Building Your Modern Data Architecture

Josh Luedeman

Today's Speaker

- Josh Luedeman
- Sr. Consultant – Premier Support for Developers at Microsoft
- Over 10 years experience with SQL Server
- Roles have included DBA, Data Architect, BI Architect, Cloud Consultant, etc.
- Technical Editor for Hadoop and Cloud books
- Speaker at Data and Developer Focused conferences

Housekeeping

- Full Day session
- Lots of breaks
- Bathroom location

Agenda

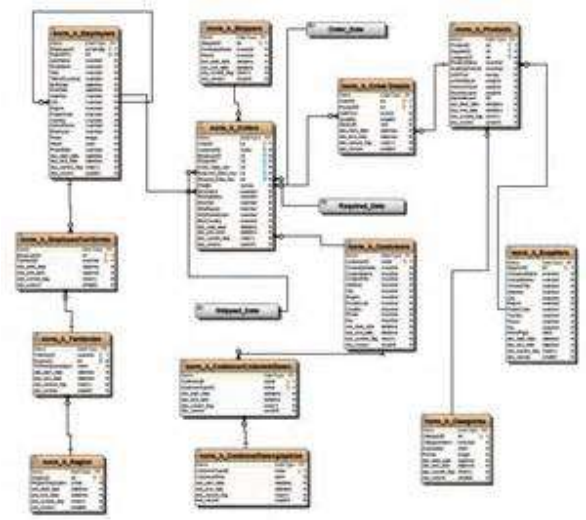
- What was/is a data warehouse?
- Star Schema
- Data Modeling 101
- ETL Architecture
- The new Future

Data Warehousing

- What?
  - Different Methodologies
    - To be normalized or NOT to be normalized….THAT is the question
- Where?
  - In house?
  - In the cloud?
- Why?

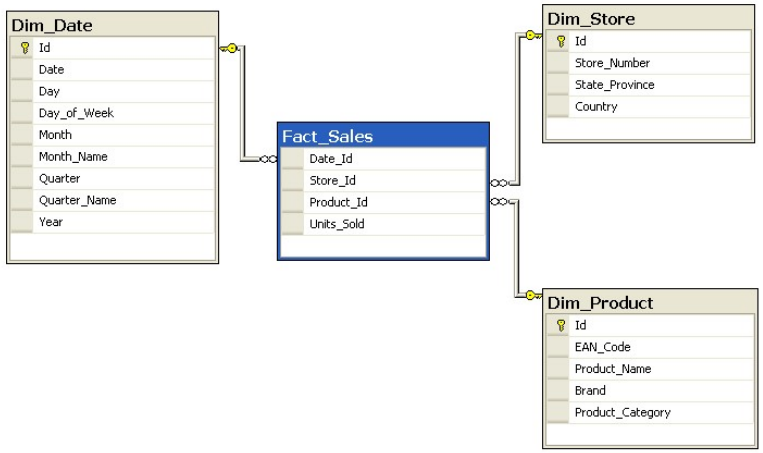
Inmon Methodology

- Normalized Database
- 3NF
- Top Down
- Waterfall Style of Data Warehousing



Kimball Methodology

- Dimensional Model (aka Star/Snowflake Schema)
- Facts and Dimensions
- Bottom-up approach
- More Agile approach to implementation



Where?

- On-premises
  - In your own data center
  - Purchase of large hardware to run your DW
  - Will need to be refreshed ever few years
  - Add resources annually, if not sooner
  - Manage the infrastructure
- In the Cloud
  - In someone else's data center
  - Provision resources as/when you need them
  - VMs are moved around to new hardware by vendor, when needed
  - Scale up or out immediately
  - It's not your infrastructure

Where should you build yours?

On-premises

- Auditing Restrictions
- Data size is stagnant

Cloud

- Your data will grow
- Resource Usage will change
- You like to save money

On-premises Data Warehouse
From looking back to now looking forward

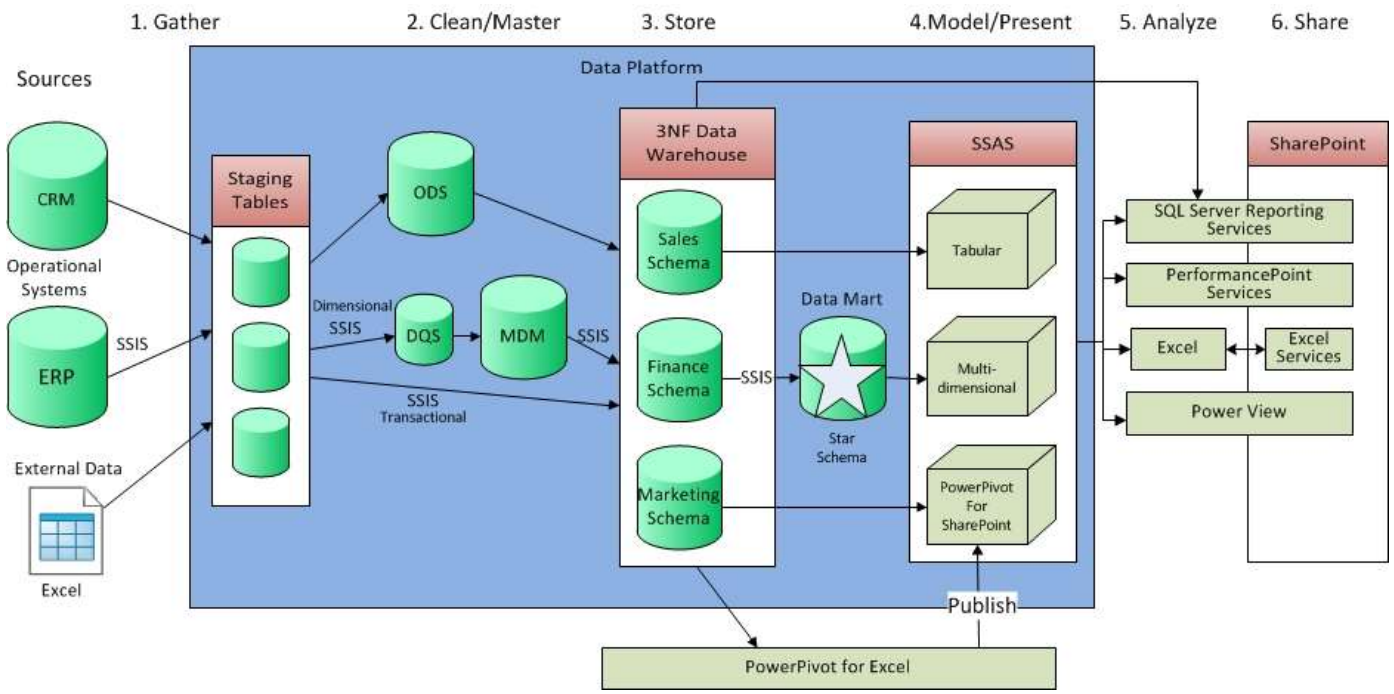
Yesterday's Data Architecture

- SQL Server Database
- SQL Server Integration Services (SSIS)
- SQL Server Analysis Services Multidimensional (SSAS)
- SQL Server Reporting Services (SSRS)
- SQL Server Analysis Services Data Modeling (about as rare as an albino deer ☺)

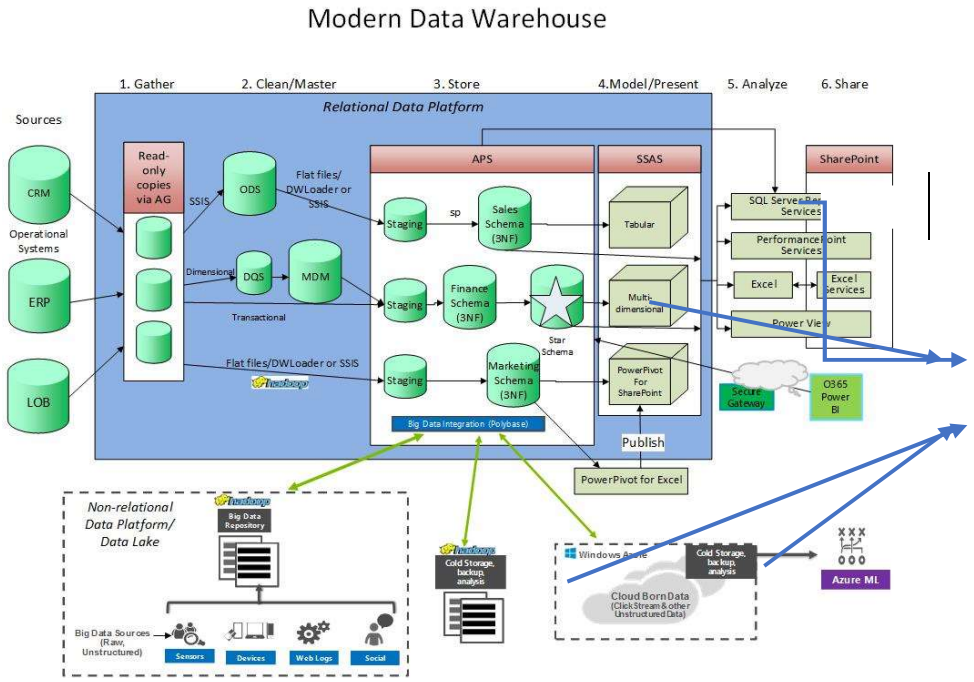
Modern Data Architecture

- SQL Server Database
  - Polybase
- SSIS or Azure Data Factory
- SSAS MDM, Tabular, PowerBI, or Azure Analysis Services
- SSRS, PowerBI, or SSRS Mobile Reporting
- Microsoft R Services, SQL Server R Services
- Hadoop, HDInsight, Azure Data Lake

Yesterday's Data Architecture

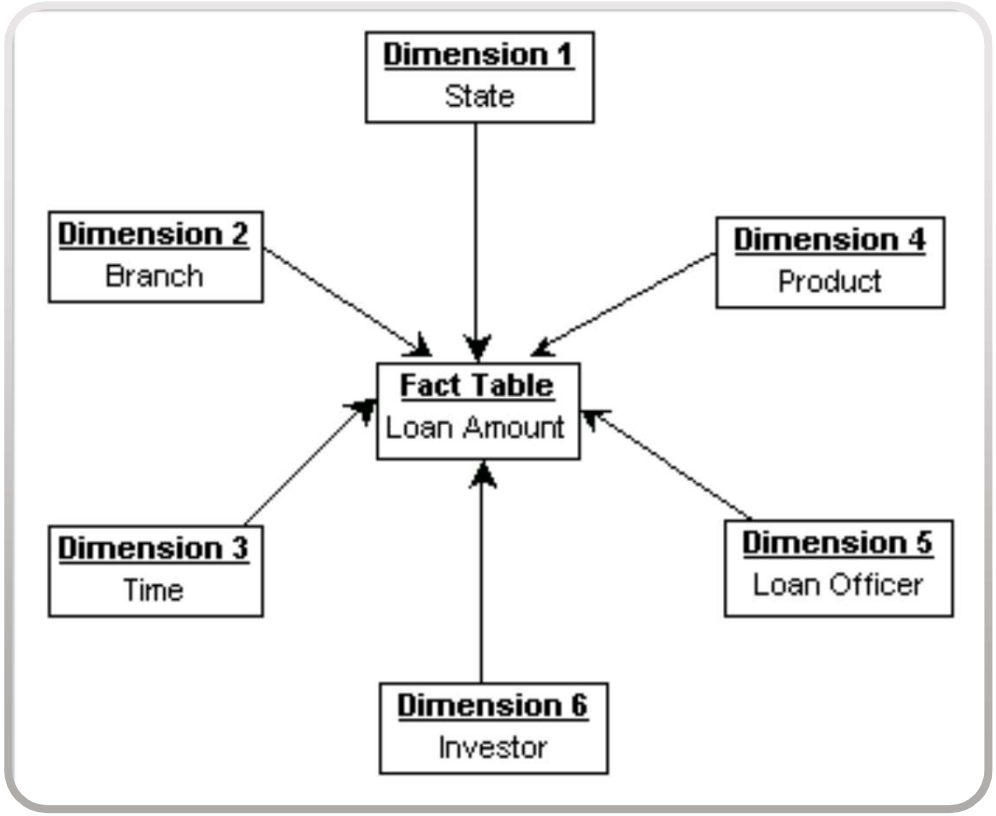


Modern Data Architecture



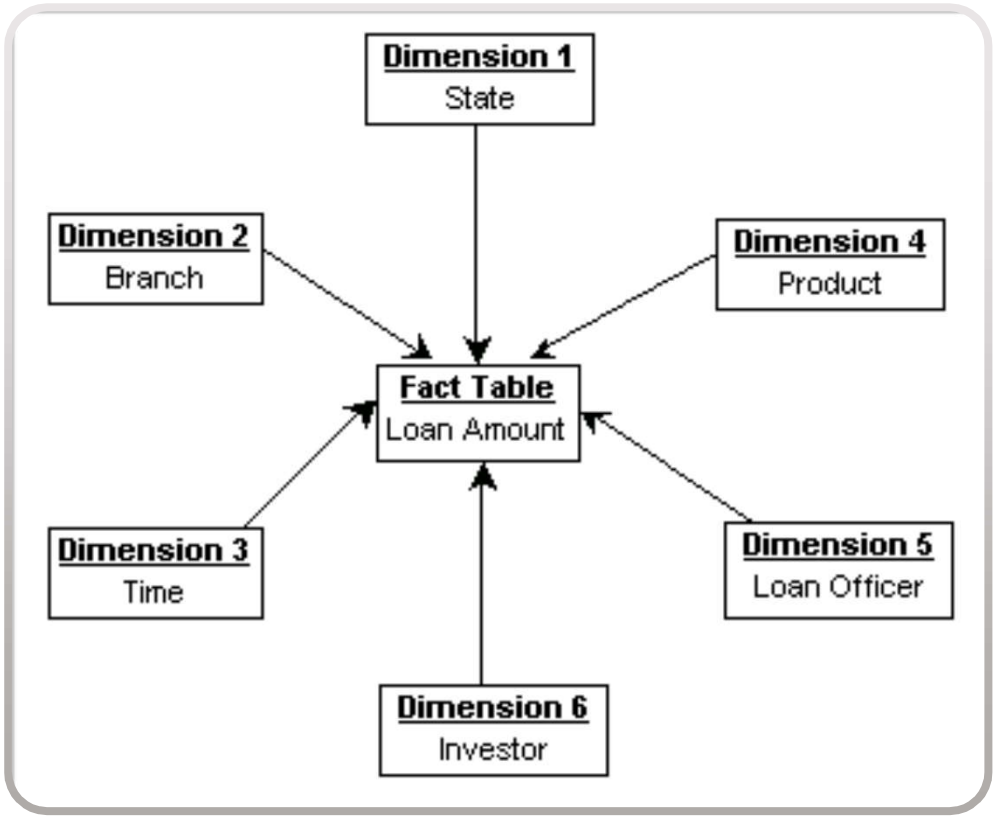
Star Schema

- Why the name? Look at it
- Iterative
- Facts
  - Actions
  - Things you count
- Dimensions
  - Nouns
  - Attributes of the action
- Designed to make it almost impossible to get the wrong value
- Dimensions and facts become the bridges between systems



Star Schema (cont.)

- One row for the lowest granularity
- Facts can be additive or snapshot, depending on the need
- Star has the ability to have dimensional versioning (i.e. Type 1 and 2)



Snowflake Schema

- Variation on the Star
- Stars on the points of a star
- Handle Many to Many relationships that can’t be designed out

Purpose

- Center is for Metrics
- Spokes for Attributes
- Properties within each Attribute

Facts

- Metrics
- Measure Types
  - Additive
  - Semi-additive
  - Non-additive
- Fact Table Types
  - Transactional
  - Periodic Snapshot
  - Accumulating Snapshot
  - Temporal Snapshots

Dimensions

- Descriptors
- One to Many relationship with Facts
- Dimension Types
  - Conformed Dimension
  - Junk Dimension
  - Degenerate Dimension
  - Role-playing Dimension

Dimension Records

- Each record has attributes
- Attribute Types
  - Each retain history differently

Slowly Changing Dimension Types

- 0 – No History
- 1 – Change Attribute, No History
- 2 – New Row, Retain History
- 3 – Add New Attribute Column for New Value
- 4 – Add History Table
- 6 - Hybrid

Data Warehouse Data Modeling

- Holistic Organizational Approach
- Start High and Work Down
  - Business Unit
  - Department
  - Metrics (high level) and granularity
  - Slicers (high level, i.e. SalesRep and not SalesRep Name and Address)

DEMO

- Data Modeling Exercise

ETL Architecture

- What are you loading
- How often do you want to load

Some ETL Structures

- Full Load
- Incremental Load
- Parent/Child Packages
- Hybrid ETL Architecture

Current Case Study – Binford Tools

- **Meet Binford Tools**
- **Current Situation:** After a recent management buyout, the once family-owned tool manufacturers, Binford Tools, have suddenly found themselves in the midst of a technical revolution. The launch of the new Binford 3000 has catapulted Binford into contention as a serious market contender. However, the market is fiercely competitive, and Binford has had to grow extremely quickly to maximize its recent success. Therefore a number of companies have been purchased, international franchises established and new alliances formed.

- However, all this expansion has not come without cost. The business has rapidly become overwhelmed with the disparate data, systems and technologies resulting from this success and is now spending a disproportionate amount of time firefighting. To make matters worse, the Binford Board of Directors are hungry for information and insights as they build their global domination strategy.

- They need insights on their acquisitions, plant efficiency, product strategy, pricing and industry trends — that is just the tip of the iceberg.

•Current Case Study– Binford Tools

- **Architecture and Configuration**
  - Still running legacy systems on Windows 2003 and SQL Server 2005
  - Other solutions running on older third party software
- **Availability and Continuity**
  - Currently using replication and log shipping since it's "what they know"
  - Only some of the applications have defined RTO and RPO goals
  - All data is in costly and non-optimized company data centers
- **Performance and Optimization**
  - Does not understand major querying patterns of applications
  - No existing benchmarks beyond "it's too slow"
  - Mix of new hardware and old hardware, no strategic upgrade path
  - Capital budget constraints stifling new platforms

•Current Case Study – Binford Tools – cont'd

- **Business Intelligence**
  - Using a mixed bag of tools and getting mixed results
  - Business is going around BI teams and IT to create shadow IT
  - No centralized data store breaks governance and insight efforts
  - DW no longer maps to business needs, provides only canned reports
  - Reports all require excel export and further analysis
- **Non-Relational and Big Data**
  - Currently exploring moving more data into enterprise governance but concerned about the already weak reputation of IT.
  - Multiple data sources provide opportunity for this
- **Advanced Analytics**
  - A business unit just hired a data scientist and he is off using new tools. IT is unaware and not involved
  - Current architects working on master's in Data Science from universities - 1 year left

•Our Vision for Binford Tools

- Update and solidify their architecture
- Prioritize and implement a new availability strategy
- Document performance expectations and needs
- Streamline business intelligence operations and standardize delivery of insights
- Find new opportunities for leveraging new data streams in a governed way
- Apply advanced analytics in an agile approach not requiring each team to have a data scientist
- Improve delivery to the business, increasing confidence
- Create a game plan for team's skills and key capability development

- Introductions
- Enterprise BI
- Our customer for the day
- Analyze Your Company’s Data Lifecycle Risk
- Big Data Architecture & Deployment
- Business & Predictive Analytics
- Creating your Personalized Action Plan
- Wrap-up, Resources etc..

ARCHITECTURE & CONFIGURATION

QUESTIONS ABOUT YOUR ARCHITECTURE	HIGH RISK	MEDIUM RISK	LOW RISK
Do you have any windows 2003 or SQL Server 2005 running in your environment?	*We are primarily still running 10 year old platforms.*	*Some of our environments.*	*Um... Why would we have those?*
Is budget usually tight for platform upgrades?	*Budget time is like squeezing blood from a stone.*	*Depends on the year.*	*My company is invested in staying competitively current.*
Do you experience issues with hardware reliability?	*Only if regular reboots or network outages are a problem - so yes!*	*Only on our older systems (you know Dev/QA)*	*We operate an effective refresh schedule for all environments.*
Do you experience delays with third party hosting centers?	*Only all the time!*	*Hmm good question - I don't think so...*	*We did once... then I "took care of it".*
Does your architecture have a cloud integration component?	*We are firmly rooted on the ground.*	*I think the business has cloud apps?*	*We have systems in the cloud or are planning on it in the next 6 months.*
Are you confident in your environment's documentation?	*Hmmm.. What is this docume-ntation you're speaking of...*	*I know it's around here somewhere.*	*Yes - I can find any information quickly.*
Is your business confident in your change control procedures?	*Sure, when we want to make a change, we do it. Confidently!*	*They are when we follow it.*	*We are using ITIL or another framework and have a good process in place.*
Do you have a security strategy and is it audited regulary?	*We put the passwords under the servers - sneaky huh?*	*We have a process until we ignore it.*	*We have a centralized security team. They don't let us do anything fun!*
Does your business have confidence that incidences get triaged, communicated and handled effectively and efficiently?	*They all call a different person in IT during an outage. That ensures sure we get it done.*	*We have a helpdesk process but it leaves something to be desired.*	*We spin up a response team and communicate effectively, then document the incident.*
Does your team contribute to business strategy or take your priorities from the business?	*What? Would you like fries with that report?*	*Sure they ask us questions, but not sure if they listen.*	*We can cleary see our influence in the organizational objectives.*
Are you and your business partners confident in your time to delivery?	*Let me get back to you on that. When? Hmmm ...*	*Yes we set general delivery timeframes, sometimes we hit them.*	*Timelines are prioritized among other projects and treated as critical.*
Do you have a confidence inspiring data governance and quality strategy?	*Yes we have data gover-what?*	*We do some data quality but Susie on that other team does that...*	*Yes data is identified, valued and reviewed for accruacy and compliance.*

Category Score
High: 25-30
Medium: 18-25
Low: 1-18

Totals

HIGH	MEDUIM	LOW

• Architecture and Configuration

- What Needs to be Accomplished
- How Much will the Project Cost
- What Else Needs to Change
- How Long will the Project Take
- How Many People will you Need
- What Don’t you Know
- Enterprise Integration

• Architecture and Configuration

- Assess Your Current Environment
- What Needs to Be Accomplished – Strategy vs. Tactics
- How Much will the Project Cost
- What Else Needs to Change - Hardware & Documentation
- Security
- Business Confidence
- Time to Delivery
- Enterprise Integration

• Assess Your Current Environment

- What are you behind on?
  - Older versions of SQL Server
  - Old hardware
  - Does your organization have Cloudphobia
- Why are you behind?
  - 3rd Party Vendor support
  - Budget constraints
- Do you have your environment documented?
- Who do you rely on as a business partner?

What Needs to be Accomplished

Strategic vs. Tactical

-High level vs the ‘Details’

- Are we Upgrading
  - What is the Cause
    - Support for a Platform
    - New Functionality
  - One System
  - The Entire Environment
- Adding a New System
- Workload Characteristics

How Much Will it Cost

- Hardware
  - Not Just Production
  - Development, QA, and DR
- Stand Up Costs
  - Additional IP’s
  - Power & Rackspace
  - Installation and Configuration
- Licensing
  - Multiple POS Locations in Stores
  - Store Servers
  - Corporate Servers

How Much Will it Cost

- What Needs to be Accomplished – Workload Analysis & ROI
- How Much will the Project Cost – Not Just Prod, QA, Dev, support hardware (Network, power supply) Cloud vs Physical? Edition Licensing
- How Long will the Project Take (Complexity of Migration*this doesn't get easier over time, training, and implementation)
- How Many People will you Need (Knowledge, Manpower, Consulting)
- What Don't you Know (Institutionalized knowledge that may have left, missing documentation)
- Enterprise Integration (How will this Integrate with your Current Environment, does it add SOX, HIPPA, Regulations, Audits, security concerns)

How Much Will it Cost

- How Do we Get to the Path of Least Resistance?
- (*Tactical)
- Hardware & Stand Up Costs
    - Utilize Cloud resources for Development/Testing
- Licensing
    - Use Standard and Express Edition where Possible
    - When Utilizing Enterprise Use Features
    - Consolidate for APS

- What Else Needs to Change

- What have we neglected
- Hardware?
- Documentation?

- Do we have proper documentation?
- Network Diagrams
- SIPS & SOPS defined
- Coding standards
- Overview for new employees

- Security

- ISO 27001/27002:2013, ISO/IEC 27018:2014, MLPS, HIPAA/HITECH, PCIDSS Level 1, FedRAMP, SOC 1, SOC 2, Australia IRAP, UK G-Cloud, Singapore MTCS, CDSA, CJIS, CSA CCCM, DIACAP, DISA Level 2, EU Model Clauses, FDA 21 CFR Part 11, FERPA, FIPS 140-2, FISC, FISMA

•Time to Delivery

- Break Projects into Realistic Phases
- Upgrades Do Not Get Easier the Longer You Wait

What Don’t You Know

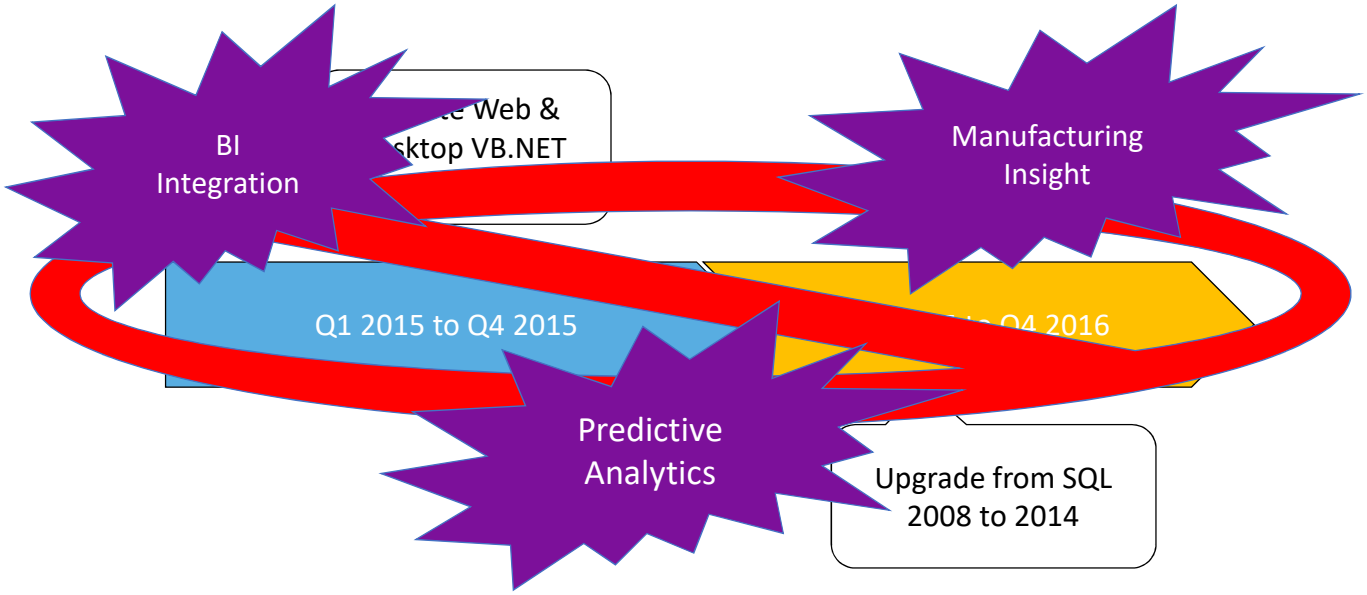
- Evaluate Systems
  - Employee Turn Over – Lost Institutionalized Knowledge
  - Incomplete Documentation
  - System Owners
  - Versioned Code Stores for Migrations
- Evaluate Development/Test/DR System Requirements
- Clean Data/Data Validation
  - Do You Have Multiple Copies in Different Systems
  - Does that Data Match

• Enterprise Integration

- New System Considerations
    - Regulatory Consideration
    - Data Governance
    - Performance Requirements
    - Reporting Considerations
    - Master Data Management

What Needs to be Accomplished

- Case 1: Binford Tools needs to upgrade applications and RDBMS, gained through acquisition and legacy systems that have not been upgraded. Integrate with new systems for Manufacturing, EDW BI, and Predictive Analytics.



•What Needs to be Accomplished

- Case 1: Binford Tools needs to upgrade applications and RDBMS, gained through acquisition and legacy systems that have not been upgraded. Integrate with new systems for Manufacturing, EDW BI, and Predictive Analytics.

How Do we Get to the Path of Least Resistance?

- Know when we should use SQL and when we should not

Data Per Machine Hourly (MB)	Machines Per Facility	# Facilities	Data Per Hour (GB)	Data Per Day (GB)
100	100	5	48.83	1171.88

What Needs to be Accomplished

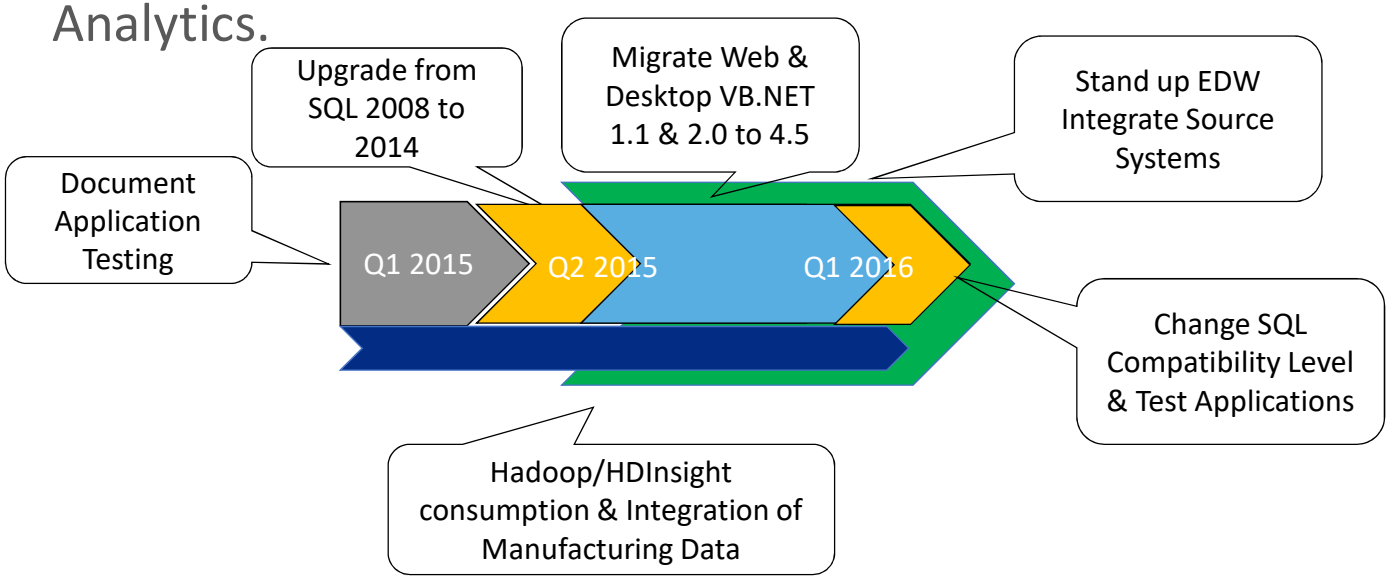
- Case 1: Binford Tools needs to upgrade applications and RDBMS, gained through acquisition and legacy systems that have not been upgraded. Integrate with new systems for Manufacturing, EDW BI, and Predictive Analytics.

How Do we Get to the Path of Least Resistance?

- SQL Server Upgrade Advisor
- Perfmon Counters Deprecated Features (SQL 2008 and UP)
- SQL Server Compatibility Mode
- MAPs

•Module 1 – Wrap Up

- Case 1: Binford Tools needs to upgrade applications and RDBMS, gained through acquisition and legacy systems that have not been upgraded. Integrate with new systems for Manufacturing, EDW BI, and Predictive Analytics.



•Wrap Up

- Assess Your Current Environment
- How Much will the Project Cost
- What Else Needs to Change - Hardware & Documentation
- Security
- Business Confidence
- Time to Delivery
- Enterprise Integration
- What Needs to Be Accomplished – Strategy vs. Tactics

- Introductions
- Enterprise BI
- Our customer for the day
- Analyze Your Company’s Data Lifecycle Risk
- Big Data Architecture & Deployment
- Business & Predictive Analytics
- Creating your Personalized Action Plan
- Wrap-up, Resources etc..

AVAILABILITY & CONTINUITY

QUESTIONS ABOUT YOUR AVAILABILITY/DR	HIGH RISK	MEDIUM RISK	LOW RISK
When was your last successful DR test?	*We have never tested our DR, the vendor said it would just work.*	*I think about a year ago.*	*4 or more times a year during Maintanance windows.*
How often do you regularly test your important business systems?	*We don't have time to test.*	*Tested a few times but not on the top of our list.*	*Test with every release.*
Do you have a plan in place for seamless availability for your most critical apps?	*Currently no plan in place.*	*Have a plan but have not executed on it yet.*	*We have a plan and it is implemented.*
Does your business have confidence in your availabiity strategy?	*Our business users want us all fired.*	*They heard our strategy but are not sold on it yet.*	*They have full confidence and we share our strategy with them.*
Do you use different technologies to manage availablity?	*We don't have any technology to manage our availability.*	*There are several technologies and it's hard manage.*	*We have a single technolgy to manage all availability.*
Do you rely on product features or vendor tools for availability?	*I wish we had at least one.*	*We use a combination of both or have several vendors.*	*We rely on product features and a single vendor tool to manage them.*
Does your business have different plans for different applications?	*Our business doesn't have any plans.*	*We have plans, but too many to count.*	*Single plan and SLA's we manage to.*
Do you leverage the cloud for availability or testing?	*I've heard of the cloud...but didn't all the celebrities get their photos hacked.*	*We want to use the cloud but not sure how or have regulatory concerns.*	*We use the cloud as part of our availabilty plan.*
Do you have an availability strategy for development and QA environments?	*We don't even have one for production yet.*	*Not sure how to make one scale for Dev and QA.*	*We use the same strategy as our Production environment.*
Do you have the same availabilty strategy for OLTP and Data Warehouse?	*What's a Data Warehouse?*	*They are all just databases so we use one strategy.*	*We have separate strategies for each one.*
How do you keep current with technology?	*If it's not broke don't replace it.*	*Always stay one version behind.*	*Constantly evaluate and upgrade.*
What training do you provide to your employees?	*What good does training do?*	*We always budget for training but its hard execute.*	*Our employees are required to attend a certain number of training courses a year.*

Category Score
High: 25-30
Medium: 18-25
Low: 1-18

Totals

HIGH	MEDUIM	LOW

• Availability and Continuity

- Disaster Recovery –It’s more than Just a Backup
- Availability Tiers
- OLA, SLA, RTO, RPO
- Uptime & the five 9’s
- Tools & Management
- Why Multiple Environments are Important
- Training & Staying Current

• Availability and Continuity

- Disaster Recovery –It’s more than Just a Backup
- OLA, SLA, RTO, RPO
- Uptime & the five 9’s
- Why Multiple Environments are Important

• Disaster Recovery –It's more than Just a Backup

- High Availability Technologies
- SQL Technologies
  - Backup and Recovery
  - Log Shipping
  - Snapshots
  - Clustering
  - AlwaysOn Availability groups
- SAN
  - SAN Replication
- Virtualization
  - Image Restoration

- Introductions
- Enterprise BI
- Our customer for the day
- Analyze Your Company’s Data Lifecycle Risk
- Big Data Architecture & Deployment
- Business & Predictive Analytics
- Creating your Personalized Action Plan
- Wrap-up, Resources etc..

PERFORMANCE & OPTIMIZATION

QUESTIONS ABOUT YOUR PERFORMANCE	HIGH RISK	MEDIUM RISK	LOW RISK
Do you know your critical workloads?	*Whoever yells the loudest.*	*I think I know what's important.*	*We have all areas categorized.*
How do you measure performance of your applications?	*When someone complains we react.*	*Production is monitored but we don't have time to look at it.*	*It's part of our development lifecycle.*
How many times a week do you get complaints about performance?	*Better question is, when don't we have complaints?*	*At least once a day.*	*Maybe a couple times.*
When you have performance issues how does your team react?	*Try everything till it works.*	*Indexes are always the answer.*	*Follows best practices for troubleshooting .*
What training have you attended?	*Google it.*	*I took a class a couple years ago.*	*Get training at least once a year.*
Have you ever used features such as Columnstore indexes or InMemory OLTP?	*The DBA's are afraid of them.*	*Have tried, but not sure when to use them.*	*Only when appropriate.*
What is your scale out strategy?	*Just put in more memory.*	*Replication or Mirroring.*	*We use Availability Groups.*
How often do you review systems for possible performance issues?	*How do you do that?*	*Gathering statistics but not sure what do do next.*	*Part of a regular review process.*
What is your business application review process?	*That's for the users to worry about.*	*The users have asked for our help but we don't have time.*	*Reviews on an regular basis.*
How is code review implemented?	*Just put it in production already!*	*Just review the basics for naming conventions.*	*Nothing gets promoted without full review from Sr. DBA.*
How do you ensure new releases will not introduce performance issues?	*It worked in Dev, that's good enough for me!*	*Our QA team does some testing.*	*It's part of our build process.*
What do you do if a release fails?	*Start looking for a new job.*	*To the backups we go!*	*We stage our release on standby servers.*

Category Score
High: 25-30
Medium: 18-25
Low: 1-18

Totals

HIGH

MEDUIM

LOW

- Agenda – Performance & Optimization

  - Monitoring Your Environment
  - Detecting Issues
  - Finding Solutions
  - Implementing Solutions

• Agenda – Performance & Optimization

- Identify Critical Workloads
- Baselines & Performance
- Solutions via New Features & Technologies
- Scale Out vs. Scale Up
- Pro-active Performance Reviews
- Code Reviews, Rollbacks, & Testing

• Identify Critical Workloads

- After evaluating your environment assign Tiers. As the following questions?

1. If a system is down what does it prevent us from doing business?
2. If a system is down who cannot do their job?
3. How much does that cost us?

• Identify Critical Workloads

- Critical Workloads Are:
- Workloads that cost the company money
- Allows us to set a dollar value to down time
- Allow IT to quantify the business value
- Could be OLTP or OLAP

• Baselines & Performance

- Baselines – Not Free
  - SQL Sentry
  - Dell/Quest
  - Microsoft – SCOM
  - SolarWinds – Database Performance Analyzer
  - MORE
- Business Intelligence Infrastructure
  - Pragmatic Works Work Bench
  - SQL Sentry

- Baselines & Performance

Baselines – Free

SQL Server

- Perfmon.exe
- SQL Server Performance Management Data Warehouse
- DMV's & Extended Event
- Community Scripts – sp_IndexAnalysis (Jason Strate)
- Query Store *SQL 2016

SQL Azure Database

- DMV's & Extended Events
- Query Store

Business Intelligence Infrastructure

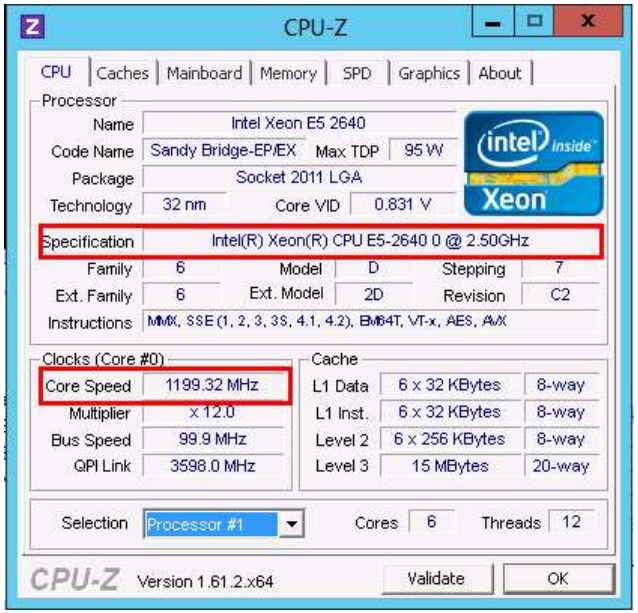
- Perfmon.exe
- Extended Events *SQL 2016

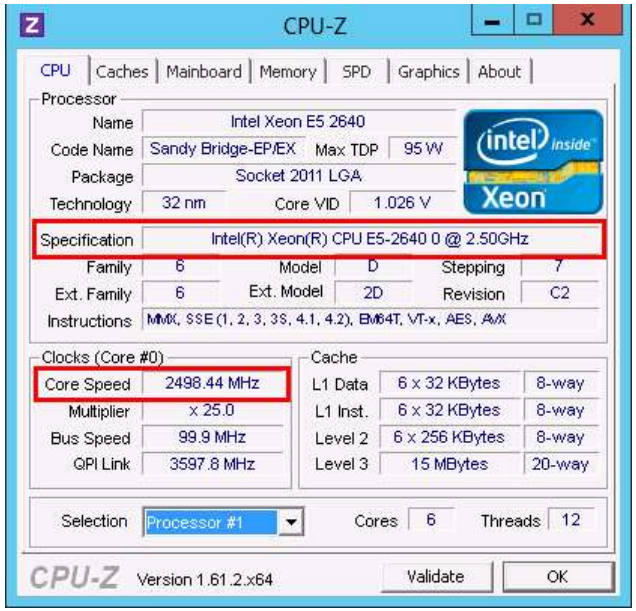
- Baselines & Performance

Setup and Configuration- Power Configuration

- Baselines & Performance

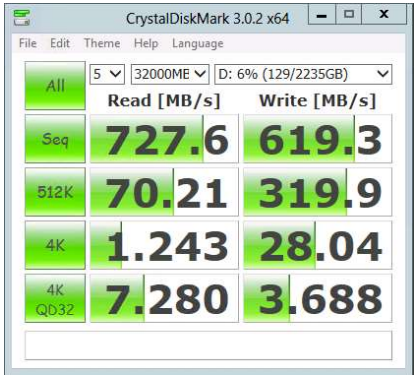
Setup and Configuration- Power Configuration



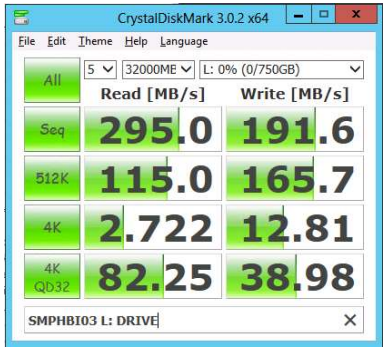


• Baselines & Performance

Setup and Configuration- Baseline Disks



(DAS raid 5): Direct Attached Storage



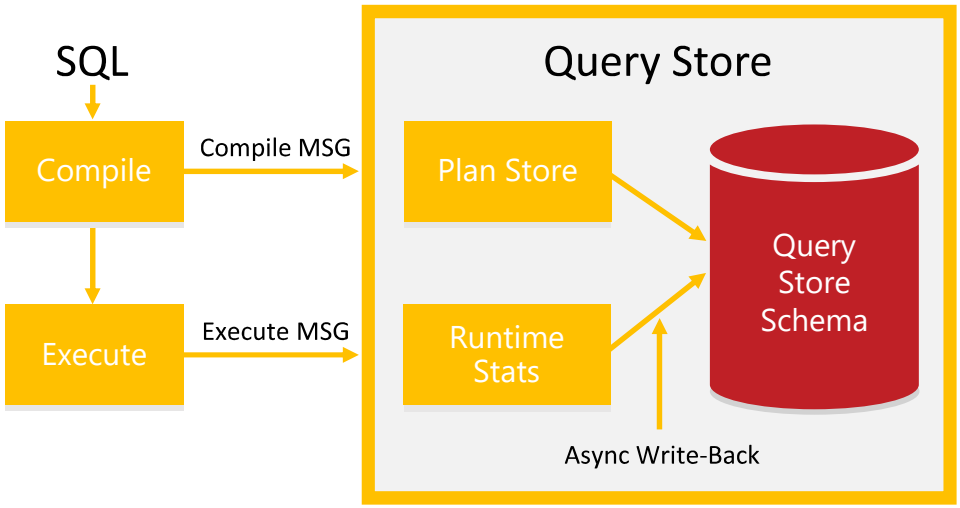
(SAN) 15K Drives RAID 5

- Solutions via New Features & Technologies
- Query Store

- I CAN get full history of query execution
- I CAN quickly pinpoint the most expensive queries
- I CAN get all queries that regressed
- I CAN easily force better plan from history with a single line of T-SQL
- I CAN safely do server restart or upgrade

Solutions via New Features & Technologies

Query Store



Durability latency controlled by DB option
DATA_FLUSH_INTERNAL_SECONDS

Collects query texts (+ all relevant properties)

Stores all plan choices and performance metrics

Works across restarts / upgrades / recompiles

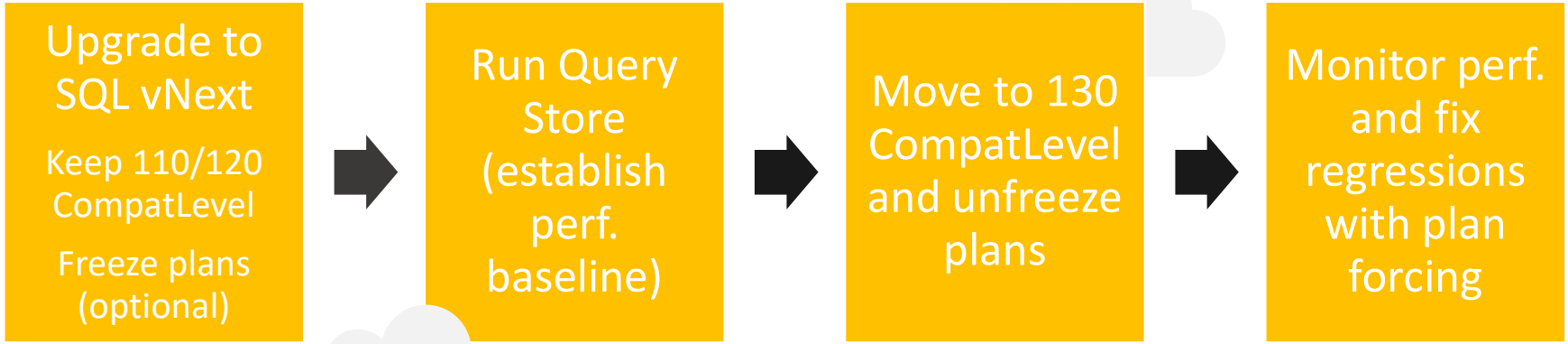
Dramatically lowers the bar for perf. Troubleshooting

New Views

Intuitive and easy plan forcing

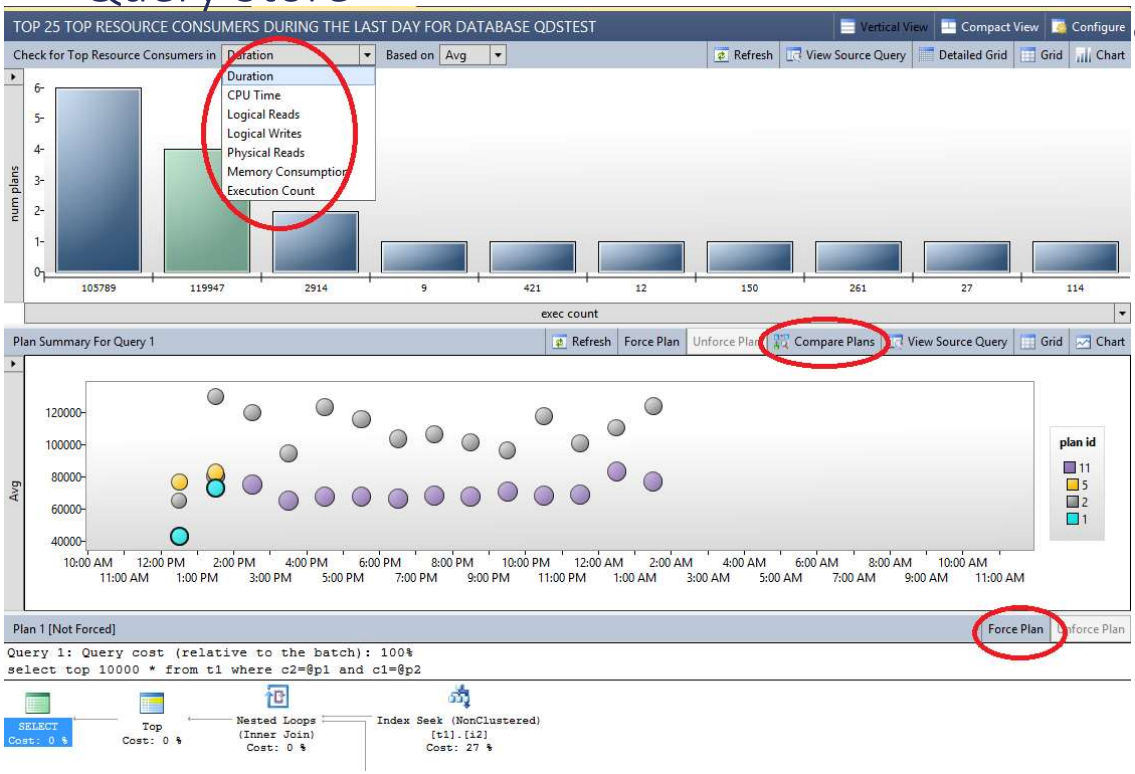
Solutions via New Features & Technologies

Query Store



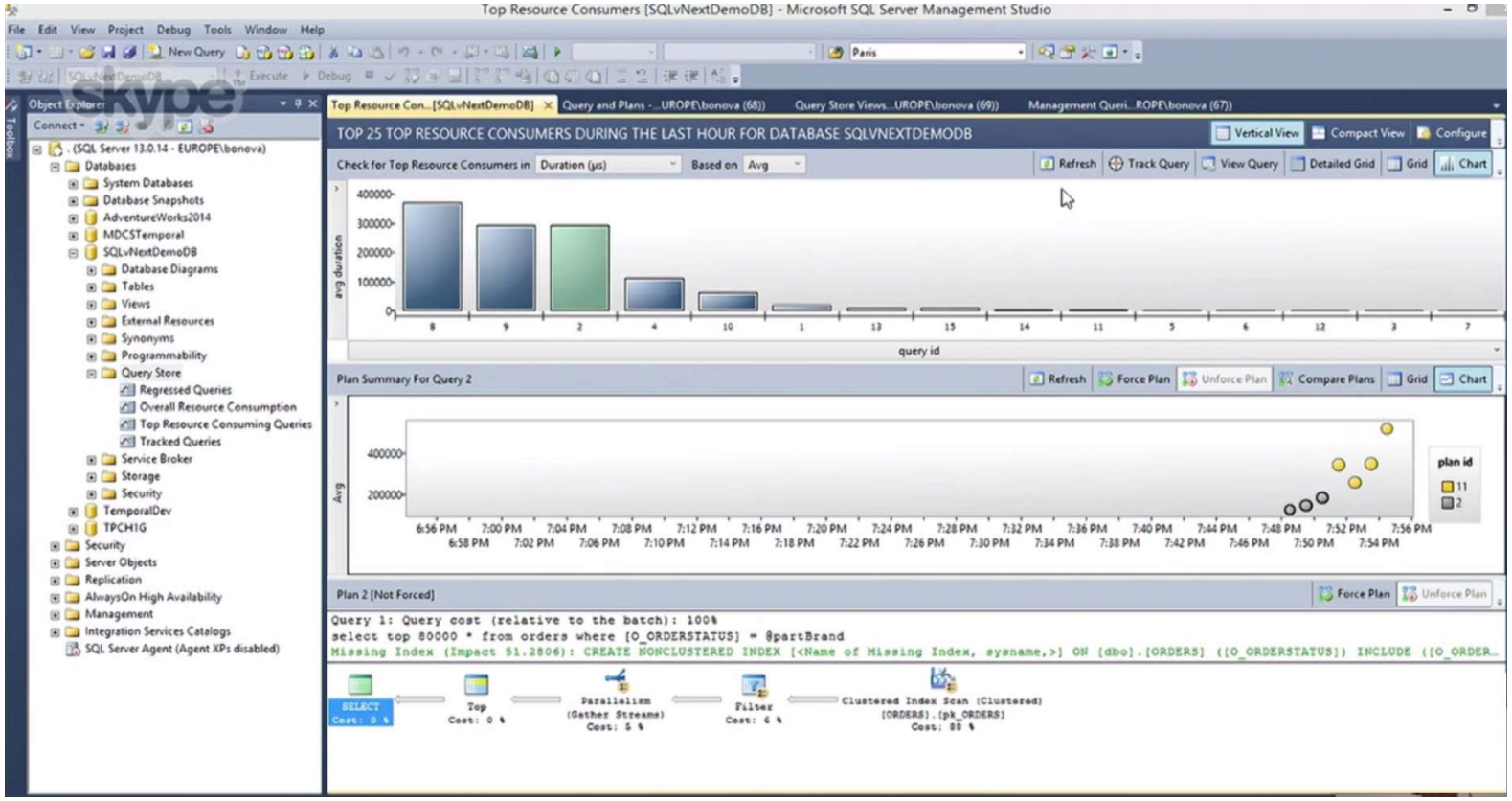
Solutions via New Features & Technologies

Query Store



- The query store feature provides DBAs with insight on query plan choice and performance

- Solutions via New Features & Technologies
- Query Store



Solutions via New Features & Technologies

Query Store

```
/* (6) Performance analysis using Query Store views*/
SELECT q.query_id, qt.query_text_id, qt.query_sql_text,
SUM(rs.count_executions) AS total_execution_count
FROM
sys.query_store_query_text qt JOIN
sys.query_store_query q ON qt.query_text_id =
q.query_text_id JOIN
sys.query_store_plan p ON q.query_id = p.query_id JOIN
sys.query_store_runtime_stats rs ON p.plan_id = rs.plan_id
GROUP BY q.query_id, qt.query_text_id, qt.query_sql_text
ORDER BY total_execution_count DESC

/* (7) Force plan for a given query */
exec sp_query_store_force_plan
12 /*@query_id*/, 14 /*@plan_id*/
```

```
);

/* (4) Clear all Query Store data */
ALTER DATABASE MyDB SET QUERY_STORE CLEAR;

/* (5) Turn OFF Query Store */
ALTER DATABASE MyDB SET QUERY_STORE = OFF;
```

DB-level feature exposed through T-SQL extensions

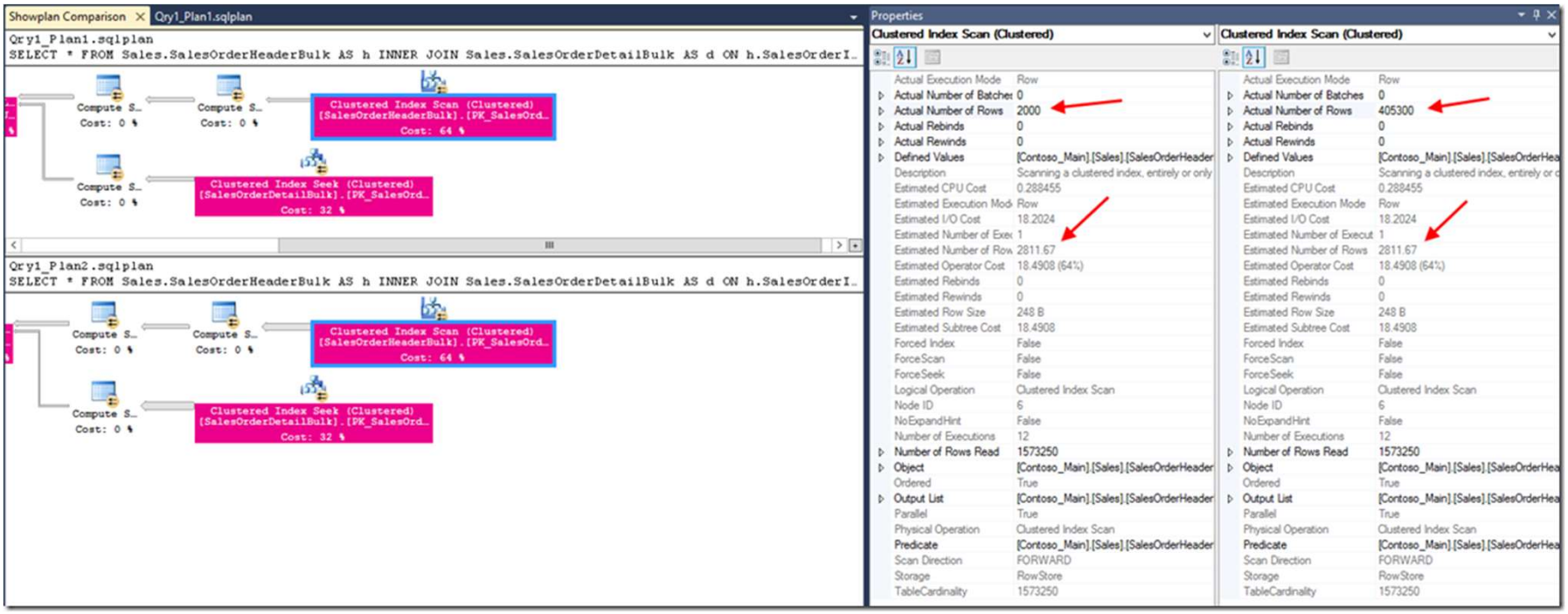
ALTER DATABASE

Catalog views (settings, compile & runtime stats)

Stored Procs (plan forcing, query/plan/stats cleanup)

Solutions via New Features & Technologies

Execution Plan Compare



http://blogs.msdn.com/b/sql_server_team/archive/2015/10/12/comparison-tool-released-with-latest-ssms.aspx#.Vi-KtLerQb0

- Polybase

- Enables interconnecting of non-SQL data sources
- Easy to setup
- Easier to use
- Some important things to know

- Scale Up vs. Scale Out

- APS
- Superdome
- Elastic Scale Cloud
- SQL DW

- Pro-active Performance Reviews

- Look for spikes in known workloads
- Look for changes in Wait Statistics
- Alerting mechanisms
- Daily reports on trends and changes

- Code Reviews, Rollbacks, & Testing

- Establish Standards
- Train to a minimum level
- Peer review all code before migration
- Content management teams
- Dedicated Testers
- Consider implementing DevOps Systems and Procedures

- Finding More Solutions

  - Microsoft Virtual Academy
  - PASS Virtual Chapters
  - SQL Saturday’s
  - Conferences (SQLBits!, PASS Summit)
  - Webinars
  - #SQLHelp

- Performance & Optimization

  - Case 3: Binford Tools is experiencing periodic outages related to performance issues. The DBA team is constantly in fire fighting mode. Proactive measures and new technology advancements are lagging.

- Performance & Optimization

  - Case 3: Binford Tools is experiencing periodic outages related to performance issues. The DBA team is constantly in fire fighting mode. Proactive measures and new technology advancements are lagging.

  - Impliment a Monitoring Strategy on Production Systems
  - Bring in a Consultant to do a Performance Tuning Review of Critical Business systems
  - Train DBA Team in Performance Tuning
  - Impliment recommendations upon review and consensuses of DBA Team
  - Impliment SQL Server 2014 Technologies to Alleviate Performance Issues

- Introductions
- Enterprise BI
- Our customer for the day
- Analyze Your Company’s Data Lifecycle Risk
- Big Data Architecture & Deployment
- Business & Predictive Analytics
- Creating your Personalized Action Plan
- Wrap-up, Resources etc..

- Your Environment

- Virtual or Hardware
- Do you have enough?
- Is the cloud available?
- If so, how can you leverage?
- Can you afford it?

- Introductions
- Enterprise BI
- Our customer for the day
- Analyze Your Company’s Data Lifecycle Risk
- Big Data Architecture & Deployment
- Business & Predictive Analytics
- Creating your Personalized Action Plan
- Wrap-up, Resources etc..

BIG DATA ARCHITECTURE & DEPLOYMENT

QUESTIONS ABOUT YOUR DATA ARCHITECTURE	HIGH RISK	MEDIUM RISK	LOW RISK
Are you aware of all the data sources that your business is using?	*Data what?*	*I'm aware, but that doesn't mean much.*	*Yes, we coordinate closely with the business.*
Do you have an agile data governance strategy?	*Why would we need that?*	*I think we do.*	*Yes, we are adapting to new business needs.*
Is your team aware or trained on these new technologies?	*I don't have what you would call a "team".*	*My team has read up on it.*	*Some of the team has been trained or has previous experience.*
Do you know about the business' needs for these technologies?	*They don't let me come to meetings...*	*I'm still trying to get other things discussed.*	*Yes we align closely with the business on data strategy.*
How mature is your data archieving and tiering strategy?	*We have a lot of thumb drives?*	*We archive backups but not much else.*	*We tier data based on performance and have an arhival policy.*
Are you performing workload identification to look for scale out opportunities?	*We know work is happening.*	*We don't review workloads specifically.*	*Applications are segmented by workload.*
Are you looking to do more in-line computing within your ETL and data pipelines?	*No we don't stand in line for anyone.*	*We would like to understand that a bit more...*	*I think we could take advantage of this.*
Are you prepared to integrate these new data sources to your master data platform?	*We don't have "master"data.*	*We will figure it out.*	*We are masters of all our data.*
Are you concerned about cleansing issues with all these new data sources?	*No, we just dump everything in one table.*	*The business will want to do that themselves.*	*We are charged with ensuring all data is clean and valuable.*
Do you organizationally understand what these types of technologies can do?	*I read the internet.*	*My CIO has told me what we're doing.*	*We are piloting several use cases.*
What's your level of architectural confidence in these technologies?	*I said... I read the internet...*	*My architect is trying to learn these right now...*	*We are comfortable with some, not with others.*
What training do you provide to your employees?	*What good does training do?*	*We always budget for training but its hard execute.*	*Our employees are required to attend a certain number of training courses a year.*

Category Score
High: 25-30
Medium: 18-25
Low: 1-18

Totals

HIGH

MEDUIM

LOW

- The “What’s?” for Big Data Success

- What kind of data do you have?
- What does you or your team know about these technologies?
- What do you know about Big Data?
- What do you think you are going to do with it?
- What policies and governance do you already have in place?

- What kind of data do you have?

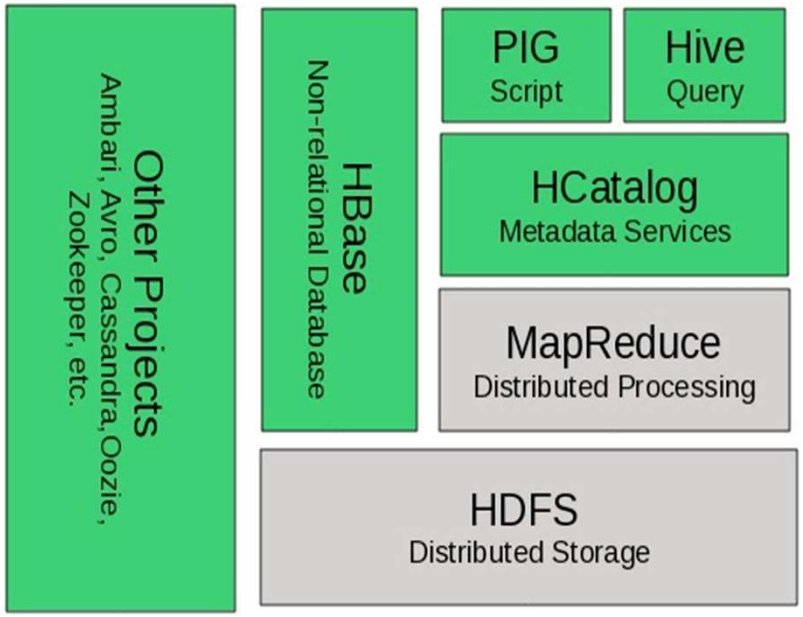
- Where is your data coming from?
- How is it stored?
- How fast is it coming at you?
- What do you think you can gain from it?

- What do you and your team know about these technologies?

- Have you used these technologies before?
- Do you know how to apply them? Correct use cases for each tool?
- Do you have the training budget for skills gaps?
- Do you have the confidence in the skills base to build the infrastructure?

Components of a Big Data Ecosystem

- Operating System Platforms
  - On Premise or Cloud?
  - Linux or Windows?
  - Amazon or Azure?
- Hadoop
- Apache Projects
  - Storm
  - Spark
- Analytic Products
- Visualization and Analysis

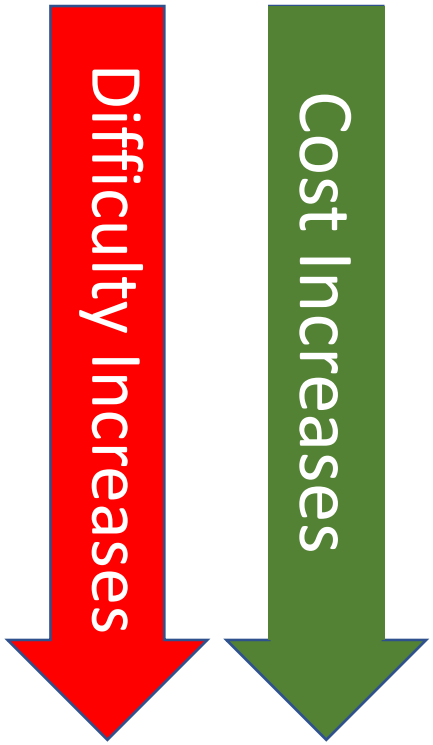


Why so many questions?

- Silver bullet mentality
  - “I have big data now. I just want to find out what it can tell me.”
- Big Data is the new cloud
  - Every boss has to have it, but ask them what it is
- Understand your business
  - You can’t be successful if you don’t understand the problems

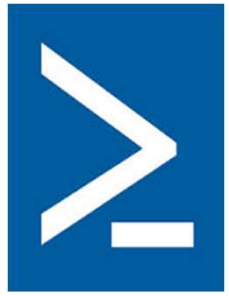
Discussion - So Which One is right for You?

Deployment Model	Use Case	Time to Deploy	Functionality	Cost
Local Deployment	Personal Development	Hours	Full Feature, limited capacity	Minimal or free
Cloud Clusters	Agile and Scalable Deployment	Hours	Basically unlimited capacity and performance	Scalable – builds over time
On Premise Cluster	For Specific Requirements	Days - Weeks	Capacity and performance maxed	Fixed-high



Development & Production Team Setup

- HDInsight
  - Deploy through Azure Portal
  - Deploy with Powershell and Windows Azure SQL Database
  - Multiple Storage Accounts and Configuration Values
- HDP
  - Deploy in Azure IAAS
  - Deploy on Premise Clusters
- Linux or Windows?
- You can build your own!



Onsite Deployment

- Multiple options
- Multiple considerations
- Options
    - Raw Apache
    - Cloudera
    - Hortonworks

Onsite Considerations

- Your current environment
- Team’s current skill set
- Licensing

Linux Deployment Options

- Been around the longest
- Apache Hadoop
    - Most difficult to deploy
- Cloudera
- Hortonworks (HDP)
- Cloudera and HDP come with their own applications

Wrong Problems for Big Data

- Fast response
- Structured data
- Highly transactive
- Application database

- What are you going to do with it?

- Don’t get in to Big Data because everyone else is
- You WILL fail without a plan
- Do you have the political capital to take your time and build it correctly?
- Do you have the correct use case?

- What policies and procedures do you have in place?

- Do you have an archiving policy?
- Master Data policy
- Data Governance

- Why?

- Why do you think you need Big Data?
- What is the benefit to the business?
- How can you calculate the ROI?

- Introductions
- Enterprise BI
- Our customer for the day
- Analyze Your Company’s Data Lifecycle Risk
- Big Data Architecture & Deployment
- Business & Predictive Analytics
- Creating your Personalized Action Plan
- Wrap-up, Resources etc..

BUSINESS & PREDICTIVE ANALYTICS

QUESTIONS ABOUT YOUR ADVANCED ANALYTICS	HIGH RISK	MEDIUM RISK	LOW RISK
How does your business perform forecasting?	*No clue what they do.*	*Custom algorithms built into our BI system.*	*We have some third party tools.*
How do you predict success of your business?	*I'd be rich if I could do that.*	*Put together some projects but never validate.*	*Compare to similar and use our past success/failures .*
What kind of Predictive Analysis do you have for sales, production, budgeting, hardware failure, schedule,etc?	*Let me break out my crystal ball.*	*Do some but not all.*	*Using Machine Learning.*
Does Data drive your business decisions?	*Business is smarter than the data.*	*Data plays a roll.*	*All decisions are validated by data.*
How do you perform Cost Reduction Analysis?	*When money is tight accounting lets us know.*	*Always looking for ways to save.*	*Effeciency is a metric we are held to.*
How do you evaluate yourself against competitors?	*If our stock price is down that's not good.*	*We get some generic industry numbers and compare.*	*Compare directly on same products/services.*
Does your organization view analytics as a cost or investment?	*All about the benjamins!*	*Everyone wants it to be an investment, but finance doesn't agree.*	*Investment and we track ROI.*
Who drives the analytics in your organization?	*A few users who love numbers.*	*Some departments but not all.*	*Senior Management drives it from top down.*
Do you have a culture of curiosity?	*Users have been told not to think out of the box.*	*They love Access and Excel!*	*They love Self Service BI tools.*
How does Senior leadership view of data?	*Still thinks BI is a buzz word.*	*Not 100 % sold on it yet but making progress.*	*We have a data driven business culture.*
Do you provide information or inform decisions?	*What does that mean?*	*We would inform them if we were asked .*	*Decisions are prioritized and the data is analyzed.*
What training do you provide to your employees?	*What good does training do?*	*We always budget for training but its hard execute.*	*Our employees are required to attend a certain number of training courses a year.*

Category Score
High: 25-30
Medium: 18-25
Low: 1-18

Totals

HIGH	MEDUIM	LOW

- The keys to successful a implementation of analytics

- Data Usage
- Skills
- Company Culture related to data

- How do you use your data? Now and moving forward

- Do you use data to make decisions?
- How do you implement those decision making level datapoints?
- How do you perform Cost Reduction Analysis?
- How do you evaluate your business currently?

- What is your existing skill set?

- Do you have team members that understand forecasting?
- Do you have data scientists or those types of personalities?

- What is your Company Culture related to Data?

- Does your company value the data is has?
- Are you already using it for decision making?
- Is your data trusted at all?
- Do you have a curious nature?

• Creating an “insightful” business platform

- What makes up an insightful platform?
  - Using the right tools for the right thing
  - Creating a platform that serves valuable data
  - Focused on solving business problems not technology problems
- How do we combine insight platforms to create business optimization?
  - Use more than what you’re using now
  - Look to new components for new capabilities
  - Realize the true “adoption point” of a business solution
- What are the components of a insights platform?
  - Highly available data stores
  - Integrated, flexible data sets that focus on critical business analysis
  - Consumption tools that just work for the business
  - Analytics that help you predict what’s coming and make the best decisions

• Finding value in your data..

- How do we “value data” ?
- How does your business value data?
- Why is Data's value is never stationary?
- How do we process different kinds of data in different ways to optimize value ?
- How does your current process impact your data’s value?
  - Access at any point?
  - Consume as desired?
  - Analyze or leverage in new ways?
- How does a modern approach increase data’s value?
  - Availability?
  - Speed and accessibility?
  - Creating “app mindset”?

- Operational analytics are a requirement

  - What do you mean operational?
  - How do we move from looking back to looking forward?
  - What are challenges in operationalizing analytics?
  - How do we make sure this is successful?

Data Mining? What is that?

- Finding meaning in data
- Types
  - Anomaly Detection
  - Association rule learning (market basket analysis)
  - Clustering
  - Classification
  - Regression
  - Summarization

What is R?

- Open Source
- Does graphing
- Statistical Calculations
- Built on top of C, C++, & Fortran
- Extensible

Uses

- Graphing Equations
  - Plotting data for identifying patterns
- Algebra
  - Solving equations
- Statistical Analysis of Data
  - Advanced statistical algorithms

R Services in SQL Server

- Utilizes Microsoft Open R
- Runs through Stored Procedures
- Utilizes server resources to produce analytics

Let’s take a look at Azure ML

- Full Data Mining set
- Pay for the processing time of your model
- Comes with canned algorithms
  - Statistical and Forecasting algorithms
- Can even run your own R scripts
- Designed to run within your applications

- Introductions
- Enterprise BI
- Our customer for the day
- Analyze Your Company’s Data Lifecycle Risk
- Big Data Architecture & Deployment
- Business & Predictive Analytics
- Creating your Personalized Action Plan
- Wrap-up, Resources etc..

DEMO

- Let’s Take a Look at a Modern Data Architecture Example

How do you compare to Binford?

1. Let's analyze your risk factors
2. Please review the questions as I go through the sections
3. Circle the answer that best describes your answer to the question
4. Add up the circles in each column and tally it up at the bottom
5. Add up all the columns and get your category score
    1. High risk = 5 points
    2. Medium = 3 points
    3. Low = 1 point
6. Transfer those numbers and add them up to get your category score
7. Add the columns to get your organizational risk scores
8. Make some notes on your observations…..

	HIGH RISK	MEDIUM RISK	LOW RISK	CATEGORY SCORE
ARCHITECTURE & CONFIGURATION				
AVAILABILITY & CONTINUITY				
PERFORMANCE & OPTIMIZATION				
ENTERPRISE BI				
BIG DATA ARCHITECTURE & DEPLOYMENT				
BUSINESS & PREDICTIVE ANALYTICS				
Organizational Risk				

Notes

How did you do?

- Please take out your handout and turn to page 5
  - **Group Discussion - Let's Analyze Your Risk factors**
    - Architecture and Configuration
      - Availability and Continuity
    - Performance and Optimization
      - Big Data and Non-relational
        - Advanced Analytics

1. Think Through Your Strategy

- **Go section by section and brainstorm your future vision for your organization**
    1. Identify gaps in your current strategy
    2. Identify new opportunities for process improvement
    3. List new tools or technologies that would help you close those gaps
    4. Identify stakeholders by section or gap
    5. Gather stakeholders and review your analysis and vision
    6. Document feedback and update vision plan

2. Understand Your Environment

- **Go section by section and create a plan for assessing your current environment**
    1. Define metrics and measurements that are important to that section's maturity
    2. Set specific target goals with the stakeholders to measure success
    3. Look for anything that stands in the way of your vision (people, process or tech)
    4. Pay special attention to team training and readiness

3. Prioritize Your Goals

- **Following the maturity and risk model of DLO, prioritize the sections and projects**
    1. Sections increase in risk from the left to the right (i.e. gaps in architecture are higher risk than gaps in big data)
    2. Get input on strategic initiatives from stakeholders (they may override DLO priority)
    3. Document these priorities and share them with your team

4. Create Your Roadmap

- **Assign and scope prioritized projects**
    1. Look for alignment between resources and objectives
    2. Assign initiatives for scoping and analysis
    3. Focus on scoping v1 of solution
    4. Gather scoping and create a timeline and roadmap for agile development
    5. Kick start any training or readiness needed

5. Start Your Journey

- **Design and deploy your pilot or phase I**
    1. Bring in the stakeholders and finalize scope for pilot
    2. Assign resources and setup access to technology solution
    3. Develop Phase I and release to business
    4. Gather feedback and integrate into planning future sprints

6. Measure Your Success

- Measure phase I client satisfaction
- Measure performance against metrics you chose with the business
- Report on those metrics and track them across future phases

7. Re-Evaluate

- Re-evaluate the roadmap and priority of next sections' objectives
- Identify new gaps and compare to existing priorities
- Review with stakeholders and discus risk and benefits of future endeavors
- Set next round of goals and move forward

- Review our case study progress

- Discussion

- How to Prioritize in your plan..

  - Let’s set some priorities – togethe
  - Let’s fill in that plan!

Today's Agenda

- Introductions
- Our customer for the day
- What is Data Lifecycle Optimization?
- Analyze Your Company's Data Lifecycle Risk
- Architecture and Configuration
- Availability and Continuity
- Performance and Optimization
- Enterprise BI
- Big Data Architecture & Deployment
- Business & Predictive Analytics
- Creating your Personalized Action Plan
- **Wrap-up, Resources etc..**

Resources

Thank You