Service Quality Assessment for NASA's Deep Space Network: No Longer a Luxury

Erik Barkley, Paul Wolgast, Silvino Zendejas
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Dr
Pasadena, CA 91101
{Erik.Barkley, Paul.Wolgast, Silvino.Zendejas}@jpl.nasa.gov

Abstract—When NASA's Deep Space Network (DSN) was established almost a half century ago, the concept of computer-based service delivery was impractical or infeasible due to the state of information technology As a result, the interface the DSN exposes to its customers tends to be equipment-centric, lacking a clear demarcation between the DSN and the mission operation systems (MOS) of its customers. As the number of customers has continued to increase, the need to improve efficiency and minimize costs has grown. This growth has been the impetus for a DSN transformation from an equipment-for-rent provider to a provider of standard services. Service orientation naturally leads to requirements for service management, including proactive measurement of service quality and service levels as well as the efficiency of internal processes and the performance of service provisioning systems. [1]

DSN System Engineering has surveyed industry offerings to determine if commercial successes in decision support and Business Intelligence (BI) solutions can be applied to the DSN. A pilot project was initiated, and subsequently executed to determine the feasibility of repurposing a commercial Business Intelligence platform for engineering analysis in conjunction with the platform's intended business reporting and analysis functions. [2]

This paper surveys the challenges, the lessons learned, and highlights the interesting results to date. It will also highlight the technologies applied to achieve these goals including:

- Business Intelligence Software
- Data Warehousing
- ETL (Extract, Translate and Load) techniques
- Complex Event Processing

TABLE OF CONTENTS

1. INTRODUCTION

The DSN was formally organized in late 1963. At that time, the concept of computer-based service delivery was impractical due to the state of information technology. As such, the interface that customers came to know had no clear demarcation between the provider of space communications services - the DSN - and the mission operation systems using these services. It was and still is common to think of the DSN in terms of its equipment as opposed to the services it delivers. This is evident in the monitor data provided to customers as services are provided -- for example, parameters are expressed with regard to internal switch settings in relation to a particular hardware asset. By contrast, a service orientation expresses information with regard to the salient characteristics important to the customer; for example, statement of service configuration including left-hand circular polarization as opposed to reporting a series of parameters which happens to include microwave switch settings. This in turn is identified as part of a telemetry service as opposed to an x-band downlink (and the telemetry service also includes frequency identification as part of its configuration statement). Figure 1 identifies the DSN services and service interfaces for a typical MOS and its spacecraft.

The DSN is moving to a services paradigm that results in a more-to-the-point interface with regard to the services requested and provided to its customers. The shift to a services approach obviates the need for customers to maintain cognizance about internal DSN equipment details while simultaneously allowing the DSN a greater degree of flexibility in deploying equipment to meet service-level commitments. However, this approach also obligates the DSN to be much more knowledgeable about how its equipment is behaving to properly analyze, assess, and report on activities at the service level. Rather than the customer monitoring the detail equipment level information, the DSN performs this activity for all missions simultaneously, assessing its overall service delivery and taking action appropriately. This is a significant departure from current DSN operations and requires collecting,

[1] 978-1-4244-3888-4/10/$25.00 ©2010 IEEE.
[2] IEEEAC paper #1631, Version 3 Updated December 15, 2009

aggregating, and analyzing a large volume of data on a continuous/on-going basis to assist with difficult prioritization and investment decisions.

Making the right decisions necessitates factual and reliable information. Current methods used to derive metrics and generate reports require significant manual efforts to extract, cleanse, integrate, and analyze data from a variety of structured and semi-structured sources. The infusion of commercial off-the-shelf (COTS) solutions to facilitate or automate this type of work needs to be analyzed against the continued use of highly manual methods and in-house solutions that are subject to human error and subjective evaluations.

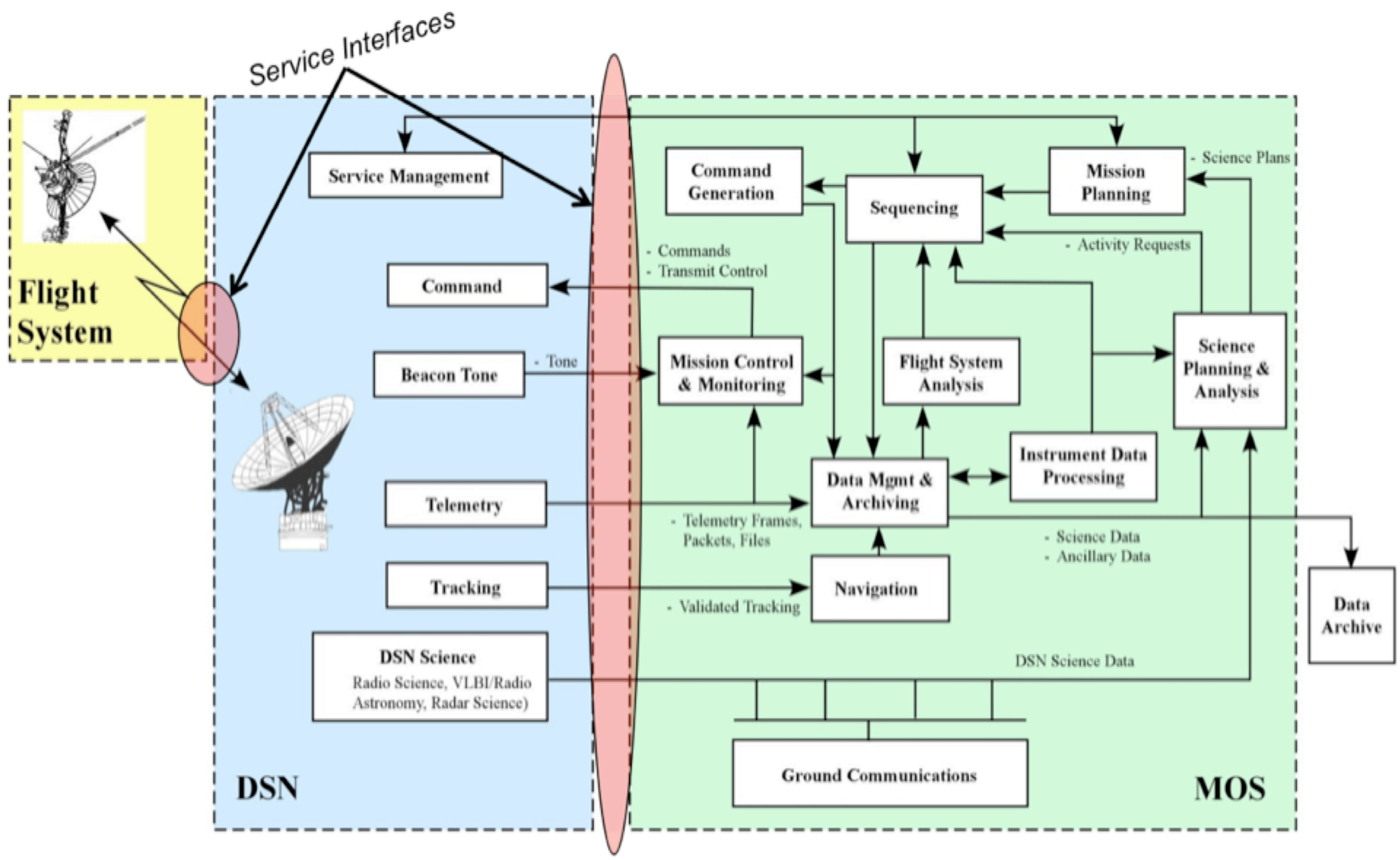


Figure 1. DSN Service Interfaces

During the pilot project, it was noticed that much of the information the DSN routinely publishes about subsystem and assembly performance could be used to support analysis and data mining activities using modern techniques in Business Intelligence (BI). Until now only operations personnel and cognizant individuals could interpret and analyze this data via labor-intensive methods. As a result of the Pilot, the DSN is now in the process of developing an analysis, trending, and data mining capability via the Service Quality Assessment (SQA) subsystem. SQA relies on a data warehouse of schedule, discrepancy, network control, and execution monitoring data automatically extracted from the various source systems and exposed to end users via a BI COTS toolset.

2. BUSINESS INTELLIGENCE

"Wikipedia[1] defines business intelligence as: "**Business Intelligence** (**BI**) refers to skills, technologies, applications and practices used to help a business acquire a better understanding of its commercial context…." This definition, when applied to the DSN, could be translated to: "BI is the skills, technologies, applications, and practices used to help DSN management acquire a better understanding of its service context". The challenges in the commercial sector are similar to the challenges faced by NASA's DSN. Both sectors need a solid and accurate understanding of their processes and systems from the point of view of performance and efficiency in order to correctly identify opportunities for improvement to the bottom line. In the case of the DSN, this includes cost, efficiency, service quality, and service levels. Traditionally, decision-making has been driven by strategically balancing risk, resources, requirements, and the perceived need to change. While risk, resources and requirements tend to be tangible and analyzed via well-developed systems engineering principles, the perception of a need to change has been much less understood and difficult to quantify and even justify. It necessitates factual and accurate information that can best be delivered via actionable metrics based on accurate data from all relevant sources. BI and its associated practices appear to address this gap in the decision support information set. BI lives at the intersection of the IT infrastructure, the business processes, and management.

In the referenced Wikipedia article, a reference is made to a Gartner paper [2] which makes strong statements about BI's role in IT and business processes. A few notable statements extracted from the Gartner paper are:

· "Because of lack of information, processes, and tools, through 2012, more than 35 per cent of the top 5,000 global companies will regularly fail to make insightful decisions about significant changes in their business and markets."

· "Organizations will expect IT leaders in charge of BI and performance management initiatives to help transform and significantly improve their business," said Nigel Rayner, research vice president of Gartner. "This year's predictions focus on the need for BI and performance management to deliver greater business value."·
"IT leaders in companies with a strong culture of information-based management should create a task-force to respond to the changing information and analysis needs of executives," said Bill Hostmann, research vice president and distinguished analyst at Gartner. "IT leaders in businesses without such a culture should document the costs and challenges of adjusting to new conditions and propose a business case for investing in the information infrastructure, process and tools to support decision making."

"By making purchases independently of the IT organization, business units risk creating silos of applications and information, which will limit cross-function analysis, add complexity, and delay to corporate planning and execution of changes," said Mr. Rayner. "IT organizations can overcome this by encouraging business units to use existing assets and create standards for purchasing classes of packaged analytic applications that minimize the impact of isolated functions."

The above statements and many similar statements and trends found in private industry hint to where information technology and modern management practices will take us in the coming years. The DSN is now attempting to apply BI solutions in response to these challenges.

Figure 2, borrowed with permission from a presentation packet by Lunexa LLC, depicts the various layers of software needed to implement a typical BI solution. It is not the intent of this paper to go into an explanation of each of these layers, as this information is readily available via a simple web search. The graphic is used to illustrate the type of investment that is needed in order to deploy a BI solution based on COTS software. The DSN is currently developing the SQA subsystem that has all of the elements in this technology stack.

Commercially Available BI Solutions

The field of Business Intelligence is under consolidation. Large players in the database and business applications arena have acquired many of the most popular and

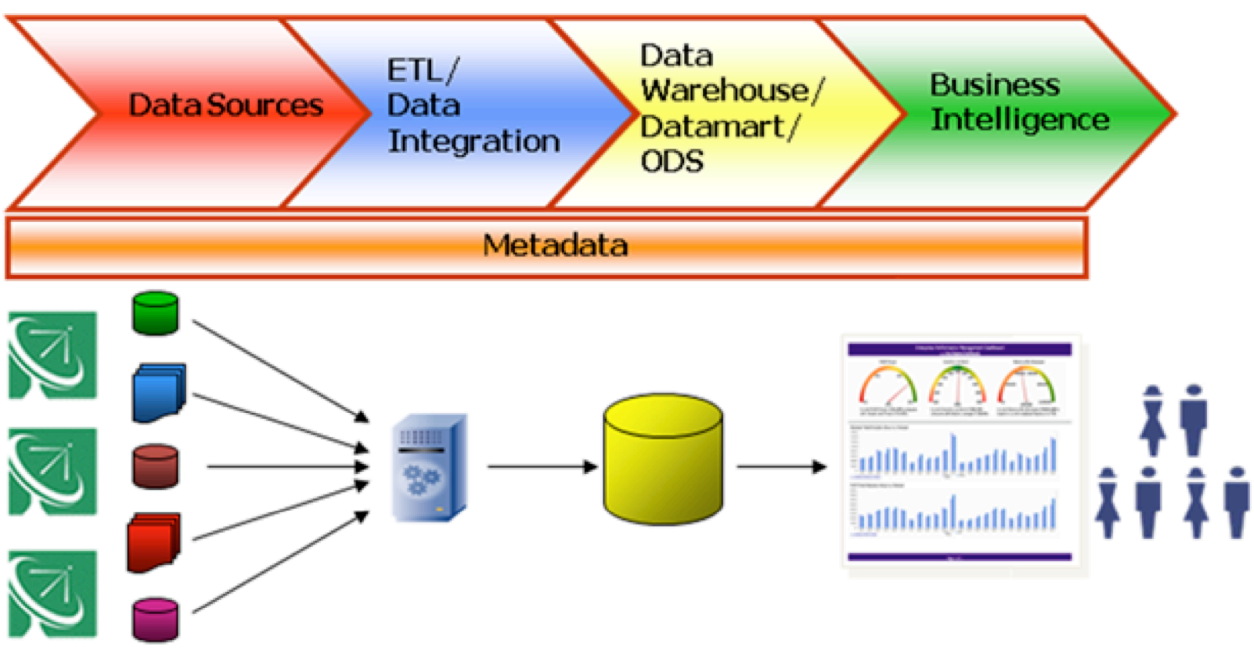


Figure 2. Typical BI Solution

successful solutions. IBM, SAP, and Oracle have acquired various products over the last five years. All appear to be leveraging for position and integrating the solutions into existing product lines. After consolidation, five companies now dominate the market:

Table 1 - Leading BI Vendors

Company	Product
IBM	Cognos
Microsoft	Analysis Services, Reporting Services, Excel PivotTables
MicroStrategy	Intelligence Server, Narrowcast, MicroStrategy Web
Oracle	Hyperion, Business Intelligence Enterprise Edition, Discoverer
SAP	Business Objects, Crystal Reports

3. PILOT DESCRIPTION

A pilot project was initiated to determine the feasibility of applying a commercial grade Business Intelligence platform in an operationally complex and highly technical environment that is not traditionally thought of as a business. The goals of the pilot project were to produce useful and typical reports for DSN management and engineering personnel within a six-month timeframe. The pilot project was approached from two perspectives more or less simultaneously: determining interesting questions to answer and understanding the data available to be analyzed.

Questions deemed appealing included ascertaining aggregate service of volume and quality across the DSN for all missions over a given sample space, understanding which missions/customers were "heavy" users of the DSN and correlatively assessing which missions were experiencing service delivery below commitment expectations. Also of interest were comparisons among the various DSN Communications Complexes, and various DSN assets to assist in more readily identifying trouble spots.

Understanding the data available for analysis proved to be a significant aspect of the engineering effort for the pilot task. In particular, the development of precise semantics with regard to large amounts of data that were generally understood by DSN engineers but intended only for operational displays as opposed to analysis consumed a fair

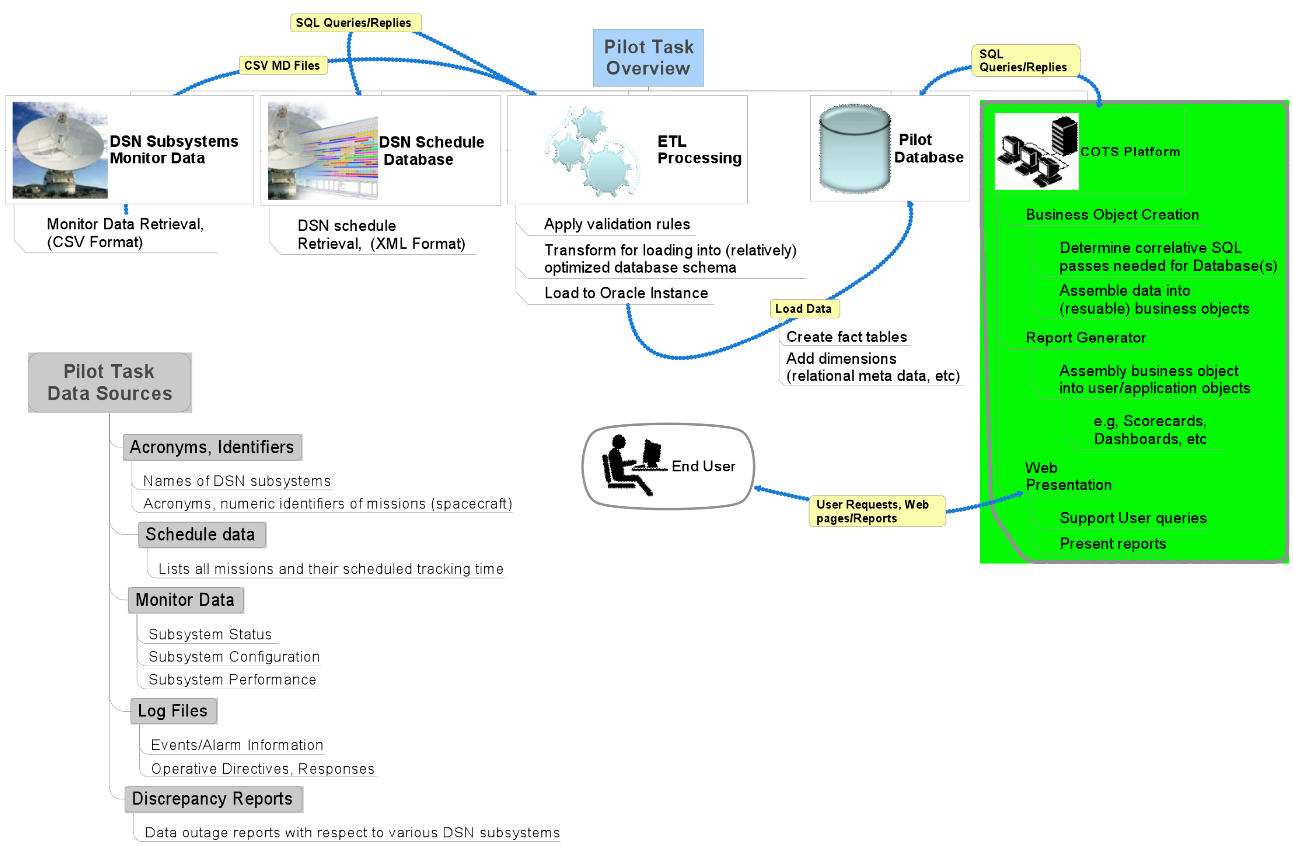


Figure 3. Overview of the Pilot Task

amount of time and required extensive "data interview" sessions. In essence the pilot task was faced with the need to retrofit an information model rapidly on top of data that had never been engineered for analysis.

An overview of the pilot task is shown in Figure 3. Focusing on questions related to assessing service volume and quality, Extract, Transform, and Load (ETL) processing was developed to extract schedule information and monitor data, apply validation and cleansing rules, and provide that information clearly associated in a relational database. At this point the data is available for more intensive semantic interrogations to properly produce reports with regard to the questions of interest. For example, in answering questions

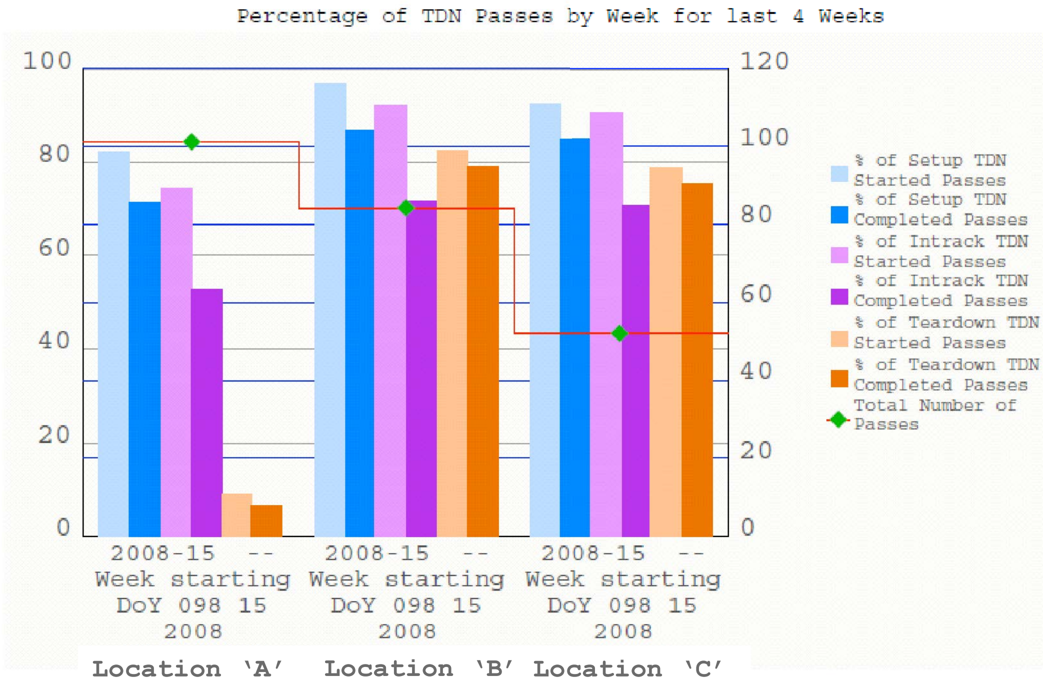


Figure 4. Initial Automation Metrics

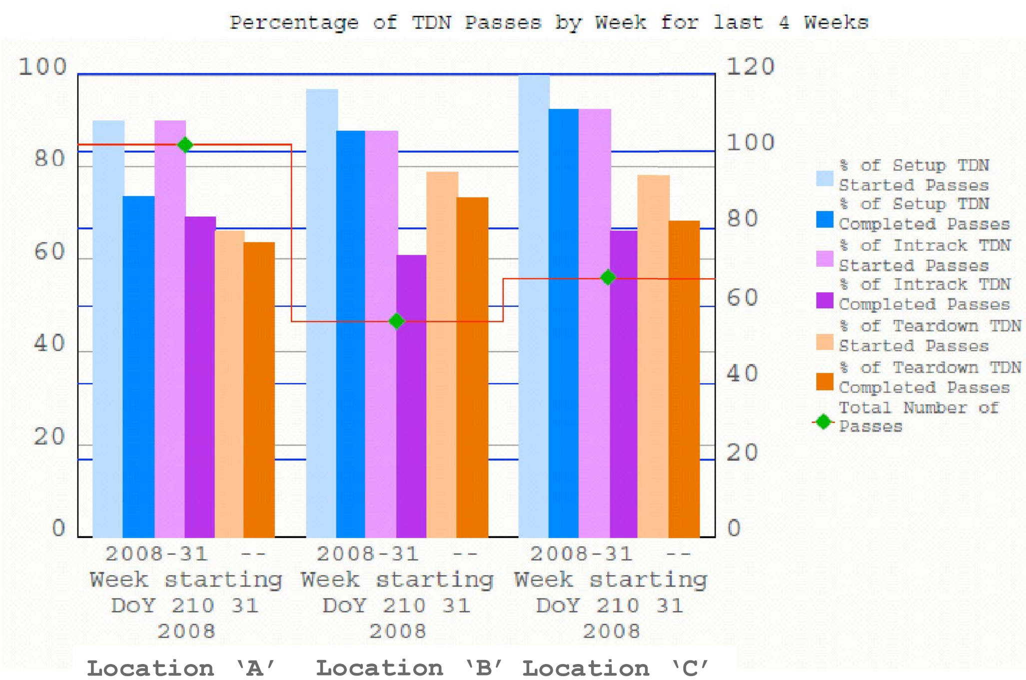


Figure 5. Follow-up Automation Metrics

with regard to aggregate telemetry service volume and quality some of the considerations include:

a) De-accumulating running total counts to allow proper aggregation over arbitrary reporting intervals
b) Tagging performance counts with signal lock status flags to provide proper quality accounting
c) Associating the schedule and performance data at multiple levels for asset or customer oriented reporting

An interesting finding from the pilot project that was not known previously is that the DSN captures and moves on the order of 150 million frames of telemetry per week.

4. Illustration of SQA Usage

As of the writing of this paper, the Service Quality Assessment subsystem has moved from its "pilot" phase to its "build" phase. While still in its build phase, SQA's initial capabilities can be used to help illustrate application of service quality assessment with regard to DSN performance. As an example we can look at internal DSN processes and overall DSN service delivery with regard to various solar system "destinations".

Analyzing Internal DSN Processes

With regard to internal DSN processes, measurement of automation software utilization proves to be feasible from the DSN internal subsystem logs files. The automation software is typically utilized in three distinct phases: *Setup* which involves configuring and calibrating the necessary pieces of equipment to provide the scheduled services, *InTrack,* which is essentially the production of the DSN services, and *Teardown* which releases and normalizes the internal equipment in preparation for subsequent use. The automation software is structured around the concept of a "Temporal Dependency Network" or TDN, which models activities that can be completed in parallel vs. sequential dependencies. The set of phases are referred to collectively as a "Tracking Pass" or "Pass". From the DSN subsystem log files, it is feasible to extract the initiation and abnormal versus normal termination of the execution of different TDNs in regard to the distinct phases of a tracking pass. This information can then be used to build up the aggregate picture of how frequently the automation software is being initiated and completing successfully at each of the three Deep Space Communication Complexes (DSCCs) of the DSN.

After collecting several weeks of data in the beginning of 2008, a distinct pattern of disparate usage emerged. In particular, it was evident, as shown in Figure 4, that the InTrack and Teardown usage of TDNs was significantly less

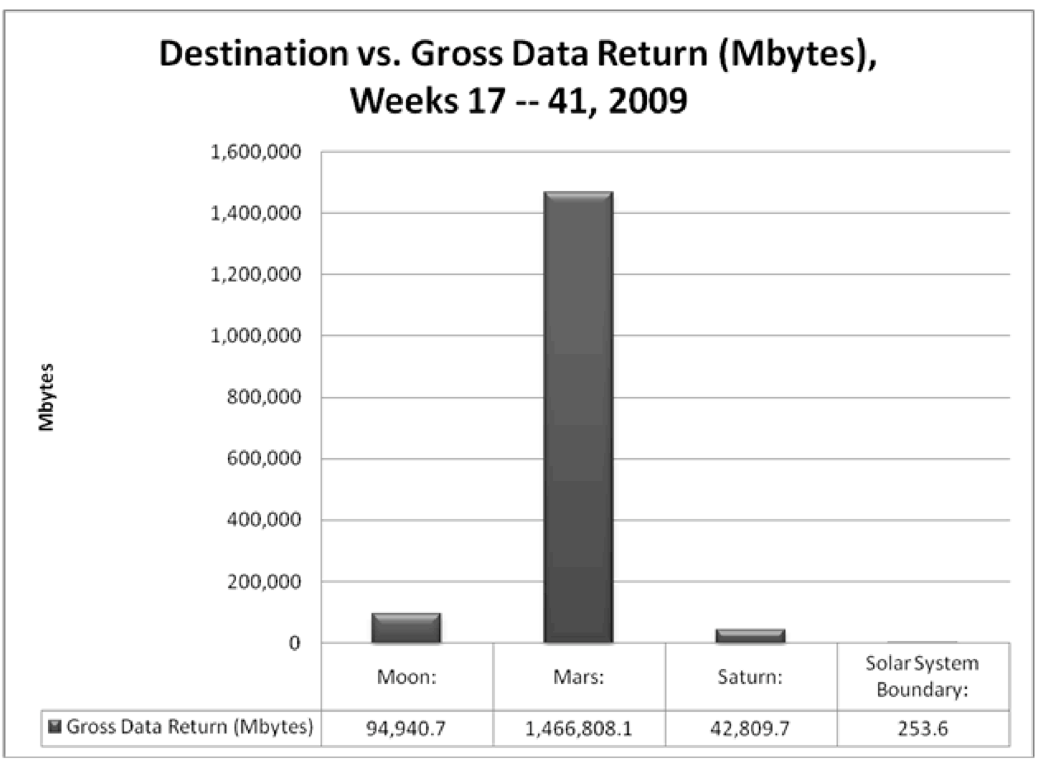


Figure 6. Gross Data Return

at one of the DSCCs versus the other two[3]. DSN management subsequently shared this information with the operations management of each DSCC. Several weeks later the usage profiles were much more closely matched across each of the DSCCs as shown in Figure 5, and continue to remain so. In this case, service quality assessment provided insight for DSN management that to increase the level of automation in providing DSN services was essentially an issue of better adherence to standard operating procedures as opposed to development and installation of additional automation software or hardware. This tends to suggest a return on investment in that a) DSN management could quickly conclude whether or not additional monies need to be expended in terms of new developments to achieve uniform utilization of automation capabilities extant (no), and b) the data to support such decisions could be automatically and continuously gathered obviating the need for analysis via conjecture.

Analyzing DSN Services for Various Solar System "Destinations"

SQA affords something previously considered infeasible: the ability to perform analysis on large amounts of intelligently aggregated data. An interesting aspect of this is the ability to aggregate data about DSN service delivery in arbitrary categorizations. For the sake of illustration assume that we are interested in determining DSN performance relative to various solar system "destinations". As of the writing of this paper some of the key solar system destinations include missions to the Moon, Mars, Saturn, and the edges or boundary of the solar system. A quick question in this regard may be something like: how much data is being sent to Earth from these different solar system destinations? If we pose this question to SQA it can readily sorted and aggregate data in relation to those missions in operation around the Moon, Mars, Saturn, etc. Sampling 24 weeks of DSN performance data in 2009 it can be seen that, for these 24 weeks, Mars is a "hot" destination as far as utilization of DSN Services goes (see figure 6), with well over 1,000 Gigabytes of data returned[4], being two orders of magnitude higher than the data returned from the Moon or Saturn.

Sticking with the aggregation of data into the different solar system destinations, we can also take a look at the degree of difficulty for the DSN in providing these services. The DSN maintains a discrepancy reporting system, which can be utilized as another source of data that can be correlated and aggregated against DSN service delivery performance. Asking SQA to perform the same type of aggregation on the discrepancy reporting system results in figure 7.

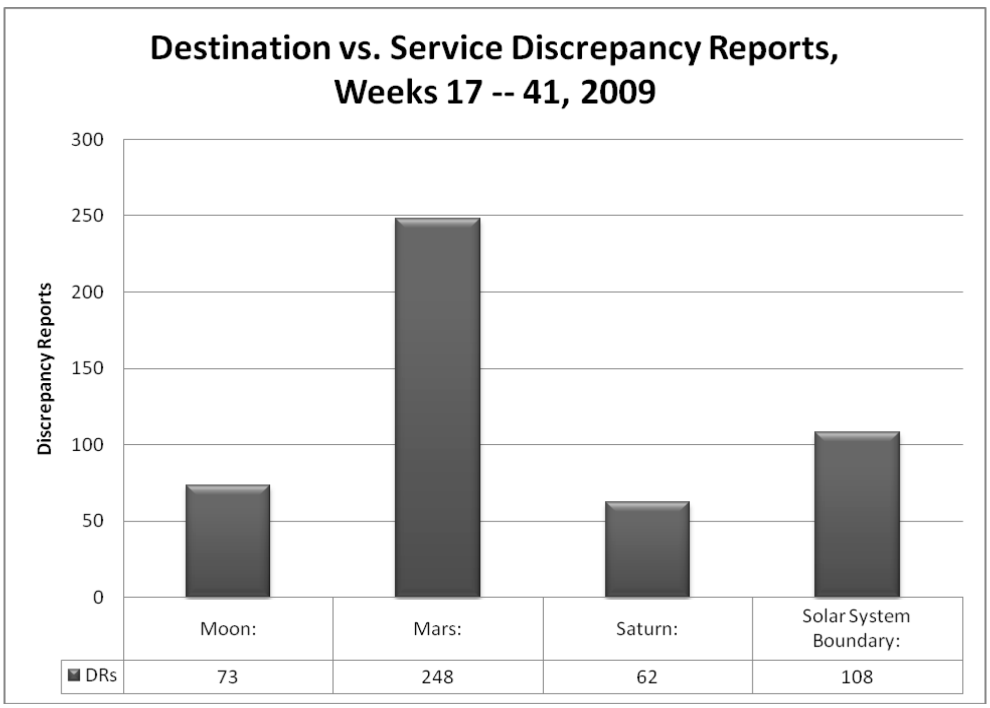


Figure 7. Service Discrepancy Reports

Looking at the DSN service volumes as indicated by data returned and the discrepancy reports, it suggests that there may be an overall DSN performance curve. Figure 8 illustrates what happens if we ask SQA to plot these two types of aggregation against one another.

[3] The specific locations have been re-labeled as 'A', 'B', and 'C', but the data are real.

[4] The volume of data is based on telemetry frame size, which includes framing "overhead" data and is therefore a measure of gross data return.

Although this tends to suggest that there is indeed a service volume versus discrepancy relationship, it should be noted that this is to be taken for illustration purposes only; the data represented here are taken from an in-development subsystem and it does not represent a vetted in-depth analysis. Rather, the intent is to illustrate how individual monitor data items / measurements representing very basic facts (e.g., received n telemetry frames from spacecraft A today; n discrepancy reports were filled this week), can be aggregated into arbitrary views of DSN performance and arbitrarily cross-correlated to help answer compound questions (e.g, how many frames did we move this week and how much did the DSN "struggle", or was it "easy", is this getting better or worse, and what is our overall performance curve in this regard? Where do we start to "sweat"?).

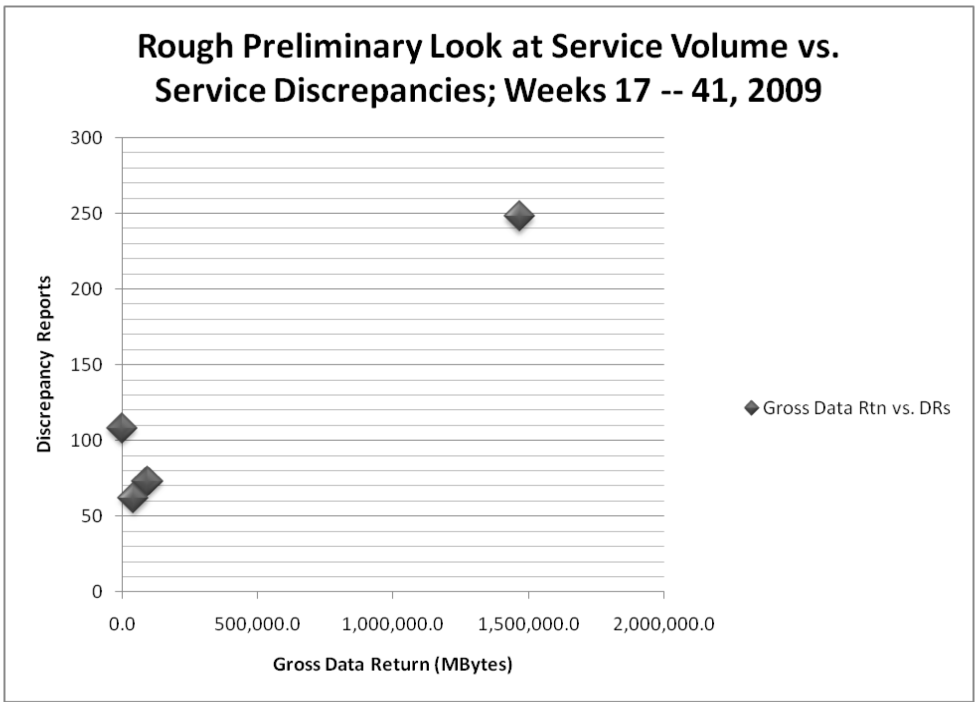


Figure 8. Service Volume vs. Service Discrepancies

5. LESSONS LEARNED

This was a new endeavor on many fronts and afforded a good amount of learning. The lessons learned covered a range of areas from the new technology approaches of complex event processing to understanding and applying the best development approaches.

REQUIREMENTS

We found that in the area of BI, most users generally know what they want but can't quite narrow it down until they see it implemented. Gathering user-level requirements in this field takes a different path than most folks with typical hardware or software systems are used to.

TECHNOLOGY

Extract, Transform and Load (ETL)

Much of the raw legacy data used to set up the final relational models used by the BI system had to be extracted from a variety of ASCII and binary formats. Further, much of the data is on old systems that needed to be extracted in an automated fashion. This led to many challenges and lessons learned in this phase of the solution.

- The extraction of data needs to be delivered to a central place. The large volumes of data involved require network bandwidth. The legacy data formats are typically not designed or intended for ease of parsing and loading into a structured database.
- Transformation is a challenge due to the many rules for filtering and interpretation of the data. The idiosyncrasies of subsystems need to be known, as data quality in a warehouse is critical.
- The need to increase the robustness of the software to handle different types and likelihood of

exceptions at the E, T, and L phases of data processing became apparent. Robust solutions are required that can be configured to: stop all processing, ignore and continue, stop after a set number of exceptions, etc.

- The pull method is most appropriate for loading data. This method is characterized by reading data as opposed to allowing external applications to directly deliver the data to the consuming system. Staging areas in file systems and/or temporary database tables is necessary when working with many disparate data inputs.

- A method is needed to roll back a new load. One must be careful not to invalidate an entire data set due to a bad ETL run. Beware of non-transactional operations such as a database table truncate.

- It is important to devise a way to reprocess a set of inputs without creating multiple entries of the same data.

- Finally, it is best to identify requirements for new metrics or its supporting data so that subsystems can prepare and deliver the data in a convenient format and delivery method for data warehousing.

Verification and Validation

V&V requires domain experts. Verification of data does not necessarily validate the results. Many times the quality of the data is not there even if data appears "normal".

Our experience has shown that weeks are need for testing under operational conditions. Data issues might only be detected after months of running reports. Some subsystems exhibit certain types of behavior only under special conditions.

Complex Event Processing

BI has traditionally been about looking at historical data looking for insights, trends, patterns, and forecasts. However, we found that users quickly begin to ask for more timely information that begins to overlap the domain of traditional system monitoring. The BI stack has not been designed with real-time analysis in mind. This is where we see the potential for incorporating a relatively new type of solution introduced in the financial trading sector. This technology is coined with the term Complex Event Processing (CEP) and was enabled by the advent of in-memory databases, availability of larger amounts of physical and virtual memory, increased processor speeds, multi-processor/multi-core architectures, messaging systems, and high speed networks. The DSN is looking to CEP as a potential solution for alarm isolation, early detection of potential issues as they develop, and a solution for adaptable and user-configurable extractors of real-time key performance indicators that traditionally are part of a Monitor and Control system. Having an adaptable CEP enables the DSN to extract the benefits of a discrete-event processing system with field-replaceable event handlers, correlators, and event pumps that feed the system with streams of events driven by schedules, service requests, geometry (e.g. occultations) and other expected as well as unexpected events such as weather, spacecraft emergencies, and equipment failures. If CEP is combined with the traditional domain of BI, a complete operations insight can be derived from historical to real-time domain.

DEVELOPMENT STRATEGY

Data warehousing requires planning but can't be driven by rigid requirements. It is important to plan for scalability, performance, and availability. However, the nature of BI tends to require agile development cycles as new opportunities are uncovered by end-users through the exploration of available data.

Metrics work best when identified by a top-down approach with actionable metrics extracting the most value out of a BI solution.

The design of Reports and the high level summary of key performance indicators, often called *Dashboards*, are best driven by specific use-cases. Often users are not able to visualize what can be done and thus unable to vocalize detailed requirements. However, if a use-case is known, BI development experts can suggest a solution to quickly iterate to a most useful design.

BI VENDORS

Only a handful of BI vendors exist after years of consolidation. Top vendors more or less can do similar things with their solutions. Some vendors have the full stack (See Figure 2). Others specialize on a specific layer of the stack. Selecting best of breed for each layer is a luxury. In the end, what works best depends on existing infrastructure, development capability, availability of resources, and requirements. There is no standard stack in BI. Even de-facto standards don't exist like the LAMP (Linux, Apache, MySQL, and PHP) stack in Open Source.

DESIGN

Data warehouse design is in critical path of a successful BI project. A data model that works well today might not work well at all tomorrow due to data volumes and usage patterns. So flexibility is important. Further, data normalization and multi-level snowflakes are not practical when cardinality of dimensions is large. As a related issue, dimensions should not be buried in large fact tables. They should be extracted or pre-populated into lookup tables with relatively small cardinalities.

5. CONCLUSIONS

This paper has described one wedge of fulfilling the service orientation goal of the DSN. Our experience has shown that the automated collection of raw metrics coupled with good data design and correlation is the basic requirement for decision support in an automated service interface.

Within the context of service provisioning the concept of quality refers to the level of compliance with which a service was delivered relative to service level agreements; timeliness refers to how quickly the measure of quality for a given service instance can be made available, and reliability refers high fidelity of the values reported, which may be derived, relative to actual, real-world values. (This is especially important in performing predictive analysis for long-term trends). The quality, timeliness, and reliability of deriving and publishing key DSN actionable metrics (that is, metrics that can be used to justify changes to work processes, for example) will become critical in order to better support the DSN in evolving a network architecture that better meets the current and future needs of its flight mission customers and NASA sponsors. However, periodic snapshots of performance provide only a subset of the capabilities needed to support management in investment decisions to address future needs in a proactive manner. Workload projections, capacity planning, new opportunities, and future trends are also needed. Until now, these needs tended to be addressed by special studies periodically chartered to address the information gap in the decision support information set.

The cost of BI for an existing operational system such as the DSN tends to be driven by the cost of ETL. 70% of the development costs of SQA deal with ETL. The system, subsystems, and assemblies were not designed with a requirement to make performance metrics easily available to consumers such as a data warehouse. To reduce the cost of identifying, retrieving, cleaning, and storing this information, it is important to identify the type of investments that can be made to reduce the overhead and complexity of metrics collection and data analysis during operations. These types of self-measuring requirements have been missing from functional and non-functional requirements. This has resulted in impeded progress on metric collection efforts due to the cost of retrofitting software systems after the fact or the cost of deriving the metrics from less than ideal data forms.

With the advent of information technology, the explosion of data, the increase in complexity of missions, and the continued oversubscription of available service provisioning assets, it has become more important to mine the information for trends, patterns, and correlations that can give better insights into the behavior and performance of the assets and processes comprising the overall service provisioning DSN system. In other words, this capability is no longer a luxury but rather a necessity for the DSN to successfully manage this service provisioning. And this capability can be delivered using Business Intelligence solutions adapted to the DSN domain. In this adaptation, the "Business" becomes "Service Provisioning Organization", "Products" become "Services" and "Profit" is replaced by "Cost Efficiencies and Service Levels". The DSN is well on its way to building this capability and has already benefited from it by increased operational efficiencies achieved by higher levels of automation.

ACKNOWLEDGEMENTS

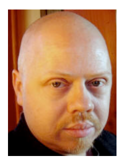
The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] "Business intelligence" *Wikipedia, the free encyclopedia.* Web. 20 Oct. 2009. <http://en.wikipedia.org/wiki/Business_intelligence>.

[2] "Gartner Reveals Five Business Intelligence Predictions for 2009 and Beyond", http://www.gartner.com/it/page.jsp?id=856714

BIOGRAPHY



Erik Barkley *is a Principal Engineer of the Software and Systems Division at NASA's Jet Propulsion Laboratory. JPL. His early professional career was focused on design and development of software intensive systems, first for commercial concerns, and from 1987 to 2001, as a contractor, for NASA's Jet Propulsion Laboratory. These efforts included several software system deliveries into the Deep Space Network (DSN). In 2001 Mr. Barkley joined JPL as an employee, and currently provides system engineering for the DSN Service Management System. Mr. Barkley is also involved with international space data systems standardization and serves as the Cross Support Services Area Director of the Consultative Committee for Space Data Systems (CCSDS). Among his professional interests are formal modeling techniques such as UML, BPMN, etc for systems specifications, Web Services protocols, and Service Oriented Architectures. Mr. Barkley holds a Bachelor of Science in Computer Science from the University of Montana..*

Paul Wolgast *is the group supervisor for a group developing DSN software systems including parts of the effort described in this paper. Mr. Wolgast has over twenty years of software development experience in systems software and information technology. Paul began his involvement with JPL developing monitor and control software for the DSN. Mr. Wolgast has also been a contributor to software architecture initiatives at JPL.*

Silvino Zendejas *is a senior member of the Systems Software Division at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California. Mr. Zendejas has worked the last 20 years designing and implementing software systems in support of the JPL Deep Space Network (DSN). His work at JPL spans various application domains, including Modeling and Simulation, Database Design and Analysis, Workflow Systems, Resource Allocation and Scheduling, Business Intelligence and Data Warehousing, and Distributed Systems. Mr. Zendejas is currently the Cognizant Design Engineer for the DSN's Service Data Management subsystem. Mr. Zendejas is also currently the task manager for the DSN's Service Quality Assessment subsystem which is responsible for the collection, analysis, and reporting of service performance and subsystem performance metrics. Mr. Zendejas holds an MS in Computer Engineering from University of Southern California and a BS in Civil Engineering from California Polytechnic State University San Luis Obispo.*