



ONNX

# Architecture & Infra SIG Update

Liqun Fu, Microsoft

# Shape Inference

- Improved at operator level:
  - QuantizeLinear
  - Transpose
  - ConstantOfShape
  - Expand
  - Gather and Shape
- InferShapes extended to expose result of data propagation
- Shape inference for local functions added in Release 1.11.0 and further enhanced in the Latest 1.12.0 release.

# Node Test Cases

- Node test case generation are now deterministic and flexible
  - since\_version is used by default
  - users can choose an opset version
- All test cases are updated with the default opset version.
- As a result of this deterministic measure, we are able to validate data generation in CI pipelines.

# CI Pipeline and Packaging

- Linux release pipeline upgraded to manylinux2014
- Dropped support for x86 32bit Linux
- Release packages now include Python 3.10
- No longer produce Python 3.6 packages

# New Utilities

- Model composer to help with creating and combining models out of several graphs or models
- FunctionBuilder to help construct function ops

# Many Miscellaneous Improvements

- Initializer decoupled from input
- Version converter becomes more stable with multiple bug fixes
- Python code modernized with type annotation
- Parser improved to support richer grammar
- Shape inference becomes much reliable
- Tons of bug fixes and document updates
- And more...

# What Next?

- Better support for quarterly releases
  - Early detection of breaking changes
  - Automatic integration test with onnxruntime
- Continue to improve shape inference
- Provide support for other SIGs

Please Get Involved!

Github: PRs, Issues, and  
Discussions

Slack channel:  
<https://slack.lfai.foundation> and join  
onnx-archinfra

Monthly SIG meetings (see slack  
channel for announcements)

Thanks for coming!!!