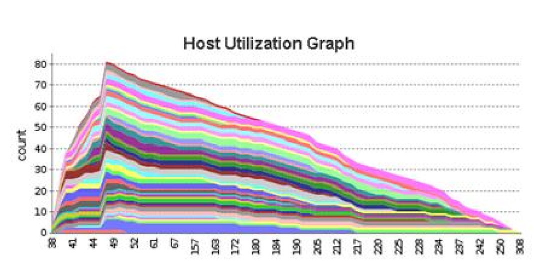
It is beyond the scope of this paper to recommend a Big Data technology to use within the verification process, but a commercially available tool would need to support multiple types of databases and be web-ready.

Single and Cross Metric Analysis

The system needs the ability to run extractors either periodically or on the occurrence of certain events to capture metrics such as the following:

- Project – Milestones, Requirements & Testplans
- Process - Source Code Management, Bug Tracking, Regression & Coverage
- Resources – Grid, File & Licensing Systems

Trending metrics detailed from the sources listed can give a very good indication of how much more time is needed to complete sign-off. The ability to trend multiple metrics on the same graph can provide increased insights.



Greater insights into the data can also be achieved by storing extra meta-data to cross domains, for example ID's to link the source code changes with fixed bugs and vice versa.

The ability to add smarts to the system allows many other functions to improve efficiency from data analysis.

- Prediction across multiple projects.
- Alarms and Notifications.
- Housekeeping & Down sampling.

Data Mining and Predictive Analysis

Maturing data analytics is driving the demand for automated decision-making based on reliable predictive analytics. (Figure 3) highlights the maturing of data and analytics, illustrating how we have moved from descriptive and diagnostic analytics to predictive and prescriptive analytics, resulting in the unleashing of the full potential of data.

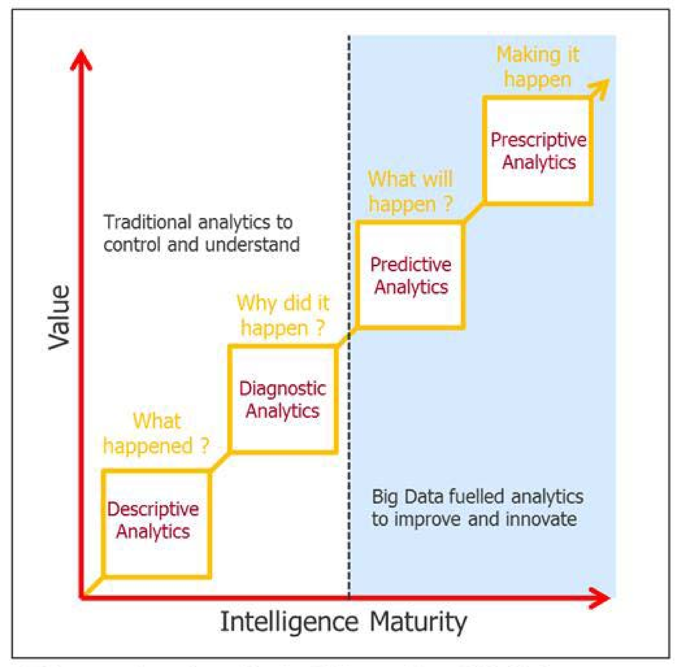
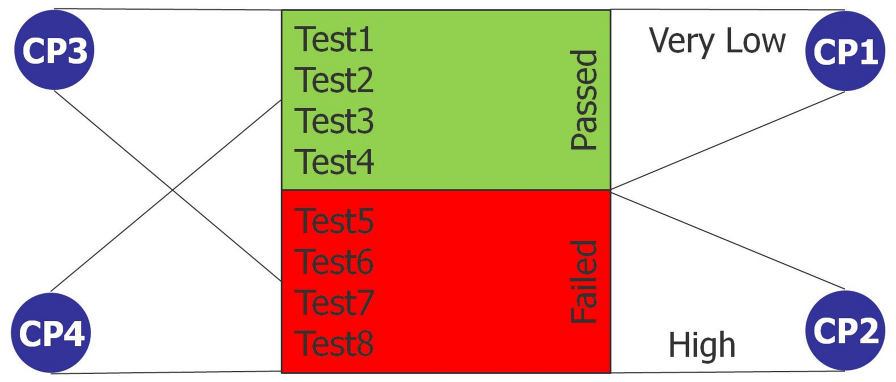


Figure 3 – Analytic Maturity Model

The debug cycle, and where to start it, are two of the most time consuming aspects within the verification process. With all coverage and assertion data in one place, it is possible to utilize the data from regression runs to suggest starting points for debug.



Conclusion

Big Data is not just about harvesting vast amounts of data but also allowing the exploration of the different relationships you can start to see by combining data captured from the verification process. Using some of the readily available technologies, it is possible to start to gather very valuable data, which can help today's and more importantly tomorrow's projects. Data mining and predictive analysis can be applied to the data within the verification process to significantly improve the predictability of the entire process.