



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps

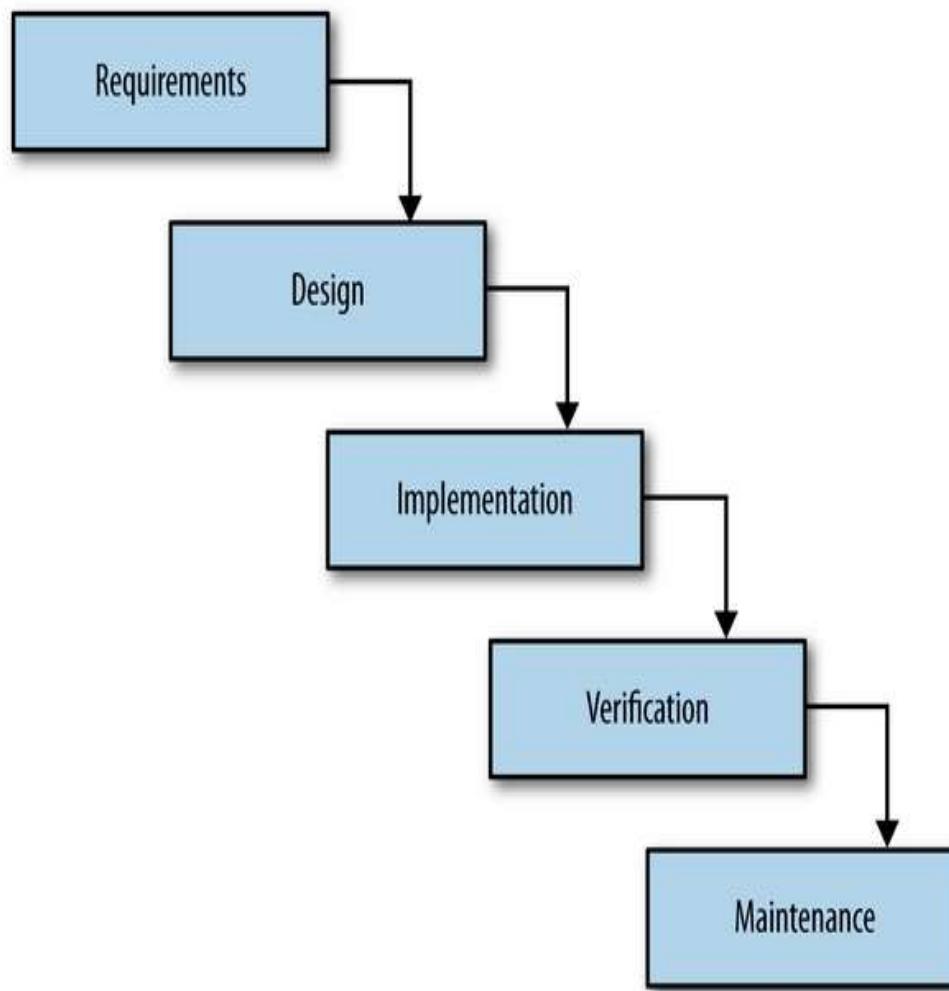




CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 1

Waterfall



Pros of Waterfall

Cons of Waterfall



Agile

- *individuals and interactions over processes and tools*
- *working software over comprehensive documentation*
- *customer collaboration over contract negotiation*
- *responding to change over following a plan*



Scrum

- **Scrum**
- In the mid-1990s, Ken Schwaber and Dr. Jeff Sutherland, two of the original creators of the Agile Manifesto, merged individual efforts to present a new software development process called Scrum. Scrum is a software development methodology that focuses on maximizing a development team's ability to quickly respond to changes in both project and customer requirements. It uses predefined development cycles called *sprints*, usually between one week and one month long, beginning with a sprint planning meeting to define goals and ending with a sprint review and sprint retrospective to discuss progress and any issues that arose during that sprint.

DSM

- What did I do yesterday that helped the team meet its sprint goals?
- What am I planning to do today to help the team meet those goals?
- What, if anything, do I see that is blocking either me or the team from reaching their goals?

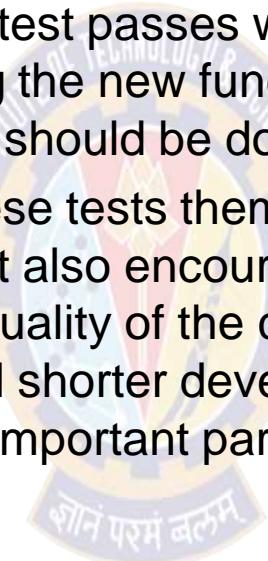


Version Control

- A version control system records changes to files or sets of files stored within the system. This can be source code, assets, and other documents that may be part of a software development project. Developers make changes in groups called *commits* or *revisions*. Each revision, along with metadata such as who made the change and when, is stored within the system in one way or another.
- Having the ability to commit, compare, merge, and restore past revisions to objects to the repository allows for richer cooperation and collaboration within and between teams. It minimizes risks by establishing a way to revert objects in production to previous versions.

Test Driven Development

- In test-driven development, the code developer starts by writing a failing test for the new code functionality, then writes the code itself, and finally ensures that the test passes when the code is complete. The test is a way of defining the new functionality clearly, making more explicit what the code should be doing.
- Having developers write these tests themselves not only greatly shortens feedback loops but also encourages developers to take more responsibility for the quality of the code they are writing. This sharing of responsibility and shorter development cycle time are themes that continue to be important parts of a devops culture.

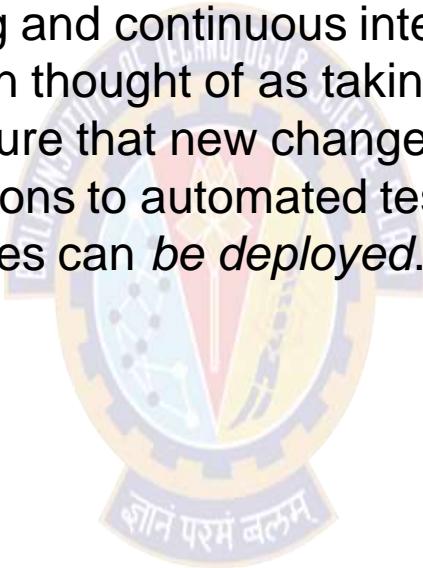


Continuous Integration

- Continuous integration (CI) is the process of integrating new code written by developers with a mainline or “master” branch frequently throughout the day. This is in contrast to having developers working on independent feature branches for weeks or months at a time, merging their code back to the master branch only when it is completely finished. Long periods of time in between merges means that much more has been changed, increasing the likelihood of some of those changes being breaking ones. With bigger changesets, it is much more difficult to isolate and identify what caused something to break. With small, frequently merged changesets, finding the specific change that caused a regression is much easier. The goal is to avoid the kinds of integration problems that come from large, infrequent merges.

Continuous Delivery

- Continuous delivery (CD) is a set of general software engineering principles that allow for frequent releases of new software through the use of automated testing and continuous integration. It is closely related to CI, and is often thought of as taking CI one step further, that beyond simply making sure that new changes can be integrated without causing regressions to automated tests, continuous delivery means that these changes can *be deployed*.

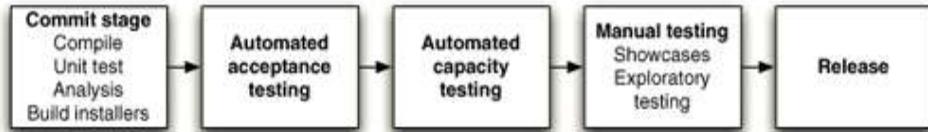


Continuous Deployment

- Continuous deployment (also referred to as CD) is the process of deploying changes to production by defining tests and validations to minimize risk. While continuous delivery makes sure that new changes can be deployed, continuous deployment means that they *get deployed* into production.



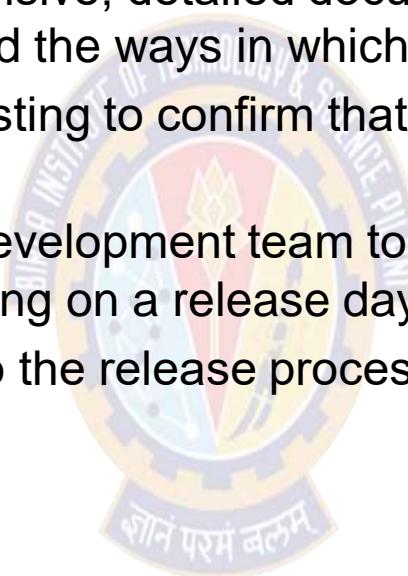
Deployment Pipeline



- The pattern that is central to this book is the *deployment pipeline*. A deployment pipeline is, in essence, an automated implementation of your application's build, deploy, test, and release process. Every organization will have differences in the implementation of their deployment pipelines, depending on their value stream for releasing software, but the principles that govern them do not vary. An example of a deployment pipeline is given in Figure

Antipattern: Deploying Software Manually

- The signs of this antipattern are:
 - The production of extensive, detailed documentation that describes the steps to be taken and the ways in which the steps may go wrong
 - Reliance on manual testing to confirm that the application is running correctly
 - Frequent calls to the development team to explain why a deployment is going wrong on a release day
 - Frequent corrections to the release process during the course of a release



Antipattern: Deploying to a Production-like Environment Only after Development Is Complete

- If testers have been involved in the process up to this point, they have tested the system on development machines.
- Releasing into staging is the first time that operations people interact with the new release. In some organizations, separate operations teams are used to deploy the software into staging and production. In this case, the first time an operations person sees the software is the day it is released into production.
- Either a production-like environment is expensive enough that access to it is strictly controlled, or it is not in place on time, or nobody bothered to create one.
- The development team assembles the correct installers, configuration files, database migrations, and deployment documentation to pass to the people who perform the actual deployment—all of it untested in an environment that looks like production or staging.

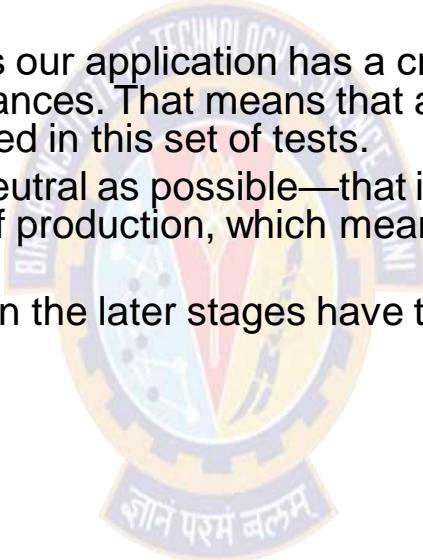
Every Change Should Trigger the Feedback Process

- Executable code changes when a change is made to the source code. Every time a change is made to the source code, the resulting binary must be built and tested. In order to gain control over this process, building and testing the binary should be automated. The practice of building and testing your application on every check-in is known as continuous integration



Possible

- They run fast.
- They are as comprehensive as possible—that is to say, they cover more than 75% or so of the codebase, so that when they pass, we have a good level of confidence that the application works.
- If any of them fails, it means our application has a critical fault and should not be released under any circumstances. That means that a test to check the color of a UI element should not be included in this set of tests.
- They are as environment-neutral as possible—that is, the environment does not have to be an exact replica of production, which means it can be simpler and cheaper.
- On the other hand, the tests in the later stages have the following general characteristics.





Thank You!

Acknowledgement:

- Significant portions of the information in the presentation is from IT Systems Management - Rich Schiesser and other books/Internet. Permission was requested from the publisher, sometime ago for use, but has not been received as yet. This is intended solely to teach BITS students enrolled for IM. I would like to acknowledge and reiterate that all the credit/rights remain with the original authors/publishers only. Mistakes if any are mine.



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps





CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 2 and 3

Agenda

DevOps Misconceptions and Dimensions

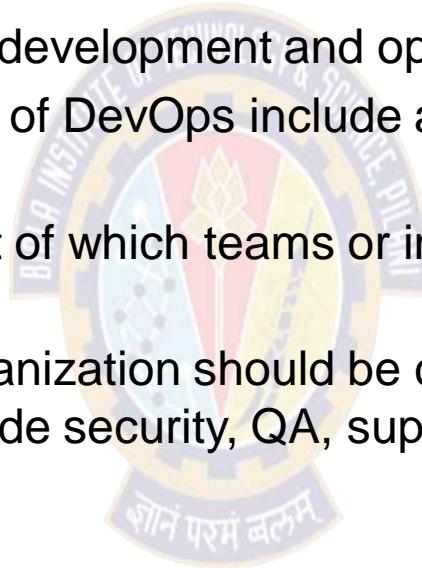
- DevOps Misconceptions
- DevOps Antipatterns
- Three Dimensions of DevOps



DevOps Misconceptions

DevOps Only Involves Developers and System Administrators

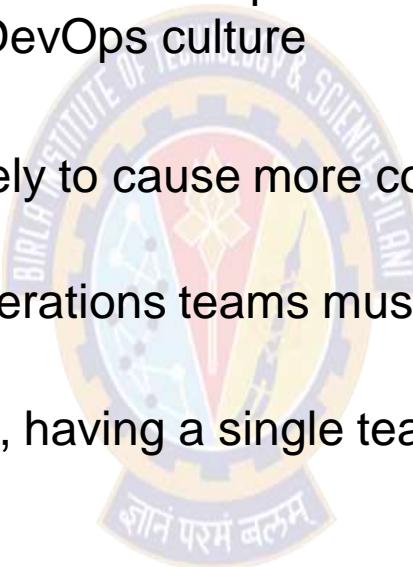
- As the name indicates Development and Operations
- **However the Truth is:**
- DevOps tagline is bring development and operations together
- The concepts and ideas of DevOps include all roles within an organization
- There is no definitive list of which teams or individuals should be involved
- Any team within the organization should be considered as a candidate for DevOps, it may include security, QA, support, and legal



DevOps Misconceptions Contd..

DevOps is a Team

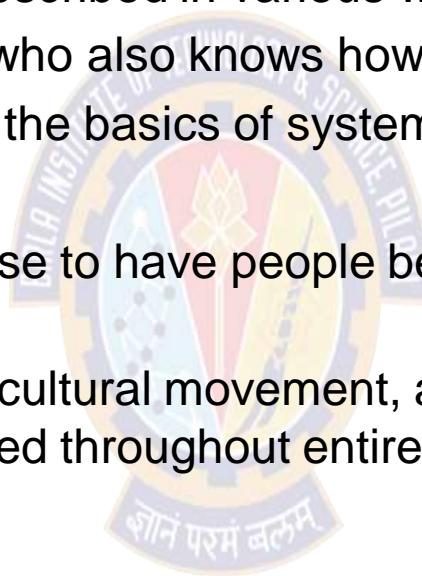
- Creating a team called DevOps, OR
- Renaming an existing team to DevOps is neither necessary nor sufficient for creating a DevOps culture
- **However truth is:**
- An additional team is likely to cause more communication issues, not fewer
- the development and operations teams must connect frequently or it should work together
- In a startup environment, having a single team that encompasses both functions can work



DevOps Misconceptions Contd..

DevOps is a Job Title

- The “DevOps engineer” job title has started a controversial debate
- The job title has been described in various ways
- A system administrator who also knows how to write code
- A developer who knows the basics of system administration
- **However truth is:**
- In DevOps it makes sense to have people become more specialized in their job role
- DevOps is at its core a cultural movement, and its ideas and principles need to be used throughout entire organizations in order to be effective



DevOps Misconceptions Contd..

You Need a DevOps Certification

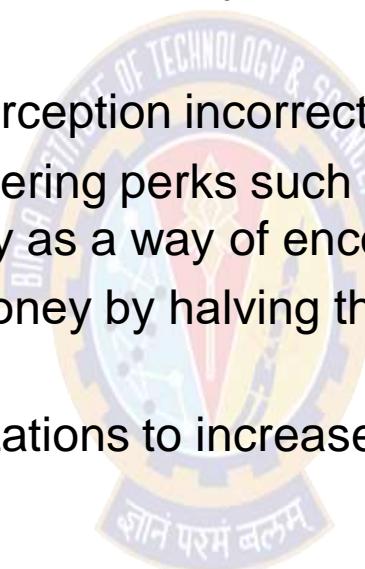
- Certification exams are testing knowledge where there are clear right or wrong answers, which DevOps generally does not have
- **However truth is:**
- A significant part of DevOps is about culture: how do you certify culture?
- DevOps doesn't have required technology or one-size-fits-all solutions



DevOps Misconceptions Contd..

DevOps Means Doing All the Work with Half the People

- Get both a software developer and a system administrator in one person and with one person's salary
- **However truth is:**
- Not only is this above perception incorrect, but it is often harmful
- At a time startups are offering perks such as three meals a day in the office and on-site laundry as a way of encouraging
- DevOps doesn't save money by halving the number of engineers your company needs
- Rather, it allows organizations to increase the quality and efficiency of their work



DevOps Misconceptions Contd..

DevOps Is About Automation

- Many innovations in DevOps-adjacent tools help to codify understanding, bridging the gaps between teams and increasing velocity through automation
- Practitioners have focused on tools that eliminate tasks that are boring and repetitive, such as with infrastructure automation and continuous integration
- **However truth is:**
- In both of these above cases, automation is a result of improved technology
- Automating repetitive tasks automation will help to human from having to do them, that automation helps that person work more efficiently
- Example: Automating server builds saves hours per server that a system administrator can then spend on more interesting or challenging work

DevOps Misconceptions Contd..

DevOps Is a FAD

- It's a Buzzword in Market
- Just a new word in place
- **However truth is:**
- As DevOps is not specific to a technology, tool, or process however, It is movement about improving organizational effectiveness and individual employee happiness
- One of the primary differences between DevOps and methodologies like ITIL and Agile is:
- It does not have strict definition
- DevOps is movement defined by stories and ideas of individuals, teams, and organizations
- DevOps is the evolution of processes and ideas that lead growth and change
- Harnessing and leveraging effective DevOps will lead to growth and evolution in tools, technology, and processes in your organization

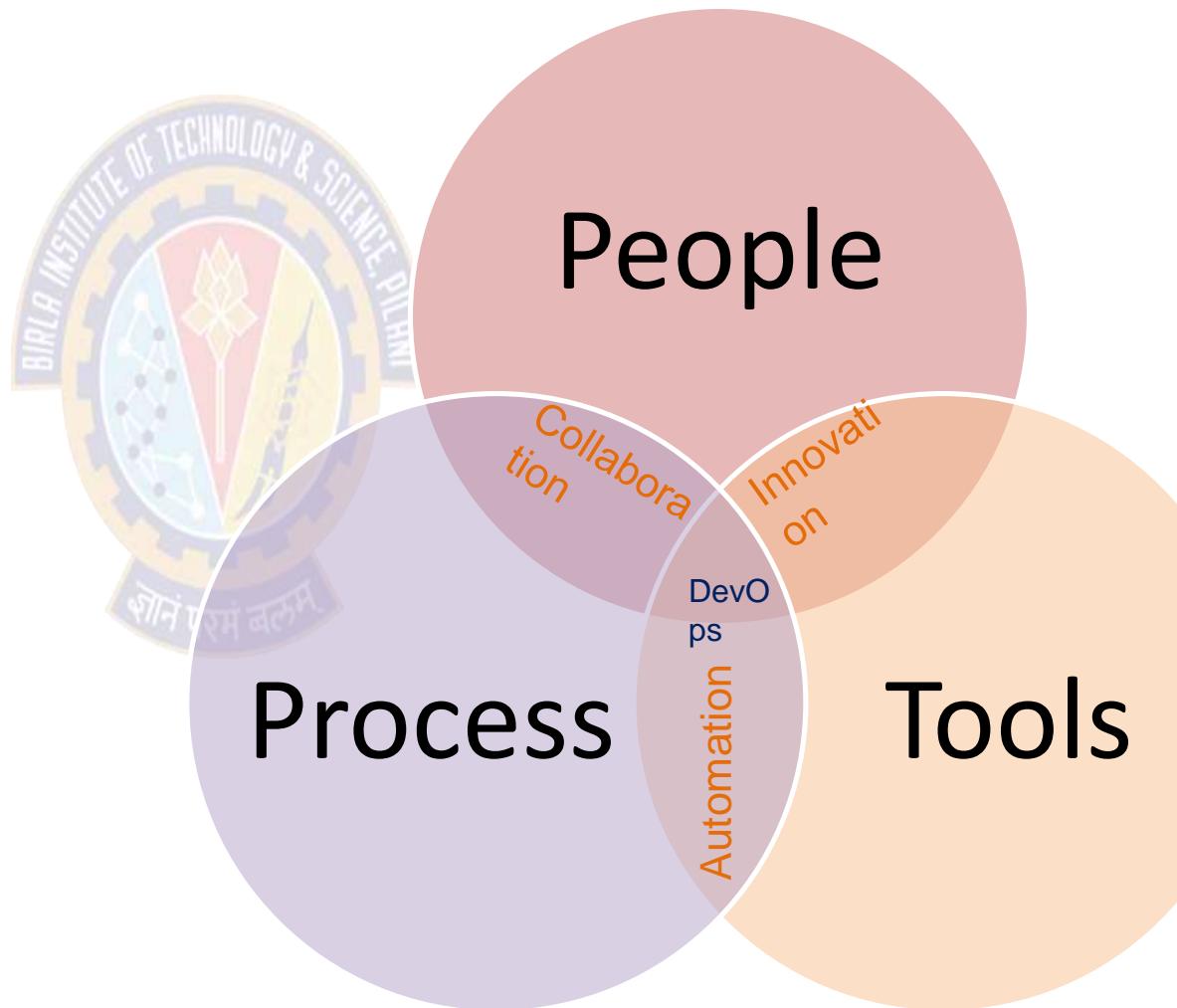
DevOps Anti-Patterns

- Blame Culture
 - A blame culture is one that tends toward blaming and punishing people when mistakes are made, either at an individual or an organizational level
- Silos
 - A departmental or organizational silo describes the mentality of teams that do not share their knowledge with other teams in the same company
- Root Cause Analysis
 - Root cause analysis (RCA) is a method to identify contributing and “root” causes of events or near-misses/close calls and the appropriate actions to prevent recurrence
- Human Errors
 - Human error, the idea that a human being made a mistake that directly caused a failure, is often cited as the root cause in a root cause analysis

Three dimensions of DevOps

3-D of DevOps

- People
- Process
- Tools / Technology



Agenda

Agile Methodology SCRUM

- Scrum



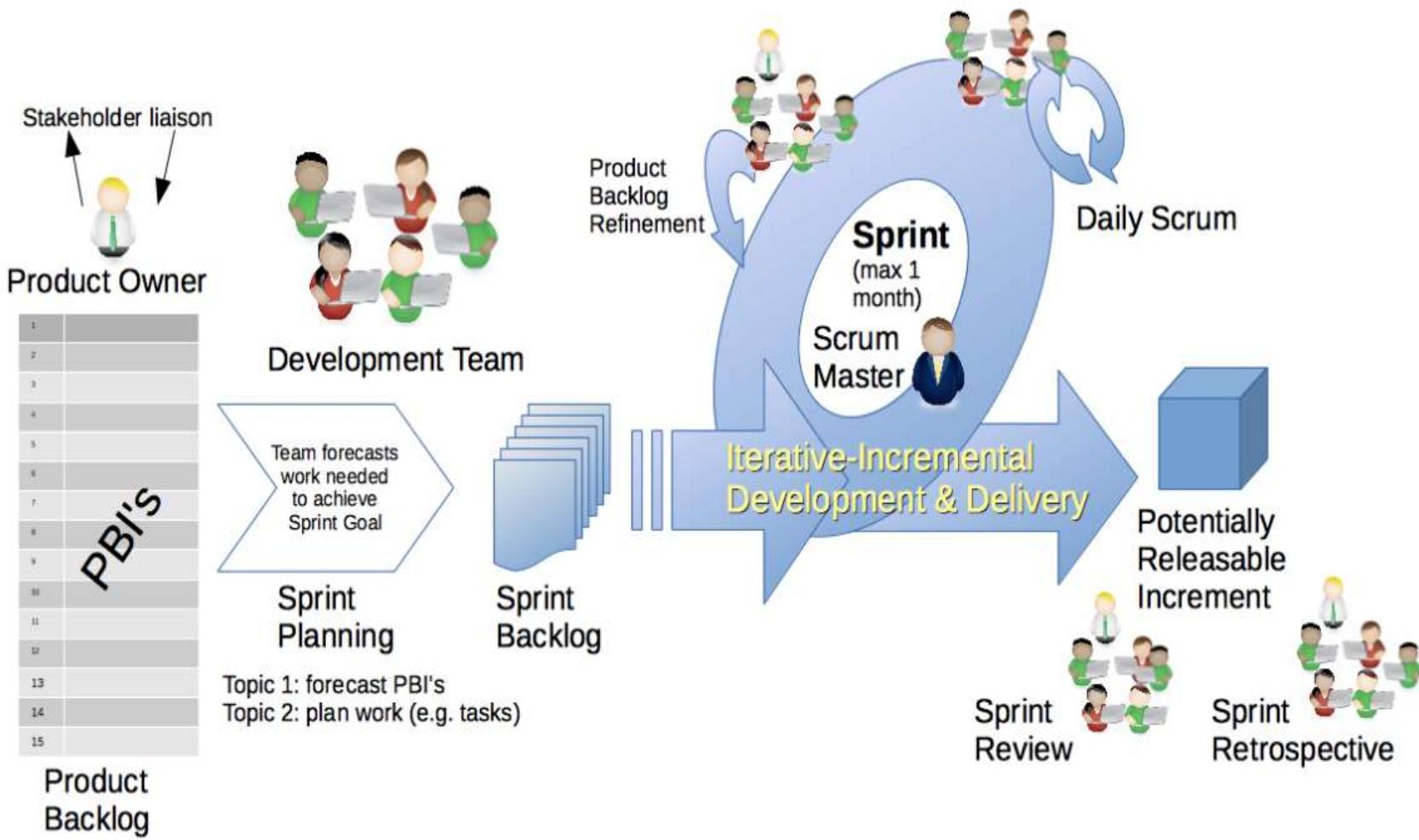
SCRUM Methodology

About Scrum

- Scrum is part of the Agile movement
- Scrum is an agile way to manage a project, usually software development
- Scrum's early advocates were inspired by empirical inspect and adapt feedback loops to cope with complexity and risk
- Scrum emphasizes decision making from real-world results rather than speculation
- By Mountain Goat Software – Pioneer of Agile
- “Agile software development with Scrum is often perceived as a methodology; but rather than viewing Scrum as methodology, think of it as a framework for managing a process”

SCRUM

Scrum Sprints



SCRUM

Benefits of Scrum

- Flexibility
- Quality
- Incremental Working Product
- Early to Market



Agenda

Agile Methodologies

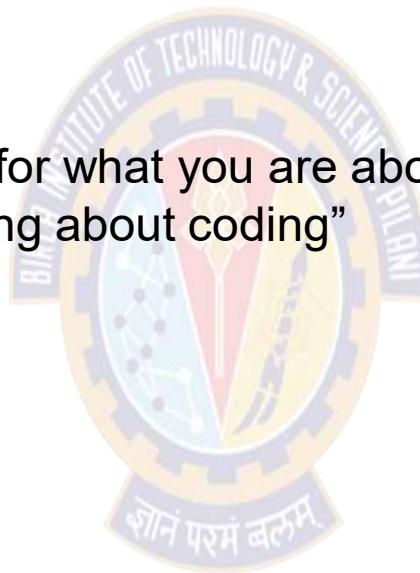
- Test Driven Development
- Feature Driven Development
- Behavior Driven Development



TDD

TDD Introduction

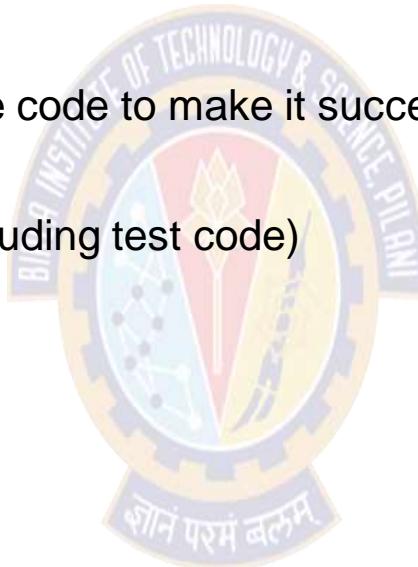
- Kent Beck said “Test-first code tends to be more cohesive and less coupled than code in which testing isn’t a part of the intimate coding cycle”
- “If you can’t write a test for what you are about to code, then you shouldn’t even be thinking about coding”



TDD

TDD Three steps

- **RED**
 - Write a new TEST which fails
- **GREEN**
 - Write simplest possible code to make it succeed
- **REFACTOR**
 - Refactor the code (including test code)



TDD

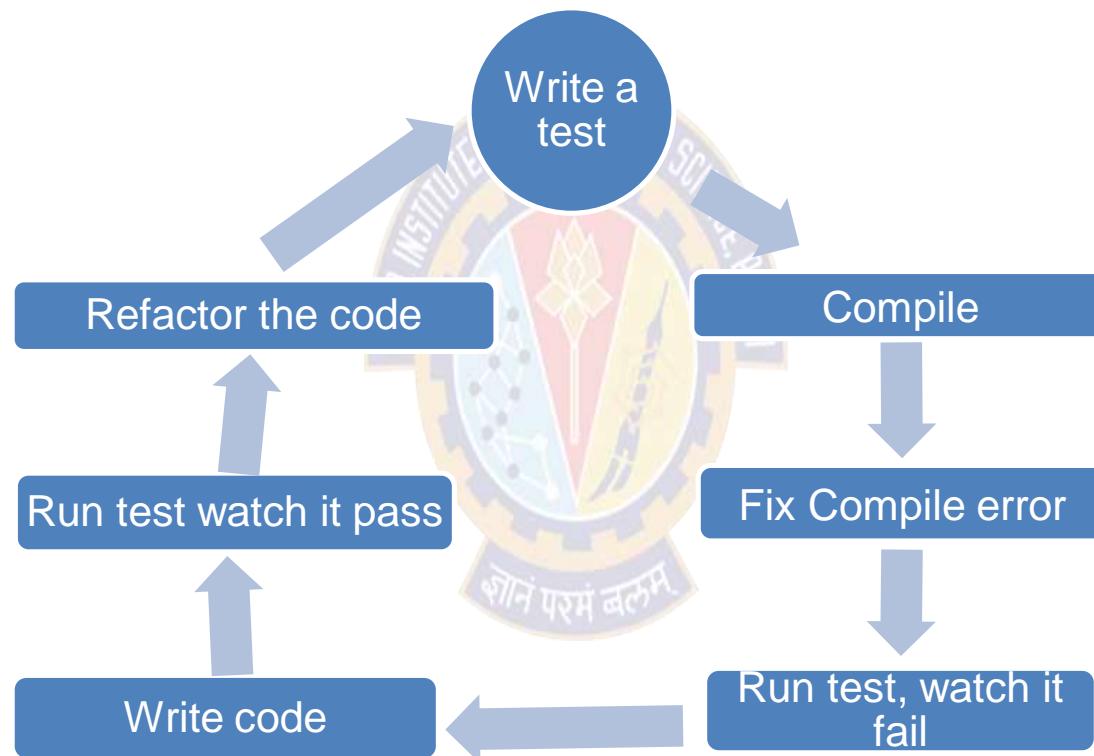
TDD Technique

- TDD is a technique whereby you write your test cases before you write any implementation code



TDD

TDD Overview



TDD

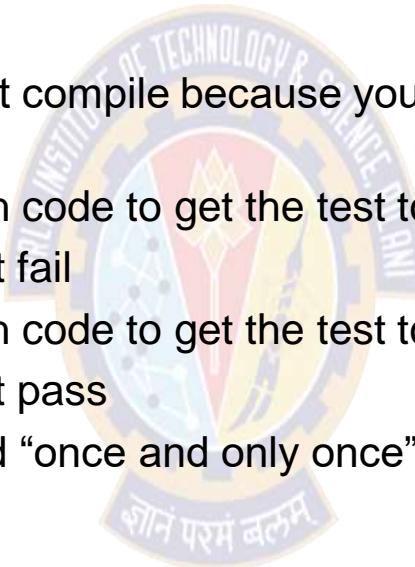
TDD Overview Contd..

- How It Works:
- Add a Test
 - Use Cases / User Stories are used to understand the requirement clearly
- Run all tests and see the new one fail
 - Ensures test harness is working correctly
 - Ensures that test does not mistakenly pass
- Write some code
 - Only code that is designed to pass the test
 - No additional functionality should be included because it will be untested
- Run the automated tests and see them succeed
 - If tests pass, programmer can be confident code meets all tested requirements
- Refactor code
 - Cleanup the code
 - Rerun tests to ensure cleanup did not break anything
- Repeat

TDD

TDD Overview Contd..

- In Extreme Programming Explored (The Green Book), Bill Wake describes the test / code cycle:
 1. Write a single test
 2. Compile it. It shouldn't compile because you've not written the implementation code
 3. Implement just enough code to get the test to compile
 4. Run the test and see it fail
 5. Implement just enough code to get the test to pass
 6. Run the test and see it pass
 7. Refactor for clarity and “once and only once”
 8. Repeat



TDD

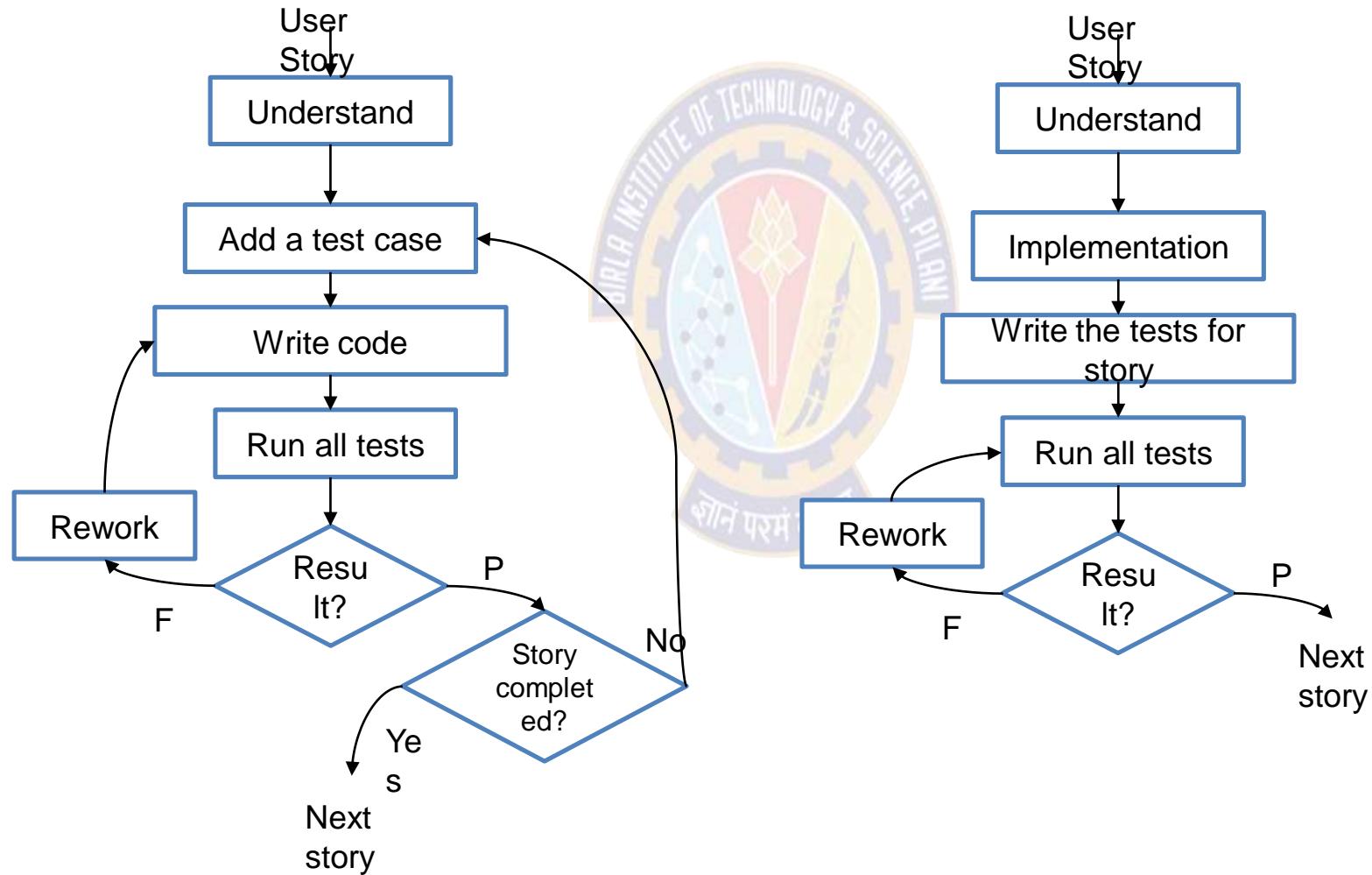
Why TDD

- TDD can lead to more modularized, flexible, and extensible code
- Clean code
- Leads to better design
- Better code documentation
- More productive



TDD

Test First vs. Test Last



Feature Driven Development

FDD History

- Original Creator: Jeff De Luca
 - Singapore in late 1997
- FDD evolved from an actual project
 - Bank Loan Automation
 - Luca was Project manager
 - 50 member developer team
 - Peter Coad : Chief Architect
 - 1990's object-oriented analysis and design expert



FDD

What is a Feature?

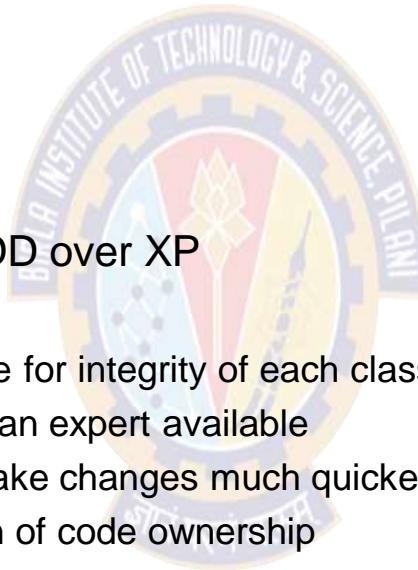
- Definition: small function expressed in client-valued terms
- FDD's form of a customer requirement



FDD

FDD Primary Roles

- Project Manager
- Chief Architect
- Development Manager
- Domain Experts
- Class Owners
 - This concept differs FDD over XP
 - Benefits
 - Someone responsible for integrity of each class
 - Each class will have an expert available
 - Class owners can make changes much quicker, if needed anytime
 - Easily lends to notion of code ownership
 - Assists in FDD scaling to larger teams, as we have one person available for complete ownership of feature.
- Chief Programmers



FDD

FDD Supporting Roles

- Release Manager
- Language Guru
- Build Engineer
- Toolsmith
- System Administrator
- Tester
- Deployers
- Technical Writer



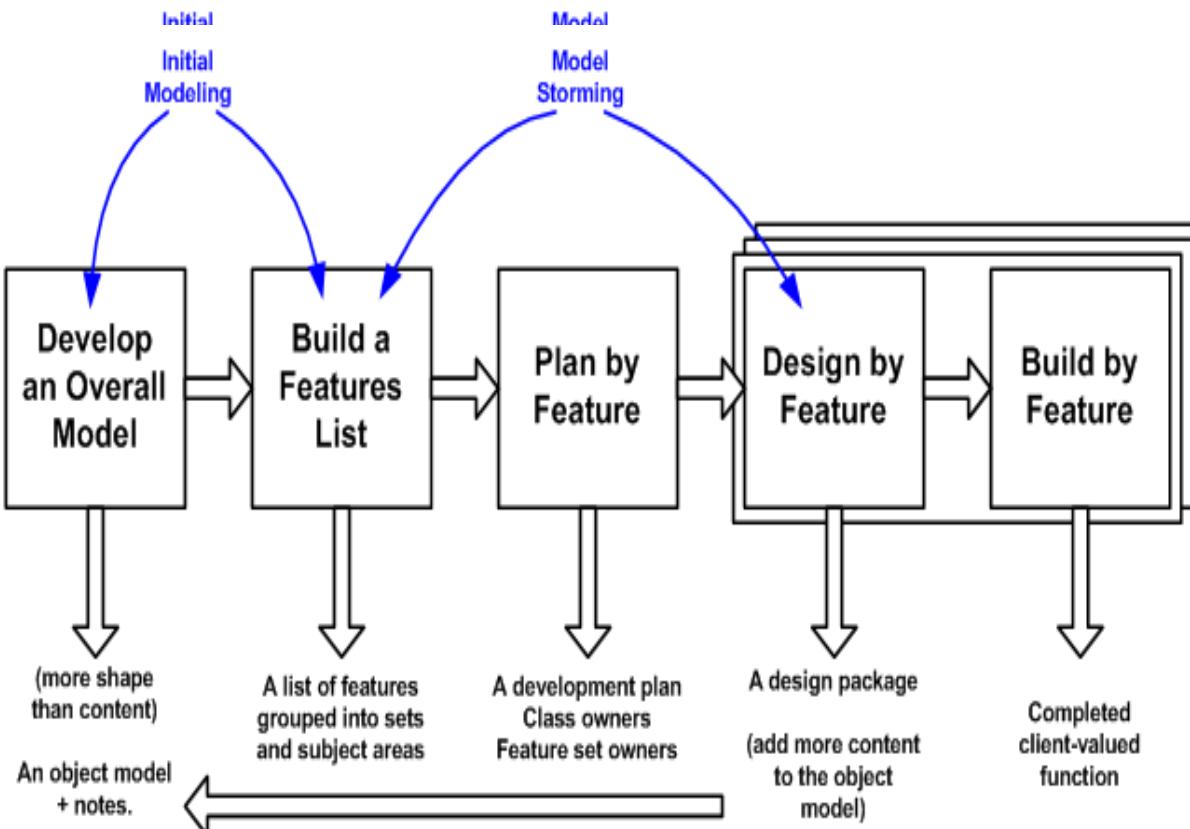
FDD

Feature Driven Development Process

- Process #1: Develop an Overall Model
- Process #2: Build a Features List
- Process #3: Plan By Feature
 - Constructing the initial schedule, Forming level of individual features, Prioritizing by business value , As we work on above factors we do consider dependencies, difficulty, and risks.
 - These factors will help us on Assigning responsibilities to team members, Determining Class Owners , Assigning feature sets to chief programmers
- Process #4: Design By Feature
 - Goal: not to design the system in its entirety but instead is to do just enough initial design that you are able to build on
 - This is more about Form Feature Teams: Where team members collaborate on the full low level analysis and design.
- Process #5: Build By Feature
 - Goal: Deliver real, completed, client-valued function as often as possible

FDD

Feature Driven Development Process Contd..



Behavior Driven Development

BDD Introduction

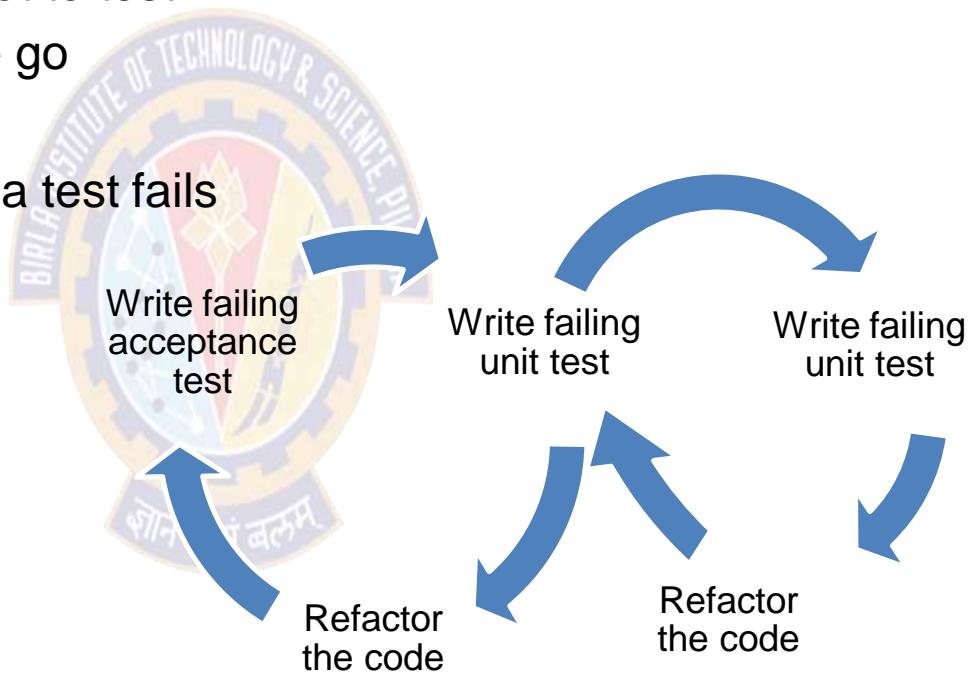
- Behavior-Driven Development (BDD) is a software development process
- BDD is considered an effective technical practice especially when the "problem space" for the business is complex
- What is BDD?:
 - General Technique of TDD
 - Follows same principle as TDD
 - Shared tools
 - Shared Process



BDD

BDD Focus

- Where to start in the process
- What to test and what not to test
- How much to test in one go
- What to call the tests
- How to understand why a test fails



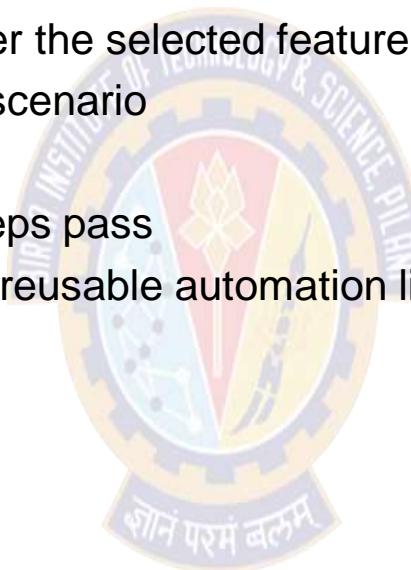
BDD

TDD

BDD

BDD Basic structure

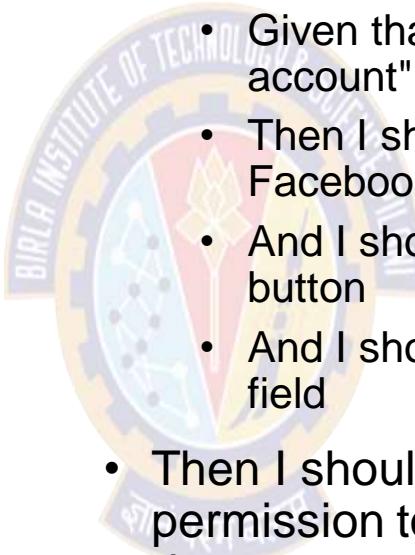
- BDD consists of cycles of a set of steps to follow.
 - Identify business feature
 - Identify scenarios under the selected feature
 - Define steps for each scenario
 - Run feature and fail
 - Write code to make steps pass
 - Refactor code, Create reusable automation library
 - Run feature and pass
 - Generate test reports



BDD

BDD Example

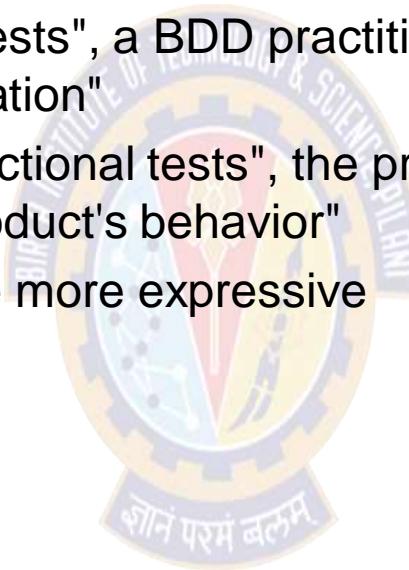
- User story:
- As someone interested in using the Mobile app, and want to sign up in app to enjoy app membership
- Mobile App Signup:
- Scenario 1:
 - Given that I am on the app's "Create new account" screen
 - Then I should see a "Sign up using Facebook" button
 - And I should see a "Sign up using Twitter" button
 - And I should see a "Sign up with email" form field
 - Then I should see a new screen that asks permission to use my Facebook account data to create my new Mobile app account



BDD

BDD Sign Use

- Functional Documentation in terms of User Stories & Executable Scenarios
- Instead of referring to "tests", a BDD practitioner will prefer the terms "scenario" and "specification"
- Rather than refer to "functional tests", the preferred term will be "specifications of the product's behavior"
- Specification Names are more expressive



References

- <http://agilemodeling.com/>



Agenda

DevOps – People

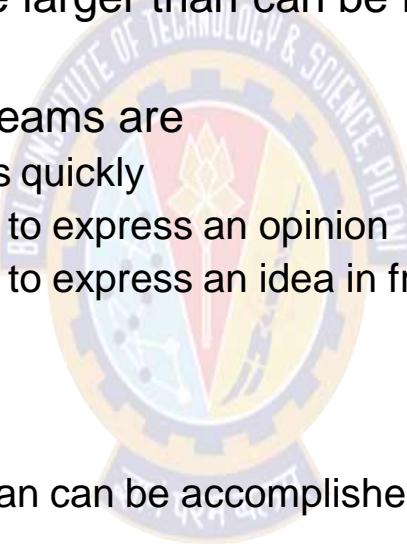
- DevOps Team Structure
- Team Coordination



Team Structure

Team Size

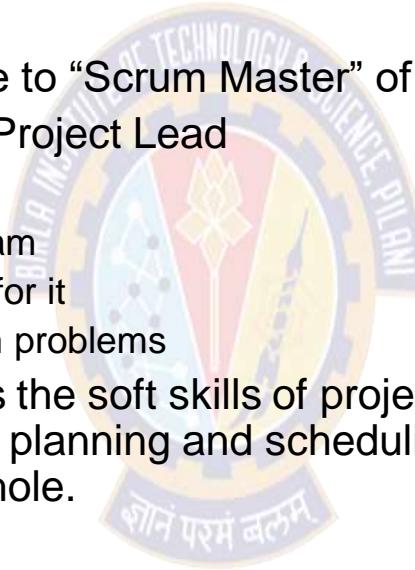
- The size of the team should be relatively small
- Amazon has a “two pizza rule”
- That is, no team should be larger than can be fed from two pizzas
- The advantages of small teams are
 - They can make decisions quickly
 - It is easier for individuals to express an opinion
 - It is easier for individuals to express an idea in front of a small group than in front of a large one
- The disadvantage
 - Some tasks are larger than can be accomplished by a small number of individuals
 - Task has to be broken up into smaller pieces
 - A small team, by necessity, works on a small amount of code



Team Structure

Team Roles

- We will study roles in the team from Scott Ambler's description of roles in an agile team
- Team lead :
 - This role you can relate to "Scrum Master" of SCRUM TEAM.
 - Also you can relate to Project Lead
 - Is responsible:
 - For facilitating the team
 - Obtaining resources for it
 - And protecting it from problems
 - This role encompasses the soft skills of project management but not the technical ones such as planning and scheduling, activities which are better left to the team as a whole.
- Team member
 - This role, sometimes referred to as developer or programmer
 - Is responsible for the creation and delivery of a system
 - This includes modeling, programming, testing, and release activities, as well as others



Team Structure

Additional roles

- Service owner
- Reliability engineer
- Gatekeeper
- And DevOps engineer



Team Structure

Service Owner

- This role on the team responsible for outside coordination
- The service owner participates:
 - System-wide requirements activities
 - Prioritizes work items for the team
 - And provides information both from the clients of the team's service and about services provided to the team.
 - The requirements gathering and release planning activities for the next iteration can occur in parallel with the conception phase of the current iteration
 - The service owner maintains and communicates the vision for the service
 - That is, the vision involves the architecture of the overall system and the team's role in that architecture

Team Structure

Reliability Engineer

- Reliability engineer monitors the service in the time period immediately subsequent to the deployment
- Second, the reliability engineer is the point of contact for problems with the service during its execution
- Google calls this role “Site Reliability Engineer.”
- Once a problem occurs, the reliability engineer performs short-term analysis to diagnose, mitigate, and repair the problem
- In any case, the reliability engineer has to be excellent at troubleshooting and diagnosis
- The reliability engineer should discover or work with the team to discover the root cause of a problem
- Increasingly, reliability engineers need to be competent developers, as they need to write high-quality programs to automate the repetitive part of the diagnosis, mitigation, and repair

Team Structure

Gatekeeper

Netflix uses the steps given in Figure 1.3 from local development to deployment.

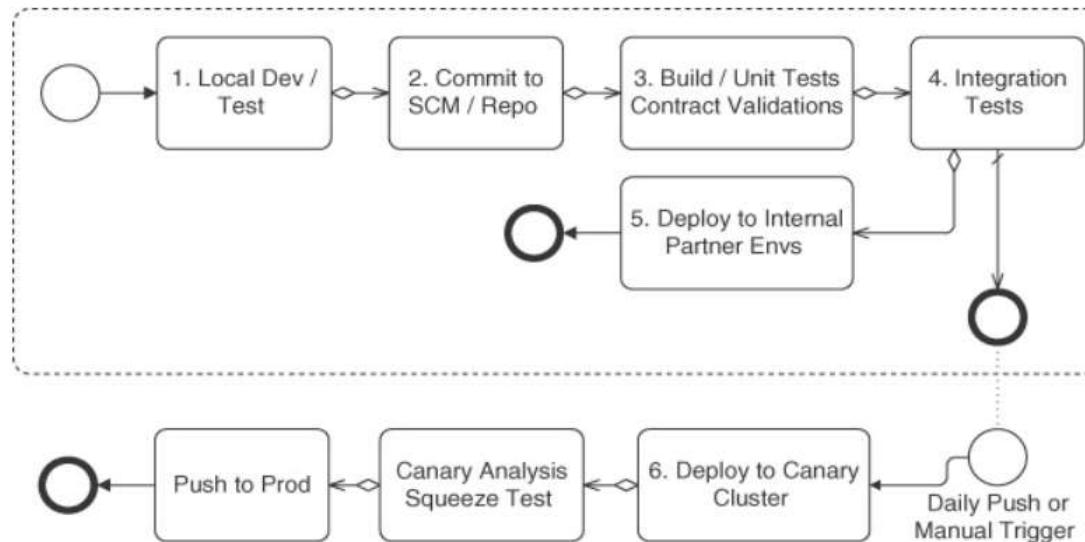
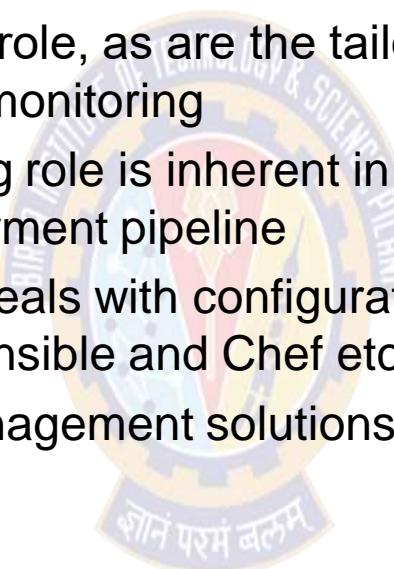


FIGURE 1.3 Netflix path to production. (Adapted from <http://techblog.netflix.com/2013/11/preparing-netflix-api-for-deployment.html>) [Notation: BPMN]

Team Structure

DevOps Engineer

- The DevOps engineer role is responsible for the care and feeding of the various tools used in the DevOps tool chain
- The DevOps engineer's role, as are the tailoring of the tool for the development team and monitoring
- The DevOps engineering role is inherent in automating the development and deployment pipeline
- DevOps engineer also deals with configuration managements tools which are like Puppet, Ansible and Chef etc.
- These configuration management solutions are applied to the source code for the service



Team Coordination

Forms of Coordination

- One goal of DevOps is to minimize coordination in order to reduce the time to market
- Two of the reasons to coordinate are, first, so that the pieces developed by the various teams will work together and, second, to avoid duplication of effort
- Forms of Coordination
 - Direct—the individuals coordinating know each other (e.g., team members)
 - Indirect—the coordination mechanism is aimed at an audience known only by its characterization (e.g. Operators)
 - Persistent—the coordination artifacts are available after the moment of the coordination (e.g., documents, e-mail, bulletin boards)
 - Ephemeral—the coordination, per se, produces no artifacts (e.g., face to face meetings, conversations, telephone/video conferencing)
 - Synchronous—individuals are coordinating in real time, (e.g., face to face)
 - Asynchronous—individuals are not coordinating in real time (e.g., documents, e-mail)

Team Coordination

Team coordination mechanisms

- Team coordination mechanisms are of two types
 - Human Processes and
 - Automated Processes
- The DevOps human processes are adopted from agile processes and are designed for high-bandwidth coordination with limited persistence
 - Examples: Stand-up meetings
- Automated team coordination mechanisms are designed to protect team members from interference of their and others' activities:
 - To automate repetitive tasks (continuous integration and deployment),
 - And to speed up error detection and reporting (automated unit, integration, acceptance, and live production tests)

Team Coordination

Cross-team Coordination

- Coordination must occur with customers, stakeholders, other development teams, and operations. Therefore, DevOps processes attempt to minimize this coordination as much as possible
- There are three types of cross-team coordination:
 - Upstream Coordination [with stakeholders and customers]
 - Downstream Coordination [with operations]
 - And Cross-Stream Coordination [with other development teams]
- There are two reasons for a development team to coordinate with other development teams
 - To ensure that the code developed by one team works well with the code developed by another
 - And to avoid duplication of effort

Agenda

DevOps People

- Transformation to Enterprise DevOps culture



Transformation to Enterprise DevOps culture

Constraints of Large Enterprise :: DevOps Transformation

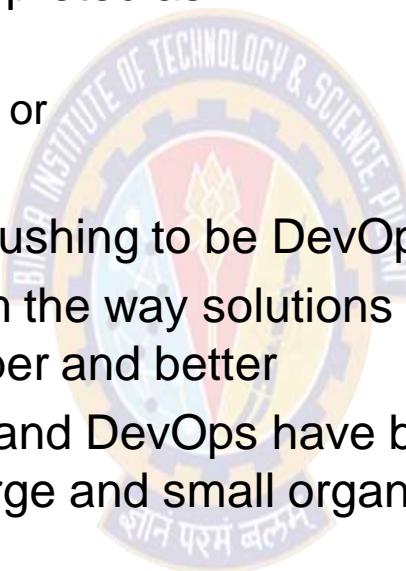
- Organization Structure
- Legacy Technology Stack
- Organization Culture



Transformation to Enterprise DevOps culture

Concepts

- Executives do not like terms such as “Transformation”
- Somehow it is often interpreted as:
 - Cost Cutting or
 - Resource Realignment or
 - Outsourcing
- Why organizations are pushing to be DevOps
- The intent is to transform the way solutions are being delivered today and to do it faster, cheaper and better
- The techniques of Agile and DevOps have been successfully implemented in many large and small organizations to enable those outcomes
- Look in to the approach that worked quite well for most enterprises



Transformation to Enterprise DevOps culture

Initial Planning for Enterprise Readiness

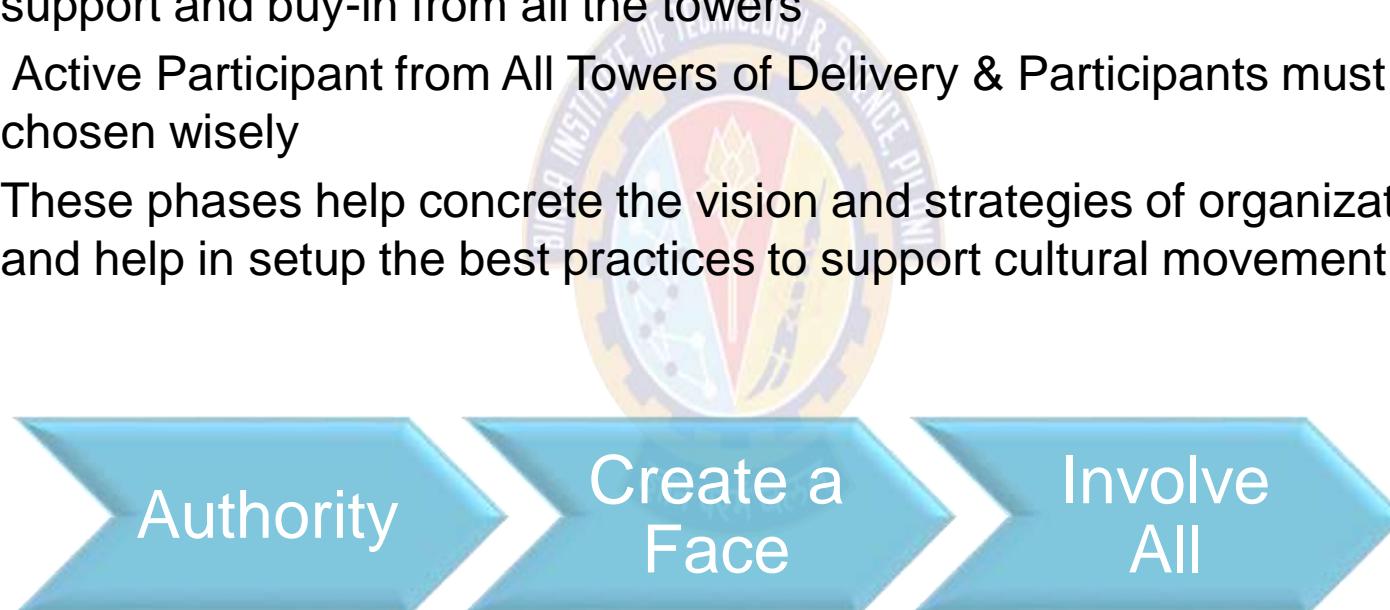
- Participants should absolutely include all the towers that make up the solution delivery
- Design Thinking : It is a great method; it leverages the expertise of all stakeholders, enables them to come to a common understanding
- High Level Output



Transformation to Enterprise DevOps culture

Establish a DevOps Center of Excellence

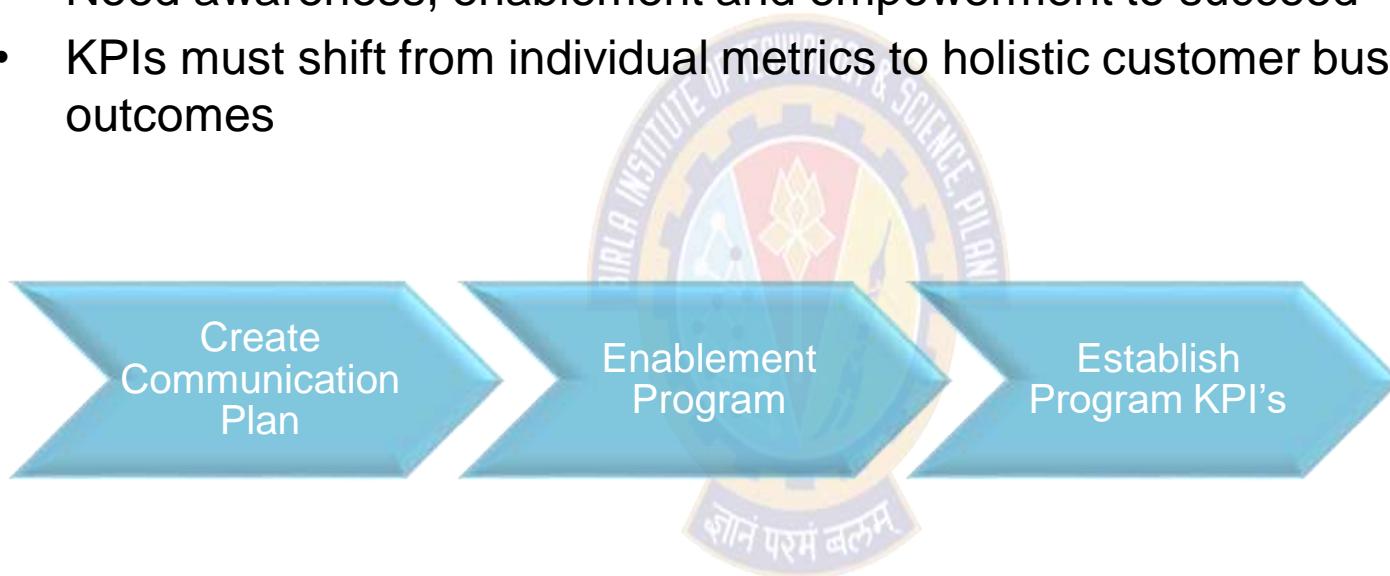
- At the right organizational level and enterprise authority
- Create a Face: It has to be led by an enterprise leader who has the support and buy-in from all the towers
- Active Participant from All Towers of Delivery & Participants must be chosen wisely
- These phases help concrete the vision and strategies of organization and help in setup the best practices to support cultural movement



Transformation to Enterprise DevOps culture

Establish Program Governance

- Agile and DevOps, practitioners' roles and responsibilities will change
- Need awareness, enablement and empowerment to succeed
- KPIs must shift from individual metrics to holistic customer business outcomes



Transformation to Enterprise DevOps culture

Establish Project In-take Process

- DevOps SME's conduct in-take workshops for Scrum Teams
- Automation scripts, Infrastructure assets
- Test Automation, Branching and Merging & Lessons Learned
- Fit for purpose tool selection and Application Criticality



Reusable Assets

Best Practices

Right Sizing

Transformation to Enterprise DevOps culture

Identify and Initiate Pilots

- Value stream-mapping exercise
- Level of detail necessary:
 - To identify end to end as-is process, tooling, manual and automated processes
 - And skills and people



Transformation to Enterprise DevOps culture

Scale Out DevOps Program

- Onboard Parallel Release Trains
- Apply Intake Process
- Support, Monitor and Manage



References

- I would like to thank you Mr. Sunil Joshi
- <https://devops.com/six-step-approach-enterprise-devops-transformation/>





Thank You!

In our next session:

Pros of Waterfall

Cons of Waterfall



Agile

- *individuals and interactions over processes and tools*
- *working software over comprehensive documentation*
- *customer collaboration over contract negotiation*
- *responding to change over following a plan*



Scrum

- **Scrum**
- In the mid-1990s, Ken Schwaber and Dr. Jeff Sutherland, two of the original creators of the Agile Manifesto, merged individual efforts to present a new software development process called Scrum. Scrum is a software development methodology that focuses on maximizing a development team's ability to quickly respond to changes in both project and customer requirements. It uses predefined development cycles called *sprints*, usually between one week and one month long, beginning with a sprint planning meeting to define goals and ending with a sprint review and sprint retrospective to discuss progress and any issues that arose during that sprint.

DSM

- What did I do yesterday that helped the team meet its sprint goals?
- What am I planning to do today to help the team meet those goals?
- What, if anything, do I see that is blocking either me or the team from reaching their goals?

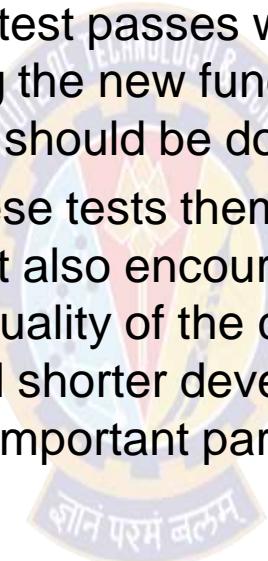


Version Control

- A version control system records changes to files or sets of files stored within the system. This can be source code, assets, and other documents that may be part of a software development project. Developers make changes in groups called *commits* or *revisions*. Each revision, along with metadata such as who made the change and when, is stored within the system in one way or another.
- Having the ability to commit, compare, merge, and restore past revisions to objects to the repository allows for richer cooperation and collaboration within and between teams. It minimizes risks by establishing a way to revert objects in production to previous versions.

Test Driven Development

- In test-driven development, the code developer starts by writing a failing test for the new code functionality, then writes the code itself, and finally ensures that the test passes when the code is complete. The test is a way of defining the new functionality clearly, making more explicit what the code should be doing.
- Having developers write these tests themselves not only greatly shortens feedback loops but also encourages developers to take more responsibility for the quality of the code they are writing. This sharing of responsibility and shorter development cycle time are themes that continue to be important parts of a devops culture.

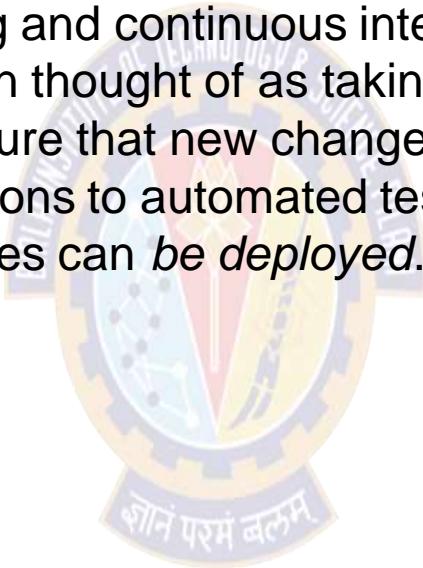


Continuous Integration

- Continuous integration (CI) is the process of integrating new code written by developers with a mainline or “master” branch frequently throughout the day. This is in contrast to having developers working on independent feature branches for weeks or months at a time, merging their code back to the master branch only when it is completely finished. Long periods of time in between merges means that much more has been changed, increasing the likelihood of some of those changes being breaking ones. With bigger changesets, it is much more difficult to isolate and identify what caused something to break. With small, frequently merged changesets, finding the specific change that caused a regression is much easier. The goal is to avoid the kinds of integration problems that come from large, infrequent merges.

Continuous Delivery

- Continuous delivery (CD) is a set of general software engineering principles that allow for frequent releases of new software through the use of automated testing and continuous integration. It is closely related to CI, and is often thought of as taking CI one step further, that beyond simply making sure that new changes can be integrated without causing regressions to automated tests, continuous delivery means that these changes can *be deployed*.

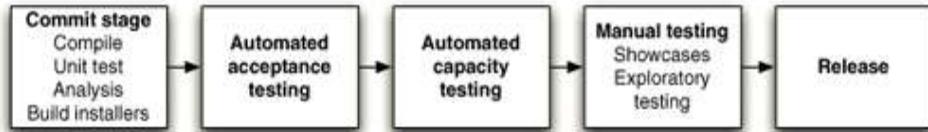


Continuous Deployment

- Continuous deployment (also referred to as CD) is the process of deploying changes to production by defining tests and validations to minimize risk. While continuous delivery makes sure that new changes can be deployed, continuous deployment means that they *get deployed* into production.



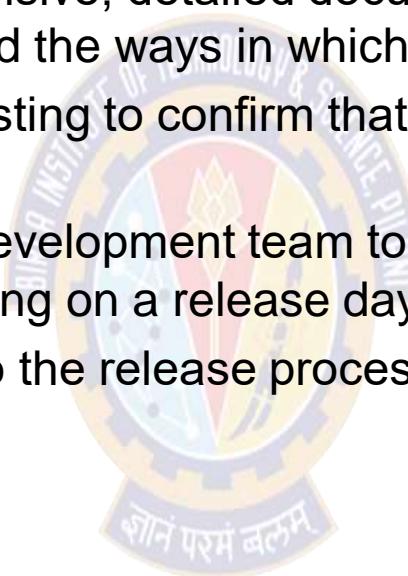
Deployment Pipeline



- The pattern that is central to this book is the *deployment pipeline*. A deployment pipeline is, in essence, an automated implementation of your application's build, deploy, test, and release process. Every organization will have differences in the implementation of their deployment pipelines, depending on their value stream for releasing software, but the principles that govern them do not vary. An example of a deployment pipeline is given in Figure

Antipattern: Deploying Software Manually

- The signs of this antipattern are:
 - The production of extensive, detailed documentation that describes the steps to be taken and the ways in which the steps may go wrong
 - Reliance on manual testing to confirm that the application is running correctly
 - Frequent calls to the development team to explain why a deployment is going wrong on a release day
 - Frequent corrections to the release process during the course of a release



Antipattern: Deploying to a Production-like Environment Only after Development Is Complete

- If testers have been involved in the process up to this point, they have tested the system on development machines.
- Releasing into staging is the first time that operations people interact with the new release. In some organizations, separate operations teams are used to deploy the software into staging and production. In this case, the first time an operations person sees the software is the day it is released into production.
- Either a production-like environment is expensive enough that access to it is strictly controlled, or it is not in place on time, or nobody bothered to create one.
- The development team assembles the correct installers, configuration files, database migrations, and deployment documentation to pass to the people who perform the actual deployment—all of it untested in an environment that looks like production or staging.

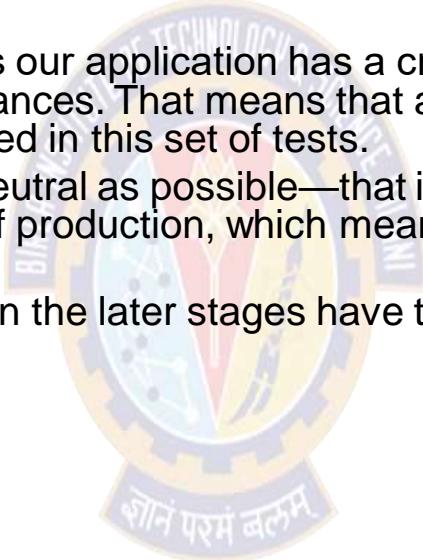
Every Change Should Trigger the Feedback Process

- Executable code changes when a change is made to the source code. Every time a change is made to the source code, the resulting binary must be built and tested. In order to gain control over this process, building and testing the binary should be automated. The practice of building and testing your application on every check-in is known as continuous integration



Possible

- They run fast.
- They are as comprehensive as possible—that is to say, they cover more than 75% or so of the codebase, so that when they pass, we have a good level of confidence that the application works.
- If any of them fails, it means our application has a critical fault and should not be released under any circumstances. That means that a test to check the color of a UI element should not be included in this set of tests.
- They are as environment-neutral as possible—that is, the environment does not have to be an exact replica of production, which means it can be simpler and cheaper.
- On the other hand, the tests in the later stages have the following general characteristics.





Thank You!

Acknowledgement:

- Significant portions of the information in the presentation is from IT Systems Management - Rich Schiesser and other books/Internet. Permission was requested from the publisher, sometime ago for use, but has not been received as yet. This is intended solely to teach BITS students enrolled for IM. I would like to acknowledge and reiterate that all the credit/rights remain with the original authors/publishers only. Mistakes if any are mine.



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps





CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 4

Agenda

Cloud as a catalyst for DevOps

- Introduction
- Characterization of the cloud
- Cloud services



Cloud as a Catalyst for DevOps

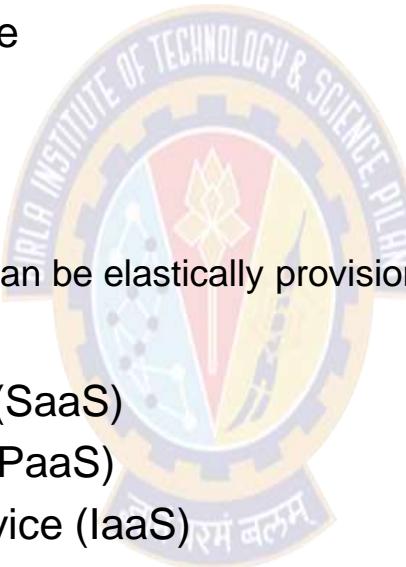
Introduction

- Unfortunately for many organizations, the cloud is still somewhat mysterious
- Not the concept of the cloud but why they are not seeing the benefits that were promised to them
- The focus is on adopting cloud-based infrastructure and managing it
- While we will discuss cost benefits primarily, one should not forget the other benefits and risks of the cloud
- Benefits:
 - Redundancy for resilience
 - The scalability of resources
 - The strong ecosystem of related
- Risks:
 - The dependency on a third-party provider
 - The challenges with data sovereignty
 - Risk of being attacked because you are on a popular platform

Cloud as a Catalyst for DevOps

Characterization of the cloud

- Characterization of the cloud by National Institute of Standards and Technology (NIST)
 - On-demand self-service
 - Broad network access
 - Resource pooling
 - Rapid elasticity
 - It is the Capabilities can be elastically provisioned and released
- Measured service
 - Software as a Service (SaaS)
 - Platform as a Service (PaaS)
 - Infrastructure as a Service (IaaS)



Cloud as a Catalyst for DevOps

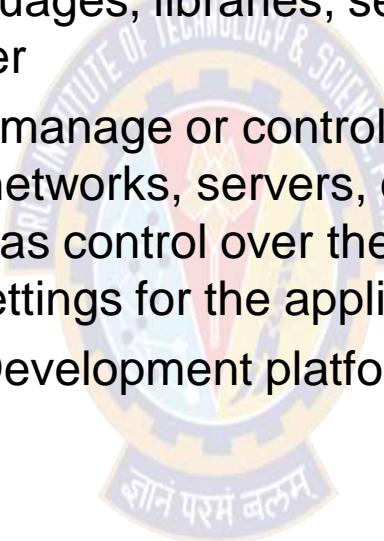
Software as a Service (SaaS)

- In this the consumer is provided the capability to use the provider's applications running on a cloud infrastructure
- The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based e-mail) or an application interface.
- The consumer does not manage or control the underlying cloud infrastructure including networks, servers, operating systems, storage.
- For an example, you can relate google apps, Cisco WebEx, as a Service, Office 365 service, where Provider deals with the licensing of software's

Cloud as a Catalyst for DevOps

Platform as a Service (PaaS)

- The consumer is provided the capability to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider
- The consumer does not manage or control the underlying cloud infrastructure including networks, servers, operating systems, or storage, but consumer has control over the deployed applications and possibly configuration settings for the application-hosting environment
- For an Example: .NET Development platform is considered as a platform



Cloud as a Catalyst for DevOps

Infrastructure as a Service (IaaS),

- The consumer is provided the capability to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications
- The consumer does not manage or control the underlying cloud infrastructure but consumer has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls). For this you can consider any Server Provisioning is IaaS,

Cloud as a Catalyst for DevOps

Example

- Think of a shopping site that provides you with personalized recommendations based on your previous purchases
- If that service does not work for some reason, rather than not showing you the site or delaying the response for your request, ecommerce site might choose to show you some static recommendation
- For you as a user, the experience is slightly worse (although many people wouldn't notice), but it is still much better than getting a page error or time-out
- Leveraging this idea of graceful degradation works in many contexts but not all, and it will usually be a cheaper alternative to keeping the whole end-to-end availability at very high availability

Cloud as a Catalyst for DevOps

Summary

- The cloud has emerged as a major trend in IT during recent years
- Its characteristics include metered usage (pay-per-use) and rapid elasticity, allowing the scaling out of an application to virtually infinite numbers of VMs
- The cloud rests on a platform that is inherently distributed and exploits virtualization to allow rapid expansion and contraction of the resources available to a given user
- From an operational perspective, controlling the VMs, managing different database management systems, and ensuring the environments meet the needs of the development and operations tasks are new considerations associated with the cloud

Agenda

Evolution of Version Control

- What is Version Control
- Version Control System
- Benefits of Version Control System



Evolution of Version Control

What is Version Control

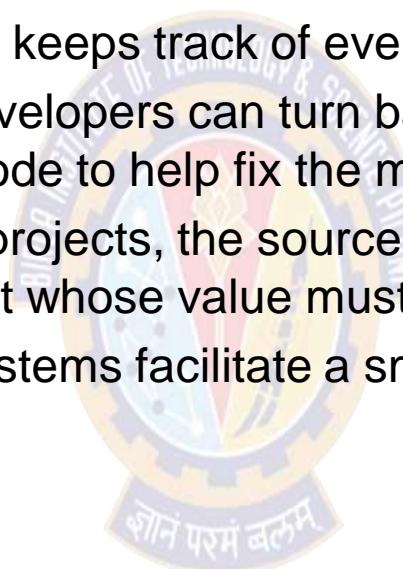
- What is “version control”, and why should you care?
- Definition: Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later
- The history of version control tools can be divided into three generations

Generation	Networking	Operations	Concurrency	Example Tool
First Generation	None	One file at a time	Locks	RCS, SCCS
Second Generation	Centralized	Multi-file	Merge before commit	CVS, Subversion
Third Generation	Distributed	Changesets	Commit before merge	Bazaar, Git

Evolution of Version Control

Version Control System

- A category of software tools that help a software team manage changes to source code over time
- Version control software keeps track of every modification to the code
- If a mistake is made, developers can turn back the clock and compare earlier versions of the code to help fix the mistake
- For almost all software projects, the source code is like the crown jewels - a precious asset whose value must be protected
- Great version control systems facilitate a smooth and continuous flow of changes to the code



Evolution of Version Control

Benefits of Version Control System (VCS)

- A complete long-term change history of every file
- Restoring previous versions
- Branching and merging; having team members work concurrently
- Traceability; being able to trace each change made to the software
- Backup



Evolution of Version Control

Summary

- The vast majority of professional programmers are using second generation tools
- However the third generation is growing very quickly in popularity
- The most popular Version Control System is Apache Subversion
- Commonly known as SVN:
 - Subversion is an open source version control system
 - Founded in 2000 by CollabNet, Inc.
 - Subversion is developed as a project of the Apache Software Foundation
 - It's a centralized version control system
 - Current Stable release is “Apache Subversion 1.11.1” as on date 11-Jan-2019

Agenda

Version control system and its types

- Centralized Version Control System
- Distributed Version Control System
- Version Control System Basics
- Version Control System Common Terminology
- Version Control System Basic Operations



Version Control System

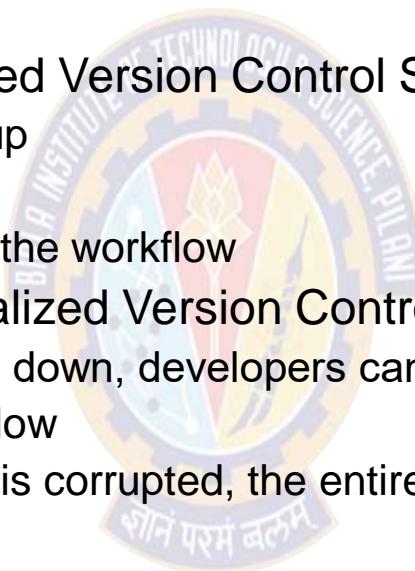
Version Control System Types

- Centralized Version Control System:[CVCS]
 - With centralized version control systems, you have a single “central” copy of your project on a server and commit your changes to this central copy
 - You pull the files that you need, but you never have a full copy of your project locally
 - Some of the most common version control systems are:
 - Subversion (SVN) by Apache
 - Perforce by Perforce Software
- Distributed Version Control System: [DVCS]
 - With distributed version control systems (DVCS), you don't rely on a central server to store all the versions of a project's files
 - Instead, you clone a copy of a repository locally so that you have the full history of the project
 - Some of most common distributed version control systems are:
 - Git
 - and Mercurial

Version Control System

Centralized Version Control System:[CVCS]

- Centralized Version Control Systems were developed to record changes in a central system and enable developers to collaborate on other systems
- Advantages of Centralized Version Control Systems:
 - Relatively easy to set up
 - Provides transparency
 - Enable admins control the workflow
- Disadvantages of Centralized Version Control Systems:
 - If the main server goes down, developers can't save versioned changes
 - Remote commits are slow
 - If the central database is corrupted, the entire history could be lost (security issues)



Version Control System

Distributed Version Control System:[DVCS]

- They allow developers to clone the repository and work on that version. Developers will have the entire history of the project on their own hard drives
- Advantages of Distributed Version Control Systems:
 - Performing actions other than pushing and pulling changesets is extremely fast because the tool only needs to access the hard drive, not a remote server
 - Everything but pushing and pulling can be done offline
 - Since each programmer has a full copy of the project repository, they can share changes with one or two other people at a time if they want to get some feedback
- Disadvantages of Centralized Version Control Systems:
 - If your project contains many large files then the space will be more on local drives
 - If your project has a very long history then downloading the entire history can take an impractical amount of time

Version Control System

Version Control System Available:[Centralized]

- Open source:
 - Subversion (SVN) – versioning control system inspired by CVS
 - Concurrent Versions System (CVS) – originally built on RCS, licensed under the GPL
 - Vesta – build system with a versioning file system and support for distributed repositories
 - OpenCVS – CVS clone under the BSD license, with emphasis put on security and source code correctness
- Commercial:
 - AccuRev – source configuration management tool with integrated issue tracking based on "Streams" that efficiently manages parallel and global development. Now owned by Micro Focus
 - Helix Core, formerly Perforce Helix - for large scale development environments
 - IBM Rational ClearCase – SCC compliant configuration management system by IBM Rational Software
 - Team Foundation Server (TFS) - Development software by Microsoft which includes revision control

Version Control System

Version Control System Available:[Distributed]

- Open source:
 - Git – written in a collection of Perl, C, and various shell scripts, designed by Linus Torvalds based on the needs of the Linux kernel project; decentralized, and aims to be fast, flexible, and robust
 - Bazaar – written in Python, originally by Martin Pool and sponsored by Canonical; decentralized, and aims to be fast and easy to use; can losslessly import Arch archives
 - Mercurial – written in Python as an Open Source replacement to BitKeeper; decentralized and aims to be fast, lightweight, portable, and easy to use
- Commercial:
 - Visual Studio Team Services - Services for teams to share code, track work, and ship software for any language by Microsoft
 - Sun WorkShop TeamWare – designed by Larry McVoy, creator of BitKeeper
 - Plastic SCM – by Codice Software, Inc
 - Code Co-op – peer-to-peer version control system (can use e-mail for synchronization)

Version Control System Basics

What do you want from your version control system?

- Store all versions of your files ("version control")
- Associate versions of each file with appropriate versions of all other files ("configuration management")
- Allow many people to work on the same files, toward a common goal or release ("concurrency")
- Allow groups of people to work on substantially the same files, but each group towards its own goal or release ("branching")
- Recover, at any time, a coherent configuration of file versions that correspond to some goal or release, either for investigation or extension like bug fixing ("release management")

Version Control System Basics

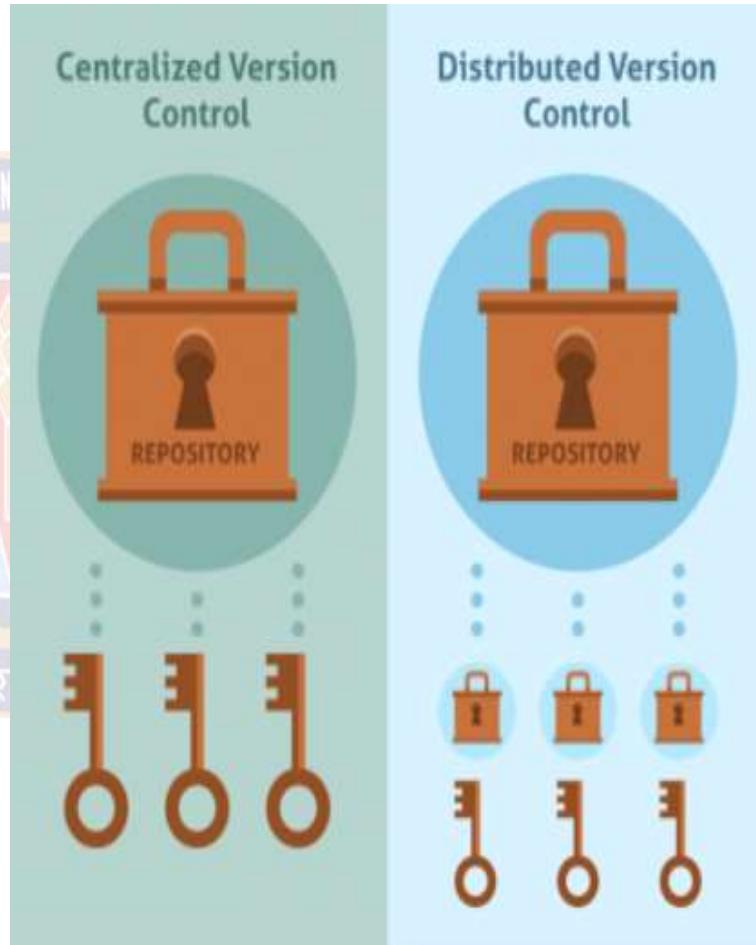
Version Control System Terminology

- Baseline : An approved revision of a document or source file from which subsequent changes can be made
- Branch : A set of files under version control may be branched or forked at a point in time so that, from that time forward, two copies of those files may develop at different speeds or in different ways independently of each other
- Change : A change represents a specific modification to a document under version control
- Checkout : To check out is to create a local working copy from the repository. A user may specify a specific revision or obtain the latest
- Clone : Cloning means creating a repository containing the revisions from another repository
- Commit : To commit is to write or merge the changes made in the working copy back to the repository
- Conflict : A conflict occurs when different parties make changes to the same document, and the system is unable to reconcile the changes
- Fetch : Fetch is initiated by the receiving repository
- Initialize : To create a new, empty repository
- Merge : A merge or integration is an operation in which two sets of changes are applied to a file or set of files
- Repository : The repository is where files' current and historical data are stored, often on a server
- Tag : A tag or label refers to an important snapshot in time, consistent across many files. These files at that point may all be tagged with a user-friendly, meaningful name or revision number

Version Control System Basics

Summary to Understand:

- Centralized version control systems are based on the idea that there is a single “central” copy of your project somewhere (probably on a server), and programmers will “commit” their changes to this central copy
- “Committing” a change simply means recording the change in the central system
- Distributed Version control systems do not necessarily rely on a central server to store all the versions of a project’s files. Instead, every developer “clones” a copy of a repository and has the full history of the project on their own hard drive
- Then you collect all copies and prepare a final copy



Version Control System Basics

Centralized VCS over Distributed VCS

- Advantages of Centralized VCS over Distributed VCS :
 - If your project contains many large, binary files that cannot be easily compressed, the space needed to store all versions of these files can accumulate quickly
 - If your project has a very long history (50,000 changesets or more), downloading the entire history can take an impractical amount of time and disk space
- Advantages of Distributed VCS over Centralized VCS :
 - Performing actions other than pushing and pulling changesets is extremely fast because the tool only needs to access the hard drive, not a remote server
 - Committing new changesets can be done locally without anyone else seeing them. Once you have a group of changesets ready, you can push all of them at once
 - Since each programmer has a full copy of the project repository, they can share changes with one or two other people at a time if they want to get some feedback before showing the changes to everyone
- Misconceptions: Distributed systems do not prevent you from having a single “central” repository, they just provide more options on top of that

Version Control System Basic Operations

Create

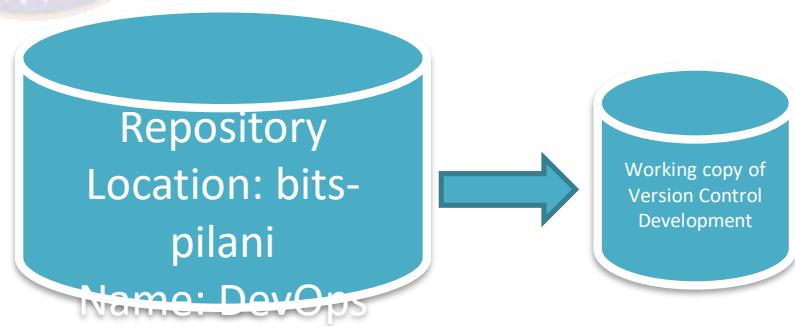
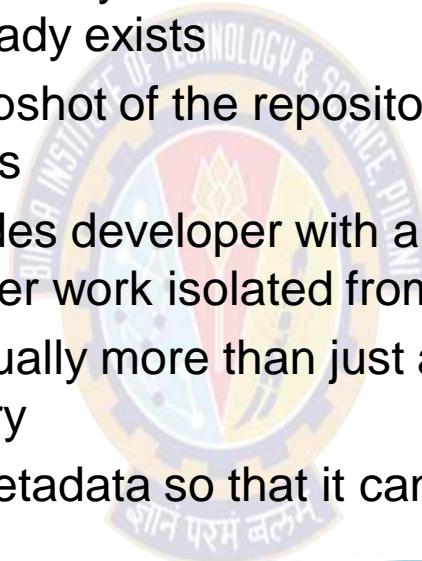
- Create a new, empty repository
- A repository is the official place where you store all your work.
- It keeps track of your tree i.e. all your files, as well as the layout of the directories in which they are stored
- A version control repository would be a ~~network~~ filesystem
- A repository is much more than that. A repository contains history
 - Repository = Filesystem * Time
- A filesystem is two-dimensional: Its space is defined by directories and files
- Whereas a repository is three-dimensional: It exists in a continuum defined by directories, files, and time
- What you need to create a Repository?
 - Location [Where you want to create it]
 - Name



Version Control System Basic Operations

Checkout

- Create a working copy
- This operation is used when you need to make a new working copy for a repository that already exists
- A working copy is a snapshot of the repository used by a developer as a place to make changes
- The working copy provides developer with a private workspace where developer can do his / her work isolated from the rest of the team
- The working copy is actually more than just a snapshot of the contents of the repository
- It also contains some metadata so that it can keep careful track of the state of things



Version Control System Basic Operations

Commit

- Apply the modifications in the working copy to the repository as a new changeset
- This is the operation that actually modifies the repository
- Several others modify the working copy and add an operation to a list we call the pending changeset, a place where changes wait to be committed
- The commit operation takes the pending changeset and uses it to create a new version of the tree in the repository
- It is typical to provide a log message (or comment) when you commit, explaining the changes you have made. This log message becomes part of the history of the repository



Version Control System Basic Operations

Update

- Update the working copy with respect to the repository
- Update brings your working copy up-to-date by applying changes from the repository, merging them with any changes you have made to your working copy if necessary
- When the working copy was first created, its contents exactly reflected a specific revision of the repository
- Update is sort of like the mirror image of commit
- Both operations move changes between the working copy and the repository
- Commit goes from the working copy to the repository; Update goes in the other direction



Version Control System Basic Operations

Add

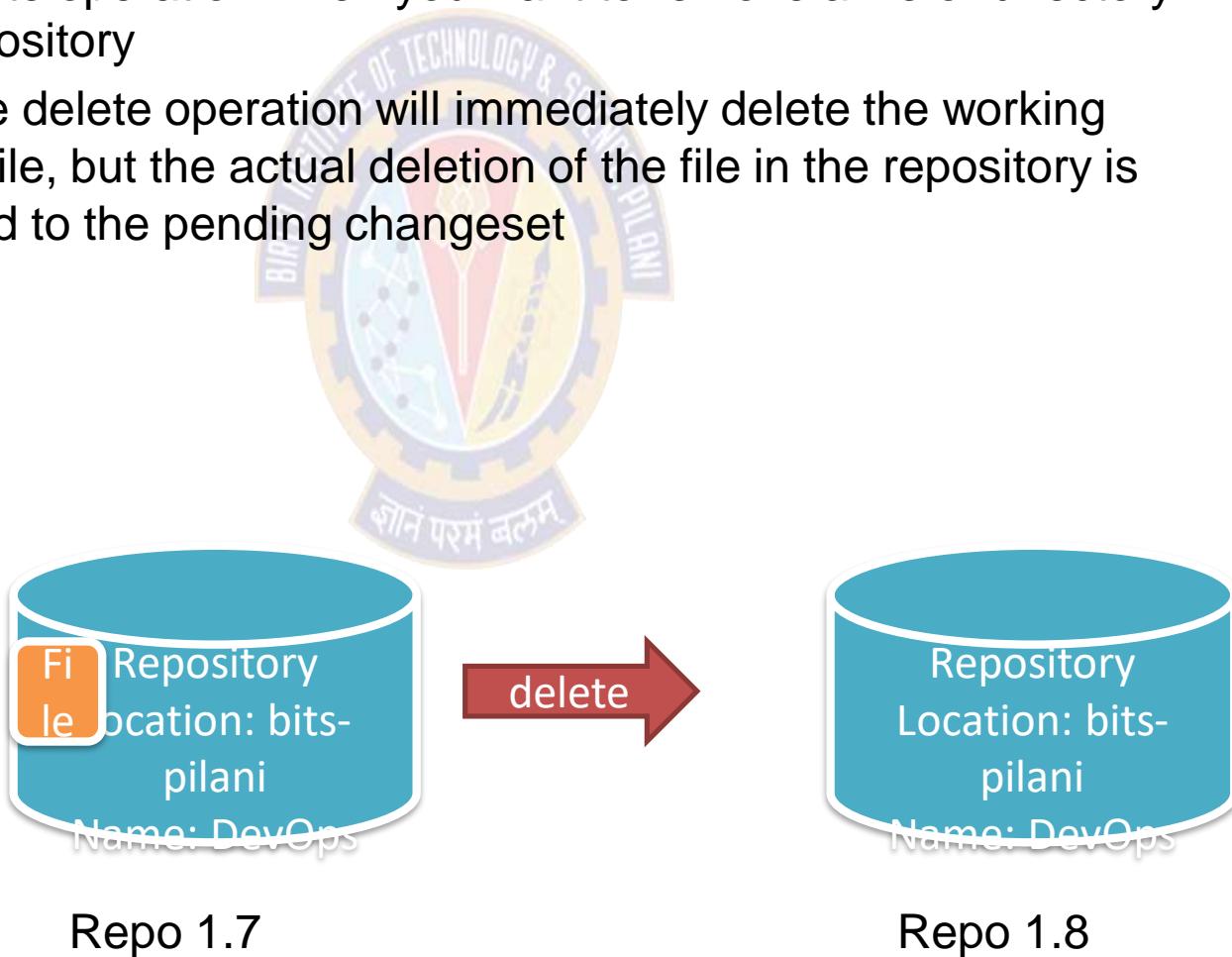
- Add a file or directory
- Use the add operation when you have a file or directory in your working copy that is not yet under version control and you want to add it to the repository



Version Control System Basic Operations

Delete

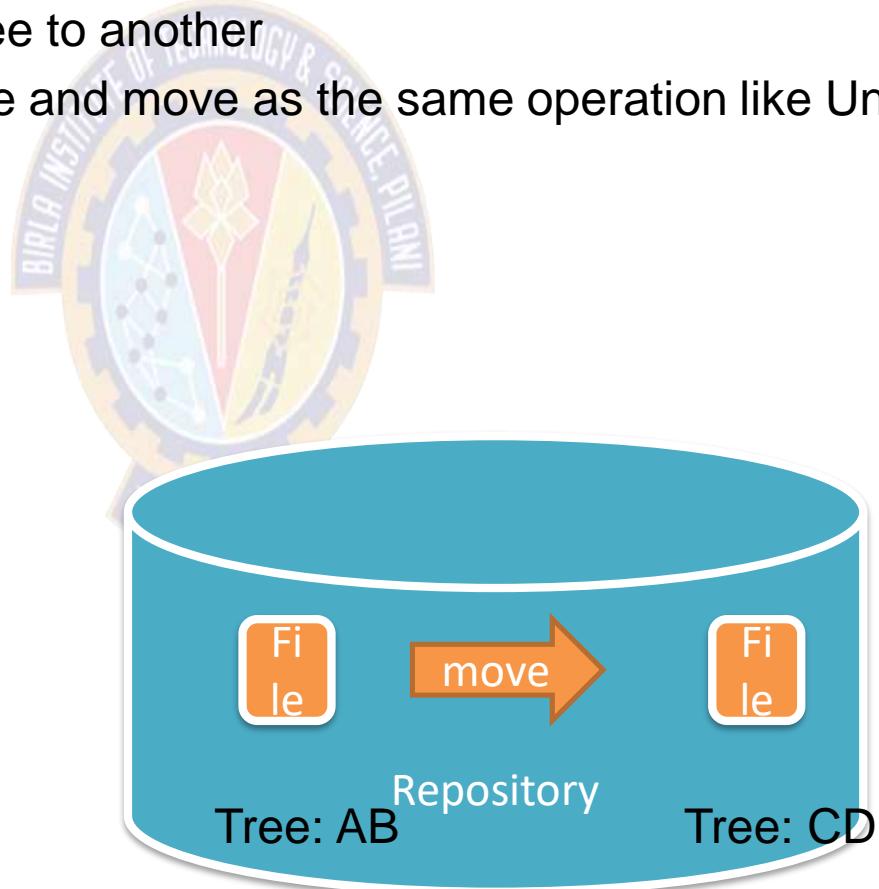
- Delete a file or directory
- Use the delete operation when you want to remove a file or directory from the repository
- Typically, the delete operation will immediately delete the working copy of the file, but the actual deletion of the file in the repository is simply added to the pending changeset



Version Control System Basic Operations

Move

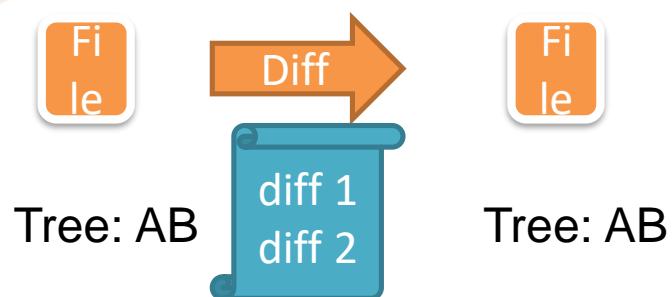
- Move a file or directory
- Use the move operation when you want to move a file or directory from one place in the tree to another
- Some tools treat rename and move as the same operation like Unix



Version Control System Basic Operations

Diff

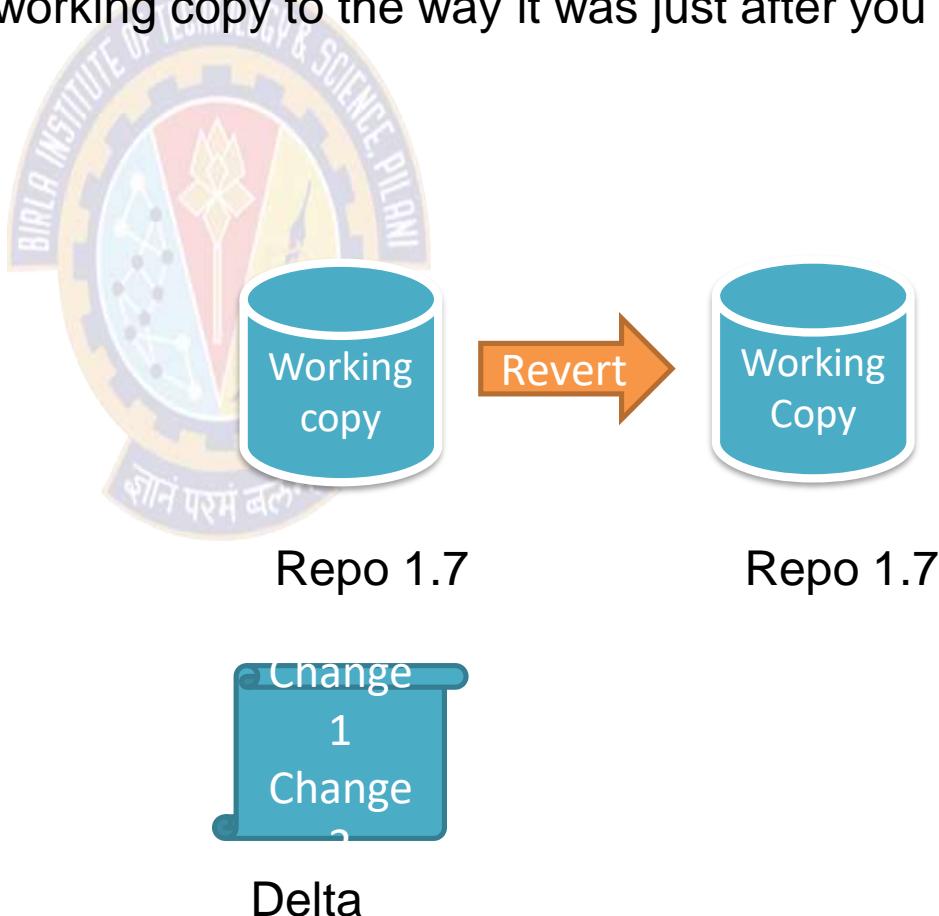
- Show the details of the modifications that have been made to the working copy
- In other hand Status provides a list of changes but no details about them
- To see exactly what changes have been made to the files, you need to use the diff operation



Version Control System Basic Operations

Revert

- Undo modifications that have been made to the working copy
- A complete revert of the working copy will throw away all your pending changes and return the working copy to the way it was just after you did the checkout



Version Control System Basic Operations

Log

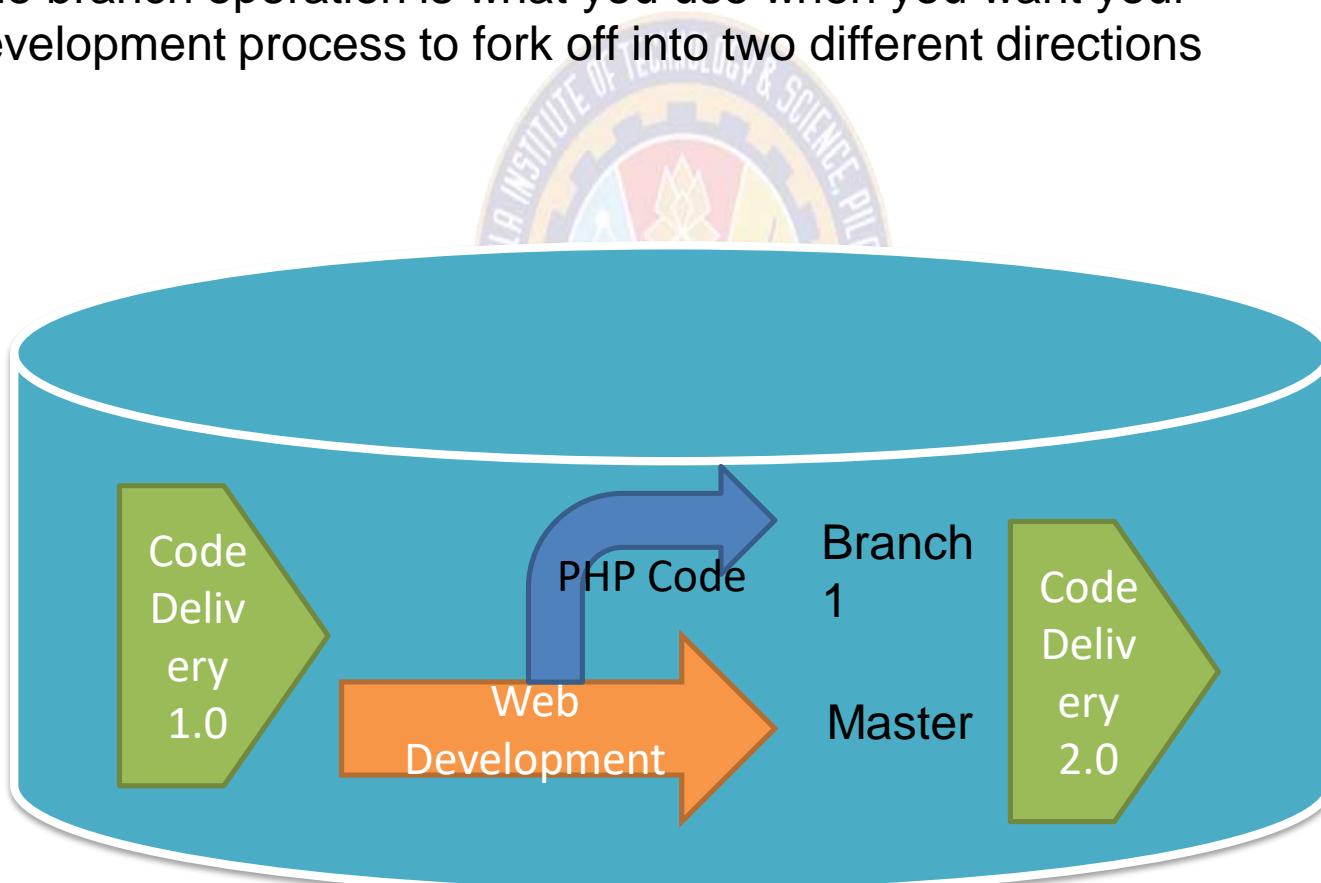
- Show the history of changes to the repository
- Your repository keeps track of every version that has ever existed
- The log operation is the way to see those records
- It displays each changeset along with additional data:
 - Who made the change?
 - When was the change made?
 - What was the log message?

- 
- Repo Created – 1 Jan 2019 – user Rathi – For Devops @ bits
 - Add File 1 – 10 Jan 2019 – user 1 – For Devops Source Code
 - Add File 2 – 12 Jan 2019 – user 1 – For Devops Source Code
 - Rename File 1 – 12 Jan 2019 – user A – For Java Ammend Code
 - Created Working Copy – 15 Jan 2019 – User B
 - Deleted File 2 – 5 Feb 2019 – User 1 – not need
 - Moved File 1 – 5 Feb 2019 – User Rathi - rename

Version Control System Basic Operations

Branch

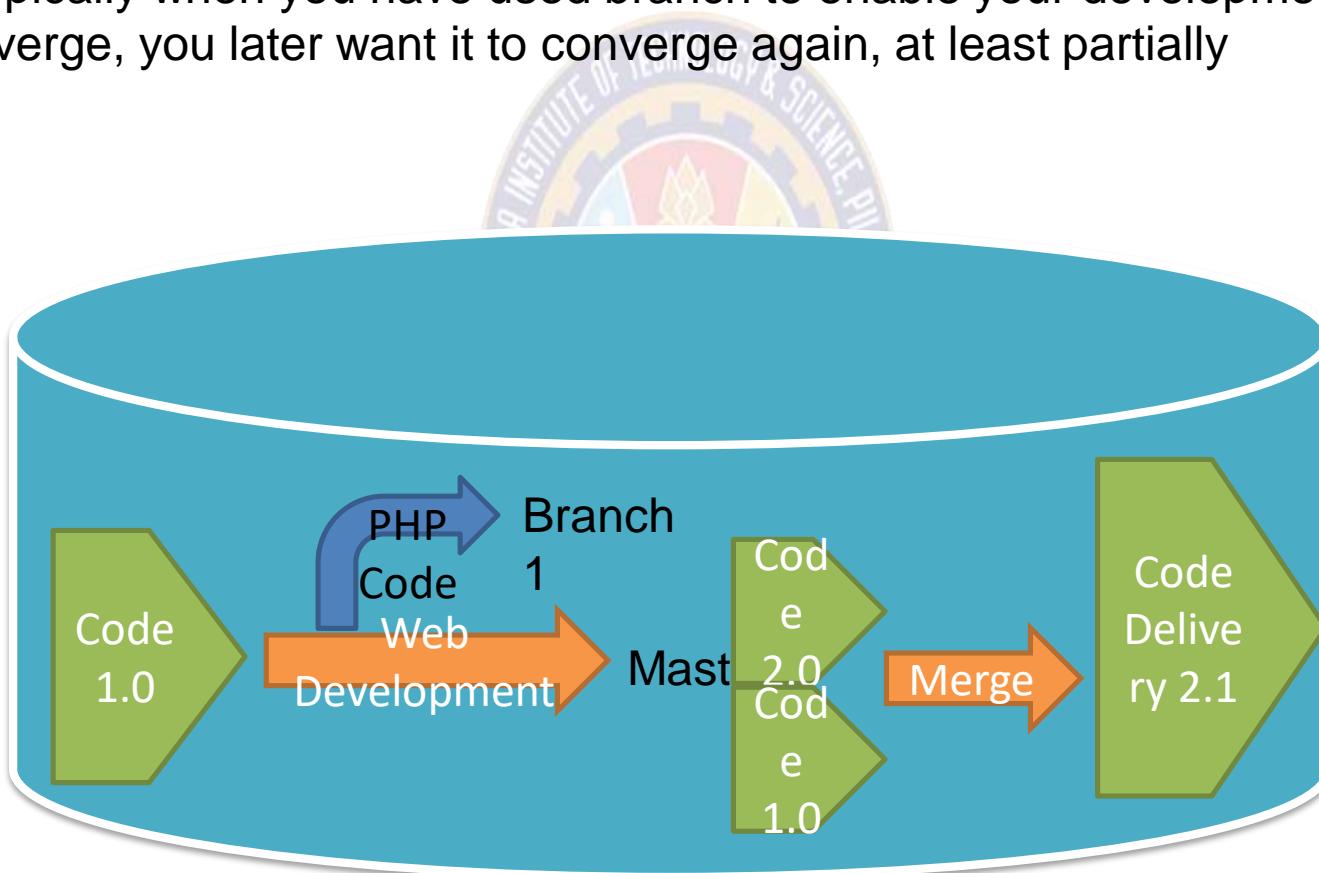
- Create another line of development
- The branch operation is what you use when you want your development process to fork off into two different directions



Version Control System Basic Operations

Merge

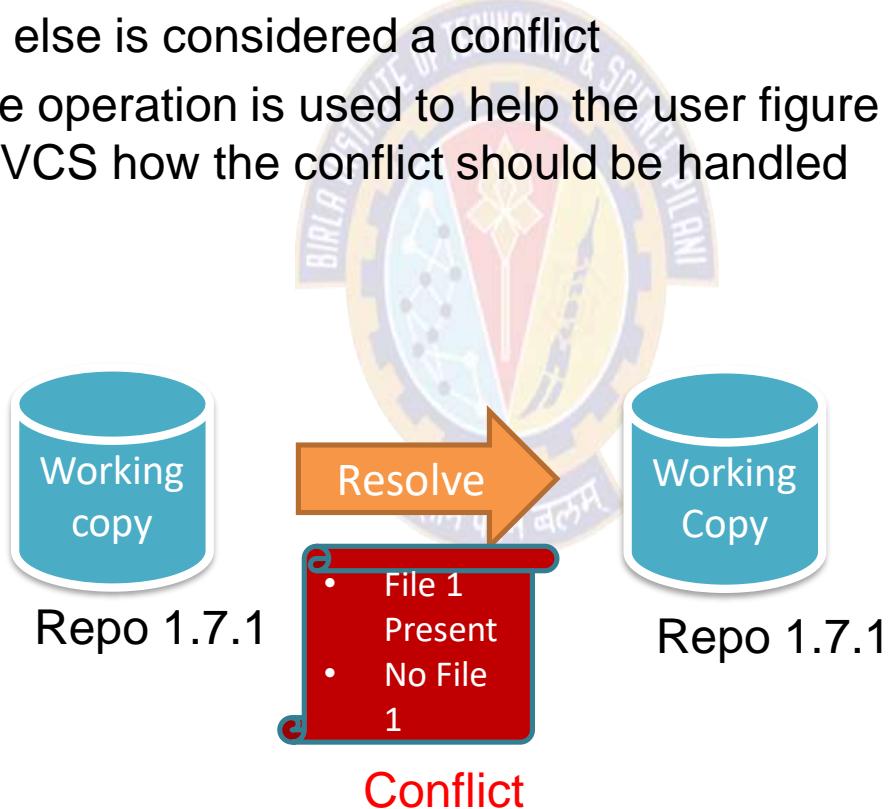
- Apply changes from one branch to another
- Typically when you have used branch to enable your development to diverge, you later want it to converge again, at least partially



Version Control System Basic Operations

Resolve

- Handle conflicts resulting from a merge
- Merge automatically deals with everything that can be done safely
- Everything else is considered a conflict
- The resolve operation is used to help the user figure things out and to inform the VCS how the conflict should be handled



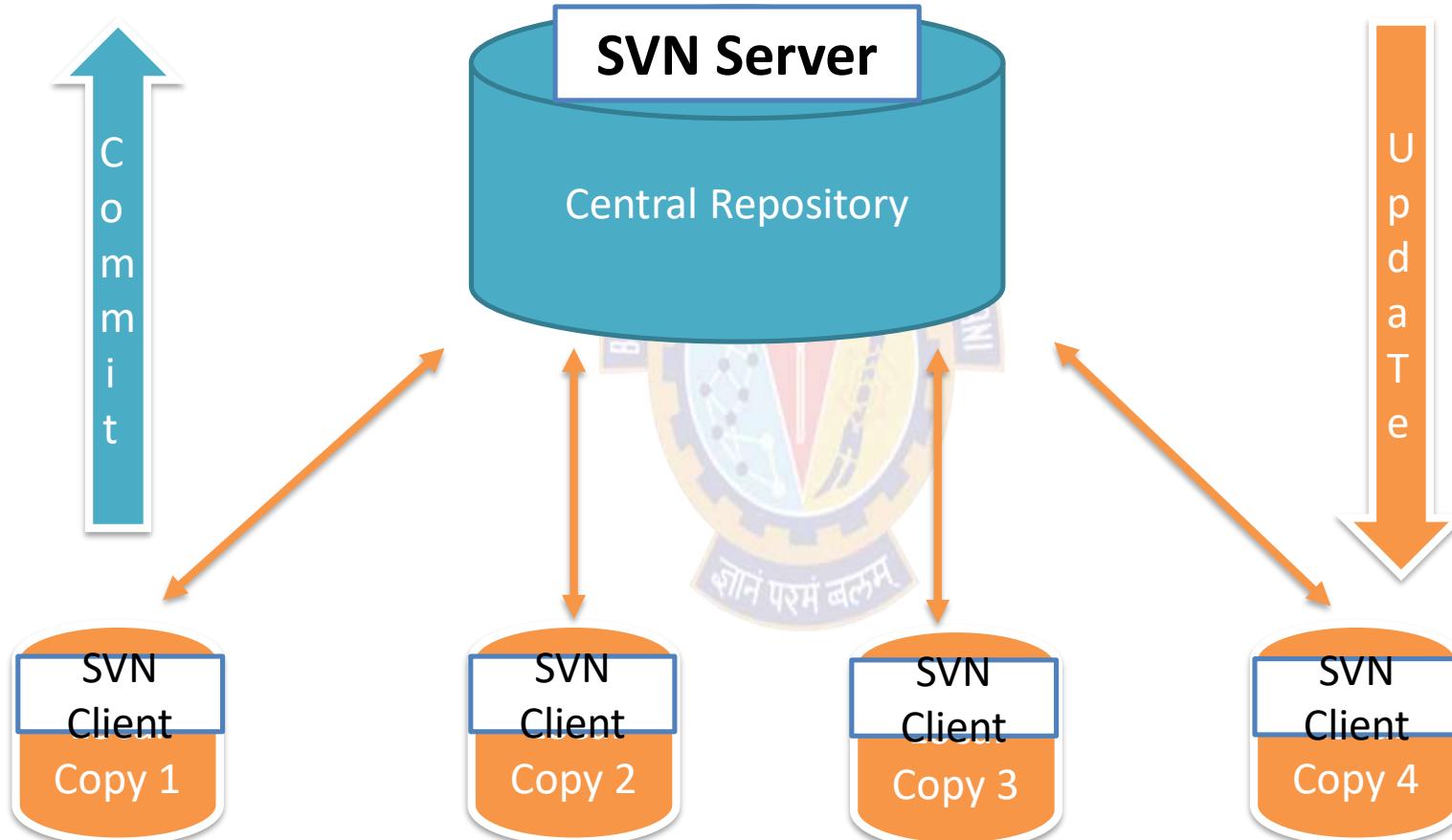
Examples of Centralized Version Control Systems

Apache Subversion (SVN)

- Subversion is a free/open source version control system
- Subversion manages files and directories, and the changes made to them, over time
- This allows you to recover older versions of your data, or examine the history of how your data changed
- Subversion is the most popular and its usage is growing larger
- Its centralized architecture make it easy to maintain a security hierarchy, access control, and backups, which is why it's preferred in many enterprise organizations over Perforce, VSS, and ClearCase
- Subversion is still widely popular amongst the open source community

Examples of Centralized Version Control Systems

Apache Subversion (SVN) Architecture



Examples of Centralized Version Control Systems

Features of Subversion

- CVS is a relatively basic version control system. For the most part, Subversion has matched or exceeded CVS's feature set where those features continue to apply in Subversion's particular design
- Subversion versions directories as first-class objects, just like files
- Copying and deleting are versioned operations. Renaming is also a versioned operation
- Subversion allows arbitrary metadata ("properties") to be attached to any file or directory
 - These properties are key/value pairs, and are versioned just like the objects they are attached to
 - These properties are not versioned, since they attach metadata to the version-space itself, but they can be changed at any time
- No part of a commit takes effect until the entire commit has succeeded
- Revision numbers are per-commit, not per-file, and commit's log message is attached to its revision
- Branches and tags are both implemented in terms of an underlying "copy" operation
- Unix users can place symbolic links under version control
- Subversion supports (but does not require) locking files so that users can be warned when multiple people try to edit the same file
- All output of the Subversion command-line client is carefully designed to be both human readable
- Subversion uses gettext() to display translated error, informational, and help messages, based on current locale settings
- The Subversion command-line client (svn) offers various ways to resolve conflicting changes, include interactive resolution prompting

Agenda

About GIT

- Examples of Distributed Version Control Systems
- Introduction to GIT
- States of GIT
- A GIT Project
- GIT workflow



About GIT

Examples of Distributed Version Control Systems

- Distributed Version Control Systems (DVCSS) don't rely on a central server
- They allow developers to clone the repository and work on that version
- Developers will have the entire history of the project on their own hard drives
- Revise:
 - Because of local commits, the full history is always available
 - No need to access a remote server (faster access)
 - Ability to push your changes continuously
 - Saves time, especially with SSH keys
 - Good for projects with off-shore developers
- Lets Discuss our Distributed Version Control System GIT



Introduction to Git

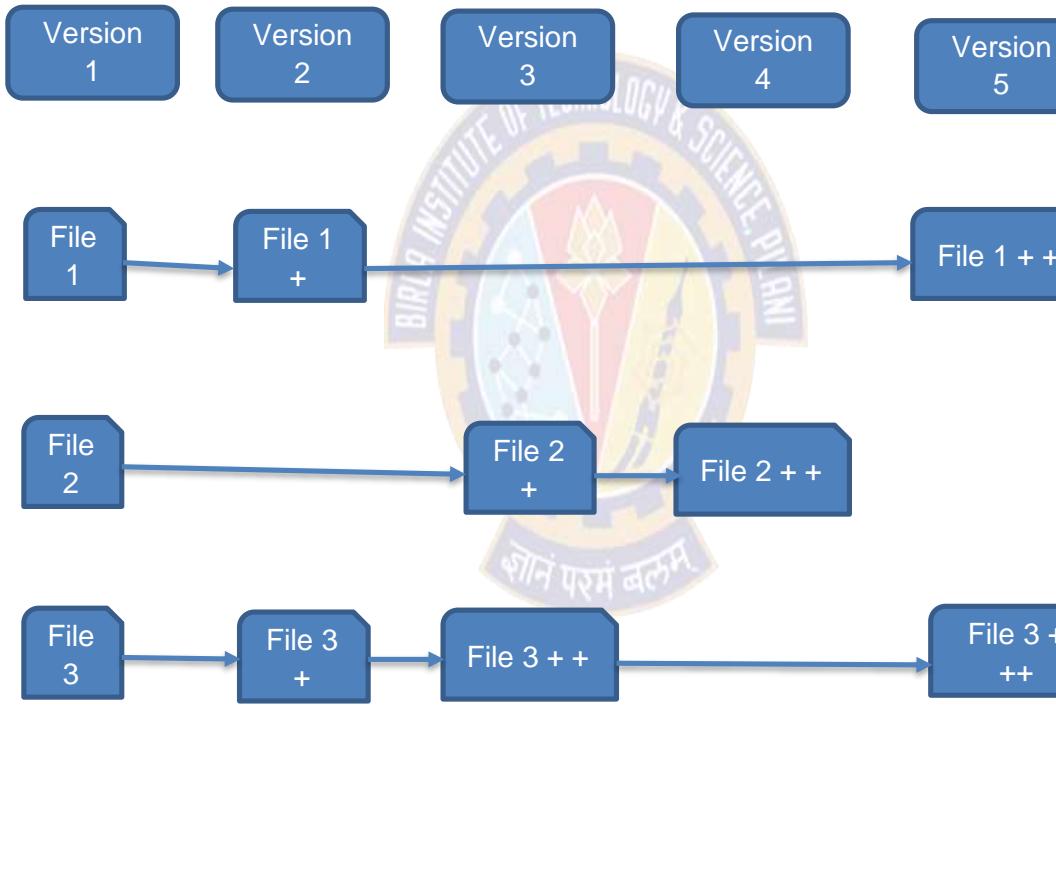
A Short History of Git

- The Linux kernel is an open source software project of fairly large scope
- For most of the lifetime of the Linux kernel maintenance (1991–2002), changes to the software were passed around as patches and archived files
- In 2002, the Linux kernel project began using a proprietary DVCS called BitKeeper
- In 2005, the relationship between the community that developed the Linux kernel and the commercial company that developed BitKeeper broke down
- This prompted the Linus Torvalds the creator of Linux to develop their own tool based on some of the lessons they learned while using BitKeeper
- Since its birth in 2005, Git has evolved and matured to be easy to use
- The qualities considered for the development of Git was:
 - Speed
 - Simple design
 - Strong support for non-linear development (thousands of parallel branches)
 - Fully distributed
 - Able to handle large projects like the Linux kernel efficiently (speed and data size)



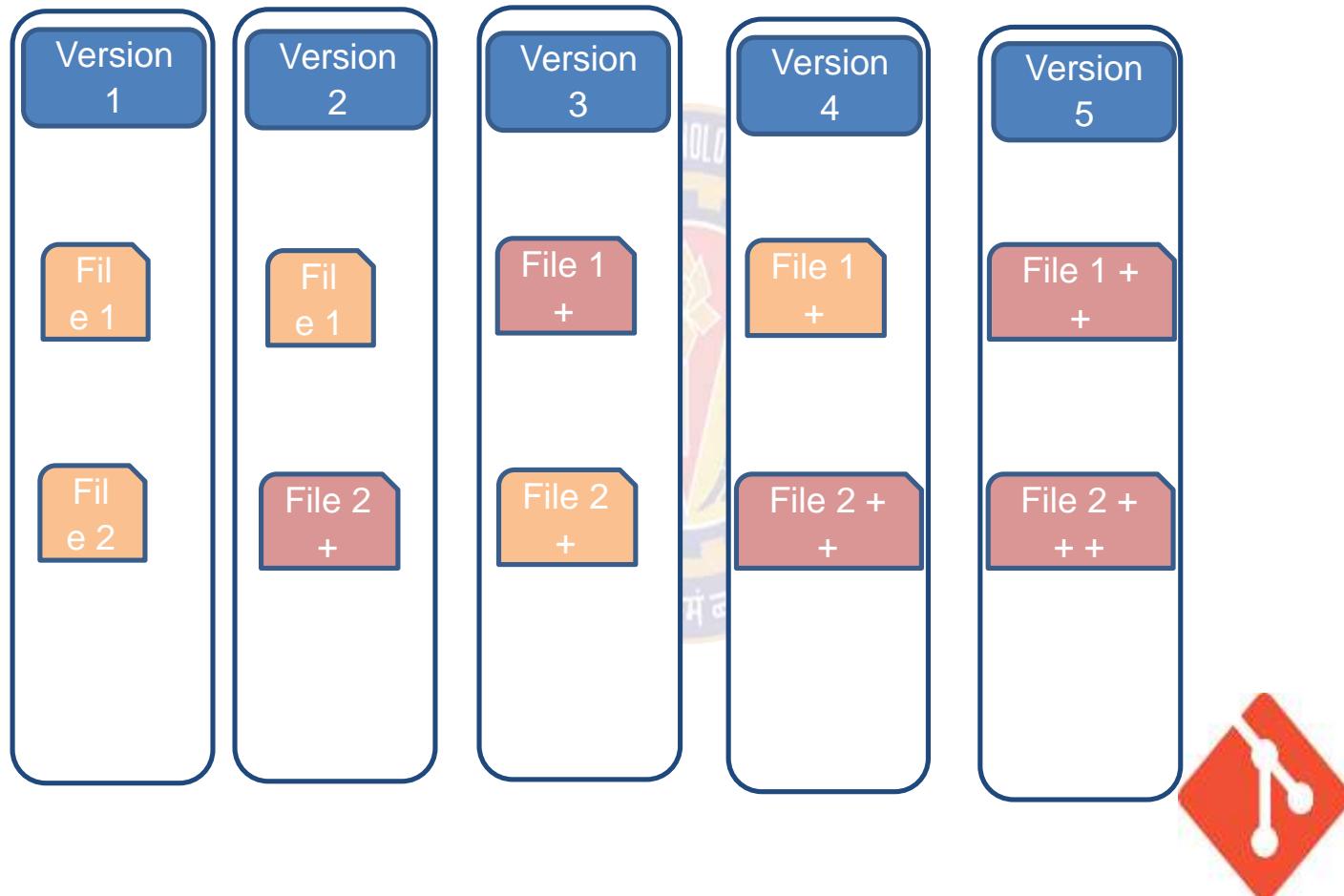
Introduction to Git

What is Different in Git?



Introduction to Git

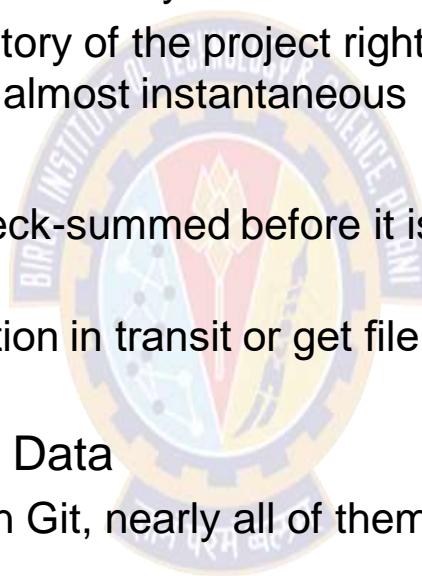
This is How Git treats:



Introduction to Git

How it works?

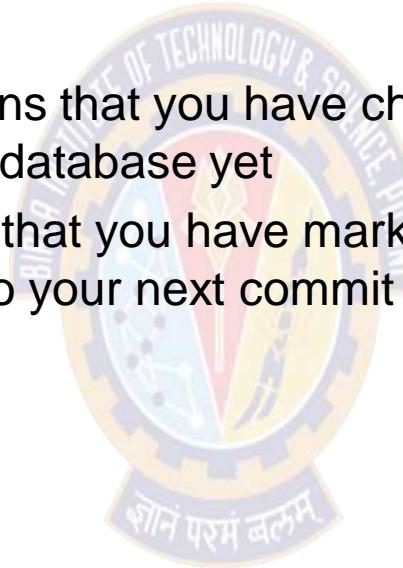
- Nearly Every Operation Is Local:
 - Most operations in Git need only local files and resources to operate
 - you have the entire history of the project right there on your local disk, most operations seem almost instantaneous
- Git Has Integrity
 - Everything in Git is check-summed before it is stored and is then referred to by that checksum
 - You can't lose information in transit or get file corruption without Git being able to detect it.
- Git Generally Only Adds Data
 - When you do actions in Git, nearly all of them only add data to the Git database
 - After you commit a snapshot into Git, it is very difficult to lose, ~~especially~~ if you regularly push your database to another repository



States of Git

The three states of Git

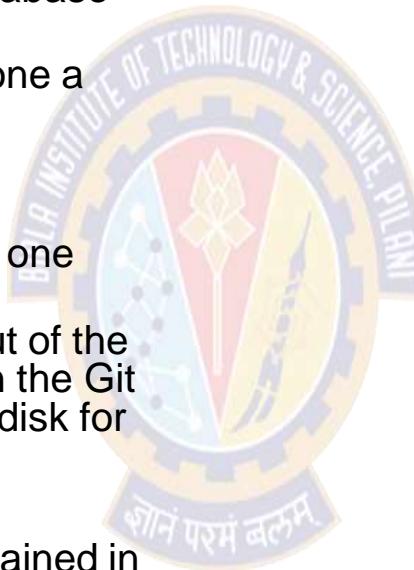
- Git has three main states that your files can reside in:
- Committed : Committed means that the data is safely stored in your local database
- Modified : Modified means that you have changed the file but have not committed it to your database yet
- Staged : Staged means that you have marked a modified file in its current version to go into your next commit snapshot



A Git Project

A Git Project will have

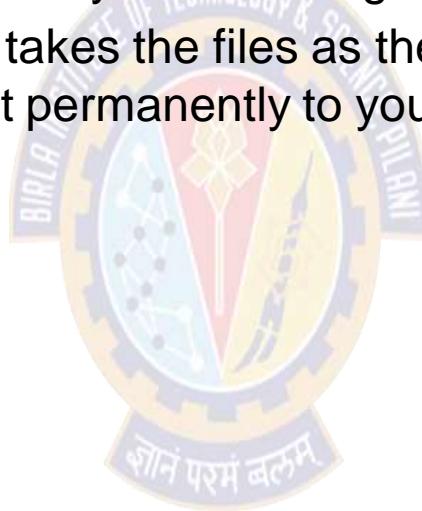
- The Git directory:
 - It is where Git stores the metadata and object database for your project
 - It is copied when you clone a repository from another computer
- The working tree:
 - It is a single checkout of one version of the project
 - These files are pulled out of the compressed database in the Git directory and placed on disk for you to use or modify
- The staging area:
 - It is a file, generally contained in your Git directory
 - It stores information about what will go into your next commit
 - Its technical name in Git parlance



Git workflow

The basic Git workflow

- You modify files in your working tree
- You selectively stage just those changes you want to be part of your next commit; which adds only those changes to the staging area
- You do a commit; which takes the files as they are in the staging area and stores that snapshot permanently to your Git directory



Agenda

Git workflow

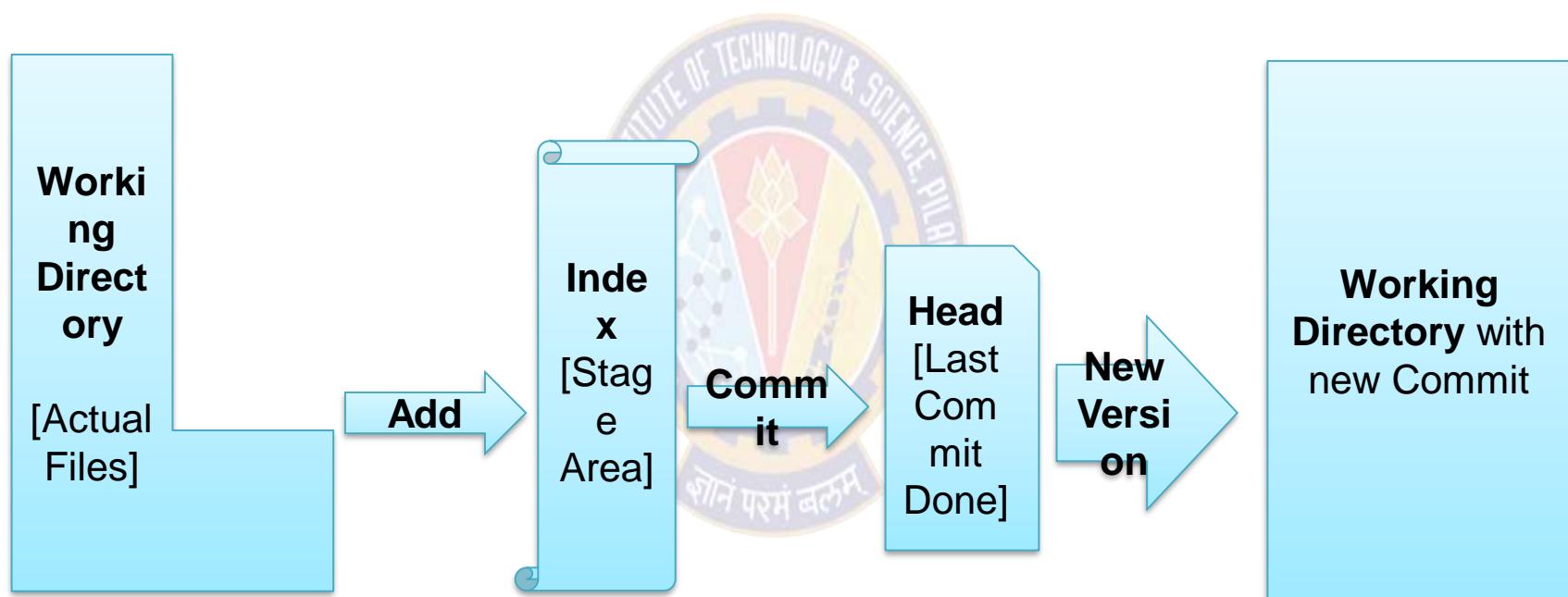
- Git tree Structure
- What is Git workflow?
- Git Centralized workflow



Git workflow

Git tree Structure

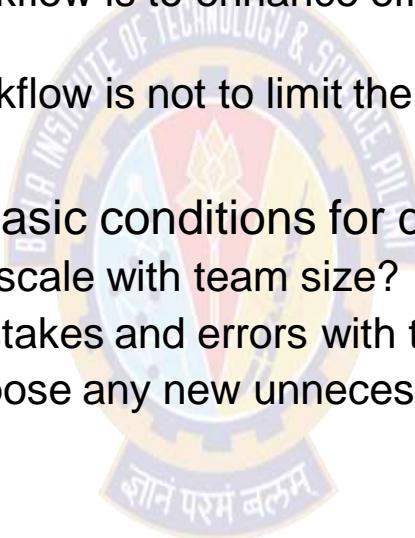
- Git local repository consists of three “trees”



Git workflow

Considerations

- To evaluate a Workflow of Git for a team you must consider:
 - Teams culture
 - Let team know the workflow is to enhance effectiveness of team and not to burden
 - Let team know the workflow is not to limit the productivity
- Selecting a Workflow? Basic conditions for decision
 - Will selected workflow scale with team size?
 - Is it easy to handle mistakes and errors with this workflow?
 - Does this workflow impose any new unnecessary cognitive overhead to the team?



Git workflow

What is Git workflow?

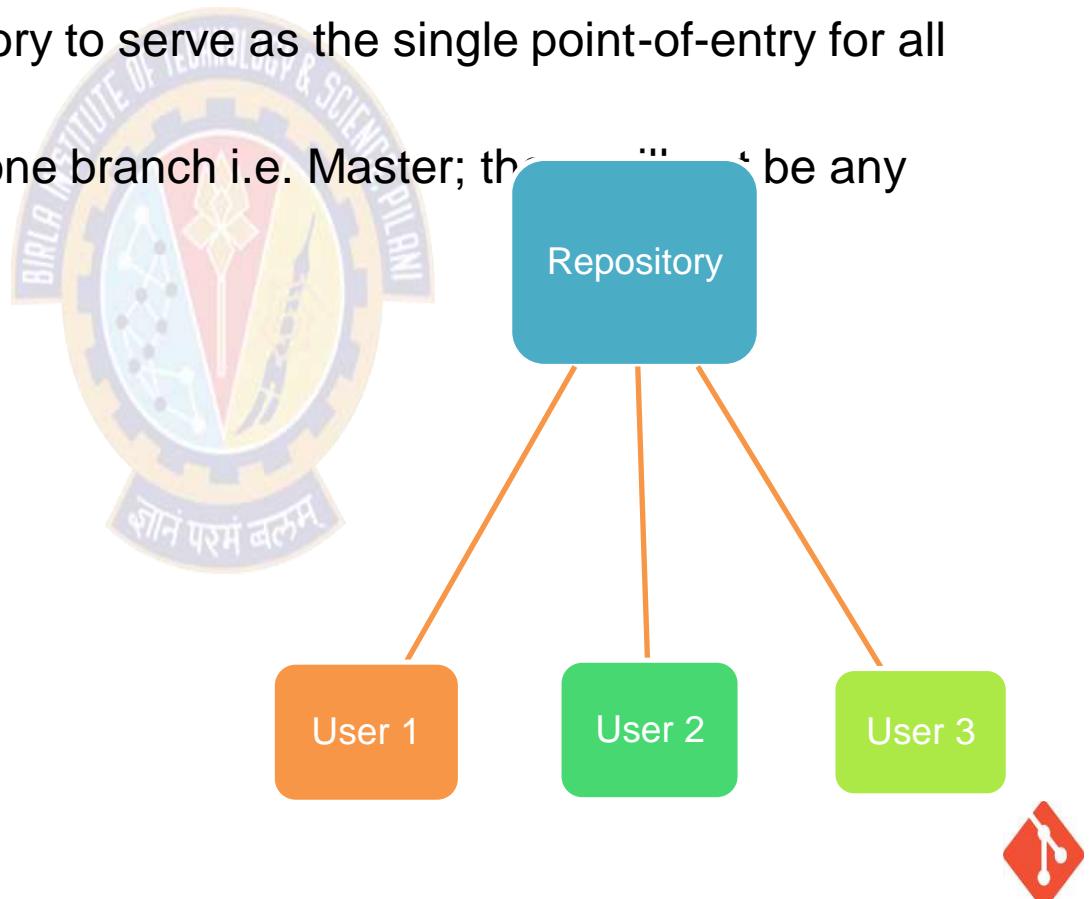
- A Git Workflow is a recipe or recommendation for how to use Git to accomplish work in a consistent and productive manner
- Git workflows encourage users to leverage Git effectively and consistently
- Git offers a lot of flexibility in how users manage changes
- There are several publicized Git workflows that may be a good fit for your team
- We will discuss:
 - Centralized Workflow
 - Feature Branching Workflow



Git Centralized workflow

Introduction

- It is a great Git workflow for teams transitioning from SVN (Subversion)
- It uses a central repository to serve as the single point-of-entry for all changes to the project
- Here there will be only one branch i.e. Master; there won't be any other branch



Git Centralized workflow

Benefits

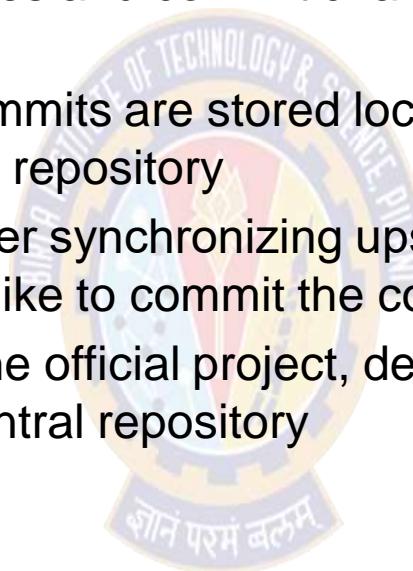
- Centralized workflow gives every developer their own local copy of the entire project
- This isolated environment lets each developer work independently of all other changes to a project
- Centralized workflow gives you access to Git's robust branching and merging model
- Git branches are designed to be a fail-safe mechanism for integrating code and sharing changes between repositories
- The Centralized Workflow is similar to other workflows in its utilization of a remote server-side hosted repository that developers push and pull from
- A Centralized Workflow is generally better suited for teams migrating from SVN to Git and smaller size teams



Git Centralized workflow

Git Centralized workflow: How it works?

- Developer Starts Cloning the central repository
- In local copy they edit files and commit changes as they would with SVN
- However, these new commits are stored locally - they're completely isolated from the central repository
- This lets developers defer synchronizing upstream until they're at a state where they would like to commit the code to the Master
- To publish changes to the official project, developers "push" their local master branch to the central repository



Git Centralized workflow

Git Centralized workflow: Example

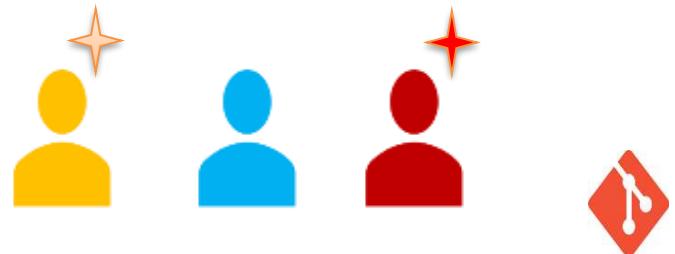
- Lets say the team of three people named Orange, Blue and Red are working on a Centralized Git Repository
- They will be collaborating with Centralized Workflow



Git Centralized workflow

Git Centralized workflow: Example Contd..

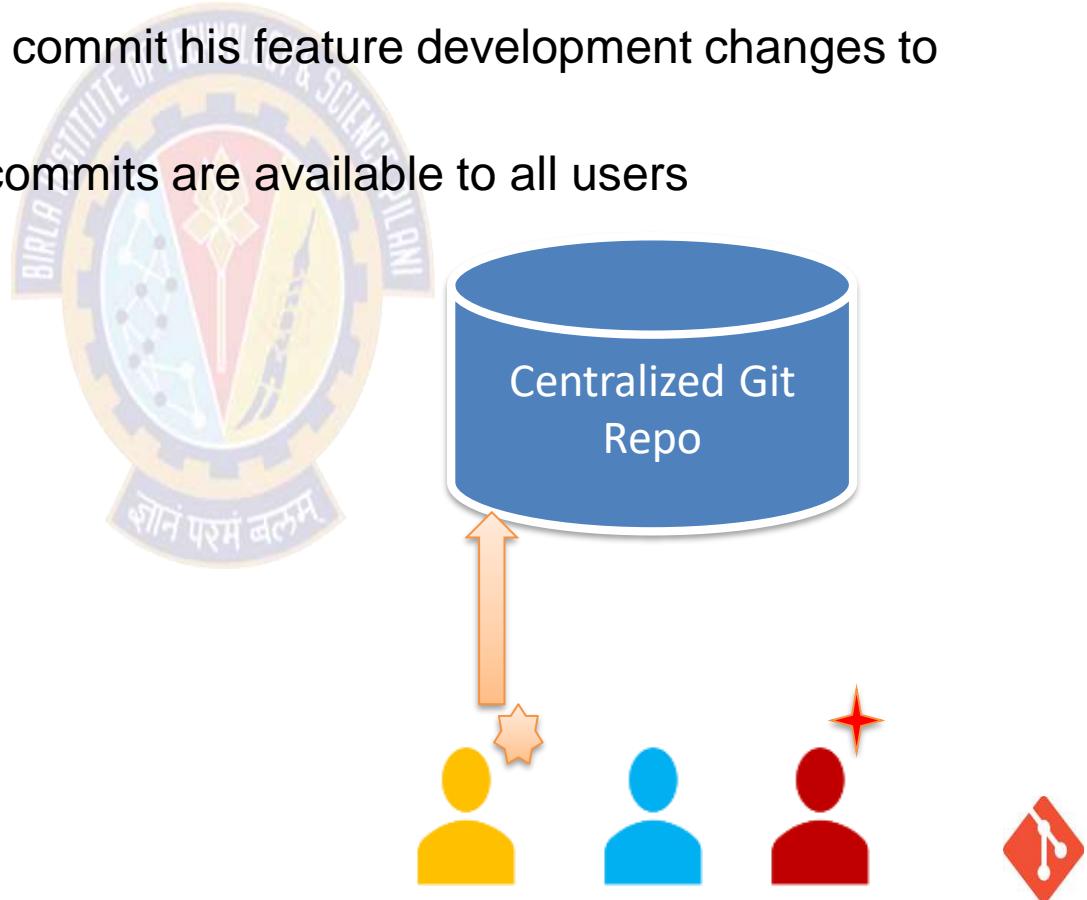
- User Orange is working on a feature development in his local repository
- At the Same time User Red is also working on a separate feature in his local repository



Git Centralized workflow

Git Centralized workflow: Example Contd..

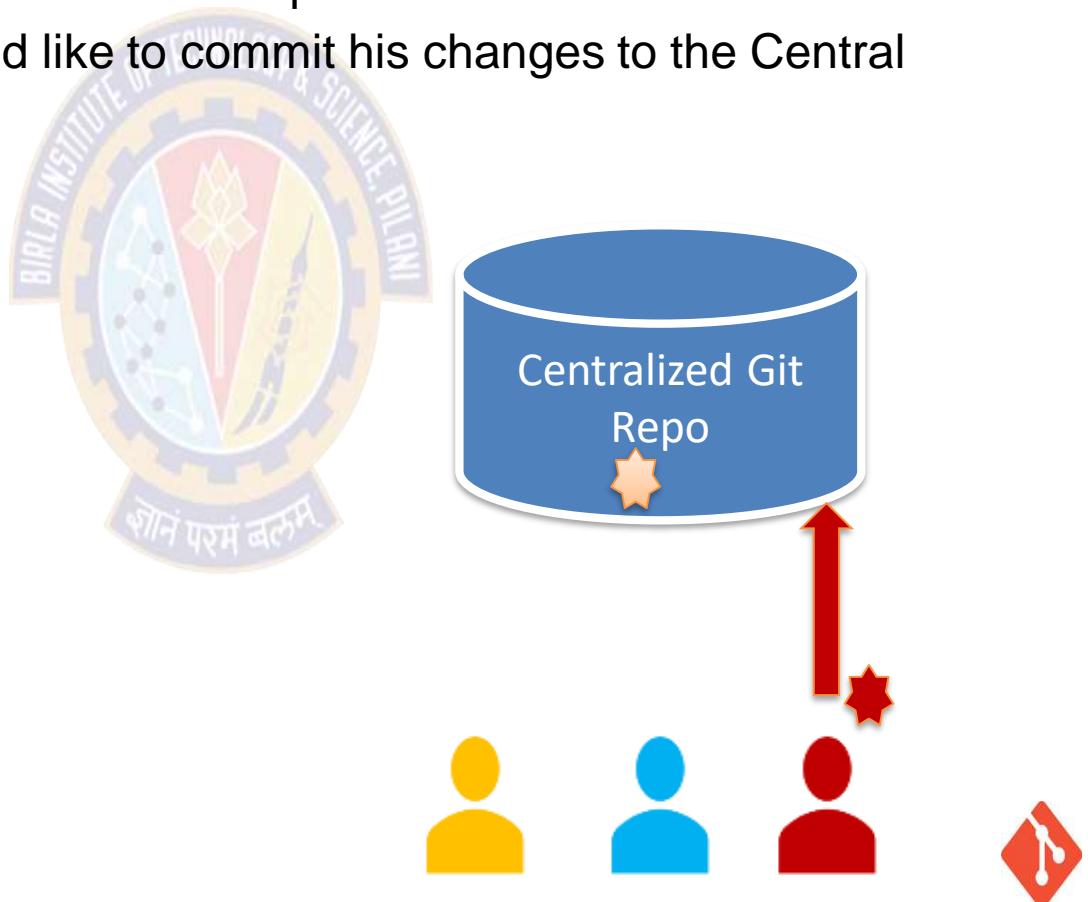
- User Orange has completed his feature development
- And user Orange publishes his feature
- I.e. User Orange should commit his feature development changes to Central repository
- So now these updated commits are available to all users



Git Centralized workflow

Git Centralized workflow: Example Contd..

- After the commit of Orange user
- User Red complete his feature development
- And now User Red would like to commit his changes to the Central Repository
- **And It Fails**



Git Centralized workflow

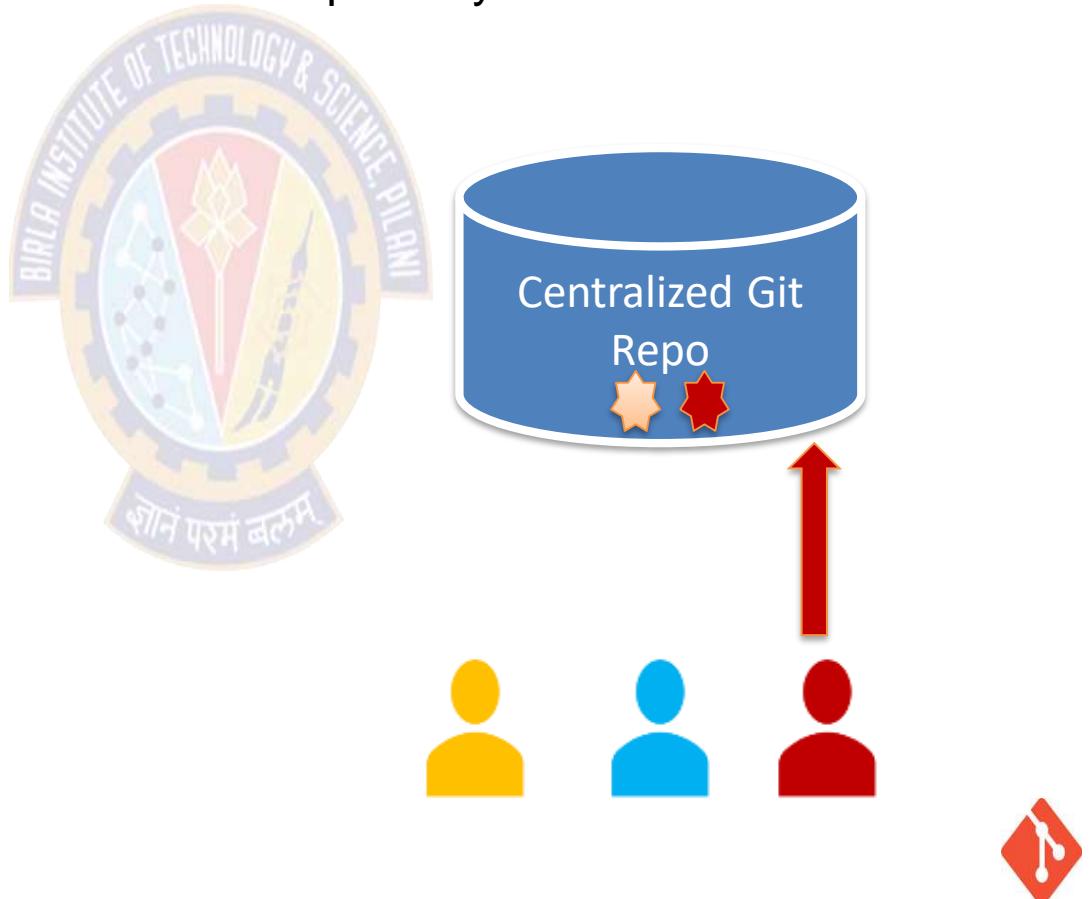
Git Centralized workflow: Example Contd..

- Here user Red need to pull the recent changes done to repository
- Incorporate his changes with recent changes done by User Orange
- I.e. Kind of rebase activity and resolve the conflicts if any



Git Centralized workflow

- Now user Red can commit his changes in local repository
- And Publish his feature to the central repository
- Its Successful



Git Centralized workflow

Summary

- The Centralized Workflow is great for small teams
- The conflict resolution process detailed above can form a bottleneck as your team scales in size
- This is great for transitioning teams off of SVN, but it doesn't leverage the distributed nature of Git



Agenda

Feature workflow

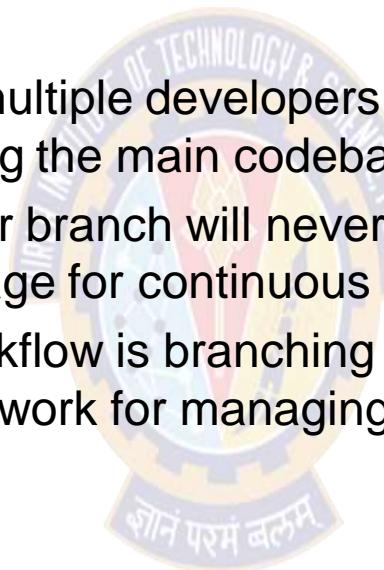
- Introduction
- How it works?



Git Feature Branch Workflow

Introduction

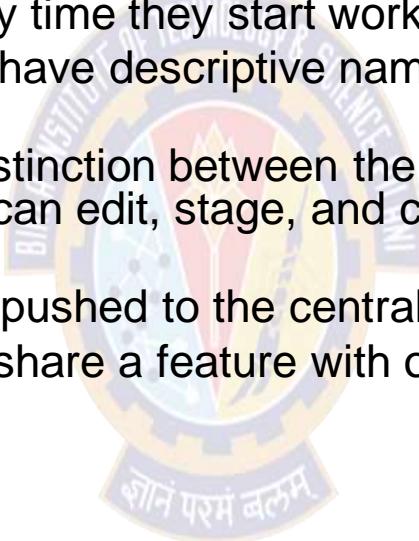
- The core idea behind the Feature Branch Workflow is that all feature development should take place in a dedicated branch instead of the master branch
- This makes it easy for multiple developers to work on a particular feature without disturbing the main codebase
- It also means the master branch will never contain broken code, which is a huge advantage for continuous integration environments
- Git Feature Branch Workflow is branching model focused, meaning that it is a guiding framework for managing and creating branches



Git Feature Branch Workflow

Git Feature Branch workflow: How it works?

- The Feature Branch Workflow assumes a central repository, and master represents the official project history
- Instead of committing directly on their local master branch, developers create a new branch every time they start work on a new feature
- Feature branches should have descriptive names like web page development #23
- Git makes no technical distinction between the master branch and feature branches, so developers can edit, stage, and commit changes to a feature branch
- Feature branches can be pushed to the central repository
- This makes it possible to share a feature with other developers without touching any official code



Git Feature Branch Workflow

Git Feature Branch workflow: Example

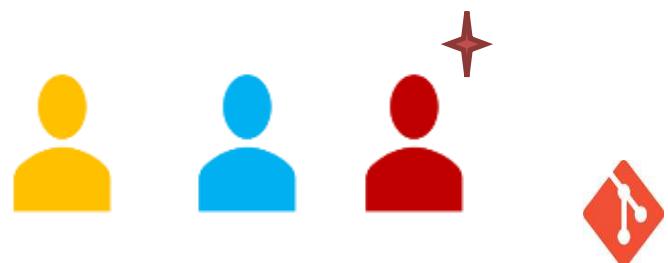
- Lets say the team of three people named Orange, Blue and Red are working on a Centralized Git Repository
- They will be collaborating with Feature Branch Workflow



Git Feature Branch Workflow

Git Feature Branch workflow: Example Contd..

- Before user Red starts working on a new feature, he request his own new branch
- User Red Starts development of his new feature
- At the end of the day user Red complete 80% of his feature development
- Before leaving for the day user Red push the commit to the Central Repository
- So the new feature work will be available for all developer and also to the member who is collaborating with user Red for feature development



Git Feature Branch Workflow

Git Feature Branch workflow: Example Contd..

- Next day User Red complete his development
- Now before merging the new Feature to master, User Red would like to let all other team member know he is done with his work by filing a pull request
- Before filing pull request User Red need to make sure the Branch is available on Central Repository



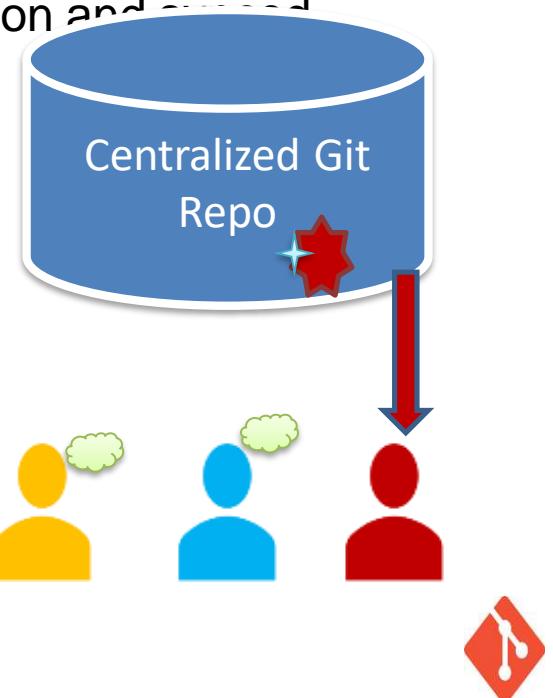
Only Push no merge
File a pull request by User Red



Git Feature Branch Workflow

Git Feature Branch workflow: Example Contd..

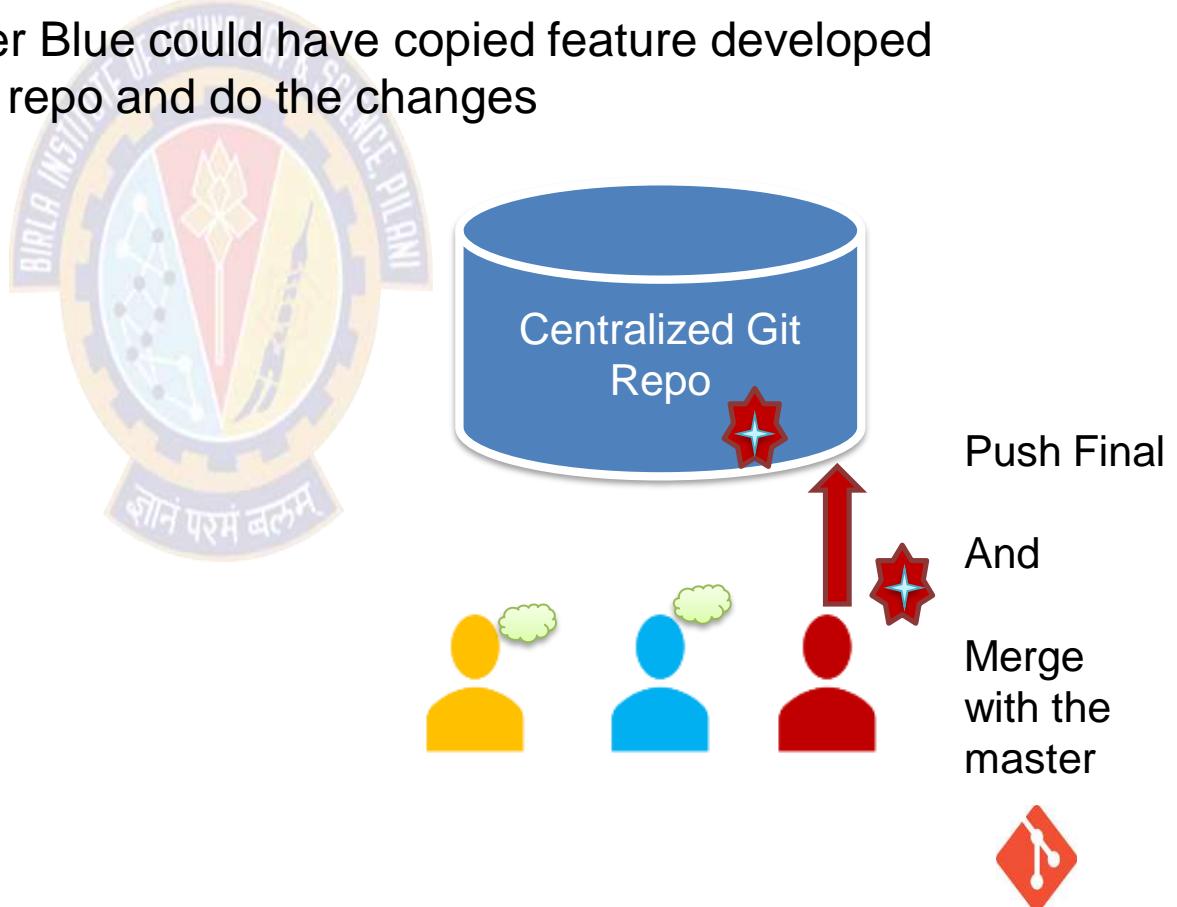
- Once User Red file a pull request Every user gets a notification that User Red is done with his feature development
- User Blue Checks the notification and visit the feature developed by User Red
- User Blue adds some more in to the feature pushed by user Red
- User Blue and User Red has discussed the addition and removed



Git Feature Branch Workflow

Git Feature Branch workflow: Example Contd..

- User Red combined the addition by user Blue and creates final Push
- And finally user Red merges feature with Master
- Even here if wanted User Blue could have copied feature developed by User Red to his local repo and do the changes



Git Feature Branch Workflow

Summary

- This workflow helps organize and track branches that are focused on business domain feature sets
- Feature Branch Workflow are focused on branching patterns
- Feature Branch Workflow promotes collaboration with team members through pull requests and merge reviews
- A feature branching model is a great tool to promote collaboration within a team environment





Thank You!

In our next session:



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps





CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 5

Agenda

Git

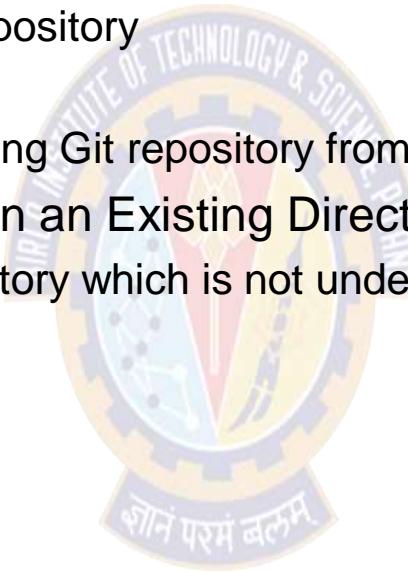
- Git basic commands



Git basic commands

Getting a Git Repository or Creating a Git Repository

- You typically obtain a Git repository in one of two ways
 - You can take a local directory that is currently not under version control, and turn it into a Git repository
 - OR
 - You can clone an existing Git repository from elsewhere
- Initializing a Repository in an Existing Directory
 - Go to the Project Directory which is not under version control
 - And type



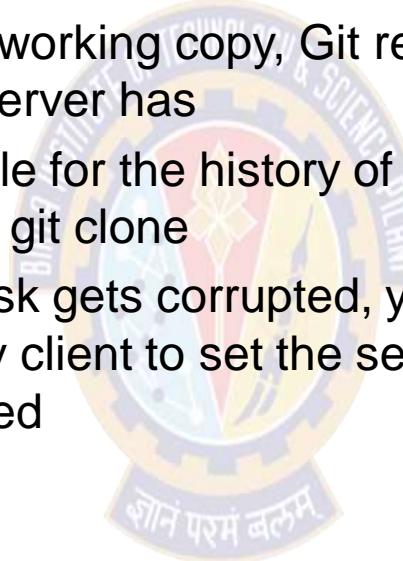
#git init



Git basic commands

Cloning an Existing Repository

- If you want to get a copy of an existing Git repository
- You should notice that the command is "clone" and not "checkout"
- Instead of getting just a working copy, Git receives a full copy of nearly all data that the server has
- Every version of every file for the history of the project is pulled down by default when you run git clone
- Benefit: if your server disk gets corrupted, you can often use nearly any of the clones on any client to set the server back to the state it was in when it was cloned



#git clone <url>



Git basic commands

Create a working copy

- Create a working copy of a local repository

```
#git clone  
/path_to_repository
```

- When using a remote server:

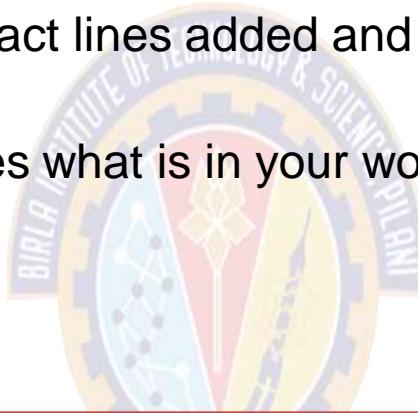
```
#git clone  
username@host:/path_to_repositor  
y
```



Git basic commands

Getting a Git Repository Status:

- The Git command to know exactly what you changed, not just which files were changed
- git diff shows you the exact lines added and removed — the patch, as it were
- That command compares what is in your working directory with what is in your staging area



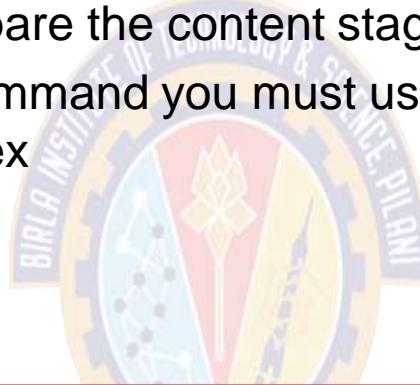
#git diff



Git basic commands

Add a file to Git Repository

- The Git command to add a file to the git repository is git add
- This command updates the index using the current content found in the working tree, to prepare the content staged for the next commit
- Before using commit command you must use the add any new or modified files to the index



```
#git add <file  
name>
```



Git basic commands

Committing your changes

- Once the staging area is set up, you can commit your changes
- Remember that anything that is still unstaged i.e. any files you have created or modified that you haven't run git add on since you edited them, those won't go into this commit
- The default commit message contains the latest output of the git status

#git commit

**#git commit –m “Commit
Message”**



Git basic commands

Removing Files from a Git Repository

- To remove a file from Git, you have to remove it from your staging area and then commit
- The git rm command does that



```
#git rm <filename>
```

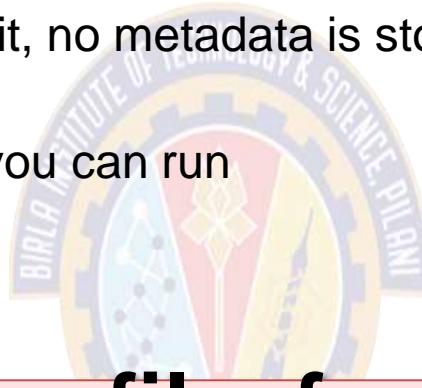


Git basic commands

Moving Files or Renaming Files in a Git Repository

- Unlike many other VCS systems, Git doesn't explicitly track file movement
- If you rename a file in Git, no metadata is stored in Git that tells it you renamed the file
- To rename a file in Git, you can run

```
#git mv <file_from>  
      <file_to>
```



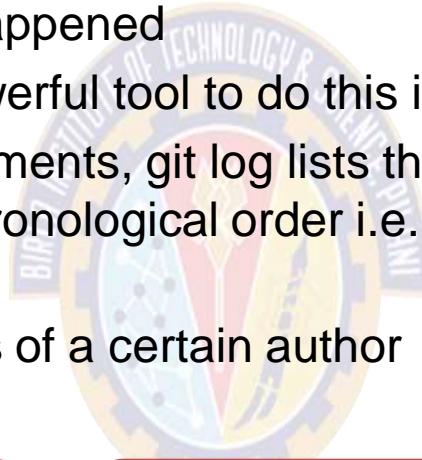
Git basic commands

Viewing the Commit History of Git Repository

- After you have created several commits, or if you have cloned a repository with an existing commit history, you'll probably want to look back to see what has happened
- The most basic and powerful tool to do this is the
- By default, with no arguments, git log lists the commits made in that repository in reverse chronological order i.e. the most recent commits show up first
- To see only the commits of a certain author

#git log

**#git log --
author=john**



Git basic commands

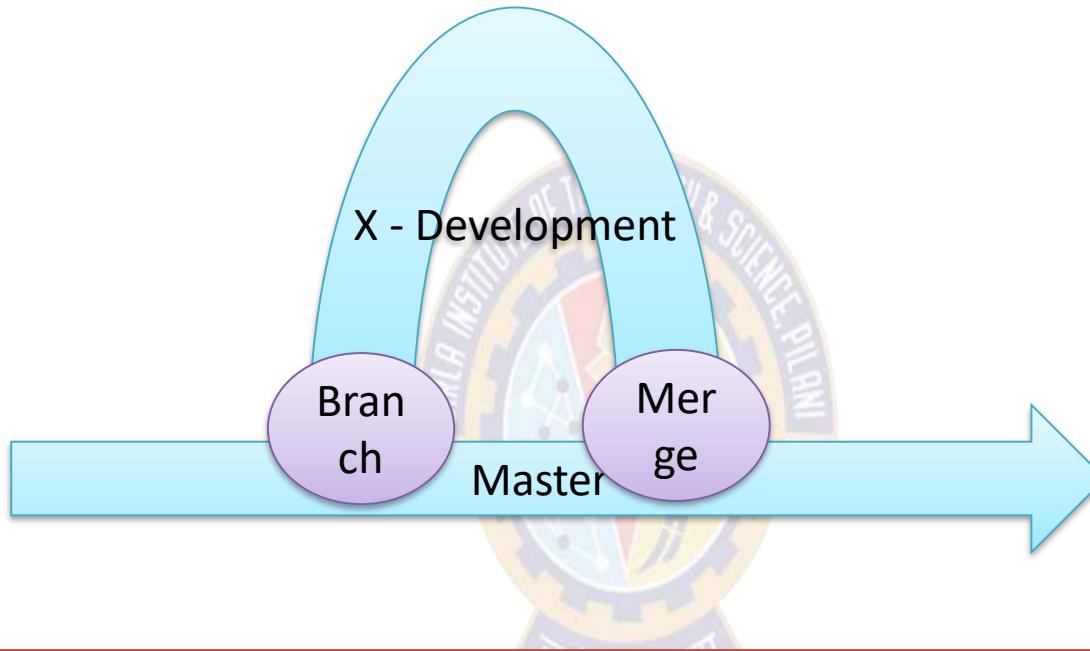
Git Branching

- As we saw nearly every Version Control System has Branching Support
- Branching means you diverge from the main line of development and continue to do work without messing with that main line
- Branches are used to develop features isolated from each other
- The way Git branches is incredibly lightweight, making branching operations nearly instantaneous
- Git doesn't store data as a series of changesets or differences
- However Git instead stores data as a series of snapshots
- Which means when you make a commit, Git stores a commit object that contains a pointer to the snapshot of the content you staged



Git basic commands

Git Branching Contd..



```
#git branch <branch name>
```



Git basic commands

Git Branching Contd..

- What happens if you create a new branch?
- Creation of Branch creates a new pointer for you to move around
- How does Git know what branch you're currently on?
- Git keeps a special pointer called HEAD; HEAD acts as a pointer to the local branch you're currently on
- The git branch command only created a new branch; it didn't switch to that branch, you will be still on master branch
- To switch to an existing branch, you run

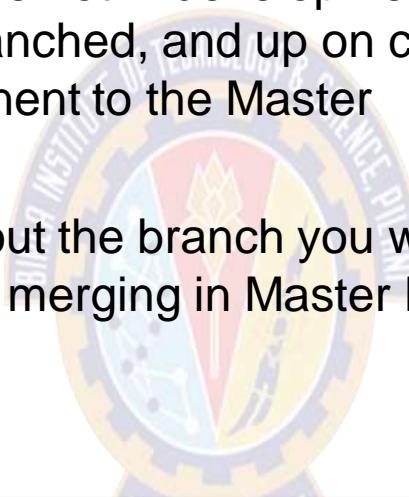
```
#git checkout <branch name>
```



Git basic commands

Git Merging

- When to Merge?
- If you have completed the hotfix development or feature development or the Individual task branched, and up on completion of that now you need to add that component to the Master
- How does it work?
- First you need to checkout the branch you wish to merge in to, for an example here we will be merging in Master Branch
- Then Run



```
#git merge <name of branch>
```



Git basic commands

Getting a Git Repository Status & check conflicts

- The main tool you use to determine which files are in which state is the git status command
- The below output means you have a clean working directory — in other words, none of your tracked files are modified

#git
status

Example:\$ git status
On branch master

No commits yet

Changes to be committed:
(use "git rm --cached <file>..."
to unstage)



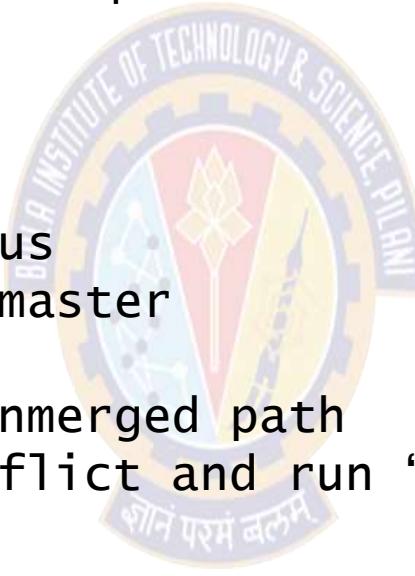
Git basic commands

Git check conflicts or status

- git status output is pretty comprehensive, it's quite self explanatory
- Lets see how the conflicts output looks like

```
$ git status  
On branch master
```

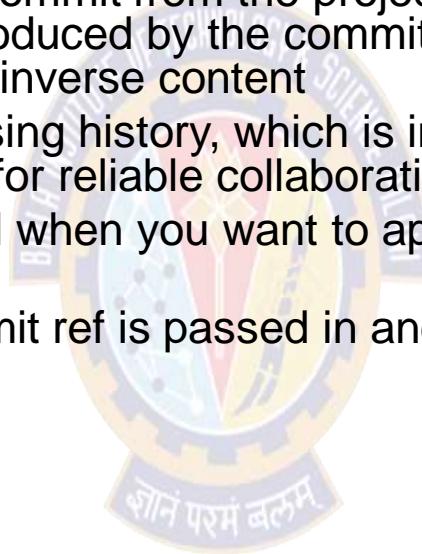
You have unmerged path
(Fix conflict and run “git-
commit”)



Git basic commands

Git Revert

- The git revert command can be considered an 'undo' type command
- however, it is not a traditional undo operation
- Instead of removing the commit from the project history, it figures out how to invert the changes introduced by the commit and appends a new commit with the resulting inverse content
- This prevents Git from losing history, which is important for the integrity of your revision history and for reliable collaboration
- Reverting should be used when you want to apply the inverse of a commit from your project history
- Git revert expects a commit ref is passed in and will not execute without one



Git basic commands

Git Revert:

- Similar to a merge, a revert will create a new commit
- It's important to understand that git revert undoes a single commit
- It does not "revert" back to the previous state of a project by removing all subsequent commits
- The git revert command is a forward-moving undo operation that offers a safe method of undoing changes
- Instead of deleting or orphaning commits in the commit history, a revert will create a new commit that inverses the changes specified
- Git revert is a safer alternative to git reset in regards to losing work

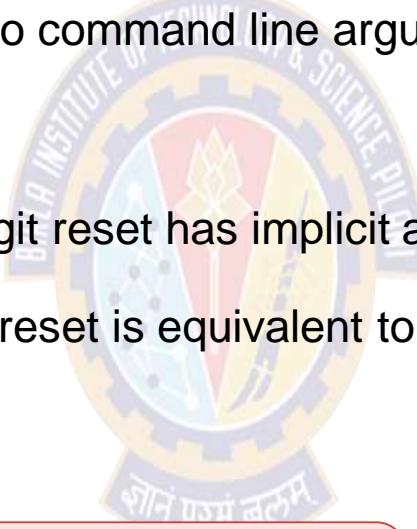
#git revert HEAD



Git basic commands

Git Reset

- The git reset command is a complex and versatile tool for undoing changes
- It has three primary forms of invocation
- These forms correspond to command line arguments
 - soft
 - mixed
 - hard
- The default invocation of git reset has implicit arguments of --mixed and HEAD
- This means executing git reset is equivalent to executing git reset --mixed HEAD



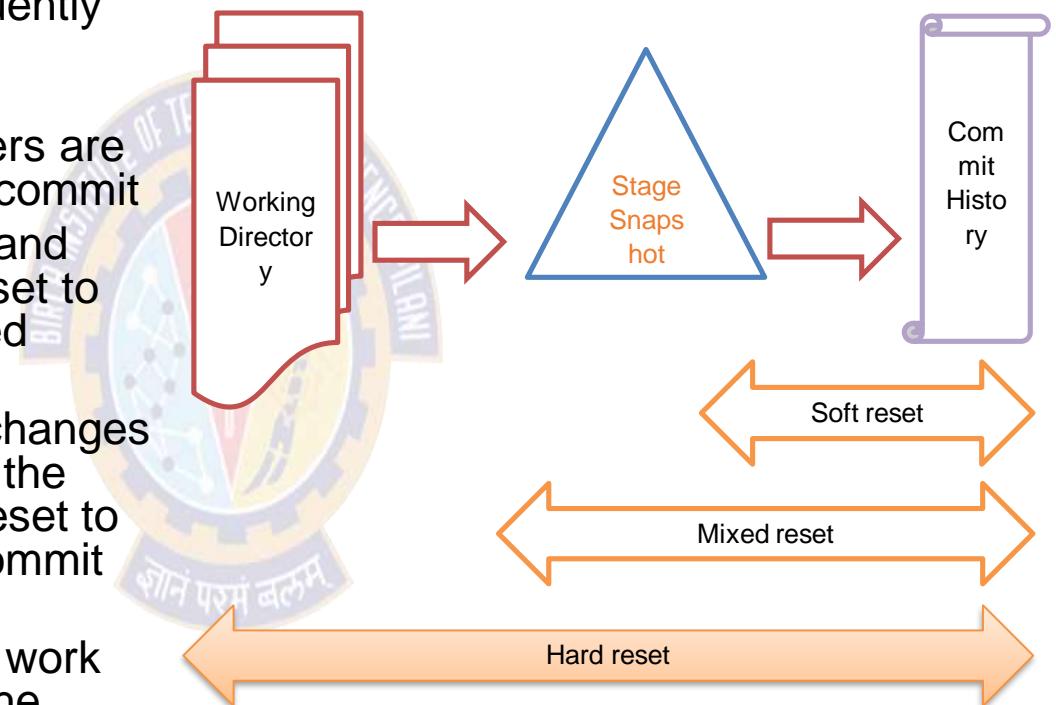
#git reset



Git basic commands

Git Reset: --hard

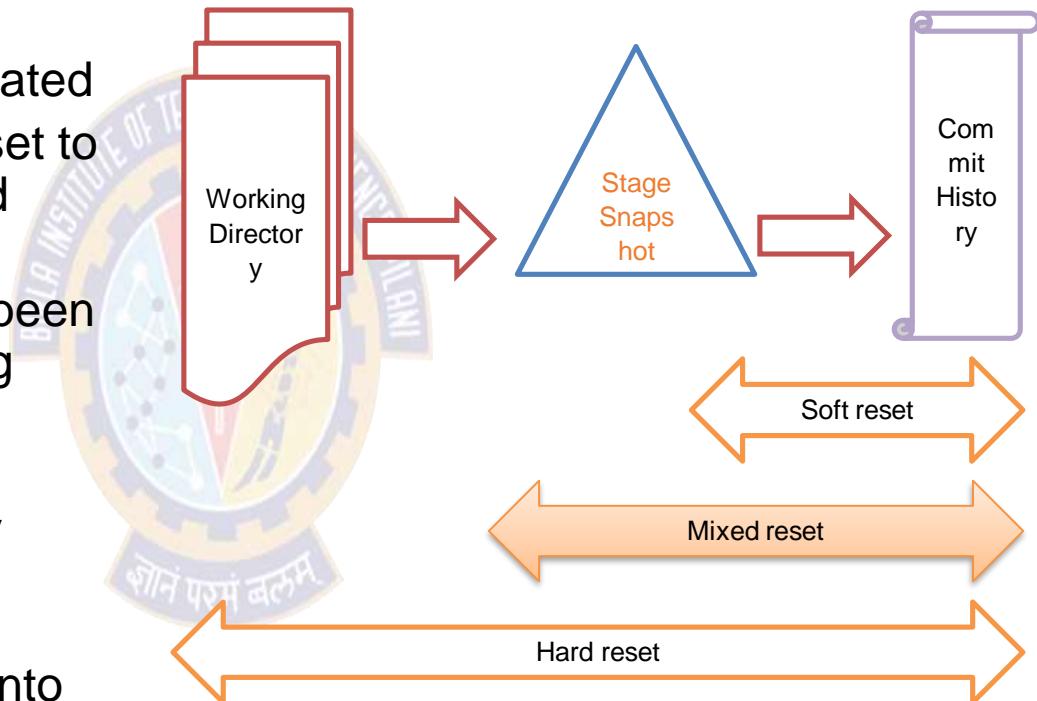
- This is the most direct, **DANGEROUS**, and frequently used option
- When passed `--hard` The Commit History ref pointers are updated to the specified commit
- Then, the Staging Index and Working Directory are reset to match that of the specified commit
- Any previously pending changes to the Staging Index and the Working Directory gets reset to match the state of the Commit Tree
- This means any pending work that was hanging out in the Staging Index and Working Directory will be lost



Git basic commands

Git Reset: --mixed

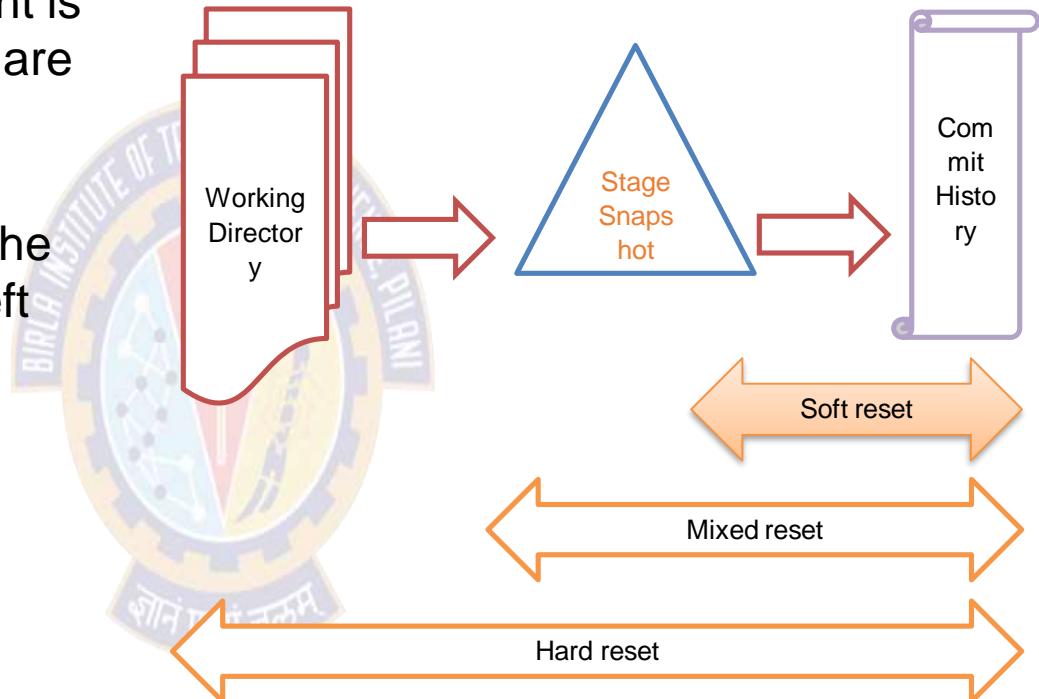
- This is the default operating mode
- The ref pointers are updated
- The Staging Index is reset to the state of the specified commit
- Any changes that have been undone from the Staging Index are moved to the Working Directory
- In other way we can say here Staging Index gets reset and the pending changes will be moved into the Working Directory



Git basic commands

Git Reset: --soft

- When the --soft argument is passed, the ref pointers are updated and the reset stops there
- The Staging Index and the Working Directory are left untouched



Git basic commands

Git Revert vs Git Reset

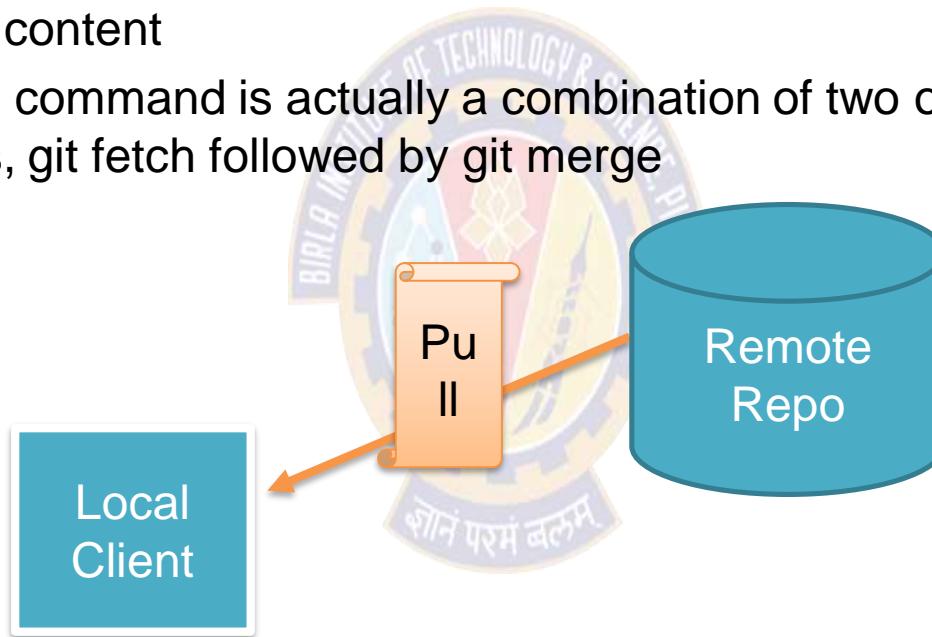
- Git Revert:
- It is a “safe” way to undo changes
- it doesn’t change the project history, which makes it a “safe” operation for commits that have already been published to a shared repository
- git revert is able to target an individual commit at an arbitrary point in the history
- git revert is a safer alternative to git reset in regards to losing work
- Git Reset:
- It is the dangerous method
- There is a real risk of losing work with git reset
- git reset can only work backward from the current commit
- git reset will never delete a commit, however, commits can become ‘orphaned’ which means there is no direct path from a ref to access them
- For example, if you wanted to undo an old commit with git reset, you would have to remove all of the commits that occurred after the target commit, remove it, then re-commit all of the subsequent commits



Git basic commands

Git Pull

- The git pull command is used to fetch and download content from a remote repository and immediately update the local repository to match that content
- The git pull command is actually a combination of two other commands, git fetch followed by git merge



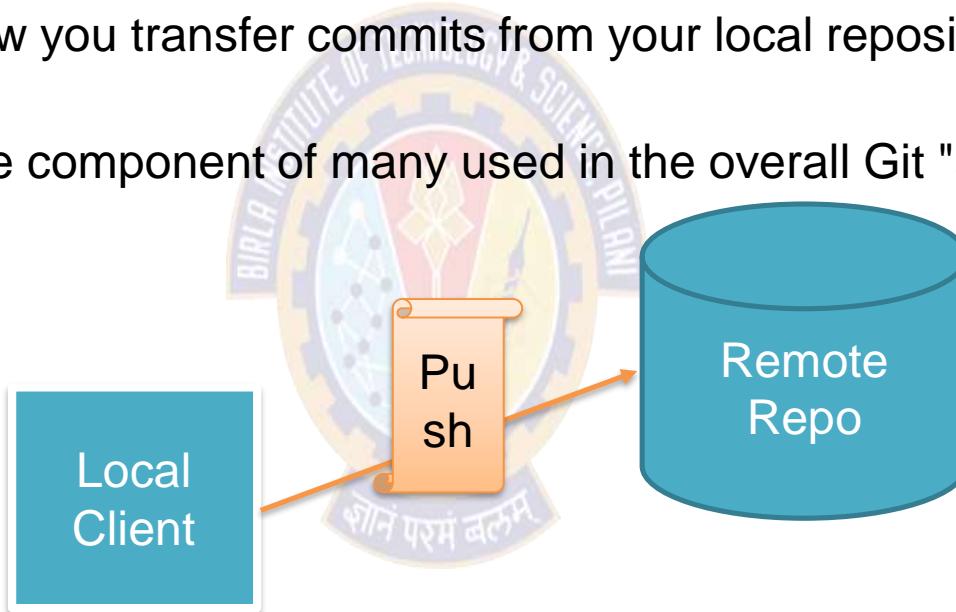
#git pull



Git basic commands

Git Push

- The git push command is used to upload local repository content to a remote repository
- Pushing is how you transfer commits from your local repository to a remote repo
- git push is one component of many used in the overall Git "syncing" process



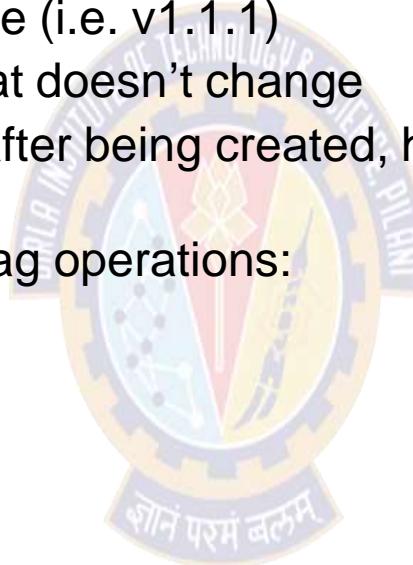
#git push



Git basic commands

Git tagging:

- Tags are ref's that point to specific points in Git history
- Tagging is generally used to capture a point in history that is used for a marked version release (i.e. v1.1.1)
- A tag is like a branch that doesn't change
- Unlike branches, tags, after being created, have no further history of commits
- We will have look in to tag operations:
 - Create tag
 - List tags
 - Delete tag
 - Sharing tag



Git Tagging

Git create tag

- To create a new tag execute the following command

```
#git tag <tag name>
```

- A common pattern is to use version numbers like git tag v1.4
- Git supports two different types of tags, annotated and lightweight tags
- A best practice is to consider Annotated tags as public, and Lightweight tags as private
- Annotated tags store extra meta data such as: the tagger name, email, and date so it helps for public release
- Lightweight tags are essentially 'bookmarks' to a commit, they are just a name and a pointer to a commit, so useful for creating quick links to relevant commits



Git Tagging

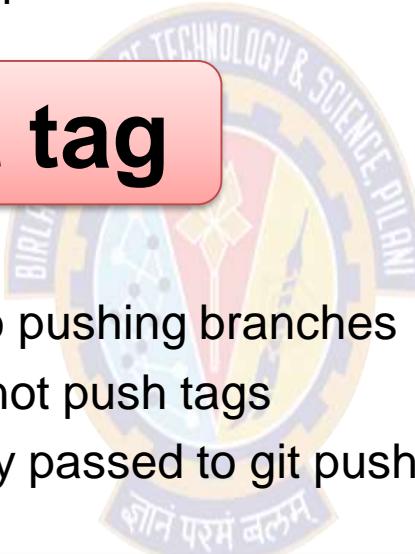
Git list tag and Git share tag

- Git list tag:
- To list stored tags in a repo execute the following command

#git tag

- Git share tag:
- Sharing tags is similar to pushing branches
- By default, git push will not push tags
- Tags have to be explicitly passed to git push

#git push origin <tag num>



Git Tagging

Git delete tag

- Deleting tags is a straightforward operation
- Passing the -d option and a tag identifier to git tag will delete the identified tag

```
#git tag -d v1
```

- Tagging is an additional mechanism used to create a snap shot of a Git repo
- Tagging is traditionally used to create semantic version number identifier tags that correspond to software release cycles



Git Tagging

Git gui

- Built in Git gui

#gitk

- Use colorful git output

#git config
color.ui true



Agenda

Best Practices of Clean Code

- Clean code
- General Rules



Clean code

Condition for a Clean Code

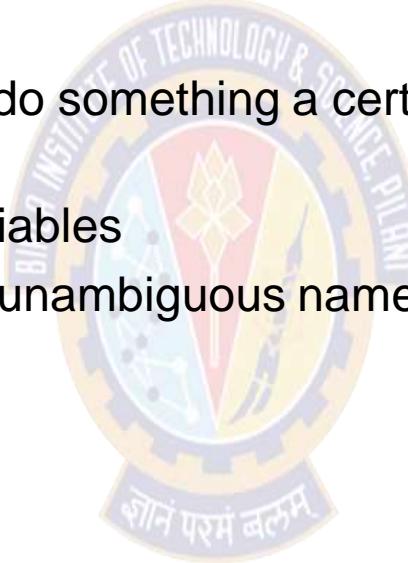
- If it can be understood easily
- It can be read and enhanced by a developer other than its original author
- If code follows readability, changeability, extensibility and maintainability



Best Practices of Clean Code

General Rules

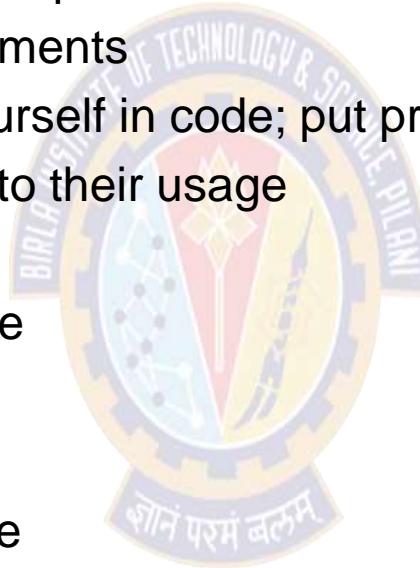
- Follow standard conventions
- Keep it simple, Simpler is always better, Reduce complexity as much as possible
- Be consistent i.e. If you do something a certain way, do all similar things in the same way
- Use self explanatory variables
- Choose descriptive and unambiguous names
- Keep functions small



Best Practices of Clean Code

General Rules Contd..

- Each Function should do one thing
- Use function names descriptive
- Prefer to have less arguments
- Always try to explain yourself in code; put proper comments
- Declare variables close to their usage
- Keep lines short
- Code should be readable
- Code should be fast
- Code should be generic
- Code should be reusable



Agenda

Dependencies

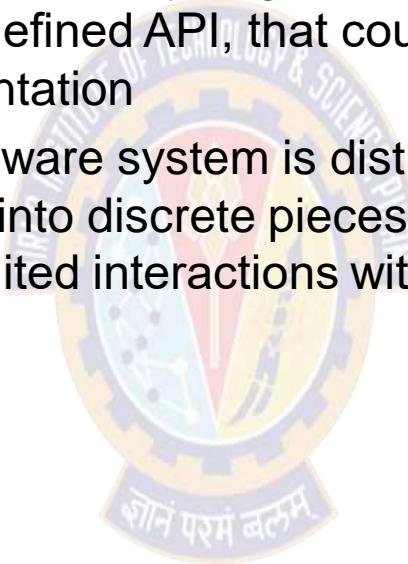
- What is component
- Best Practices for component based design
- Overview of Dependencies



Component

What is component?

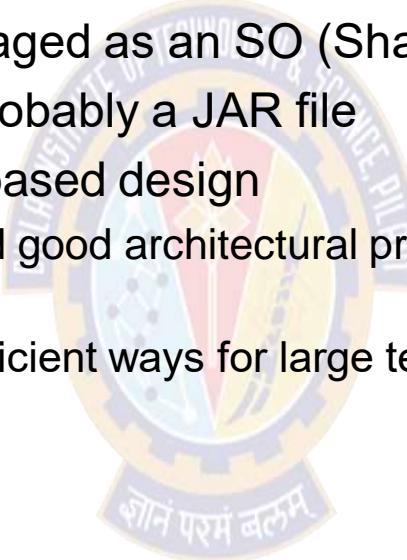
- This is a overloaded term in software
- Components, mean a reasonably large-scale code structure within an application, with a well-defined API, that could potentially be swapped out for another implementation
- A component-based software system is distinguished by the fact that the codebase is divided into discrete pieces that provide behavior through well-defined, limited interactions with other components



Component

What is component? Contd..

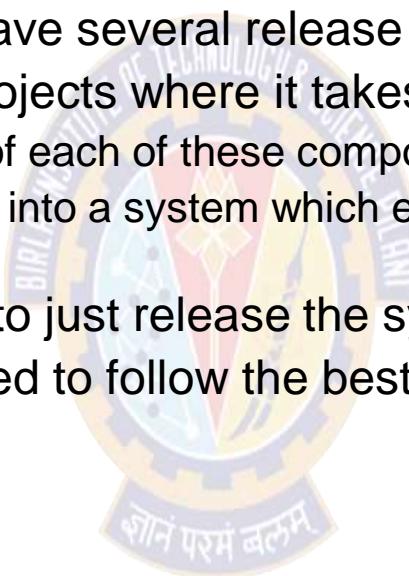
- Components also refer as “modules”
- In Windows, a component is normally packaged as a DLL
- In UNIX, it may be packaged as an SO (Shared object) file
- In the Java world, it is probably a JAR file
- Benefits of component-based design
 - Encouraging reuse and good architectural properties such as loose coupling
 - It is one of the most efficient ways for large teams of developers to collaborate



Component

Challenges of component based design

- In large system, components form a series of dependencies, which in turn depend on external libraries
- Each component may have several release branches
- So, we have heard of projects where it takes months
 - finding good versions of each of these components
 - that can be assembled into a system which even compiles is an extremely difficult process
- Which results into ages to just release the system
- To overcome this we need to follow the best practices



Component

Keeping Your Application Releasable

- During development, teams continue an implementation of features, and sometimes need to make major architectural changes
 - In such cases the application can't be released, although it will still pass the commit stage of continuous integration
 - Usually, before release, teams will stop developing new functionality and focus only on bug fixing
 - When the application is released, a release branch is created in version control, and new development begins again
 - However, this process generally results in weeks or months between releases
- The aim of continuous delivery is for the application to always be in a releasable state
- How can we achieve this?
- One approach is to create branches in version control that are merged when work is complete, so that mainline is always releasable (feature workflow)

Component

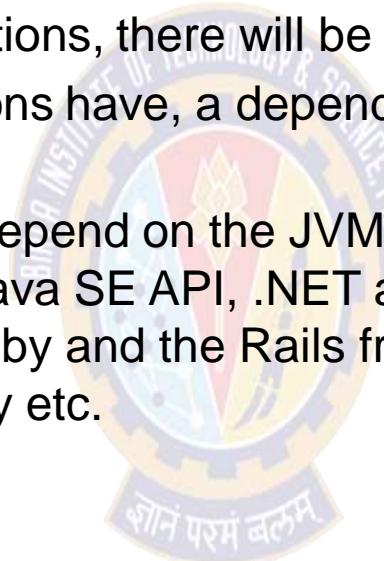
Strategies to keep your application releasable during change

- Hide new functionality until it is finished
 - problem with continuous deployment of applications is that a feature, or a set of features, can take a long time to develop which might prevent it being released
 - One solution is to put in the new features, but make them inaccessible to users
- Make all changes incrementally as a series of small changes, each of which is releasable
 - Here, analysis plays an important role to make large changes as a series of small changes
- Use branch by abstraction to make large-scale changes to the codebase
 - to make a large-scale change to an application
 - Instead of branching, an abstraction layer is created over the piece to be changed
 - A new implementation is then created in parallel with the existing implementation
 - Once it is complete, the original implementation and (optionally) the abstraction layer are removed

Dependencies

Overview of Dependencies

- A dependency occurs whenever one piece of software depends upon another in order to build or run
- In most trivial of applications, there will be some dependencies
- Most software applications have, a dependency on their host operating environment
- Like Java applications depend on the JVM which provides an implementation of the Java SE API, .NET applications on the CLR, Rails applications on Ruby and the Rails framework, C applications on the C standard library etc.



Dependencies

Types of Dependencies

- Build time dependencies and Run time dependencies
- Libraries and components

Libraries

- Software packages that your team does not control, other than choosing which to use
 - Libraries are Usually updated rarely
- This distinction is important because when designing a build process, there are more things to consider when dealing with components than libraries.
- For example, do you compile your entire application in a single step, or compile each component independently when it changes? How do you manage dependencies between components, avoiding circular dependencies?

Components

- your application depends upon, but which are also developed by your team, or other teams in your organization
- Components are usually updated frequently

Dependencies

Build time dependencies and Run time dependencies

Build-time dependencies must be present when your application is compiled and linked (if necessary)

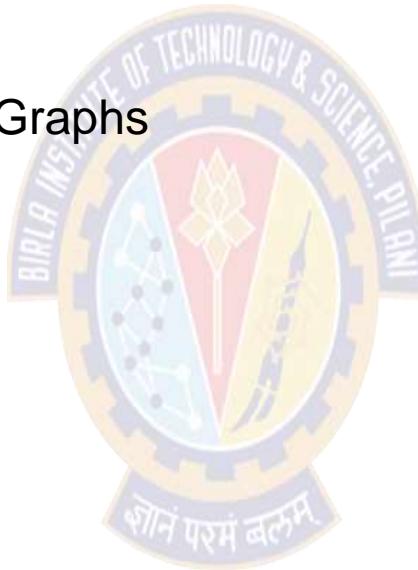
Runtime dependencies must be present when the application runs, performing its usual function

- Ex. In C and C++, your build-time dependencies are simply header files, while at run time you require a binary to be present in the form of a dynamic-link library (DLL) or shared library (SO)
- Managing dependency can be difficult

Agenda

Managing Dependency Problems

- Common dependency problem with libraries at run time
- Managing Libraries
- Managing Component
- Managing Dependency Graphs



Dependency Problems

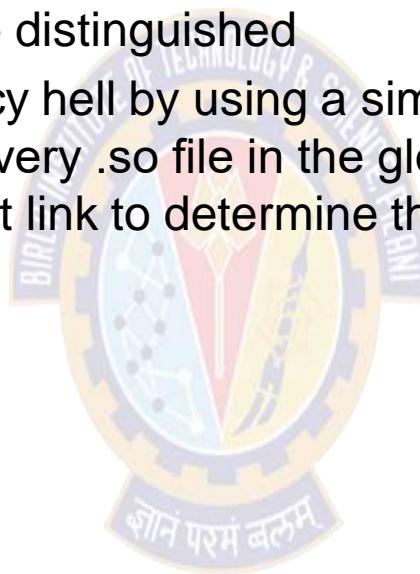
Common dependency problem with libraries at run time

- Dependency Hell also refer as DLL Hell
- Occurs when an application depends upon one particular version of something, but is deployed with a different version, or with nothing at all
- DLL hell was common problem in earlier versions of Windows,
 - as all shared libraries (DLLs), stored in a system directory (windows\system32) without any versioning; new versions would simply overwrite old ones
 - In versions of Windows prior to XP the COM class table was a singleton, so applications that required a particular COM object would be given whichever version had been loaded first

Dependency Problems

Common dependency problem with libraries at run time contd..

- The introduction of the .NET framework resolves the DLL hell problem by introducing the concept of assemblies that allow different versions of the same library to be distinguished
- Linux avoids dependency hell by using a simple naming convention: It appends an integer to every .so file in the global library directory (/usr/lib), and uses a soft link to determine the canonical system-wide version



Managing Libraries

Implementing Version Control

- One is to check them into version control
 - Is the simplest solution, and will work fine for small projects
 - a lib directory is created in your project's root to put libraries and by adding three further subdirectories: build, test, and run—for build-time, test-time, and runtime dependencies
 - using a naming convention for libraries that includes their version number; you know exactly which versions you're using
- Benefit:
 - Everything you need to build your application is in version control
 - Once you have a local check-out of the project repository, you know you can repeatably build the same packages that everybody else has
- Problems:
 - your checked-in library repository may become large and it may become hard to know which of these libraries are still being used by your application
 - Another problem crops up if your project must run with other projects on the same platform
 - Manually managing transitive dependencies across projects rapidly becomes painful

Managing Libraries

Automated

- Another solution is to declare libraries and use a tool like Maven or Ivy to download libraries from Internet repositories or (preferably) your organization's own artifact repository [Automated]



Managing Component

Overview

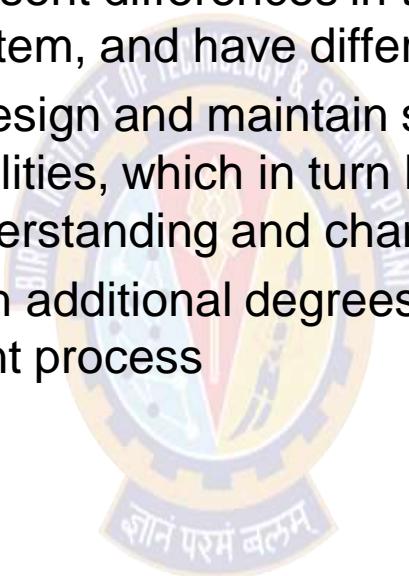
- Almost all modern software systems consist of a collection of components
- These components may be DLLs, JAR files, OSGi bundles, Perl modules, or something else
- Components have a relatively long history in the software industry



Managing Component

Components make development processes efficient

- They divide the problem into smaller and more expressive chunks
- Components often represent differences in the rates of change of different parts of the system, and have different lifecycles
- They encourage us to design and maintain software with clear description of responsibilities, which in turn limits the impact of change, and makes understanding and changing the codebase easier
- They can provide us with additional degrees of freedom in optimizing our build and deployment process



Managing Component

Components to be separated from Codebase

- Part of your codebase needs to be deployed independently (for example, a server or a rich client)
- You want to turn a monolithic codebase into a core and a set of plugins, perhaps to replace some part of your system with an alternative implementation, or to provide user extensibility
- The component provides an interface to another system (for example a framework or a service which provides an API)
- It takes too long to compile and link the code
- It takes too long to open the project in the development environment
- Your codebase is too large to be worked on by a single team

Managing Component

Pipelining Components

- Split your system into several different pipelines
- Parts of your application that have a different lifecycle (perhaps you build your own version of an OS kernel as part of your application, but you only need to do this once every few weeks)
- Functionally separate areas of your application that are worked on by different (perhaps distributed) teams may have components specific to those teams
- Components that use different technologies or build processes
- Shared components that are used by several other projects
- Components that are relatively stable and do not change frequently
- It takes too long to build your application, and creating builds for each component will be faster

Managing Component

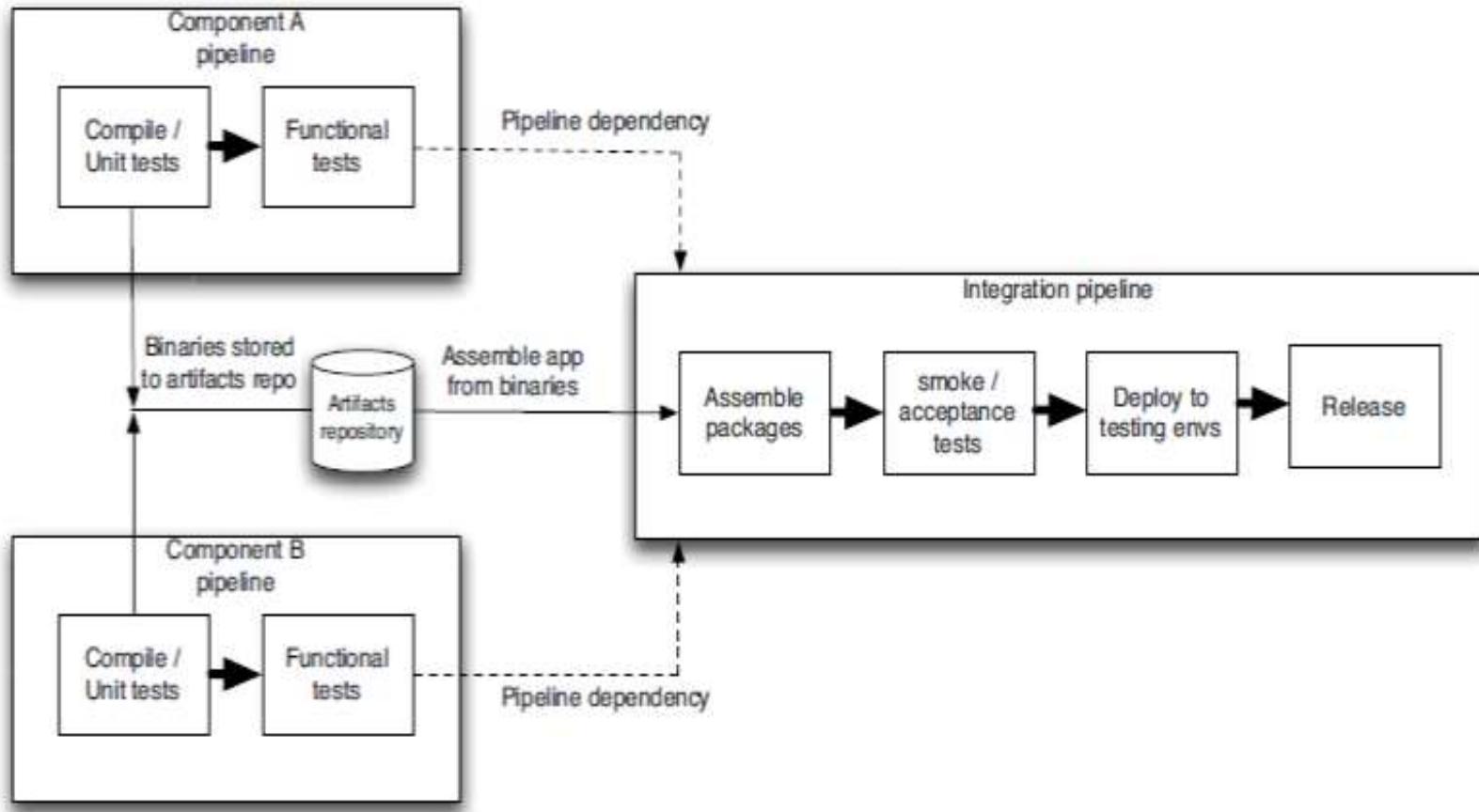
Pipelining Components

- The build for each component or set of components should have its own pipeline to prove that it is fit for release
- This pipeline will perform the following steps
 - Compile the code, if necessary
 - Assemble one or more binaries that are capable of deployment to any environment
 - Run unit tests
 - Run acceptance tests
 - Support manual testing, where appropriate



Managing Component

Integration Pipeline



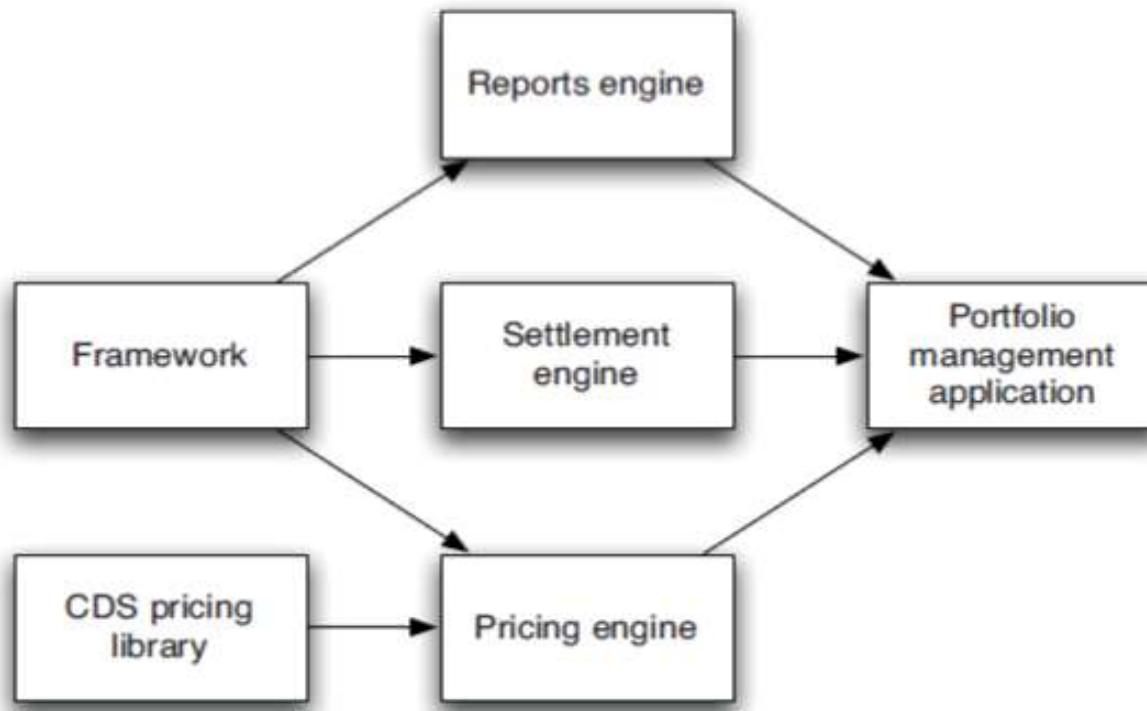
Managing Component

Managing Dependency Graphs

- Having set of components, each with their own pipeline, feeding into an integration pipeline which assembles the application and runs automated and manual tests on the final application
- However, things are often not quite this simple:
- Components can have dependencies on other components, including third-party libraries
- If you draw a diagram of the dependencies between components, it should be a directed acyclic graph (DAG)
- If this is not the case (and in particular, if your graph has cycles) you have a pathological dependency problem

Managing Component

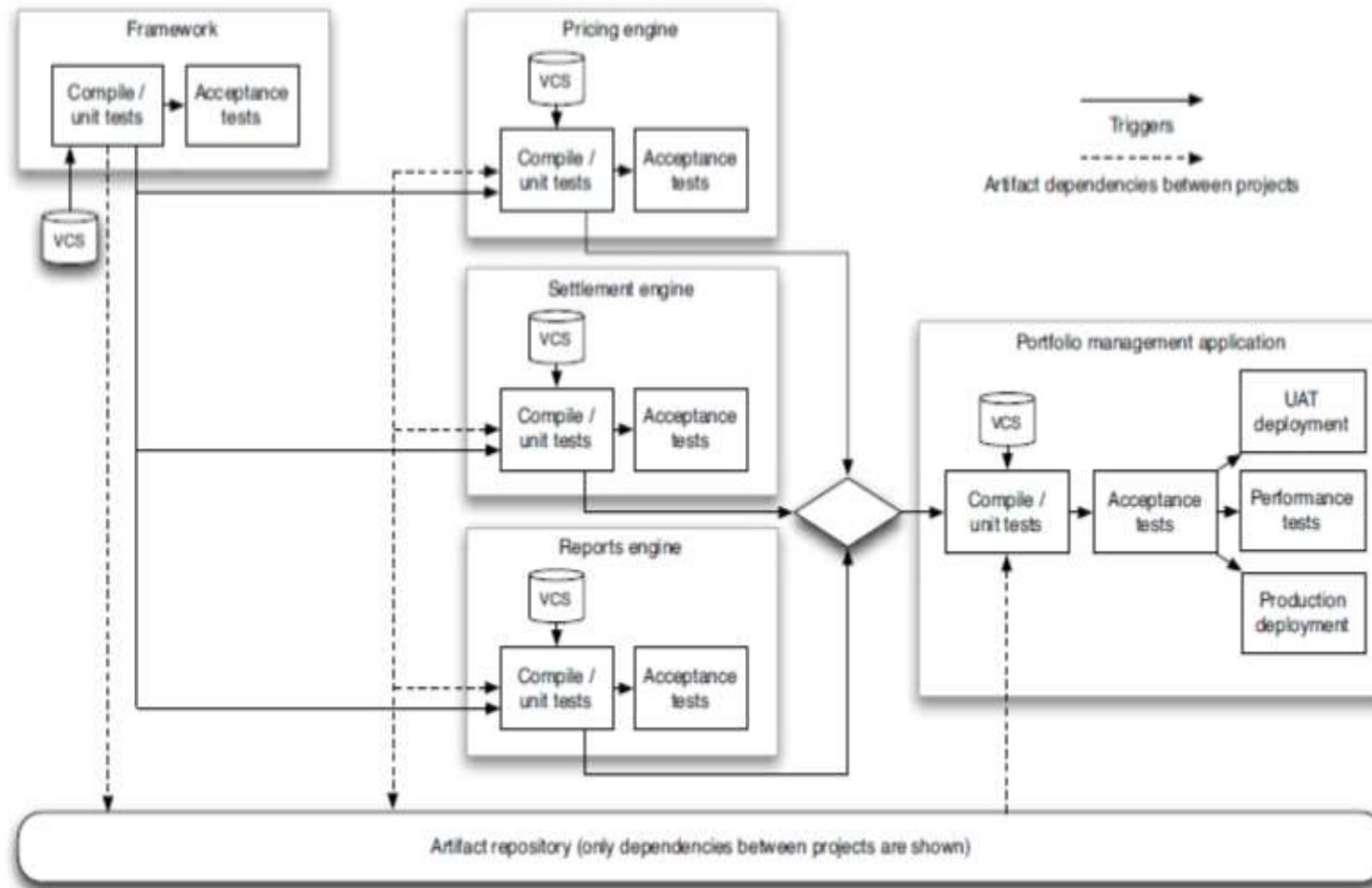
Building Dependency Graphs



credit default swap (CDS)
library that is provided by third
party

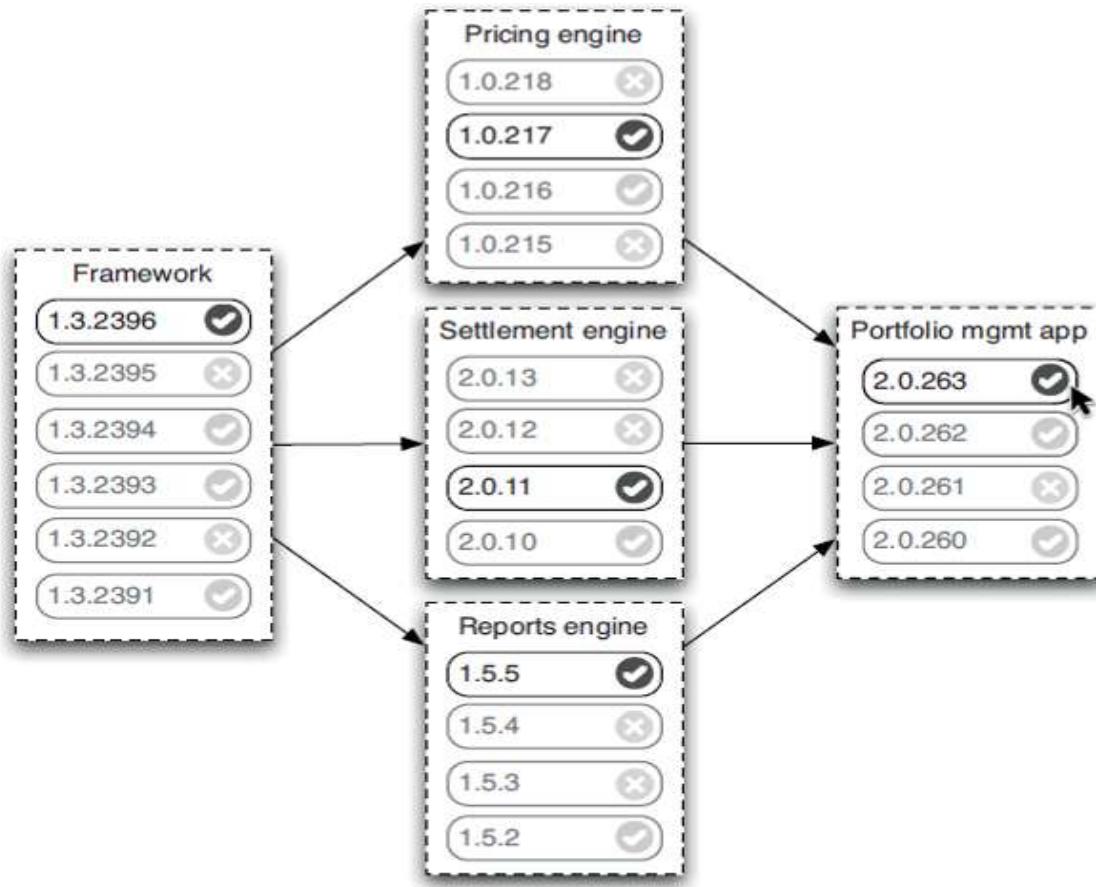
Managing Component

Pipelining Dependency Graphs



Managing Component

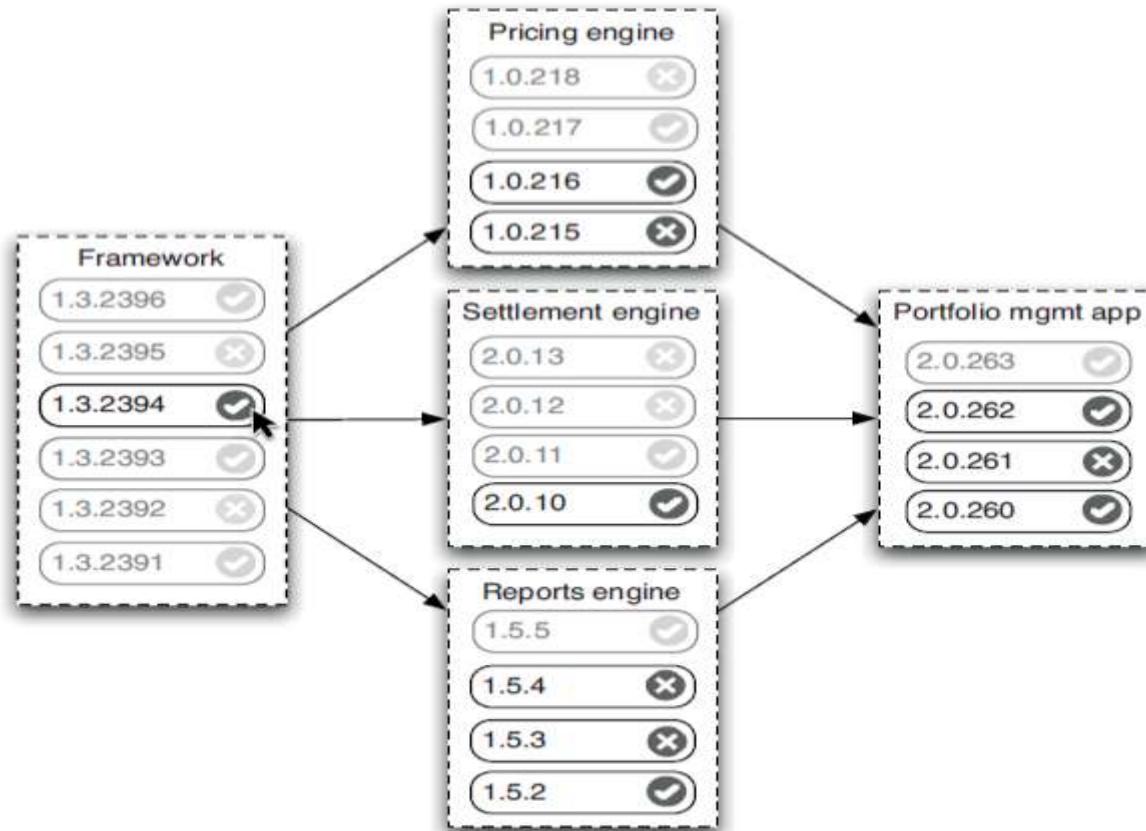
Pipelining Dependency Graphs Contd..



Visualizing upstream dependencies

Managing Component

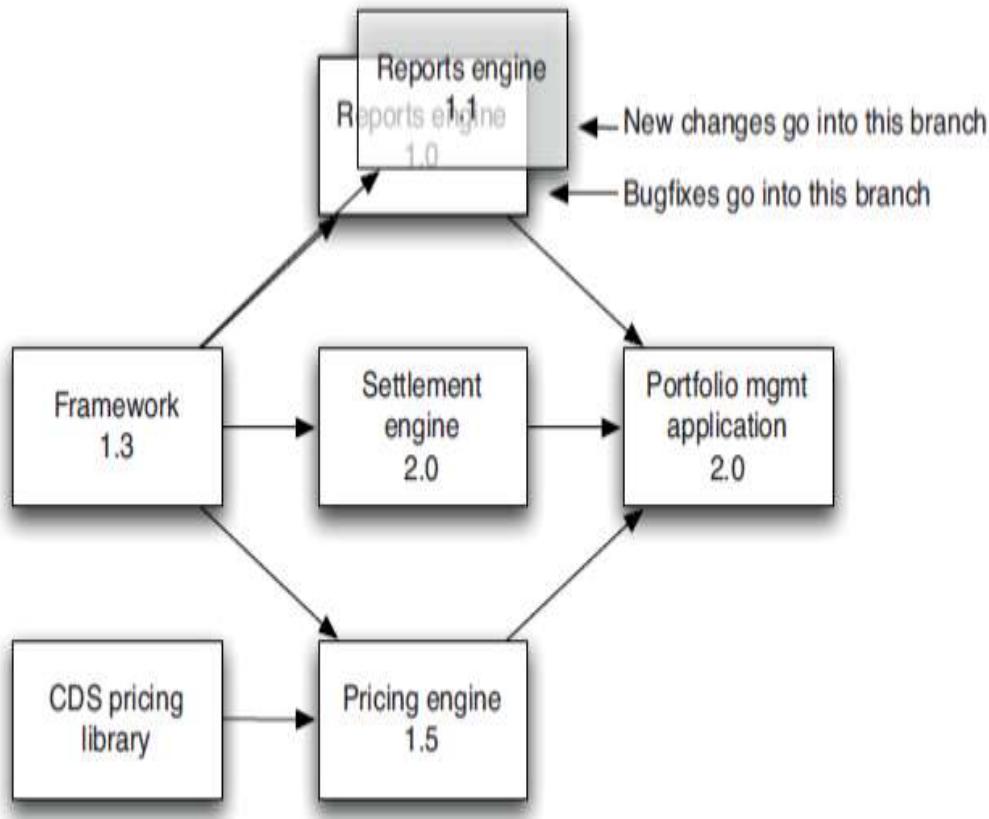
Pipelining Dependency Graphs Contd..



Visualizing downstream dependencies

Managing Component

Pipelining Dependency Graphs Contd..

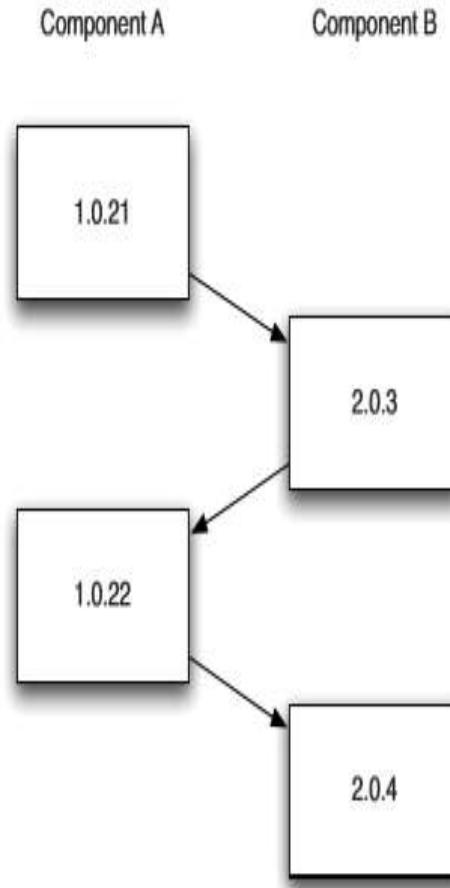
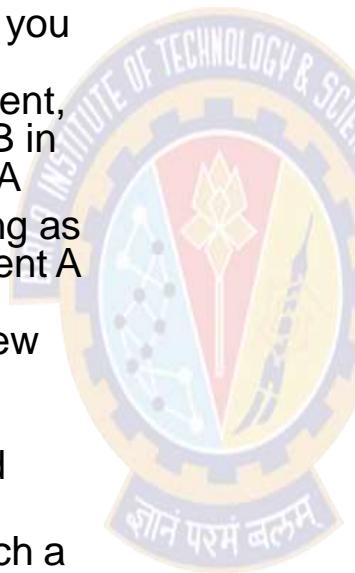


Branching Components

Managing Component

Circular Dependencies

- This occurs when the dependency graph contains cycles
- The simplest example is that you have a component, A, that depends on another component, B, unfortunately component B in turn depends on component A
- to survive this problem so long as there is a version of component A that you can use to build component B then use the new version of B to build the new version of A
- This results in a kind of “build ladder”
- No build system supports such a configuration out of the box, so you have to hack your toolchain to support it.
- also have to be cautious in how the parts of your build interact



**Circular dependency
build ladder**



Thank You!

In our next session:



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps





CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 6

Agenda

Introduction to build

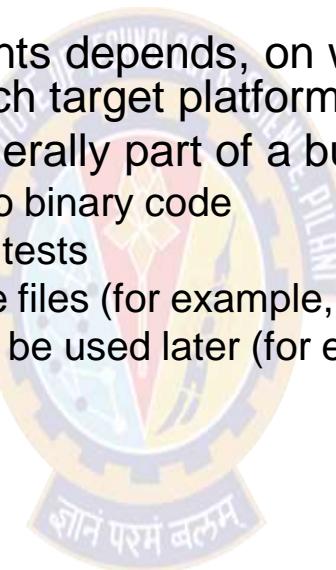
- Build automation and Build tools



Build automation and Build tools

Build tools

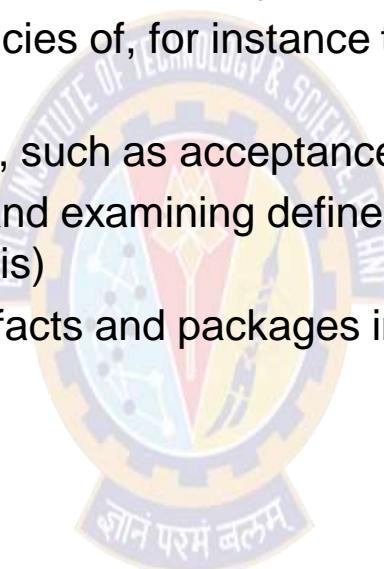
- Build tools automate software builds
- Usually, a software build consists of a number of steps that are dependent on each other
- In detail the sequence of events depends, on which programming language is used and for which target platform software is developed
- The following phases are generally part of a build:
 - Compiling the source code to binary code
 - Running and evaluating unit tests
 - Processing existing resource files (for example, configurations)
 - Generating artifacts that can be used later (for example, WAR or EAR files, Debian packages)



Build automation and Build tools

Build tools Contd..

- Additional steps that are often executed in the a build and that can likewise be automated with the employed build tools are:
 - Administering dependencies of, for instance the libraries that are used in the project
 - Running additional tests, such as acceptance and load tests
 - Analyzing code quality and examining defined conventions in the source code (static code analysis)
 - Archiving generated artifacts and packages in a central repository



Build automation and Build tools

Build tools Contd..

Technology	Tool
Rails	Rake
.Net	MsBuild
Java	Ant, Maven, Buildr, Gradle
C,C++	SCons
• Maven	
• Gradle	



Agenda

Build Tools – Maven and Gradle

- Maven
- Gradle



Build Tool - Maven

History

- Maven, a Yiddish word meaning accumulator of knowledge
- Originally started as an attempt to simplify the build processes in the Jakarta Turbine project
- Concept behind building Maven?
 - There were several projects each with their own Ant build files that were all slightly different and JARs were checked into CVS
 - Need of a standard way to build the projects, a clear definition of what the project consisted of, an easy way to publish project information and a way to share JARs across several projects
- The result is a tool that can now be used for building and managing any Java-based project

Build Tool - Maven

History

- With maven, day-to-day work of Java developers get easier and generally it help with the understanding of any Java-based project
- Maven addresses two aspects of building software
 - describes how software is built
 - describes its dependencies
- Maven can also be used to build and manage projects written in C#, Ruby, Scala, and other languages
- The Maven project is hosted by the Apache Software Foundation, where it was formerly part of the Jakarta Project
- And Initial Release of Maven was : 13 July 2004; 14 years ago

Build Tool - Maven

Software

- Maven is a software which collects all your dependency find it from all over internet compile your classes and provide you the final jar/war
- JAR: Java Archive Files
- WAR: Web Application Resource or Web Application Archive Files
- EAR: Enterprise Archive



Build Tool - Maven

Objective

- Making the build process easy
- Providing a uniform build system
- Providing quality project information
- Providing guidelines for best practices development
- Allowing transparent migration to new features



Build Tool - Maven

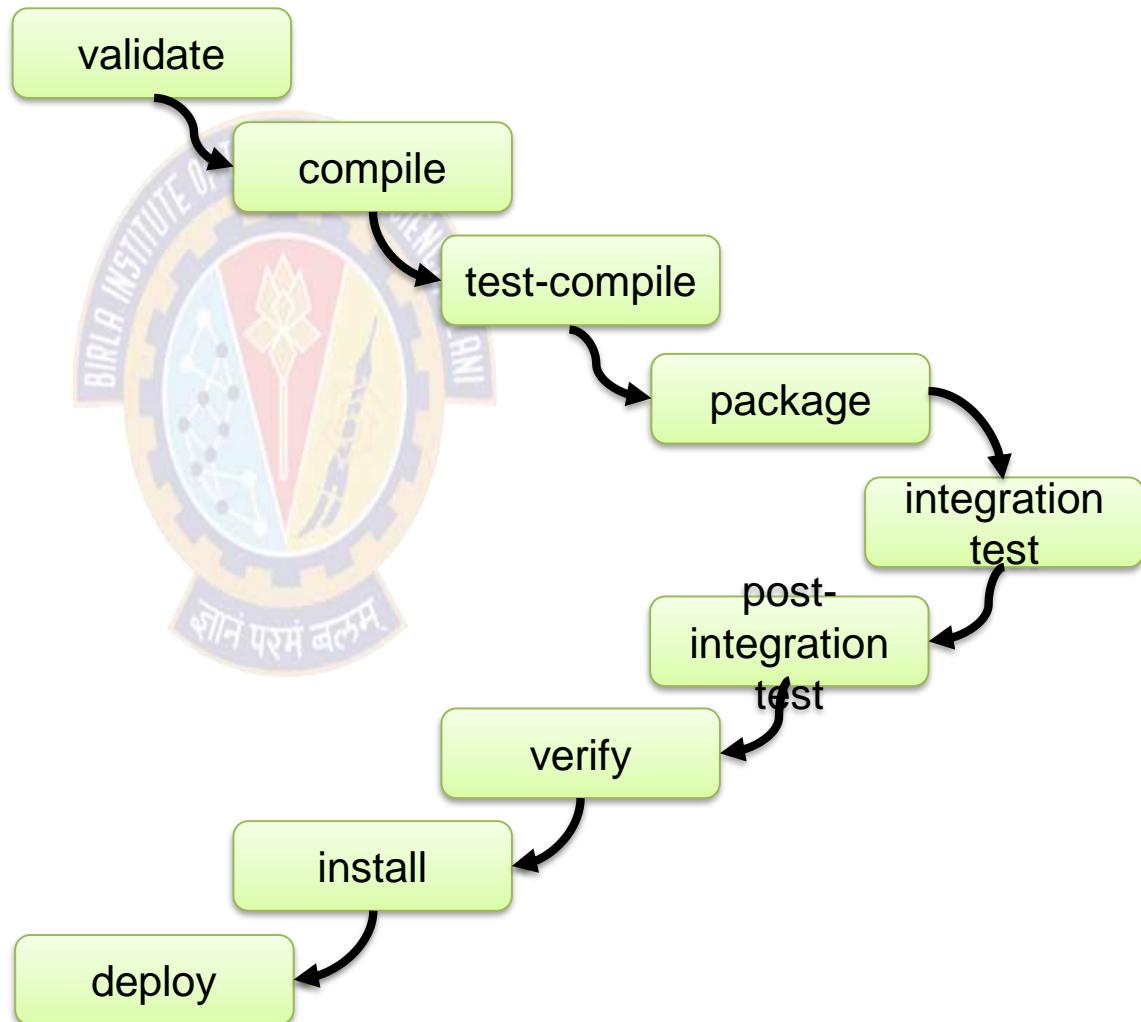
Phases of a standard Maven build

- Currently, Maven is likely the most widely used build tool for Java especially in large enterprises
- Maven default life cycle comprises a number of predefined and fixed phases
 - First, the source files are verified, and the build is initialized
 - Subsequently, the sources and resources are compiled and processed
 - The same happens with test sources and resources
 - Afterwards, the result is packaged, and optionally an integration test is run
 - Finally, the result can be verified, installed, and deployed in an artifact repository
- Not all actions are necessarily executed in each phase
- In this way Maven offers a skeleton along which your own build logic can be oriented

Build Tool - Maven

Phases of a standard Maven build

- Standardized phases, the commands for the compilation and testing of software are the same in each Maven project:
- *mvn package* executes all phases up to packaging and creates a build
- *mvn test* stops at the phase “test.”



Build Tool - Maven

Maven POM

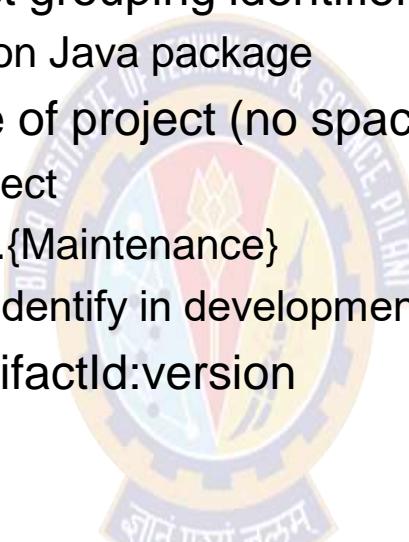
- Stands for Project Object Model
- Describes a project
 - Name and Version
 - Artifact Type
 - Source Code Locations
 - Dependencies
 - Plugins
 - Profiles (Alternate build configurations)
 - Uses XML by Default



Build Tool - Maven

Project Name (GAV)

- Maven uniquely identifies a project using
- groupId: Arbitrary project grouping identifier (no spaces or colons)
 - Usually loosely based on Java package
- artifactId: Arbitrary name of project (no spaces or colons)
 - version: Version of project
 - Format {Major}.{Minor}.{Maintenance}
 - Add '-SNAPSHOT' to identify in development
- GAV Syntax: groupId:artifactId:version



```
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>

  <groupId>Maven_BITS_Training</groupId>
  <artifactId>Maven_BITS_Training</artifactId>
  <version>0.0.1-SNAPSHOT</version>
```

Build Tool - Maven

Packaging

- Build type identified using the “packaging” element
- Tells Maven how to build the project
- Example packaging types: – pom, jar, war, ear, etc.
- Default is jar



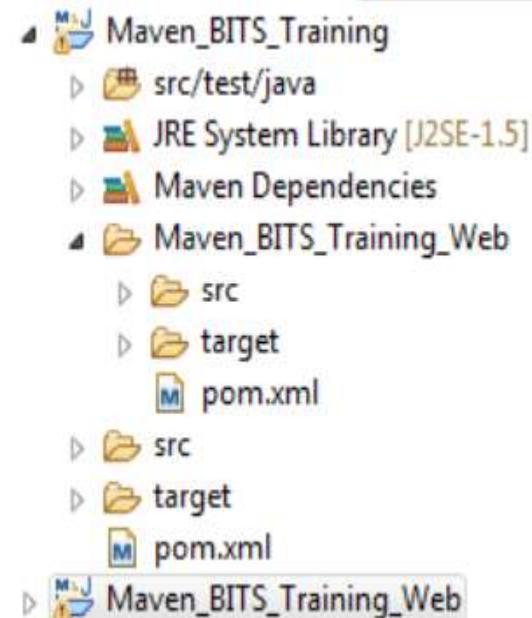
```
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>

  <groupId>Maven_BITS_Training</groupId>
  <artifactId>Maven_BITS_Training</artifactId>
  <version>0.0.1-SNAPSHOT</version>
  <packaging>pom</packaging>
```

Build Tool - Maven

Multi Module Projects

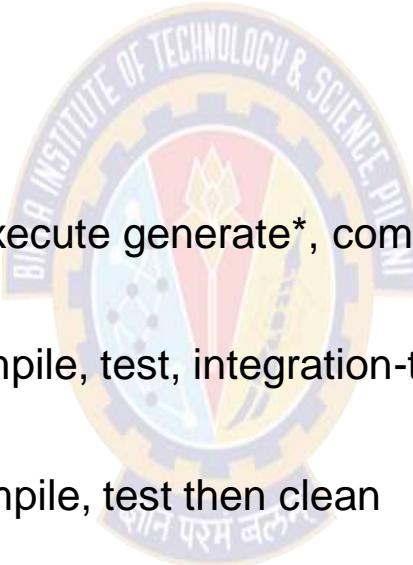
- Maven has 1st class multi-module support
- Each maven project creates 1 primary artifact
- A parent pom is used to group modules



Build Tool - Maven

Example Maven Goals

- *mvn install*
 - Invokes generate* and compile, test, package, integration-test, install
- *mvn clean*
 - Invokes just clean
- *mvn clean compile*
 - Clean old builds and execute generate*, compile
- *mvn compile install*
 - invokes generate*, compile, test, integration-test, package, install
- *mvn test clean*
 - Invokes generate*, compile, test then clean



Build Tool - Gradle

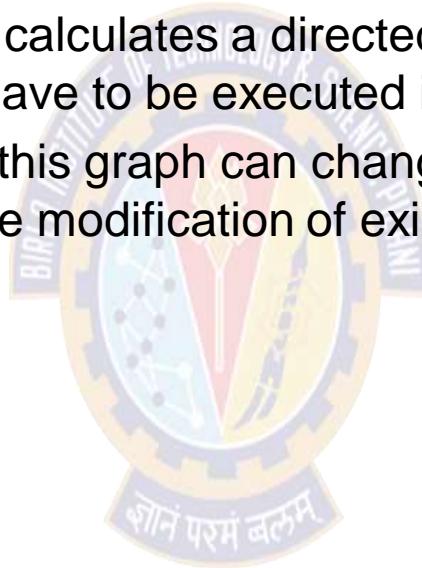
About Gradle

- Gradle is an open-source build automation tool focused on flexibility and performance
- Gradle build scripts are written in Domain Specific Language [DSL], i.e. Groovy
- Groovy resembles Java
- Every Java program is also a valid Groovy program, Groovy offers options for writing programs with substantially easier syntax
- Why Consider Gradle?
 - High performance
 - It is very specific on running only required tasks; which have been changed
 - Build cache helps to reuse tasks outputs from previous run
 - It has ability to have shared build cache within different machine
 - JVM foundation
 - It has JVM Foundation; where using Java Development Kit is Must
 - Benefits to use of standard Java API's in build logic
 - However it is not limited to Java

Build Tool - Gradle

Gradle concepts

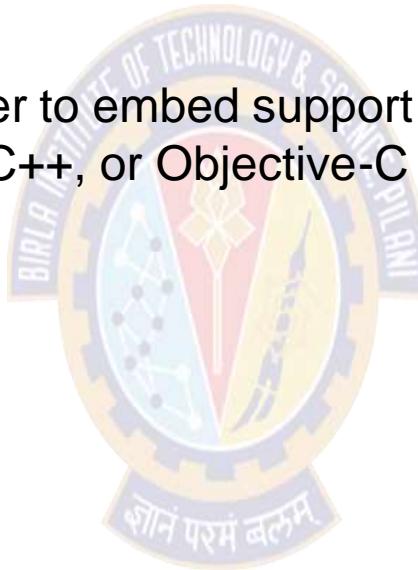
- The core concepts of Gradle are tasks and the dependencies between them
- Based on these, Gradle calculates a directed, acyclic graph to determine which tasks have to be executed in which order
- This is necessary since this graph can change through custom tasks, additional plug-ins, or the modification of existing dependencies



Build Tool - Gradle

Gradle offerings

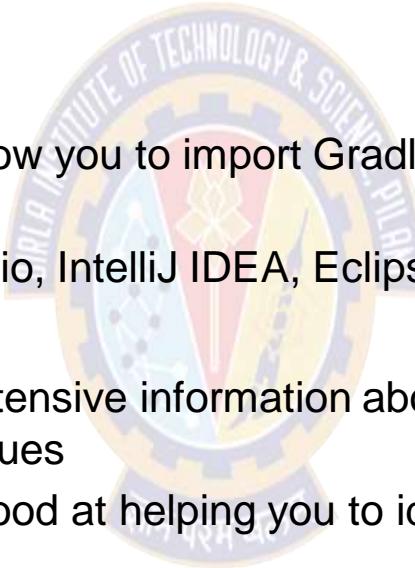
- Gradle also offers the option to extend its functionality by embedding plug-ins
- This allows the developer to embed support of other programming languages like Groovy, C++, or Objective-C in Gradle



Build Tool - Gradle

Gradle benefits

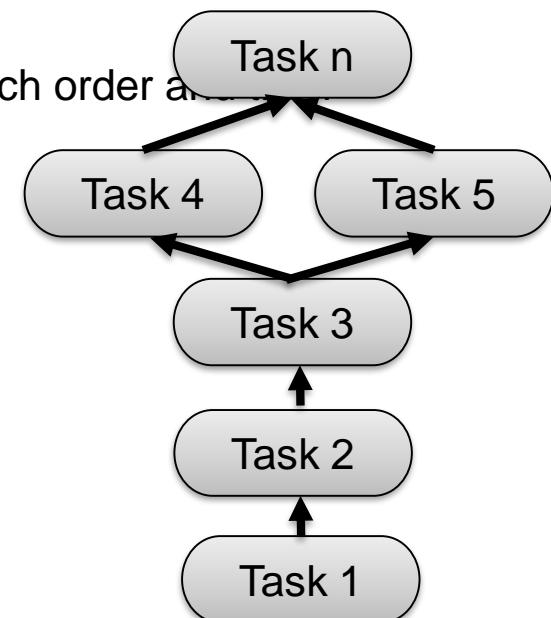
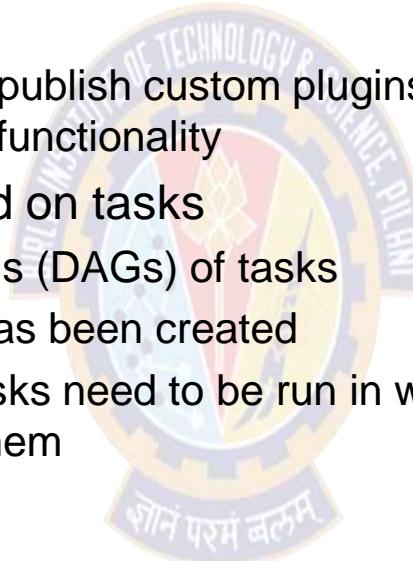
- Extensibility
 - You can readily extend Gradle to provide your own task types or even build model
- IDE Support
 - Several major IDEs allow you to import Gradle builds and interact with them
 - Example: Android Studio, IntelliJ IDEA, Eclipse, and NetBeans
- Insight
 - Build scans provide extensive information about a build run that you can use to identify build issues
 - They are particularly good at helping you to identify problems with a build's performance



Build Tool - Gradle

Gradle DAG

- Gradle is a general-purpose build tool
 - Currently it supports Maven- and Ivy-compatible repositories and the filesystem
 - It supports create and publish custom plugins to encapsulate own conventions and build functionality
- The core model is based on tasks
 - Directed Acyclic Graphs (DAGs) of tasks
 - Once the task graph has been created
 - It determines which tasks need to be run in which order and proceeds to execute them
 - Example of a Graph
 - Tasks Consists of:
 - Action : To perform something
 - Input : values to action
 - Output : generated by



Build Tool - Gradle

Summary

- Gradle has several fixed build phases
 - It evaluates and executes build scripts in Three Phases
 - Initialization : Set-up environment & Identify which Projects will be part of it
 - Configuration: Construct Task graph and then Identify task to be executed in defined order
 - Execution: Run the task
- Gradle is extensible in more ways than one
 - It allows you to Custom Task Types – Define your own Task Types
 - Custom Task Actions
 - Extra Properties on Projects and Tasks
 - Custom Conventions – User friendly

Build Tool

Gradle vs Maven

Gradle	Maven
Flexibility: <ul style="list-style-type: none">Flexibility on ConventionsUser Friendly & Customized	Flexibility: <ul style="list-style-type: none">No flexibility on ConventionsIt is rigid
Performance: <ul style="list-style-type: none">It process only the files has been changedReusability by working with Build CacheShipping is faster	Performance: <ul style="list-style-type: none">It process the complete buildNo Build Cache conceptShipping is slow as compared to Gradle
User Experience: <ul style="list-style-type: none">IDE Support : Is in evolving StageCLI : Modern CLI	User Experience: <ul style="list-style-type: none">IDE Support : Is matureCLI solution is classic in comparison with Gradle

Agenda

Selenium - Test Automation

- Introduction
- Test login page – example
- Selenium Components
- Selenium History
- Selenium Supports
- Selenium Architecture
- Selenium Benefits



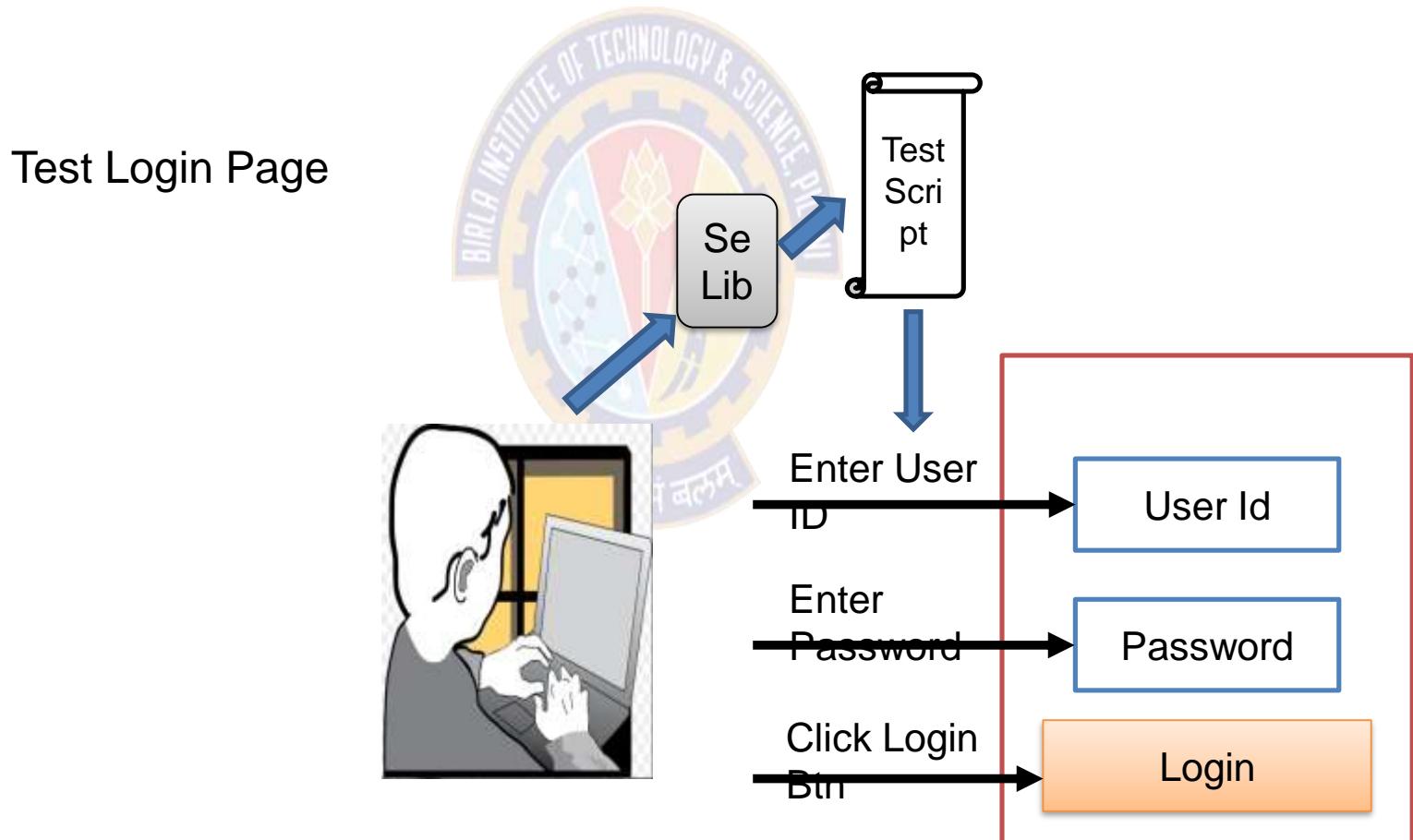
Selenium - Test Automation

Introduction

- Selenium helps to automate web browser interaction
- Scripts perform the same interactions that any user can perform manually
- Selenium can perform any sort of automated interaction; however it was originally designed for automated web application testing
- Common benefits of Test Automation:
 - Frequent regression testing
 - Rapid feedback to developers
 - Virtually unlimited iterations of test case execution
 - Support for Agile and extreme development methodologies
 - Disciplined documentation of test cases
 - Customized defect reporting
 - Finding defects missed by manual testing

Selenium - Test Automation

Example

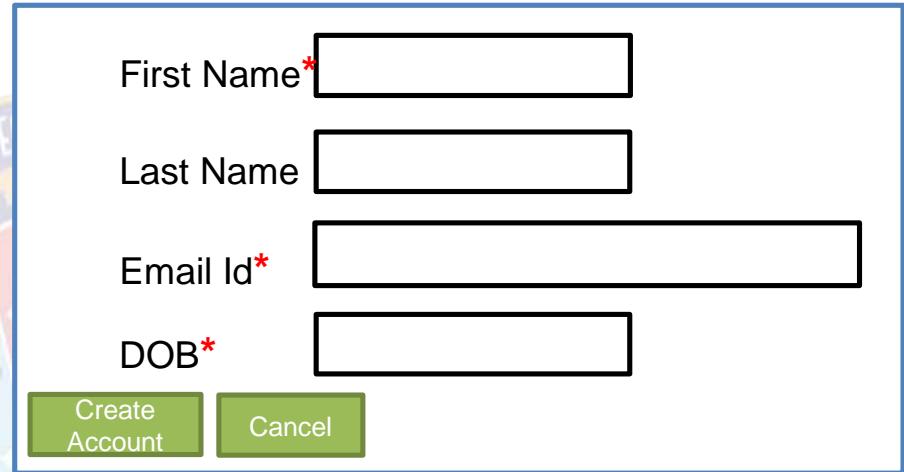
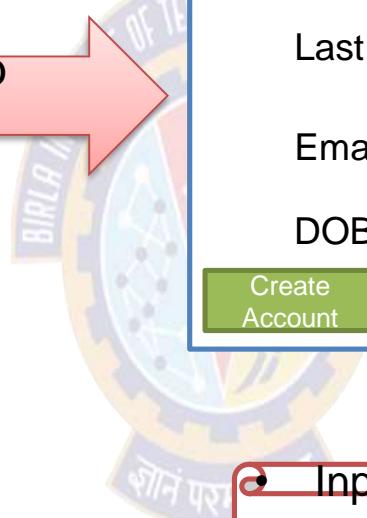


Selenium - Test Automation

Working with Selenium

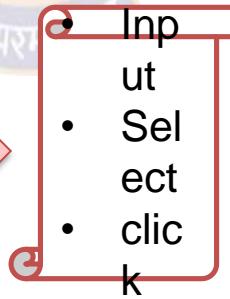
- At a high level you will be doing three things with Selenium

1. Recognize web elements



A screenshot of a registration form. It contains four text input fields labeled "First Name*", "Last Name", "Email Id*", and "DOB*". Below the input fields are two green buttons: "Create Account" and "Cancel". The entire form is enclosed in a blue border.

2. Add Actions (Script using preferred lang)



3. Run the Test

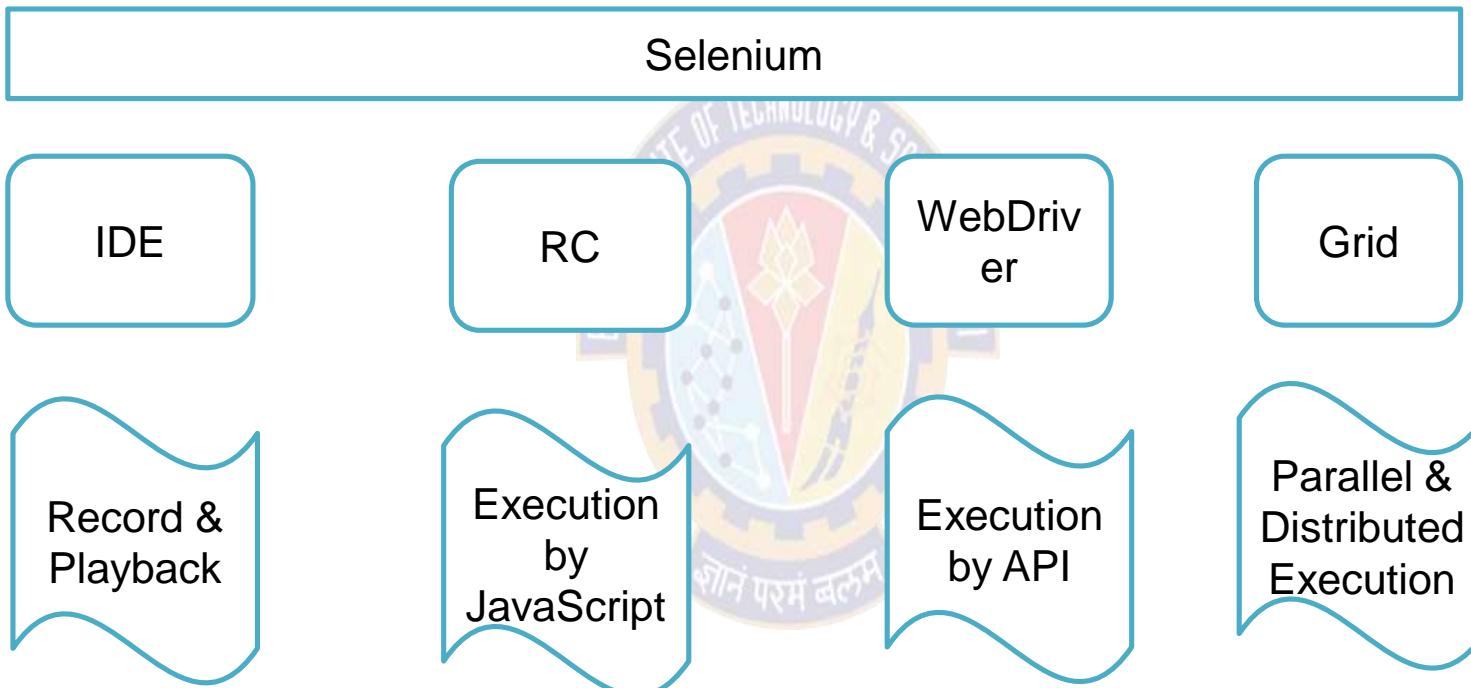
Selenium - Test Automation

Selenium Components

- Selenium IDE
 - A Record and playback plugin for Firefox add-on
 - Prototype testing
- Selenium RC (Remote Control)
 - Also known as selenium 1
 - Used to execute scripts (written in any language) using Javascript
 - Now Selenium 1 is deprecated and is not actively supported
- WebDriver
 - Most actively used component today
 - An API used to interact directly with the web browser
 - Is a successor to Selenium 1 / Selenium RC
 - Selenium RC and WebDriver are merged to form Selenium 2
- Selenium Grid
 - A tool to run tests in parallel across different machines and different browser simultaneously
 - Used to minimize the execution time

Selenium - Test Automation

Selenium Components Contd..



Selenium - Test Automation

Selenium History

- Selenium is a set of different software tools each with a different approach to support test automation
- Selenium support executing one's tests on multiple browser platforms
- Selenium first came to life in 2004 when Jason Huggins was testing an internal application at ThoughtWorks
- It started with a Javascript library
- In 2006 at Google, Simon Stewart started work on a project WebDriver
- The merging of Selenium and WebDriver provided a common set of features for all users and brought some of the brightest minds in test automation under one roof in 2009

Selenium - Test Automation

Selenium - Supports

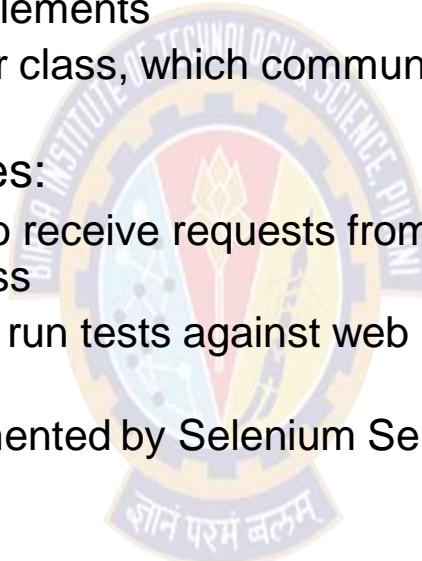
- Browsers
- OS
- Language



Selenium - Test Automation

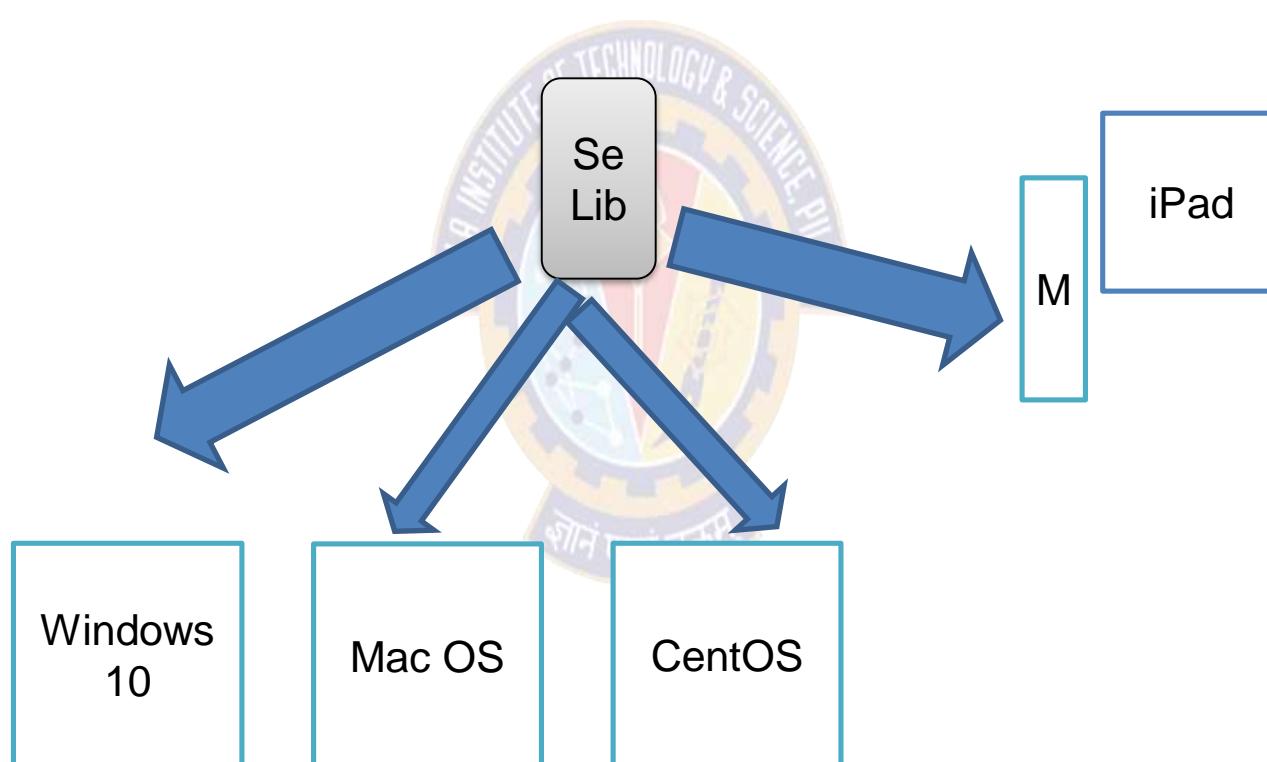
Selenium client-server architecture

- Selenium Client includes:
 - The WebDriver API, which is used to develop test scripts to interact with page and application elements
 - The RemoteWebDriver class, which communicates with a remote Selenium server
- Selenium Server includes:
 - A server component, to receive requests from Selenium Client 's RemoteWebDriver class
 - The WebDriver API, to run tests against web browsers on a server machine
 - Selenium Grid, implemented by Selenium Server in command-line options for grid features



Selenium - Test Automation

Selenium Grid



Selenium - Test Automation

Benefits

- Faster Feedback
 - It enhances communication between product owners, developers and designers
 - Quick fix of malfunctions & bugs in real-time
- Accelerated Results
 - Testing can be executed on a repetitive basis, which in turn will deliver speedier results every time with minimized time and reduced effort
- Reduced Business Expenses
 - Time needed to execute tests will decrease dramatically
 - Larger quantity of work can be undertaken in the same amount of time with greatly increased levels of accuracy
 - This will directly reduce project costs

Selenium - Test Automation

Benefits Contd..

- Testing Efficiency Improvement
 - Testing occupies a considerable part of the whole Software Development Lifecycle
 - So even the smallest improvement in the overall efficiency can result in a massive difference to the overall project timeline
- Reusability of Automated Tests
- Earlier Detection of Defects
 - Early defect detection leads to cost-effective code rework
- Faster Time-to-Market
 - Time to market is a critical factor in product planning and success with regard to software applications
 - Efficiency of test automation will reduce the time of project; which results you to hit market early

References

- <https://www.seleniumhq.org/docs/index.jsp/>



Continuous Code Inspection

Code quality

- Continuous code inspection = Constantly scanning code
- Continuous code inspection is to Identify if any defects
- It is a process of code review; however the goal is to prevent defects
- Its been proved 90% of defects can be addressed using code inspections tools
- Code Inspection & Testing?
 - Testing verifies functionality and improves software quality
 - Testing is expensive if you have to go through it over and over again
 - By minimizing defects during code inspections, you can make your testing efforts more efficient
 - If your code is complex, it may not be fully testable
 - A code inspection, however, can find defects at the code level

Note: Even with automated testing, it takes time to verify functionality; by resolving defects at the code level, you'll be able to test functionality faster

Continuous Code Inspection

Code Inspection Measures

- Code inspections must be well-defined as per requirements:
 - Functional requirements : User Needs : Cosmetic
 - Structural requirements : System Needs : Re-engineering
- Run Time Defects:
 - Identify run time errors before program run
 - Examples: Initialization (using the value of unset data), Arithmetic Operations (operations on signed data resulting in overflow) & Array and pointers (array out of bounds, dereferencing NULL pointers), etc.,
- Preventative Practices:
 - This help you avoid error-prone or confusing code
 - Example: Declarations (function default arguments, access protection), Code Structure (analysis of switch statements) & Safe Typing (warnings on type casting, assignments, operations), etc.,
- Style:
 - In-house coding standards are often just style, layout, or naming rules and guidelines
 - Instead using a proven coding standard is better for improving quality

Continuous Code Inspection

How to Improve Your Code Inspection Process

- Involve Stakeholders
 - Developer, Management & Customer
- Collaborate
 - Collaboration — both in coding and in code inspections
- Recognize Exceptions
 - Sometimes there are exceptions to the rule
 - In an ideal world, code is 100% compliant to every rule in a coding standard
 - The reality is different
- Document Traceability
 - Traceability is important for audits
 - Capture the history of software quality
- What to Look For in Code Inspection Tools
- Automated inspection
- Collaboration system

Agenda

SonarQube - Continuous Code Inspection Tool

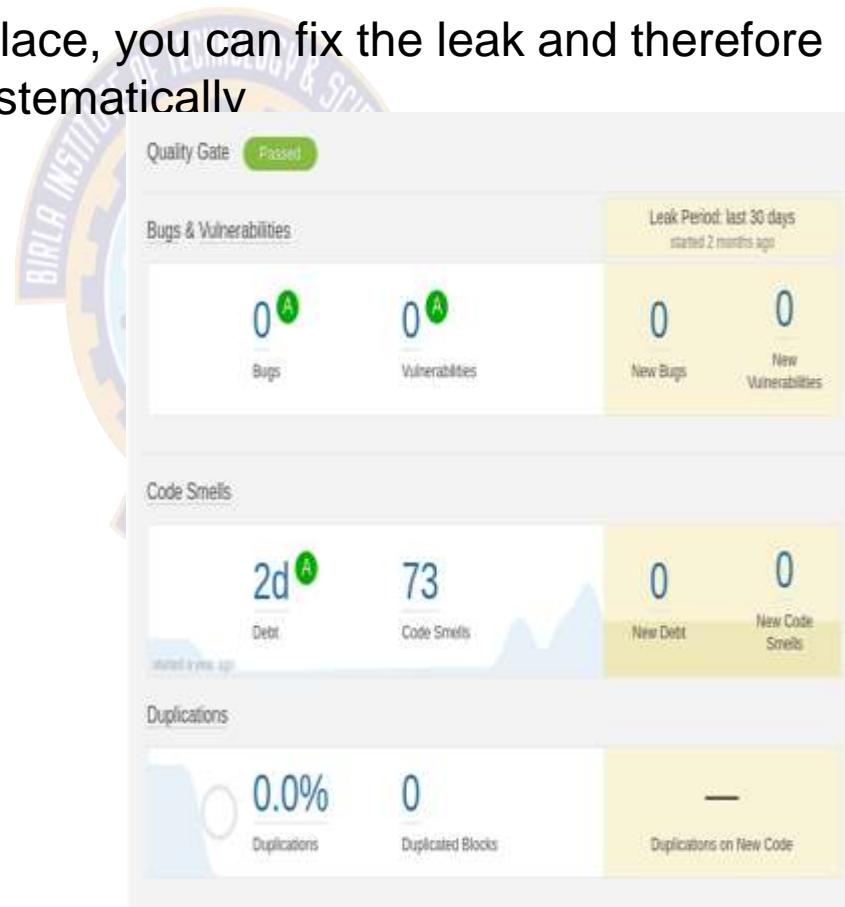
- SonarQube Quality Gate
- Continuous Inspection
- Focus on the Leak
- Enforce Quality Gate
- Benefits of SonarQube



SonarQube - Continuous Code Inspection Tool

SonarQube Quality Gate

- SonarQube provides the capability to not only show health of an application but also to highlight issues newly introduced
- With a Quality Gate in place, you can fix the leak and therefore improve code quality systematically



SonarQube - Continuous Code Inspection Tool

Continuous Inspection

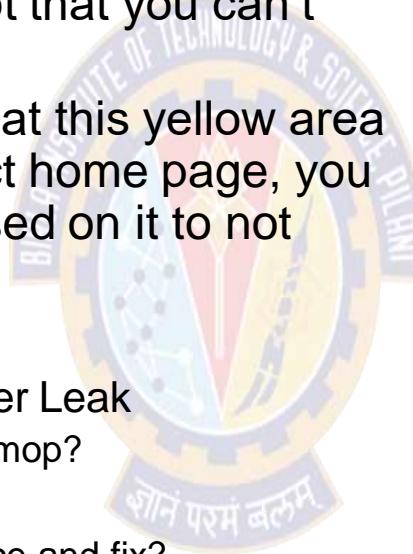
- Overall health
 - Tools shows overall health of the application
 - Also it shows good results achieved over time
- Terminology:
 - Bug:
 - An issue that represents something wrong in the code
 - If this has not broken yet, it will, and probably at the worst possible moment
 - Code Smell:
 - A maintainability-related issue in the code
 - Leaving it as-is means that at best maintainers will have a harder time than they should making changes to the code
 - At worst, they'll be so confused by the state of the code that they'll introduce additional errors as they make changes
 - Vulnerability:
 - A security-related issue which represents a backdoor for attackers



SonarQube - Continuous Code Inspection Tool

Focus on the Leak

- Once that Leak is under control, code quality will start improving systematically
- Leak is a built-in concept that you can't miss
- Once you've had a look at this yellow area on the left of your project home page, you will always remain focused on it to not miss any new issues
- Water Leak Paradigm:
 - Situation: Kitchen Water Leak
 - Do you reach out to mop?
 - OR
 - Do you find the source and fix?

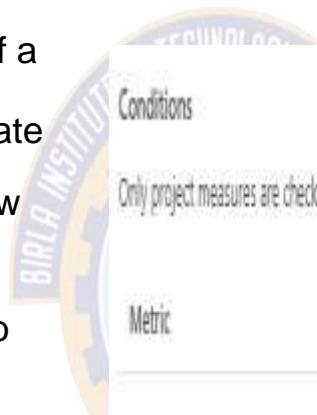


Note: Same logic should be used for Code

SonarQube - Continuous Code Inspection Tool

Enforce Quality Gate

- To fully enforce a code quality practice across all teams, you need to set up a Quality Gate
- A set of requirements that tells whether or not a new version of a project can go into production
- SonarQube's default Quality Gate checks what happened on the Leak period and fails if your new code got worse in this period
- A quality gate is the best way to enforce a quality policy in any organization
- Define a set of Boolean conditions based on measure thresholds against which projects are measured
- It supports multiple quality gate definitions



Metric	Over Leak Period	Operator	Warning	Error
Coverage on New Code	Always	is less than	80.0%	
Duplicated Lines on New Code (%)	Always	is greater than	3.0%	

SonarQube - Continuous Code Inspection Tool

Analyze pull requests

- To shorten the feedback loop, you can set up the analysis of your pull requests
- Analyses will be run on your feature branches without being pushed to SonarQube
- This gives you the opportunity to fix issues before they ever reach SonarQube



Benefits of SonarQube

Branch Analysis

- Track the quality of short-lived and long-lived code branches in SonarQube to ensure that only clean, approved code gets merged into master



References

- <https://docs.sonarqube.org>





Thank You!

In our next session:



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps





CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 7 and 8

Agenda

Using Continuous Integration Software

- Basic Functionality of Continuous Integration Software
- Basic Operations
- How it was before Continuous Integration
- Prevention to be adhere



Continuous Integration Software

Basic functionality of continuous integration software

- Poll version control system
- Check if any commits have occurred
- If so, check out the latest version of the software
- Trigger / Run build script to compile the software
- Run the tests
- And then notify you of the results
- There are many products on the market that can provide the infrastructure for automated build and test process

Continuous Integration Software

Basic Operation

- Continuous Integration server software has two components
 - The first is a long-running process which can execute a simple workflow at regular intervals
 - The usual CI workflow polls revision / version control system at regular intervals
 - If it detects any change, it will check out a copy of project to a directory on the server
 - It will then execute the commands (these commands build an application and run the relevant automated tests)
 - The second provides a view of the results of the processes that have been run, notifies for the success or failure of build and test runs, and provides access to test reports, installers, and so on
 - Most CI servers include a web server that shows you a list of builds that have run
 - Allows look at the reports that define the success or failure of each build

Continuous Integration Software

Bells and Whistles

- Apart from basic operations we can get the status of the most recent build sent to an external device
- Organizations use red and green lava lamps to show the status of the last build
- Even few organization go with all in one health board for all the tools used; where you can see health of each build, git commits and configuration management results
- Some Advanced experiments:
 - Flashing lights and sirens for build progress
 - Text-to-speech to read out the name of the person who broke the build
 - Display the status of the build, along with the avatars of the people who checked in

Note: These are just a great way to allow everyone to see the status of the build at a glance and Visibility is one of the most important benefits of using a CI server

How it was before Continuous Integration

Nightly Build

- It was a common practice for many years
- If I break it; I will monitor until next break by someone
- The idea is that a batch process will compile and integrate the codebase every night when everybody goes home
- If it failed again during the night run, can do changes to fix it however need to wait for next night run
- So verification of System Integration will be on hold until next run

Note: this strategy will not be a good Idea when you have a geographically dispersed team working on a common codebase from different time zones

Prevention to be adhere

Don't Check In on a Broken Build

- Per process if a build breaks, developer team need to fix it ASAP
- If any check-in breaks the build then try to fix it before moving ahead
- At the Broken Stage; nobody should be allowed to check-in
- What if we do Check In on Broken Build:
 - If any new check-in or build trigger during broken state will take much longer time to fix
 - Frequent broken build will encourage team not to care much about working condition

Note: The CI System we are building is to make sure green build and working software

Prevention to be adhere

Commit locally than direct to Production

- Per process a commit triggers the creation of a release candidate
- Try to keep Check-in lightweight
- Generally regular check-in with 20 min interval is a achievable state
- To achieve this we need some commitment from Development Team
 - Commit locally
 - Always keep refreshed local copy
 - If feasible run local build and commit tests
 - If successful; then push to Production
- Why Check-in locally and Commit test run?
 - If there are check-in's before our last update; high chance of conflict :: So if we check out and run tests locally, we have chance to identify problem without breaking the build
 - If new artifacts inclusion has been missed in the Check In, it will result to common errors :: running locally will help to Identify missing configuration

Prevention to be adhere

Wait for Commit Tests to Pass before Moving On

- The CI System is a Shared resource
- At time of Check-in, we should monitor the build progress – Its Developers default duty
- Until the check-in is compiled and passed commit tests, Do not start any new task
- Its not good to say however “Developer should not even leave out for Lunch or Meeting”
- If the commit succeeds the developers are then and only then free to move
- If it fails “Either fix it or Revert it to previous working version”
- Never go home with broken build

Prevention to be adhere

Wait for Commit Tests to Pass before Moving On Contd..

**Geo – Distributed
Teams Working on
CI**



**Hardly it will be 3 hrs
offline time to fix
build**

Prevention to be adhere

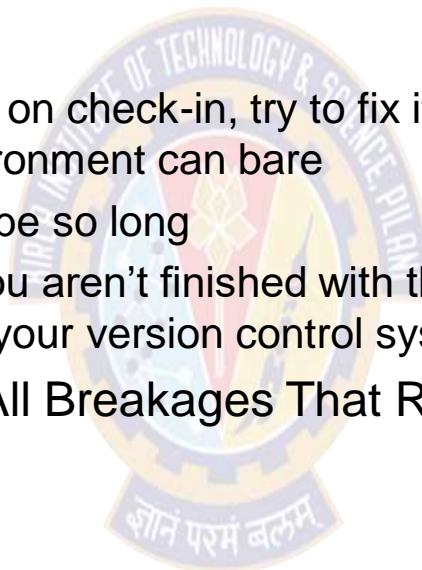
Always Be Prepared to Revert to the Previous Revision

- In a scenario of failed commit stage, it is important that you get everything working again quickly
- If you can't fix the problem quickly, for whatever reason, you should revert to the previous change-set held in revision control
- And remedy the problem in our local environment
- Example:
 - Airplane pilots are taught that every time they land, they should assume that something will go wrong
 - So they should be ready to abort the landing attempt and “go around” to make another try
 - Use the same mindset when checking in
 - Assume that it may break something that will take more than a few minutes, and know what to do to revert the changes and get back to the known-good revision in version control

Prevention to be adhere

Always Be Prepared to Revert to the Previous Revision Contd..

- If you try to revert every time then; how can you make progress?
- Time Boxing
 - Establish a team rule
 - When the build breaks on check-in, try to fix it for ten minutes or any approximate time environment can bare
 - However it should not be so long
 - If, after ten minutes, you aren't finished with the solution, revert to the previous version from your version control system
- Take Responsibility for All Breakages That Result from committed Changes



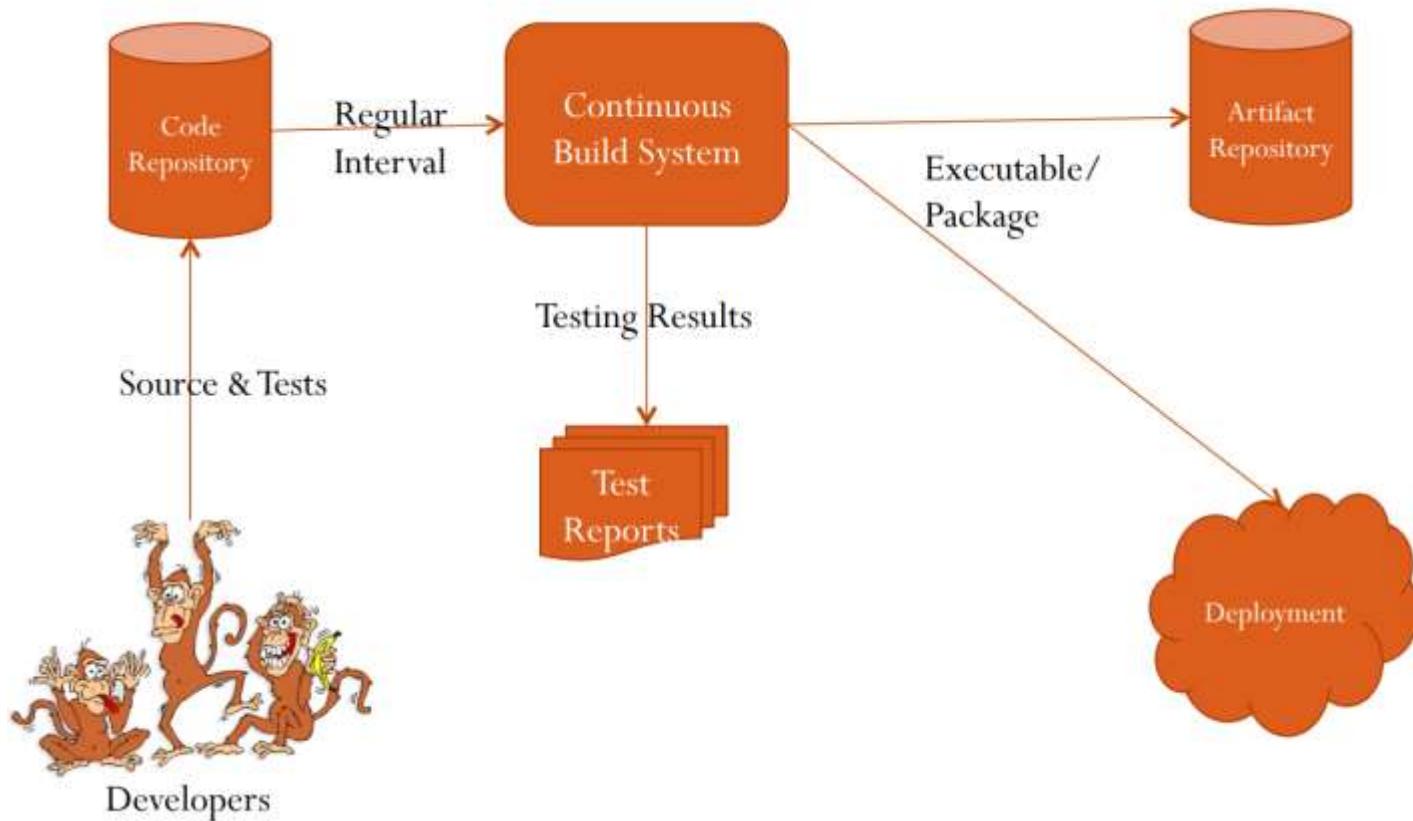
“Continuous Integration is a software development practice where members of a team integrate their work frequently, usually each person integrates at least daily - leading to multiple integrations per day. Each integration is verified by an automated build (including test) to detect integration errors as quickly as possible”

~Martin Fowler

What does CI Mean

- At a regular frequency (ideally at every commit), the system is:
- **Integrated:** All changes up until that point are combined into the project
- **Built:** The code is compiled into an executable or package
- **Tested:** Automated test suites are run
- **Archived:** Versioned and stored so it can be distributed as is, if desired
- **Deployed:** Loaded onto a system where the developers can interact with it

CI workflow





Benefits of using CI and CD

- Reduces overhead across the development and deployment process
- Reduces the time and effort for integrations of different code changes
- Enables a quick feedback mechanism on every change
- Allows earlier detection and prevention of defects
- Helps collaboration between team members so recent code is always shared
- Reduces manual testing effort
- Building features more incrementally saves time on the debugging side so you can focus on adding features
- First step into fully automating the whole release process
- Prevents divergence in different branches as they are integrated regularly
- If you have a long running feature you're working on, you can continuously integrate but hold back the release with **feature flags**

Best practice workflow

1. Maintain a code repository
2. Automate your build
3. Make your build self-testing
4. Daily commits to the baseline by everyone on the team
5. Every commit (to the baseline) should be built
6. Keep your builds fast
7. Clone the production environment and test there
8. Make it easy to get the latest deliverables
9. Everyone on the team can see the results of your latest build
10. Automate build deployment

CD checklist

1. Before submitting changes, check to see if a build is currently in the "Successful" status. If not, you should assist in fixing a build before submitting new code.
2. If the status is currently "Successful", you should rebase your personal workspace to this configuration.
3. Build and test locally to ensure the update doesn't break functionality.
4. If Successful, check in new code.
5. Allow CI to complete with new changes.
6. If build **fails**, **stop** and **fix** on your machine. Return to step 3.
7. If build **passes**, **continue** to work on the next item.



Areas of CD Process

- Source Control
- Build Process
- Testing & Q&A
- Deployment
- Visibility

CD Maturity Matrix

	Novice	Beginner	Intermediary	Advanced	Expert
Build	Verification before commit run in developer's Workspace Common nightly build	CI server builds on commit Artifacts are managed	No build scripts -only configurations Dependencies are managed	Distributed builds Staged build sequence	Build from VM CI server orchestrate VMs
Test + QA	Unit Test Code Coverage	Metrics on technical debt & compliance Mock-up's & proxies	Peer-reviews Automated Functional Test	Test Data Test in target	Automated Acceptance Test
SCM	"Early Branching" Branches used for releases Merges are rare	"Late branching" Branches used for work isolation Merges are common	Pre-tested Commits Integration branch is pristine	All commits are tied to tasks Individual history rewrites in DVCS	Release notes & traceability analysis are generated automatically
Visibility	Build status is notified to committer	Latest build status is available to all stakeholders	Trend reports Build status can be subscribed to (pull vs push)	Monitors in work areas show real-time status	Build reports and statistics are shared with customer and public

Automation - Principles of Continuous Delivery



Principle #1

Every build is a potential release



Principle #3

Automate wherever possible



Principle #2

Eliminate manual bottlenecks



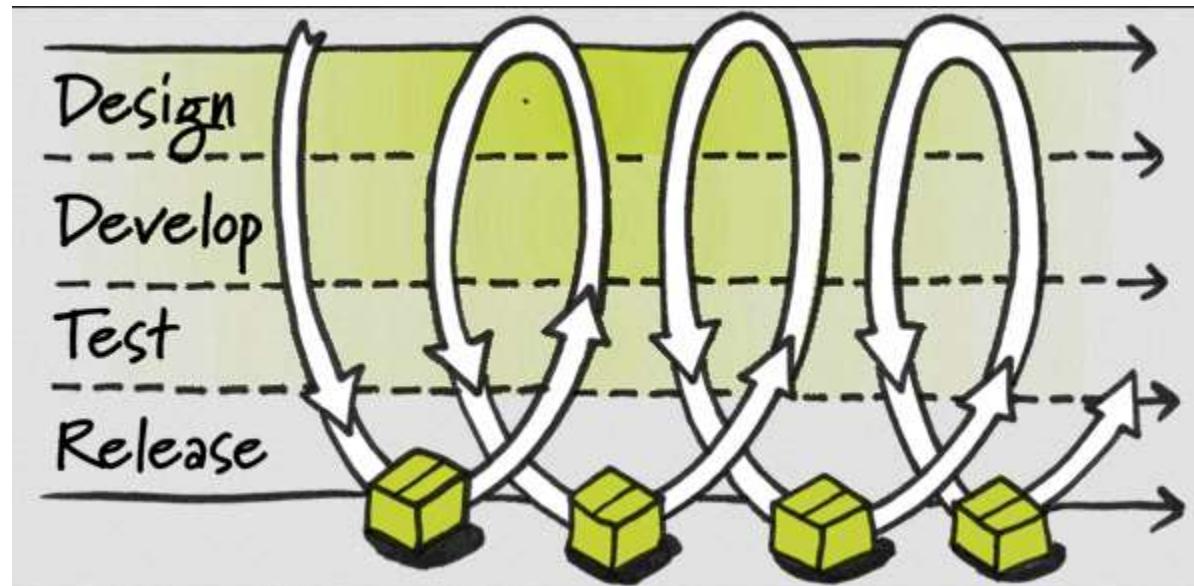
Principle #4

Have automated tests you can trust



Continuous Delivery

- Recall: Build software that is always ready to be deployed into production



Continuous Delivery cont.

- Is this a new idea?
- 1st principle of the Agile Manifesto:
 - “Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.”
- Goal: Release to production more often
 - Monthly
 - Weekly
 - Daily

Continuous Delivery cont.

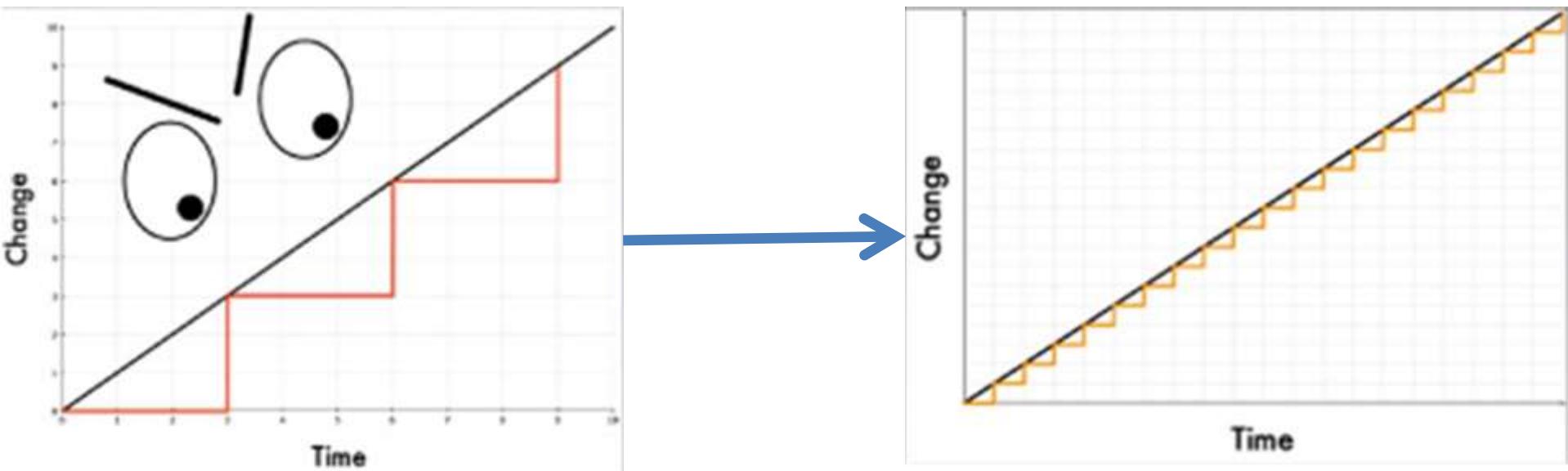
- Pros: Receive many benefits from the Agile process as well as others
 - Build the right product
 - Constant feedback from customers
 - Write thorough acceptance tests
 - Earlier benefits
 - Get product out before competitors
 - Find bugs earlier

Continuous Delivery cont.

- Ability to react quickly to change
 - Not a huge amount of time and money can be abandoned to adopt new requirements
- Save money and time
 - Save money and time if project goes bad
 - Save money and time with automation
 - As deploying to production becomes easier, more time can be spent developing valuable features

Continuous Delivery cont.

- Pros cont:
 - Reliability



Practice Continuous Delivery

- Basics to start practicing Continuous Delivery
 - Configuration Management
 - What needs to be kept under CM?
 - Everything!
 - Code
 - Tests
 - Configuration Files
 - Build Scripts
 - Environments
 - Documentation
 - Etc.

CM Branching

- Stay away from branching except in special cases
 - Branch on releases
 - Horror story example
- You must always check into the trunk!
 - Otherwise you're not continuously integrating



CM Branching

- What if your project team is releasing to production every week, but you're working on a feature that will take longer than the release cycle to implement?
- First option:
 - Gradually release feature into production
- Second option:
 - Feature toggle

Managing Environments

- Must have multiple environments when developing software
- Need to be able to duplicate environments with ease
- Environments configurations to take consideration of:
 - Operating systems including their framework and settings
 - Packages needed to be installed for the application to function properly
 - Network settings

Managing Environments

- Tools to help configure environments in an automated fashion:



Continuous Integration

- Vital step when practicing CD
- Recall: To practice CI, every developer on a project must integrate their work daily with every other developer
 - Everyone on the team needs to practice this for it to work
- Continuous integration is not a tool, but a technique
 - But there are many open source tools to help practice CI

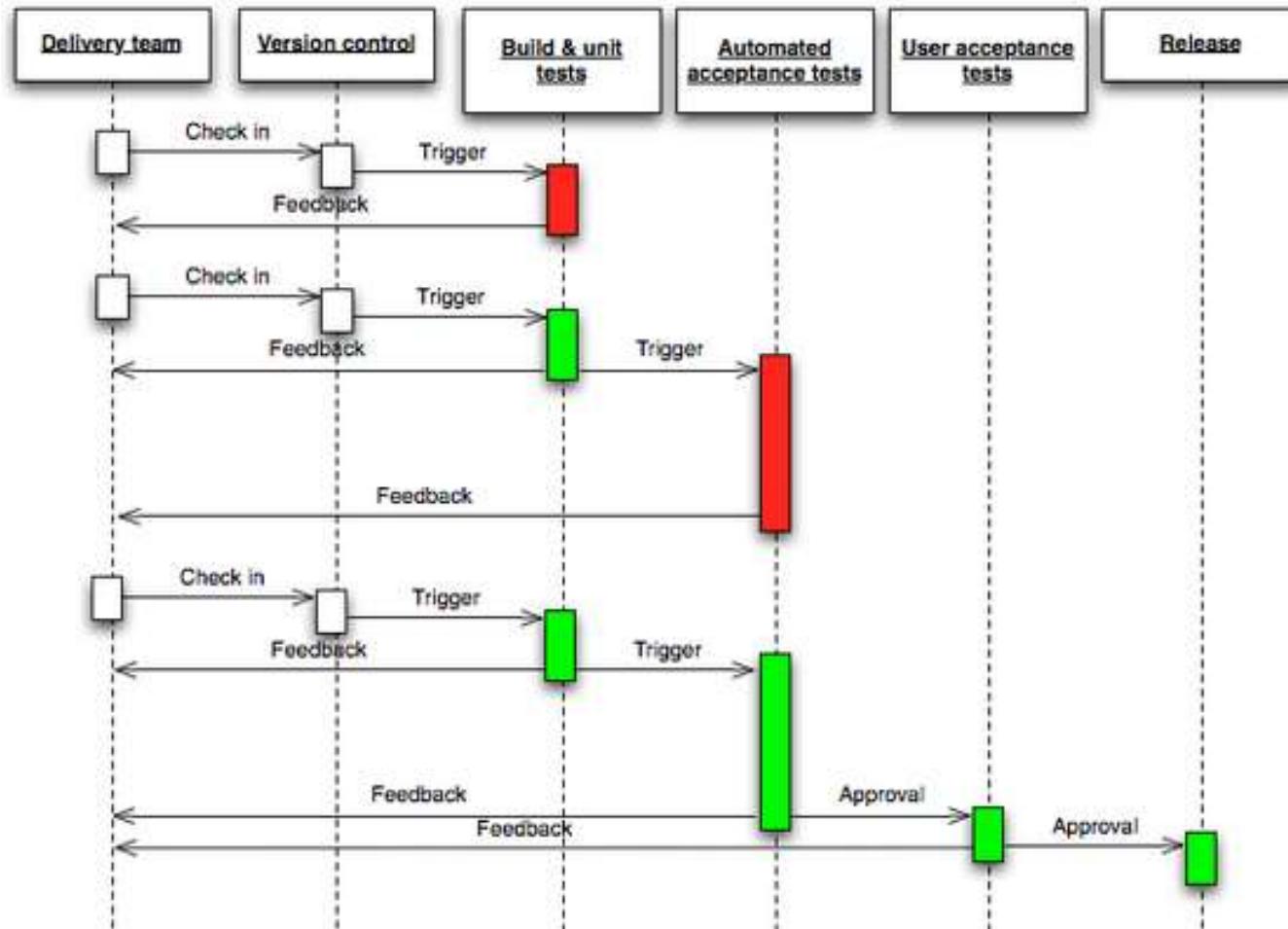
Continuous Integration

- Things a CI server can do:
 - After you check-in code it can:
 - Run build/test scripts
 - Notify developers of a failed build or failed tests
 - Can show all past check-ins and if they failed or passed all the tests
 - Can keep track of multiple projects
 - Plus much more!!

Deployment Pipeline

- Heart of Continuous Delivery
- Will give immediate feedback for how ready your software is for production
- Automate everything(unit tests, acceptance tests, performance tests,...) to the point of just being able to click a button to deploy to production

Deployment Pipeline cont.



Deployment Pipeline in Go

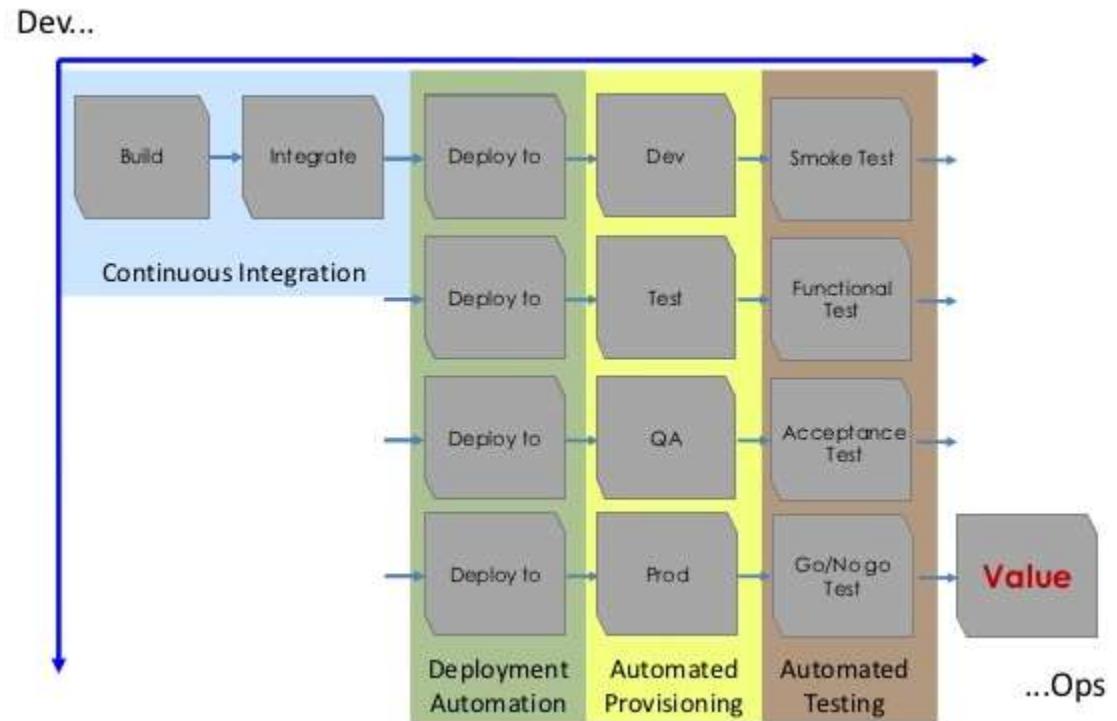
Pipeline Activity

	Commit	Dist	Smoke	Acceptance	Dist-All	Performance
2.3.0.6689 revision: a3a5f8920c48... 1 day ago Triggered by ssudheen						
2.3.0.6688 revision: 472ea60c48ac... 5 days ago Triggered by changes						
2.3.0.6687 revision: 5780e01322cf... 5 days ago Triggered by changes						
2.3.0.6686 revision: 733818836c8b... 5 days ago Triggered by changes						
2.3.0.6685 Git - trunk - https://cruise_builder:*****@blinstdscm01.thoughtworks.com/git/cruise.git jmm <jmonahan@thoughtworks.com> Prototypes: v6 finished 733818836c8b85cdab463406d75558c548e2d Mercurial - twist - https://cruise_builder:*****@blinstdscm01.thoughtworks.com/hg/cruise_qa Rajeshvaran Appasamy <rappasamy@thoughtworks.com> Committing the completed script for #5349 477813cbecd8865d6ec3042b2a12aa1b5c95605 Triggered by changes						

Automation – Automate all the things

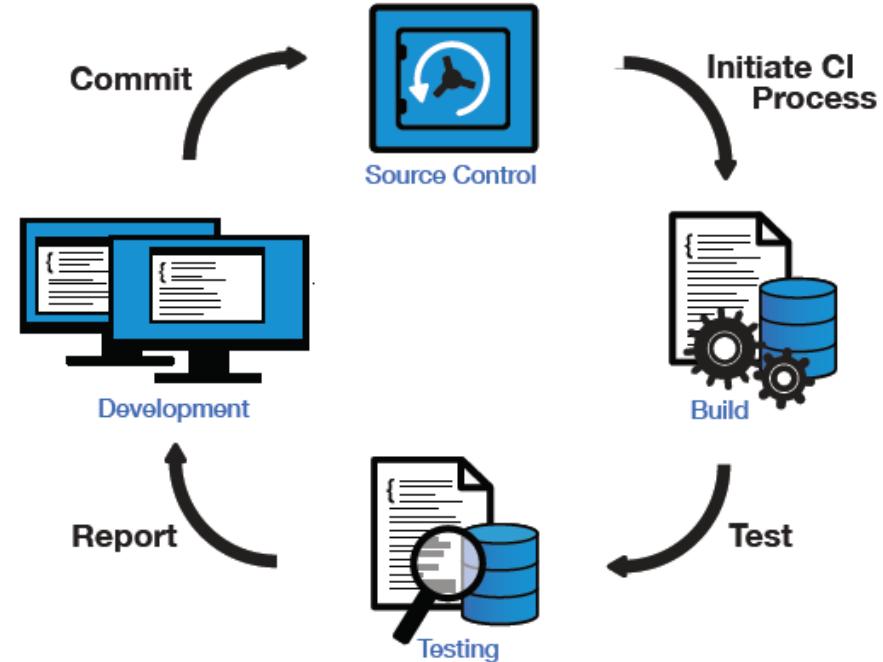
- Reproducible Build
 - Build on a Prod-Like environment
 - No more “Works on my machine”
- Test
 - Testing reduces risks
 - Make you more confident
- Deploy
 - Deploy to Dev, Deploy to QA
 - Deploy to Pre-Prod, Prod

Automation – Continuous Deployment



Automation – Continuous Integration

- Build each “Commit” – check the integration code
- Static Code Analysis
- Perform set of Tests – Unit Test ,Automation Tests , Regression, TDD and compatibility Test
- Nightly and Weekly
- Packaging
- Delivery
- Deploy



Automation – Continuous Deployment

- Deployment Automation/Continuous Delivery
 - Distributing the components of the releasable package into target environment (Dev, Test, Prod)
- Provisioning
 - Create and readying the target Env. And middleware required by your application.
- Automated Testing
 - Verify:
 - Functional
 - Non functional
- Feedback



TESTS = DEPLOYMENT HEALTH



Automation – IAC

Infrastructure As A Code

- Put Infra under **version control**
- History , Melodize your infra
- Deploy monitoring , backups and apps
- **IT IS NOT SCRIPTING !**
- Think Puppet, Chef, Ansible, Udeploy....

Automation – High Availability



Automation – Scalability



Measurements



Measurements - Metrics

- Easy to Create new Metrics
- Build Dashboards
- Learn from logs
- Learn from *

Measurements - Monitoring

- Monitor Every Platform – Dev, QA, Prod, see troubles early
- Run Time:
 - OS
 - Disk, CPU, I/O, Memory
 - Middleware
 - Queues
 - API calls
 - Connections
 - Application
 - Response time
 - Users
 - Objects
 - Usage

Measurements - Monitoring





Measurements – Measure Everything

- Deployments
- Commits
- Tickets
- Bugs

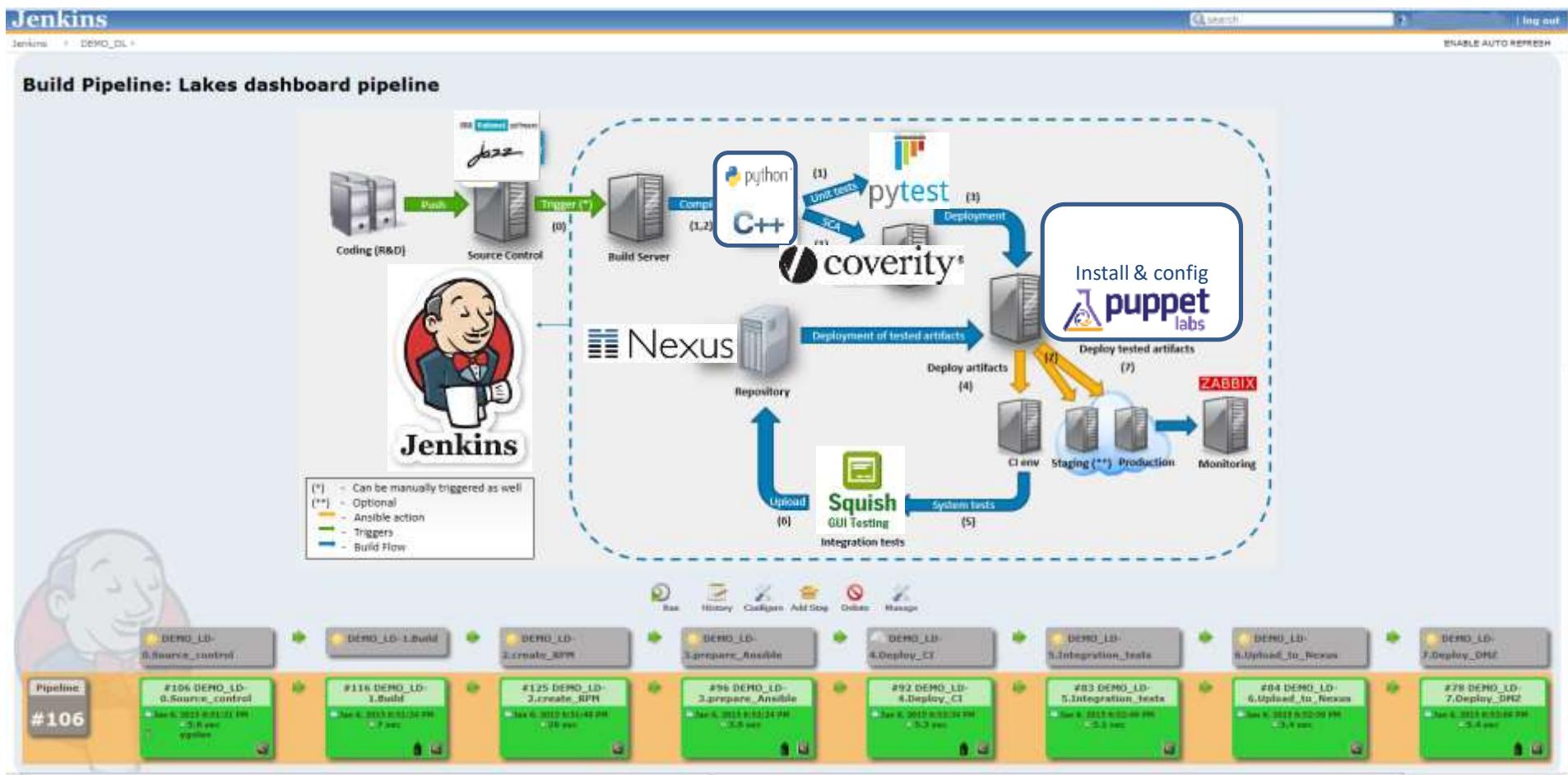
Measurements – Read Metrics

- Get Business Keys
- Share dashboard with dev ops mgmt.
- Find metrics that matter
- Learn from metrics

Sharing

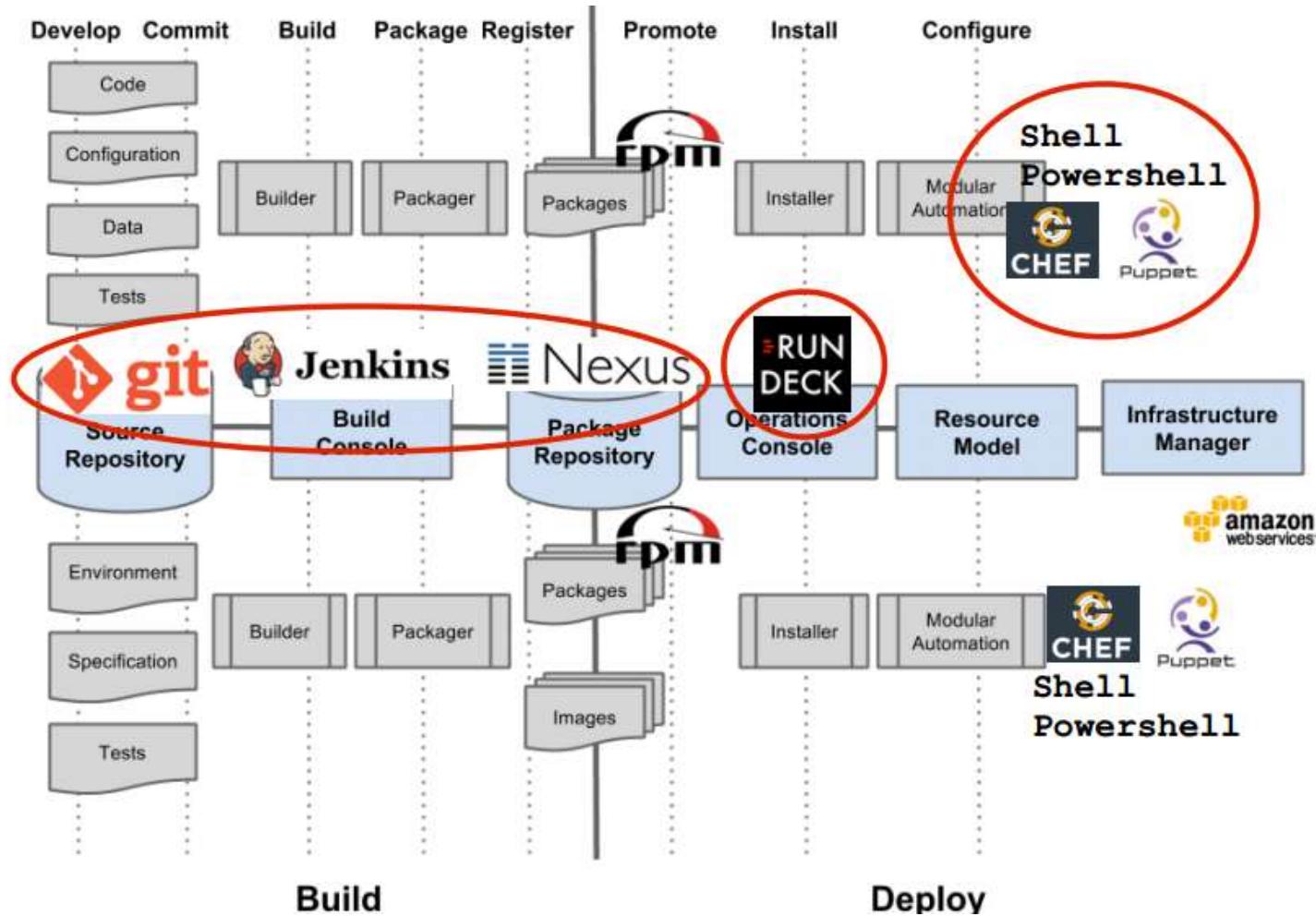


Implementation example



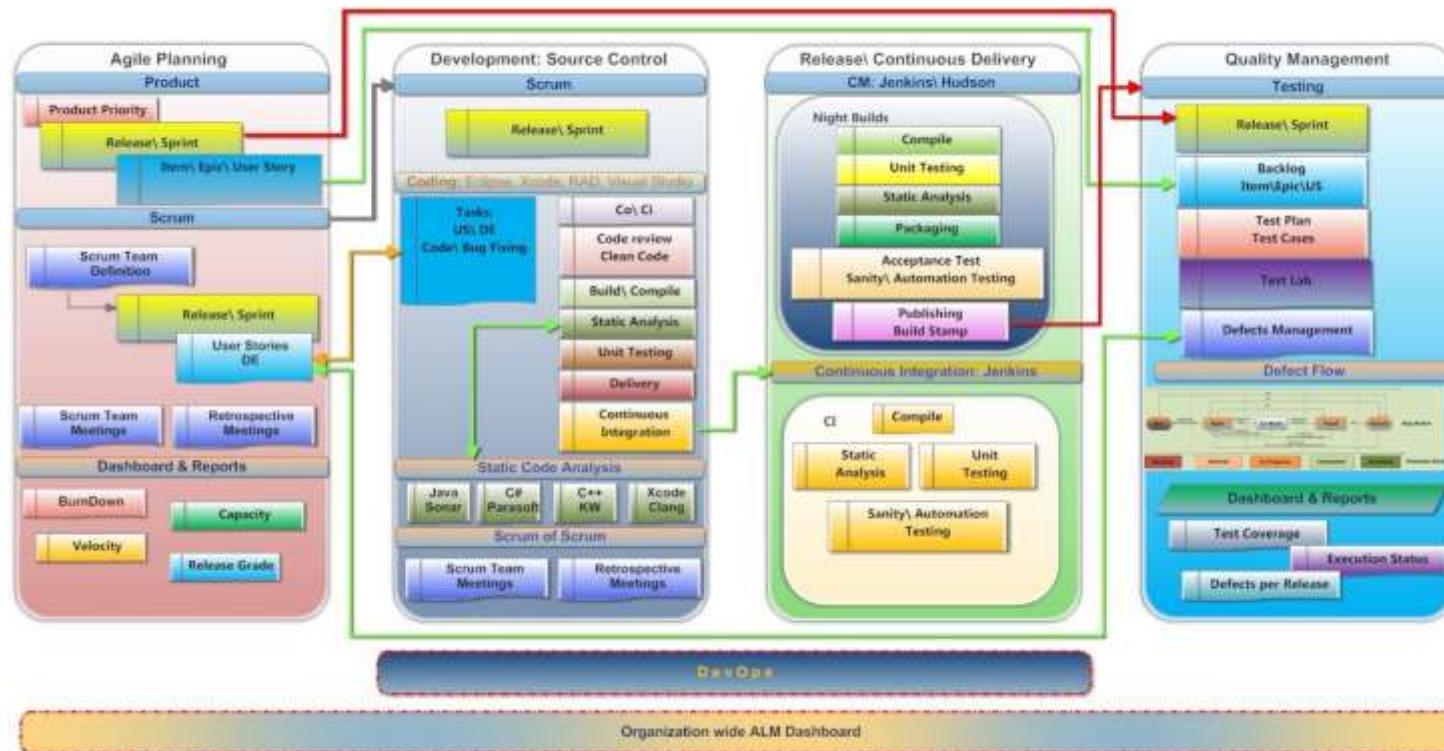
Application

Infrastructure



Full ALM Solution - CI

ALM and Agile



Jenkins

- Jenkins, a continuous build tool, enables teams to focus on their work by automating the build, artifact management, and deployment processes
- Jenkins' core functionality and flexibility allow it to fit in a variety of environments and can help streamline the development process for all stakeholders involved

CI Tools

- Code Repositories: SVN, Mercurial, Git
- Continuous Build Systems: Jenkins, Bamboo, Cruise Control
- Test Frameworks: JUnit, Cucumber, CppUnit
- Artifact Repositories: Nexus, Artifactory, Archiva

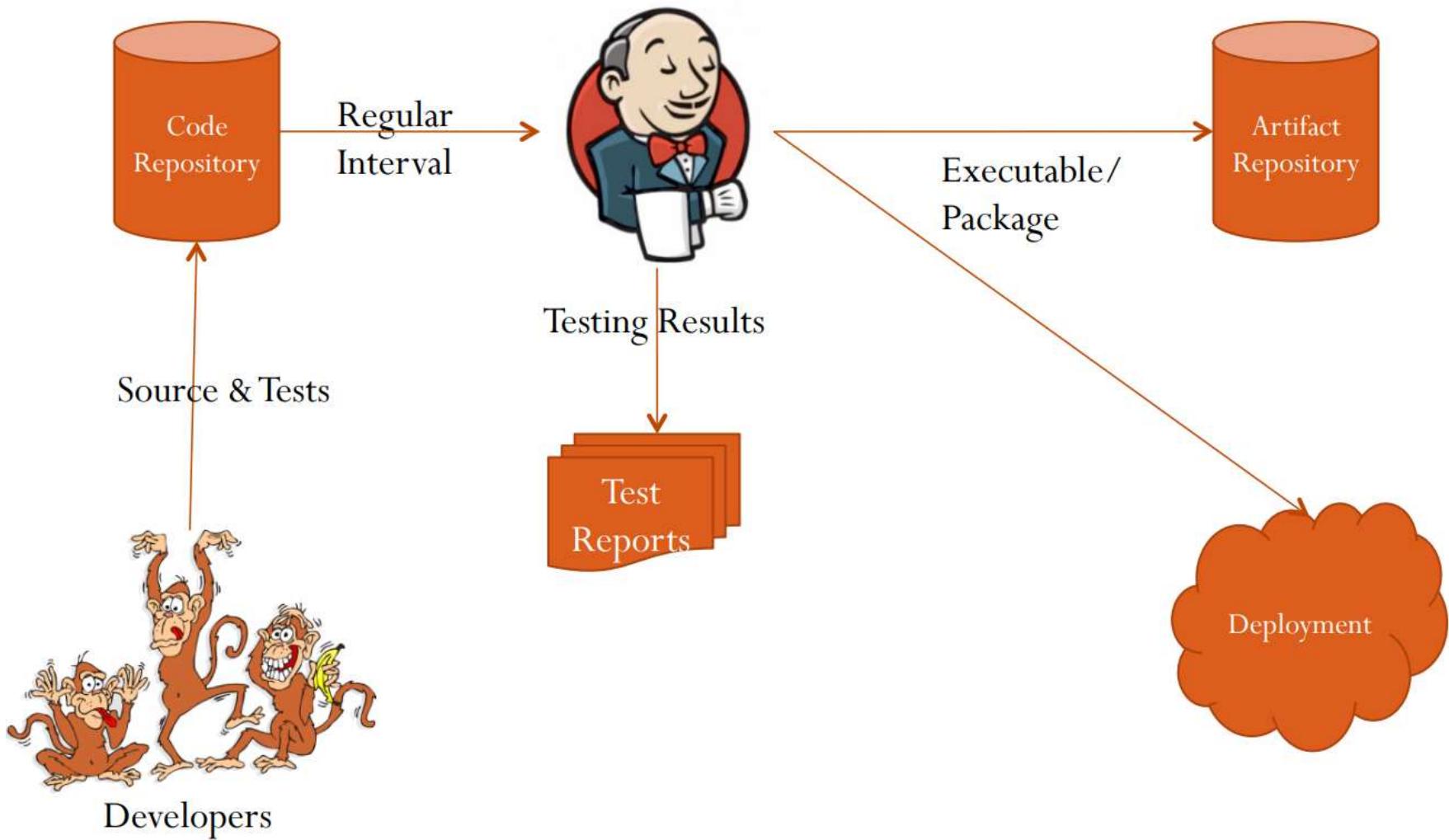
Jenkins

- Branched from Hudson
- Java based Continuous Build System
- Runs in servlet container
 - Glassfish, Tomcat
- Supported by over 400 plugins
 - SCM, Testing, Notifications, Reporting, Artifact Saving, Triggers, External Integration
- Under development since 2005
- <http://jenkins-ci.org/>

History

- 2005 - Hudson was first release by Kohsuke Kawaguchi of Sun Microsystems
- 2010 – Oracle bought Sun Microsystems Due to a naming dispute, Hudson was renamed to Jenkins Oracle continued development of Hudson (as a branch of the original)

Jenkins – Fitting in



Why Jenkins?

- Jenkins is a highly configurable system by itself
- The additional community developed plugins provide even more flexibility
- By combining Jenkins with Ant, Gradle, or other Build Automation tools, the possibilities are limitless



Why Jenkins? Free/OSS

- Jenkins is released under the MIT License
- There is a large support community and thorough documentation
- It's easy to write plugins
- Think something is wrong with it? You can fix it!

What can Jenkins Do?

- Generate test reports
- Integrate with many different Version Control Systems
- Push to various artifact repositories
- Deploys directly to production or test environments
- Notify stakeholders of build status
- ...and much more



How Jenkins works?

- When setting up a project in Jenkins, out of the box you have the following general options:
 - Associating with a version control server
 - Triggering builds
 - Polling, Periodic, Building based on other projects
 - Execution of shell scripts, bash scripts, Ant targets, and Maven targets
 - Artifact archival
 - Publish JUnit test results and Javadocs
 - Email notifications
- Plugins expand the functionality even further



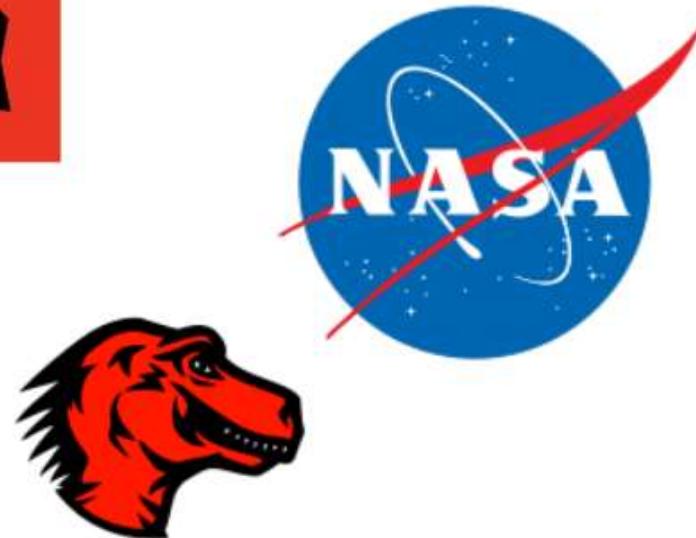
How Jenkins works?

- Jenkins comes with basic reporting features
 - Keeping track of build status
 - Last success and failure
 - “Weather” – Build trend
- These can be greatly enhanced with the use of pre-build plugins
 - Unit test coverage
 - Test result trending
 - Findbugs, Checkstyle, PMD

Enhancing Jenkins

- Jenkins plugin system can enable a wide range of features including (but certainly not limited to)
 - SCM
 - Mercurial, Git, Subversion
 - Testing
 - Selenium, Windmill, TestLink
 - Notifications
 - IRC, Twitter, Jabber
 - Reporting
 - Doxygen, PMD, Findbugs
 - Artifact Saving
 - Artifactory, Amazon S3, SCP
 - Triggers Jabber, Directory Watchers
 - External Integration GitHub, Bugzilla, JIRA
 - And most importantly – The CI Game A points based game where developers compete against each other to develop the most stable, welltested code

Who uses Jenkins



Running Jenkins yourself

- Jenkins is packaged as a WAR, so you can drop it into whichever servlet container you prefer to use
- Jenkins comes pre-packaged with a servlet if you just want a lightweight implementation
- Native/Supported packages exist for
 - Windows
 - Ubuntu/Debian
 - Redhat/Fedora/CentOS
 - Mac OSX
 - openSUSE
 - FreeBSD
 - OpenBSD
 - Solaris/OpenIndiana
 - Gentoo

Running Jenkins updates

Jenkins has two release lines

- **Standard releases**
 - Weekly bug fixes and features
- **Long-Term Support releases**
 - Updates about every 3 months
 - Uses a “Stable but older” version from the standard release line
 - Changes are limited to backported, well-tested modifications



Tying it into Agile

- For an Agile team, Jenkins provides everything needed for a robust continuous build system
- Jenkins supports Agile principles by constantly providing access to working copies of software
- Jenkins' extensibility allows the system to adapt to many different pre-existing environments

Putting together

- While an integral part of a CI system, Jenkins is by no means the only component
- In order for a CI system to function, a common repository for the codebase needs to exist
- A database of artifacts needs to exist, so deliveries can be made at past iterations
- The last step in a CI process is the deployment of the components built
- ...and none of this matters if the developers don't use the system; procedures need to ensure the system is used as intended



Self Management

Self
configuring

Self
Optimizing

Self
Healing

Self
Adaptive

Self
Protective

Self
Organizing

Self * (Selfware)

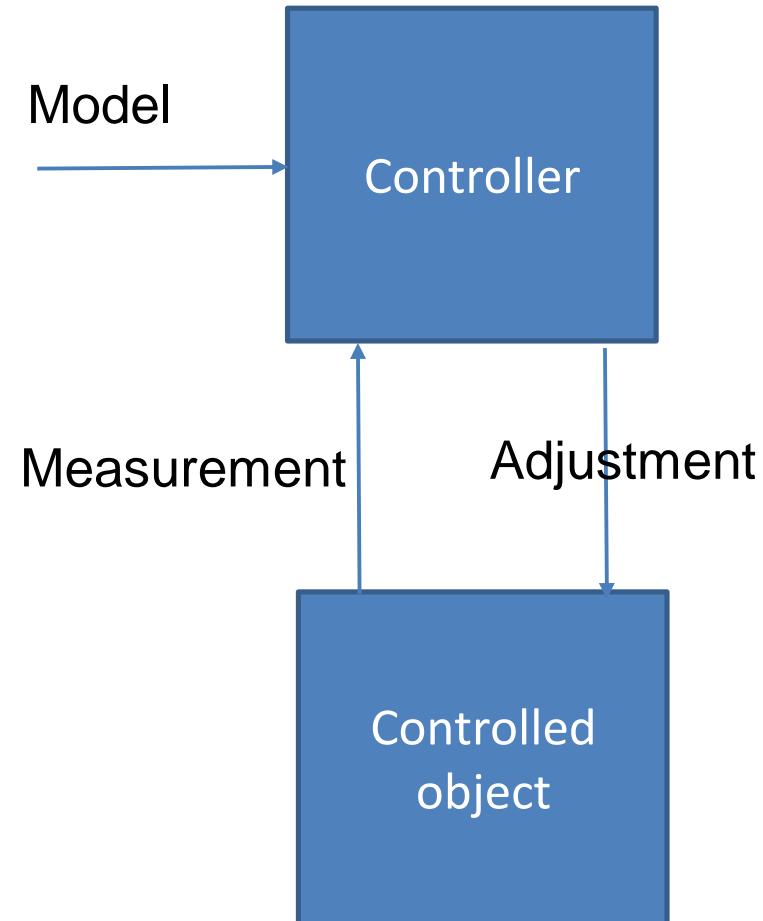
- Self-configuring
- Self-healing
- Self-optimising
- Self-protecting
- Self-aware
- Self-monitor
- Self-adjust
- Self-adaptive
- Self-governing
- Self-managed
- Self-controlling
- Self-repairing
- Self-organizing
- Self-evolving
- Self-reconfiguration
- Self-maintenance

Eight goals of a system

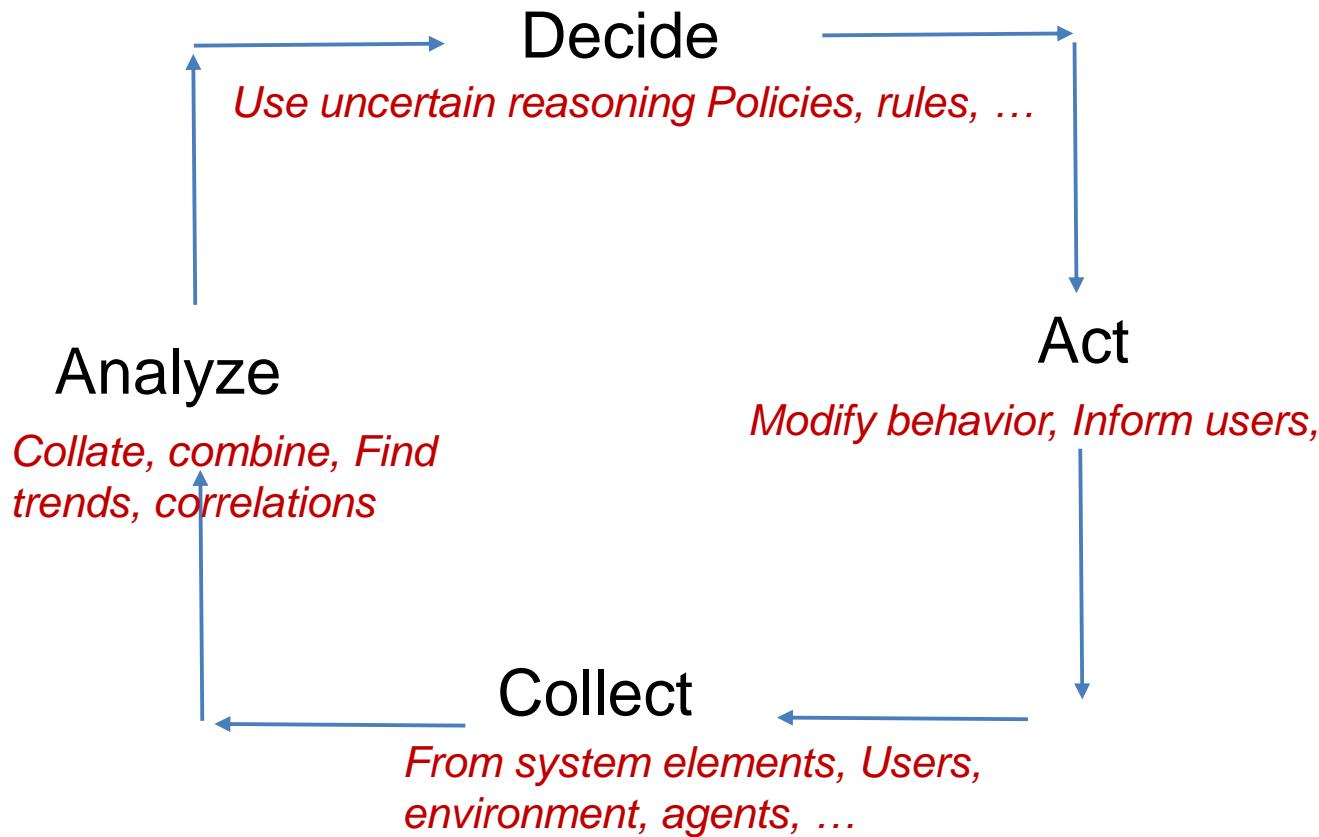
1. System must know itself
2. System must be able to reconfigure itself within its operational environment
3. System must pre-emptively optimize itself
4. System must detect and respond to its own faults as they develop
5. System must detect and respond to intrusions and attacks
6. System must know its context of use
7. System must live in an open world
8. System must actively shrink the gap between user/business goals and IT solutions

Autonomic Computing

- Basic model: closed control loops
 - Based on Process Control Theory
- Controller continuously compares the actual and expected behavior and makes needed adjustments



Autonomic Control Loop



Elements of Self healing

Fault Model	System Response	System Completeness	Design Context
<ul style="list-style-type: none">• Fault duration• Fault manifestation• Fault Source• Granularity• Fault profile expectations	<ul style="list-style-type: none">• Fault Detection• Degradation• Fault Response• Fault recovery• Time constants Assurance	<ul style="list-style-type: none">• Architectural completeness• Designer Knowledge• System Self-knowledge• System evolution	<ul style="list-style-type: none">• Abstraction level• Component homogeneity• Behavioral pre determination• User involvement in healing• System linearity• System Scope

Size of the self healing unit

Component

- Focus on connectors and component discovery
 - Service
 - Service interfaces, Service discovery, restart
 - Node
 - Network and interface failures, change to new connection

Architectural approach

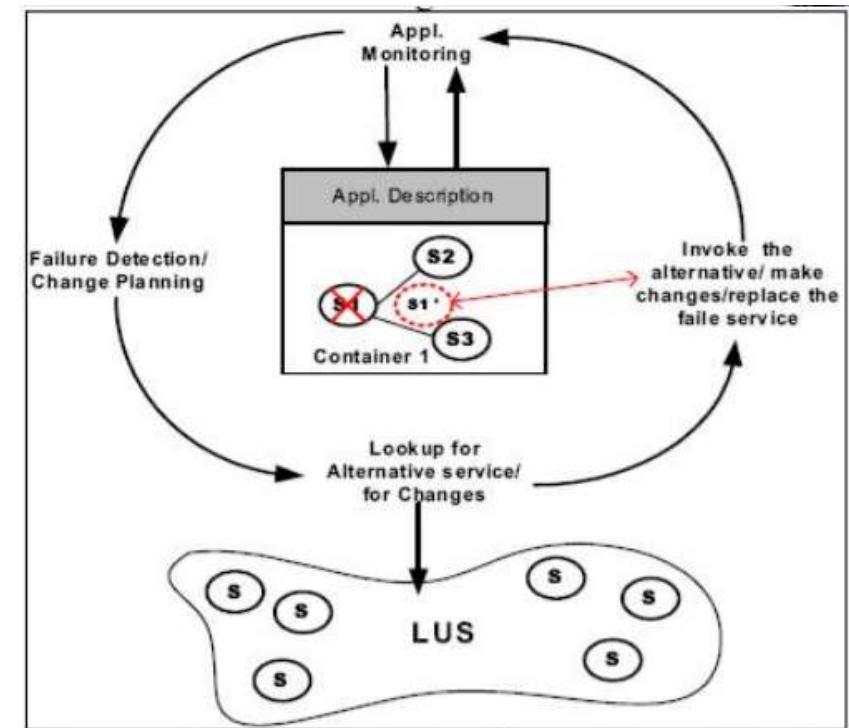
- The healing or recovery part often requires reconfiguration and adaptation
- They change the architecture
 - Locate and use alternative component
 - Restart (or rejuvenation or resurrection) the failed component
- Self-healing can be build on reflective middleware

Experiments

- OSAD – model (On-demand Service Assembly and Delivery)
- MARKS – Middleware Adaptability for Resource discovery, Knowledge usability and Self-healing
- PAC – Autonomic Computing in Personal Computing Environment
- Using self-healing components and connectors

Life-cycle of Self-Healing

- OSAD – On-demand Service Assembly and Delivery
- Prototype in JINI environment
- Looking for alternatives only by name





Thank You!

In our next session:



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps



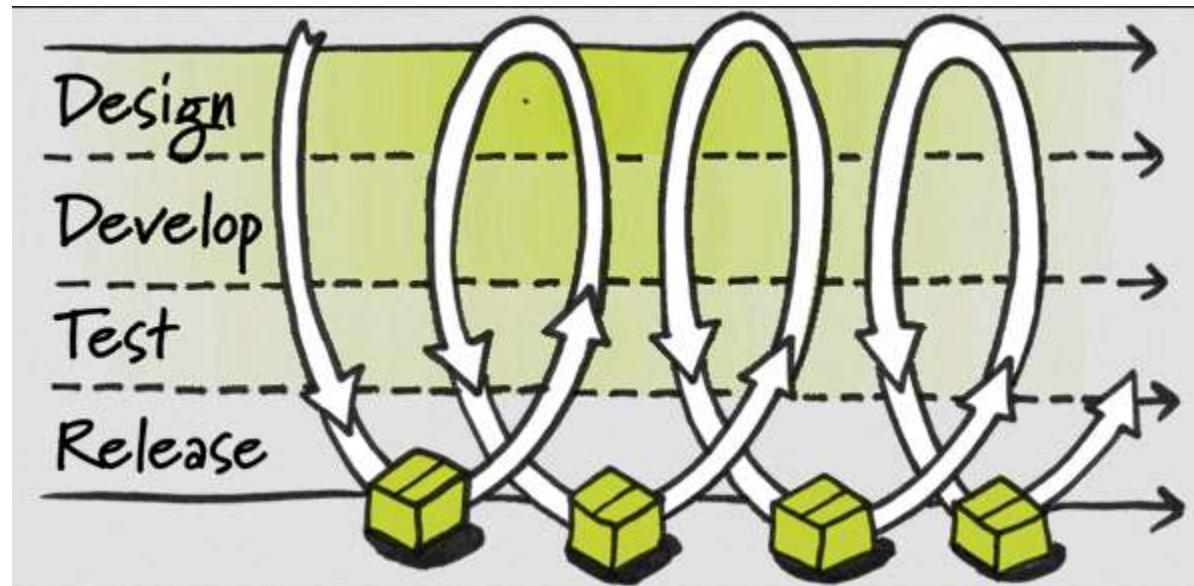


CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 9 and 10

Continuous Delivery

- Recall: Build software that is always ready to be deployed into production



Continuous Delivery cont.

- Is this a new idea?
- 1st principle of the Agile Manifesto:
 - “Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.”
- Goal: Release to production more often
 - Monthly
 - Weekly
 - Daily

Continuous Delivery cont.

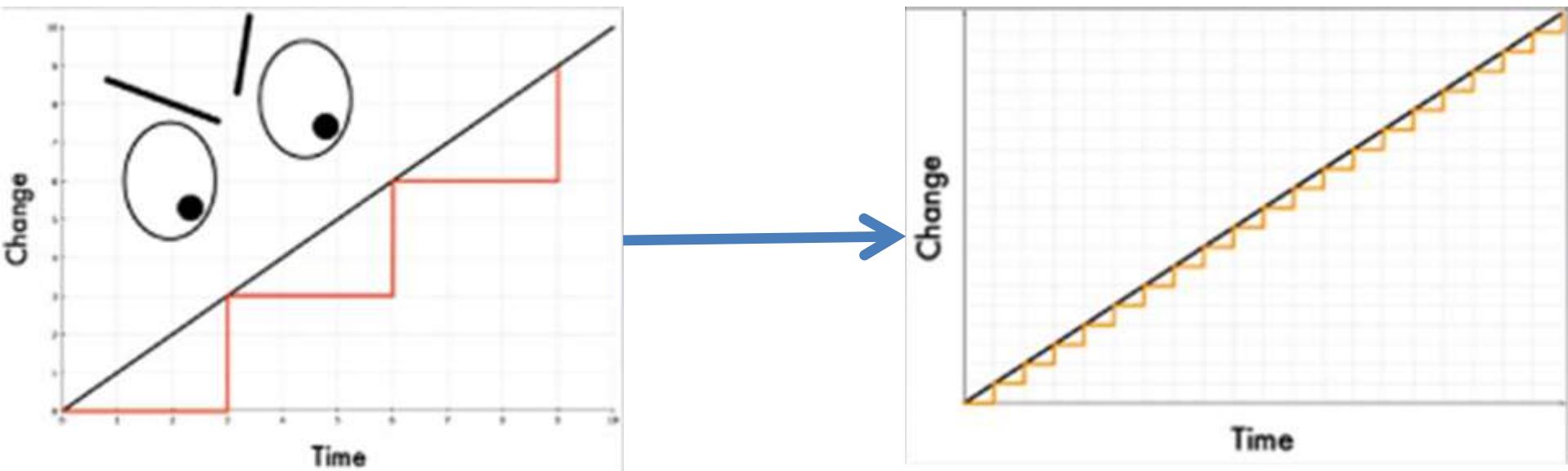
- Pros: Receive many benefits from the Agile process as well as others
 - Build the right product
 - Constant feedback from customers
 - Write thorough acceptance tests
 - Earlier benefits
 - Get product out before competitors
 - Find bugs earlier

Continuous Delivery cont.

- Ability to react quickly to change
 - Not a huge amount of time and money can be abandoned to adopt new requirements
- Save money and time
 - Save money and time if project goes bad
 - Save money and time with automation
 - As deploying to production becomes easier, more time can be spent developing valuable features

Continuous Delivery cont.

- Pros cont:
 - Reliability



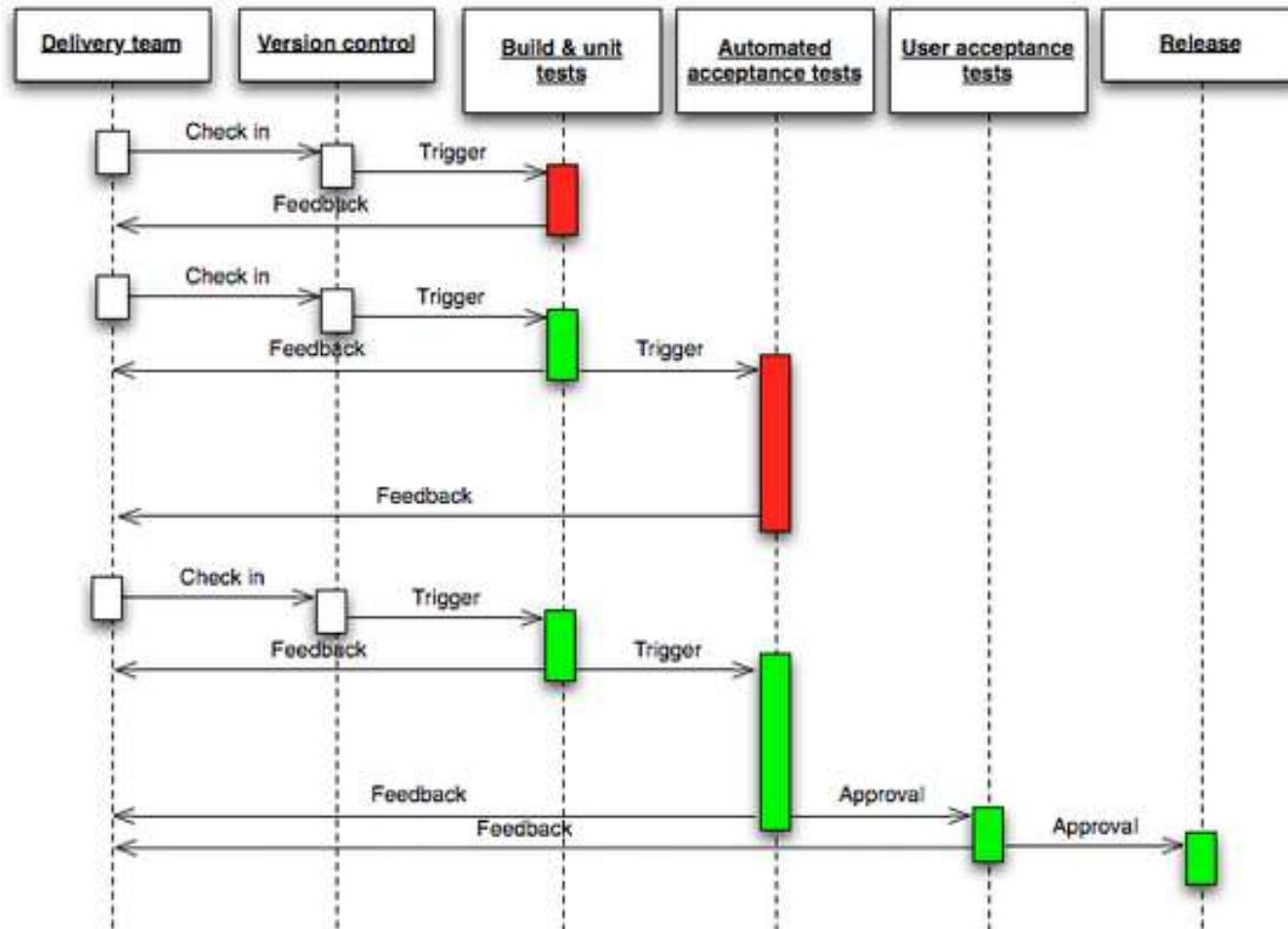
Practice Continuous Delivery

- Basics to start practicing Continuous Delivery
 - Configuration Management
 - What needs to be kept under CM?
 - Everything!
 - Code
 - Tests
 - Configuration Files
 - Build Scripts
 - Environments
 - Documentation
 - Etc.

Deployment Pipeline

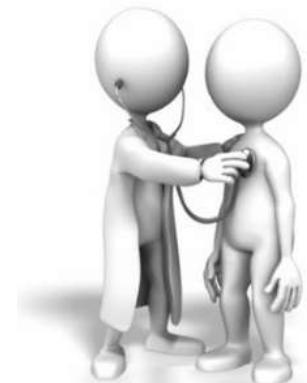
- Heart of Continuous Delivery
- Will give immediate feedback for how ready your software is for production
- Automate everything(unit tests, acceptance tests, performance tests,...) to the point of just being able to click a button to deploy to production

Deployment Pipeline cont.



Automation – Continuous Deployment

- Deployment Automation/Continuous Delivery
 - Distributing the components of the releasable package into target environment (Dev, Test, Prod)
- Provisioning
 - Create and readying the target Env. And middleware required by your application.
- Automated Testing
 - Verify:
 - Functional
 - Non functional
- Feedback



TESTS = DEPLOYMENT HEALTH



Automation – IAC

Infrastructure As A Code

- Put Infra under **version control**
- History , Melodize your infra
- Deploy monitoring , backups and apps
- **IT IS NOT SCRIPTING !**
- Think Puppet, Chef, Ansible, Udeploy....

Eight goals of a system

1. System must know itself
2. System must be able to reconfigure itself within its operational environment
3. System must pre-emptively optimize itself
4. System must detect and respond to its own faults as they develop
5. System must detect and respond to intrusions and attacks
6. System must know its context of use
7. System must live in an open world
8. System must actively shrink the gap between user/business goals and IT solutions

Elements of Self healing

Fault Model	System Response	System Completeness	Design Context
<ul style="list-style-type: none">• Fault duration• Fault manifestation• Fault Source• Granularity• Fault profile expectations	<ul style="list-style-type: none">• Fault Detection• Degradation• Fault Response• Fault recovery• Time constants Assurance	<ul style="list-style-type: none">• Architectural completeness• Designer Knowledge• System Self-knowledge• System evolution	<ul style="list-style-type: none">• Abstraction level• Component homogeneity• Behavioral pre determination• User involvement in healing• System linearity• System Scope

Size of the self healing unit

Component

- Focus on connectors and component discovery
 - Service
 - Service interfaces, Service discovery, restart
 - Node
 - Network and interface failures, change to new connection

Architectural approach

- The healing or recovery part often requires reconfiguration and adaptation
- They change the architecture
 - Locate and use alternative component
 - Restart (or rejuvenation or resurrection) the failed component
- Self-healing can be build on reflective middleware

Course Objectives



9

Continuous Deployment

- Deployment pipeline
- Human-free deployments
 - Implementing and Automating the deployment process
 - Deploying it to testing environments

T1- Chapter 6, 12

T2- Chapter 10

R2- Chapter 3, 4

Error Code 725: It works on my machine.

—RFC for HTTP Status Code 7XX: Developer Errors

Problems

Time killers

Manual testing

Manual deployment

Integrations done at the end

Code horror

Small change requires full regression testing

Release “jams”

Release packages various from team to team

On-site participation of 1-2 team members during release

Release “jams”

Team releases Release manager



PROD

Awaiting teams





Eliminating time killers

Integrations done at the end

- Automatic builds on commit

- Integration tests

Manual deployment

- Automatic deployment to DEV (x2) and INT environments from Tools

Manual testing

- Unit and integration tests

- Functional and UI tests (SoapUI, Selenium)



Eliminating code horror

Small change requires full regression testing

Unit tests

Best to have experienced person in place

Will require code refactoring

Separate data storage from the business logic

Are part of the task DONE criteria

Integration tests

Working with DB, Web services, file system, etc.



Eliminating release “jams”

Unified release packages

MSBuild scripts with custom tasks

One-click principle

Single package for all environments

Initially takes some time to create, but later is semi- or fully-
automatically created on Tools

Centralized release server

Single place to deploy packages to remote servers

Deployment

- Process of placing a version of a service into production
- **The initial deployment of a service can be viewed as going from no version of the service to the initial version of the service.**
- Because an initial deployment happens only once for most systems and new versions happen frequently

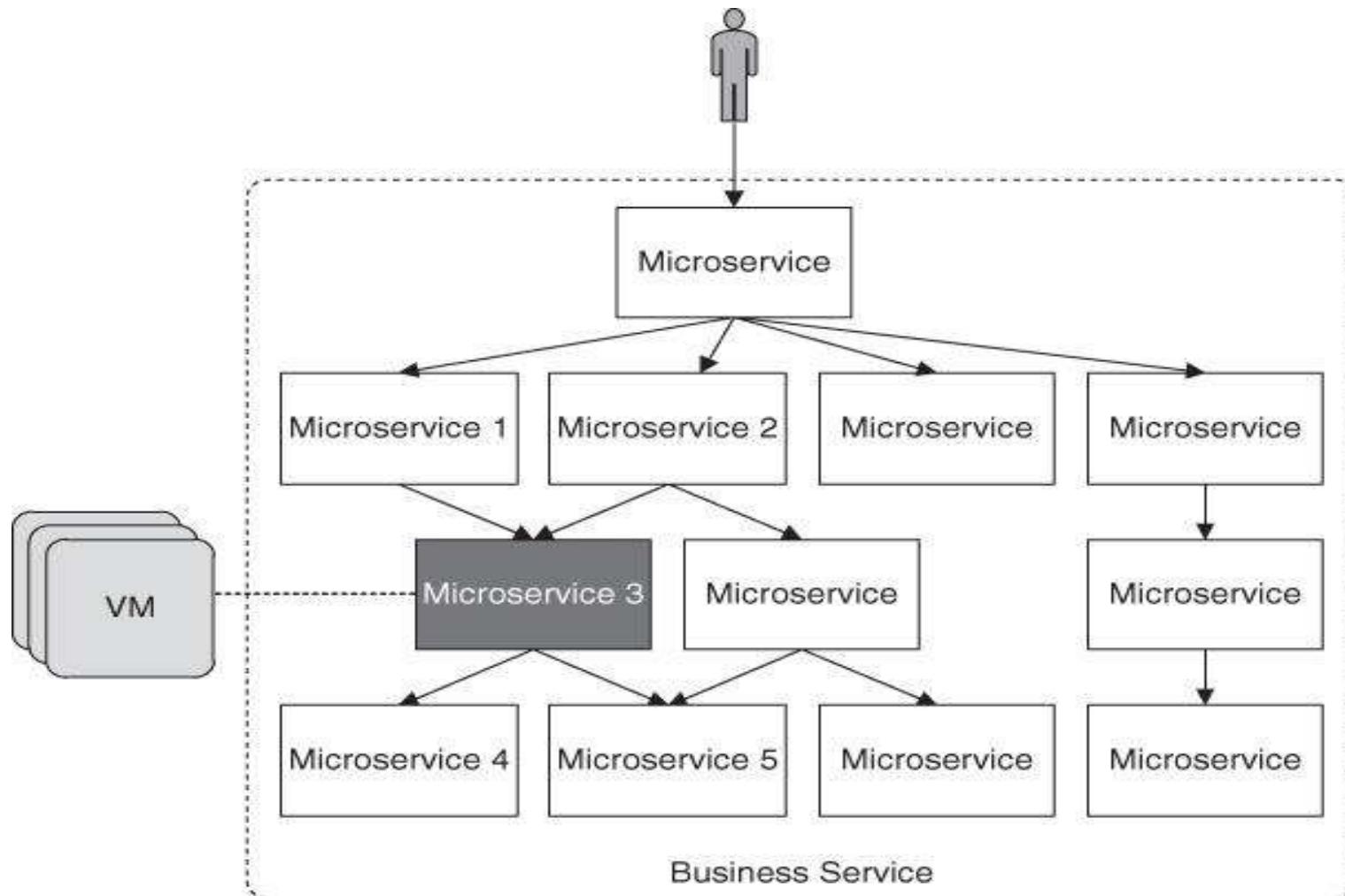
Goal of a deployment is to place an upgraded version of the service into production with minimal impact to the users of the system, be it through failures or downtime



Reasons for changing a service

- To fix an error
- To improve some quality of the service
- To add a new feature

Business Service

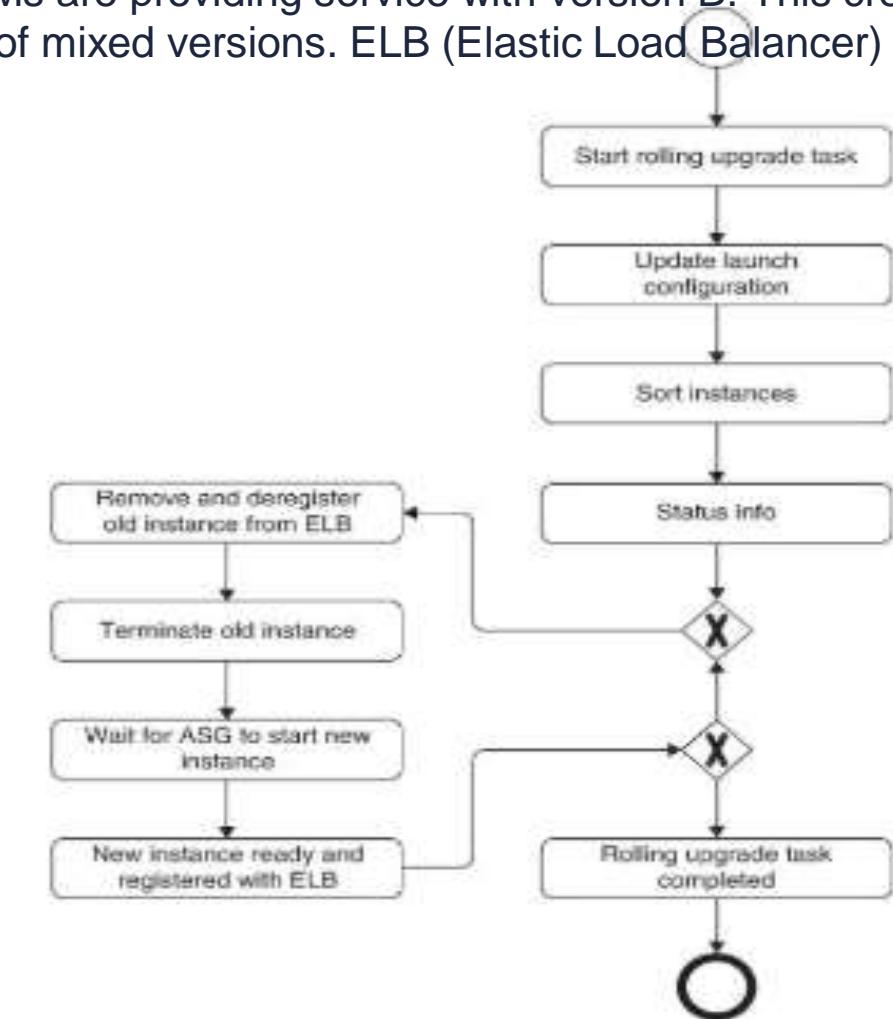


Strategies for managing Deployments

- Blue/green deployment (aka big flip or red black)
 - consists of maintaining the N VMs containing version A in service while provisioning N VMs of virtual machines containing version B. Once N VMs have been provisioned with version B and are ready to service requests, then client requests can be routed to version B
- Rolling upgrade
 - A rolling upgrade consists of deploying a small number of version B VMs at a time directly to the current production environment, while switching off the same number of VMs running version A.

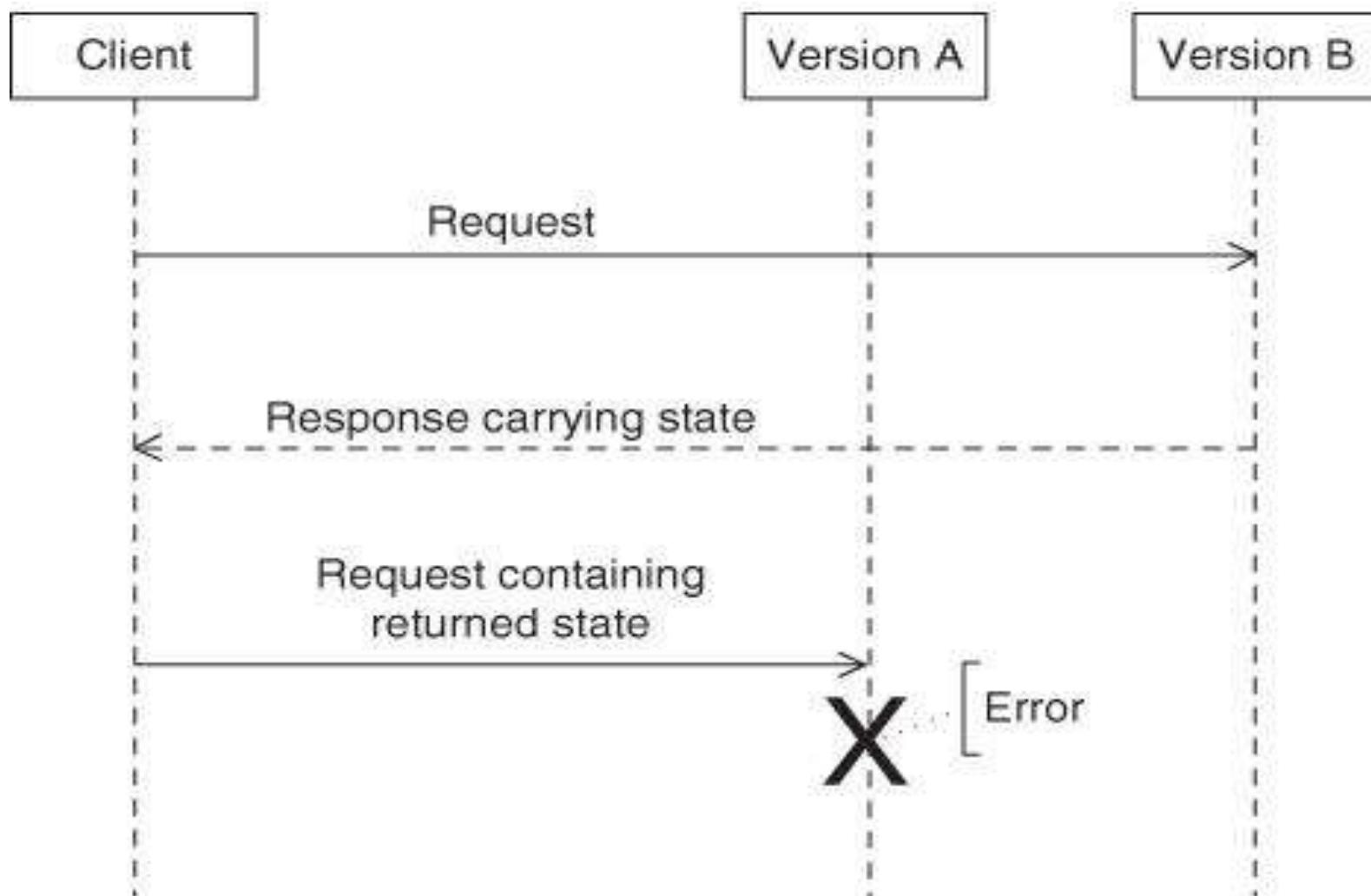
Rolling Upgrade in Amazon Cloud

During a rolling upgrade, one subset of the VMs is providing service with version A, and the remainder of the VMs are providing service with version B. This creates the possibility of failures as a result of mixed versions. ELB (Elastic Load Balancer)



Logical consistency

1. Assuming that the deployment is done using a rolling upgrade introduces one type of logical inconsistency—multiple versions of the same service will be simultaneously active
2. Assuming that a service is being deployed without synchronous coordination with its client or dependent services – a second possible source of logical inconsistency—inconsistency in functionality between a service and its clients
3. A third source of logical inconsistency is inconsistency between a service and data kept in a database.



Feature Toggling

- If services are developed by a single small team, the features of the services will be limited. This means that features are likely to span multiple services. In turn, this means that you must coordinate the activation of the feature in two directions
- First, all of the VMs for the service you just deployed must have the service's portion of the feature activated.
- Second, all of the services involved in implementing the feature must have their portion of the feature activated
- Feature toggles, can be used to control whether a feature is activated.

Feature Toggling Contd

A feature toggle, to repeat, is a piece of code within an *if* statement where the *if* condition is based on an externally settable feature variable. Using this technique means that the problems associated with activating a feature are

- (a) Determining that all services involved in implementing a feature have been sufficiently upgraded
- (b) Activating the feature in all of the VMs of these services at the same time.

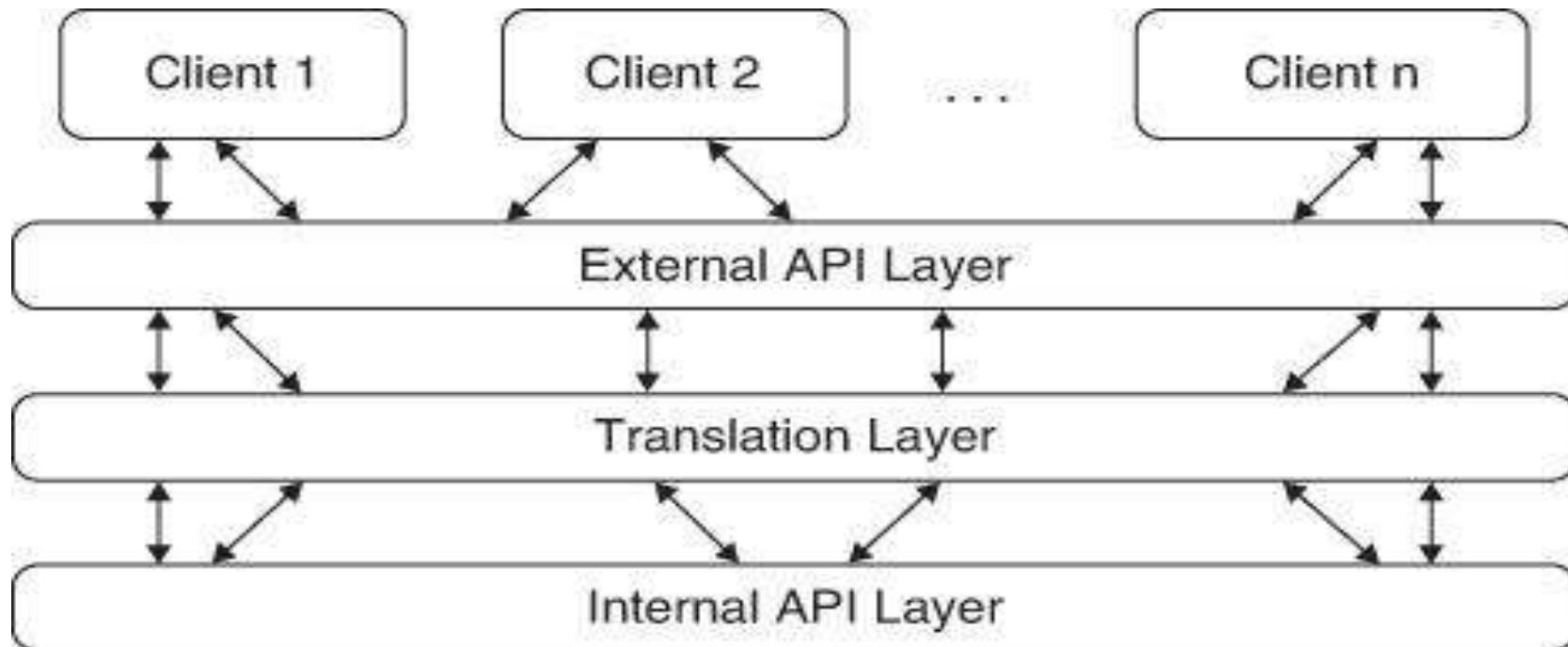
Backward and Forward Compatibility

Using feature toggles to coordinate the various services involved in a new feature is one option for preventing failures as a result of multiple versions

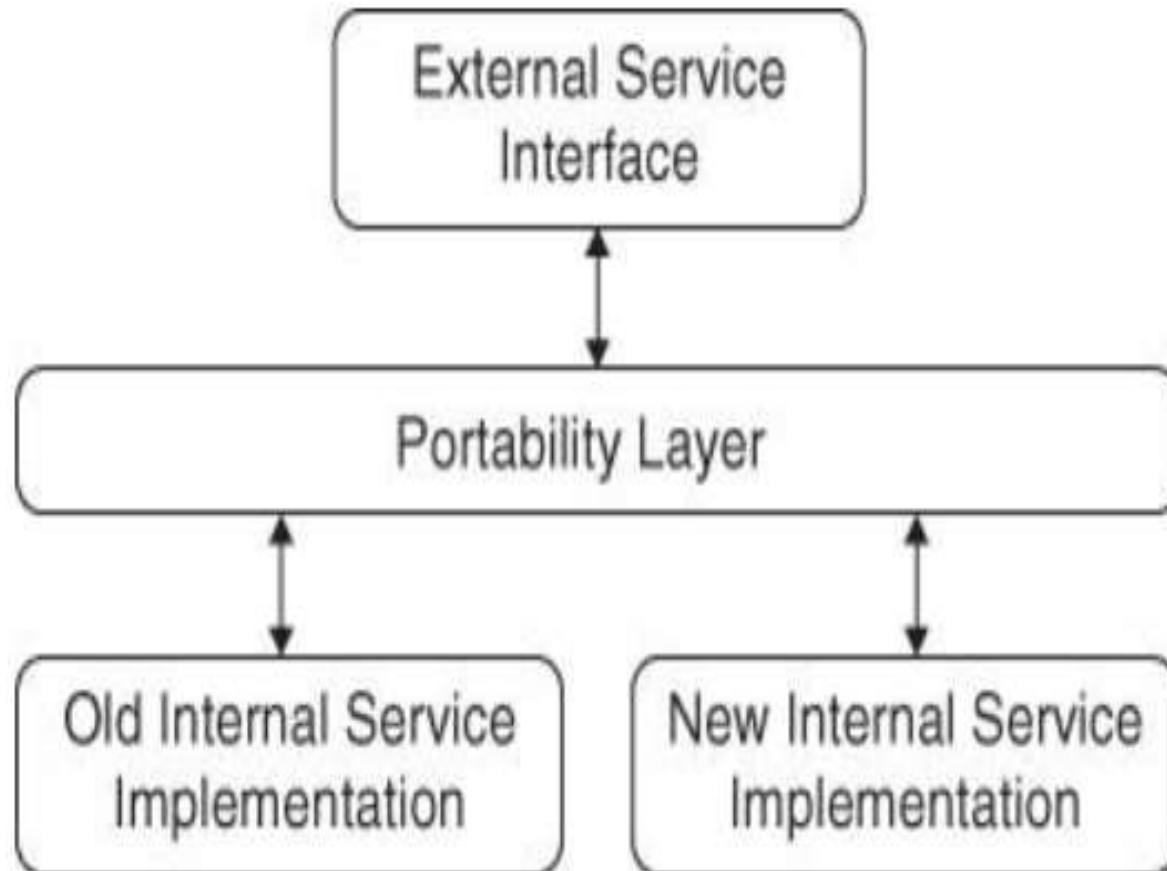
Forward compatibility means that a client deals gracefully with error responses indicating an incorrect method call. Suppose a client wishes to utilize a method that will be available in version B of a service but the method is not present in version A. Then if the service returns an error code indicating it does not recognize the method call, the client can infer that it has reached version A of the service. Requiring backward compatibility might seem at first to preclude many changes to a service. If you cannot change an interface, how can you add new features or, for example, refactor your service?

A service is backward compatible if the new version of the service behaves as the old version. For requests that are known to the old version of a service, the new version provides the same behavior. In other words, the external interfaces provided by version B of a service are a superset of the external interfaces provided by version A of that service.

Maintaining backward compatibility



Portability layer with 2 versions of the external system co existing

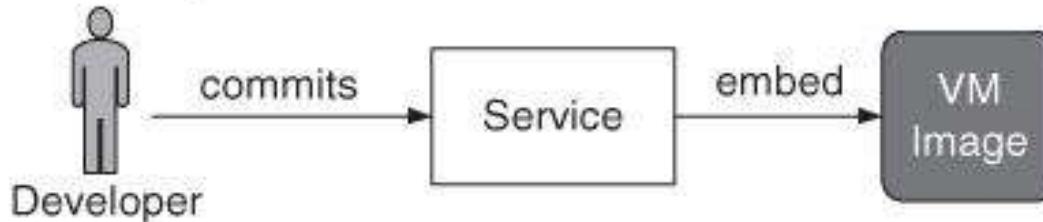


Packaging

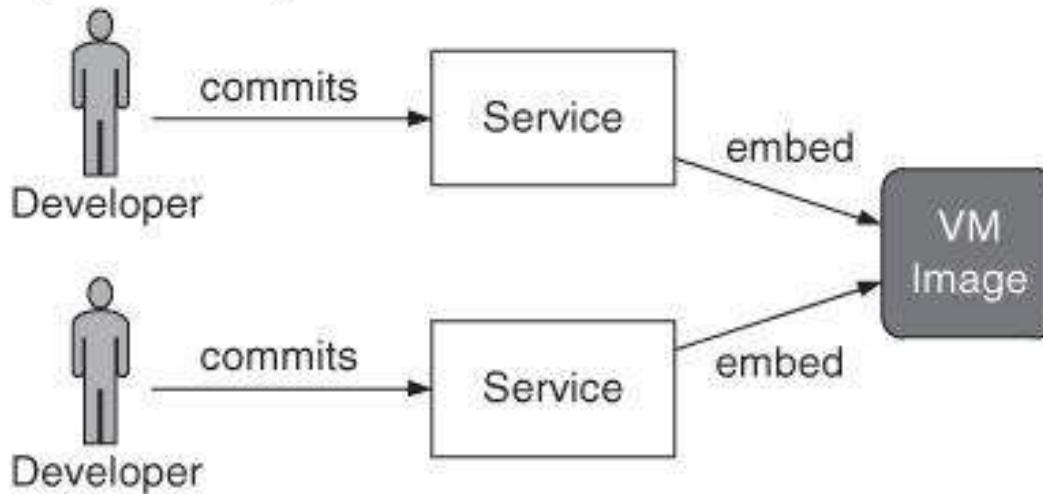
- Packaging components onto a VM image is called *baking* and the options range from *lightly baked* to *heavily baked*
- A VM is an image that is running on top of a hypervisor that enables sharing a single bare metal processor, memory, and network among multiple tenants or VMs
- The image of the VM is loaded onto the hypervisor from which it is scheduled. A VM image could include multiple independent processes—each a service

Options for packaging service

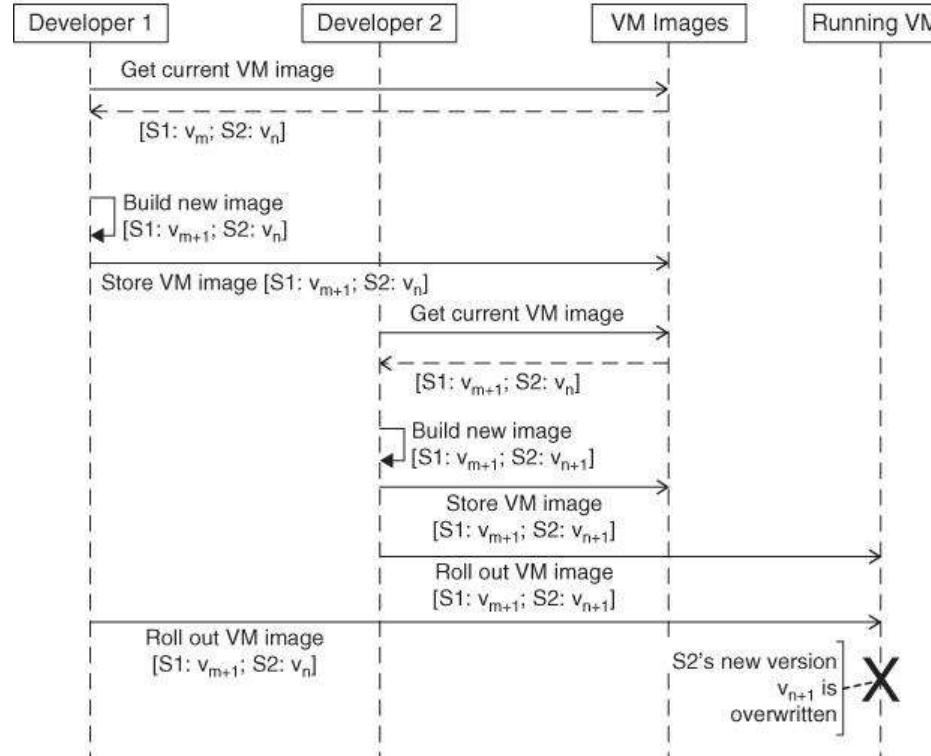
One service per VM



Multiple services per VM

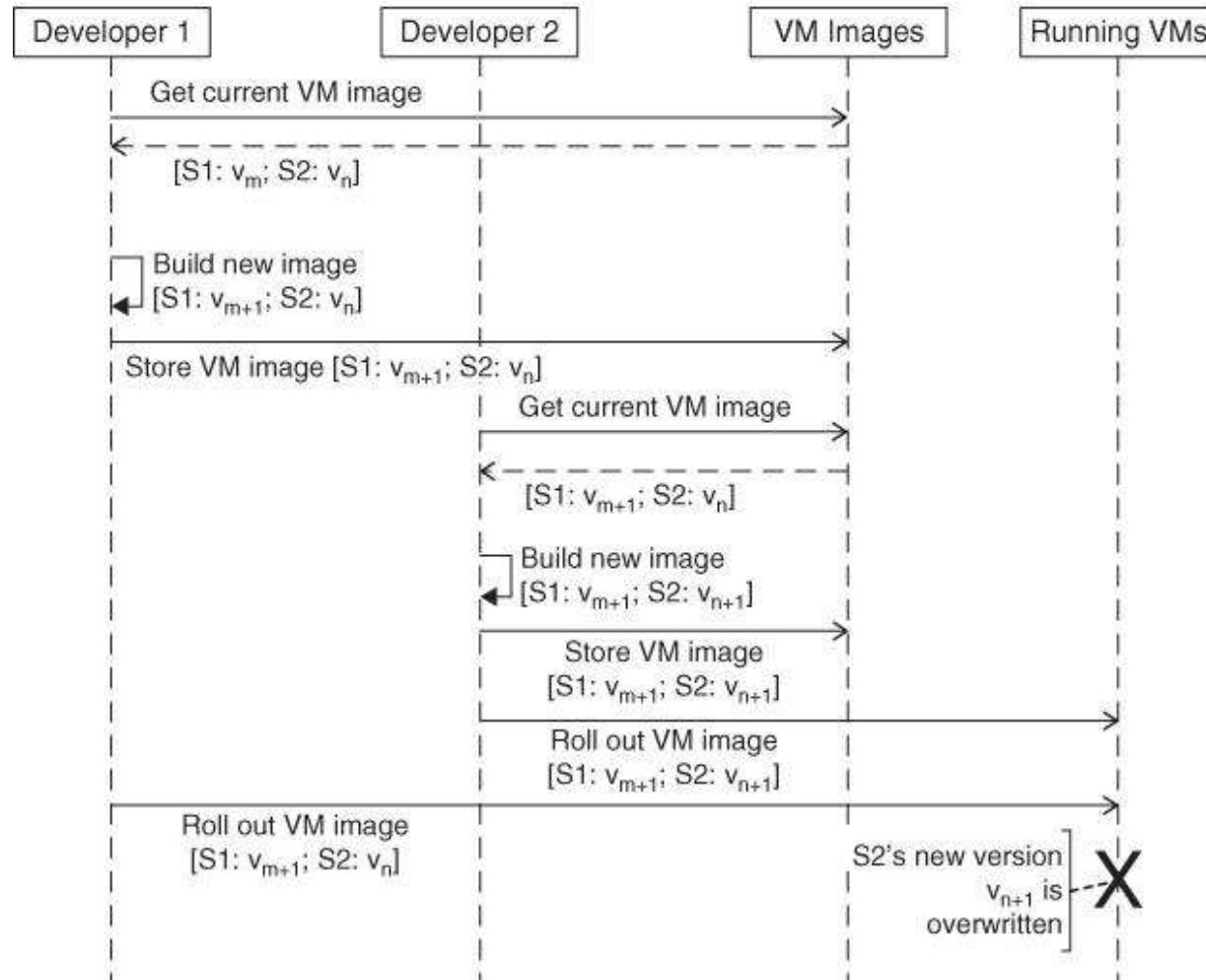


Race condition



One type of race condition when two development teams deploy independently

Race condition 2



- Deploying to testing and production environments and rolling back—need to form part of your deployment pipeline implementation
- It should be possible to see a list of builds available for deployment into each of these environments and run the automated deployment process by pressing a button or clicking a mouse to select the version to deploy and the environment to deploy it to

Creating a release strategy

- Parties in charge of deployments to each environment, as well as in charge of the release
- An asset and configuration management strategy
- A description of the technology used for deployment. This should be agreed upon by both the operations and development teams
- A plan for implementing the deployment pipeline
- An enumeration of the environments available for acceptance, capacity, integration, and user acceptance testing, and the process by which builds will be moved through these environments

Creating a release strategy

- A description of the processes to be followed for deployment into testing and production environments, such as change requests to be opened and approvals that need to be granted.
- Requirements for monitoring the application, including any APIs or services
the application should use to notify the operations team of its state.
- A discussion of the method by which the application's deploy-time and runtime configuration will be managed, and how this relates to the automated deployment process.
- Description of the integration with any external systems. At what stage and
how are they tested as part of a release? How do the operations

Creating a release strategy

- Details of logging so that operations personnel can determine the application's state and identify any error conditions.
- A disaster recovery plan so that the application's state can be recovered following a disaster.

The service-level agreements for the software, which will determine whether the application will require techniques like failover and other high-availability strategies.

- Production sizing and capacity planning: How much data will your live application create? How many log files or databases will you need? How much bandwidth and disk space will you need? What latency are clients expecting?

Creating a release strategy

- An archiving strategy so that production data that is no longer needed can be kept for auditing or support purposes
- How the initial deployment to production works
- How fixing defects and applying patches to the production environment will be handled
- How upgrades to the production environment will be handled, including data migration
- How application support will be managed

Release Plan

- The steps required to deploy the application for the first time
 - How to smoke-test the application and any services it uses as part of the deployment process
 - The steps required to back out the deployment should it go wrong
 - The steps required to back up and restore the application's state
 - The steps required to upgrade the application without destroying the application's state
 - The steps to restart or redeploy the application should it fail
 - The location of the logs and a description of the information they contain
 - The methods of monitoring the application
 - The steps to perform any data migrations that are necessary as part of the release
 - An issue log of problems from previous deployments, and their solutions
-

Releasing Products

- Pricing model
- Licensing strategy
- Copyright issues around third-party technologies used
- Packaging
- Marketing materials—print, web-based, podcasts, blogs, press releases, conferences, etc.
- Product documentation
- Installers
- Preparing sales and support teams

The first deployment

- Your deployment pipeline's commit stage
- A production-like environment to deploy to
- An automated process that takes the binaries created by your commit stage

and deploys them into the environment

- A simple smoke test that verifies that the deployment worked and the application is running

Characteristics of a prod like system



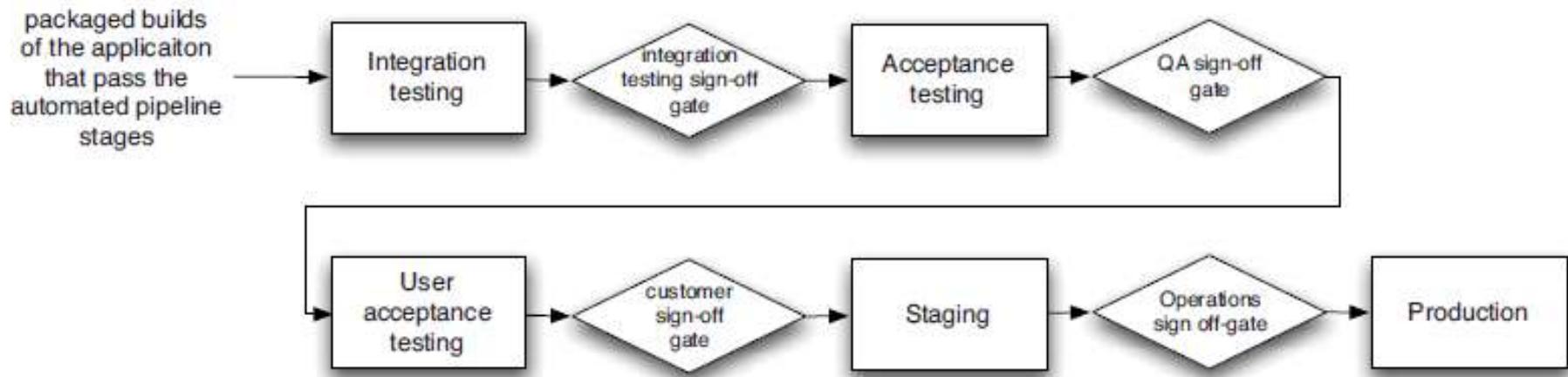
- It should run the same operating system as the production system will.
- It should have the same software installed as the production system will—and in particular, none of the development toolchain (such as compilers or IDEs) should be installed on it.
- This environment should, as far as is reasonable, be managed the same way as the production environment
- In the case of client-installed software, your UAT environment should be representative of your clients' hardware statistics, or at least someone else's
 - real-world statistics

Modeling release process and promoting builds



- What stages a build has to go through in order to be released (for example, integration testing, QA acceptance testing, user acceptance testing, staging, production)
- What the required gates or approval are
- For each gate, who has the authority to approve a build passing through that gate

Test and release process diagram



PULL system

- The other essential facility that must be provided by the tool you use to manage your deployment pipeline is the ability, for each stage, to see which builds have passed all the previous stages in the pipeline and are hence ready for the next stage.
- It should then be possible to choose one of these builds and press a button to have it deployed. This process is known as promotion.
- Promoting builds at the press of a button is what turns the deployment pipeline into a pull system, giving everybody involved in the delivery process the ability to manage their own work.
- Analysts and testers can self-service deployments for exploratory testing, showcasing, or usability testing.
- Operations personnel can deploy any version of their choice to staging or production at the press of a button

Workflow for testing stages

- The person or team doing the testing should have a way to select which version of the application they want to deploy into their testing environment. This list will include all versions of the application that have passed all the prior stages of the deployment pipeline. Selecting a particular build should cause the following steps, up to the actual testing, to be performed automatically.
- Prepare the environment and associated infrastructure (including middleware) so it is in a clean state, ready for deployment of the application. This should be done in a fully automated fashion
- Deploy the application's binaries. These binaries should always be fetched from the artifact repository, never built from scratch for each deployment.

Workflow for testing stages

- Configure the application. Configuration information should be managed in a consistent way across all applications and applied at deployment time or run time, using a tool like Escape
- Prepare or migrate any data managed by the application
- Smoke-test the deployment
- Perform the testing (this might be manual or automated)
- If this version of the application passes the tests, approve its promotion to the next environment.
- If this version of the application does not pass the tests, record why

Planning at beginning of project



- Ensure your production, capacity testing, and staging environments are commissioned. In particular, on a green field project, have your production environment ready some time before the release, and deploy to it as part of your pipeline.
- Have an automated process for configuring your environment, including networks, external services, and infrastructure.
- Ensure the deployment process is adequately smoke-tested.
- Measure the warm-up period for your application. This is especially applicable if your application uses caching. Incorporate this into your deployment plan.
- Test integration with external systems. You don't want your application's release to be the first time you run against the real external systems.

Planning at beginning of project



- If possible, get your application into its production environment well before release. If “release” can be as simple as reconfiguring some router to direct traffic from a holding page to your production environment, so much the better. This technique, known as blue-green deployment, is described a little later in this chapter.
- If possible, try rolling your system out to a small group of users before you roll it out to everybody. This technique is known as canary releasing, and is also described later in this chapter.
- Deploy every change that passes acceptance tests to your staging environment (although not necessarily to production).

Rolling back deployments & Zero downtime releases

- It is essential to be able to roll back a deployment in case it goes wrong
- Debugging problems in a running production environment is almost certain to result in late nights, mistakes with unfortunate consequences, and angry users
- You need to have a way to restore service to your users when things go wrong, so you can debug the failure in the comfort of normal working hours

Zero downtime releases AND Rollbacks

- Blue-green deployments
- Canary releasing

Constraints

- The first is your data. If your release process makes changes to your data, it can be hard to roll back.
- Another constraint is the other systems you integrate with. With releases involving more than one system (known as orchestrated releases), the rollback process becomes more complex too

Principles for rolling back

- The first is to ensure that the state of your production system, including databases and state held on the filesystem, is backed up before doing a release
- The second is to practice your rollback plan, including restoring from the backup or migrating the database back before every release to make sure it works

Create the env & deploy from scratch

- If you do not have an automated rollback process but you do have an automated deployment process, then redeploying the last version is a fixed time operation that poses a lower risk (because there is less to go wrong).
- It is the same process you have tested (hopefully) hundreds of times before. Rollbacks are performed much less frequently, and therefore are more likely to contain bugs

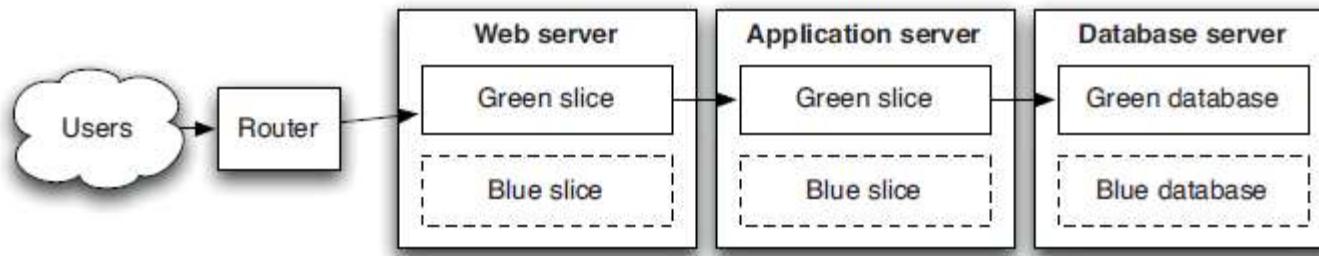
Disadvantages

- Even though the time it takes to redeploy the old version is fixed, it is nonzero. It will thus lead to a downtime.
- It makes it harder to debug what went wrong. Redeploying the old version often overwrites the new version, thereby removing the opportunity to work out what happened. This can be mitigated if your production environment is virtual, which we describe later on. With relatively simple applications, it's often easy to keep the old version around by deploying each version to a new directory and using symbolic links to point to the current version.
- If you restore from the database backup you took before deploying the latest version, you will lose any data created following the deployment. This may not be a big deal if you roll back reasonably quickly, but in some situations this is not acceptable

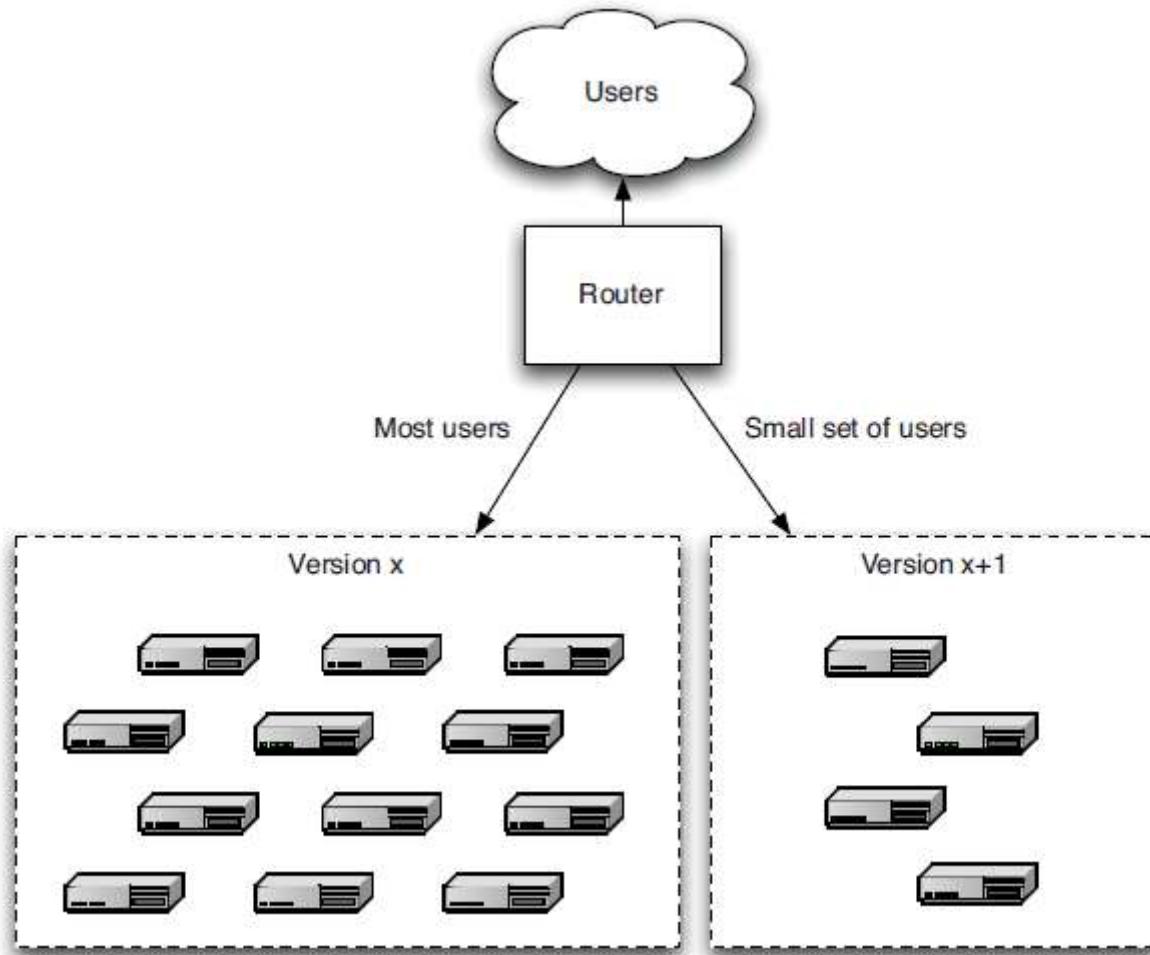
Zero downtime releases

- A zero-downtime release, also known as hot deployment, is one in which the actual process of switching users from one release to another happens nearly instantaneously
- It must also be possible to back users out to the previous version nearly instantaneously too, if something goes wrong
- The key to zero-downtime releases is decoupling the various parts of the release process so they can happen independently as far as possible
- It should be possible to put in place new versions of shared resources your applications depend on, such as databases, services, and static resources, before you upgrade your applications.

Blue Green Deployments



Canary Releasing



Canary releasing, involves rolling out a new version of an application to a subset of the production servers to get fast feedback. Like a canary in a coal mine, this quickly uncovers any problems with the new version without impacting the majority of users. This is a great way to reduce the risk of releasing a

Benefits of Canary Releasing

1. It makes rolling back easy: Just stop routing users to the bad version, and you can investigate the logs at your leisure.
2. You can use it for A/B testing by routing some users to the new version and some to the old version. Some companies measure the usage of new features, and kill them if not enough people use them. Others measure the actual revenue generated by the new version, and roll back if the revenue for the new version is lower. If your software generates search results, you might compare the quality of the results obtained by real users in the new version versus the old version. You don't need to route a large number of users to the new version to do A/B testing; a representative sample is sufficient.

Benefits of Canary Releasing

3. You can check if the application meets capacity requirements by gradually ramping up the load, slowly routing more and more users to the application and measuring the application's response time and metrics like CPU usage, I/O, and memory usage, and watching for exceptions in logs. This is a relatively low-risk way to test capacity if your production environment is too large to create a realistic production-like capacity testing environment

Dealing with defects in PROD



- Never do them late at night, and always pair with somebody else.
- Make sure you have tested your emergency fix process.
- Only under extreme circumstances circumvent the usual process for making changes to your application.
- Make sure you have tested making an emergency fix using your staging environment.
- Sometimes it's better to roll back to the previous version than to deploy a fix. Do some analysis to work out what the best solution is. Consider what happens if you lose data or face integration or orchestration problems

Assumptions

1. Service to the clients should be maintained while the new version is being deployed.

Maintaining service to the clients with no downtime is essential for many Internet e-commerce businesses. Their customers span the globe and expect to be able to transact business around the clock. Certainly, some periods of a day are going to be busier than others, but service must be available at all times. Organizations that have customers primarily localized in one geographic area can afford scheduled downtime—but why have downtime if it is avoidable? Scheduled off-hours during downtime requires system administrators and operators to work in the off-hours. This is another reason to avoid downtime.

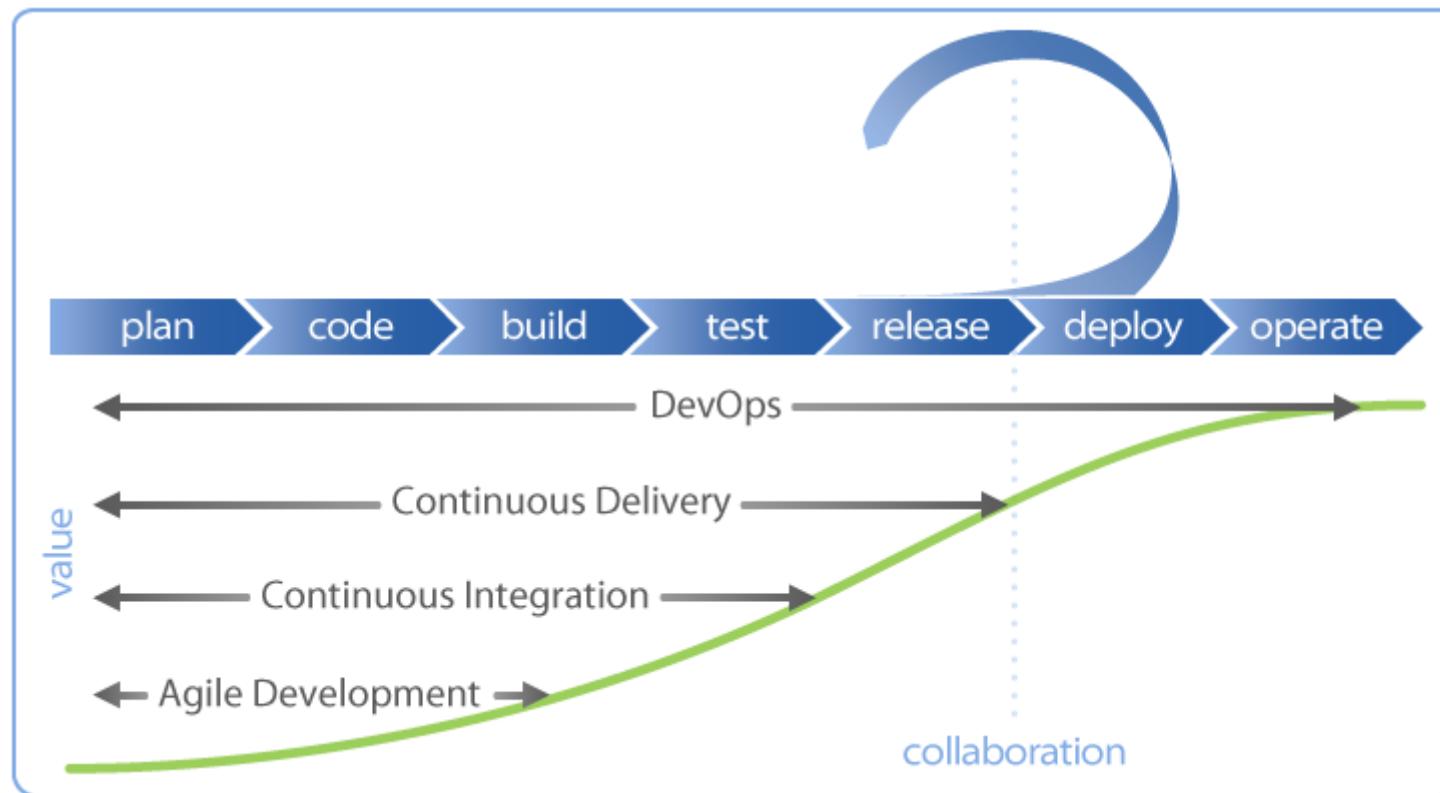
2. Any development team should be able to deploy a new version of their service at any time without coordinating with other teams. This may certainly have an impact on client services developed by other teams. We have previously discussed the relationship between synchronous co-ordination of development teams and the time to release new features. Allowing a development team or individual developer to release a new version of their service without coordinating with teams developing client services removes one cause for synchronous coordination.

CD vs CD?

"Continuous Delivery" means that you are ready and able to deploy any version to any supported platform at any time.

"Continuous Deployment" means that you are engaging in actual deployment.

Phases in Solution Delivery Pipeline



Continuous Delivery

Process improvement questions

- Did you finish what you thought you would finish?
- Did you estimate properly?
- Did it take the same amount of time you thought it would take?
- Were the tasks done the actual tasks the team had to complete during Sprint execution?
- Were there a lot of tasks that had to be done that were not planned?

In looking at Continuous Delivery, the preferred goal is to be ready and able to deploy 10 or more times per day. If you hit that goal, then deployment can be done every 45 minutes

Continuous deployment

Continuous, successful and repeatable
methodology to deploying code



Who is doing?

WordPress.com

“Everyone has access to a deploy button”

“Deployment to 400 production servers in our web tier in less than 30 seconds”

Facebook

Flickr

Atlassian

others...

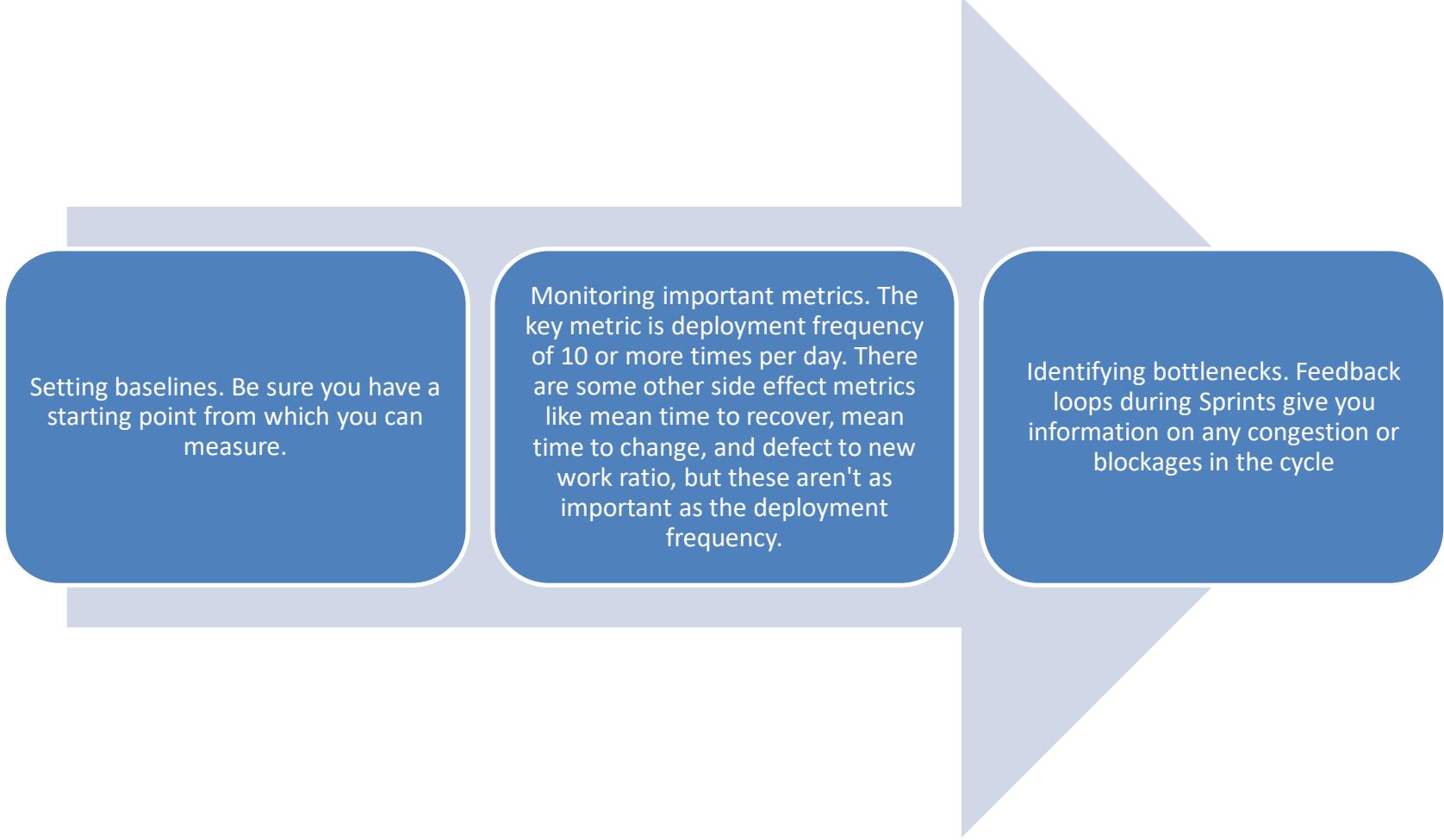


Continuous Deployment

- Continuous Deployment is actually deploying; however, it doesn't mean you must deploy to production or to the customer every time. For example, you can deploy to QA or to a deployment slot. A deployment slot is an A/B deployment method that allows you to get valuable information without inconveniencing users in the event that a defect leaks through a gap in the testing process. This takes some of the fear and anxiety out of the deployment.
- One way of deploying to a slot might be to offer an "opt-in" choice for the customer to get an updated version, providing you with a smaller group of users that can serve as a "beta." You always give these users the ability to go back to the former version if they want to do so.

Continuous Improvement

Continuous Improvement requires that you amplify feedback loops by:



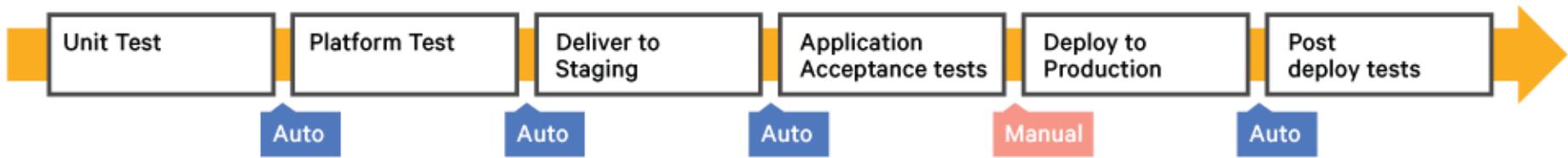
Setting baselines. Be sure you have a starting point from which you can measure.

Monitoring important metrics. The key metric is deployment frequency of 10 or more times per day. There are some other side effect metrics like mean time to recover, mean time to change, and defect to new work ratio, but these aren't as important as the deployment frequency.

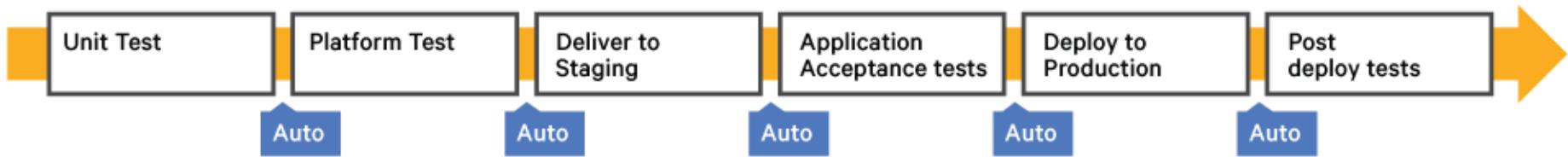
Identifying bottlenecks. Feedback loops during Sprints give you information on any congestion or blockages in the cycle

Delivery vs Deployment

Continuous Delivery



Continuous Deployment



Continuous Deployment

- if it hurts, do it more often—the logical extreme is to deploy every change that passes your automated tests to production
- I take my pipeline and make the final step—deployment to production—automatic. That way, if a check-in passes all the automated tests, it gets deployed directly to production
- Continuous deployment can be combined with canary releasing by using an automated process that rolls out a new version to a small group of users first, rolling it out to all users once it has been determined (probably as a manual step) that there are no problems with the new version. The added safeguards provided by a good canary releasing system make continuous deployment an even less risky proposition.

Continuously Releasing User-Installed Software: Issues

- Managing the upgrade experience
- Migrating binaries, data, and configuration
- Testing the upgrade process
- Getting crash reports from users

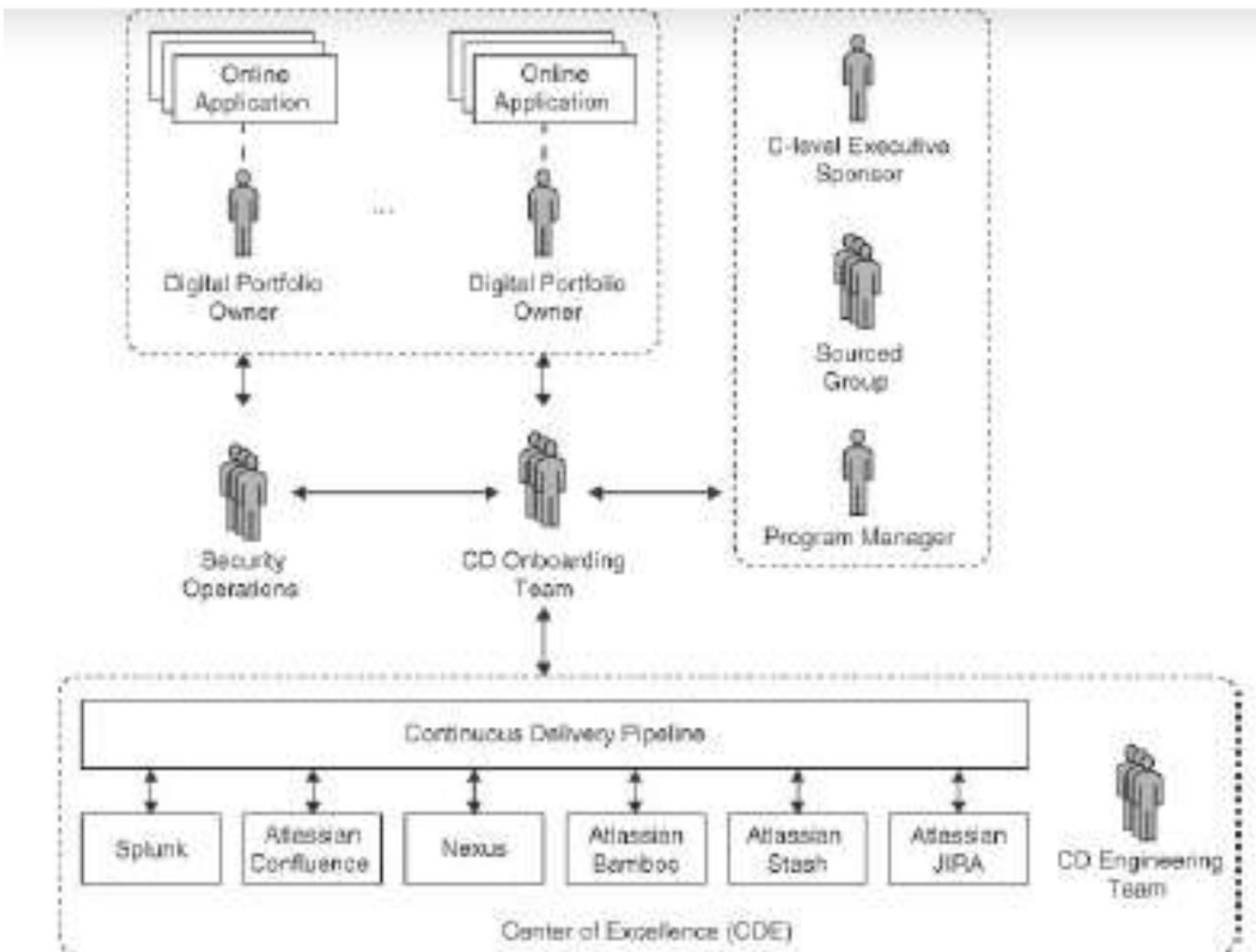
Upgrade process by clients

1. Have your software check for new versions and prompt the user to download and upgrade to the latest version. This is the easiest to implement, but the most painful to use. Nobody wants to watch a download progress bar
2. Download in the background and prompt for installation. In this model, your software periodically checks for updates while running and downloads them silently. After the download is successful, it keeps prompting the user to upgrade to the latest version
3. Download in the background and silently upgrade the next time the application is restarted. Your application might also prompt you to restart now if you'd like to upgrade (as Firefox does)

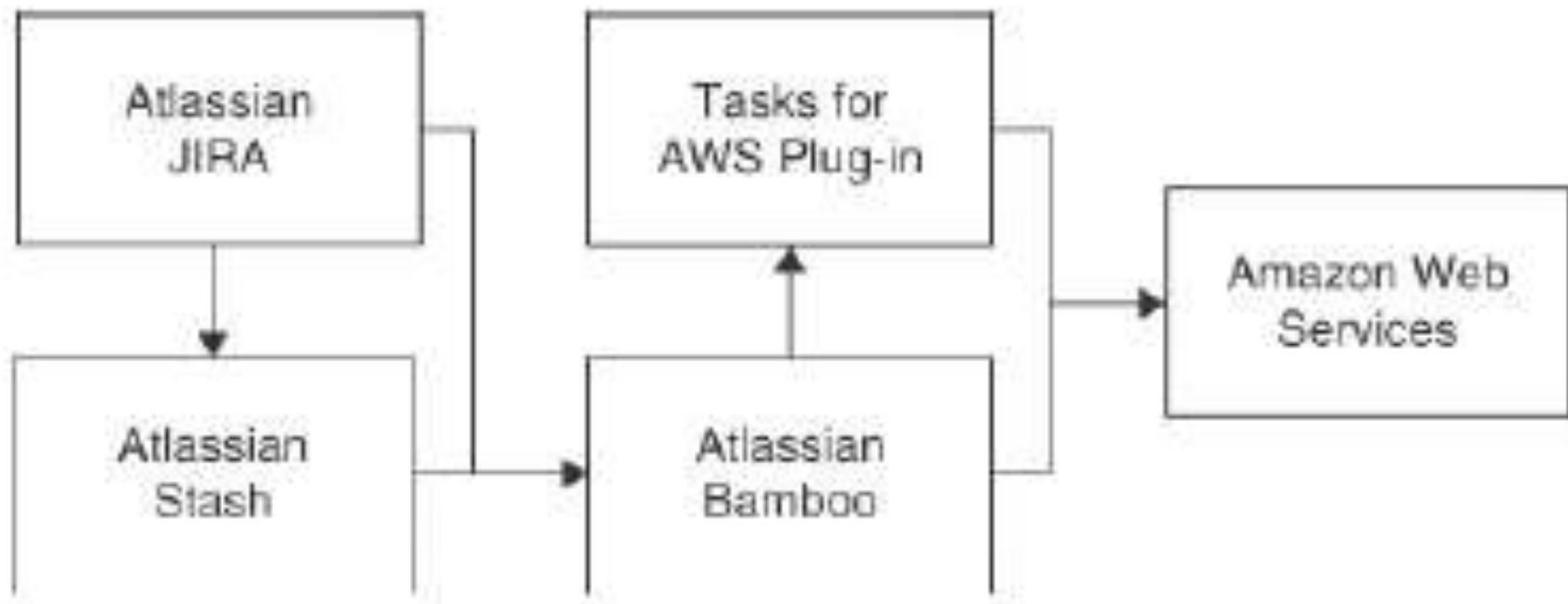
Tips & Tricks

- The People Who Do the Deployment Should Be Involved in Creating the Deployment Process
- Log Deployment activities
- Don't Delete the Old Files, Move Them
- Deployment Is the Whole Team's Responsibility
- Server Applications Should Not Have GUIs
- Have a Warm-Up Period for a New Deployment
- Fail Fast
- Don't Make Changes Directly on the Production Environment

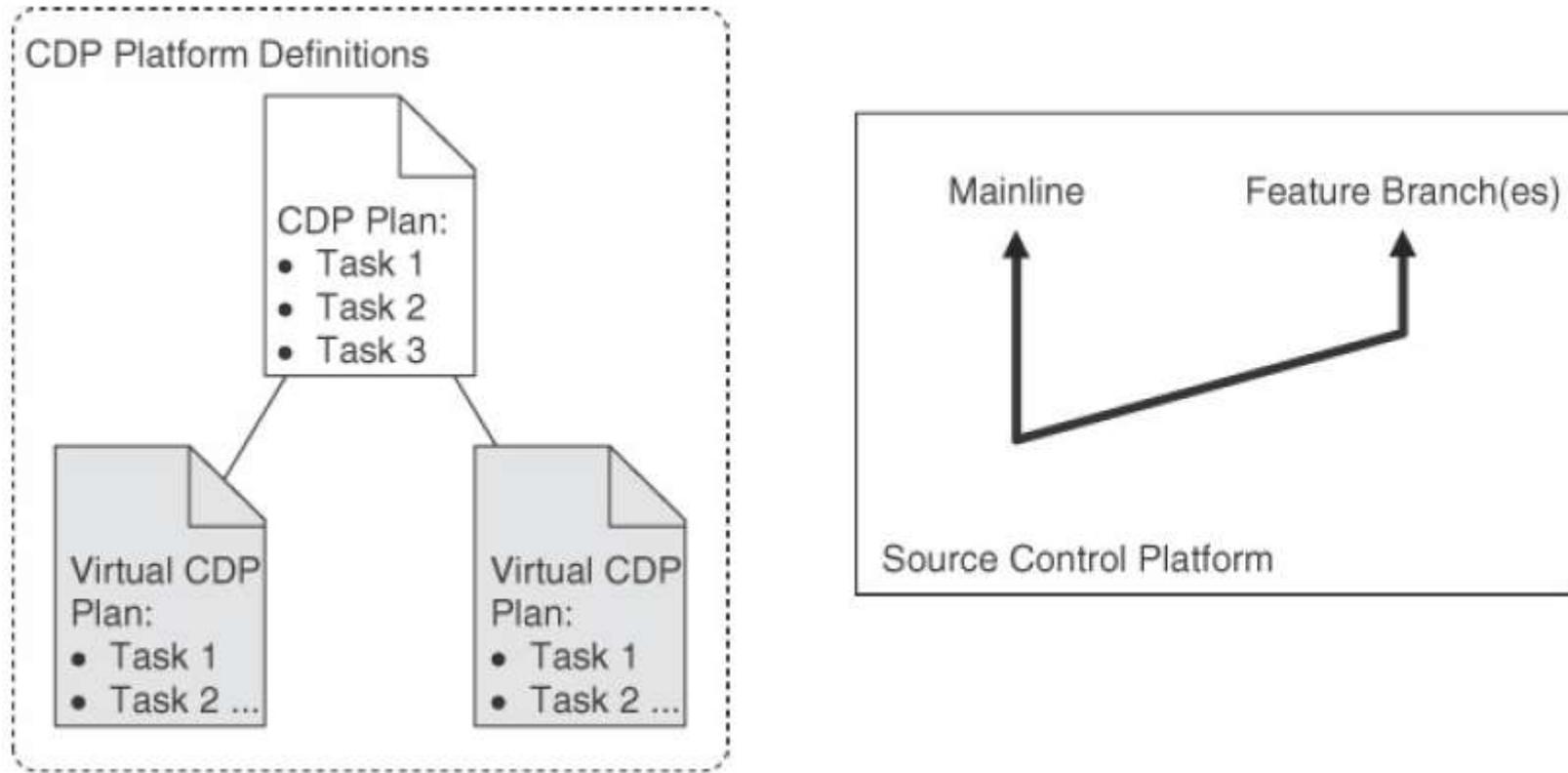
Project team structure

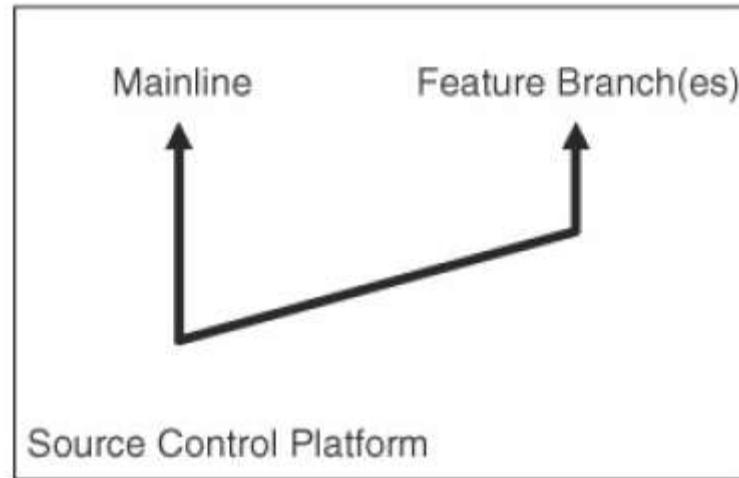
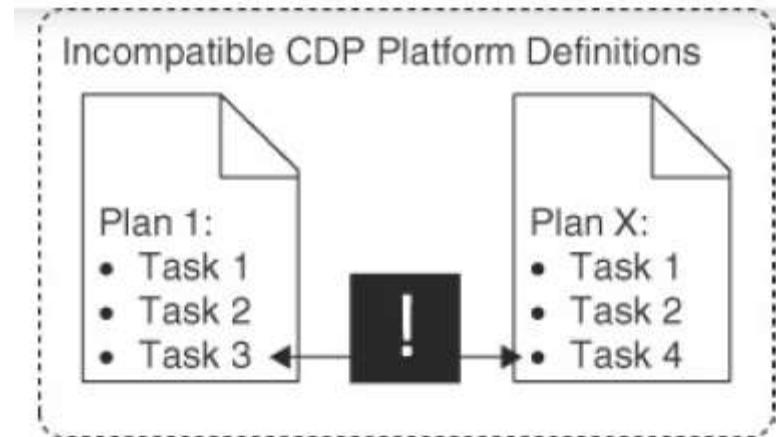


The CDP continuous deployment pipeline

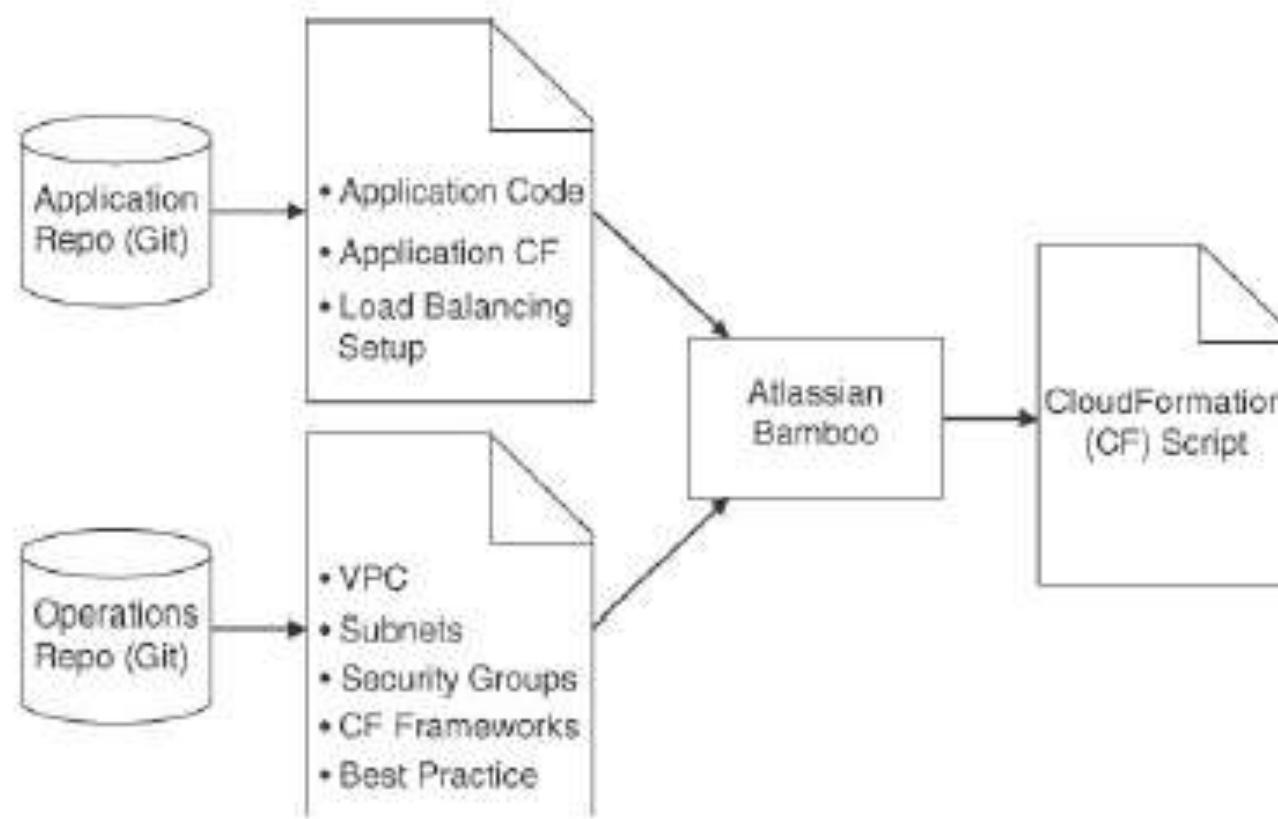


Virtual Plans for branches





Merging operations templates and application configuration into a single CF script





Thank You!

In our next session:



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps





CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 11 and 12

Course Objectives

11

Continuous Monitoring

- Need for continuous monitoring
- Goals of monitoring
- Challenges of monitoring under continuous change
- Alert management
- Analytics
- Continuous customer feedback and optimization
- Use of ELK (Elasticsearch, Logstash, and Kibana) Stack

T1- Chapter 7

R1- Chapter 11

First get your facts; then you can distort them at your leisure

Introduction

- Monitoring has a long history
- Earliest monitors were hardware devices like oscilloscop
- Monitoring Activities:
 - Collecting metrics at various levels(resources, OS, middleware, application- level)
 - Graphing and analyzing metrics, logging, generating alerts concerning system health status
 - Measuring user interactions



Monitoring refers to the process of observing and recording system state changes and data flows

Monitoring

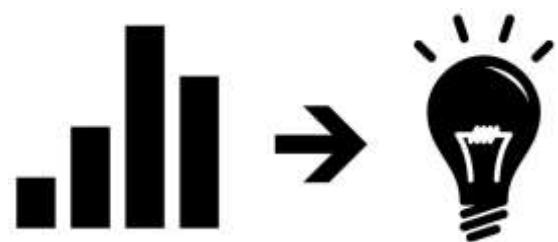
- The methods of information collection include metrics and logs
- Monitoring includes gathering basic system-level metrics such as if a server is up or down, how much memory and CPU are being used, and how full each disk is
- Higher-level application monitoring, which can range from how many user requests a web server is handling, how many items are queued in a queuing system, how long a given web page takes to load, and what are the longest-running queries going into a database
- **While once solely the domain of system and network administrators, as software grows more complex and teams collaborate more, people are beginning to realize that monitoring is a core reflection of product health**

Monitoring Contd

- Monitoring is the process of tracking the current state of your systems and environment, usually with the goal of checking whether or not they meet some predefined conditions of what constitutes the desired state. Often monitoring, alerting, and testing are conflated
- This leads to confusion about what we are trying to accomplish or build. **As mentioned, monitoring usually runs on a predefined schedule while tests are run in response to changes**
- Alerts are automated communications sent to humans about the results of a test or a monitor

Insights from Monitoring

1. Identifying failures and the associated faults both at runtime and during postmortems held after a failure has occurred.
2. Identifying performance problems of both individual systems and collections of interacting systems.
3. Characterizing workload for both short- and long-term capacity planning and billing purposes.
4. Measuring user reactions to various types of interfaces or business offerings. A/B testing
5. Detecting intruders who are attempting to break into the system



Monitoring System

- State changes can be expressed by direct measurement of the state or by logs recording updates that impact part of the state
- Data flows can be captured by logging requests and responses between both internal components and external Systems
- The software supporting such a process is called a *monitoring system*



Challenges in monitoring

- Monitoring under continuous changes is difficult
- The cloud environment introduces different levels from application programming interface (API) calls to VM resource usage
- Managing logs becomes a challenge in large-scale distributed systems



Goal of monitoring

Goal of Monitoring	Source of Data
Failure detection	Application and infrastructure
Performance degradation detection	Application and infrastructure
Capacity planning	Application and infrastructure
User reaction to business offerings	Application
Intruder detection	Application and infrastructure



Monitoring changes to configurations and resource specification files
enables the detection of such errors

Items to be monitored?

- Inputs, resources, and outcomes
- The **resources** can be hard resources such as CPU, memory, disk, and network—even if virtualized
- They can also be soft resources such as queues, thread pools, or configuration specifications
- The **outcomes** include items such as transactions and business-oriented activities



Failure Detection

- Any element of the physical infrastructure can fail
- The cause can be anything from overheating to mice eating the cables.
- Total failures are relatively easy to detect: No data is flowing where data used to flow
- Partial failures that are difficult to detect, for instance: a cable is not firmly seated and degrades performance; before a machine totally fails because of overheating it experiences intermittent failure
- Software can also fail, either totally or partially. Total failure is relatively easy to detect – Partial software failures have myriad causes

Detecting Software Failures

1. The monitoring software performs health checks on the system from an external point
2. A special agent inside the system performs the monitoring
3. The system itself detects problems and reports them

Performance degradation detection

1. Latency
2. Throughput
3. Utilization

Capacity Planning

- Long term capacity planning involves humans and has a time frame on the order of days, weeks, months, or even years
- Short-term capacity planning is performed automatically and has a time frame on the order of minutes

User interaction

- Latency of a user request
- Reliability of the system with which the user is interacting
- Effect of a particular business offering or user interface modification
- Organization's particular set of metrics



User interaction monitoring

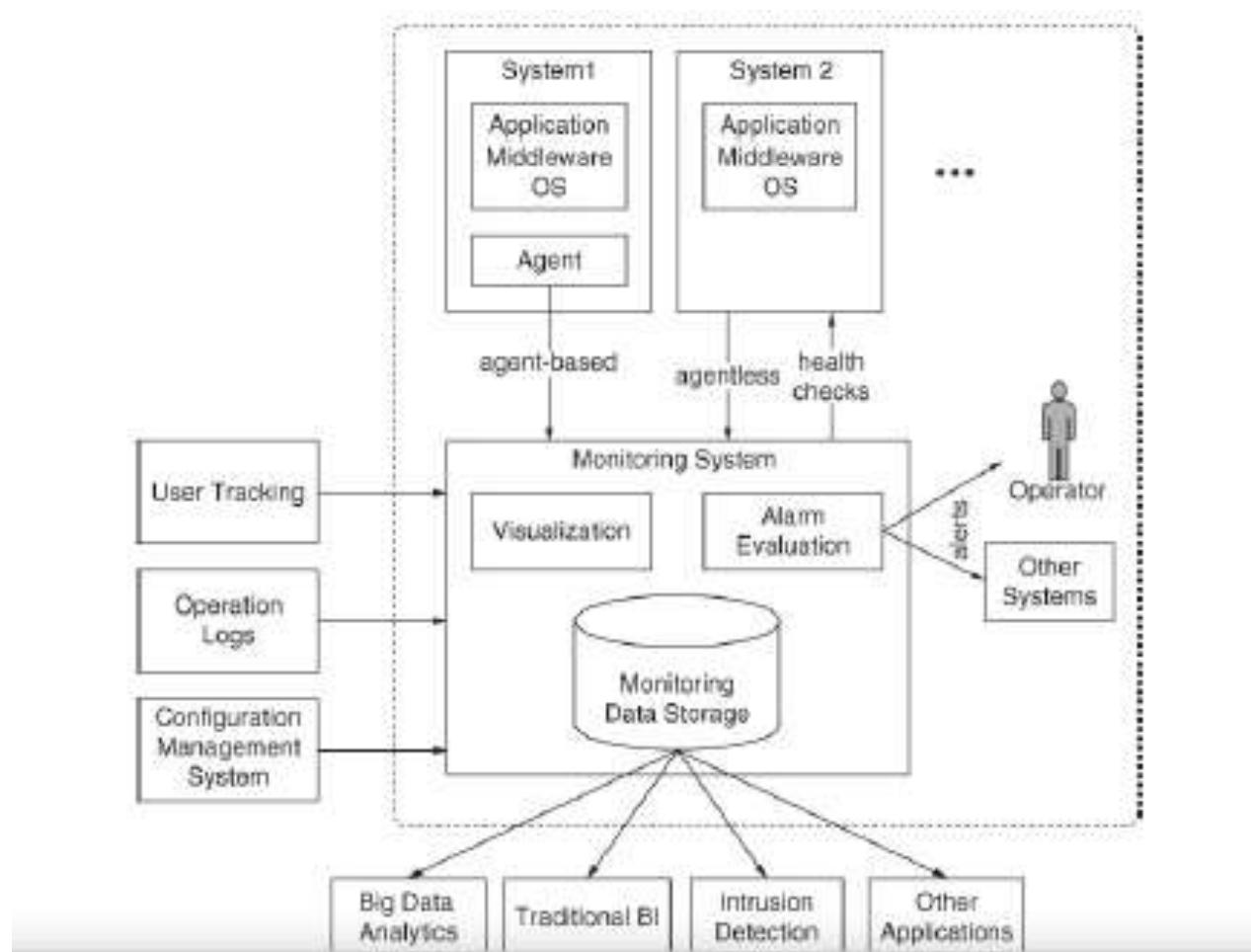
- Real user monitoring
- Synthetic monitoring

Intrusion Detection

- Intruders can break into a system by subverting an application, for example, through incorrect authorization or a man-in-the-middle attack
- Applications can monitor users and their activities to determine whether the activities are consistent with the users' role in the organization or their past behaviour

For instance, if user John has a mobile phone using the application, and the phone is currently in Australia, any log-in attempts from, say, Nigeria should be seen as suspicious.

How to monitor



Agent Based & Agent Less Monitoring



The system to be monitored already has internal monitoring facilities that can be accessed through a defined protocol

For example, the Simple Network Management Protocol (SNMP) is a common mechanism for gathering metrics from servers and network equipment

It is especially useful on network equipment because that equipment often comes as a closed system and you cannot install monitoring agents.

Windows Management Instrumentation (WMI) provides access to management data for Windows systems. You can use protocols like Secure Shell (SSH) to remotely access a system and retrieve available data. Agentless monitoring is particularly useful when you cannot install agents, and it can simplify the deployment of your monitoring system

Agent Based & Agent Less Monitoring



Agent Based	Agent Less
Periodically check a system or conduct performance monitoring from an external user's point of view	Better in terms of deployment & maintenance effort
An agent on a host can communicate with the OS and applications locally and send all collected information over a single channel. Agent-based approach optimizes network traffic and processing overhead.	Less secure if the collection repository is outside of your network because more ports need to be opened and firewall rules relaxed to allow different layers of a system to communicate its data to the external world

Collection & Storage

- The core of monitoring is recording and analyzing time series data, namely, a sequence of time-stamped data points.
- The system being monitored will generate time-stamped event notifications(logs) at various levels of severity
- The monitoring system can conduct direct measurement or collect existing data, statistics, or logs and then turn them into metrics, which have a set of properties usually indicating time and space.

Key challenges

- Collating related items by Time
- Collating related items by context
- Handling the volume of monitoring data

Round-Robin Database

- Popular time series database
- Designed for storing and displaying time series data with good retention policy configuration capabilities
- As we are moving into the big data age, big data storage and processing solutions are increasingly used for monitoring data.
- You can treat your monitoring data as data streams feeding into streaming systems for real-time processing, combined with (big) historical data.
- You can load all your data into big data storage systems such as Hadoop Distributed File System (HDFS) or archive it in relatively inexpensive online storage systems such as Amazon Glacier.

When to change the monitoring configuration?



- An alert
- Deployment
- Canary Deployment
- Rolling upgrade
- Feature activation or deactivation
- Changes to infrastructure software including devops tools
- Changes to any configuration parameters

Interpreting Monitoring data

- Logs
- Logs are used during operations to detect and diagnose problems.
- Logs are used during debugging to detect errors
- Logs are used during post problem forensics to understand the sequence that led to a particular problem

Logging

- Logging is the generation, filtering, recording, and analysis of events that occur on a system due to operating system or software messages. When tracking down the source of a software issue, one of the first things that engineers often do is to check the logs for any relevant error messages.
- Logs can be treasure troves of useful information, and with storage getting cheaper and cheaper, just about any log you might want can be saved and stored for later use.

Logging

- Logs can come from the apps you develop, from third-party tools you use, and even from the OS itself. As there is no logging standard across software, it can be difficult to categorize and qualify events within logs to identify patterns of concern
- A single system can generate hundreds or even thousands of lines of logs a day
- In modern environments that have dozens of applications running on hundreds or thousands of servers, the sheer volume of log data can be overwhelming; it is no longer a simple matter of searching through one logfile. Much work has gone into developing applications that handle the storage and searching of logs
- The complexities of logging solutions are beyond the scope of this chapter, but their value should not be underestimated

Rules about writing logs

- Logs should have a consistent format
- Logs should include an explanation for why this particular log message was produced. Tags such as “error condition detected” or “tracing of code” can be used
- Log entries should include context information. Context is more than date and time; it also includes information to support tracking the log entry such as:
 - Source of the log entry within the code
 - Process ID for the process executing when the log entry was produced
 - Request ID for the request that caused that process to execute this log producer
 - VM ID for the VM that produced this message
 - Logs should provide screening information

Alerts & Alarms

- Monitoring systems inform the operator of significant events
- This information can be in the form of either an alarm or an alert. Technically, alerts are raised for purposes of informing and may be in advance of an alarm (e.g., the datacenter temperature is rising), whereas alarms require action by the operator or another system (e.g., the datacenter is on fire).
- Alarms and alerts are generated based on configurations set by the operators.
- Alarms and alerts can be triggered by events (e.g., a particular physical machine is not responding), by values crossing a threshold (e.g., the response time for a particular disk is greater than an acceptable value), or by sophisticated combinations of values and trends.

Issues with alarms and alerts

- How do you configure your monitoring system to reduce false positives (alarms without the necessity for action) and false negatives (the necessity for action without an alarm being raised)?
- How do you configure your monitoring system so that the alerts provide necessary information to diagnose an alarm?

Improve usefulness of alerts and alarms



- Introduce context to your alarms
- Alarms can not only go off if something happens, they can also be set to go off if an expected event did not happen
- Aggregate different alerts that are likely referring to the same events.
- Set clear severity levels and urgency levels so people or systems receiving the alerts can act accordingly

Diagnosis & Reaction

- Operators often use monitoring systems to diagnose the causes and observe the progress of mitigation and recovery
- Operators will try to correlate events, dive into details and execute queries, and examine logs
- Concurrently, they manually trigger more diagnostic tests and recovery actions (such as restarting processes or isolating problematic components) and observe their effects from the monitoring system.

Monitoring DevOps Processes

Damon Edwards lists five things that are important to monitor:

1. A business metric
2. Cycle time
3. Mean time to detect errors
4. Mean time to report errors
5. Amount of scrap (rework)

Challenges

Challenge 1: Monitoring Under Continuous Changes

Challenge 2: Bottom-Up vs. Top-Down and Monitoring in the Cloud

Challenge 3: Monitoring a Microservice Architecture

Challenge 4: Dealing with Large Volumes of Distributed (Log) Data

Tools

- Nagios
- Sensu & Icinga
- Ganglia
- Graylog2, Logstash, and Splunk
- CloudWatch and the like
- Kafka
- Stream Processing tools (Storm, Flume, S4)
- Apdex (Application Performance Index)

Metrics

- Metrics are the collection of qualitative & quantitative measurements
- Compared against some benchmark or established standard, tracked for analytics, or tracked for historical purposes. Often metrics are siloed within functional organizations, which can impact choosing the right direction for product development
- Metrics are one of the key parts of monitoring; data can be gathered and stored for nearly any part of even the most complex web software, and different teams can have different metrics that they keep track of and use in their work
- **StatsD and Graphite are very commonly used and a powerful combination for tracking, storing, and viewing metrics**

Infrastructure automation

Infrastructure automation is provisioning elements of infrastructure through code—treating this code just like the rest of your software, with the ability to recover business through data backups, code repository, and compute resources.

“Treat infrastructure code like the rest of your software” means the code is developed using a common local development environment, versioned in version control, versioned in artifacts in an artifact repository, tested, and verified before

Terms for infra mgmt

- Configuration drift (Servers will change or drift away from desired config over time)
- MTBF (Mean Time Between Failures)
- MTTR (Mean Time To Repair)
- Availability $\text{Availability} = \text{MTBF} / (\text{MTBF} + \text{MTTR})$.
- Capacity management
- **Snowflake server** – A snowflake server is a server that has gotten to its current, desired configuration by way of many manual changes, often a combination of command-line sorcery, configuration files, hand-applied patches, and even GUI configurations and installations.

Infrastructure automation should provide:



- Prevent Configuration drift
- Elimination of snowflake servers
- Versioned artifacts of infrastructure code
- Minimizing complexity

Important Points

- If the various infrastructure automation tools available do not cater to your exact needs, it's still more cost-effective to help expand the fundamental features or reliability of software that exists in the space rather than rolling your own
- Infrastructure automation leads to repeatable, consistent, documented, auditable, and resilient processes. This frees up time, improves efficiency of staff, allows for more flexibility, and facilitates risk measurement
- Infrastructure automation also increases the degree of confidence that the machine setup and deployment are identical, reducing the amount of time spent debugging problems based on system differences

Continuous customer feedback and optimization



- Continuous customer feedback provides the visual evidence and full context for analyzing customer behavior and pinpointing customer pain points
- Feedback can be applied during both pre- and post-production phases to maximize the value of every customer visit and ensure that more transactions are completed successfully
- This allows immediate visibility into the sources of customer struggles that affect their behavior and impact business



How do Netflix, Facebook, Microsoft, LinkedIn, and Cisco monitor their logs?

With ELK

Elasticsearch, Logstash, and Kibana

Why ELK?

Rapid on-premise (or cloud) installation and easy to deploy

- Scales vertically and horizontally
- Easy and various APIs to use
- Ease of writing queries, a lot easier than writing a MapReduce job
- Availability of libraries for most programming/scripting languages
 - Elastic offers a host of language clients for Elasticsearch, including Ruby, Python, PHP, Perl, .NET, Java, and Javascript, and more
- Tools availability
- It's free (open source), and it's quick

Why ELK?

Logstash

Data From Any Source



Elasticsearch

Instantly Search & Analyze



Kibana

Actionable Insight

Logstash is a log pipeline tool that accepts inputs from various sources, executes different transformations, and exports the data to various targets

→ collects and parses logs

Elasticsearch is a NoSQL database that is based on the Lucene search engine

→ indexes and stores the information

Kibana is a visualization layer that works on top of Elasticsearch

→ presents the data in visualizations that provide actionable insights

The Elastic Stack

Plugins

Monitoring

Security

Alerting

User Interface

Kibana

Store, Index,
& Analyze

Elasticsearch

Ingest

Logstash

Beats

Cloud/Hosted

Cloud: Hosted Elasticsearch

ELK Modules

Open Source —

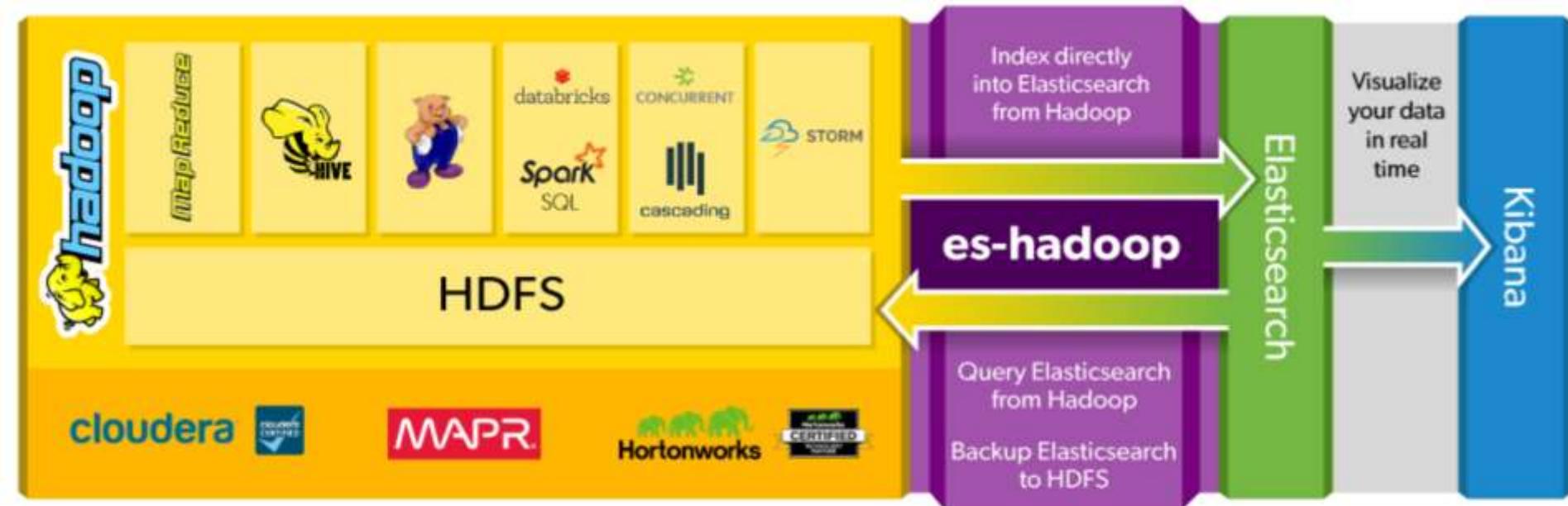
- ElasticSearch
- Logstash
- Kibana
- Beats
 - data shippers – collect, parse & ship

Extension plugins —

- Alerting (Watcher)
 - Proactively monitoring and alerting based on elasticsearch queries or conditions
- Security (Shield)
 - Protect and provide security to elastic stack
- Monitoring (Marvel)
 - Monitor and diagnose health and performance of elastics cluster
- Graph
 - discover and explore the relationships live in data by adding relevance to your exploration

Connect Speedy Search with Big Data Analytics – Elasticsearch for Apache Hadoop

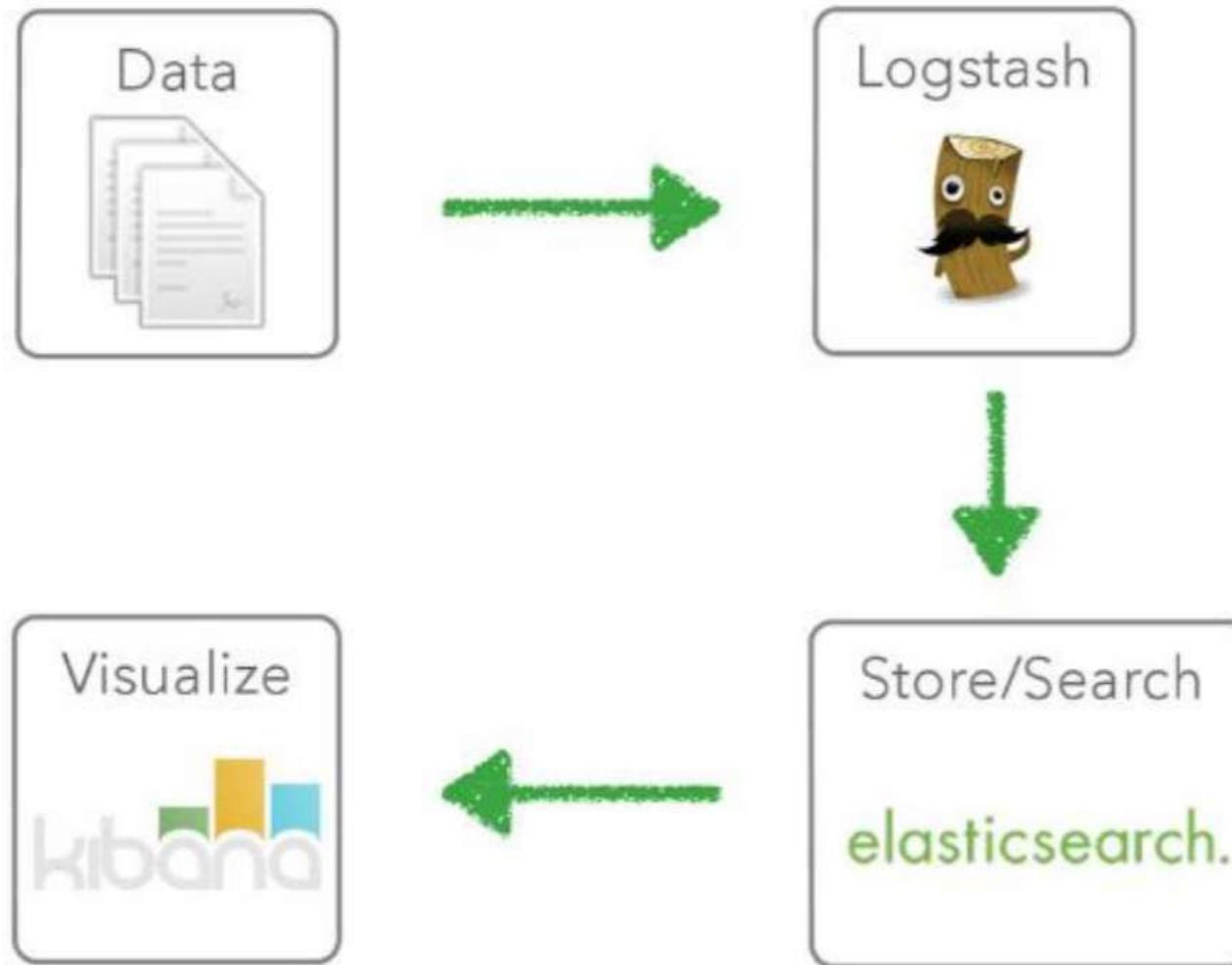
innovate achieve lead



ES-Hadoop -- a two-way connector

- **Read and write data to ES and query it in real time**

The ELK Stack

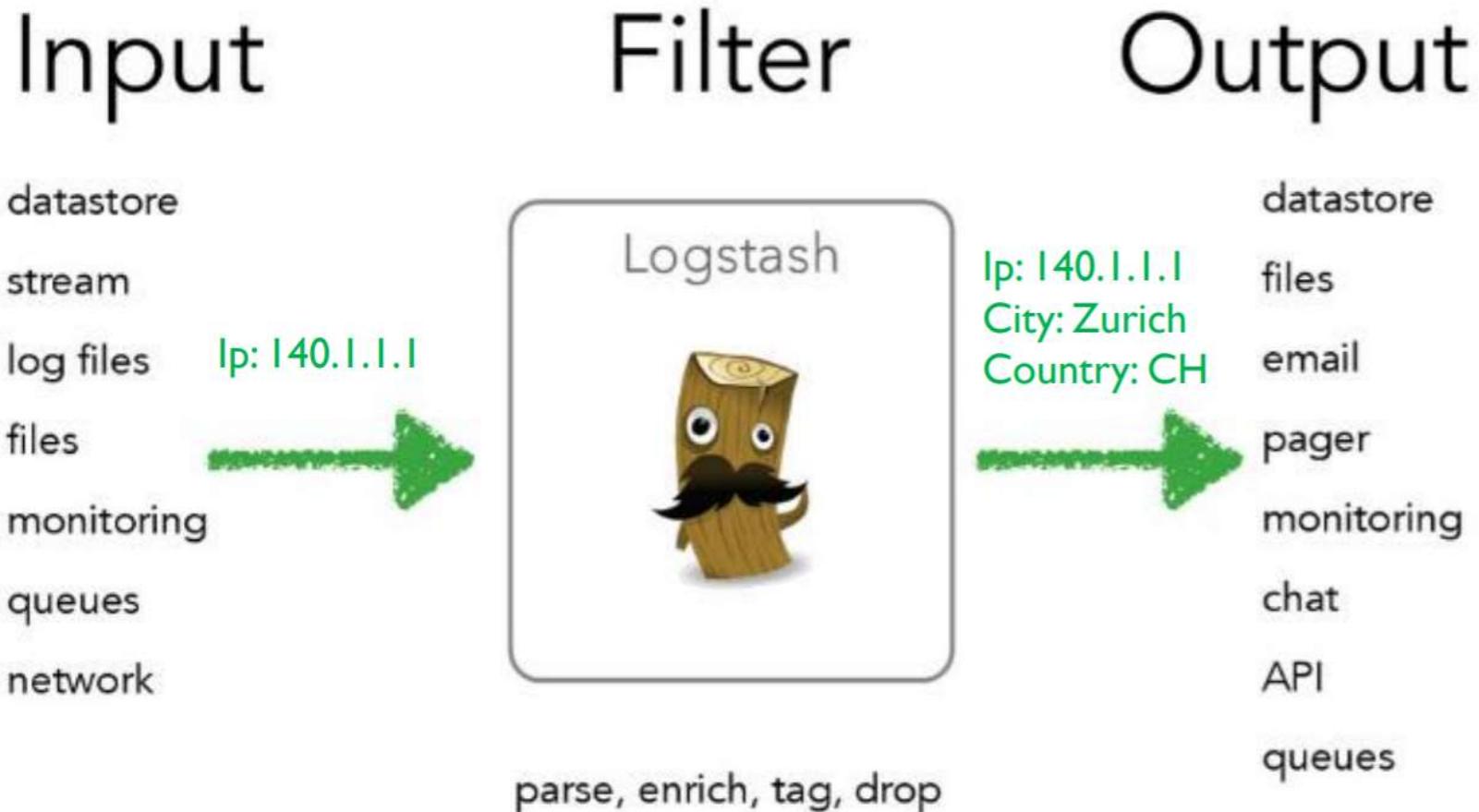


Logstash

- Managing events and logs
- Collect data
- Parse data
- Enrich data
- Store data
- Open Source: Apache License 2.0



Logstash Arch.



How Logstash works?

```
input {  
  file {  
    path => "/tmp/access_log"  
    start_position => "beginning"  
  }  
}
```

```
filter {  
  grok {  
    match => { "message" => "%{COMBINEDAPACHELOG}" }  
  }  
}
```

```
output {  
  elasticsearch {  
  }  
}
```

Logstash Input plugins

- Stdin – Reads events from standard input
- File – Streams events from files (similar to “tail -0F”)
- Syslog – Reads syslog messages as events
- Eventlog – Pulls events from the Windows Event Log
- Imap – read mail from an IMAP server
- Rss – captures the output of command line tools as an event
- Snmptrap – creates events based on SNMP trap messages
- Twitter – Reads events from the Twitter Streaming API
- Irc – reads events from an IRC server
- Exec – Captures the output of a shell command as an event
- Elasticsearch – Reads query results from an Elasticsearch cluster
-

Logstash Filter Plugins

- grok – parses unstructured event data into fields
- Mutate – performs mutations on fields
- Geoip – adds geographical information about an IP address
- Date – parse dates from fields to use as the Logstash timestamp for an event
- Cidr – checks IP addresses against a list of network blocks
- Drop – drops all events
- ...

Logstash Output Plugins

- Stdout – prints events to the standard output
- Csv – write events to disk in a delimited format
- Email – sends email to a specified address when output is received
- Elasticsearch – stores logs in Elasticsearch
- Exec – runs a command for a matching event
- File – writes events to files on disk
- mongoDB – writes events to MongoDB
- Redmine – creates tickets using the Redmine API
-

Dec 23 14:30:01 louis CRON[619]: (www-data) CMD (php /usr/share/cacti/site/poller.php
>/dev/null 2>/var/log/cacti/poller-error.log)

```
filter {
  if [type] == "syslog" {
    grok {
      match => { "message" => "%{SYSLOGTIMESTAMP:syslog_timestamp}
%{SYSLOGHOST:syslog_hostname} %{DATA:syslog_program}(?:\[ %{POSINT:syslog_pid}\])?:
%{GREEDYDATA:syslog_message}" }
      add_field => [ "received_at", "%{@timestamp}" ]
      add_field => [ "received_from", "%{host}" ]
    }
    date {
      match => [ "syslog_timestamp", "MMM d HH:mm:ss", "MMM dd HH:mm:ss" ]
    }
  }
}
```

```
{  
    "message" => "Dec 23 14:30:01 louis CRON[619]: (www-data) CMD (php  
/usr/share/cacti/site/poller.php >/dev/null 2>/var/log/cacti/poller-error.log)",  
    "@timestamp" => "2013-12-23T22:30:01.000Z",  
    "@version" => "1",  
    "type" => "syslog",  
    "host" => "0:0:0:0:0:0:1:52617",  
    "syslog_timestamp" => "Dec 23 14:30:01",  
    "syslog_hostname" => "louis",  
    "syslog_program" => "CRON",  
    "syslog_pid" => "619",  
    "syslog_message" => "(www-data) CMD (php /usr/share/cacti/site/poller.php  
/>/dev/null 2>/var/log/cacti/poller-error.log)",  
    "received_at" => "2013-12-23 22:49:22 UTC",  
    "received_from" => "0:0:0:0:0:0:1:52617",  
    "syslog_severity_code" => 5,  
    "syslog_facility_code" => 1,  
    "syslog_facility" => "user-level",  
    "syslog_severity" => "notice"  
}
```

Example: Web server logs

```
    "message" => "83.149.9.216 - - [28/May/2014:16:13:42 -0500] \\"GET / presentations/logstash-monitorama-2013/images/kibana-search.png HTTP/1.1\\" 200 203023 \\"http://semicomplete.com/presentations/logstash-monitorama-2013/\\" \\"Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36\\\"",  
  
    "@version" => "1",  
    "@timestamp" => "2014-05-28T21:13:42.000Z"  
    "host" => "kryptic.local",  
    "clientip" => "83.149.9.216",  
    "ident" => "-",  
    "auth" => "-",  
    "timestamp" => "28/May/2014:16:13:42 -0500",  
    "verb" => "GET",  
    "request" => "/presentations/logstash-monitorama-2013/images/kibana-search.png",  
    "httpversion" => "1.1",  
    "response" => "200",  
    "bytes" => "203023",  
    "referrer" => "\\"http://semicomplete.com/presentations/logstash-monitorama-2013/\\"",  
    "agent" => "\\"Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36\\\""
```

grok

date

```
"geoip" => {
    "ip" => "83.149.9.216",
    "country_code2" => "RU",
    "country_code3" => "RUS",
    "country_name" => "Russian Federation",
    "continent_code" => "EU",
    "region_name" => "48",
    "city_name" => "Moscow",
    "latitude" => 55.75219999999999,
    "longitude" => 37.6156,
    "timezone" => "Europe/Moscow",
    "real_region_name" => "Moscow City",
    "location" => [
        [0] 37.6156,
        [1] 55.75219999999999
    ]
},
"useragent" => {
    "name" => "Chrome",
    "os" => "Mac OS X 10.9.1",
    "os_name" => "Mac OS X",
    "os_major" => "10",
    "os_minor" => "9",
    "device" => "Other",
    "major" => "32",
    "minor" => "0",
    "patch" => "1700"
}
```

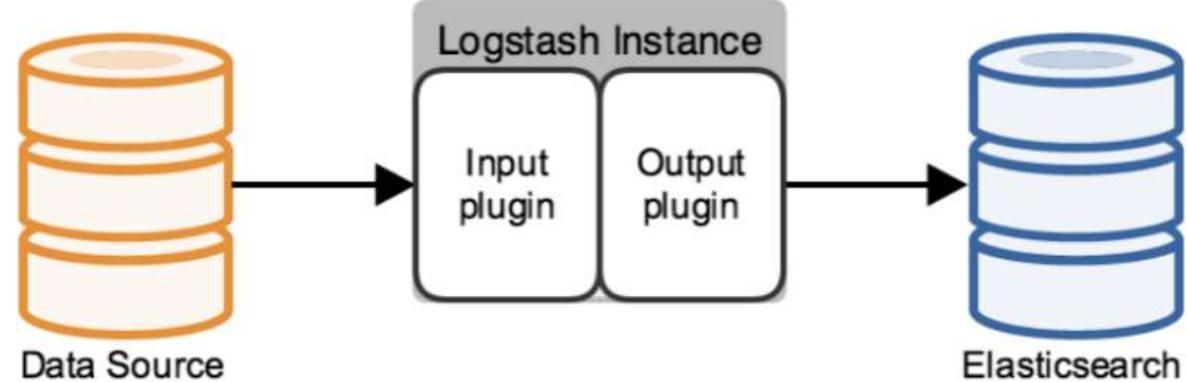
geoip

useragent

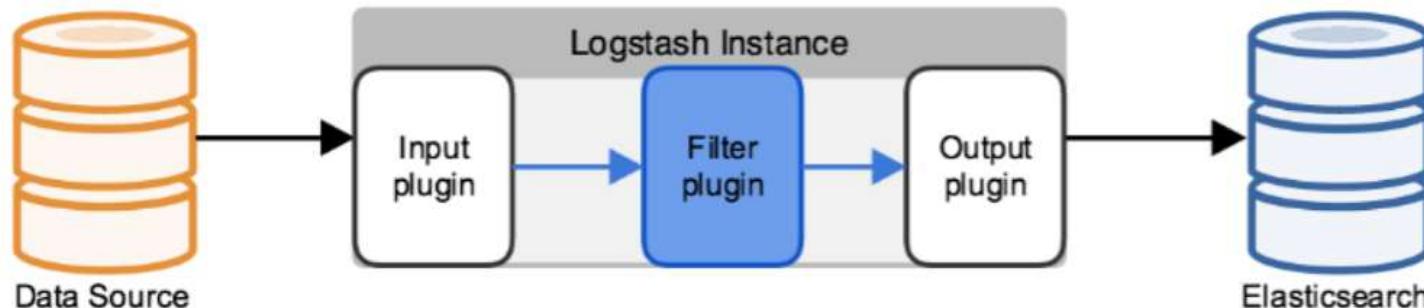
Deploying & Scaling Logstash



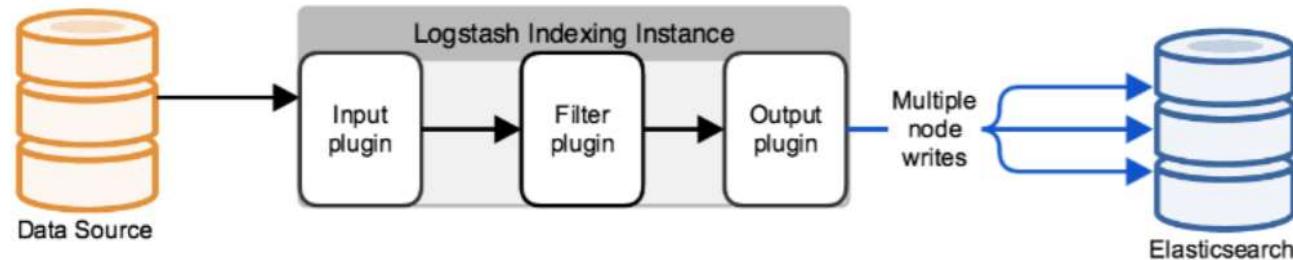
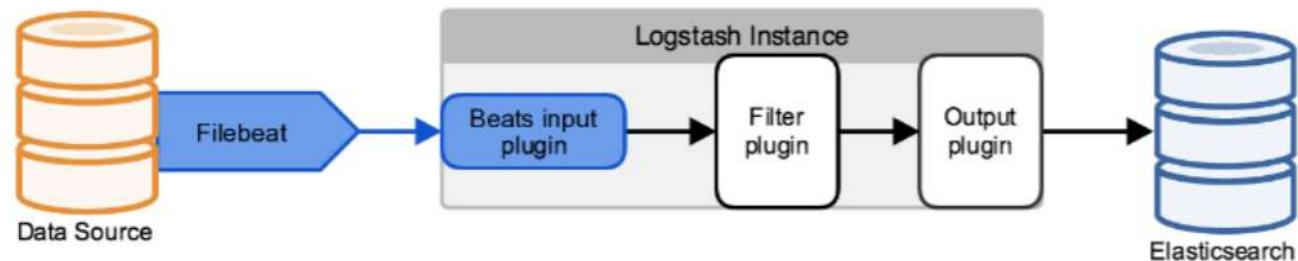
Minimal installation



Using Filters

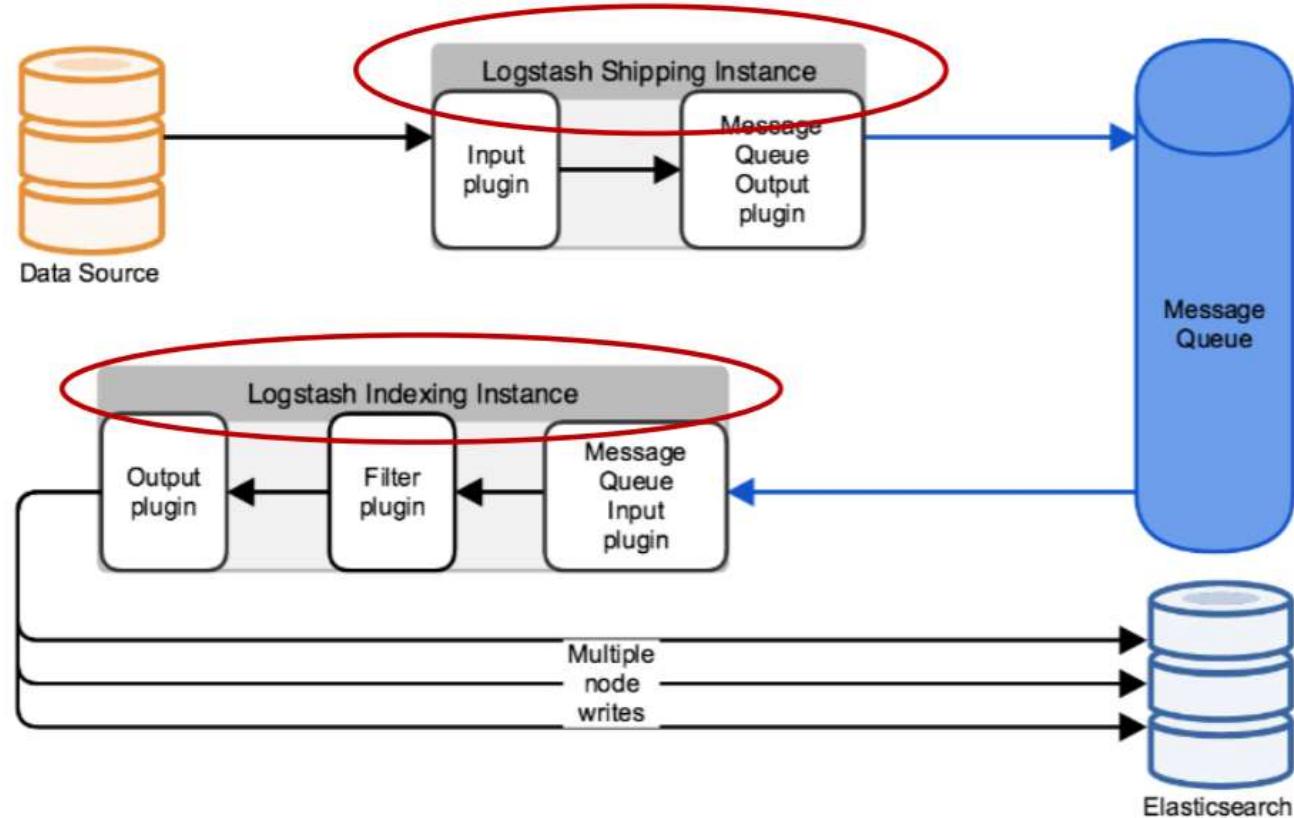


Using log shipper to minimize the resource demands on Logstash

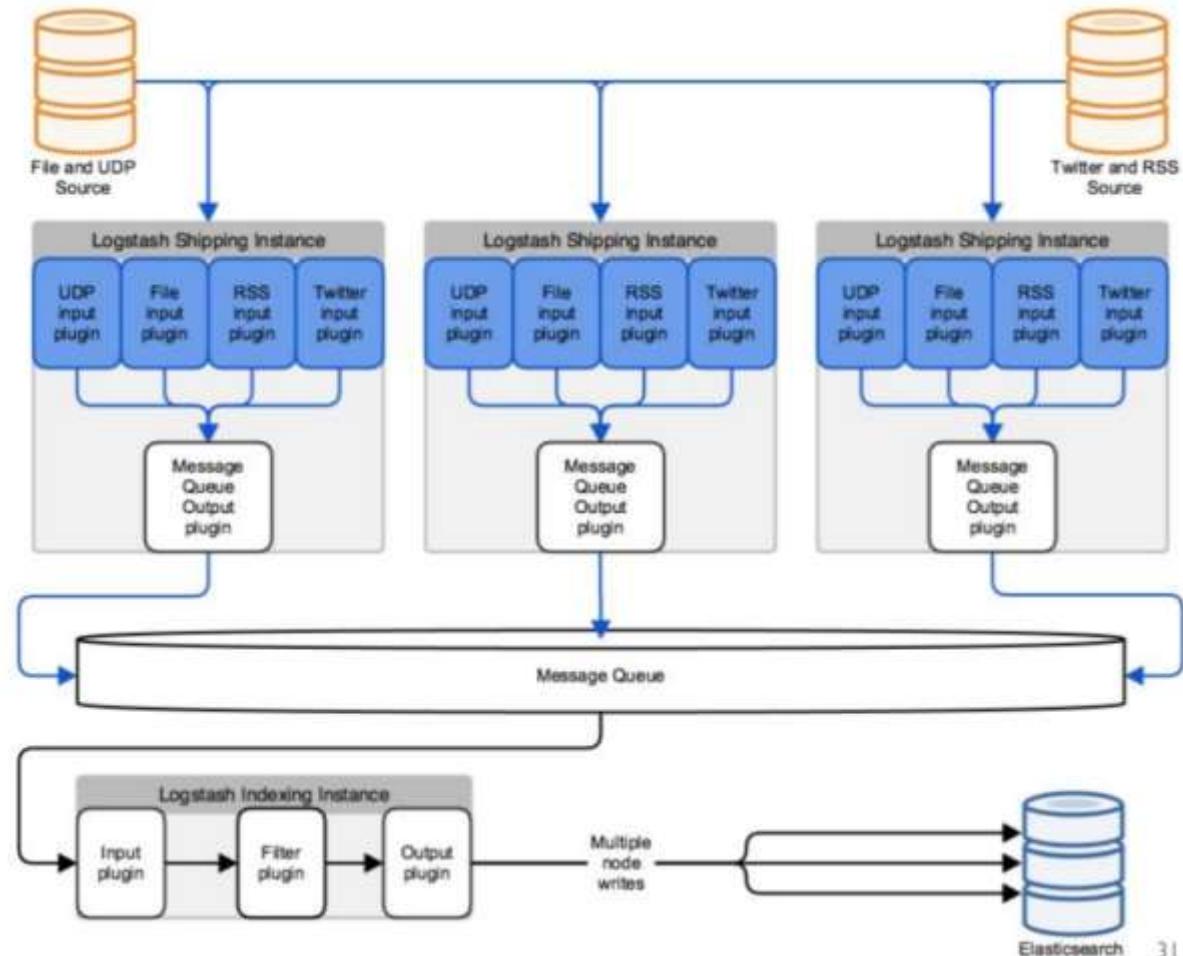


Scaling to a Larger Elasticsearch Cluster

Managing Throughput Spikes with Message Queuing



Multiple Connections for Logstash High Availability



Elastic Search

Schema-flexible

- Built on top of [Apache Lucene™](#), a full-text search-engine library
- A Schema-free, REST & JSON based distributed search engine with real-time analytics
- Capable of scaling to hundreds of servers and petabytes of structured and unstructured data
- Open Source: Apache License 2.0

Real scalability comes from horizontal scale

- [Wikipedia](#) uses Elasticsearch to provide full-text search with highlighted search snippets, and `search-as-you-type` and `did-you-mean` suggestions
- [The Guardian](#) uses Elasticsearch to combine visitor logs with social-network data to provide real-time feedback to its editors about the public's response to new articles
- [Stack Overflow](#) combines full-text search with geolocation queries and uses `more-like-this` to find related questions and answers
- [GitHub](#) uses Elasticsearch to query 130 billion lines of code

Elastic Search vs RDB

ElasticSearch	Relational DB
Index	Database
Type	Table
Document	Row
Field	Column
Shard	Partition
Mapping - (everything is indexed)	Schema Index
Query DSL (<i>domain specific language</i>)	SQL

Shards are how
Elasticsearch distributes
data around your
cluster

Shard?

- a shard is a single instance of Lucene, and is a complete search engine in its own right
- Documents are stored and indexed in shards □ shards are allocated to nodes in your cluster
- As your cluster grows or shrinks, Elasticsearch will automatically migrate shards between nodes so that the cluster remains balanced
- A shard can be either a primary shard or a replica shard
- Each document in your index belongs to a single primary shard
- A replica shard is just a copy of a primary shard

Elastic Search Clustering – Single node cluster

CLUSTER

NODE 1 - ★ MASTER

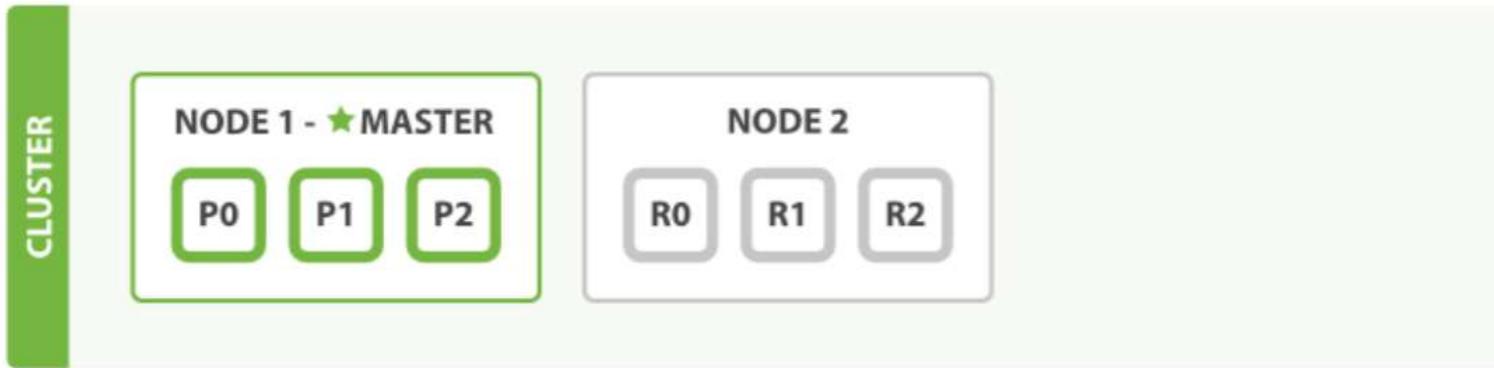
- **Node** = running instance of ES
- **Cluster** = 1+ nodes with the same cluster.name
- Every cluster has **1 master node**
- 1 Cluster can have any number of indexes

CLUSTER

NODE 1 - ★ MASTER

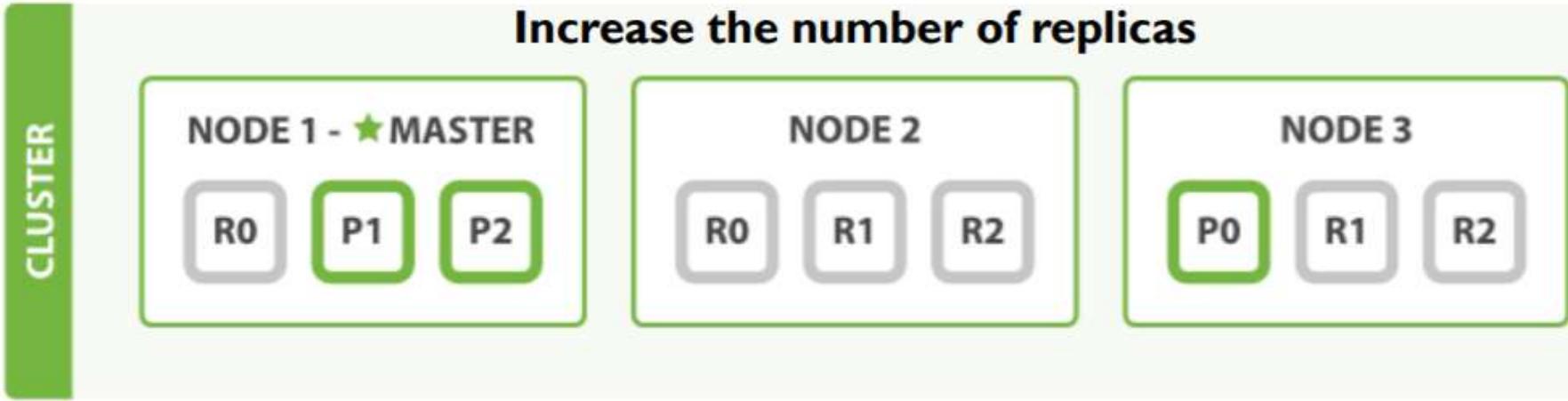
P0 P1 P2

Elastic Search Clustering – Adding a second node



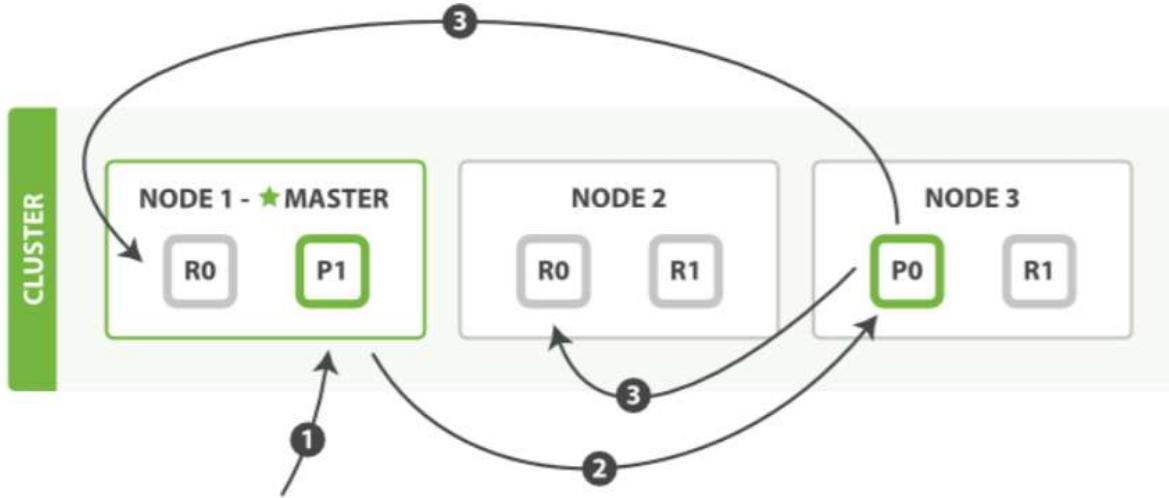
- A cluster consists of one or more nodes with the same cluster.name
 - All primary and replica shards are allocated
 - Each index has one primary (P) and one replica (R) shard
 - Clients talk to any node in the cluster

Elastic Search Clustering – Adding a third node



- More primary shards:
 - faster indexing
 - more scale
- More replicas:
 - faster searching
 - more failover

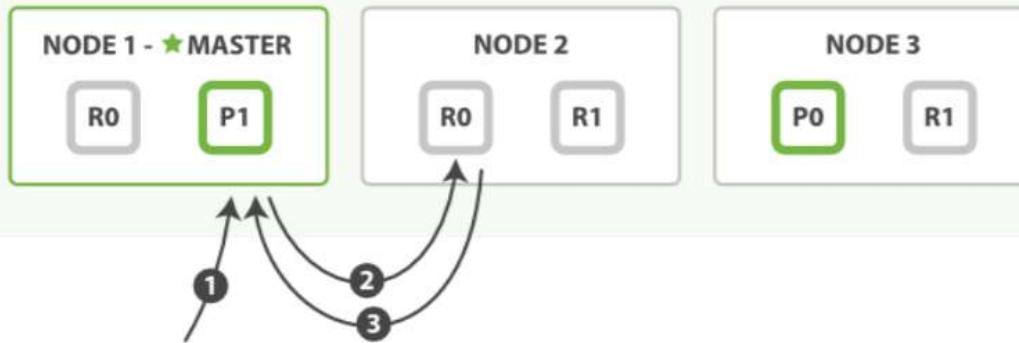
Creating indexing and deleting a document



1. The client sends a create, index, or delete request to Node 1
2. The node uses the document's `_id` to determine that the document belongs to shard 0. It forwards the request to Node 3, where the primary copy of shard 0 is currently allocated
3. Node 3 executes the request on the primary shard. If it is successful, it forwards the request in parallel to the replica shards on Node 1 and Node 2. Once all of the replica shards report success, Node 3 reports success to the coordinating node, which reports success to the client.

Retrieving a document

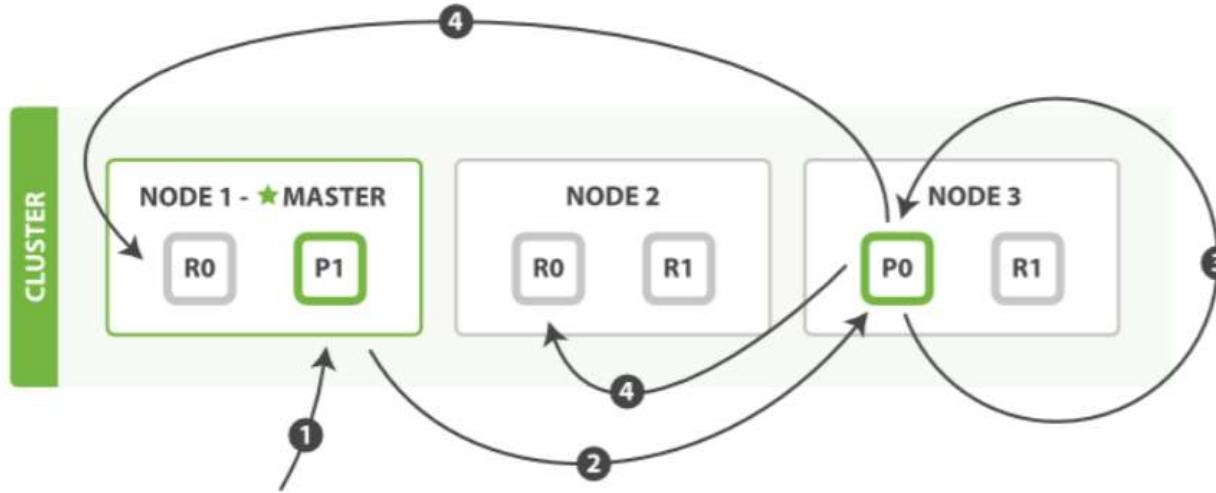
CLUSTER



For read requests, the coordinating node will choose a different shard copy on every request in order to balance the load

1. The client sends a get request to node 1
2. The node uses the document's `_id` to determine that the document belongs to shard 0. Copies of shard 0 exist on all three nodes. On this occasion, it forwards the request to node 2.
3. Node 2 returns the document to node 1, which returns the document to the client.

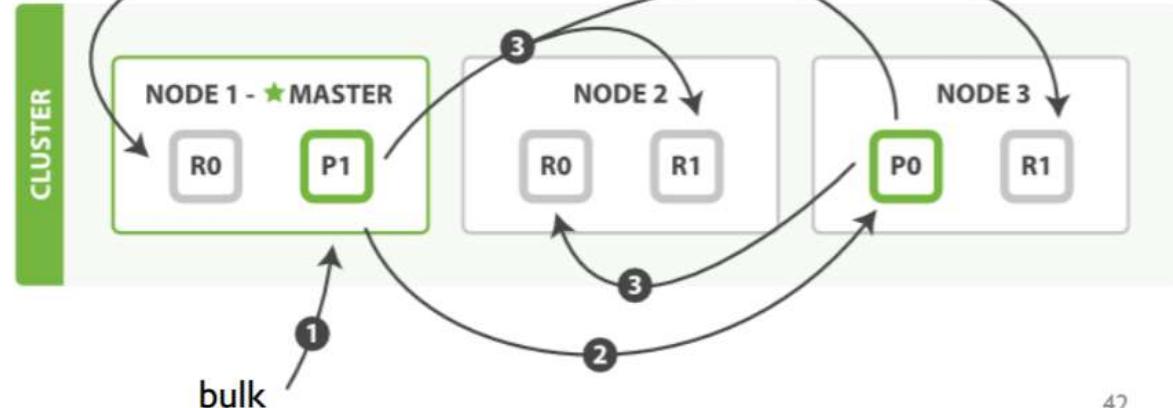
Partial update to a document



When a primary shard forwards changes to its replica shards, it doesn't forward the update request. Instead it forwards **the new version of the full document**.

Multidocument patterns

CLUSTER



- the coordinating node knows in which shard each document lives.
- It breaks up the multidocument request into a multidocument request **per shard**, and forwards these in parallel to each participating node
- Once it receives answers from each node, it collates their responses into a single response

Talking to Elasticsearch

- RESTful API with JSON over HTTP

- Over port 9200
- Access via web client, or command line by `curl` command

```
curl -X<VERB> '<PROTOCOL>://<HOST>:<PORT>/<PATH>?<QUERY_STRING>' -d '<BODY>'
```

- JSON (JavaScript Object Notation) ← the standard format used by NoSQL

```
curl -XGET 'http://localhost:9200/_count?pretty' -d '  
{  
  "query": {  
    "match_all": {}  
  }  
}'
```

- Elasticsearch clients

- Java API, Java REST client, JavaScript API, PHP API, Python API, Perl API...

Indexing a document

Index	Type	Id
curl -XPUT localhost:9200/test/product/1 -d '{"category": "electronics", "price": 129.99, "id": 1, "name": "ipod"}'	product	1
curl -XPUT localhost:9200/test/product/1 -d '{"category": "electronics", "price": 129.99, "id": 1, "name": "ipod"}'	product	1

Document

```
curl -XPOST localhost:9200/test/product -d '{"category": "electronics", "price": 129.99, "name": "ipod"}'  
{"ok":true,"_index":"test","_type":"product","_id":"9wrADN4eS8uXm3gNpDvEJw","_version":1}
```

- Store a document in an index so that it can be retrieved and queried
- Like the **INSERT** keyword in SQL

Retrieving documents

```
curl -XGET 'localhost:9200/test/product/1?pretty'
---
{
  "_index" : "test",
  "_type" : "product",
  "_id" : "1",
  "_version" : 2,
  "exists" : true, "_source" : {"category": "electronics", "price": 129.99, "name": "ipod"}
}
```

- Using GET method to retrieve document
- We can retrieve a specific document if we happen to know its id

Performing Queries

- Using the `q=<query>` form performs a full-text search by parsing the query string value

```
curl -XGET 'localhost:9200/test/product/_search?  
q="ipod"&format=yaml'  
---  
took: 104  
timed_out: false  
_shards:  
  total: 1  
  successful: 1  
  failed: 0  
hits:  
  total: 1  
  max_score: 0.15342641  
  hits:  
    - _index: "test"  
      _type: "product"  
      _id: "1"  
      _score: 0.15342641  
      _source:  
        category: "electronics"  
        price: 129.99  
        name: "ipod"
```

- Query with **query DSL**, which is specified using a JSON request body

```
GET /megacorp/employee/_search  
{  
  "query" : {  
    "match" : {  
      "last_name" : "Smith"  
    }  
  }  
}
```

Domain Specific Language

Query DSL – combining filters

```
SELECT product  
FROM products  
WHERE (price = 20 OR productID = "XHDK-A-1293-#fJ3")  
AND (price != 30)
```



```
GET /my_store/products/_search  
{  
  "query" : {  
    "constant_score" : { ①  
      "filter" : {  
        "bool" : {  
          "should" : [  
            { "term" : {"price" : 20}}, ②  
            { "term" : {"productID" : "XHDK-A-1293-#fJ3"}} ③  
          ],  
          "must_not" : {  
            "term" : {"price" : 30} ④  
          }  
        }  
      }  
    }  
}
```

Bool Filter

```
{  
  "bool" : {  
    "must" : [],  
    "should" : [],  
    "must_not" : [],  
    "filter": []  
  }  
}
```

Query DSL – nesting Boolean queries

```
SELECT document
FROM   products
WHERE  productID      = "KDKE-B-9947-#kL5"
       OR (    productID = "JODL-X-1937-#pV7"
           AND price      = 30 )
```

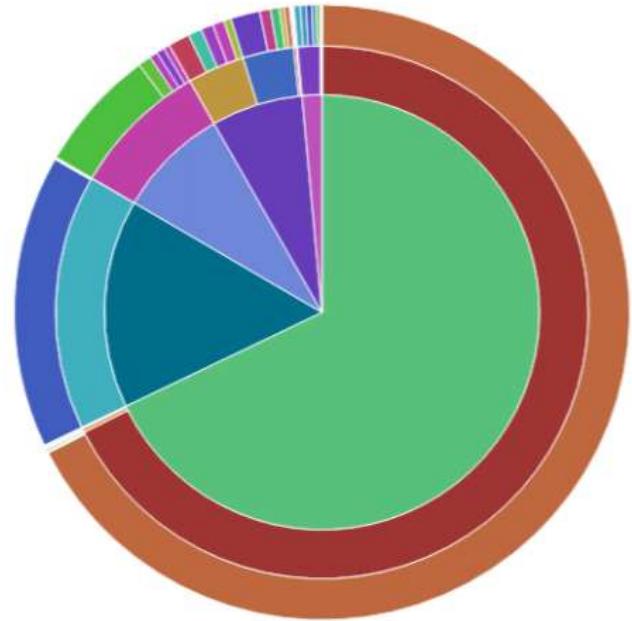


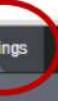
```
GET /my_store/products/_search
{
  "query" : {
    "constant_score" : {
      "filter" : {
        "bool" : {
          "should" : [
            { "term" : { "productID" : "KDKE-B-9947-#kL5" } }, ❶
            { "bool" : { "must" : [
              { "term" : { "productID" : "JODL-X-1937-#pV7" } }, ❷
              { "term" : { "price" : 30 } } ❸
            ] } } } } } } }
```

The diagram illustrates the mapping between an SQL query and its corresponding Elasticsearch Query DSL. On the left, an SQL SELECT statement is shown, which is then mapped by a blue arrow to the right, where the equivalent Elasticsearch Query DSL is displayed. The Elasticsearch query uses nested Boolean queries to achieve the same result as the SQL query, specifically handling the OR condition and the nested AND condition.

Kibana

- Search, view, and interact with data stored in Elasticsearch indices
- Execute queries on data & visualize results in charts, tables, and maps
- Add/remove widgets
- Share/Save/Load dashboards
- Open Source:Apache License 2.0



kibana Discover Visualize Dashboard **Settings** 

Indices Advanced Objects Status About

Index Patterns

- ★ mail-vpn-log-*
- dailystats-*
- dailystats2-*
- msgstats_*
- msgtrk_*
- nswaf2_event-*
- nswaf_appfw-*
- nswaf_event-*
- ssvpn-access-*
- ssvpn-loginfail-*
- ssvpn-webrequest-*

Configure an index pattern

In order to use Kibana you must configure at least one index pattern. Index patterns are used to identify the Elasticsearch index to run search and analytics against. They are also used to configure fields.

Index contains time-based events
 Use event times to create index names [DEPRECATED]

Index name or pattern

Patterns allow you to define dynamic index names using * as a wildcard. Example: logstash-*

logstash-*

Do not expand index pattern when searching (Not recommended)

By default, searches against any time-based index pattern that contains a wildcard will automatically be expanded to query only the indices that contain data within the currently selected time range.

Searching against the index pattern `logstash-*` will actually query elasticsearch for the specific matching indices (e.g. `logstash-2015.12.21`) that fall within the current time range.

Unable to fetch mapping. Do you have indices matching the pattern?

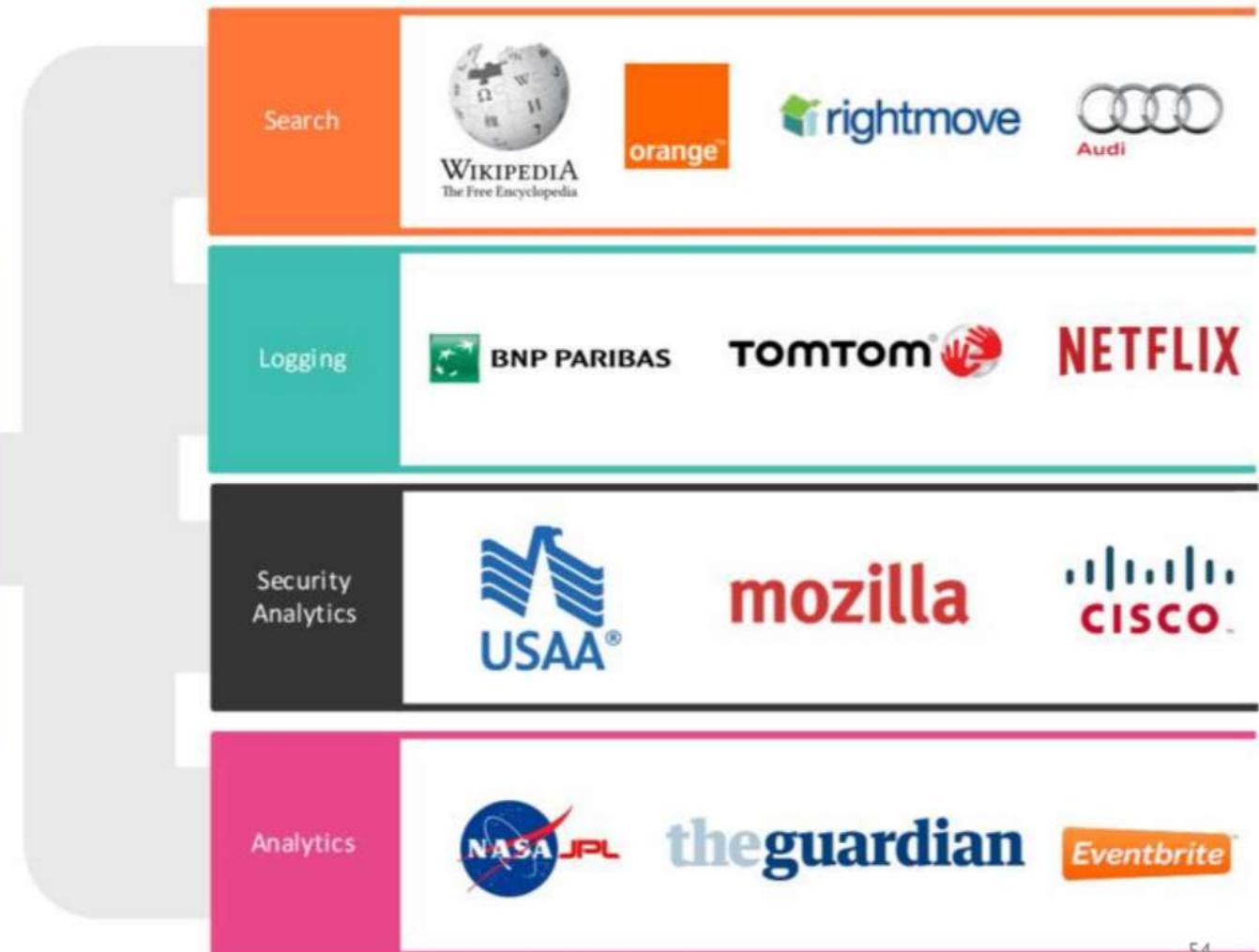
kibana Discover **Visualize** Dashboard Settings ■

Create a new visualization Step 1

 Area chart	Great for stacked timelines in which the total of all series is more important than comparing any two or more series. Less useful for assessing the relative change of unrelated data points as changes in a series lower down the stack will have a difficult to gauge effect on the series above it.
 Data table	The data table provides a detailed breakdown, in tabular format, of the results of a composed aggregation. Tip, a data table is available from many other charts by clicking grey bar at the bottom of the chart.
 Line chart	Often the best chart for high density time series. Great for comparing one series to another. Be careful with sparse sets as the connection between points can be misleading.
 Markdown widget	Useful for displaying explanations or instructions for dashboards.
 Metric	One big number for all of your one big number needs. Perfect for showing a count of hits, or the exact average a numeric field.
 Pie chart	Pie charts are ideal for displaying the parts of some whole. For example, sales percentages by department. Pro Tip: Pie charts are best used sparingly, and with no more than 7 slices per pie.
 Tile map	Your source for geographic maps. Requires an elasticsearch geo_point field. More specifically, a field that is mapped as type:geo_point with latitude and longitude coordinates.
 Timeseries	Create timeseries charts using the timeline expression language. Perfect for computing and combining timeseries set with functions suchs as derivatives and moving averages
 Vertical bar chart	The goto chart for oh-so-many needs. Great for time and non-time data. Stacked or grouped, exact numbers or percentages. If you are not sure which chart you need, you could do worse than to start here.

ELK Use cases

Use Cases

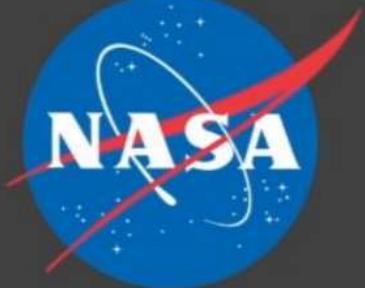


User cases



verizon

Use Case Logging, Analytics
Products Elasticsearch, Logstash



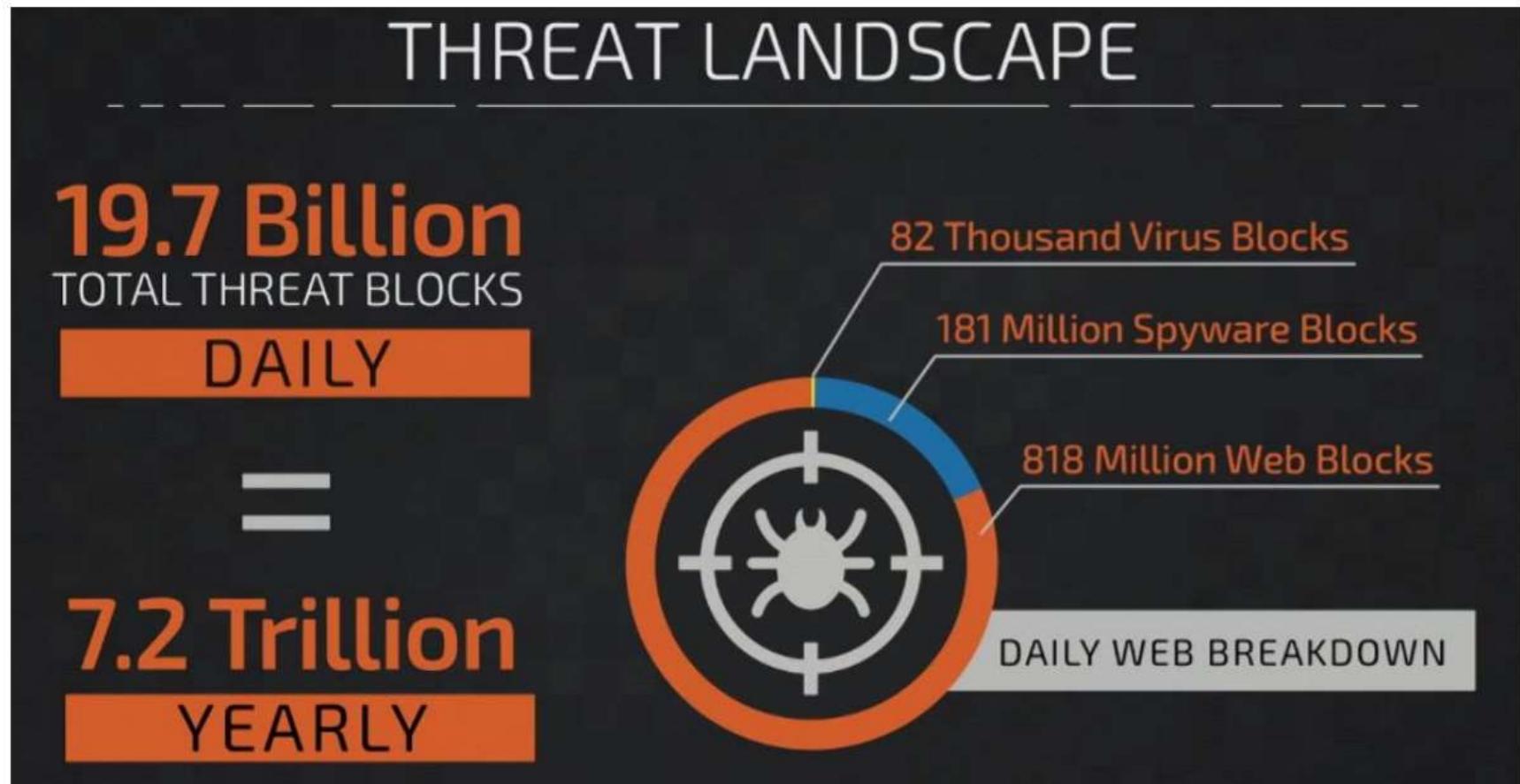
Use Case Search, Logging, Analytics
Products Elasticsearch, Logstash, Kibana

“ Elasticsearch, Logstash, and Kibana allow for real-time

“ With the ELK stack, we log more than 30K messages and 100K documents four times every day from the Mars Rover to optimize our space missions.”

Dan Isla, Data Scientist

Cisco Talos Security Intelligence and Research Group: Hunting for Hackers



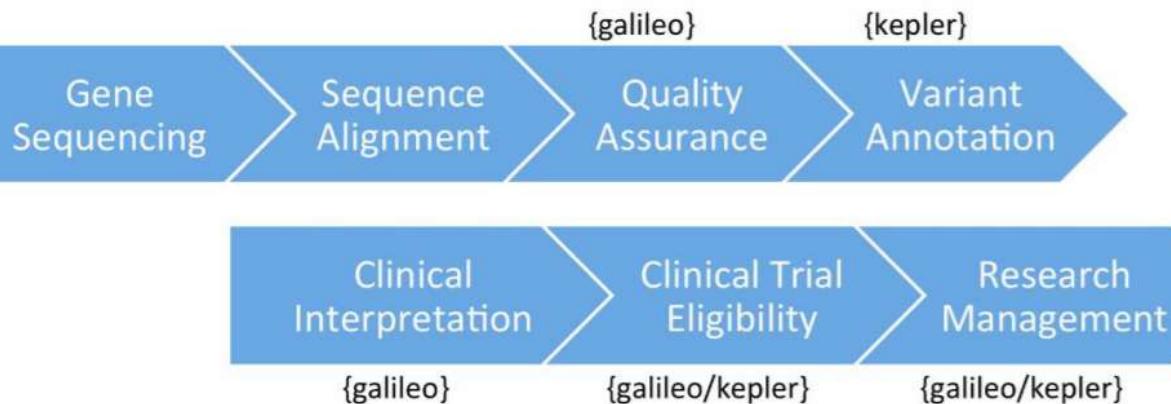
Cisco Talos use ELK to analyze...

- Sandbox data cluster
 - Dynamic malware analysis reports
 - Search for related pattern, malewares
 - ES stats
 - 10 nodes
 - 3 TB
 - 100k reports/day
 - ~8 months of data
- Honeypot cluster
 - Collect attackers' attempt
 - { Account, password } pair
 - Executed commands
 - url of download files
 - Suspicious command center for report back

```
_index": "logstash-telnet-sqs-2016.02.10",
"_type": "telnet-sqs",
"_source": {
  "Event.Type": "ConnectionLost",
  "@timestamp": "2016-02-10T23:51:10.000Z",
  "Event.Session": "1272f0ccd05111e5bb400242ac
110001",
  "Net.IP.Src": "117.158.195.59",
  "User.Name": "admin",
  "User.Pass": "1234",
  "Net.Port.Src": 23,
  "@version": "1",
  "type": "telnet-sqs".
```

Yale's {elastic}SEARCH – The Search for Cancer's Causes and Cures

Sequencing and Interpretation Pipeline



- With Next generation sequencing technology, the lab can process 8 million patients specimens yearly
- How to interpret this amount of data → what software can be used

<https://www.elastic.co/elasticon/2015/sf/videos/search-for-cancer-causes-and-cures>



Over 60 million variant annotations

Nearly 10 million documents related to cancer-associated mutations

NYC restaurants inspection @ELK

kibana

Discover Visualize Dashboard Settings

DOHMH: Restaurant

DOHMH: Description

NYC Restaurant Inspection

Dashboards

- Overview
- Restaurant Metrics

Data Source: <https://data.cityofnewyork.us/>

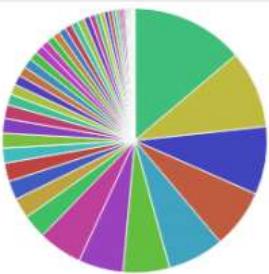
DOHMH: Inspections and Flags (vs time)



DOHMH: Violation Details

Time	Violation_Code	Violation_Description	Critical_Flag
March 19th 2016, 08:00:00.000	D6F	Wiping cloths soiled or not stored in sanitizing solution.	Critical
March 19th 2016, 08:00:00.000	D8C	Pesticide use not in accordance with label or applicable laws. Prohibited chemical used/stored. Open bait station used.	Not Critical
March 19th 2016, 08:00:00.000			Not Applicable
March 18th 2016, 08:00:00.000	D2H	Food not cooled by an approved method whereby the internal product temperature is reduced from 140° F to 70° F or less within 2 hours, and from 70° F to 41° F or less within 4 additional hours.	Critical
March 18th 2016, 08:00:00.000	D4H	Raw, cooked or prepared food is adulterated, contaminated, cross-contaminated, or not discarded in accordance with HACCP plan.	Critical
March 18th 2016, 08:00:00.000	D4L	Evidence of mice or live mice present in facility's food and/or non-food areas.	Critical
March 18th 2016, 08:00:00.000	D8A	Facility not vermin proof. Harborage or conditions conducive to attracting vermin to the premises and/or allowing vermin to exist.	Not Critical
March 18th 2016, 08:00:00.000	D10F	Non-food contact surface improperly constructed. Unacceptable material used. Non-food contact surface or equipment improperly maintained and/or not properly sealed, raised, spaced	Not Critical

DOHMH: Violation Code Distribution



60

References

<http://www.tp1rc.edu.tw/tpnet2016/training/10513.pdf>



BITS Pilani
Pilani Campus

BITS Pilani presentation

Yogesh B
Introduction to DevOps





CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 13 and 14

Agenda

On-demand Infrastructure Management

- Introduction to On-demand Infrastructure Management
- On-Demand Infrastructure Architect
- Security still a Major Concern?
- Service Providers
- Benefits of On-demand Infrastructure
- Auto Scaling
- Auto Scaling Approach
- Auto Scaling Providers



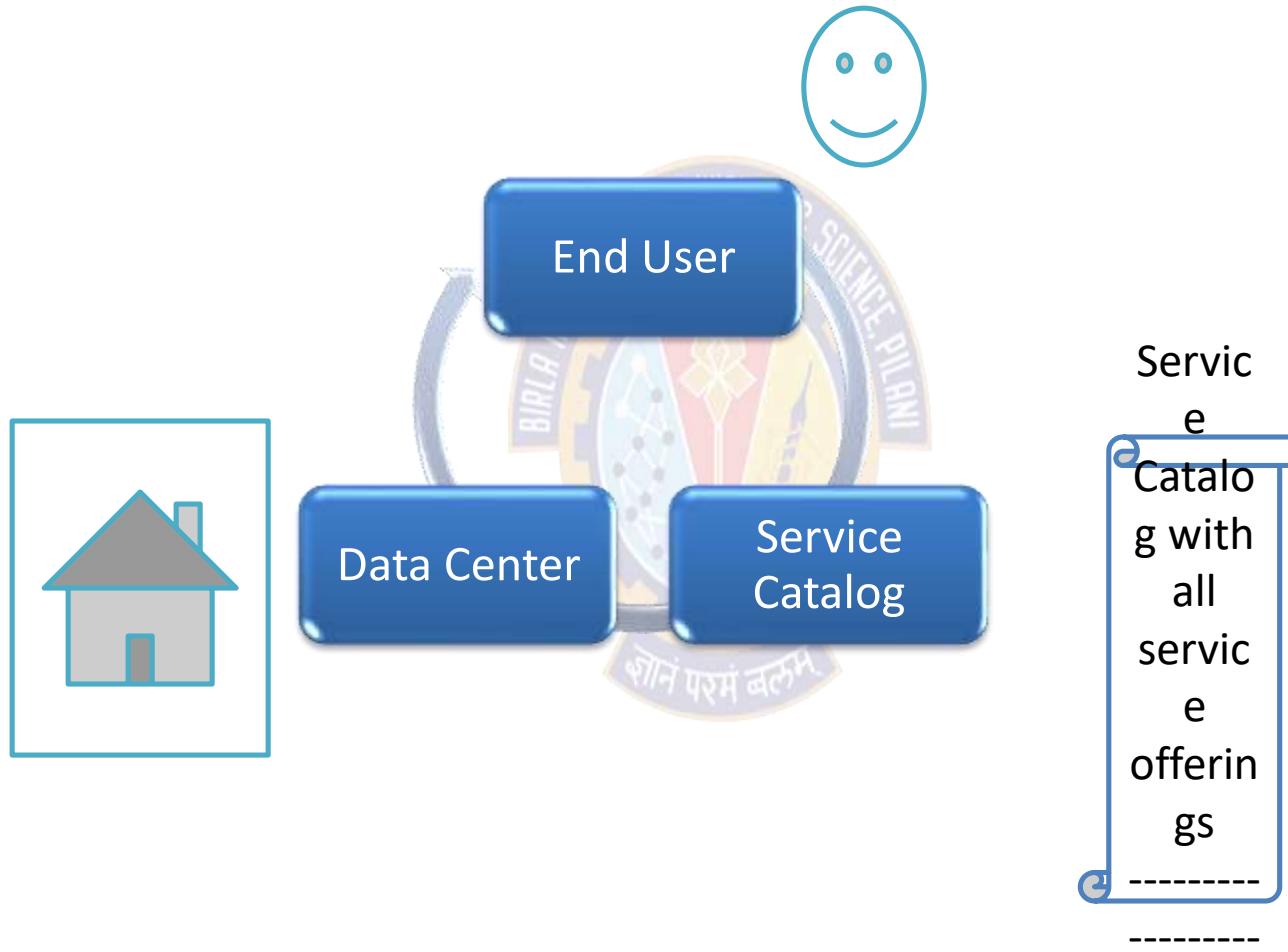
On-demand Infrastructure Management

Introduction to On-demand Infrastructure Management

- New IaaS technology is making it for on-demand infrastructure
- IaaS option provides users with on-demand access, while still giving them a system that is predictable and easy to control
- With IaaS, small businesses no longer have to build infrastructure for data storage internally
- And with IaaS they can pay for what they need as they grow
- Companies like Titanium aims to not only replace cloud storage, but also to provide processor power and memory, so that resource-intensive projects such as artificial intelligence or virtual reality can be accomplished by anyone in the world, at an affordable price

On-demand Infrastructure Management

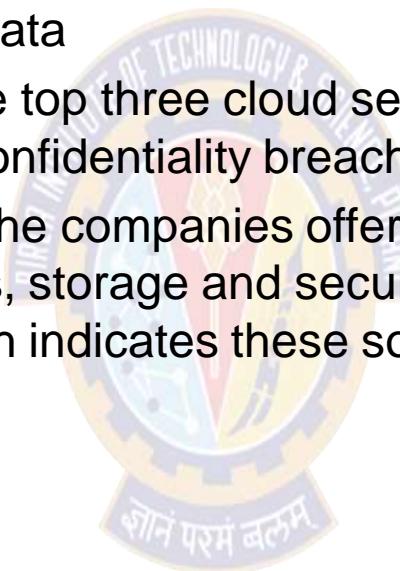
On-demand Infrastructure Architect



On-demand Infrastructure Management

Security Still A Major Concern?

- According to a recent release by Delta Risk, a cyber security and risk management firm, 90 percent of organizations are currently using cloud services to store data
- Delta Risk found that the top three cloud security concerns are data loss, data privacy and confidentiality breaches
- These are just a few of the companies offering IaaS solutions that address the data access, storage and security needs of small businesses and research indicates these solutions are hitting the mark



On-demand Infrastructure Management

Service Providers

- IaaS Providers:

- bmc
- DellEMC
- Dropbox
- Cisco
- Juniper
- Vmware

- PaaS Providers:

- Redhat
- Azure
- at & t
- Google
- Verizon

- SaaS Providers:

- Salesfroce
- Netapp
- Coupa
- liveops



On-demand Infrastructure Management

Benefits of On-demand Infrastructure

- Cost savings
- Scalability and flexibility
- Faster time to market
- Support for DR and high availability
- Focus on business growth



On-demand Infrastructure Management

Auto Scaling

- Auto scaling, also spelled auto scaling or auto-scaling, and sometimes also called automatic scaling
- It is a method used in cloud computing
- Auto Scaling Offers:
 - For companies running their own web server infrastructure
 - Auto Scaling typically means allowing some servers to go to sleep during times of low load
 - Saving on Power and Energy
 - For companies using infrastructure hosted in the cloud, Auto Scaling can mean lower bills, because most cloud providers charge based on total usage rather than maximum capacity
 - Auto Scaling can help by allowing the company to run less time-sensitive workloads on machines that get freed up by Auto Scaling during times of low traffic
 - Auto Scaling can offer greater uptime and more availability in cases where production workloads are variable and unpredictable

On-demand Infrastructure Management

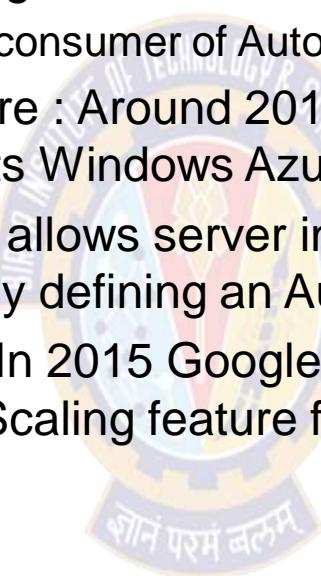
Auto Scaling Approaches

- Scheduled Auto Scaling Approach:
 - This is an approach to Auto Scaling where changes are made to the minimum size, maximum size, or desired capacity of the Auto Scaling group at specific times of day
 - Scheduled scaling is useful, for instance, if there is a known traffic load increase or decrease at specific times of the day
 - E.g. E-commerce sites: Flipkart Big Billion Day, Amazon's Great India Sale
- Predictive Auto Scaling:
 - The idea is to combine recent usage trends with historical usage data as well as other kinds of data to predict usage in the future, and Auto Scale based on these predictions
 - Identifying huge spikes in demand in the near future and getting capacity ready a little in advance
 - E.g. Online Social and Media Hosting

On-demand Infrastructure Management

Auto Scale Providers

- Amazon Web Services (AWS) : In 2009, Amazon launched its own Auto Scaling feature along with Elastic Load Balancing
 - Netflix is the well known consumer of Auto Scaling at AWS
- Microsoft's Windows Azure : Around 2013, Microsoft announced that Auto Scaling support to its Windows Azure cloud computing platform
- Oracle Cloud Platform: It allows server instances to automatically scale a cluster in or out by defining an Auto Scaling rule
- Google Cloud Platform : In 2015 Google Compute Engine announced a public beta of its Auto Scaling feature for use



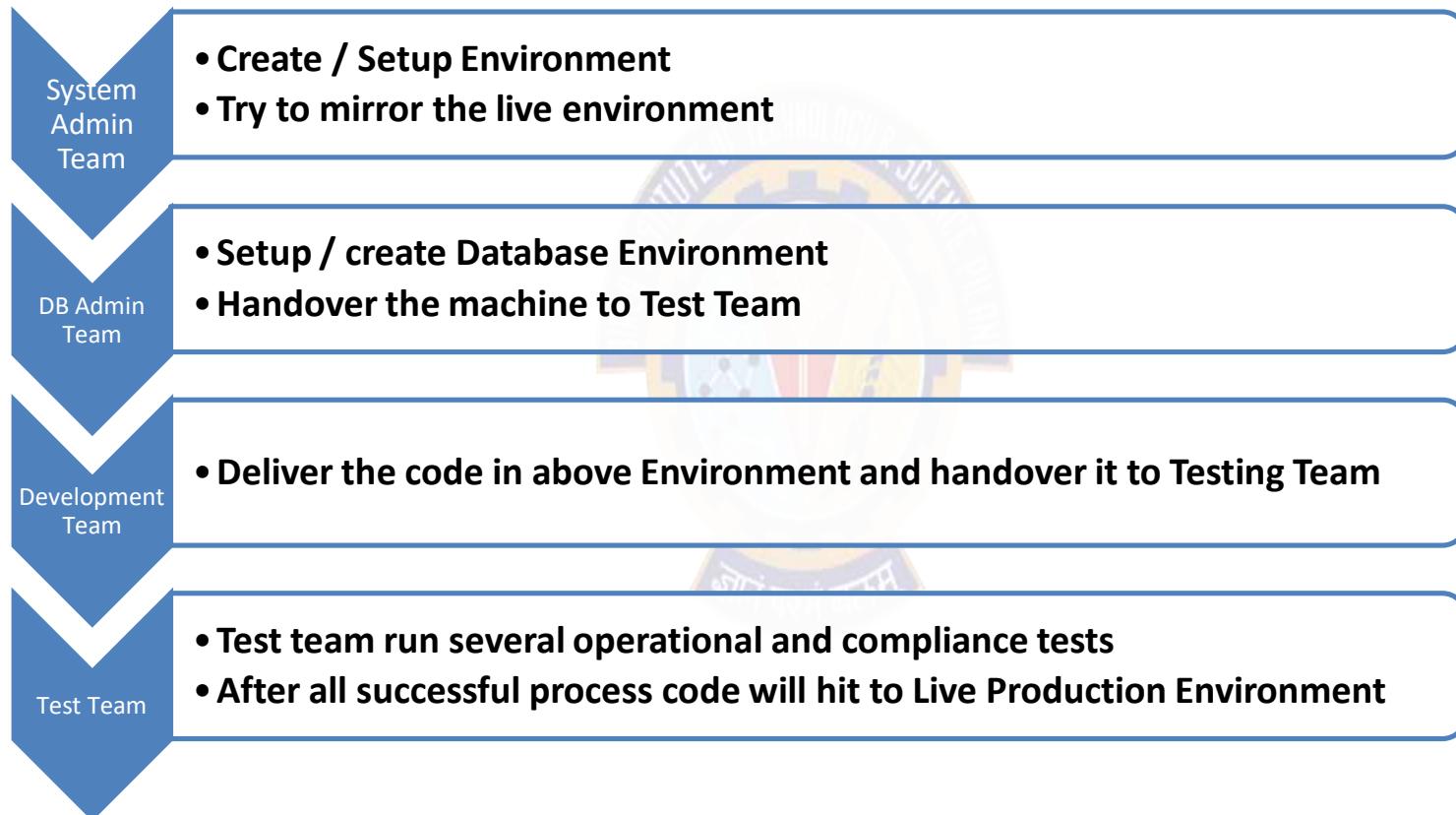
Infrastructure as Code [IaC]

Introduction to Infrastructure as Code

- Infrastructure as Code, or IaC
- It is a method of writing and deploying machine-readable definition files that generate service components
- IaC helps IT operations teams manage and provision IT infrastructure automatically through code without relying on manual processes
- IaC is often described as “programmable infrastructure”
- Infrastructure as Code is especially critical for cloud computing, Infrastructure as a Services (IaaS), and DevOps
- DevOps requires agile work processes which can be achieved through automated workflows for IT Infrastructure

Infrastructure as Code [IaC]

Before IaC and DevOps



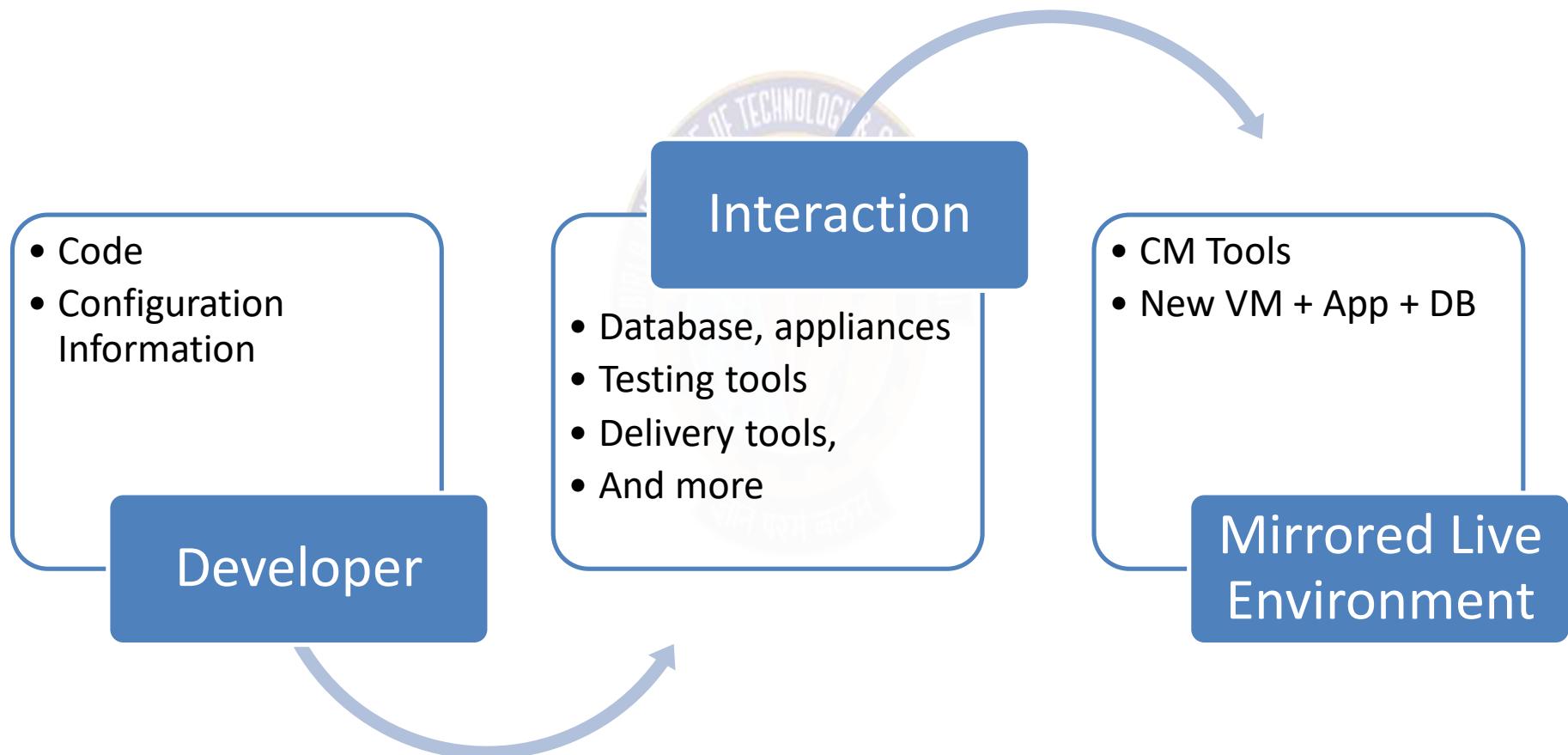
Infrastructure as Code [IaC]

IaC for DevOps

- In generic software development, a fundamental constraint is the need for the environment
- i.e. Testing or Staging environment = Live Environment
- Note: This is the only way of assuring that the new code will not collide with existing code definitions – generating errors or conflicts that may compromise the entire system
- Manual recreation of a live environment leaves doors open to a multitude of most likely minor but potentially quite important human errors, e.g. OS version, patch level, time zone etc
- Whereas clone created with the help of IaC can provide absolute guarantee of mirror of Live Production; now imagine a Software Delivery process involving DEV, UAT and Production environments using IaC

Infrastructure as Code [IaC]

Achievement using DevOps and IaC



Infrastructure as Code [IaC]

IaC Stand

- The general IaC concept, as well as available tools, has reached a very mature state with a lot of organizations having defined their roadmaps for adopting it
- There are now a number of tools available to adopt Infrastructure as Code
 - Puppet
 - Chef
 - Ansible
 - Terraform
 - Cloud formation etc



Infrastructure as Code [IaC]

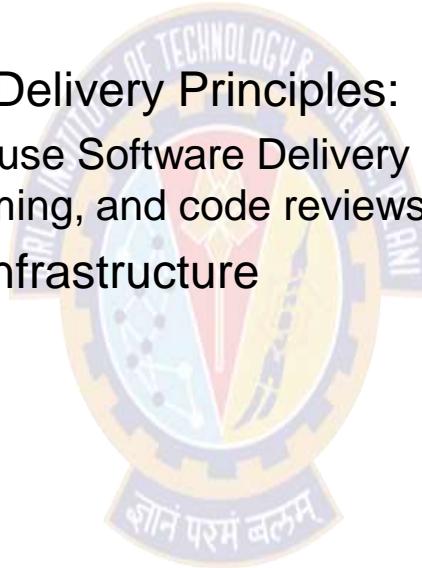
Benefits of IaC

- Reducing Shadow IT:
 - Shadow IT within organizations is due to the inability of IT departments to provide satisfactory and timely answers to operational areas concerning IT infrastructure and systems enhancements
 - This will result to potential unforeseen costs for the organization
 - IaC assures security and compliance with corporate IT standards and also it is helpful with budgeting and cost allocation
- Improving Customer Satisfaction:
 - Being able to deliver a quality service component within a short period of time contributes to customer satisfaction
- Operating Expenses [OPEX]:
 - Save in work time and avoiding human error will help to reduce financial risk
- Capital Expenditure [CAPEX]:
 - Reduce funds used by a company to acquire, upgrade, and maintain physical assets i.e. Single engineer with automation can perform task of multiple people
- Standardization:
 - Assurance of a consistent set of instructions and standardization

Infrastructure as Code [IaC]

Benefits of IaC Contd..

- Safer Change Management:
 - Standardization assurance enables safer changes to take place, with lower risk
- Application of Software Delivery Principles:
 - Ability to promote and use Software Delivery principles, such as version control, peer programming, and code reviews
- Scalable and Absolute Infrastructure



Infrastructure as Code [IaC]

Risk Involved in IaC

- Missing proper planning:
 - Once a company decides to move towards having an IaC; there is the mandatory need to define Infrastructure that will allow the implementation, configuration, and operation of IaC tools
- IaC requires new skills:
 - Most existing IaC tools require expertise to be handled, and reaching such levels requires significant time in learning and training
- Error replication:
 - Since the initial code is developed by humans, there is always the chance that it contains minor errors; several machines may have been automatically created where such errors exist
- Configuration Drift:
 - Once a machine is created via an IaC workflow; Manual or external updates (even if just security patching) may result in configuration drift
- Accidental Destruction:
 - Some IaC tools that maintain state have the ability to automatically destroy resources
 - IaC in an automation pipeline can sometimes have undesired outcomes

Why Is Software Difficult to Build?

- Lack of control
- Lack of monitoring
- Lack of traceability
- Uncontrolled changes

Conflicts in Team Software Development



- Simultaneous updates – how to prevent one person from undoing the changes of another
- Shared and common code – how to notify everyone who needs to know about a change
- Versions – how to make changes to all affected versions when needed

Configuration management – Why

- New versions of software systems are created as they change
 - For different machines/OS
 - Offering different functionality
 - Tailored for particular user requirements
- Configuration management is concerned with managing evolving software systems
 - System change is a team activity
 - CM aims to control the costs and effort involved in making changes to a system



Configuration management – Why

- Involves the development and application of procedures and standards to manage an evolving software product
- May be seen as part of a more general quality management process
- When released to CM, software systems are sometimes called *baselines* as they are a starting point for further development

Software Configuration Management is a Project Function



- SCM is a Project Function (as defined in the SPMP) with the goal to make technical and managerial activities more effective.
- Software Configuration Management can be administered in several ways:
 - A single software configuration management team for the whole organization
 - A separate configuration management team for each project
 - Software Configuration Management distributed among the project members
 - Mixture of all of the above

Configuration Management Activities

- Software Configuration Management Activities:
 - Configuration item identification
 - Promotion management
 - Release management
 - Branch management
 - Variant management
 - Change management
- No fixed rules:
 - Activities are usually performed in different ways (formally, informally) depending on the project type and life-cycle phase (research, development, maintenance)

Configuration Management Activities (continued)



- Configuration item identification
 - modeling of the system as a set of evolving components
- Promotion management
 - is the creation of versions for other developers
- Release management
 - is the creation of versions for the clients and users
- Change management
 - is the handling, approval and tracking of change requests
- Branch management
 - is the management of concurrent development
- Variant management
 - is the management of versions intended to coexist

Configuration Management Roles

- Configuration Manager
 - Responsible for identifying configuration items. The configuration manager can also be responsible for defining the procedures for creating promotions and releases
- Change control board member
 - Responsible for approving or rejecting change requests
- Developer
 - Creates promotions triggered by change requests or the normal activities of development. The developer checks in changes and resolves conflicts
- Auditor
 - Responsible for the selection and evaluation of promotions for release and for ensuring the consistency and completeness of this release

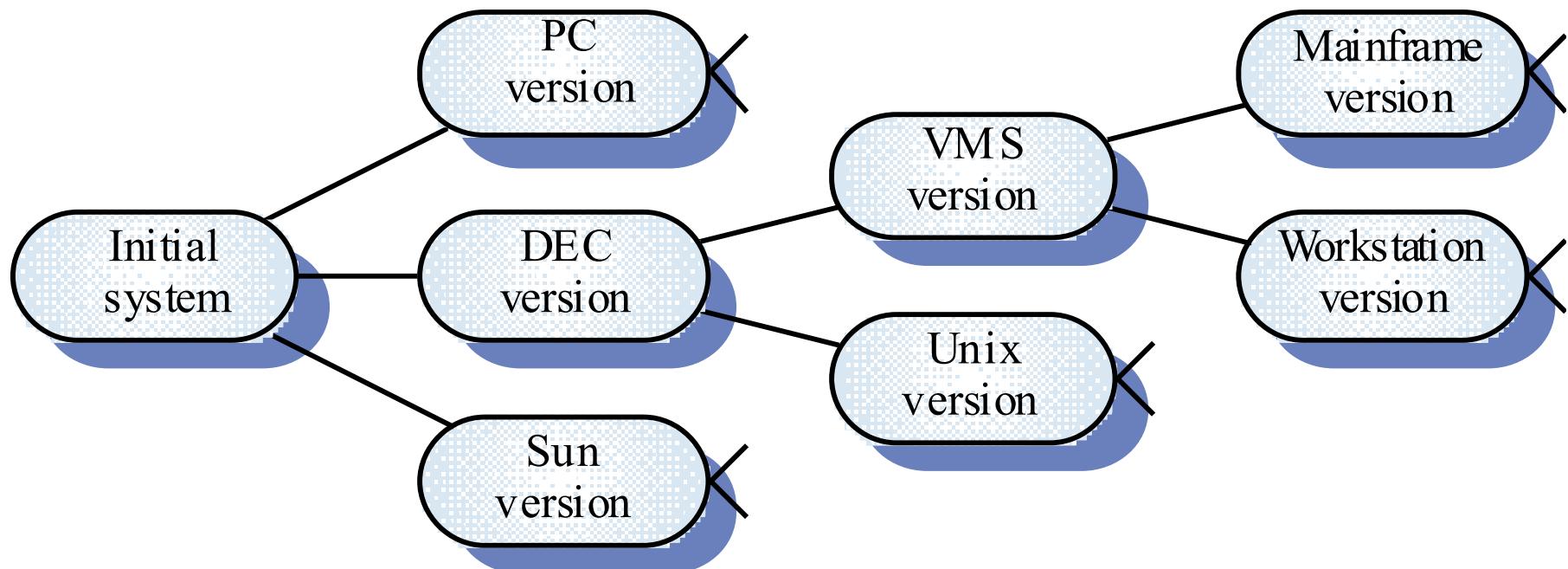
Possible Selection of Configuration Items



- Problem Statement
- Software Project Management Plan (SPMP)
- 4 Requirements Analysis Document (RAD)
- 4 System Design Document (SDD)
- Project Agreement
- 4 Object Design Document (ODD)
- Dynamic Model
- Object model
- Functional Model
- 4 Unit tests
- Integration test strategy
- 4 Source code
 - API Specification
- 4 Input data and data bases
 - Test plan
- 4 Test data
- 4 Support software (part of the product)
 - Support software (not part of the product)
 - User manual
 - Administrator manual

Once the Configuration Items are selected, they are usually organized in a tree

System families



Configuration management planning

- Starts during the early phases of the project
- All products of the software process may have to be managed
 - Specifications
 - Designs
 - Programs
 - Test data
 - User manuals
- Thousands of separate documents may be generated for a large software system



The CM plan

- Defines the types of documents to be managed and a document naming scheme
- Defines who takes responsibility for the CM procedures and creation of “baselines”
- Defines policies for change control and version management
- Defines the CM records which must be maintained
- Describes the tools which should be used to assist the CM process and any limitations on their use

Configuration item identification

- Large projects typically produce thousands of documents which must be uniquely identified
- Some of these documents must be maintained for the lifetime of the software
- Document naming scheme should be defined so that related documents have related names
- A hierarchical scheme with multi-level names is probably the most flexible approach

The configuration database

- All CM information should be maintained in a configuration database
- This should allow queries about configurations to be answered
 - Who has a particular system version?
 - What platform is required for a particular version?
 - What versions are affected by a change to component X?
 - How many reported faults in version T?
- The CM database should preferably be linked to the software being managed

Change management

- Software systems are subject to continual change requests
 - From users
 - From developers
 - From market forces
- Change management is concerned with keeping strack of these changes and ensuring that they are implemented in the most cost-effective way

The change management process

Request change by completing a change request form

Analyze change request

if change is valid **then**

 Assess how change might be implemented

 Assess change cost

 Submit request to change control board

if change is accepted **then**

repeat

 make changes to software

 submit changed software for quality approval

until software quality is adequate

 create new system version

else

 reject change request

else

 reject change request

Change tracking tools

- A major problem in change management is tracking change status
- Change tracking tools track the status of each change request
 - Automatically ensure that change requests are sent to the right people at the right time
- Integrated with E-mail systems allowing electronic change request distribution

Derivation history

- Record of changes applied to a document or code component
- Should record, in outline, the change made, the rationale for the change, who made the change and when it was implemented
- May be included as a comment in code
 - If a standard prologue style is used for the derivation history, tools can process this automatically

Component header information

```
// PROTEUS project (ESPRIT 6087)
//
// PCL-TOOLS/EDIT/FORMS/DISPLAY/AST-INTERFACE
//
// Object: PCL-Tool-Desc
// Author: G. Dean
// Creation date: 10th November 1998
//
// © Lancaster University 1998
//
// Modification history
// Version          Modifier Date      Change          Reason
// 1.0      J. Jones   1/12/1998  Add header  Submitted to CM
// 1.1      G. Dean    9/4/1999  New field  Change req. R07/99
```

Version and release management

- Invent identification scheme for system versions
- Plan when new system version is to be produced
- Ensure that version management procedures and tools are properly applied
- Plan and distribute new system releases

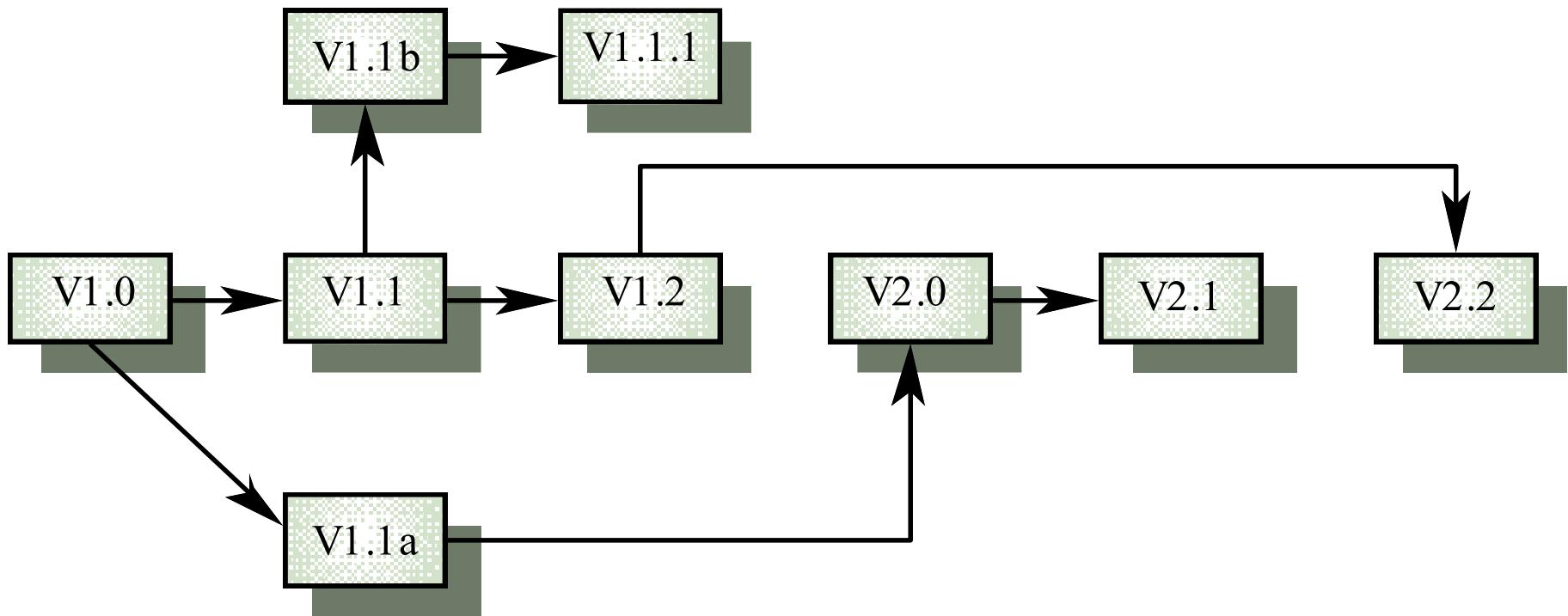
Versions/variants/releases

- *Version* An instance of a system which is functionally distinct in some way from other system instances
- *Variant* An instance of a system which is functionally identical but non-functionally distinct from other instances of a system
- *Release* An instance of a system which is distributed to users outside of the development team

Version identification

- Procedures for version identification should define an unambiguous way of identifying component versions
- Three basic techniques for component identification
 - Version numbering
 - Attribute-based identification
 - Change-oriented identification

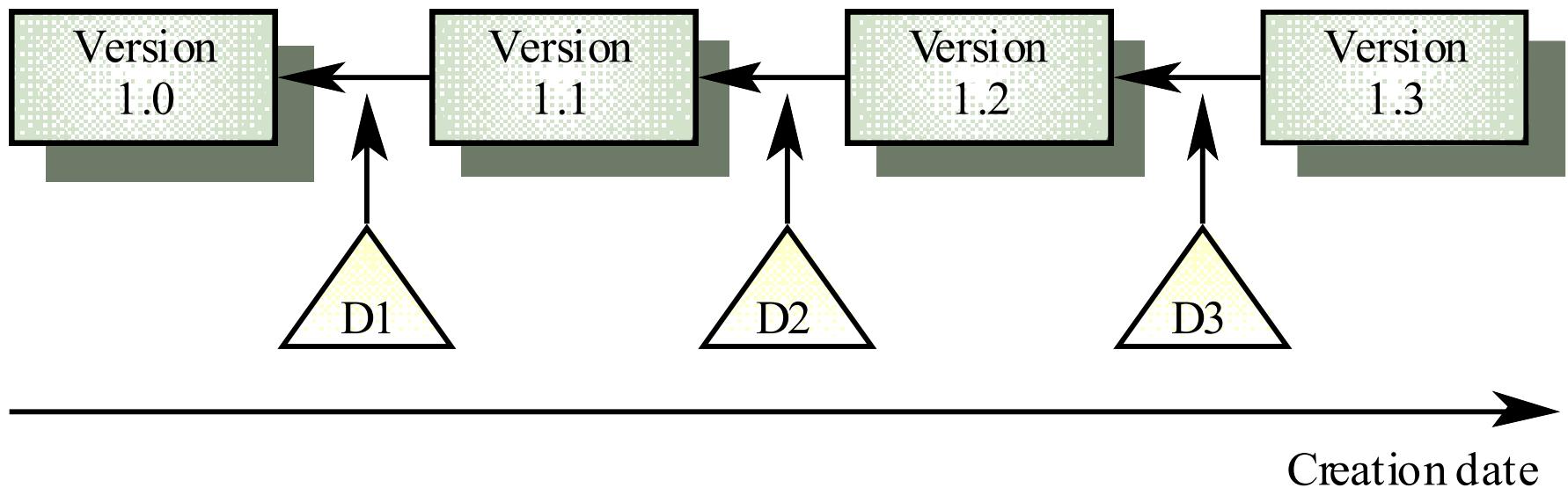
Version derivation structure



Version management tools

- Version and release identification
 - Systems assign identifiers automatically when a new version is submitted to the system
- Storage management.
 - System stores the differences between versions rather than all the version code
- Change history recording
 - Record reasons for version creation
- Independent development
 - Only one version at a time may be checked out for change
 - Parallel working on different versions

Delta-based versioning



Release management

- Releases incorporate changes forced on the system by errors discovered by users and by hardware changes
- They also incorporate new system functionality
- Release planning is concerned with when to issue a system version as a release

System releases

- Not just a set of executable programs
- May also include
 - Configuration files defining how the release is configured for a particular installation
 - Data files needed for system operation
 - An installation program or shell script to install the system on target hardware
 - Electronic and paper documentation
 - Packaging and associated publicity
- Systems are now normally released on CD-ROM or as downloadable installation files from the Web

Release problems

- Customer may not want a new release of the system
 - They may be happy with their current system as the new version may provide unwanted functionality
- Release management must not assume that all previous releases have been accepted
 - All files required for a release should be re-created when a new release is installed

Release creation

- Release creation involves collecting all files and documentation required to create a system release
- Configuration descriptions have to be written for different hardware and installation scripts have to be written
- The specific release must be documented to record exactly what files were used to create it
 - This allows it to be re-created if necessary

System building

- The process of compiling and linking software components into an executable system
- Different systems are built from different combinations of components
- Invariably supported by automated tools that are driven by 'build scripts'

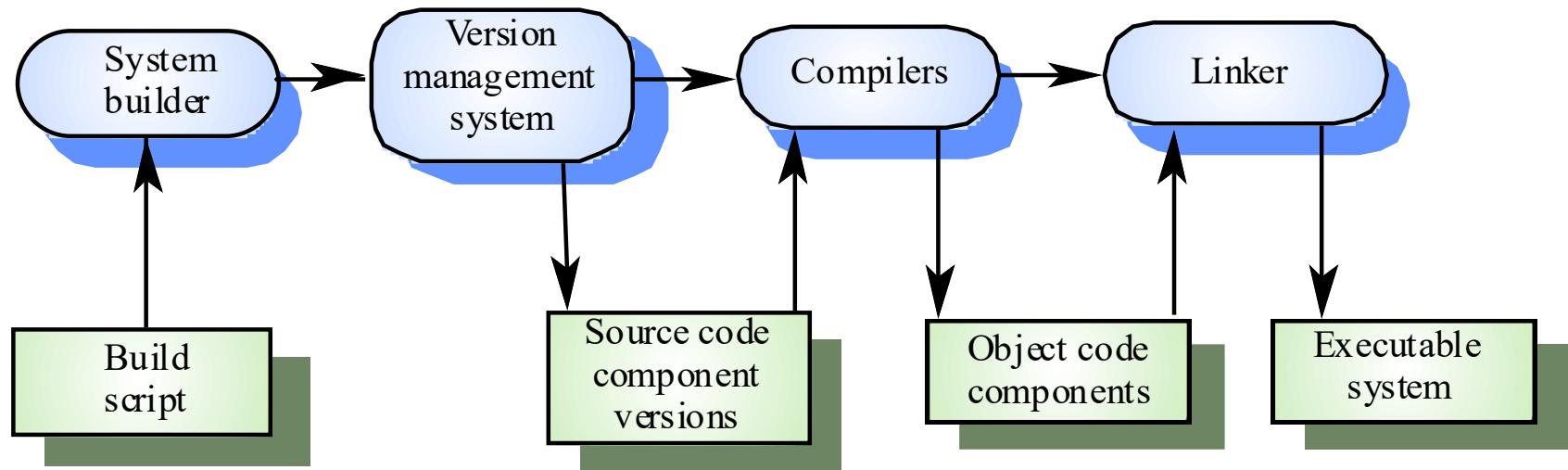
System building problems

- Do the build instructions include all required components?
 - When there are many hundreds of components making up a system, it is easy to miss one out
 - This should normally be detected by the linker
- Is the appropriate component version specified?
 - A more significant problem. A system built with the wrong version may work initially but fail after delivery
- Are all data files available?
 - The build should not rely on 'standard' data files
 - Standards vary from place to place

System building problems

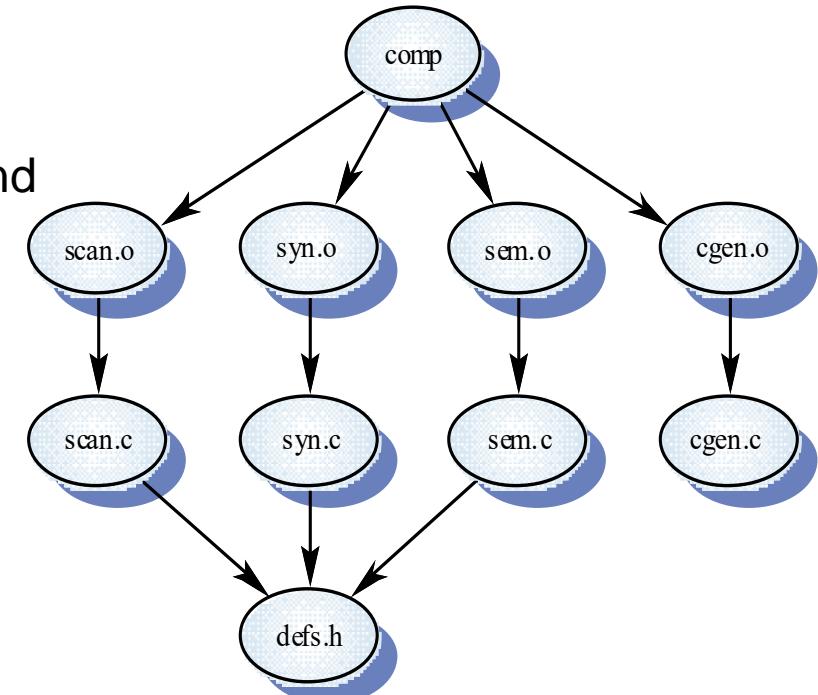
- Are data file references within components correct?
 - Embedding absolute names in code almost always causes problems as naming conventions differ from place to place
- Is the system being built for the right platform
 - Sometimes must build for a specific OS version or HW configuration
- Is the right version of the compiler and other software tools specified?
 - Different compiler versions may actually generate different code and the compiled component will exhibit different behaviour

System building



System building tools

- Building a large system is computationally expensive and may take several hours
- Hundreds of files may be involved
- System building tools may provide
 - A dependency specification language and interpreter
 - Tool selection and instantiation support
 - Distributed compilation
 - Derived object management



Key points

- Configuration management is the management of system change to software products
- A formal document naming scheme should be established and documents should be managed in a database
- The configuration database should record information about changes and change requests
- A consistent scheme of version identification should be established using version numbers, attributes or change sets
- System releases include executable code, data, configuration files and documentation
- System building involves assembling components into a system

Configuration Management

Configuration management refers to the process by which all artifacts relevant to your project, and the relationships between them, are stored, retrieved, uniquely identified, and modified

Will determine how you manage all changes that happen within the project

Governs how team members collaborate

Good Config Mgmt Strategy

- Can I exactly reproduce any of my environments, including the version of the operating system, its patch level, the network configuration, the software stack, the applications deployed into it, and their configuration?
- Can I easily make an incremental change to any of these individual items and deploy the change to any, and all, of my environments?
- Can I easily see each change that occurred to a particular environment and trace it back to see exactly what the change was, who made it, and when they made it?
- Can I satisfy all of the compliance regulations that I am subject to?
- Is it easy for every member of the team to get the information they need, and to make the changes they need to make? Or does the strategy get in the way of efficient delivery, leading to increased cycle time and reduced feedback?

Using Version Control

- Version control systems, also known as source control, source code management systems, or revision control systems, are a mechanism for keeping multiple versions of your files, so that when you modify a file you can still access the previous revisions.

Aim of VCS/ CVS:

- It retains, and provides access to, every version of every file that has ever been stored in it. Such systems also provide a way for metadata—that is, information that describes the data stored—to be attached to single files or collections of files.
- It allows teams that may be distributed across space and time to collaborate

Questions to answer

- What constitutes a particular version of your software? How can you reproduce a particular state of the software's binaries and configuration that existed in the production environment?
- What was done when, by whom, and for what reason? Not only is this useful to know when things go wrong, but it also tells the story of your application

Version Control Use

Developers should use it for source code, of course, but also for tests, database scripts, build and deployment scripts, documentation, libraries and configuration files for your application, your compiler and collection of tools, and so on—so that a new member of your team can start working from scratch

Who stores what

- It is even helpful to keep the configuration files for the development team's development environments in version control since it makes it easy for everyone on the team to use the same settings
- Analysts should store requirements documents
- Testers should keep their test scripts and procedures in version control
- Project managers should save their release plans, progress charts, and risk logs here

In short, every member of the team should store any document or file related to the project in version control

What ALL?

- Apart from storing source code and configuration information
- Many projects also store binary images of their application servers, compilers, virtual machines, and other parts of their toolchain in version control.
- This speeds up the creation of new environments and, ensures that base configurations are completely defined
- Simply checking everything you need out of the version control repository assures a stable platform for development, test, or even production environments.
- You can then store whole environments, including base operating systems with configuration baselines applied, as virtual images for an even higher level of assurance and deployment simplicity

What NOT?

- Binary output of your application's compilation shouldn't be put.
- First, they are big, and unlike compilers, they proliferate rapidly (we create new binaries for every check-in that compiles and passes the automated commit tests).
- Second if you have an automated build system, you should be able to recreate them easily from source by rerunning the build script
- The combination of your build system and source code is all that should be required to re-create an instance of your application in an emergency
- Storing the binary output of the build breaks the idea of being able to identify a single version of your repository for each application version because there may be two commits for the same version, one for source code and another for the binaries

Managing Software Config

- Configuration is one of the three key parts that comprise an application, along with its binaries and its data.
- Configuration information can be used to change the behavior of software at build time, deploy time, and run time.
- Delivery teams need to consider carefully what configuration options should be available, how to manage them throughout the application's life, and how to ensure that configuration is managed consistently across components, applications, and technologies

Types of Configuration

- Your build scripts can pull configuration in and incorporate it into your binaries at **build time**
- Your packaging software can inject configuration at **packaging time**, such as when creating assemblies, ears, or gems
- Your deployment scripts or installers can fetch the necessary information or ask the user for it and pass it to your application at **deployment time** as part of the installation process
- Your application itself can fetch configuration at **startup time or run time**

Principles of Managing Application Configuration

- Consider where in your application's lifecycle it makes sense to inject a particular piece of configuration—at the point of assembly where you are packaging your release candidate, at deployment or installation time, at startup time, or at run time. Speak to the operations and support team to work out what their needs are.
- Keep the available configuration options for your application in the same repository as its source code, but keep the values somewhere else.
- Configuration settings have a lifecycle completely different from that of code, while passwords and other sensitive information should not be checked in to version control at all.
- Configuration should always be performed by automated processes using values taken from your configuration repository, so that you can always identify the configuration of every application in every environment

Principles of Managing Application Configuration

- Your configuration system should be able to provide different values to your application (including its packaging, installation, and deployment scripts) based on the application, its version, and the environment it is being deployed into. It should be easy for anyone to see what configuration options are available for a particular version of an application across all environments it will be deployed into.
- Use clear naming conventions for your configuration options. Avoid obscure or cryptic names. Try to imagine someone reading the configuration file without a manual—it should be possible to understand what the configuration properties are.
- Ensure that your configuration information is modular and encapsulated so that changes in one place don't have knock-on effects for other, apparently unrelated, pieces of configuration.

Principles of Managing Application Configuration



- Use the DRY (don't repeat yourself) principle. Define the elements of your configuration so that each concept has only one representation in the set of configuration information
- Be minimalist: Keep the configuration information as simple and as focused as possible. Avoid creating configuration options except where there is a requirement or where it makes sense to do so
- Avoid overengineering the configuration system. Keep it as simple as you can
- Ensure that you have tests for your configuration that are run at deployment or installation time. Check that the services your application depends upon are available, and use smoke tests to assert that any functionality depending on your configuration settings works as it should

Reproducing environments

- It removes the problem of having random pieces of infrastructure around whose configuration is only understood by somebody who has left the organization and cannot be reached. When such things stop working, you can usually assume a significant downtime. This is a large and unnecessary risk.
- Fixing one of your environments can take many hours. It is always better to be able to rebuild it in a predictable amount of time so as to get back to a known good state.
- It is essential to be able to create copies of production environments for testing purposes. In terms of software configuration, testing environments

should be exact replicas of the production ones, so configuration problems

Environment Config

- The various operating systems in your environment, including their versions, patch levels, and configuration settings
- The additional software packages that need to be installed on each environment to support your application, including their versions and configuration
- The networking topology required for your application to work
- Any external services that your application depends upon, including their versions and configuration
- Any data or other state that is present in them (for example, production databases)

Steps to deploying software

- Creating and managing the infrastructure in which your application will run (hardware, networking, middleware, and external services)
- Installing the correct version of your application into it
- Configuring the application, including any data or state that it requires

Attributes Describing Env

- The hardware configuration of the servers that form the environment (such as the number and type of CPUs, amount of memory, spindles, NICs, and so on) and the networking infrastructure that connects them
- The configuration of the operating system and middleware (such as messaging systems, application and web servers, database servers) required to support the applications that will run within it

Holistic Approach to managing Infra



- The desired state of your infrastructure should be specified through version-controlled configuration.
- Infrastructure should be autonomic—that is, it should correct itself to the desired state automatically.
- You should always know the actual state of your infrastructure through instrumentation and monitoring.

Reducing risk to managing prod

like env



- The operating system and its configuration, for both testing and production environments
- The middleware software stack and its configuration, including application servers, messaging systems, and databases
- Infrastructural software, such as version control repositories, directory services, and monitoring systems
- External integration points, such as external systems and services
- Network infrastructure, including routers, firewalls, switches, DNS, DHCP, and so on
- The relationship between the application development team and the infrastructure management team

Needs of Ops team

- Almost all medium and large companies separate the activities of development and infrastructure management (or operations as it is often known) into different groups or silos
- Development teams are incentivized to deliver software as rapidly as possible, whereas operations teams aim for stability
- Continuous delivery ensures that there is as little change as possible between releases
- Releasing can and should be an activity that can be performed in a few minutes

- Operations teams measure their effectiveness in terms of the key quality-of service metrics such as mean time between failures (MTBF) and mean time to repair failures (MTTR)
- Operations teams have service-level agreements (SLAs) they have to meet
- Operations managers want to ensure that any change to any environment they control is documented and audited, so that, if things go wrong, they can find the relevant changes that caused the problem
- One of the most important processes an organization will have in place is a change management process, which is used to manage every change made to any controlled environments—and often operations will control both production and production-like testing

- A request for change needs to include details on the risk and impact of the change, and how it will be remediated if it fails
- The request should be submitted before work starts on the new version to be deployed, not a couple of hours before the business expects it to go live
- Member of the software development team have a responsibility to familiarize themselves with any such systems and processes that the operations team has in place, and comply with them
- Identifying the procedures that need to be followed to release your software should be part of your development team's release plan
- The deployment system forms an integral part of the application—it should be tested and refactored with the same care and attention as the rest of the application, and kept in version control

Types of Servers and their Config

App Server	
Apps / services / components	Application configuration
Middleware	Middleware configuration
Operating system	Operating system configuration
Hardware	

Database Server	
Database	Database configuration
Operating system	Operating system configuration
Hardware	

Infrastructure Server	
DNS, SMTP, DHCP, LDAP, etc.	Infrastructure configuration
Operating system	Operating system configuration
Hardware	

Maintaining Infra Version Control



- Operating system install definitions (such as those used by Debian Preseed, RedHat Kickstart, and Solaris Jumpstart)
- Configuration for data center automation tools like Puppet or CfEngine
- General infrastructure configuration, such as DNS zone files, DHCP and SMTP server configuration files, firewall configuration files, and so forth
- Any scripts you use for managing your infrastructure

Job of Deployment Pipeline

- First, it should verify that all applications will work with any infrastructural changes before they get pushed out to production environments, ensuring that every affected application's functional and nonfunctional tests pass against the new version of the infrastructure
- Second, it should be used to push changes out to operations-managed testing and production environments
- Finally, the pipeline should perform deployment tests to ensure that the new infrastructure configuration has been deployed successfully

Controlling access to infrastructure

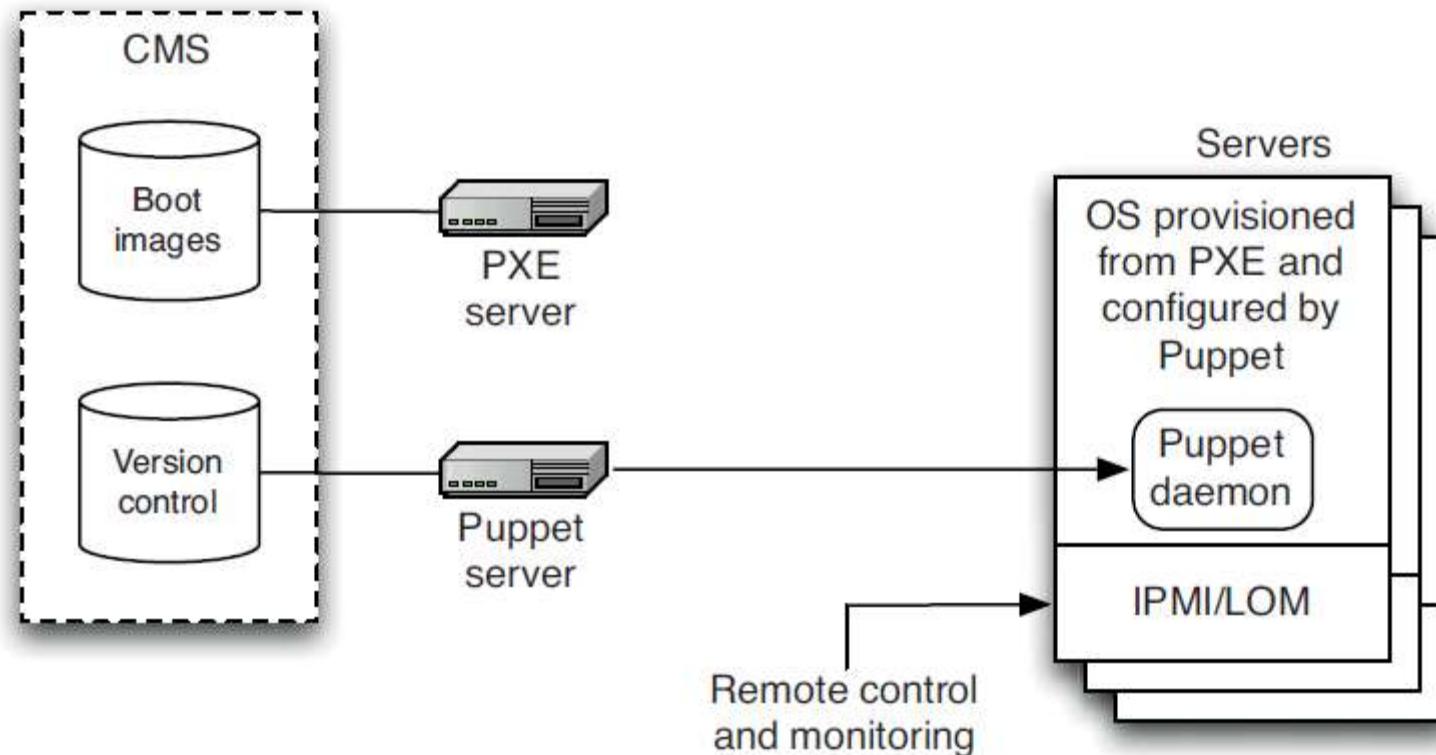
- Controlling access to prevent anyone from making a change without approval
- Defining an automated process for making changes to your infrastructure
- Monitoring your infrastructure to detect any issues as soon as they occur
- It is essential to lock down the production environments to prevent unauthorized access not only from people outside your organization, but also from people within it—even operations staff
- Requests to make changes to your production and testing environments should go through a change management process

Making changes to Infrastructure



- Every change, whether it's updating firewall rules or deploying a new version of your flagship service, should go through the same change management process
- This process should be managed using a single ticketing system that everybody can log into and which generates useful metrics such as average cycle time per change
- The exact change that is made should be logged so it can be easily audited later
- It should be possible to see a history of changes made to every environment, including deployments
- The change you want to make should first have been tested on one of your production-like testing environments, and automated tests should be run to ensure that it doesn't break any of the applications that use the environment
- The change should be made to version control and then applied through your

Automatic provisioning & config of servers



Provisioning servers

Ways to create operating system baselines:

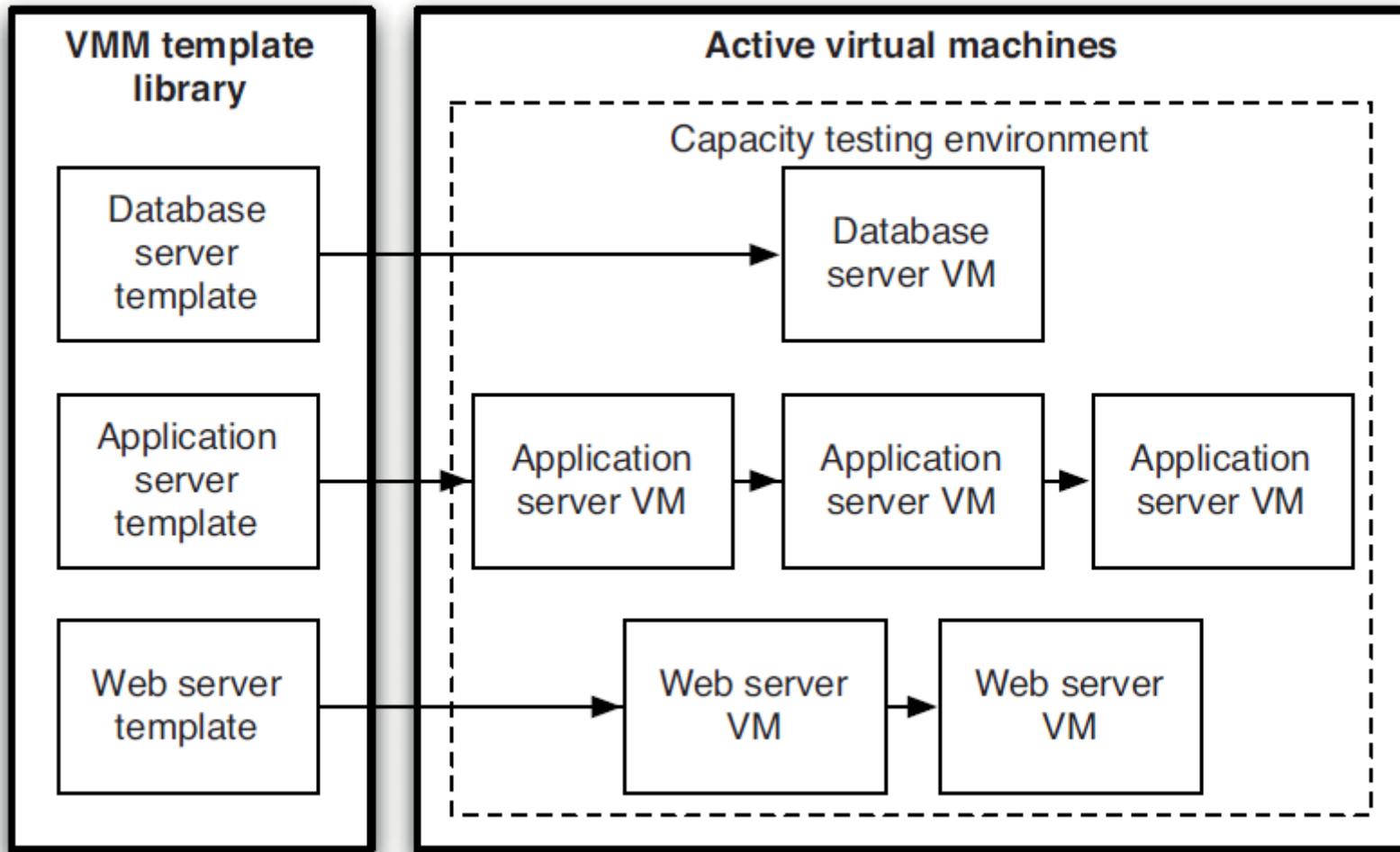
- A fully manual process
- Automated remote installation
- Virtualization



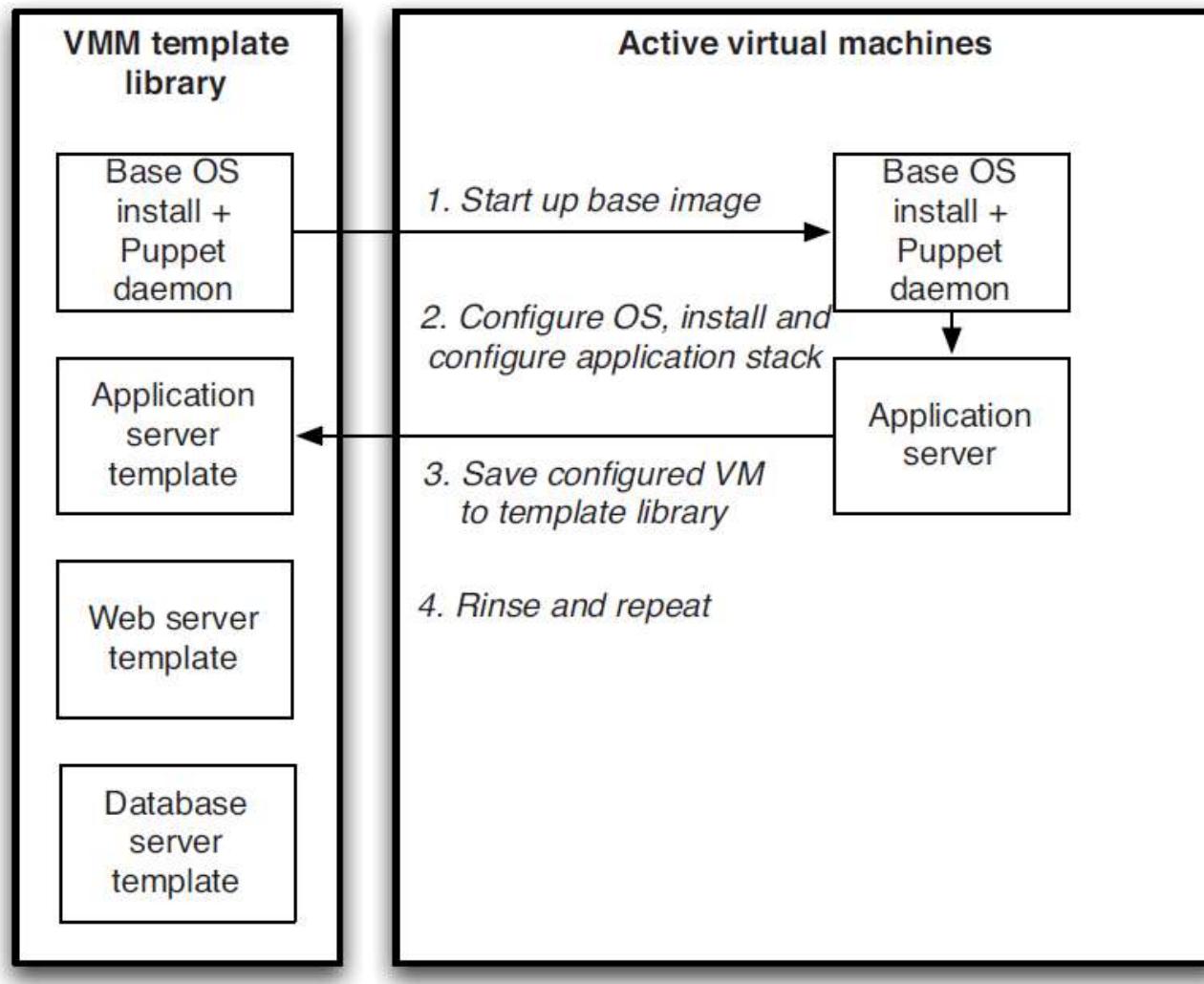
Test driven changes to environments

- In your monitoring system, write a service that monitors the problem you are trying to solve, and make sure the service shows red on your dashboard.
- Implement the configuration change, and have Puppet roll it out to your test system
- Once the service shows green on your dashboard, have Puppet roll out the change to production.

Creating Virtual Env from Templates



Creating VM templates



Virtual Env & Deployment pipeline

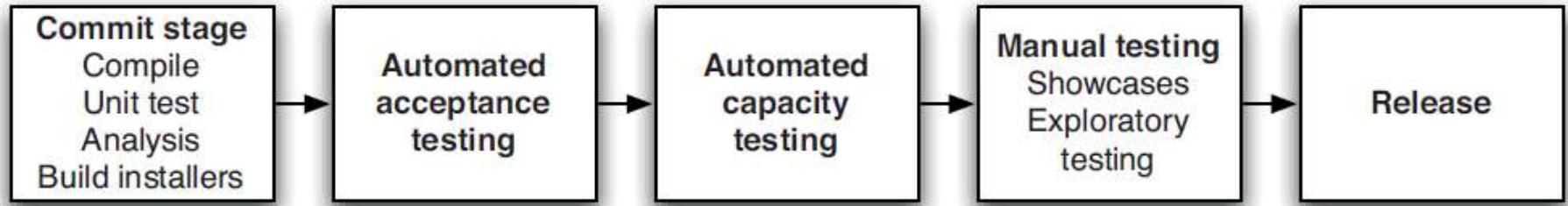


Figure 11.6 *A simple pipeline*

Tools for Software Configuration Management



- Software configuration management is normally supported by tools with different functionality.
- Examples:
 - RCS (very old but still in use; only version control system)
 - CVS (Concurrent Version Control)
 - based on RCS, allows concurrent working without locking
 - <http://www.cvshome.org/>
 - CVSWeb: Web Frontend to CVS
 - Perforce
 - Repository server; keeps track of developer's activities
 - <http://www.perforce.com>
 - ClearCase
 - Multiple servers, process modeling, policy check mechanisms
 - <http://www.rational.com/products/clearcase/>

Tasks for the Configuration Managers



SCMP following the IEEE 828-1990 standard

Define configuration items

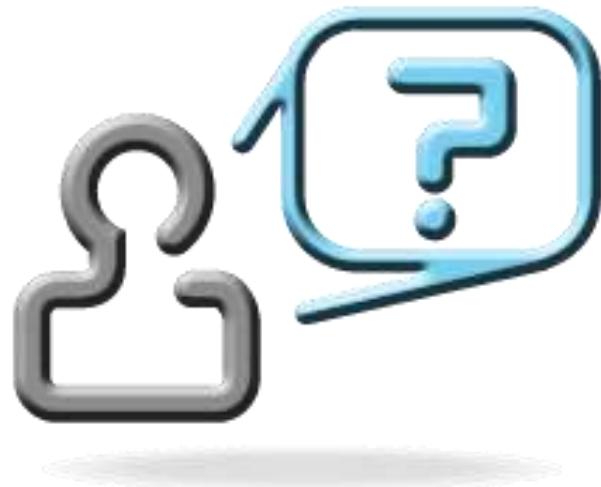
Define promote /release policies

Define activities and responsibilities

Set up configuration management system

References

- Readings used for this lecture
 - [Bruegge-Dutoit] Chapter 13 Configuration Management
 - [IEEE Std 828] Software Configuration Management
 - [IEEE Std 1042] Guide to Configuration Management Plan (SCMP)
- Additional References
 - CVS
 - Homepage: <http://www.cvshome.org/>
 - Online Documentation: <http://www.cvshome.org/docs/manual/cvs.html>
 - Jikes: Open Source Java Compiler maintained with CVS
 - Source tree (read only): <http://sourcery.org/jikes/anoncvs.html>
 - Jikes project portal <http://sourcery.org/jikes>
 - CVSWEB example
 - <http://stud.fh-heilbronn.de/~zeller/cgi/cvsweb.cgi/>





Thank You!

Baseline Management

- What baselines are required to be defined and managed?
 - Typically aligned with major milestones
 - Applies to documents as well as code
- How is the current software configuration defined?
 - A snapshot of everything the product has produced at some point in time

Baseline Management (cont'd)

- Who must approve changes to baselines?
 - Usually the Change Control Board (CCB)
- How and when are baselines created and physically controlled?
 - Through the use of document control systems, code control tools, and procedures to prevent the making of unapproved changes

Baseline Management (cont'd)

- How are people informed of changes?
 - The CCB disseminates change information
- How are baselines verified?
 - By reviews, inspections, and the testing of code
- Are baselines tied to project milestones?
 - Many are, but during coding many may not be

Baseline Management (cont'd)

- What information is required to process a change to a baseline?
 - A description of the proposed changes
 - Reasons for making the changes
 - List of other items affected by the changes
- What tools, resources, and training are required to perform baseline change assessment?
 - File comparison tools to identify changes
 - Resources and training depend on size and complexity of project

Baseline Management (cont'd)

- What metrics should be used to assess changes to a baseline?
 - Complexity
 - Average module size
 - Number of modules changed
 - Number of bugs fixed and verified
 - Code coverage

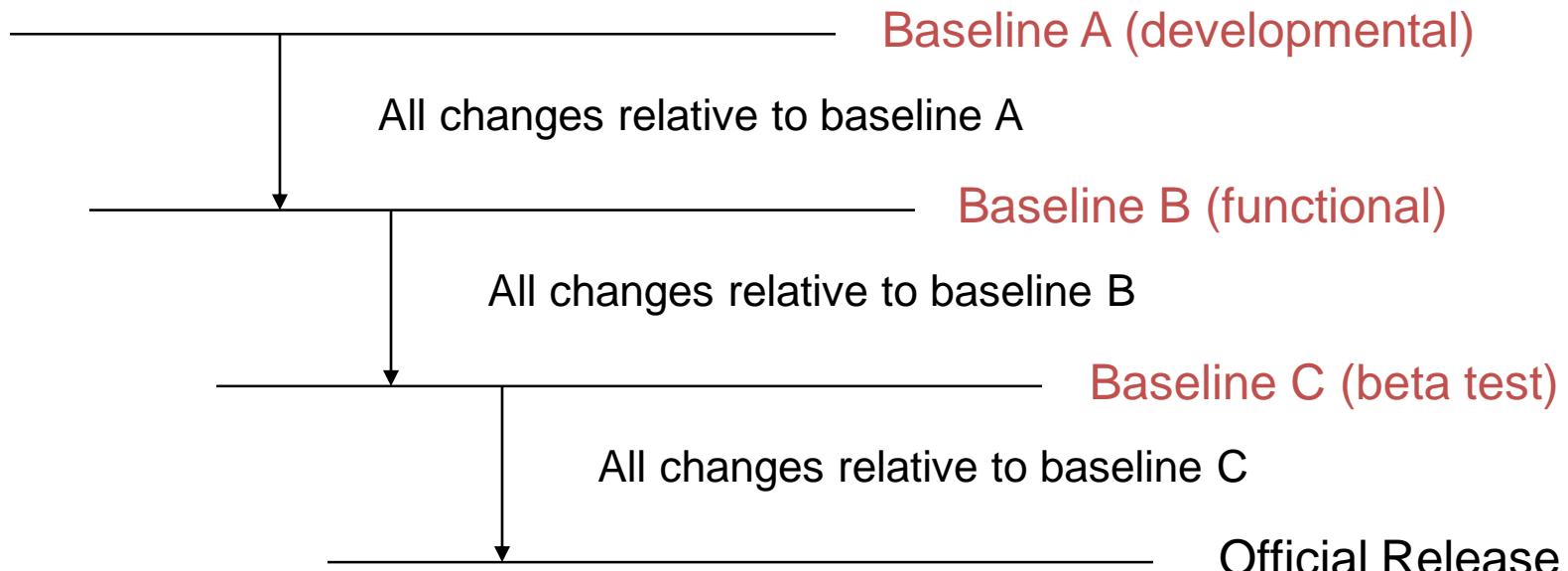
Baseline Management (cont'd)

- How are unauthorized changes to source code prevented, detected, and corrected?
 - No way to prevent unauthorized changes
 - Provide software engineers with training
 - A commercial available SCM systems provide adequate protection
 - Unauthorized changes should be caught during assessment procedures

Baseline Management (cont'd)

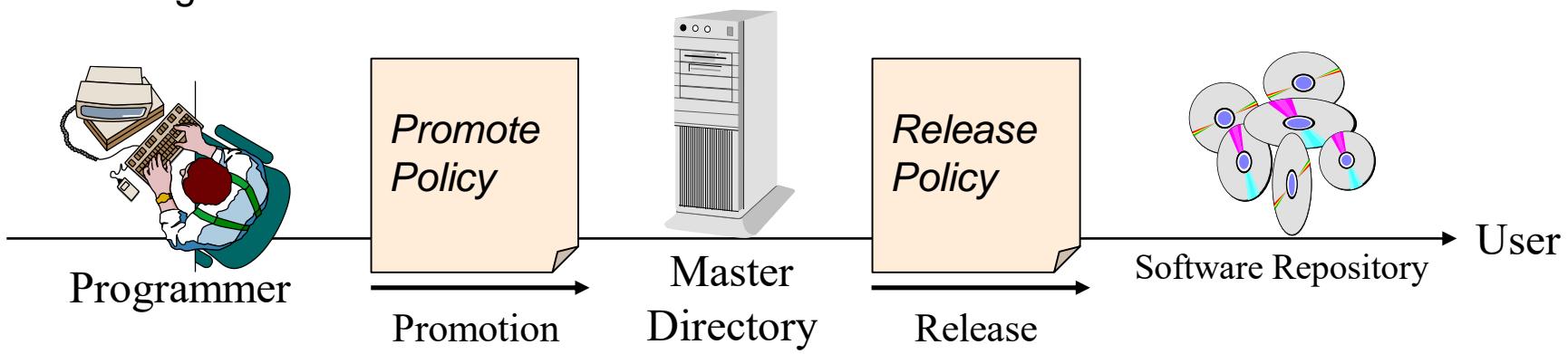
- What tools, resources, and training are required to perform baseline management?
 - A fully featured SCM tool
 - On large projects a separate SCM group may be needed
 - SCM training is required for all involved in the process

Baselines in SCM



Controlling Changes

- Two types of controlling change:
 - *Promotion*: The internal development state of a software is changed.
 - *Release*: A set of promotions is distributed outside the development organization.



- Approaches for controlling change to libraries (Change Policy)
 - Informal (good for research type environments)
 - Formal approach (good for externally developed CIs and for releases)

Change Policies

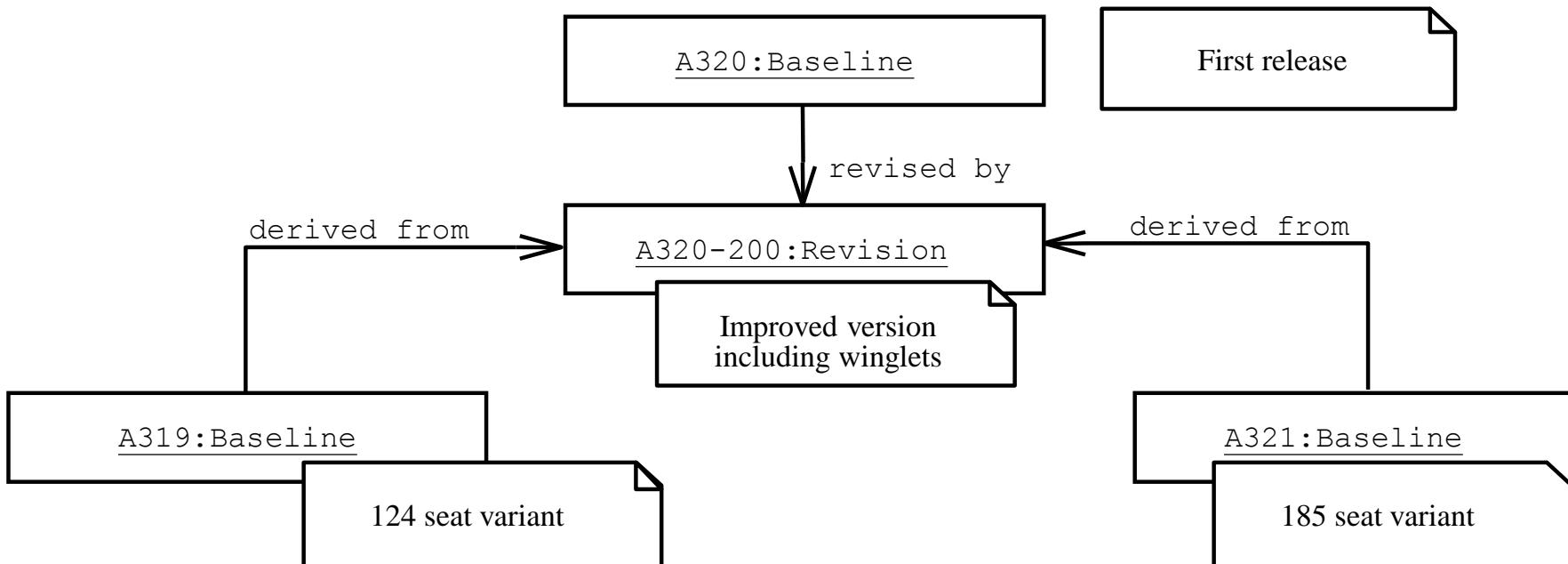
- Whenever a promotion or a release is performed, one or more policies apply. The purpose of change policies is to guarantee that each version, revision or release (see next slide) conforms to commonly accepted criteria.
- Examples for change policies:
 - “No developer is allowed to promote source code which cannot be compiled without errors and warnings.”
 - “No baseline can be released without having been beta-tested by at least 500 external persons.”

Version vs. Revision vs. Release

- Version:
 - An *initial* release or re-release of a configuration item associated with a *complete compilation* or recompilation of the item. Different versions have different functionality.
- Revision:
 - *Change to a version that corrects a documented functional error, does not affect the documented functional requirements.*
- Release:
 - The *formal distribution* of an approved version.

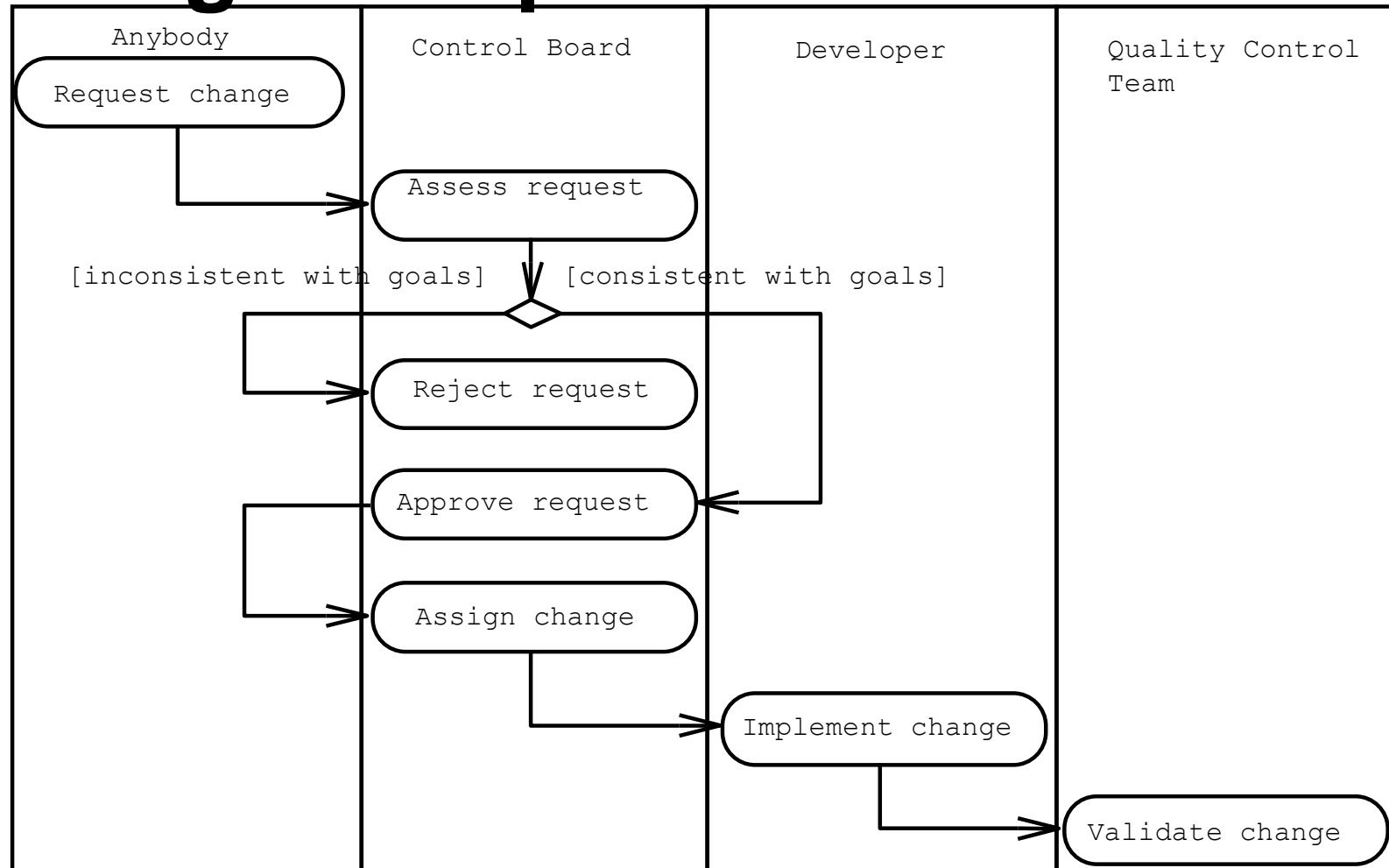
Quiz: Is Windows98 a new version or a new revision compared to Windows95 ?

Examples of baselines, revisions, and variants.



The A319, A320, and the A321 are all based on the same design. They vary mostly by the length of their fuselage.

An example of change management process





CSIZG514/SEZG514, Introduction to DevOps

Lecture No. 15 and 16

Agenda

On-demand Infrastructure Management

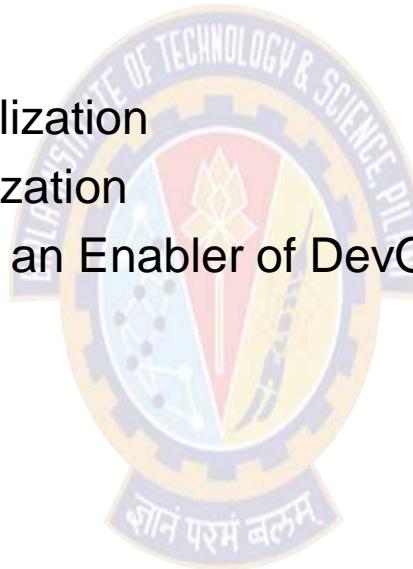
- Introduction to On-demand Infrastructure Management
- On-Demand Infrastructure Architect
- Security still a Major Concern?
- Service Providers
- Benefits of On-demand Infrastructure
- Auto Scaling
- Auto Scaling Approach
- Auto Scaling Providers



Agenda

Virtualization

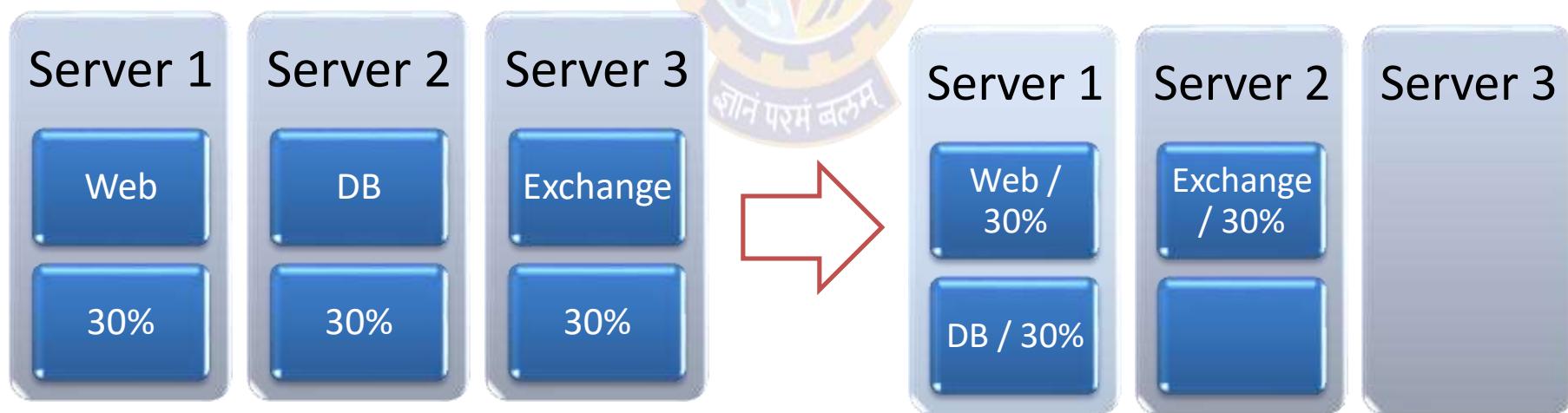
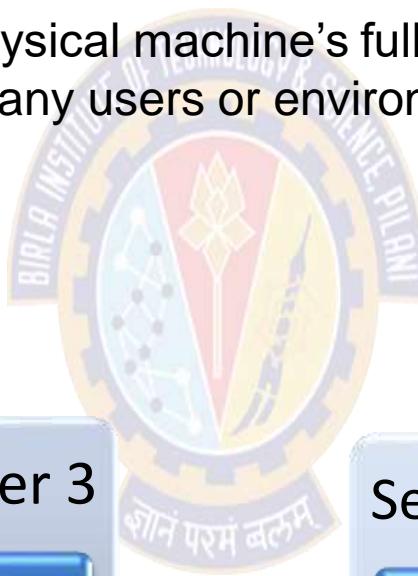
- Introduction to Virtualization
- Virtualization History
- Types of Virtualization
- Hypervisor based Virtualization
- Container based Virtualization
- Service Virtualization as an Enabler of DevOps



Virtualization

Introduction to Virtualization

- Virtualization is technology that lets you create useful IT services using resources that are traditionally bound to hardware
 - It allows you to use a physical machine's full capacity by distributing its capabilities among many users or environments



Virtualization

History

- While virtualization technology can be sourced back to the 1960s, it wasn't widely adopted until the early 2000s
- Fast forward to the 1990s. Most enterprises had physical servers and single-vendor IT stacks, which didn't allow legacy apps to run on a different vendor's hardware



Virtualization

Types of Virtualization

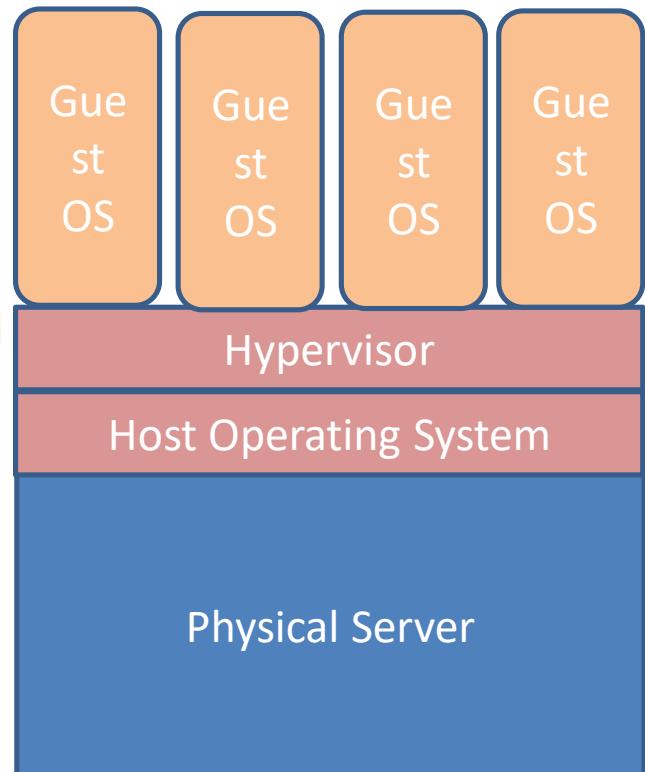
- Data virtualization
- Desktop virtualization
- Server virtualization
- Network functions virtualization



Virtualization

Hypervisor based Virtualization

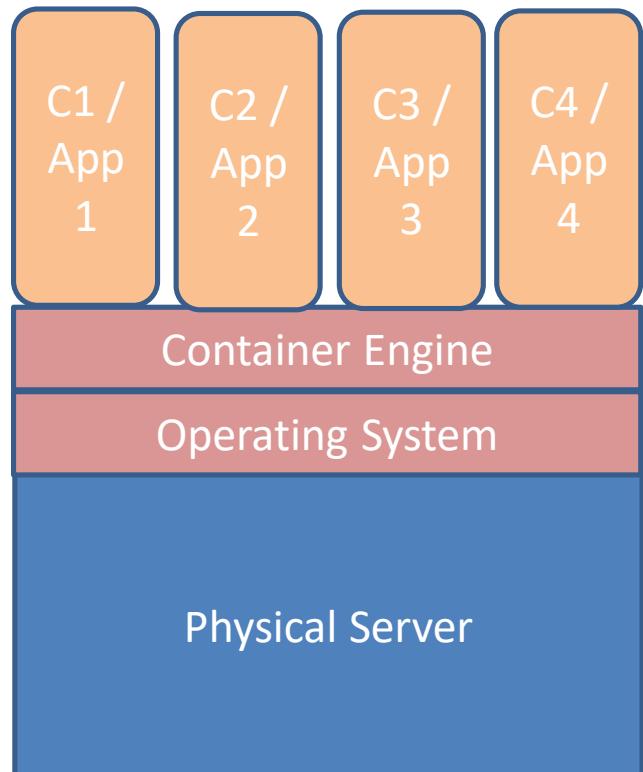
- Benefits:
 - More cost effective
 - Easy to scale
- Limitations:
 - Kernel Resource Duplication
 - Application portability issue



Virtualization

Container based Virtualization

- Benefits:
 - More cost effective
 - Faster deployment speed
 - Great portability



Virtualization

Service Virtualization as an Enabler of DevOps

- DevOps engagements have the primary goal of accelerating the time to market
- Service virtualization and DevOps
 - Service virtualization is the process of simulating the behavior of selected components within an application to enable end-to-end delivery of the application as a whole
 - Application development teams can use virtual services in lieu of the production or real services to conduct integration testing earlier in the development process
 - Virtualized assets look and act like the real thing but may be duplicated and available at times when the real assets are not available

Containerization

Introduction to Docker

- Docker is a platform for developers and sysadmins to develop, deploy, and run applications with containers
- The use of Linux containers to deploy applications is called containerization
- Containerization is increasingly popular because containers are :
 - Flexible: Even the most complex applications can be containerized
 - Lightweight: Containers leverage and share the host kernel
 - Interchangeable: Can deploy updates and upgrades on-the-fly
 - Portable: Can build locally, deploy to the cloud, and run anywhere
 - Scalable: Can increase and automatically distribute container replicas
 - Stackable: Can stack services vertically and on-the-fly

Containerization

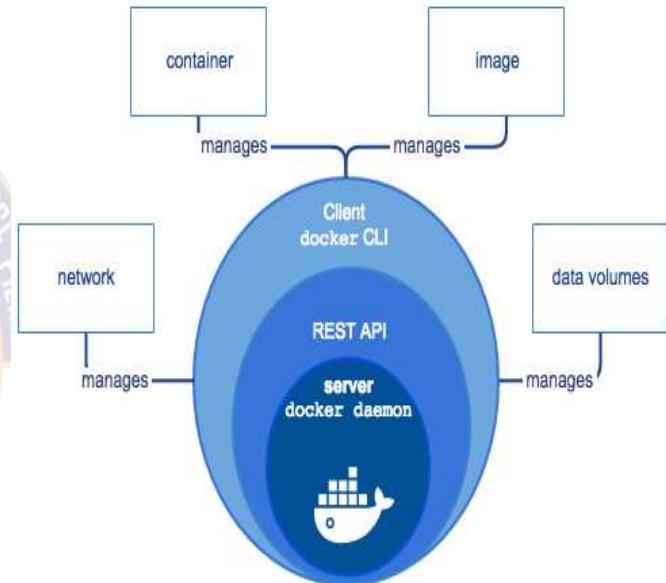
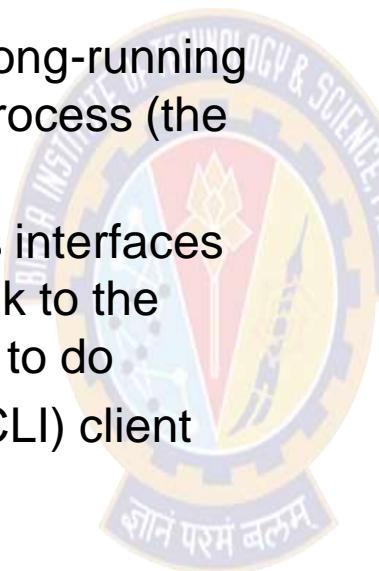
Docker Overview

- Docker is an open platform for developing, shipping, and running applications
- Docker is licensed under the open source Apache 2.0 license
- Docker enables to separate applications from infrastructure so we can deliver software quickly
- With Docker, we can manage our infrastructure in the same ways we manage our applications
- By taking advantage of Docker's methodologies for shipping, testing, and deploying code quickly, we can significantly reduce the delay between writing code and running it in production

Containerization

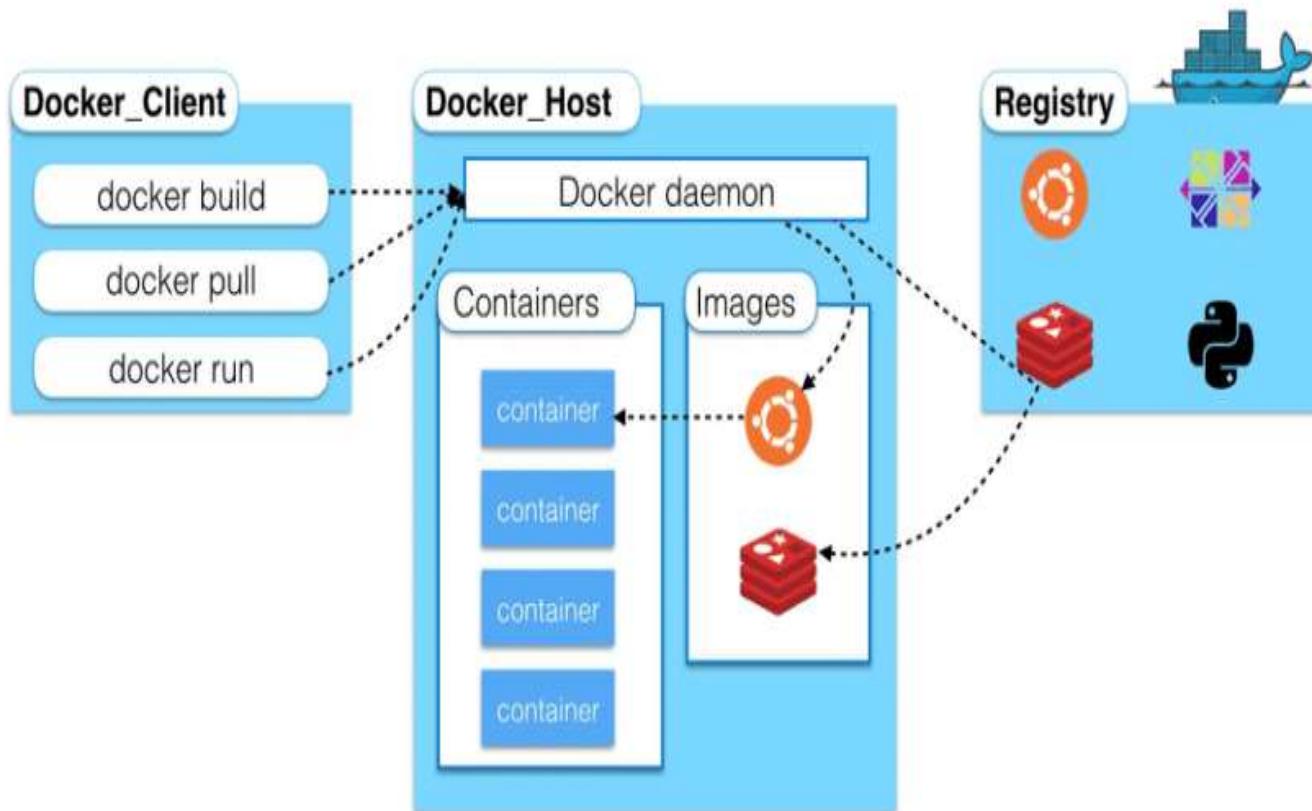
Docker Engine

- Docker Engine is a client-server application
- A server which is a type of long-running program called a daemon process (the dockerd command)
- A REST API which specifies interfaces that programs can use to talk to the daemon and instruct it what to do
- A command line interface (CLI) client (the docker command)



Containerization

Docker Architecture



Containerization

Docker Concept : Swarm and Stack

- Docker Swarm:
 - A swarm is a group of machines that are running Docker and joined into a cluster
 - Docker Command then will be executed on a cluster by a swarm manager
 - The machines in a swarm can be physical or virtual
 - After joining a swarm, machines are referred to as nodes
- Docker Stack:
 - A stack is a group of interrelated services that share dependencies, and can be orchestrated and scaled together
 - A single stack is capable of defining and coordinating the functionality of an entire application

Containerization

Docker File

- Docker file defines what goes on in the environment inside a container
- This file has information about all resources
 - Network interface and
 - Disk drives



***Bits Pilani Software installation
Docker file:***

FROM centos:latest

MAINTAINER "Bits Pilani"

RUN yum install -y bits-pilani-client

EXPOSE 9999

Containerization

Docker Cloud

- Docker Cloud provides a hosted registry service with build and testing facilities for Dockerized application images
- Access to Docker Cloud is managed by Docker ID
- Manage Builds and Images
- Manage Swarms
- Manage Infrastructure
- Manage Nodes and Apps
- Integration with AWS and Microsoft Azure Services



Containerization

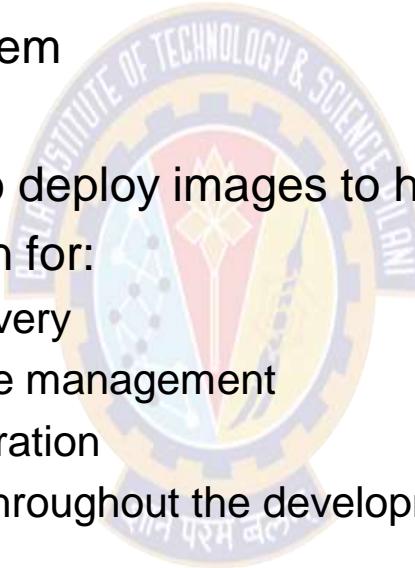
Docker Store

- For developers and operators, Docker Store is the best way to discover high-quality Docker content
- Independent Software Vendors (ISVs) can utilize Docker Store to distribute and sell their Dockerized content
- Store Experience:
- Access to Docker's large and growing customer-base
 - Customers can try or buy your software
 - Use of Docker licensing support
 - Docker handle checkout
 - Seamless updates and upgrades for your customers
 - Become Docker Certified

Containerization

Docker Hub

- Docker Hub is a cloud-based registry service
- Allows link to code repositories
- Build images and test them
- Stores Images
- Links to Docker Cloud to deploy images to hosts
- It is a Centralize Solution for:
 - Container image discovery
 - Distribution and change management
 - User and team collaboration
 - Workflow automation throughout the development pipeline



Containerization

How is Docker Store different from Docker Hub

Docker Hub	Docker Store
Contents by Docker community	Contents submitted for approval by qualified Store Vendor Partners
Anyone can push new images to the community	Contents are published and maintained by Entity
No guarantees around the quality or compatibility	Docker Certified quality assurance
<i>Common Thing: The Docker official Images are Available in both</i>	

Containerization

Overview of Docker Editions

- Docker is available in two editions:
 - Community Edition (CE)
 - Enterprise Edition (EE)
- Docker Community Edition has three types
 - Stable
 - Test
 - Nightly



References

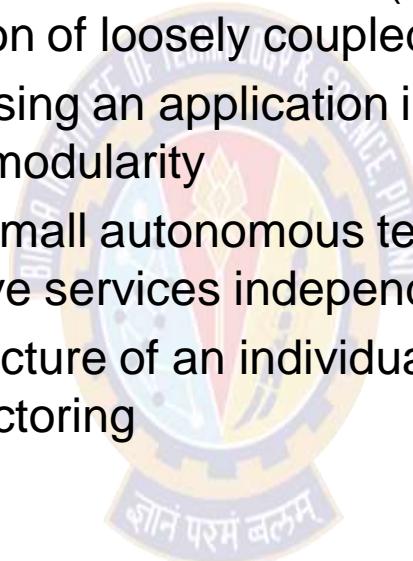
- Docker Documentation [<https://docs.docker.com/>]



Microservices

Introduction to Microservices

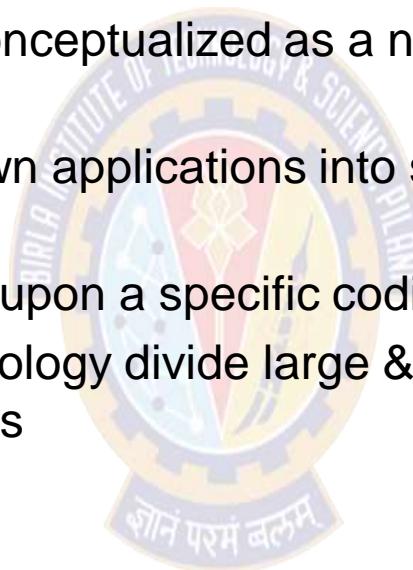
- Microservices are a software development technique
- A type of the service-oriented architecture (SOA); that structures an application as a collection of loosely coupled services
- The benefit of decomposing an application into different smaller services is; it improves modularity
- Microservices enables small autonomous teams to develop, deploy and scale their respective services independently
- It also allows the architecture of an individual service to emerge through continuous refactoring



Microservices

Introduction to Microservices Contd..

- Microservices and their related architectures are gaining power in enterprises of all sizes
- Microservices can be conceptualized as a new way to create corporate applications
- The Idea is to break down applications into smaller, independent services
- That are not dependent upon a specific coding language
- Using Microservices Ideology divide large & complex applications in to smaller building blocks



Microservices

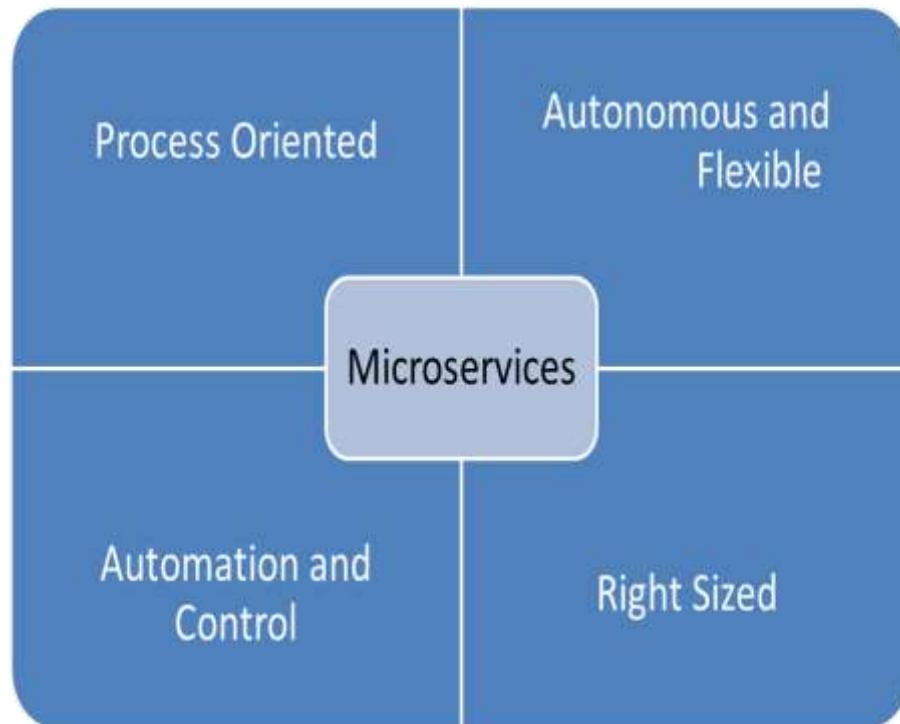
List of Organizations who are using Microservices

- Uber
- Netflix
- Amazon
- Ebay
- Gilt
- Tesla



Microservices

Common characteristics for Microservices



Microservices

Need of Microservices for DevOps

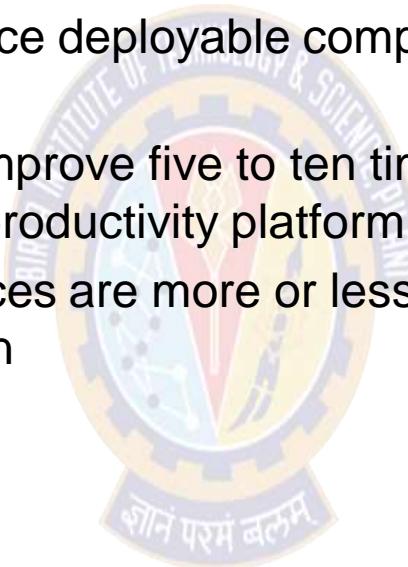
- DevOps promotes small, more empowered and autonomous teams, along with automation and measurements
- It has great potential to be a fantastic improvement in the way IT and business work together
- Struggling Area:
- Setup Pipeline for new automation tooling
- The mindset of Architects
- Flow we followed to Emerge:
 - Centralized IT Components and Deliverables – [2005]
 - Cloud and DevOps [2010 – 2013]



Microservices

DevOps and Microservices

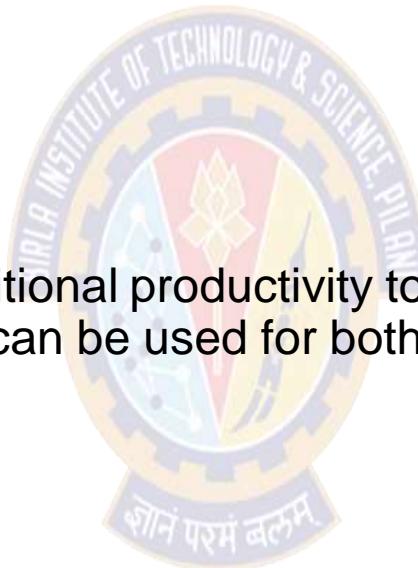
- With small autonomous DevOps teams, the world will start producing small autonomous components, called microservices
- It makes sense to produce deployable components for a business function
- This will help in speed improve five to ten times by moving to the cloud and using a high productivity platform
- DevOps and microservices are more or less inseparable, they can hardly exist in separation



Microservices

Benefits of DevOps and Microservices

- Deployability
- Reliability
- Availability
- Scalability
- Modifiability
- Management
- Microservices bring additional productivity to DevOps by embracing a common toolset, which can be used for both development and operations



Reference

- <https://www.mendix.com/blog/the-microservices-you-need-for-devops/>
- <https://www.mulesoft.com/resources/api/what-are-microservices>



AWS Lambda

What is AWS Lambda?

- AWS Lambda is a compute service that lets you run code without provisioning or managing servers
- AWS Lambda executes your code only when needed and scales automatically
- You pay only for the compute time you consume - there is no charge when your code is not running
- Often referred to as “serverless” computing. AWS Lambda allows developers to think only about the code, and not have to worry about the layers below
- You can think of PaaS service to co-relate

AWS Lambda

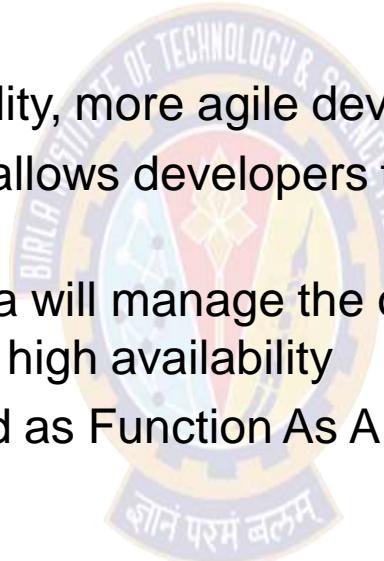
AWS Lambda : Application Support

- Can any application run in Lambda?
- Some applications may need to be modified to run with Lambda
- Serverless is a specific architecture where state and compute are decoupled
- Lambda functions can write state to external data stores via web requests
- External Data store can be of AWS or Any
- AWS Solutions: S3, Dynamo, and Redshift
- Other Solutions: PostgreSQL, Cassandra and Kafka

AWS Lambda

Function as a Service [FaaS]

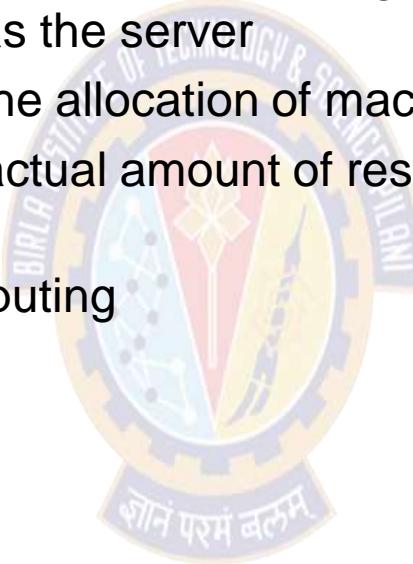
- Microservices and AWS Lambda
- Microservices where applications are broken into much Smaller pieces
- To enable better scalability, more agile development
- Similarly AWS Lambda allows developers to deploy a single function at a time
- Once Deployed, Lambda will manage the operational aspects of that function, from scaling to high availability
- AWS Lambda is referred as Function As A Service



AWS Lambda

Serverless?

- Its an Architecture
- Serverless computing is a cloud-computing execution model in which the cloud provider acts as the server
- Dynamically managing the allocation of machine resources
- Pricing is based on the actual amount of resources consumed by an application
- It is a form of utility computing



AWS Lambda

Success Story of The Seattle Times

- The Challenge:
 - After maintaining on-premises hardware and custom publishing software for nearly two decades, The Seattle Times sought to migrate its website publishing to a contemporary content management platform.
 - To avoid the costs of acquiring and configuring new hardware infrastructure and the required staff to maintain it, the company initially chose a fully managed hosting vendor.
 - But after several months, The Times' software engineering team found it had sacrificed flexibility and agility in exchange for less maintenance responsibility.
 - As the hosted platform struggled with managing traffic under a vastly fluctuating load, The Seattle Times team was constrained in its ability to scale up to meet customer demand.

AWS Lambda

Success Story of The Seattle Times Contd..

- Why Amazon Web Services:
 - To address these core scalability concerns, The Seattle Times engineering team considered several alternative hosting options, including self-hosting on premises, more flexible managed hosting options, and various cloud providers.
 - The team concluded that the available cloud options provided the needed flexibility, appropriate architecture, and desired cost savings. The company ultimately chose Amazon Web Services (AWS), in part because of the maturity of the product offering and, most significantly, the auto-scaling capabilities built into the service.
 - The Seattle Times deployed its new system in just six hours. The website moved to the AWS platform between 11 p.m. and 3 a.m. and final testing was completed by 5 a.m. — in time for the next news day.

AWS Lambda

Success Story of The Seattle Times Contd..

- End Result:
 - With AWS, The Seattle Times can now automatically scale up very rapidly to accommodate spikes in website traffic when big stories break, and scale down during slower traffic periods to reduce costs.
 - “Auto-scaling is really the clincher to this,” Seattle Times says. “With AWS, we can now serve our online readers with speed and efficiency, scaling to meet demand and delivering a better reader experience.”
 - “AWS Lambda provides us with extremely fast image resizing,” Seattle Times says. “Before, if we needed an image resized in 10 different sizes, it

Reference Architecture: [Sample Code](#)



AWS Lambda

Benefits Summary

- No servers to Manage
- Continuous Scaling
- Subsecond Metering



Reference

- <https://aws.amazon.com/lambda>





Thank You!

In our next session: